May 20, 2022

Participation and Multifamily Census Study (RI-21-RX-Participation)

National Grid (Rhode Island)

Developed For

National Grid
40 Sylvan Road
Waltham, MA 02451

Developed By

Cadeo Group
107 SE Washington Street, Suite 450
Portland, OR 97214

Contributors

Doug Bruchs, Jun Suzuki, Eric Dimperio, Jess Siegel, and Lima Hossain

Please refer questions to:

Doug Bruchs, dbruchs@cadeogroup.com 503.660.4213

Executive Summary

National Grid contracted with Cadeo to complete the Participation and Multifamily Census Study (RI-21-RX-Participation) in its Rhode Island territory. This study had two objectives:

1. Analyze historical participation in National Grid's Rhode Island residential energy efficiency program between 2016 and 2020
2. Create a comprehensive database of multifamily (MF) buildings in Rhode Island with 5+ units to support delivery of National Grid's EnergyWise and Income Eligible Multifamily programs

To meet both objectives, Cadeo began by creating an account-level dataset that combined National Grid's active customer list (as of March 2021), associated energy use data, historical program tracking data (2016–2020) with census block group data from the American Community Survey and third-party customer and tax parcel data. We used the resulting dataset to analyze historical program trends, model the drivers of participation, develop propensity scores for nonparticipating customers, and identify multifamily customers.

Because our analysis focused on assessing and modeling customer-specific (i.e., account and/or building-level) participation over time, this study covers the residential programs in Rhode Island for which National Grid collected customer-specific tracking data. For this reason, the customer-level analyses in this report exclude upstream programs, most notably Residential Lighting and National Grid's Home Energy Report program, because of the behavioral program's random customer selection and opt-out process. While these programs contribute significantly to National Grid's total residential portfolio savings in the state, they are not relevant for understanding what drives customer-level participation or determining the propensity that any given nonparticipant will participate in a future National Grid opt-in energy efficiency program.

Assessing Historical Program Participation

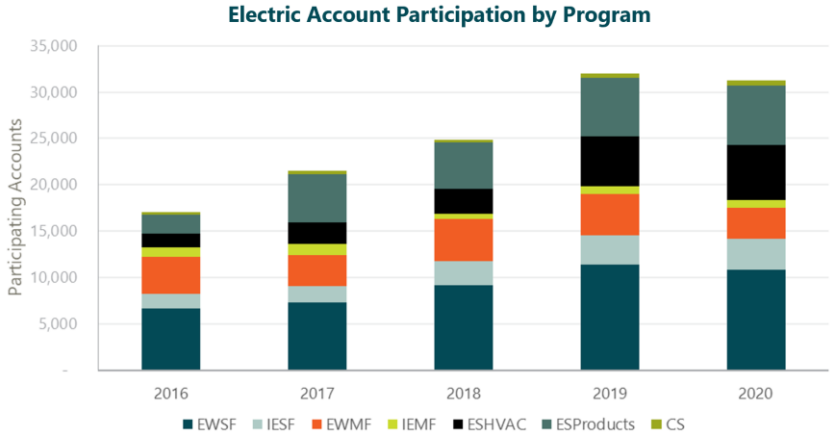
To assess cumulative participation in National Grid's residential energy efficiency programs, Cadeo developed two complementary, customer-specific metrics:

1. **Participation.** The percent of total eligible accounts or buildings that have participated in each program.
2. **Savings-to-consumption (SC).** The percent of total account or building energy consumption saved due to program participation.

Electric Accounts

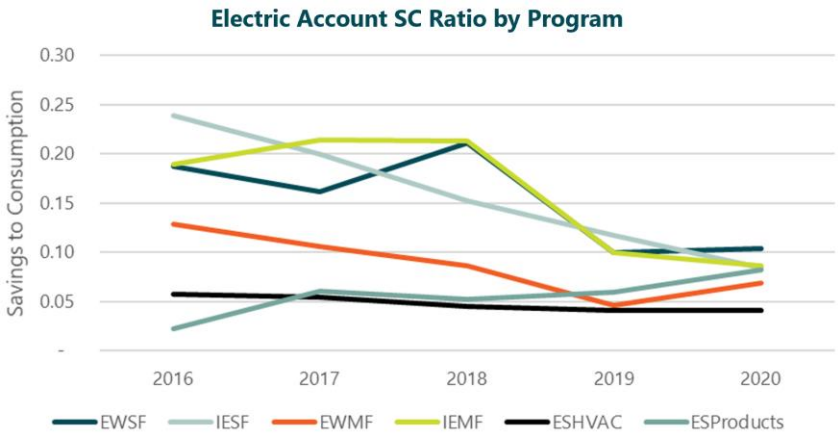
Participation

We found that, at the portfolio-level, electric account participation increased steadily from 2016 until 2019, nearly doubling over the four-year period. Unsurprisingly, participation dropped in 2020, although only slightly and likely because of COVID-19.



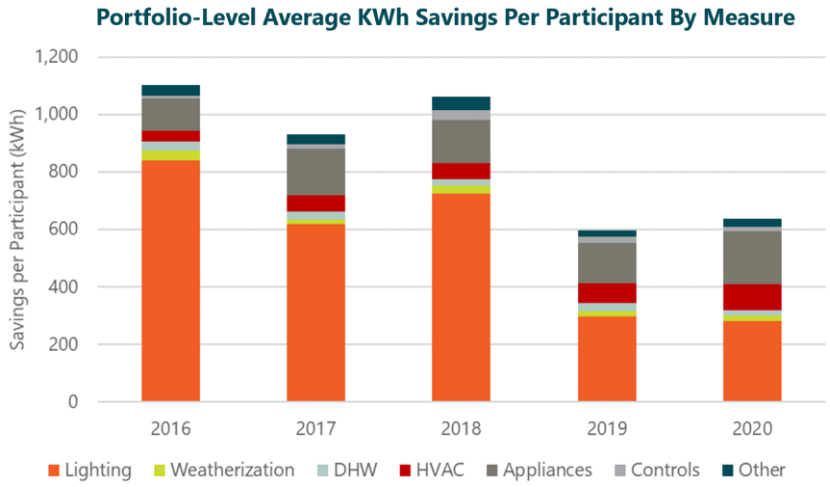
Savings-to-consumption ratio

However, all the programs, except for ENERGY STAR Products (ESProducts), showed a declining trend in SC ratios during the same period. This means that while larger numbers of customers participated each year, the average program savings generated by each participant (relative to their consumption, which was relatively consistent across the five-year analysis window) declined.



Per Participant Savings by Measure

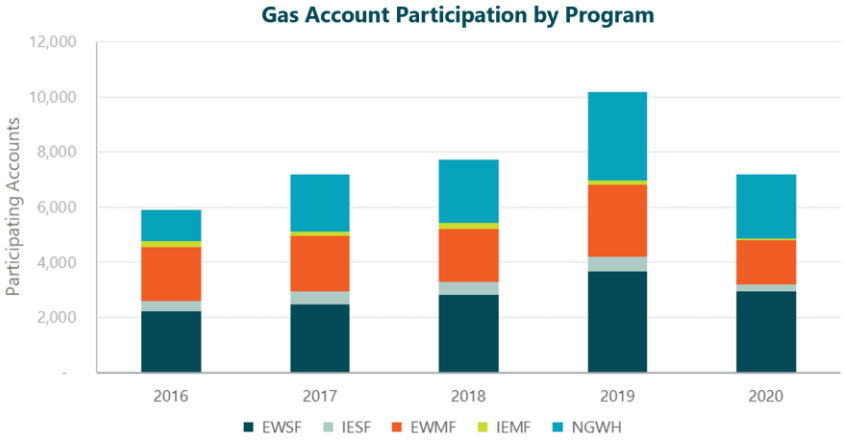
By taking a closer look at measure-level savings, we reveal that the downward trend in SC ratios is largely attributable to reduced average savings from direct install lighting measures, while the contributions of other measures stayed almost constant. We looked closer at reduction in lighting savings over time and found that, for single family programs, the average number of lighting measures per participant (28 to 16 lamps) and the average savings per lighting measure (48 to 35 kWh) dropped from 2016 to 2020.



Gas Accounts

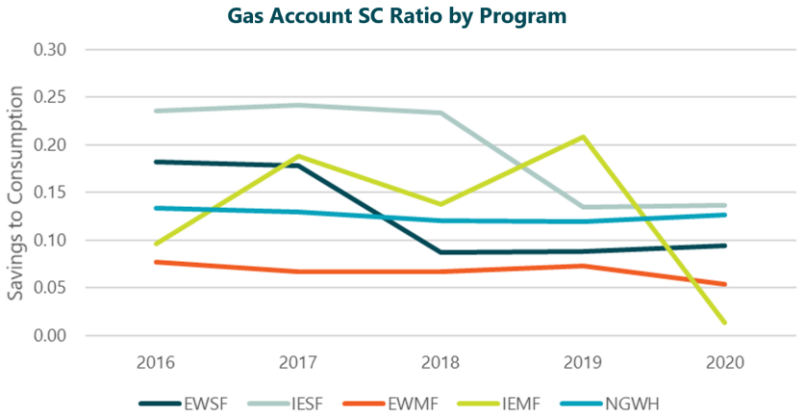
Participation

Portfolio-level gas account participation exhibited similar trends as electric: increasing account-level annual participation, especially for EnergyWise Single Family (EWSF), EnergyWise Multifamily (EWMF), and Natural Gas Heating and Water Heating (NGWH). All the gas programs also showed a similar dip in participation in 2020. Like the drop in participation for electric programs in 2020, the decline is likely due to COVID-19.



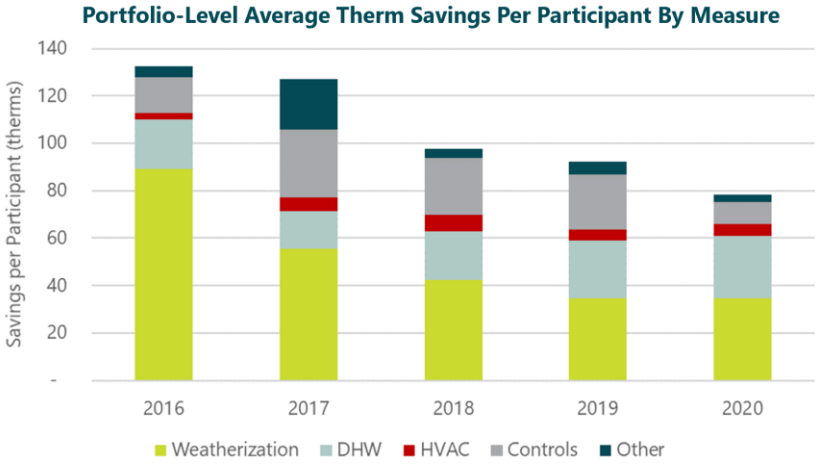
Savings-to-consumption ratio

Gas programs also showed an overall declining trend in SC ratios. SC ratios of EWSF and Income Eligible Services - Single-Family (IESF) dropped nearly 50% during this period, while other large contributors to participation such as EWMF and NGWH stayed almost constant. COVID-19 likely affected Income Eligible Services – Multifamily (IEMF)'s sharp decline in 2020.



Portfolio-level average therm savings per participant by measure

A measure-level look reveals that steadily lowering average savings from weatherization measures drove the decline during this period. When looking closer at reduction in weatherization savings over time, we found that savings per weatherized customer was steady across the five years, but the number of customers weatherizing their home was much higher in 2016 than any of the other years.



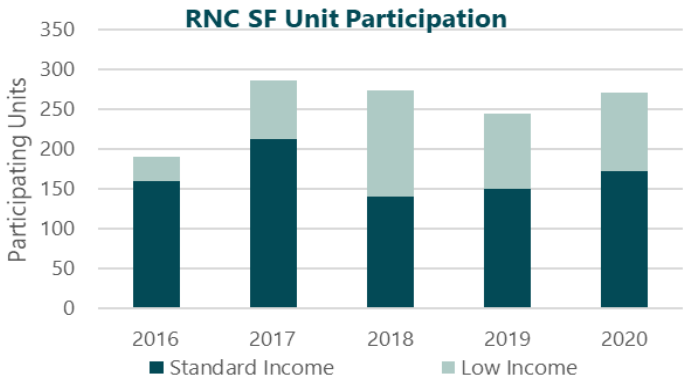
Residential New Construction

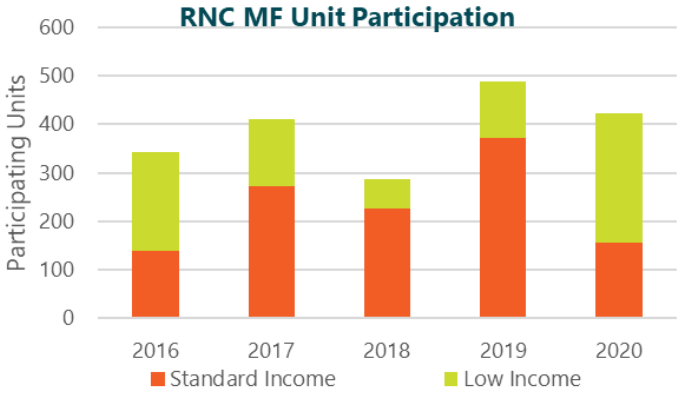
Participation

We assessed residential new construction (RNC) separately since, unlike the other retrofit programs, it focuses on improving the efficiency of new buildings.

Among single-family new construction units where market data is available, the program reached 20% of the new construction single-family units in Rhode Island.

We found that overall RNC participation peaked in 2019, but no clear annual trend in participation between 2016 and 2020. Across all five years, low-income housing units made up 34% of single-family homes and 40% of multifamily homes. Among these low-income units the program reached, 60% were multifamily.





Modeling Drivers of Participation

To understand what customer and building features drive participation, Cadeo developed a logistic model that explains how each feature is associated with participation. We also developed a complementary linear regression model that similarly examined how these same factors influence participants' SC ratio (i.e., the magnitude of program savings relative to their consumption). We included features and interactions between features that proved significant at explaining participation trends in our final specification, as well as key equity features such as income, homeownership, and primary language. We found:

	Participation	SC Ratio
Strong Drivers	**Household Income.** Greater income is associated with higher participation. **Age of head of household.** Older households tend to participate more.	**Living Area.** Larger homes are associated with greater electric savings, but less gas savings. **Total units in building.** More units in multifamily buildings were associated with a higher degree of savings.
Weak Drivers	**Primary language.** The model did not find language to strongly influence participation, even though we see a clear increase in the participation rates for English speakers.	
Important Interactions	**Age x Income.** Increases in income have a stronger impact on younger customers than on older customers. **Homeownership (x Income & x Age).** Increases observed in income and age lead to greater participation for everyone, but the increase is more pronounced among homeowners as compared to renters.	

Nonparticipants' Propensity to Participate

To obtain insight into which nonparticipants were more likely to participate in the future—ultimately to support the sampling of nonparticipant survey of the concurrent Nonparticipant Market Barriers Study—we developed a predictive model to score nonparticipants on how similar they are to customers who have participated previously. We separated the range of scores into low, medium, and high based on how well the score differentiated participants from nonparticipants.

Score	Nonparticipating Electric and Gas Accounts	Implication
Low scores (0.0–0.3) are associated with accounts that have features less commonly observed in participants.	178,762 (**56%**)	More than half of nonparticipating accounts look very different than past participants and are the least likely to participate in programs as designed, marketed, and delivered today.
Middle scores (0.3–0.6) are assigned to accounts that are equally similar to participants and nonparticipants.	78,322 (**25%**)	The reasons for nonparticipation are likely due to factors for which we do not have data.
High Scores (0.6–1.0) are most like the observed participants.	60,602 (**19%**)	Less than one-fifth of nonparticipating accounts are more likely to participate in a current program; these accounts may need more time or may face fewer barriers to participation.

Identifying Multifamily Customers

To develop a data-driven algorithm that could accurately classify each building—and associated electric and natural gas accounts—as either a single-family (SF, 1–4 units) or multifamily (MF, 5+ units), Cadeo tested a wide variety of algorithm specifications. We determined that the most accurate algorithm prioritized using the land use from tax parcel data, the number of unique gas or electric accounts associated with the building, and building-level gas consumption to identify multifamily buildings. When we assessed our final algorithm against the National Grid's historical multifamily program participation data, we found the algorithm identifies multifamily buildings with 85% accuracy. This performance was strong given limitation in the data. The previous National Grid Rhode Island Participation Study identified approximately 60% of multifamily participants without relying on participation data.

We applied the final algorithm to the aggregated dataset and estimated that there are **24,012 Multifamily buildings in Rhode Island.** This is approximately 7% of the residential buildings in the state, and 19% and 16% of National Grid's electric and natural gas residential accounts, respectively.

	Single-Family (≤4 units)		**Multifamily** (5+ units)	
	Count	**Percent**	**Count**	**Percent**
Buildings	321,178	93%	24,012	7%
Electric Accounts	360,328	81%	83,409	19%
Gas Accounts	207,470	84%	40,649	16%

We compared our list of identified multifamily buildings to 2016–2020 tracking data for the EnergyWise and Income-Eligible Multifamily programs. We found that at least one account associated with 2,288 different buildings (10% of all identified multifamily buildings) had participated in the EnergyWise program, while participants for the Income-Eligible program came from 639 (3%) different multifamily buildings. Customers from 6,052 (25%) multifamily buildings participated in any program.

Analysis Dashboard

In addition to this report, our team developed an interactive PowerBI dashboard to view results from the analysis, intended for use as an internal tool for National Grid and supporting contractors to market and deliver programs. The dashboard provides the ability to interactively explore in-depth participants and nonparticipants data on a geographic information system–assisted map by filtering the data by program and customer characteristics. The image below shows a use case scenario to obtain insights on EWMF program participants who are renters living in small multifamily buildings.

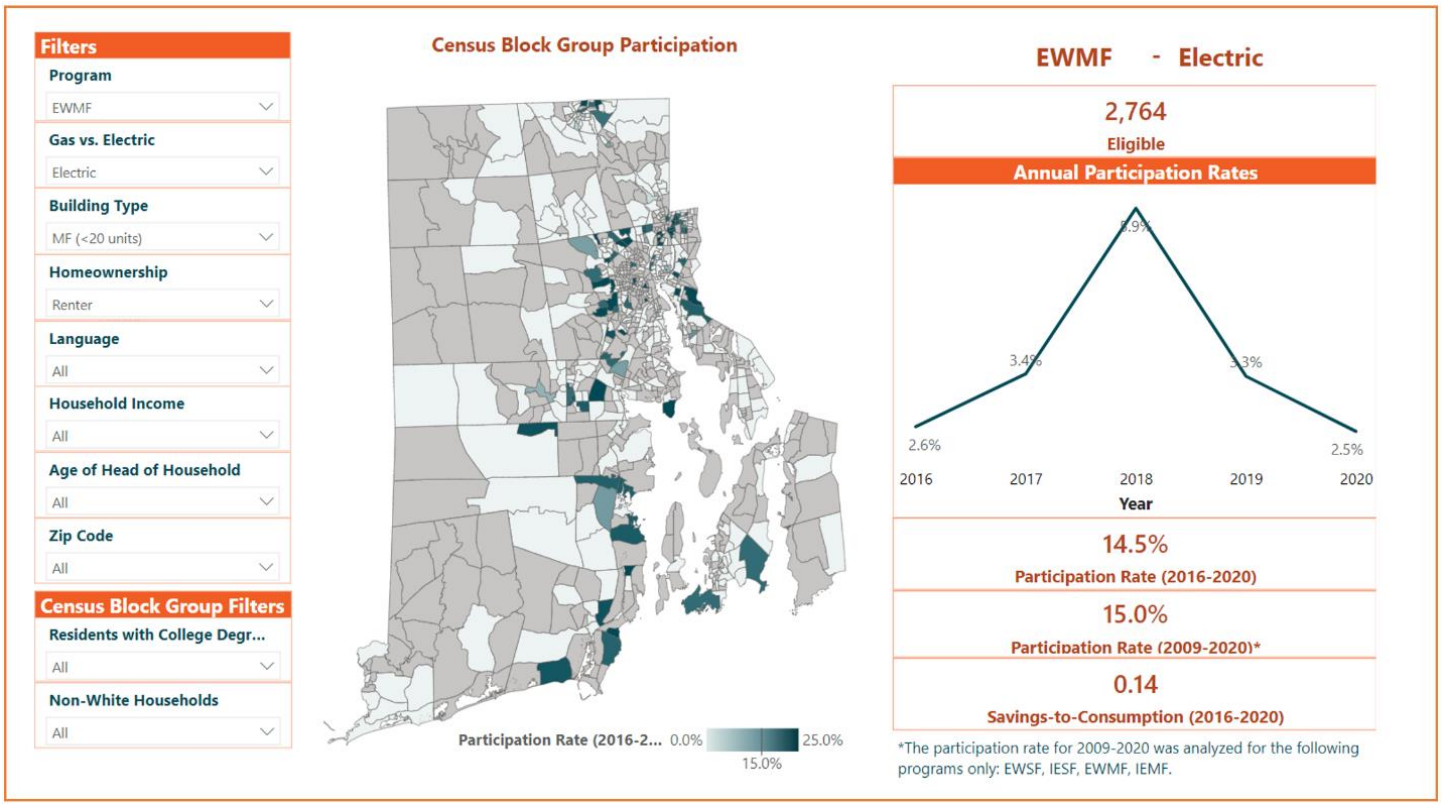

Filters
Program
EWMF
Gas vs. Electric
Electric
Building Type
MF (<20 units)
Homeownership
Renter
Language
All
Household Income
All
Age of Head of Household
All
Zip Code
All
Census Block Group Filters
Residents with College Degr...
All
Non-White Households
All
Census Block Group Participation
Participation Rate (2016-2... 0.0% 25.0%
15.0%
EWMF - Electric
2,764
Eligible
Annual Participation Rates
2.6%
3.4%
5.9%
3.3%
2.5%
2016
2017
2018
2019
2020
Year
14.5%
Participation Rate (2016-2020)
15.0%
Participation Rate (2009-2020)*
0.14
Savings-to-Consumption (2016-2020)
*The participation rate for 2009-2020 was analyzed for the following programs only: EWSF, IESF, EWMF, IEMF.

Table of Contents

Section 1 Introduction

1.1 Background

This study explores historical program participation in National Grid's residential energy efficiency programs in Rhode Island to a) determine cumulative participation rates and identify participation trends, and b) use those insights to inform future program design.

The study is built off a previous National Grid effort in Rhode Island, which documented historical participation in residential programs from 2009–2015,[1] as well as a recent study in National Grid's Massachusetts service territory that created a comprehensive list of multifamily (MF) buildings and analyzed historical program participation.[2] National Grid combined these previously separate efforts as part of this Participation and Multifamily Census Study.

National Grid also conducts an annual internal study of program participation in Rhode Island, (published in the Energy Efficiency Year-End Report). The Year-End Report and this study have similar scopes: both examine program participation. However, the Year-End Report only provides counts of participants by program.[3] This study goes significantly further by bringing together additional data sources such as billing data, census data, demographic, and property data. Our team used this aggregated dataset to undertake a series of modeling exercises aimed at understanding the drivers of participation and the likelihood that nonparticipants would participate in the future.

This study is a part of the portfolio evaluation of National Grid Rhode Island's (NGRI) residential energy efficiency programs. Another related study under the portfolio is Nonparticipant Market Barriers Study, which focuses on understanding participation barriers and strategies to recruit program participants who are historically underrepresented. This study supports the concurrent Nonparticipant Market Barriers Study by supplying a sample frame of nonparticipating customers to facilitate that study team's sampling efforts and data collection activities.

1.2 Study Objectives

This study has two complementary goals:

1. **Participation Analysis** assessed the portion of residential customers that have participated in National Grid's residential energy efficiency programs for which they are eligible between 2016 and 2020. We also completed a modeling analysis to understand

[1] Navigant Consulting, *Energy Efficiency Program Customer Participation Study, National Grid Rhode Island* (October 19, 2017), http://rieermc.ri.gov/wp-content/uploads/2018/03/national-grid-2017-ri-ee-customer-participation-study-final.pdf.
[2] Navigant Consulting, *Census of Massachusetts Multifamily Buildings (RES 43)* (May 31, 2018), https://ma-eeac.org/wp-content/uploads/RES43-Final-Report-2019-05-31.pdf.
[3] Appendix E provides a table comparing National Grid's previously reported Year-End Report's participation counts (by program and year) with our team's counts. As evident in the table, our totals were similar.

the distinguishing demographic and housing characteristics among program participants.

2. **Multifamily Census Study** developed a data-driven algorithm to identify residential customer accounts that are associated with multifamily buildings in Rhode Island, and then used that algorithm to create a comprehensive multifamily buildings database. In addition to identifying multifamily accounts, the database includes available data on building characteristics and an indicator of whether each building has participated in National Grid efficiency programs.

1.3 About Study Scope

To further define the scope of this study, it is important to discuss the following decision points for establishing the study.

1.3.1 Summary of National Grid's Residential Efficiency Portfolio

Between 2016 and 2020, National Grid implemented eleven different energy efficiency or demand reduction programs to residential customers in Rhode Island.[4,5] These included:

1. EnergyWise Single-Family (EWSF), electric and gas
2. Income-Eligible Services - Single-Family (IESF), electric and gas
3. EnergyWise Multifamily (EWMF), electric and gas
4. Income-Eligible Services – Multifamily (IEMF), electric and gas
5. Natural Gas Heating and Water Heating (NGWH), gas
6. ENERGY STAR® HVAC (ESHVAC), electric
7. ENERGY STAR Products (ESProducts), electric
8. ConnectedSolutions (CS), electric
9. Residential Lighting (ResLighting), electric
10. Home Energy Report (HER), electric and gas
11. Residential New Construction (RNC), electric and gas

Collectively, these programs saved a total of 807,610 megawatt hours, 8,921,634 therms, and 66,907 Metric Million British Thermal Units (MMBTUs) of heating oil and propane during this five-year period (74%, 24%, and 2% of the total savings, respectively).[6] As evident in FF, total residential portfolio savings peaked in 2019 with an annual MMBTU savings of 808,829. Program-specific annual savings are shown in Appendix C.

CS is a demand response program, therefore, the program does not attribute energy savings.

[4] Excludes pilot or demonstration projects.
[5] See Appendix B for detailed program descriptions and measure types offered by each program.
[6] Based on the residential program savings data between 2016 and 2020 National Grid provided.

Figure 1-1. Total Residential Program Annual Savings by Fuel by Year (in MMBTUs)

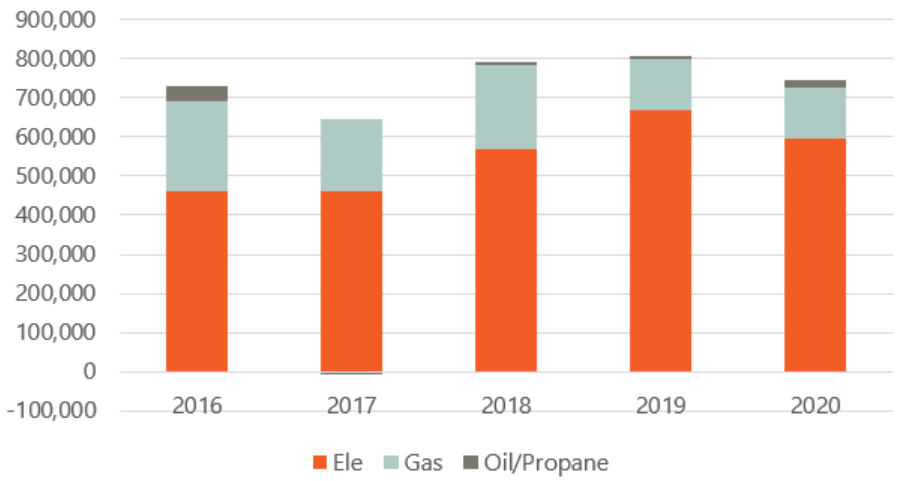


Figure1-2 shows how each of the 11 programs contributed to National Grid's residential savings achievement annually, as well as how program-specific savings have changed over time. As shown below, National Grid's ResLighting program (an upstream, retail-based offering) and the HER program (a behavioral program that randomly selects participants and uses an opt-out design) generated most of the portfolio-level savings in each year.

Figure1-2. Annual Savings by Program by Year (MMBTUs)

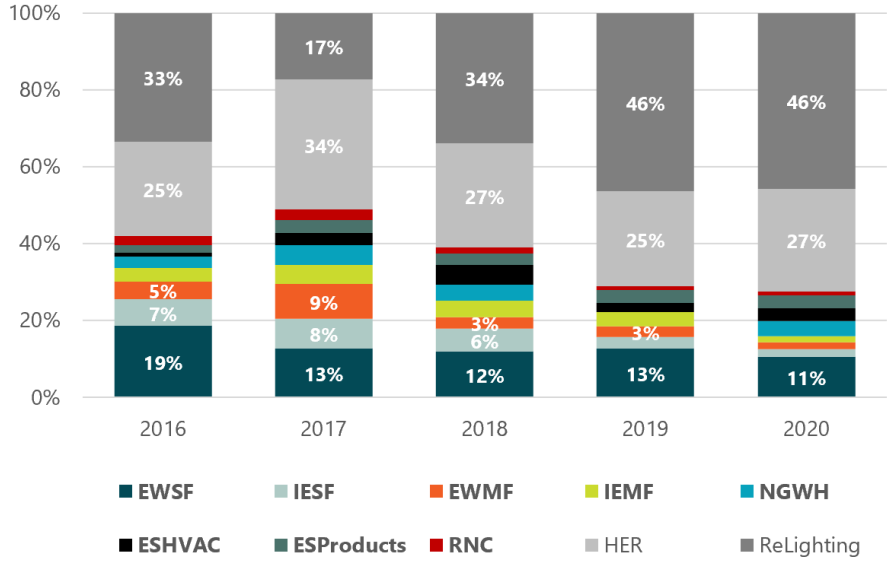


Table 1-1 provides a different, fuel-specific perspective on National Grid's residential portfolio in Rhode Island. It shows, separately for electricity and natural gas and oil/propane, each program's contribution to portfolio savings over the five-year period. Again, the ResLighting and HER program feature prominently especially for electric and gas savings, collectively, these programs are responsible for 63% of National Grid's total savings between 2016 and 2020.

Table 1-1. Fuel-Specific Savings by Program (2016–2020)

Program	Electric	Gas	Oil & Propane	Total
EWSF	7%	20%	190%	13%
IESF	2%	6%	100%	5%
EWMF	2%	11%	12%	4%
IEMF	2%	10%	3%	4%
NGWH	0%	13%	0%	3%
ESHVAC	1%	4%	59%	3%
ESProducts	4%	0%	5%	3%
RNC	1%	4%	8%	2%
HER	18%	57%	0%	27%
ResLighting	64%	-26%	-277%	36%
Total	**100%**	**100%**	**100%**	**100%**

Note: Negative percents resulted from interactive penalties.

1.3.2 Programs Included and Excluded

As shown in Section 1.3.1, we cannot understate the importance of the ResLighting and HER programs and their historical contributions to National Grid's total residential portfolio. However, the purpose of this study was not to reiterate previously reported overall portfolio savings, but to combine program tracking with customer, Census, and third-party data to provide new customer- and building-level insights. Specifically, we sought to:

- Analyze trends in customer and building participation
- Determine what customer and building characteristics drove participation
- Model how similar to—or different than—identified nonparticipating customers are to historical participants

The customer- and building-centric nature of our analysis required that we focus on the National Grid programs with customer-specific tracking data (i.e., where the specific account that participated is known) and when the customer made the decision to participate (i.e., an opt-in program design). These requirements meant our team excluded ResLighting (upstream, no customer-specific data) and HER (where National Grid randomly selects and enrolls participants).

Again, although the exclusion of these significant programs is notable, it is appropriate given the customer-specific nature of our analysis. Indeed, though both programs contributed significantly to National Grid's total energy savings between 2016 and 2020, they are fundamentally different program designs, incongruent with the other nine programs above and with the stated goals of this study: to analyze customer and building-level participation.

Consequently, our analysis focused on these eight programs, which represent 37% of the total portfolio savings between 2016 and 2020 but 100% of the customer-specific, opt-in program design savings.[7]

1. EWSF
2. IESF
3. EWMF
4. IEMF
5. NGWH
6. ESHVAC
7. ESProducts
8. CS[8]

We included RNC in this analysis but is assessed separately from these eight programs because the data from this program cannot be tied to existing electric and gas accounts, since most are new construction buildings.

1.3.3 Customer-centric Analysis Based on the Snapshot in Time Customer Data as of March 2021

All the analysis performed in this study is customer-centric analyses based on the snapshot in time active customer list as of March 2021. Different pieces of data provided by National Grid span different periods of time. For example, program participation data is a historical record of participants' accounts from 2016 to 2020. As a result, not all accounts in the program participation data match with the account database provided by National Grid. For example, if a customer participated in 2016 but has since canceled service and relocated, this account is not in the active customer list.

1.3.4 Focus on Reporting Aggregated Data

As charged by National Grid, Cadeo's primary focus of this study was to collect, aggregate, and report the data necessary to examine the participation trends. Wherever possible, we included high-level takeaways in each section into *why* a particular value may be or potential reasons for some trends we observed. However, it was outside the scope of this study to analyze the genesis of, rationale for, or specifics related to each count, metric, or trend. We anticipate National Grid will (and encourage them to) use this study and its summary analyses as the jumping-off point for deeper exploration into specific results.

[7] Our team cannot speak to how the results shown in this report for these nine programs potentially extend or apply to the two excluded programs given the fundamental differences in the program designs.
[8] As mentioned above, there are no savings associated with CS in the data provided by National Grid. In subsequent sections, CS is included with participation counts only,

1.4 Organization of This Report

This report provides comprehensive views of methodological approaches used and results yielded from this study. The report itself is the culmination of three task-specific findings memos that Cadeo shared with National Grid throughout the study.

We have organized this report organized as follows:

Cross-Cutting Data Assembly**Section 2 – Cross-Cutting Data Assembly:** data sources and data assembly activities that resulted in the core dataset for all the subsequent data analysis activities.

Section 3 – Understanding Historical Participation Trends: key takeaways, analytical objectives, methodology of participation metrics calculations, and results of portfolio- and program-level key participation metrics, as well as examination of participation results through the equity perspectives. This section also includes an analysis of participation in the RNC program.

Section 4 – Modeling Drivers of Participation: key takeaways, analytical objectives, methodology of the key-driver model development, and results of participation-drivers model.

Section 5 – Estimating Nonparticipants Propensity to Participate: key takeaways, the purpose and analytical objectives, methodology of propensity score development, and results of the nonparticipants propensity to participate.

Section 6 – Identifying Multifamily Customers: key takeaways, the purpose and analytical objectives, methodology to develop multifamily customers/buildings identification algorithm, and results of the multifamily identification algorithm.

Appendices: glossary of terms (Appendix A); descriptions of the residential energy efficiency programs covered by this study (Appendix B); program-specific annual savings (Appendix C); details of program eligibility criteria (Appendix D); comparison of National Grid's Year-End Report's and row program data's participant counts (Appendix E); details of participation metrics calculation methodology (Appendix F); program-specific participation analysis results (Appendix G); results of program sequence analysis (Appendix H); correlation analysis results that informed feature reduction of the key-driver model (Appendix I); model coefficients of the causal key-driver model (Appendix J); examples of condominium buildings and algorithm classifications (Appendix K); and implementation notes for modeling (Appendix L).

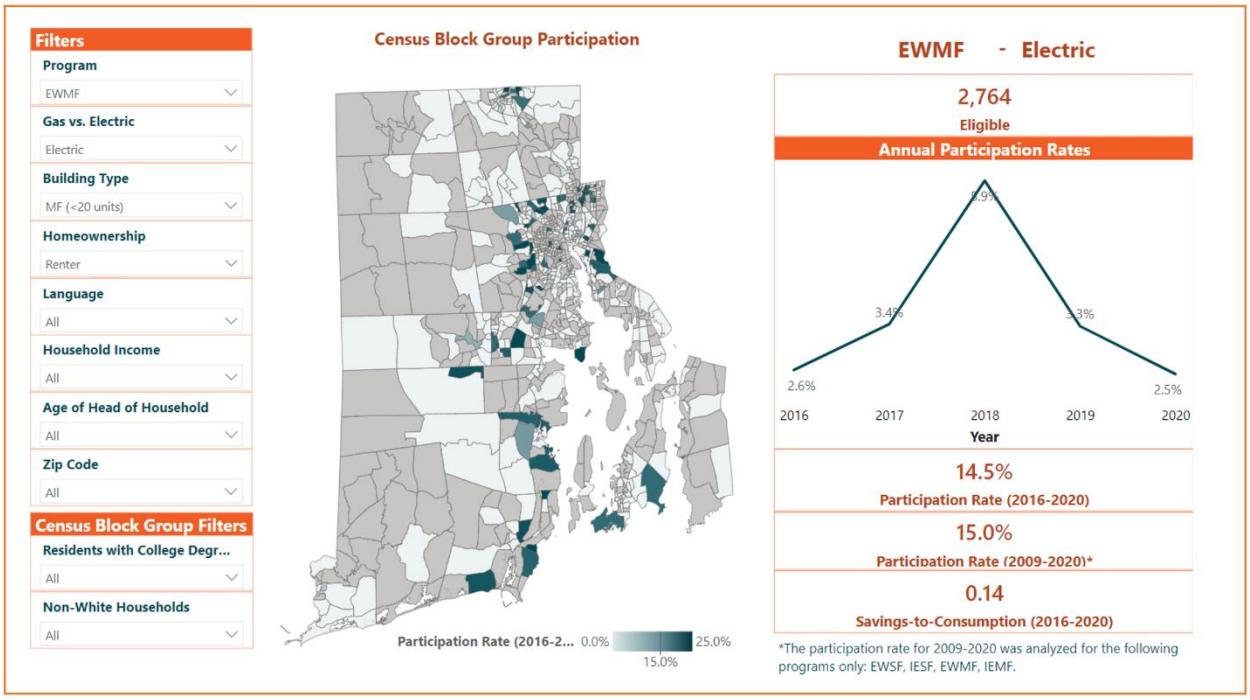
1.5 PowerBI Dashboard

Cadeo also created a supplemental interactive dashboard using Microsoft PowerBI that National Grid can use to dynamically investigate questions regarding historical participation (or nonparticipation) in the future. The dashboard includes the entire datasets our team created and used for this study, as well as user-friendly interface and integrated geographic information system (GIS) functionality. We envision the dashboard will function as an internal tool for National Grid and support contractors to market and deliver programs.

We organized the dashboard tabs by two components: participants and nonparticipants. Users can analyze the participation metrics at the account and building level to summarize the data in the same way as presented in this report or more in depth. For the nonparticipant metrics, users can explore nonparticipants' geographic locations and their participation propensity. Users can interactively filter all these data and map visuals by program and customer characteristics to get dynamic insights on participants and nonparticipants. Users can also export participants and nonparticipants data based on applied filters.

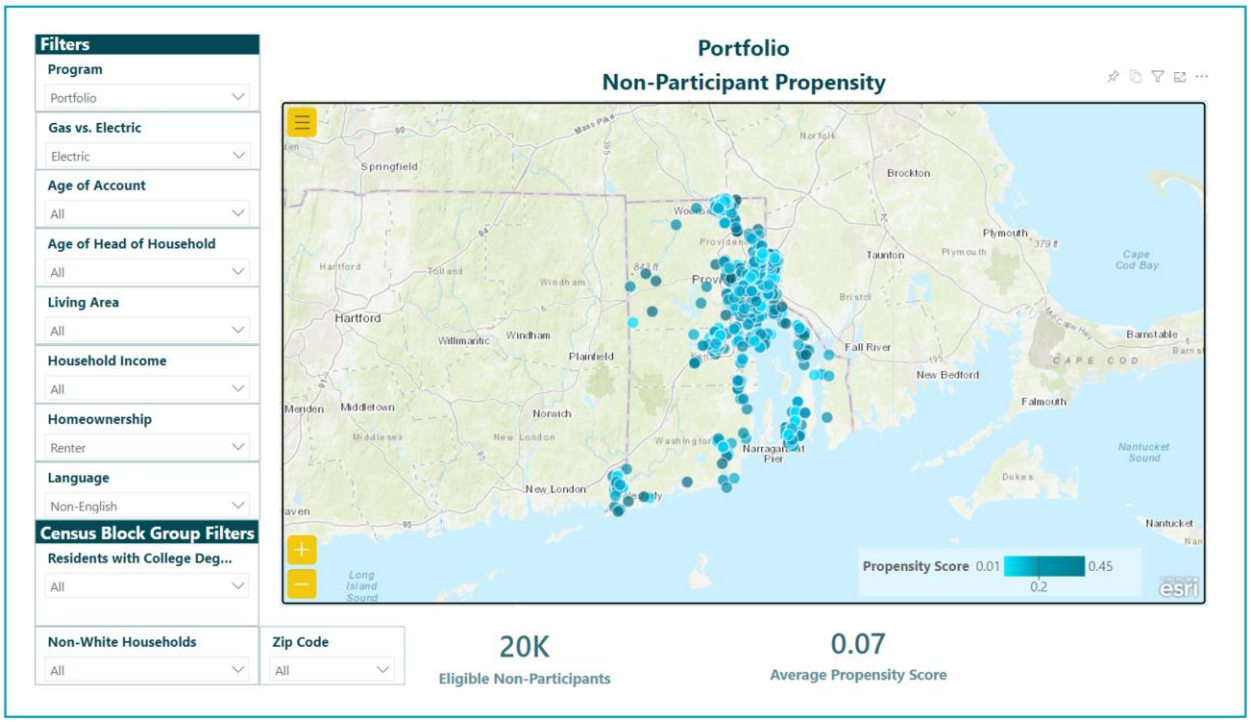
As an example, Figure 1-3 shows a use case to obtain insights on EWMF program participants who are renters living in small multifamily buildings. It shows all the participation metrics as well as census block group-level participation rates among the filtered participant characteristics.

Figure 1-3. Use Case of Interactive Participation Analysis in Dashboard



Another use case example, shown in Figure 1-4, analyzes nonparticipant participation propensity among those who are non-English-speaking renters who can potentially participate in any National Grid program. It provides street-level locations of these nonparticipants with color-coded propensity score to participate based on the applied filters. Users can readily export the filtered nonparticipant data.

Figure 1-4. Use Case of Interactive Nonparticipation Analysis in Dashboard



To obtain access to this dashboard, please contact the NGRI Evaluation Manager. The dashboard will not be publicly available and is only for National Grid' s internal planning purposes.

Section 2 Cross-Cutting Data Assembly

This section details the methodology and approach Cadeo used to develop the core dataset that became the basis of all subsequent analyses.

2.1 Data Sources

At the outset of the study, National Grid provided Cadeo with Rhode Island customer account details, program participation tracking data, billing data, and property data. We supplemented these data sources with additional property data purchased from third-party data aggregator First American (FA) and publicly available census block group data from the American Community Survey (ACS). We used these collective files to create a detailed dataset that ties each account to information regarding the customer's residence, demographics, energy consumption, and energy efficiency program participation history. We used this comprehensive dataset to identify the accounts associated with multifamily buildings (i.e., buildings/properties with 5+ residential units).

Key data sources included:

- **Active Customer Account List.** This list contains NGRI's electric and gas customer accounts active in March of 2021. It includes customers' billing account, customer number, premise number, building key, rate code, property address, and contact information. It also included third-party customer demographic data (previously merged on by National Grid), including household income, homeownership, age of householder, number of householders, preferred language, and marital status. This list contains all the residential accounts as well as commercial and industrial accounts associated with residential locations, such as common areas of multifamily buildings or mixed-use buildings where accounts are master metered.
- **Property Tax Data.** Cadeo purchased Rhode Island tax property data from FA Data Tree in August 2021.[9] The purchased data covered the entire state and contained dozens of fields regarding the customer's property including building descriptions, building and property size, building age, among others listed in Figure 2-1.
- **Energy Consumption Data.** Cadeo received monthly energy consumption (i.e., billing data) from National Grid for every electric and gas account active in Rhode Island between January 1, 2016, and December 31, 2020. Consumption of delivered fuels are not part of these billing data.
- **Building-level Data.** National Grid provided a compilation of building-level data for Rhode Island customers, which National Grid derived by aggregating account-level data to the building-level, as well as using other customer and tax parcel data. This included

[9] National Grid originally provided previously acquired third-party property tax data to support this study. However, the dataset was missing data for one of the five counties in the state and included far fewer complete demographic fields to support analysis and modeling (than the First American data). For these reasons, we relied exclusively on the property tax data that the team acquired from First American.

information such as building-level annual electric and gas usage, and the number of electric and gas accounts at the building level.

- **Program Participation Tracking Data.** National Grid also provided program participation tracking data in the form of a list of electric and gas accounts that participated in the following residential programs from 2016–2020:
    - EWSF
    - IESF
    - EWMF
    - IEMF
    - NGWH
    - ESHVAC
    - ESProducts
    - CS

  These tracking data included participating customers' account number, date of participation, information about the installed measures (quantity and associated gross savings), and incentives paid.

  Additionally, to enable a longer-term participation tracking, National Grid also provided program participation summary data used in the previous NGRI Energy Efficiency Program Customer Participation Study that assessed 2009–2015.[10]

  National Grid also included program participation tracking data for the RNC program, listing participating properties and units by project ID and rebate number for 2016 to 2020. This data includes fuel type, a low-income flag, building type, project name, address, and information about measures installed (quantity and savings) and the building rebate amount provided.

- **Census Bureau's ACS data.** To complement the data provided by National Grid, Cadeo extracted additional demographic data (not available in the Active Customer Account list above) from the 2020 ACS 5-year estimates at the census block group-level. This data reflects the proportion of the population within each of the census block group associated with a given census characteristic. These characteristics include information such as race, ethnicity, and education. We merged ACS data onto the customer account data provided by National Grid after we geocoded each provided customer address. It is important to note that since ACS data is not customer-specific, our team associated the same census block group-level information with every National Grid customer residing within the census block group. Each census block contains approximately 250 to 550 households and the ACS values assigned to each account are based on the households of the census block group in which they fall and are not specific to each account. For

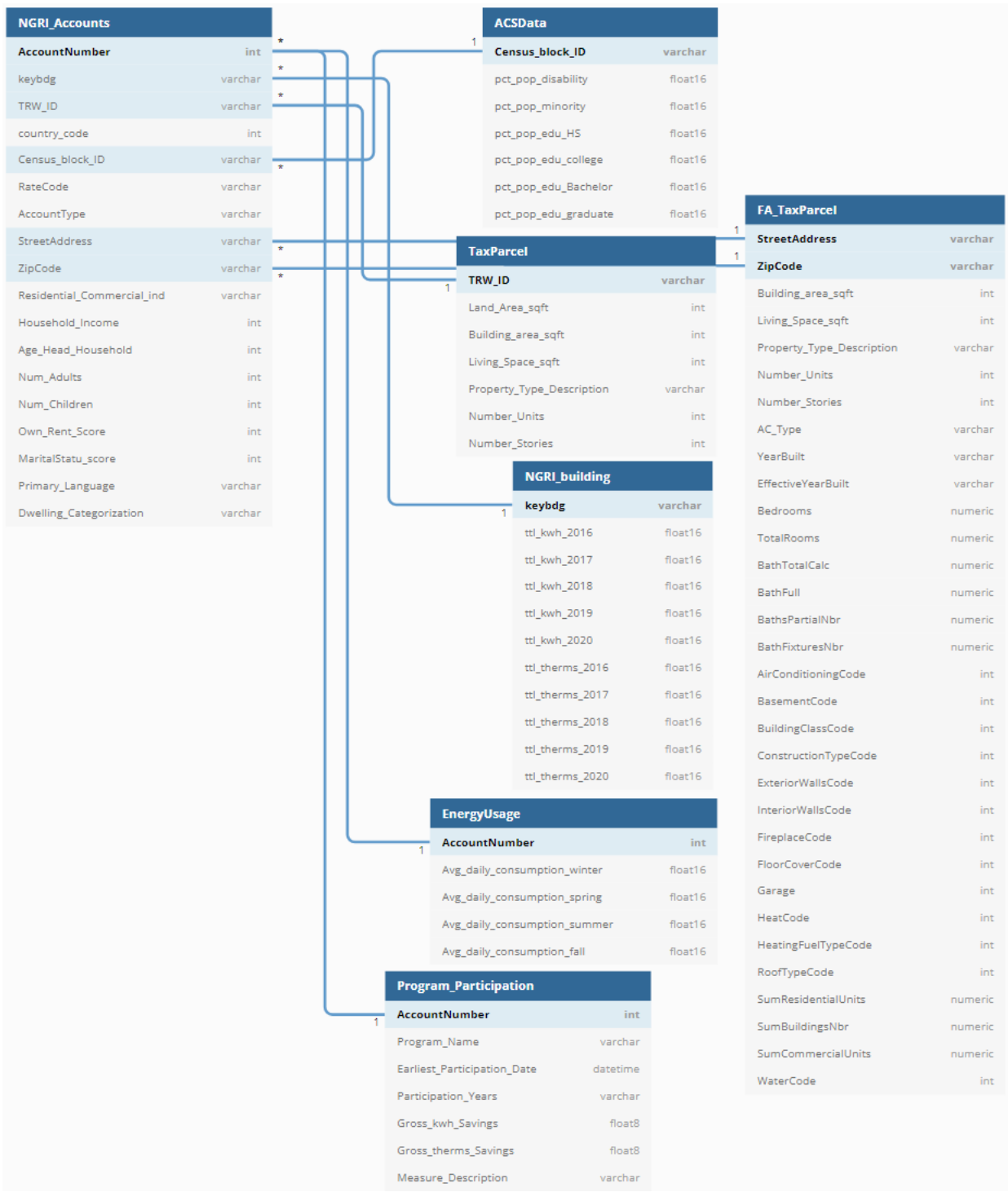
[10] Navigant Consulting, *Energy Efficiency Program Customer Participation Study, National Grid Rhode Island* (October 19, 2017), http://rieermc.ri.gov/wp-content/uploads/2018/03/national-grid-2017-ri-ee-customer-participation-study-final.pdf.

example, if an account is in a census block group for which 40% of the population has a bachelor's degree, we assigned this value to that account.

2.2 Join and Match Rates

Figure 2-1 provides a visual representation of how we combined the above data sources. We identified appropriate join keys in the active customer account data (NGRI_Accounts) to facilitate joins with the other data sources listed above. We determined that the **account number** joined the billing data (EnergyUsage) and program participation tracking data (Participation_Summary), whereas **building key (keybdg)**—a string variable resulting from concatenation of account's ZIP code and street address to represent a unique value per building—joined the building data (NGRI_building). We also used TRW_ID connected the National Grid–sourced property tax data (TaxParcel) and merged the ACS data (ACSData) using the census block group ID (Census_block_ID). We standardized street addresses and parsed to match the FA tax parcel data by address and ZIP code.

Figure 2-1. Visual Representation of Data Joins



After each join, we checked the data health and assessed the accuracy of the data joins performed. **Error! Reference source not found.** summarizes the residential electric and gas account records, and match rates of the merged data tables.

As shown in Table 2-1**Error! Reference source not found.**, there were 443,737 residential electric accounts and 248,119 residential gas accounts in the active customer database.[11] We found that building-level data matched to 100% of the accounts and that billing data matched for 97% of both electric and gas accounts. We expected these high match rates between accounts, buildings, and billing data given all three datasets originated within National Grid's internal data management system.

Next, Cadeo merged on the third-party tax data. Since there is not a common unique numeric identifier in both datasets (e.g., an account number), our team used a combination of address and name to merge the tax data with the aggregated National Grid dataset. The variability inherent in address and name fields result in a lower, but still strong, match rate of 76% and 77% for electric and gas accounts, respectively.

Our final join was demographic data from the Census ACS. We used the US Census Geocoder tool to identify the census block group based on the provided customer address. [12] The tool was able to identify 86% and 90% of the addresses associated with the electric and gas accounts, respectively. The non-matches with ACS data represent addresses the Geocoder tool could not identify (i.e., the address was not successfully geocoded). Rural properties, newer properties, and multi-unit buildings are at a greater risk of failing to match by address.

As shown in Table 2-1 we have complete data (i.e., successful join between the building-level data, billing data, property tax data, and ACS data) for 67% of residential electric accounts and 69% of gas accounts. The property tax and ACS match rates, which had lower individual match rates, primarily drove the these overall match rates. Our overall match rates for both fuels are similar to that of the previous Rhode Island Participation study, which also used an addressed-based approach for joining non-National Grid data.[13]

Table 2-1. Summary of Residential Accounts and Data Merge Match Rates

	Residential Electric	Residential Gas
Total unique active accounts in customer database	**443,737**	**248,119**
Percent of building-level data matched	100%	100%
Percent of billing data matched	97%	97%
Percent of FA property tax data matched	76%	77%
Percent of ACS data matched	86%	90%
Matched with All Joined Data	**67%**	**69%**

Cadeo also assessed how many of the customers provided in the 2016–2020 program participation tracking data we could match to the active customer account data. As shown in Table 2-2, we found the highest match rate occurred for the EWSF program with an 80% match

[11] The account is the basic unit of representation in our raw data and the level at which we will be modeling participation, therefore, we're reporting match rates at the account level.
[12] US Census Geocoder tool, https://geocoding.geo.census.gov/geocoder.
[13] Previous Rhode Island Participation Study had ~70% match rate.

rate for electric participants and a 79% match rate for gas participants. Conversely, the IESF program showed lower match rates (77% is the maximum possible for electric or gas). Since the program participation went back to 2016, we anticipated that some subset of program participants would no longer be active National Grid customers. The lower match rates for IESF indicates greater transiency for the customers that participate in this program.

Table 2-2. Participants in March 2021 Active Customer Account List by Program

Program	Electric Participants	Gas Participants
EWSF	80%	79%
IESF	77%*	
EWMF	100%**	
IEMF	100%**	
NGWH	N/A	74%
ESHVAC	83%	N/A
ESProducts	84%	N/A
CS	89%	N/A

* IESF data does not indicate which type of account (electric or gas).

** We attributed MF program data to facilities. NGRI provided a map from current accounts to facility ID. All facility IDs were in the mapping which is why the match rate is 100%, but that does not mean that the current mapping would match the mapping at the time of program participation.

As discussed above, we assessed RNC separately and did not include this program in the charts above because we cannot tie the data to active electric and gas accounts.

Section 3 Understanding Historical Participation Trends

3.1 Key Takeaways

- Among active accounts, participation grew annually from 2016 to 2019 across nearly every program and for both electric and gas customers alike.
- Overall participation declined in 2020, most likely due to COVID-19.
- Most programs exhibited a year-over-year decline in savings-to-consumption (SC) ratios, meaning that participants saved less of their total consumption over time. This was true for both electric and gas savings.
- Looking deeper, we found that participants' consumption was relatively constant over the five-year period and that the decreased participant savings over time drove declines in SC. Specifically, we observed lower average annual lighting savings (electric) as participants installed fewer lamps and as per-lamp savings dropped due to changing baseline assumptions. On the gas side, per-participant weatherization savings dipped slightly over time, but the primary driver of the lower SC was fewer participants opting to weatherize their home.
- Higher income groups, homeowners, and English speakers consistently had higher participation rates.

3.2 Goals

Participation trend analysis assesses the portion of residential active accounts that have participated in National Grid's residential energy efficiency programs for which they are eligible during the study period. As discussed in Section 2, Cadeo conducted the participation trend analysis on accounts active as of March 2021 (active accounts). The analysis does not include accounts that participated in 2016–2020 if that account was no longer active as of March 2021.

As discussed in Section 1.3, the team analyzed participation for all the eight residential programs for which National Grid collected customer-specific tracking data, and excluded the upstream ResLighting (due to no customer-specific data) and HER programs (due to random customer selection and opt-out program design).

RNC enrolls customers of new construction buildings, many of whom do not have an existing electric or gas account; therefore, we analyzed the RNC program separately. See Section 3.4.2 for RNC-specific methodology and results.

Cadeo concurrently provided National Grid with the participation trend analysis results using PowerBI-based dashboard. This interactive tool uses GIS information to show participation and other trends, including demographic, property, and usage characteristics, at a detail level.

3.3 Methodology

3.3.1 For Determining Program Eligibility

Determining whether an active account is eligible for each of the National Grid's energy efficiency programs is a necessary step to correctly establish a participation status for each program. This subsection discusses our methodology for determining program eligibility.

Using historical program data, our team quickly denoted that any active account that had participated in a program was, by virtue of their previous participation, eligible for that program. However, the program eligibility process was not as simple for nonparticipating accounts. Also, it was possible that a participating account was concurrently eligible for, but had not participated in, another program.

We determined the full set of program eligibility for each active account based on the following three relevant account characteristics:

1. **Building type** – whether we determined the account is in a multifamily (5+ units) or single-family (1–4 units) building as discussed in Section 6
2. **Income Eligibility** – income eligible or not based on rate code and household income less than 60% statewide area median income[14]
3. **Account type** – electric or gas

For example, an income-eligible electric account in an multifamily building would be eligible for IEMF, ESHVAC, and ESProducts. It is important to note that the number of accounts eligible for each program is the same for each year of the study period (i.e., 2016–2020) because our eligibility analysis relied on account characteristics from the active account list as of March 2021.

Error! Reference source not found. Table 3-1 summarizes the specific criteria our team used to determine eligibility for each program. Appendix D provides a breakdown of the specific data fields.

[14] National Grid Rhode Island considers income eligible households to be those with 60% state median income or less. Based on FY2021 Rhode Island Income Limits for Low- and Moderate-Income Households, https://www.rihousing.com/wp-content/uploads/FY-21-HUD-Income-Limits.pdf.

Table 3-1. Account Eligibility Criteria by Residential Program

Program	Eligibility Criteria
EWSF	• Account associated with single-family building • Non-income-eligible or recent past participant
IESF	• Account associated with single-family building • Income-eligible or recent past participant
EWMF	• Account associated with MF building • Non-income eligible and in a building in which less than 50% of accounts are income eligible or recent past participant
IEMF	• Account associated with MF building • Income eligible or in a building in which 50% or more of the accounts are income eligible or in a building considered public housing or recent past participant
NGWH	• Gas account
ESHVAC	• Electric account
ESProducts	• Electric account
CS	• Electric accounts with central air conditioning or central heat pump with Wi-Fi supported thermostat. We estimated the number of electric accounts eligible for this program with the assumptions, per NGRI's directive, that 30% of electric accounts have air conditioning or central heat pump and of those 19% have an eligible Wi-Fi thermostat[15]

Note: For RNC-specific methodology and results, see Section 3.4.2.

In theory, the EWMF, IEMF, EWSF, and IESF are mutually exclusive programs (i.e., each account can only be eligible for one program) based on their building type and income status. However, since income can fluctuate over time, we observed a small number of instances where an account participated in both a standard-income and income-eligible program between 2016 and 2020. In these rare instances, we considered the account is eligible for the program they most recently participated in.

The team determined program eligibility at the building level through aggregation of account eligibility using the building identifier field, keybdg. As detailed in Section 6.3.1, keybdg, a National Grid generated data field, is based on the account address, excluding unit number. If a

[15] Data driven determination of presence of AC or central heat pump at the account level was not feasible given the lack of HVAC data in the customer database or property tax data.

building contained an account eligible for a program, then we considered the building is eligible for that program.[16]

Table 3-2 summarizes the number of active electric and gas accounts that we determined are eligible for each National Grid offering.

Table 3-2. Eligible Accounts by Residential Program

	Electric		Gas	
	Accounts	Share	Accounts	Share
Total Accounts	**443,737**	**100%**	**248,119**	**100%**
EWSF	267,224	60%	141,065	57%
IESF	93,104	21%	66,405	27%
EWMF	50,835	12%	26.631	11%
IEMF	32,574	7%	14,018	6%
NGWH	-	0%	248,119	100%
ESHVAC	443,737	100%	-	0%
ESProducts	443,737	100%	-	0%
CS	25,293	6%	-	0%

Note: For RNC-specific methodology and results, see Section 3.4.2.

[16] The term "building" is used loosely as a catchall for properties with the same address that could include multiple buildings.

Table 3-3 provides a similar eligibility summary as above but at the building level by program and account type.

Table 3-3. Eligible Buildings by Residential Program[17]

Programs	Electric		Gas	
	Buildings	Share	Buildings	Share
Total Buildings	**336,408**	**100%**	**195,302**	**100%**
EWSF	245,114	73%	128,942	66%
IESF	77,019	23%	54,458	28%
EWMF	15,819	5%	12,539	6%
IEMF	8,681	3%	6,211	3%
NGWH	-	0%	195,302	100%
ESHVAC	336,408	100%	-	0%
ESProducts	336,408	100%	-	0%
CS	20,184	6%	-	0%

Note: For RNC-specific methodology and results, see Section 3.4.2.

3.3.2 For Calculating Participation Metrics

To understand participation of active accounts in each of the energy efficiency programs, Cadeo developed two complementary metrics—**participation** and **SC ratio**—that offer insight into the historical reach of National Grid's residential efficiency programs. We calculated these metrics at both the account and building level, as well as for each year of the study period and cumulatively over the five-year period.[18] It is important to note that savings are not consistently calculated from year to year due to changes in program implementation and changes in the predicted measure savings. Additionally, the mix of measures delivered through each program may differ from year to year leading to some of the variation in the SC ratio. In Appendix G, we have included specific information about the measure mix delivered in each program and each year.

Table 3-4 summarizes the use cases and limitations of each metric.

[17] The eligibility percentages of buildings for EWSF, IESF, EWMF, and IEMF add up to slightly more than 100% because there are a small number of buildings that are eligible for both the standard rate and income eligible program. This is because there are accounts within the building that are eligible for each program.

[18] The SC ratio was not calculated for the CS program because the focus of the program is demand savings, not energy savings.

Table 3-4. Participation Metrics Use Cases and Limitations

Metric	Best Used to Answer	Limitations
Account Participation (Account-level binary participation status)	How many accounts did the program reach in each year and over the 5-year study period?	• Not stable over time (people moving in/out of buildings with a consistent account) • Does not provide any insight into the magnitude of the savings achieved through participation
Building Participation (Building-level binary participation status)	How many buildings, regardless of size, did the program touch in each year and over the 5-year study period?	• Using binary approach, MF buildings count the same as single-family homes, although their consumption is typically far greater • Potential to overstate MF participation (i.e., not all units participate but building considered a participant)
Account Savings-to-Consumption (Ratio of savings achieved and consumption of participants at account level)	What was the impact of participation of each account in each program in each year and over the 5-year study period? What are the characteristics of participants with a lower or higher savings-to-consumption ratio?	• Only relevant for electric and natural gas consumption (due to inability to access delivered fuel records) • Only possible for participating accounts active in March 2021[19] • Accounts with large consumption may have a low savings-to-consumption ratio despite significant savings
Building Savings-to-Consumption (Ratio of savings achieved and consumption of participating buildings)	How well did programs reduce consumption at the building level in each year and over the 5-year study period?	• Large MF buildings with only a few participating accounts will have a very low savings-to-consumption ratio while single-family buildings may have much higher savings-to-consumption ratios

The calculation methodology for each metric is discussed in depth in Appendix F.

3.4 Results – Tracking Participation Trends

The following section details the results of the participation analysis, primarily focusing on the participation rates and SC ratios at the portfolio level and by program for active accounts. The

[19] National Grid was able to provide historical billing data for all customers active in March 2021, but not for accounts that participated between 2016 and 2020 but closed their account prior to March 2021.

section examines the savings and consumption trends that lead to a declining pattern in the SC ratio, including an examination of the measure mix and savings assigned to measures. We also examine the participation metrics by a few equity variables of interest. The final subsection provides the analysis results of National Grid's RNC program. Appendix G provides additional program-level participation analysis results.

3.4.1 Participation Analysis

3.4.1.1 Portfolio Participation Metrics

To assess the overall reach of National Grid's energy efficiency programs, the team first analyzed the portfolio-level participation of all active accounts and buildings.

Table 3-5 summarizes the results of portfolio-level annual account and building participation (participated in one program or more in each year) and annual SC ratios, as well as the results of cumulative participation (unique count of participants that participated in any program in any year) metrics during 2016–2020 and 2009–2020.

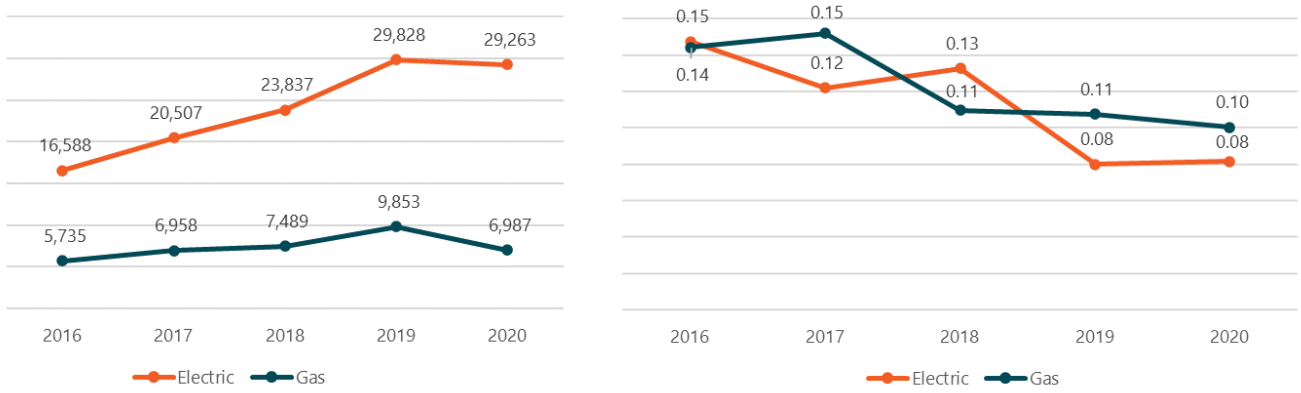
Table 3-5. Annual and Cumulative Participation at Portfolio Level

Participation Metric		Annual					Cumulative	
		2016	2017	2018	2019	2020	2016–2020	2009–2020 *
Electric								
Account	Count	16,588	20,507	23,837	29,828	29,263	**94,753**	**111,214**
	% of accounts	4%	5%	5%	7%	7%	**21%**	**26%**
	SC Ratio	0.15	0.12	0.13	0.08	0.08	**0.11**	**NA**
Building	Count	13,617	17,051	20,521	25,971	25,477	**80,318**	**95,812**
	% of buildings	4%	5%	6%	8%	8%	**24%**	**29%**
	SC Ratio	0.11	0.09	0.09	0.07	0.05	**0.09**	**NA**
Gas								
Account	Count	5,735	6,958	7,489	9,853	6,987	**29,997**	**35,265**
	% of accounts	1%	2%	2%	3%	2%	**12%**	**14%**
	SC Ratio	0.14	0.15	0.11	0.11	0.10	**0.12**	**NA**
Building	Count	4,871	5,838	6,410	8,561	6,149	**26,248**	**31,415**
	% of buildings	2%	3%	3%	4%	3%	**13%**	**16%**
	SC Ratio	0.13	0.13	0.09	0.09	0.09	**0.11**	**NA**

* Cadeo received 2009–2015 account participation records for EWSF, IESF, EWMF, and IEMF programs only. As a result, our 2009–2020 cumulative totals do not include participation in NGWH, ESHVAC, ESProducts, and CS from 2009 to 2015. The 2009–2015 participation data included binary participation data only, no savings data; therefore, SC ratios are not available for 2009–2020.

Figure 3-1 shows account participation and SC ratios at portfolio level.

Figure 3-1. Account Participation and SC Ratio at Portfolio level



Between 2016 and 2020, 21% of active electric accounts and 12% of active gas accounts participated in at least one of the programs. Relatedly, during the same period, 24% of buildings with active electric accounts and 13% of buildings with active gas accounts had at least one unit that participated in a program. Cumulative account-level SC ratios were 0.11 and 0.12 for electric and gas accounts respectively. While the number of participants at the portfolio level were in an increasing trend (except for 2020, likely the effect of COVID-19), the SC ratios for electric and gas accounts were in a declining trend between 2016 and 2020.

3.4.1.2 Account Participation by Program

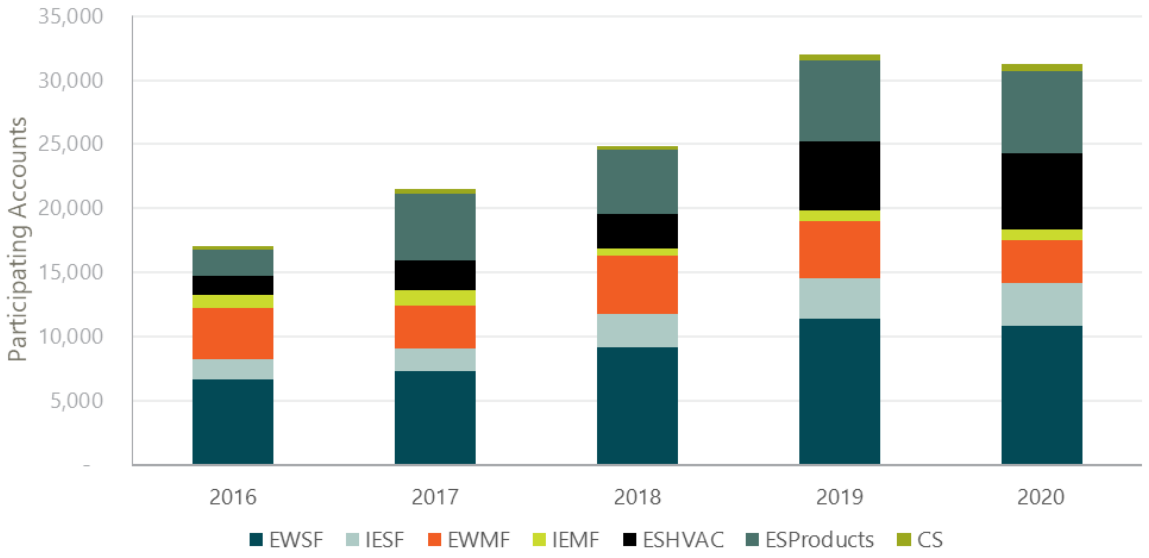
This section presents account-level participation and SC ratios by program. Unlike portfolio-level results, accounts that participated in multiple programs and years are counted in multiple times. Additional program-level analysis can be found in Appendix G

Electric

Figure 3-2**Error! Reference source not found.** shows participation of active electric accounts during 2016–2020 by program. As seen above, portfolio-level participation was increasing year over year until 2019 then slightly dropped in 2020. Programs such as EWSF, IESF, ESHVAC, and ESProducts had notable increases in participation during this period.[20]

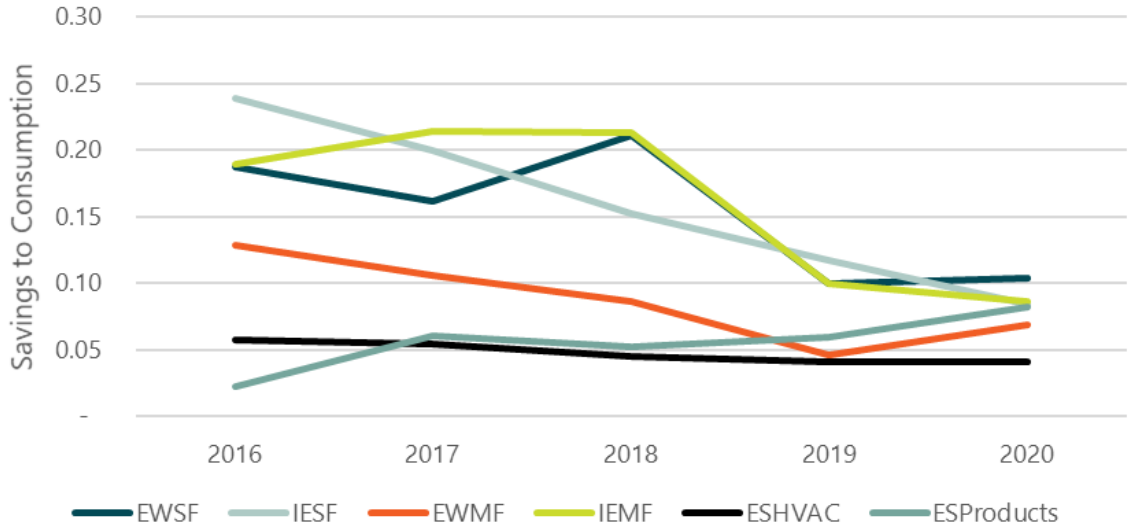
[20] As noted in the methodology section, these participation counts include only the participants that had program data that we could tie to electric and gas active accounts.

Figure 3-2. Participation of Active Electric Accounts by Program, 2016–2020



When looking at account SC ratios of active electric accounts by program (Figure 3-3**Error! Reference source not found.**), all the programs, except for ESProducts, had declining SC ratio over the study period. We discuss this declining trend further in Section 3.4.1.3 and in greater detail in the program-specific participation results in Appendix G.

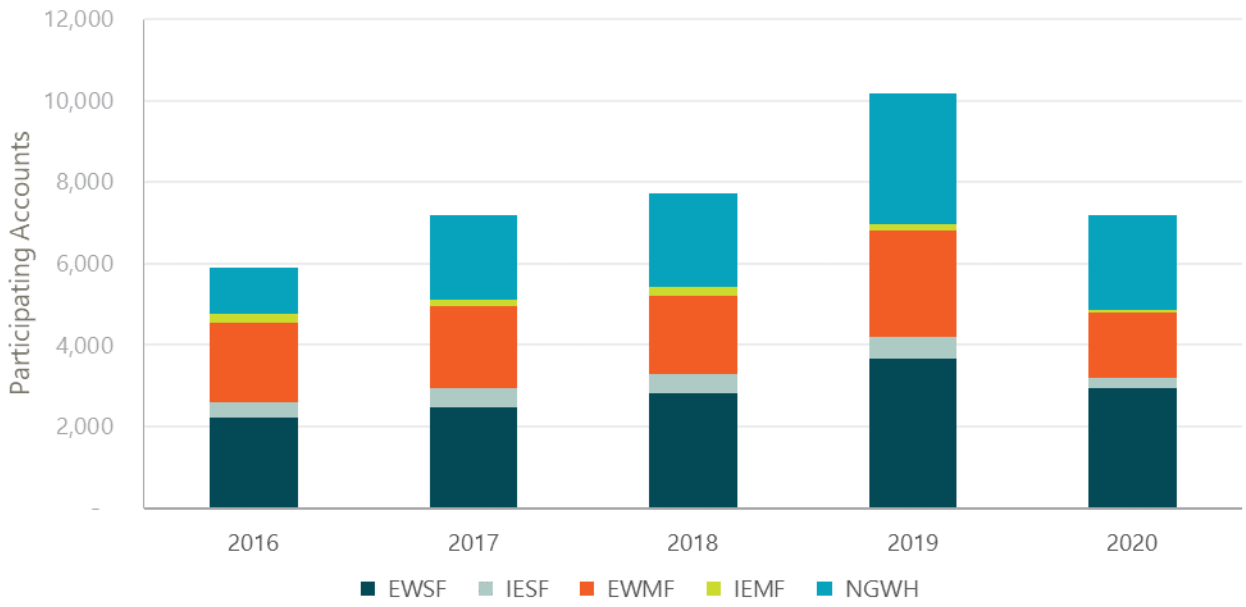
Figure 3-3. SC Ratio of Participating Electric Accounts by Program, 2016–2020



Gas

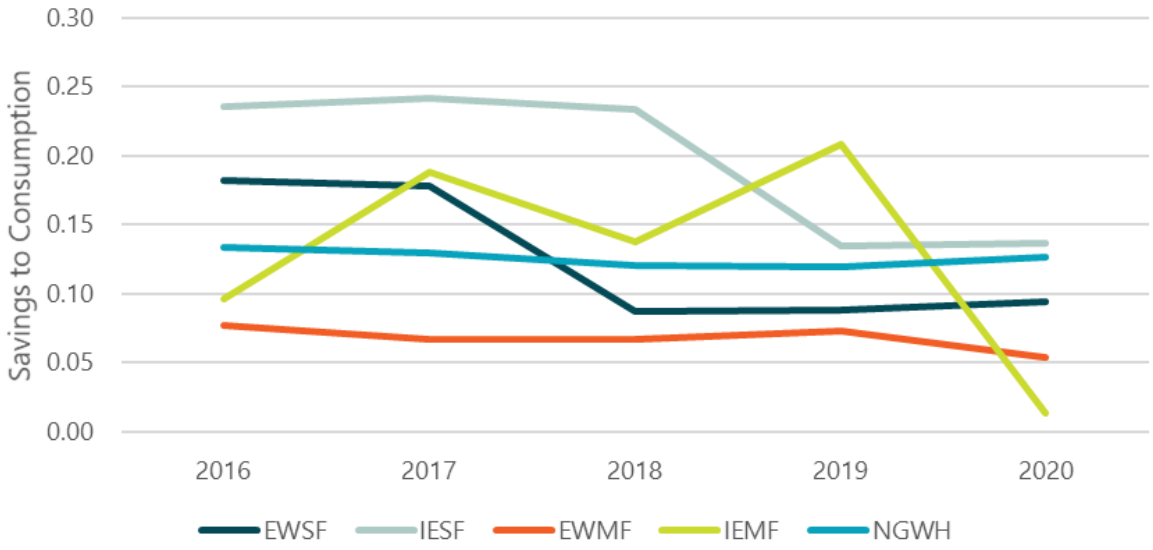
Figure 3-4 shows participation of active gas accounts during 2016–2020 by program. Similar to the trends of electric accounts, portfolio-level participation of gas accounts increased during the study period, except for 2020. Increases in participation in EWSF and NGWH programs are most notable.

Figure 3-4. Participation of Gas Accounts by Program, 2016–2020



As shown in Figure 3-5, SC ratios of participating active gas accounts are generally in a declining trend, most notably for EWSF and IESF programs. The dramatic drop-off in SC ratio for IEMF gas accounts is likely due to a decrease in the number of measures installed in multifamily buildings during the COVID-19 pandemic. Again, we discuss these trends in Section 3.4.1.3 and in greater detail in the program-specific participation results in Appendix G.

Figure 3-5. SC Ratio of Participating Gas Accounts by Program, 2016–2020



3.4.1.3 Examining the Declining SC Ratio

To understand the reasons behind the declining SC ratio, the team examined the savings and consumption trends throughout the study period. We found that savings per participant fell between 2016 and 2020 while consumption remained relatively constant (Table 3-6).

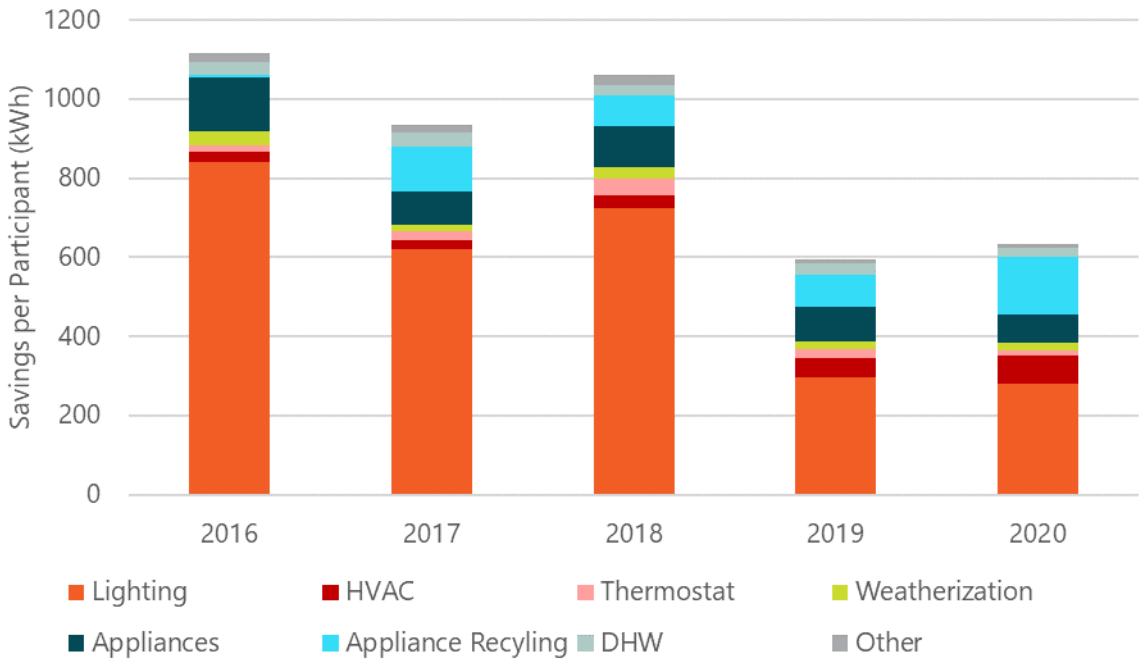
For electric accounts at the account level, the average percent change in savings per participant between 2016 and 2020 was -10%, while that in consumption was 1%. For gas accounts, the average percent change in savings per participant during the same period was -12%, while that in consumption was only -4%.

Table 3-6. Average Portfolio Savings and Consumption per Participant by Year

Avg. savings and consumption per participant	2016	2017	2018	2019	2020	**Avg. annual % change,** 2016–2020
Account Level						
Electric						
Kilowatt-hour (kWh) Savings	1,120	936	1,062	598	639	**-10%**
kWh Consumption	7,597	7,671	8,015	7,488	7,840	**1%**
Gas						
Therm Savings	132	128	97	92	78	**-12%**
Therm Consumption	913	841	889	856	780	**-4%**
Building Level						
Electric						
kWh Savings	1,422	1,179	1,250	696	770	**-11%**
kWh Consumption	14,175	12,709	13,395	11,599	12,719	**-2%**
Gas						
Therm Savings	155	151	112	104	89	**-12%**
Therm Consumption	1,205	1,197	1,307	1,195	1,133	**-1%**

The decline in savings for participating electric accounts was primarily due to a decline in the average kilowatt-hour (kWh) savings per participant from direct install lighting measures, which made up over 60% of the electric savings during the study period. We attribute the observed decline to National Grid, like many other utilities, transitioning away from reliance on lighting savings, which buoyed residential programs and portfolios for many years. Meanwhile, there was a slight increase in average savings per participant from appliance recycling and HVAC measures. Figure 3-6 illustrates these findings.

Figure 3-6. Average kWh Savings per Participant at the Portfolio Level



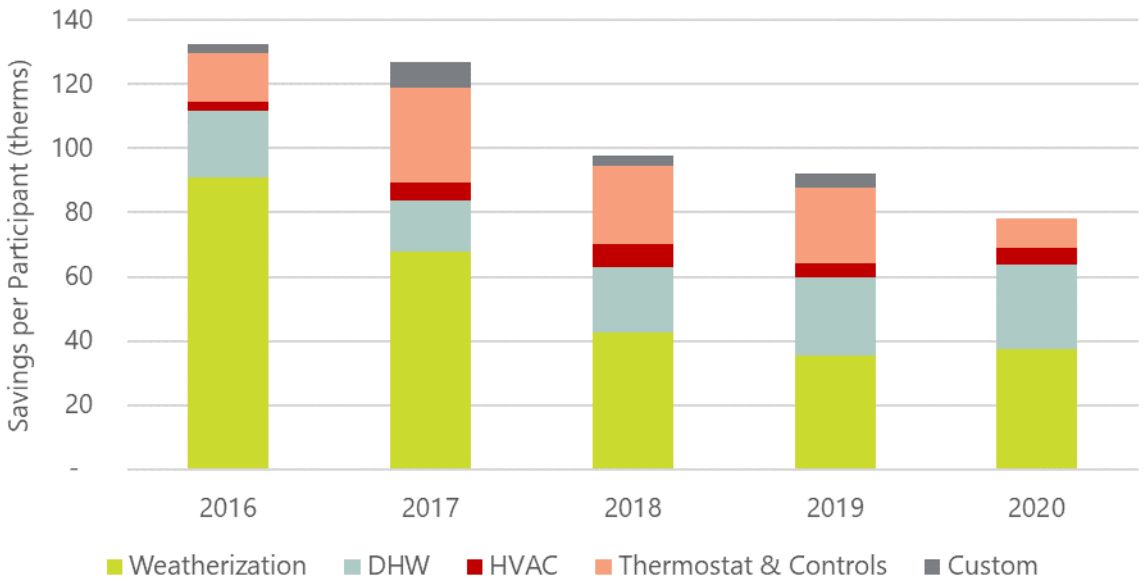
The decline in kWh savings per participant from lighting measures is due to both a decline in the average number of lighting measures installed per participant and a decline in the average kWh savings attributable to each measure installed. Between 2016 and 2020, the average annual percent change in number of lighting measures per participant was -7%. The average annual percent change in kWh savings per lighting measure was -7% during the same period. Table 3-7 provides the year-over-year details.

Table 3-7. Lighting Measures Installed in Single-family Homes

	2016	2017	2018	2019	2020	Avg. Annual % change, 2016–2020
Average # of lighting measures per participant	27.8	24.3	31.3	15.6	16.3	-7%
Average kWh savings per lighting measure (kWh)	47.8	47.2	42.9	35.9	35.1	-7%

In contrast, the decline in average therm savings per participating gas account is primarily due to fewer participants weatherizing their home, which, on average, results in less therm savings per participant. Weatherization measures accounted for nearly half of the average therm savings per participant in the study period. Average participant gas savings from domestic hot water measures has increased slightly over the study period. Figure 3-7 illustrates these findings.

Figure 3-7. Average Gas Savings per Participant at the Portfolio Level



As with lighting above, we dug into what was driving lower weatherization savings for the average gas participant. We found that savings per weatherized customer increased between 2016 and 2020, but the number of participating customers weatherizing their home declined between 2016 and 2020. The average percent change in the number of gas customers with weatherization measures was -16% (Table 3-8).[21]

Table 3-8. Weatherization Measure Installed in Single-family Homes

	2016	2017	2018	2019	2020	Avg. annual % change, 2016–2020
Gas participants with weatherization measures	6,968	3,814	3,137	4,103	2,822	-16%
Savings per Weatherization Participant (therms)	65.2	92.1	90	75.9	76.5	6%

3.4.1.4 Examining Portfolio Participation through an Equity Lens

The team also investigated differences in cumulative participation rates at the portfolio level by equity-related demographic characteristics, including by household income, homeownership, and English proficiency.

[21] See the following recent impact evaluations for more details about the changes in gross and net weatherization savings: EWSF Impact & Process Evaluation (September 2020), http://rieermc.ri.gov/wp-content/uploads/2020/10/ng-ri-ewsf-impact-and-process-comprehensive-report_final_04sept2020.pdf and IESF Impact Evaluation (August 2018), http://rieermc.ri.gov/wp-content/uploads/2019/04/ng-ri-ies-impact-evaluation-report_final_30aug2018.pdf.

Figure 3-8 summarizes cumulative participation rates (2016–2020) among the electric and gas customer accounts by household income. This shows strong positive correlation in which as accounts' household income increases the likelihood of participation increases. The participation rate for the highest income category ($120,000 or more) is nearly twice the rates as for the lowest income category for electric and nearly three times the rates for gas.

Figure 3-8. Participation at the Portfolio Level by Income Level

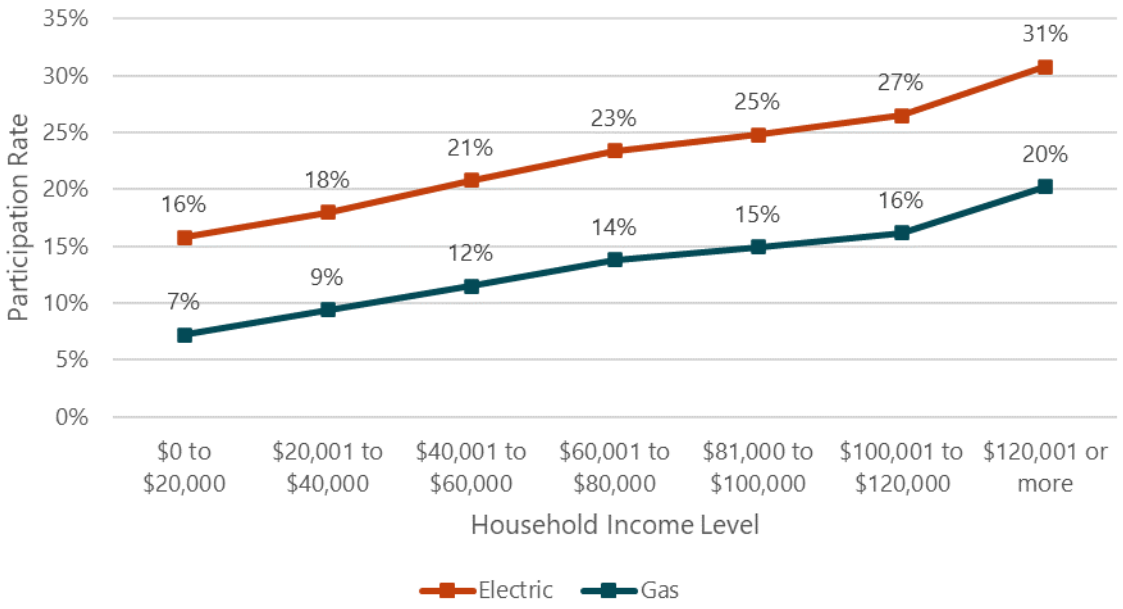


Figure 3-9 shows cumulative portfolio-level participation rates among active electric and gas accounts by homeownership. During the study period, homeowners had higher rates of participation than renters. Participation among renters with gas accounts was particularly low.

Figure 3-9. Participation at the Portfolio Level by Homeownership Status

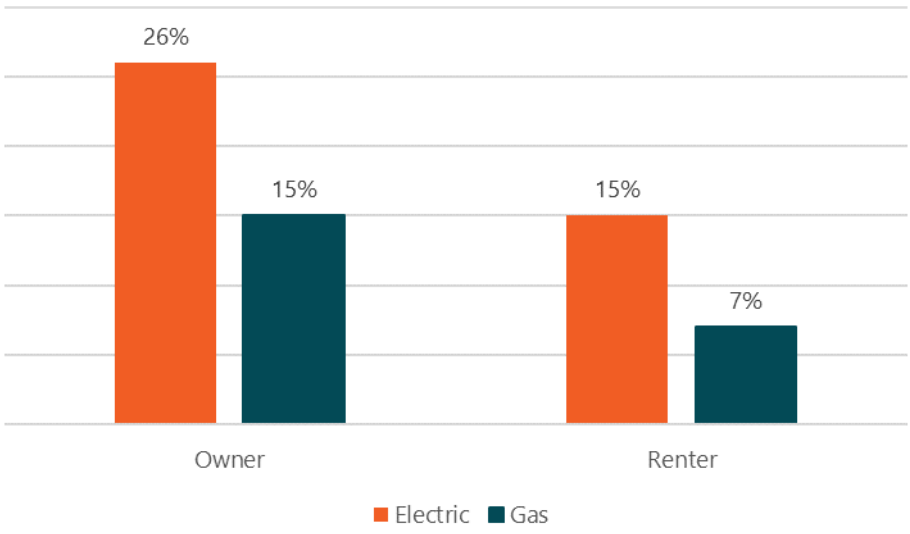
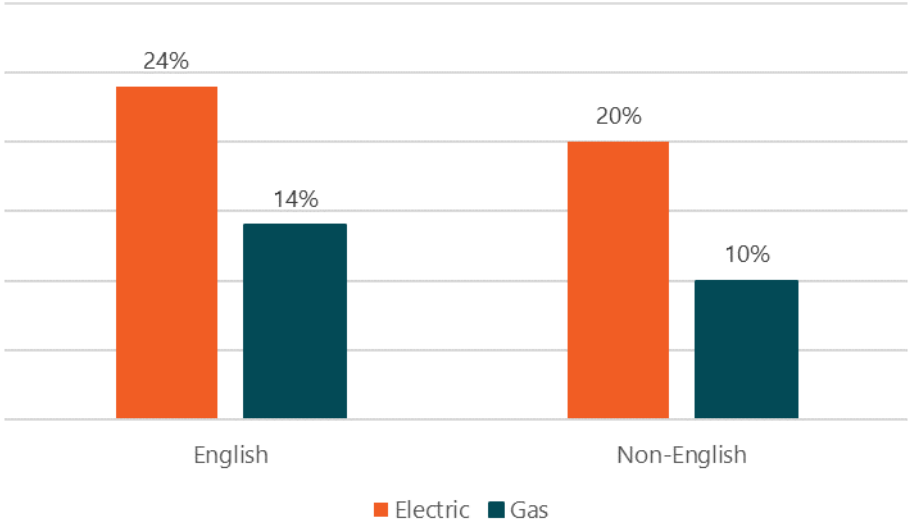


Figure 3-10 shows portfolio-level annual participation rates among active electric and gas accounts by language spoken. For both fuel types, participation among English-speaking households outpaced participation for non-English-speaking households.

Figure 3-10. Participation at the Portfolio Level by English Proficiency



Even though we found gaps in cumulative participation by these equity-related variables, we did not find changes in the participants mix by these variables over time during the study period.

The team also analyzed the SC ratios by these equity-related variables but did not find meaningful differences. Overall, this analysis suggests that the participation is higher among

higher income groups, homeowners, and English-speaking households, but among those who participate, the programs delivered equitable level of savings to the subgroups we examined.

3.4.2 New Construction and Renovation

Cadeo analyzed participation in the RNC program separately from the other NGRI programs because this program is fundamentally different from the others covered in this study. All other programs enroll existing electric and gas customers with the goal of retrofitting their homes. RNC is different: the program enrolls customers of new construction buildings, many of whom do not have an existing electric or gas account. Furthermore, there is no baseline energy consumption and therefore no savings-to-consumption ratio for this program. For these reasons, our team opted not to combine the analysis of retrofit programs and RNC.

Using the data provided by NGRI, the team assessed participation by individual units, project sites, renovation versus new construction, fuel type, multifamily versus single-family units, and income classification.

Between 2016 and 2020, the RNC program served 3,129 new construction and renovation units at 391 unique project sites according to the program data.[22] Program participation peaked in 2019, with 733 units served at 78 project sites, about two-thirds in multifamily buildings. Table 3-9 summarizes the annual and cumulative participation in the RNC program between 2016 and 2020.

[22] A project site is synonymous with a planned development site, facility, building, or group of buildings at the same address while units are synonymous with households.

Table 3-9. RNC Unit and Project Site Participation by Single-family and Multifamily[23]

		Annual					Cumulative
		2016	2017	2018	2019	2020	2016–2020
SF	Units	190	286	274	245	271	**1,263**
	Projects	78	89	60	66	96	**334**
MF	Units	343	412	288	488	424	**1,866**
	Projects	15	23	12	14	14	**70**
Total	**Units**	**533**	**698**	**562**	**733**	**695**	**3,129**
	Projects	**91**	**106**	**70**	**78**	**107**	**391**

3.4.2.1 Participation by New Construction vs. Renovation

Table 3-10 shows the number of new construction versus renovation units served by the RNC program in single-family and multifamily buildings. Most of the participating units (78%) were new construction projects.

Table 3-10. RNC Unit Participation (New and Renovation), Single-family and Multifamily, 2016–2020

		Annual					Cumulative
		2016	2017	2018	2019	2020	2016–2020
SF	New	179	265	272	189	262	**1,164**
	Renovation	11	21	2	56	9	**99**
MF	New	191	113	256	425	379	**1,290**
	Renovation	152	299	32	63	45	**576**
Total	**New**	**370**	**378**	**528**	**614**	**641**	**2,454**
	Renovation	**163**	**320**	**34**	**119**	**54**	**675**

3.4.2.2 Participation by Account Type

Table 3-11 summarizes the number of units served by account type (electric or gas). We counted units with both gas and electric accounts in both categories.

[23] There are some projects and units that were served in multiple years. The cumulative account is a unique count of the projects and units served. Additionally, some projects contained both single-family and MF units. This is why the total count of projects is greater than the sum of the single-family and MF projects.

Table 3-11. Units served in RNC by Account Type and Housing Type

		Annual					Cumulative
		2016	2017	2018	2019	2020	2016–2020
SF	Electric	185	253	176	193	261	**1,068**
	Gas	129	173	176	139	78	**692**
MF	Electric	242	327	248	414	416	**1,647**
	Gas	216	184	133	173	53	**681**
Total	**Electric**	**427**	**580**	**424**	**607**	**677**	**2,715**
	Gas	**345**	**357**	**309**	**312**	**131**	**1,373**

3.4.2.3 Participation of Low-Income Housing Units

Table 3-12 summarizes the number of low-income units versus standard-income units reached by the RNC program, including both renovation and new construction and by housing type. Between 2016 and 2020, the RNC program reached 1,201 low-income housing units. In single-family homes, low-income units made up 34% of participants, and in multifamily homes, low-income units made up 40% of participants. Between 2016 and 2020, 60% of the low-income units reached were in MF buildings.

Table 3-12. Participation of Affordable Housing Units in RNC Program

		Annual					Cumulative
		2016	2017	2018	2019	2020	2016–2020
SF	Low Income	30	73	134	94	98	**429**
	Standard Income	160	213	140	151	173	**834**
	% Low Income	16%	26%	49%	38%	36%	**34%**
MF	Low Income	204	139	62	115	267	**772**
	Standard Income	139	273	226	373	157	**1,094**
	% Low Income	59%	34%	22%	24%	63%	**40%**
Total	**Low Income**	**234**	**212**	**196**	**209**	**365**	**1,201**
	Standard Income	**299**	**486**	**366**	**524**	**330**	**1,928**
	% Low Income	**44%**	**30%**	**35%**	**29%**	**53%**	**38%**

3.4.2.4 Market Share of Single-family New Construction Participation

Table 3-13**Error! Reference source not found.** shows the market share of new construction, single-family units in Rhode Island that the program reached. We used building permit data for

new, authorized units from the US Census as a proxy for market size.[24] We did not include multifamily and renovation units in this analysis because market data for these subgroups was not available. Between 2016 and 2020, the program reached 20% of the new construction, single-family units in Rhode Island. That is a decrease from the 28% market share reported in the last participation analysis study completed by Navigant for 2009–2015. The share of single-family units reached through the RNC program peaked at 25% in 2017–2018.

Table 3-13. Single-family New Construction Market Share reached through the RNC Program, 2016–2020

	Annual					Cumulative
	2016	2017	2018	2019	2020	2016–2020
Participant Units	179	265	272	189	262	**1,164**
Housing Permits*	1,125	1,051	1,095	1,237	1,267	**5,775**
RNC Market Share	16%	25%	25%	15%	21%	**20%**

* Data from the Building Permits Survey, US Census Bureau New Privately Owned Housing Units Authorized. Includes permits for units in buildings with four units or less and does not include low-income housing data https://www.census.gov/construction/bps/stateannual.html.

[24] This number is an imperfect measurement of new construction activity for two reasons: 1) Not all approved permits result in a completed building, and 2) the permits approved in a particular year may not be a completed building until following years.

Section 4 Modeling Drivers of Participation

4.1 Key Takeaways

- Among active customers, household income is the strongest driver of participation, especially for homeowners.
- Income and age are positively correlated with participation among active accounts; however, older customers with higher incomes are less likely to participate.
- Home vintage is a meaningful participation driver; homes built between 1980 and 1999 have the highest participation rate.
- Homeownership and English proficiency of the household have limited impact on modeling participation once accounting for other variables.
- The effect of homeownership on participation is most pronounced for those aged 25–35.
- Larger homes (based on total square footage) are associated with more overall electric savings.
- More units in multifamily buildings are associated with a higher degree of savings, while additional units in attached single-family buildings are associated with a higher degree of savings.

4.2 Goals

Cadeo modeled the drivers of participation to assess how known customer and building characteristics influence participation. We modeled the demographic, housing, and behavioral characteristics that were most important in predicting both participation metrics discussed in detail in Section 0:

1. **Participation** of an eligible account
2. **SC ratio** of participating accounts

Both metrics offer complementary perspectives on customer engagement with National Grid's programs. Specifically, modeling the **Participation** metric, a straightforward binary indicator of participation in a program, provides insight into the account characteristics that lead to any level of participation—whether the customer is insulating their entire home or just purchasing a single piece of rebated equipment. In contrast, modeling the **SC ratio** offers insight into the

characteristics of accounts that take more comprehensive energy savings actions through National Grid's programs.

The role of equity-related variables in predicting participation, especially customers' income, primary language, and whether they own their home, is especially important. These variables provide insight that National Grid can use to better serve these customer types in pursuit of equity goals.

The participation models described in this section differ from the propensity models described in Section 5. The participation models particularly focused on identifying how specific customer variables influence participation, and the model generates interpretable coefficients that provide insight about how factors influence participation.

In contrast, the goal of the predictive propensity model used in Section 5 is to maximize predictive performance of the model and develop a score that most accurately predicts an account's likelihood of participation, even if the model provides less interpretable insight into *how* each customer variable effects the prediction.

4.3 Methodology

To develop models that highlight the specific drivers of participation, Cadeo engaged in a four-step process to prepare data, eliminate less useful variables, discover important interactions, and specify the final model:

1. **Value Transformations.** Manipulate raw variables to get better mathematical properties in preparation for the model.
2. **Variable Reduction.** Identify and remove variables that are not predictive or are highly correlated with other variables.
3. **Discovering Interactions.** Test two-way interaction between remaining variables to determine whether an interaction term is relevant in the final model.
4. **Model Specification.** Determine the final set of variable and most appropriate model type and fit.

4.3.1 Value Transformations

We converted numerical variables into categorical variables by collapsing all values to a bin label where a bin matches to a specific range of values. For example, when provided a specific build year, we matched it to an appropriate range from the following values: {pre-1900, 1900–1919, 1920–1939, 1940–1959, 1960–1979, 1980–1999, 2000–2020}. Figure variables and the binned values are shown in Figure 4-1.

Binning numerical values provides us with several advantages. First, binned values are robust against outliers that are binned with their nearest values. Second, binning allows us to overcome assumptions regarding linear relationships. We converted categorical variables so that individual variable values are split into different columns and coded as either being true (given a value

of 1) or not (value of 0) for each participant. Each binned value gets a coefficient with the model. Finally, since we discretized all variables and converted all values to binary values, we can directly compare and interpret model coefficients without additional transformations to account for scale.

4.3.2 Variable Reduction

To examine how key equity variables affect participation, we kept variables that represent income, homeownership, and primary language in the final specification. Other remaining variables were tested to determine whether they contributed a significant influence on participation.

We used correlation analysis as well as **LASSO regression** to identify related (i.e., colinear) variables that impact participation and select the most strongly predictive variable for inclusion in our model. For example, if we believe that the relative size of a multifamily building may impact participation, there are multiple related variables that are associated with size (square footage, number of units, number of stories) to potentially include in the model.

During the variable selection stage, we encoded values using a one-hot encoding scheme when converting categorical variables to binary columns. This means each value received its own column. This allowed us to include every value into a logistic regression model (with LASSO regularization). We treated each value independently. A variable was considered predictive if at least half of the values were significant predictors ($p<0.05$) or a single predictor was significant ($p<0.1$) but had a coefficient that was larger than most.

What is LASSO regression?

Coefficients for regression models are traditionally fit to minimize a sum of squared error (SSE) term measuring how model predictions deviate from actual known results.

Least Absolute Shrinkage and Selection Operator—or LASSO—adds a penalty based on the sum of the magnitudes of coefficients. The coefficients to a model are fit to minimize:

$$\left\{SSE + \sum_j |\beta_j|\right\}$$

It is one of several regularization techniques used to overcome overfitting and for reducing parameters.

LASSO is specifically valuable because it tends to force coefficients of less predictive variables to zero and is used to select predictive subsets of variables.

The LASSO method can be used with both logistic regression and linear regression.

If a single value of a variable proved to be predictive at this stage, we included all values of the variable in the next stage. For example, when looking at building vintage, if we discovered that only buildings built between 1980 and

2000 predicted participation, the model specification would continue to include all building vintages.

Technical note: We fit the model using the Python package scikit learn and its linear model Logistic Regression. We implemented LASSO regularization by using the l1 regularization and we performed a grid search to identify the optimal C parameter value of 10 (inverse of lambda or alpha parameter in other packages). See Appendix L for a full list of model parameters

4.3.3 Discovering Interactions

Once we identified key variables, we investigated interactions between these variables. We specified the model to include all two-way interactions between variables. Interactions identify special relationships between variables that have more than a simple linear relationship with participation. For example, we identified that the rate at which a household's income influences participation depends on whether one is a homeowner or a renter. Without interactions we could only look at the average influence of each variable without consideration for how they work together. We kept interactions in the final model specification if one combination of values was a significant predictor.

We included the following interactions in all the model specifications:

- Home Ownership x Income
- Home Ownership x Age
- Age x Income

4.3.4 Model Specification

Our final model specification included variables and variable interactions we found to be predictive of participation (or SC ratio) and used *dummy coding* to represent categorical values. This method requires one of the values of a representative variable to be the one against which we compare all others. We use this more concise representation so that we can look at variables as a whole and better interpret coefficients. Table 4-1 shows the final variables selected and the values associated with the reference customers. All coefficients show impact of the variable's value compared to the value in the reference customer. The interactions mentioned in the previous section are also part of the model specification. Education and race data were not available in the third-party data we purchased, however, the exclusion of these variables does not invalidate the model.

Table 4-1. Properties of the Reference Customer

Variable	Selection Criteria	Value
Age of Head of Household	LASSO	18–24
Household Income	Equity	Less than $20k per year
Primary Language	Equity	English

Home Ownership	Equity	Renter
Total building units	LASSO	1
Living Area	LASSO	1,000 to 1,500 square feet
Year Built	LASSO	<1900

While performing the variable reduction, variable significance was consistent across both participation and SC models except for the living area. We did not identify living area as significant for the participation model. Therefore, we did not include living area in the specification of the participation model but left it for the SC model. This is important as the degree of savings seems more highly contingent on building properties. Variables we evaluated but excluded from the final specification include marital status, average building gas consumption, total building area, average building electricity consumption, number of building stories, and census block group education level.

We used a logistic regression specification for our participation model over more accurate but less interpretable models, such the random forest model. The outputs of a logistic regression are between 0 and 1 and are well suited for binary outcomes. Alternatively, we used a standard linear regression to model the savings-to-consumption metric. Neither final model used LASSO regularization because we wanted to include all possible values of a variable in the final model specification to have a more complete understanding of how values change in regard to participation.

4.4 Results

4.4.1 Overall Participation (Portfolio Level – Both Fuels)

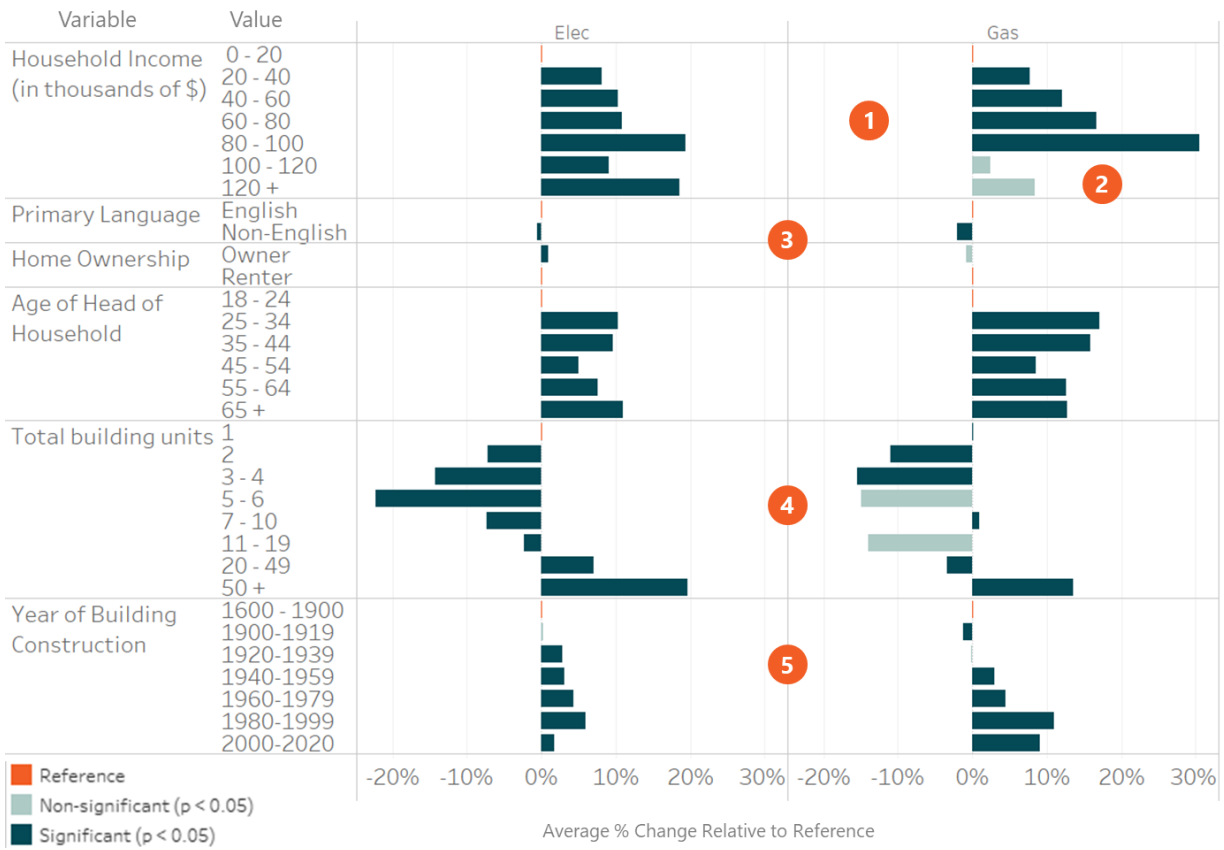
All key equity-related variables—homeownership, household income, and primary language—show clear differentiation of participation rate (See Appendix J for details). However, the modeling results show only specific values of each variable are significantly different from the reference values. Figure 4-1 and Figure 4-2 visually present the main results and interactive results of our participation model, respectively. We present model coefficients (representing log odds) in Appendix J. To aid interpretability, our team converted the logit model's coefficients to reflect the percent change in participation rate relative to the reference group. Please note the threshold for statistical significance (indicated in dark blue) was $p < 0.05$.

The results below specifically represent the contributions to predicting portfolio performance. The results impacted two programs that catered specifically to low-income participants slightly differently. Within the low-income programs, higher income did in fact predict higher performance up to a point (~$40k household income). Participants in such programs that have higher income levels may represent data errors.

Figure 4-2 shows the significant interactive effects that occur between variables. Some variable combinations result in increased participation, such as income and homeownership. Other variable interactions result in lower participation rates than when we consider those variables alone, which is the case for certain age groups and income.

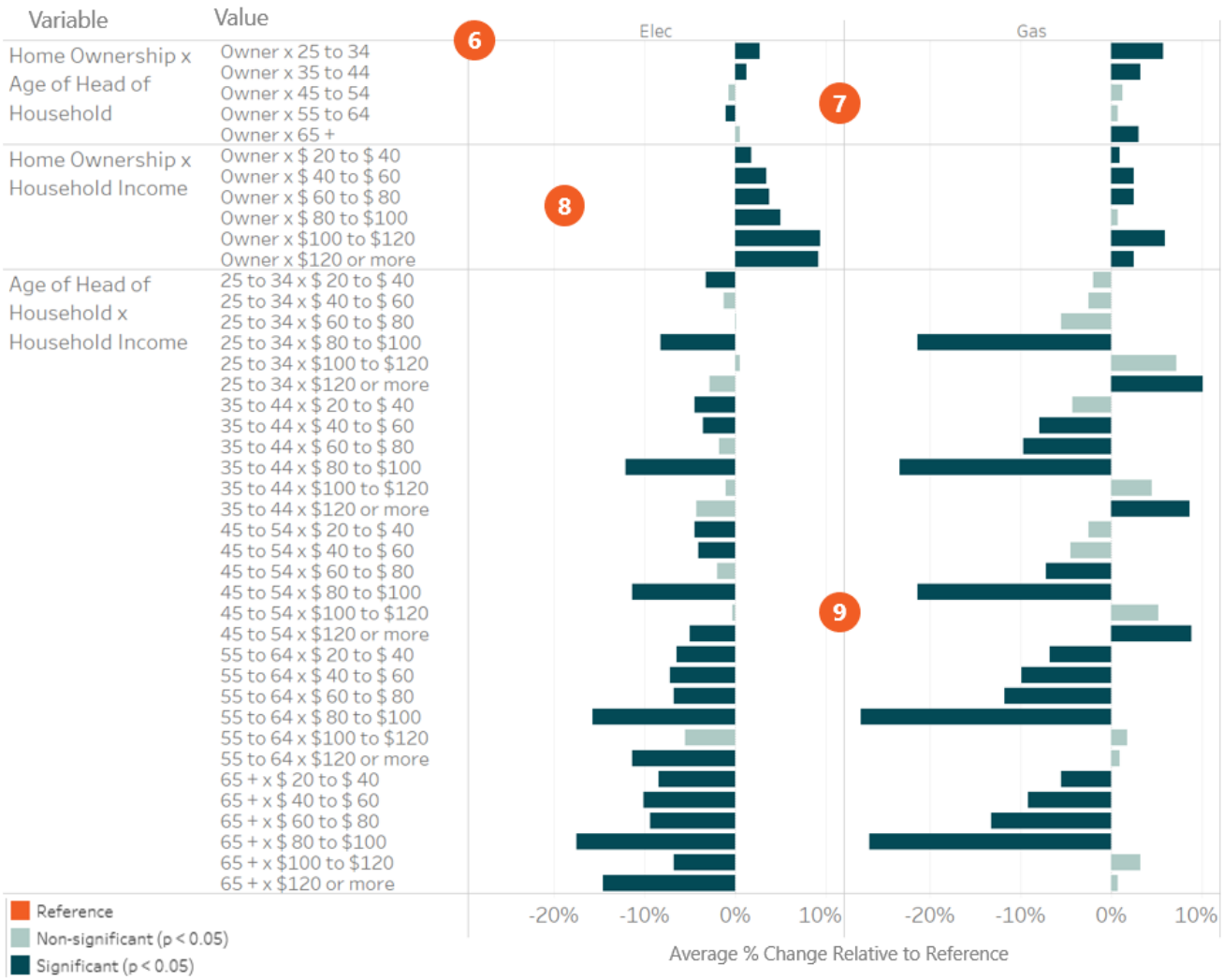
We do not present model results at the program-level. Overall programs tended to follow similar properties with slight differences. Participants in the programs that involve direct installation of measures tended to match the distribution of properties we observed in the portfolio participation with the exception that they were younger. MF programs had a higher proportion of renters.

Figure 4-1. Average Main Effects on Participation



1. Higher **household income** is associated with a higher participation rate.
2. However, for gas, once accounting for the other variables, **income's** direct influence on participation drops off once household income hit $100k.
3. **Homeownership** was not a significant predictor of participation on its own but did have important interactions with age and income.
4. Participation was **negatively associated with smaller multi-unit buildings** (i.e., non-detached SF homes under 20 units), but a strong predictor for large (50+ unit) buildings.
5. **Newer homes** had a much higher likelihood of participation; homes built after 1980 had highest participation rates.

Figure 4-2. Average Main Interactive Effects on Participation



6 Some variable combinations result in **increased participation rates** while others result in **lower participation rates than when those variables are considered alone.** These interactive effects are added on to the effects of each variable alone.

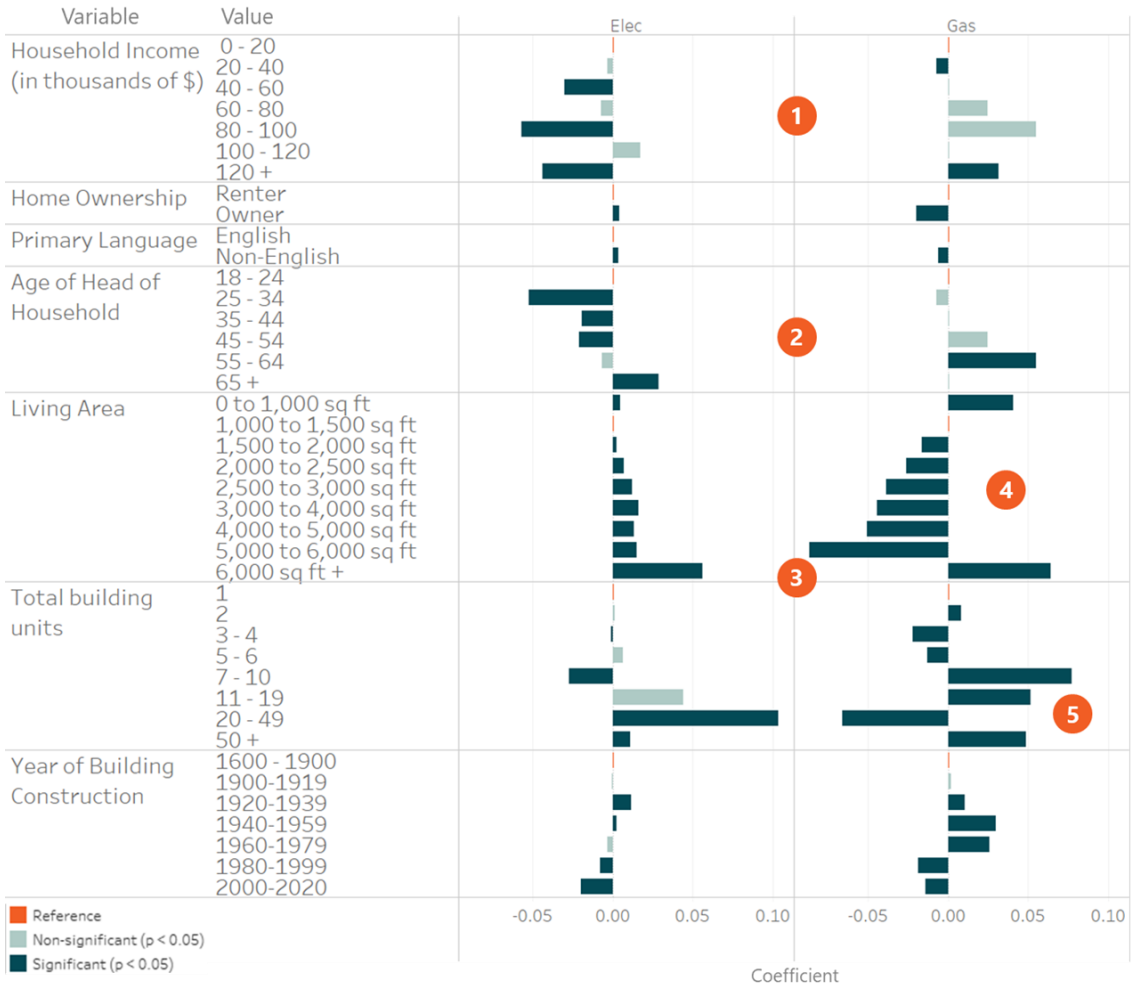
7 For example, while homeownership alone was not important, **homeownership increased participation for younger homeowners**, aged 25-34, with a diminishing increase in participation at higher ages until 65+.

8 **Homeowners had even higher participation rates in higher income households,** with the effect being more consistent for electric accounts.

9 Although **age** and **income** both had overall positive effects on participation, **the interaction between the two generally decreased the participation rate.** The exception is for gas accounts with household incomes greater that $100K, where the interaction increased participation rates for certain age groups.

4.4.2 SC Ratio (Portfolio Level – Both Fuels)

When predicting the degree of savings, Cadeo found that most personal demographic properties were not significant predictors for gas savings (Figure 4-3Figure 4-3. Average Main Effects on SC Ratio

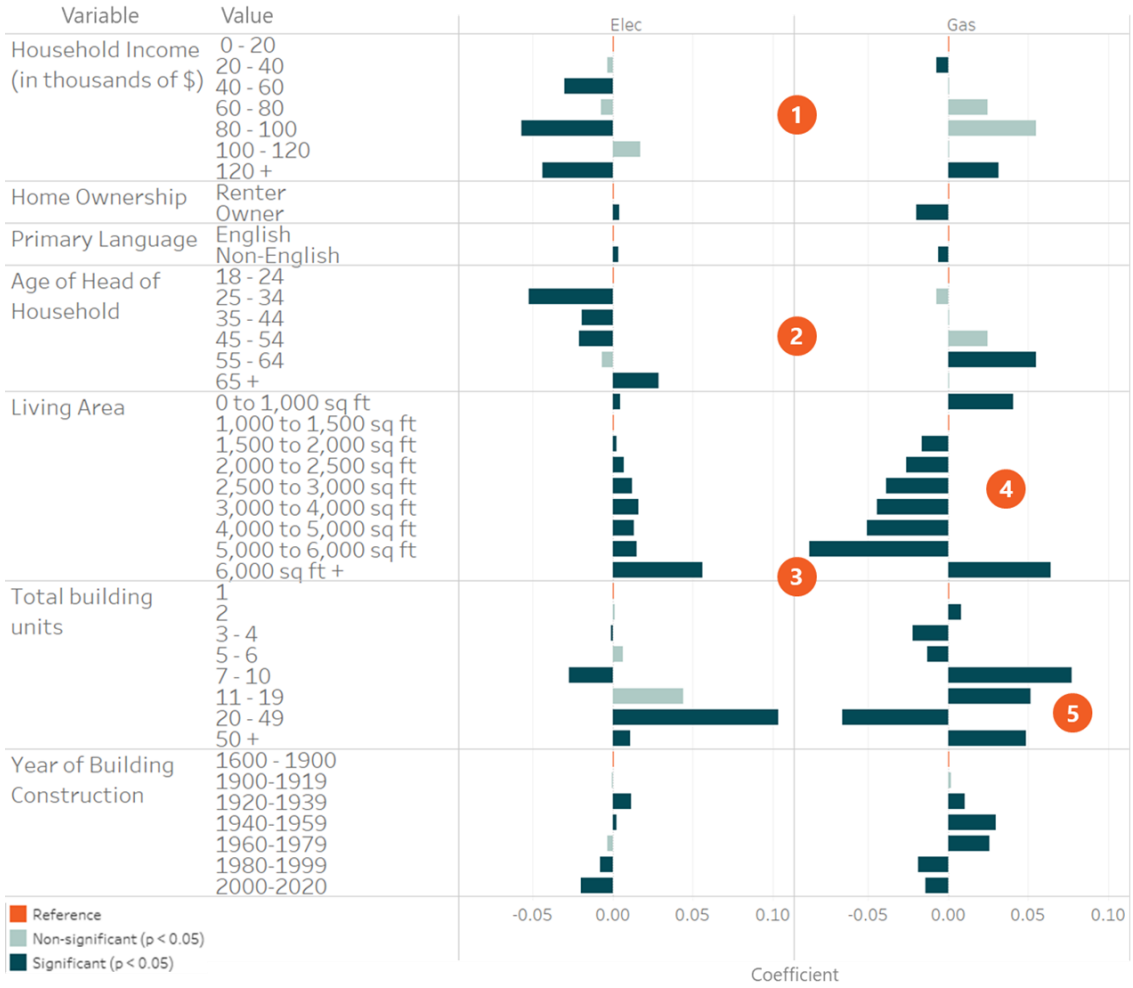


1. **Household Income** has a smaller impact overall without a clear consistent trend. This is likely due to collinearity with Living Area.
2. **Age** has a definite trend where older participants are expected to have greater savings
3. Building characteristics such as **Living Area** and **Total building units** have the largest influence on the degree of savings
4. Larger living areas generally decrease the expected gas savings while increasing the expected electric savings
5. Buildings units has some inconsistencies in the trends predicting gas savings-to-consumption. MF building size is highly heterogenous (in terms of building types/configurations). This may reflect underlying assumptions about how to distribute savings across MF buildings.

). Square footage was not adding explanatory power to predict participation, but we included it in the model of SC ratios. Building size in number of units, living area in square feet, and the year the building was constructed are the primary drivers of the SC ratio. Larger living area is associated with more overall electric savings with less overall gas savings.

Among single-family homes, a larger number of units (2–4) is associated with a decrease in the degree of savings, while more units in multifamily buildings is associated with a higher degree of savings.

Figure 4-3. Average Main Effects on SC Ratio



1. **Household Income** has a smaller impact overall without a clear consistent trend. This is likely due to collinearity with Living Area.
2. **Age** has a definite trend where older participants are expected to have greater savings
3. Building characteristics such as **Living Area** and **Total building units** have the largest influence on the degree of savings
4. Larger living areas generally decrease the expected gas savings while increasing the expected electric savings
5. Buildings units has some inconsistencies in the trends predicting gas savings-to-consumption. MF building size is highly heterogenous (in terms of building types/configurations). This may reflect underlying assumptions about how to distribute savings across MF buildings.

Section 5 Estimating Nonparticipants Propensity to Participate

5.1 Key Takeaways

- More than half of identified nonparticipants (56% or 178,762 customers) had a "low" propensity score (less than 0.3), which means they have different features (e.g., personal and housing characteristics) than historical participants. The low score suggests these nonparticipants are less likely to participate in a current National Grid efficiency offering and that engaging this group will require more significant programmatic redesign and/or different outreach efforts.
- Only 19% of identified nonparticipants (60,602 customers) had "high" modeled score (greater than 0.6), which suggests they are similar to participants and more likely to participant in current programs given more time, outreach, and/or more modest adjustments to programmatic offerings.

5.2 Goals

To support the concurrent Nonparticipant Barrier Study's survey sampling design, our team developed a list of nonparticipants, each with a modeled propensity score reflecting their likelihood of participating in an existing National Grid program (based on their similarity to known, recent participants). The Nonparticipant Barrier Study team used the nonparticipant sample and customer-specific propensity score as part of their stratified random sampling strategy, which sought to include the full range of nonparticipants and oversample the customers least likely to participate (i.e., those with the lowest propensity scores).

Nonparticipants with high scores are more likely to respond to current programs and their

How does this propensity model differ from the logit model in Section 4?

The model we developed in the previous section is a causal model used to identify the variables and variable values that drive (or cause) participation. In this section, we focused less on articulating causal relationships between variables and participation and more on maximizing the predictive power of random forest modeling to generate the most accurate propensity score for each account.

outreach. Those with low propensity scores are unlike current participants and will expected respond to different programs and/or messaging. The propensity does not tell us what sort of messaging low scoring participants might respond to, but the Nonparticipant Barrier Study is meant to explore such questions.

To best identify the nonparticipants, we modeled portfolio-level participation for electric and natural gas accounts separately. The model is trained on all known data for participants and nonparticipants. A model's score is its best guess as to whether any account has participated using only customer and account properties. Another interpretation is that the model score is a measure of how similar any account is to the "typical" participants observed in the past. We scored both participants and nonparticipants to measure the performance of the model, but only used the nonparticipant scores for follow-up in the nonparticipant study.

5.3 Methodology

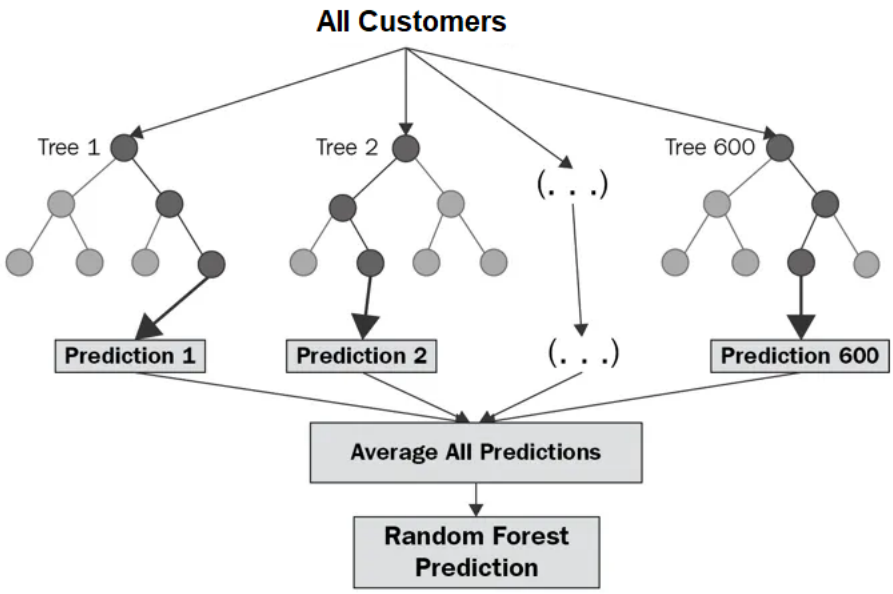
We used a **random forest model** to assign propensity scores to nonparticipants based on the features they share with participants. Below, we describe how we selected the variables for the model specification and how we measured performance to show improvement in prediction over the more interpretable model presented in Section 4.

5.3.1 About Random Forest Modeling

We chose to use a random forest model because random forests are one of a few models that consistently and robustly generate accurate predictive models without being dependent on much data preparation.[25] The reason is because a random forest model is an ensemble model where many (often thousands) of imperfect decision tree models are created. The forest's output prediction reflects an average of the individual decision tree decisions created as illustrated in Figure 5-1.

[25] In a comparison of logistic regression to random forest across over 19,000 machine learning datasets for random forest have superior performance 69% of the time. See R. Couronné, P. Probst, and A. L. Boulesteix, AL, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC Bioinformatics* 19, 270 (2018). https://doi.org/10.1186/s12859-018-2264-5.

Figure 5-1. Random Forest Model Predictions



To create each decision tree in a random forest, we select a random subset of customers to create the initial population and a random subset of variables for building the tree. Thus, each individual decision tree is a restricted model that provides a unique view of the problem, decoupled from the other trees. This allows the forest to build unique predictions based on complex, nonlinear relationships between the predictor variables by predicting the average of all the unique models.

Just as with a logistic regression model, customer accounts receive a score between 0 and 1, with 0 indicating the account is dissimilar to known participants and is unlikely to participate in the programs as they currently exist. A score of close to 1 indicates the account is similar to

Why don't all participants have a propensity of 1?

Predictive models, like the random forest specification our team used, are "trained" using datasets with participants (where participants have an actual "score" of 1) and nonparticipants (which have an actual "score" of 0).

During training, the model uses mathematical or computational methods to estimate a known outcome (dependent variable) using other values in the dataset (independent variables). The model score is a value from 0 to 1 that is an estimate of participation.

To measure model performance (i.e., the ability to correctly differentiate between participants and nonparticipants), we enter, as input, the properties of known participants and nonparticipants into the trained model. A strong model should score participants closer to 1 and nonparticipants closer to 0. However, we expect there to also be participants that score closer to 0 because they have properties that are more associated with the observed nonparticipants.

participants and has a higher estimated likelihood to participate in current programs.

It is possible for the model to assign both participants and nonparticipants scores across the entire range, but we expect the average score of participants to be higher than that of nonparticipants (See "*Why don't all participants have a propensity of 1?*" above for more details).

A "perfect" model would assign all nonparticipants a score of 0 and all participants a score of 1. However, such an idealized model is not useful as it would not differentiate between nonparticipants that are more or less like the observed participants. A model that can better identify actual participants is expected to be better at identifying the properties of participants observed in the population. As such, nonparticipants with a higher score are those customers with similar characteristics as known participants.

Technical note: We implemented the random forest model using the Python scikit learn package and the two ensemble models RandomForestClassifier and RandomForestRegressor. Scikit learn uses an optimized version of the classification and regression trees (CART) decision tree algorithm, minimizing Gini impurity with each branch selection. Please see Appendix L for model parameter details and a link to details regarding the specific tree model used.

5.3.2 Variable Selection

The team selected variables to include in the model based on National Grid interest in variables pertaining to questions of equity, our team's review of the correlation of variables (see Appendix I for more detail results of correlation analysis), and several iterations of running the predictive models.

Whenever possible, the team chose to model continuous variables instead of categorical variables. For example, rather than using building type which includes values like "duplex," "triplex," and "apartment," the team used the building unit count, which provides similar information on a more granular level. The algorithm will automatically find the best value to split the variable into two categories (above and below the split value). Any binning done beforehand limits the algorithm to identify the mathematically optimal value to split on. The benefits we saw from manually binning values for building the key-driver model in Section 4 are not realized by random forests.

The team also chose to exclude geographic information (count and census block) to eliminate collinearity between geographic areas and household characteristics.

Table 5-1 summarizes the final set of participant features that the team selected as predictive variables in the models as well as their data source.

Table 5-1. Account Characteristics included in Predictive models

Account Characteristic	Variable Type	Source
Account Age (Years since account opened)	Continuous	NGRI
Homeownership	Categorical	NGRI
Income	Categorical	NGRI
Marital Status	Categorical	NGRI
Customer Age	Categorical	NGRI
Building Age	Continuous	FA
Building Units	Continuous	FA
Living Area (Sq. Ft.)	Continuous	FA
Language (Non-English vs. English)	Categorical	NGRI
% non-white households in block group	Continuous	ACS
% with bachelor's degree or higher in block group	Continuous	ACS
Electricity consumption (kWh)*	Continuous	NGRI
Gas consumption*	Continuous	NGRI

* Calculated by Cadeo based on billing account data.

5.3.3 Model Performance

We conducted a comparison of the performance of our selected random forest model and the logistic regression model presented in Section 4. We created the logistic regression for its explanatory abilities. We created the random forest model to provide better performance in labeling participants.

We used a model performance metric called the "area under the curve" (AUC) statistic, which we describe below. It is a traditional metric used when a model provides a score for a problem with two possible outcomes such as participant or nonparticipant.

Any binary classification model that uses a score must set a score threshold to make a final prediction. Setting a threshold very low will increase the probability that participants are labeled as participants (known as the true positive rate [TPR] or Sensitivity) but will also increase the probability that nonparticipants are labeled as participants (known as the false positive rate [FPR] or 1-Specificity). The ideal model/threshold would have a high TPR and low FPR.

The Receiver Operating Characteristic (ROC) curve is a graph that shows the trade-off between TPR and FPR at all possible thresholds. The AUC is a measure of how well the model can find thresholds that get closer to that ideal. An AUC of 0.5 is as good as a random guess, and an AUC score of 1 can perfectly differentiate participants from nonparticipants. Another mathematically equivalent way of interpreting the AUC statistic is the probability assigned to a randomly selected participant will be a higher propensity score than a randomly selected nonparticipant.

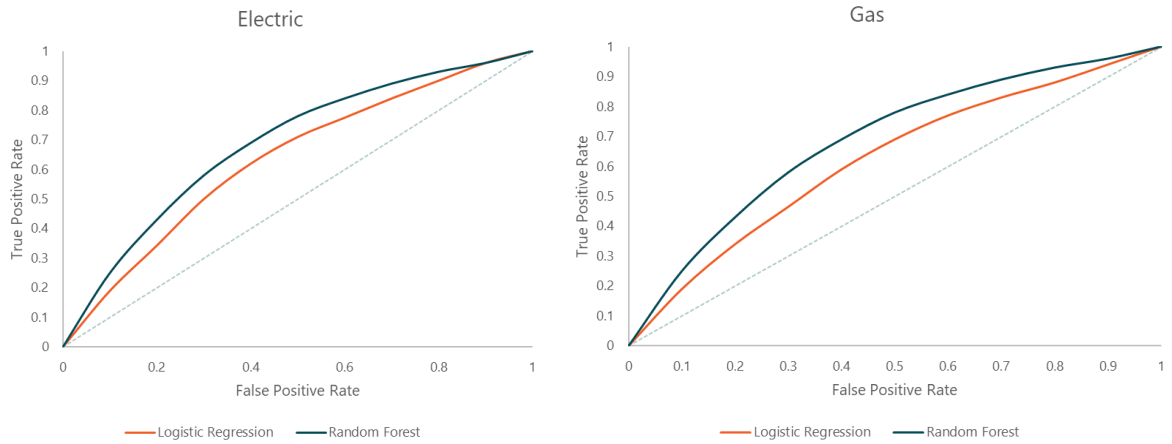
Figure 5-2 shows the ROC curve comparing the logistic regression and random forest models at predicting participation of electric and gas accounts. We use the AUC statistic to determine the most accurate model.

The AUC for the random forest model for electric participants was 0.777 compared to 0.760 for the linear regression model. The AUC for the random forest model for gas participants was 0.785 compared to 0.766 for the linear regression model. Although these numbers show a modest improvement, they are in line with what we would expect when comparing differing models trained on the same data. Larger improvements in performance will only come with more detailed data. A score above 0.7 shows good discrimination and a respectable score for the application. An AUC score of 0.85 would be quite high and unusual for predicting human behaviors such as participation in energy efficiency programs.

The AUC statistic has another mathematical interpretation that speaks to how well the model can identify members of two groups. An AUC of.785 means if we randomly selected on of the scored participants and randomly selected one of the scored nonparticipants from the data, there isa 78.5% chance that the participant's score will be higher than the score of the nonparticipant.

The model performance analysis shows us that the model scores differentiate participants from nonparticipants, and that the random forest model does so better than the linear regression model presented in Section 4.

Figure 5-2. Electric and Gas Portfolio Participation Propensity Score Model Comparison – ROC Curve



5.4 Results

Figure 5-3 (electric) and Figure 5-4 (gas) show the distribution of propensity scores for participants and nonparticipants at the portfolio-level (i.e., the estimate of participating in any National Grid residential program they are eligible for). As evident in both figures, nonparticipants (in green) tend toward lower scores, while the participants (in orange) tended

toward higher scores although—as expected—both participant types span a wide range of propensity values. The histograms of nonparticipants scores looks as if they are bimodal (have two peaks). This shows the ability of the model to separate those that look more like the participants from those that are very dissimilar.

Because the random forest model is more difficult to interpret how it is arriving at any prediction, we defer to Section 4 for understanding the characteristics driving participation. The Nonparticipant Barrier Study team used the nonparticipant sample and customer-specific propensity score as part of their stratified random sampling strategy, which sought to include the full range of nonparticipants and oversample the customers least likely to participate. The results of the Nonparticipant Barrier Study will inform strategies for reducing barriers for customers in the lower score range.

Figure 5-3. Propensity Score Distribution of Electric Participants and Nonparticipants

Figure 5-4. Propensity Score Distribution of Gas Participants and Nonparticipants

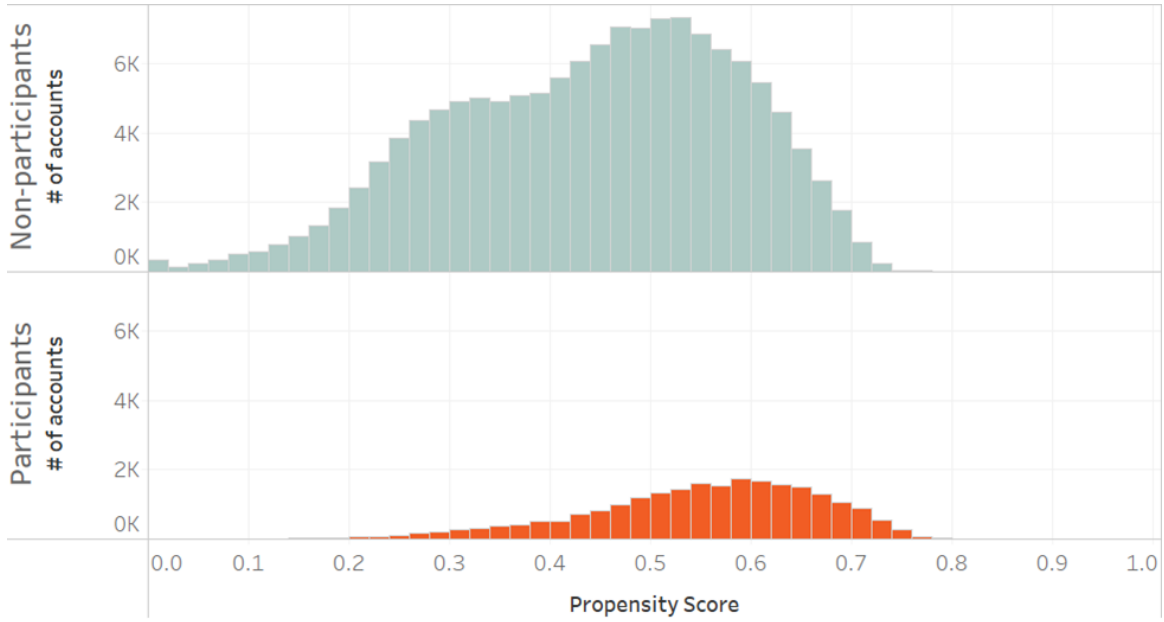


Table 5-2 takes the customer-specific propensity scores shown in distribution above and puts them into order from the lowest to highest scores, which allows our team to look at ranked scores (i.e., what are the scores associated with lowest, middle, and high scores?).

Scores in the lowest range received a propensity score below 0.3. A low score suggests the customer has very different features than observed in past program participants. These individuals are the largest opportunity, but likely require the largest changes to improve participation.

Scores in the range of 0.3–0.6 indicate customers are neither very similar nor very dissimilar to known participants. In fact, the model had difficulty discriminating participants from nonparticipants in this range. Across the population, we observed that customers with features that yielded these scores did not participate any more or less than the overall population at large and so the model is not adding any useful information.

Customer with scores above 0.6 show a strong similarity to known participants. These nonparticipating customers share the same properties as many participants but have not actually participated. These individuals are the most likely to participate in the existing programs but have some barrier stopping them.

Table 5-2. Model Results

Score	Average Score	Nonparticipating Electric and Gas Accounts	
Low scores (0.0–0.3) are associated with accounts that have features less commonly observed in participants.	0.22	178,762 (**56%**)	More than half of nonparticipating accounts look very different from past participants and least likely to participate in programs as designed, marketed, and delivered today.
Middle scores (0.3–0.6) are assigned to accounts that are equally similar to participants and nonparticipants.	0.46	78,322 (**25%**)	The reasons for nonparticipation are likely due to factors for which we do not have data.
High Scores (0.6–1.0) are most like the observed participants.	0.65	60,602 (**19%**)	Less than one-fifth of nonparticipating accounts are more likely to participate in a current program; these accounts may just need more time or face lesser barriers to participation.

Other potential use cases of the propensity scores include customizing marketing messages and channels to reflect the propensity scores, and integration of the propensity scores to National Grid's Customer Information System (CIS) to trigger appropriate ads or messaging when customers log into their online account.

Section 6 Identifying Multifamily Customers

6.1 Key Takeaways

- Our algorithm identified **24,012 multifamily buildings** in Rhode Island, which equates to approximately 7% of residential buildings in the state.
- We also estimate that **19%** and **16%** of National Grid's electric and natural gas accounts are associated with multifamily buildings.
- Our estimated number of accounts associated with multifamily buildings is very similar to the estimate from the prior Participation Study.
- Our most accurate algorithm specification (i.e., the one that most consistently identified known single-family and multifamily program participants as single family and multifamily) prioritizes **tax parcel land use codes** and the number of unique electric or gas accounts associated with each building.

6.2 Goals

We had two goals when identifying which National Grid residential customer accounts are associated with multifamily buildings:

- Develop a **data-driven algorithm** that accurately classifies each residential customer account as either "single-family" (part of a building with 1–4 units) or "multifamily" (part of a building with 5+ units) in a manner that is consistently with National Grid's program definition.
- Create a **comprehensive database of multifamily buildings** that includes building characteristics and program participation status.

6.3 Methodology

6.3.1 Defining a Building

Our team first sought to define a "building" by connecting all accounts (one or more) associated with a given building.[26] To do so, we leveraged the key building field (keybdg) provided by National Grid to serve as a unique building identifier. National Grid created this internal key,

[26] We use the term "building" casually; the definition could include accounts in one or more structures associated with a single property.

which uses accounts' ZIP code and Line 1 of the address (leaving out apartment and unit numbers) as part of a previous internal effort.

We found the keybdg field worked well for identifying most unique buildings; however, we encountered a handful of instances when it did not. These instances included:

- **Inconsistent Addresses.** For example, the following addresses – "10A Coogan Ct" and "10 Coogan Ct Unit B" – are clearly associated but have different keybdg values because of how National Grid's system lists each address.
- **Mobile Homes.** Using keybdg was also problematic for mobile homes in a park associated with a single address, as the mobile homes would appear as one building even though they are, in some cases, independently owned single-family homes.
- **Condominiums.** Some condominiums within the same building have unique addresses (rather than being differentiated with a unit number). In these cases, we assigned a unique keybdg value for each of the condominium units in the same building.[27]

Despite these issues,[28] we determined keybdg was the best available building identifier as it aligned with National Grid's existing building definition.[29]

Once the team defined buildings, we aggregated information associated all accounts with the same keybdg. This enabled us to assess participation at the building level and calculate building-level consumption metrics.

6.3.2 Building Classification Algorithm Development

Based on the aggregated building-level data, our team developed an algorithm to systematically identify multifamily buildings, i.e., a building or property with 5+ units.

Knowing that no algorithm will perfectly identify all multifamily buildings, we sought to develop an algorithm that minimized two key types of errors:

- **False Positives**, i.e., falsely identifying single-family homes as multifamily buildings.
- **False Negatives,** i.e., failing to correctly identify multifamily buildings as multifamily buildings.

These two types of errors are at odds. For example, an algorithm that limits the number of false positives will result in a smaller (albeit more accurate) list of multifamily accounts. However, that

[27] Appendix K provides examples of condominium buildings found in the database and discusses the challenges we faced in properly classifying them into single or multifamily buildings.
[28] Although such issues can be easy to locate by finding conflicting information among data sources and manually evaluating some properties, we are unable to quantify each of these issues without a manual survey of a random subset of the population.
[29] Other identifiers explored involved conditionally using address components based on other field values. Although we could fix specific issues, such rules often introduced other complexities with no overall increase of accuracy.

same algorithm, by virtue of its more conservative approach to identifying multifamily accounts, will also result in more false negatives.

For guidance on which type of error to minimize when selecting our final multifamily identification algorithm, our team met with National Grid's evaluation team and multifamily program managers. They acknowledged the inverse relationship between the two types of errors and the importance of each regarding their expected use cases for the database (i.e., reporting historic multifamily participation rates as well as prospectively targeting future multifamily participants).

They instructed our team to balance these competing factors (i.e., minimizing both sources of error) while prioritizing a specification that would enable effective targeted outreach by the multifamily program. This led our team to seek an algorithm that would result in a multifamily census list that exhaustively covers multifamily accounts even if it includes some misclassified single-family accounts.

To do so, our team employed an iterative process to develop our multifamily identification algorithm, testing a wide variety of algorithm specifications that relied on a range of data fields and prioritization strategies.

To assess the efficacy of each iteration and identify the "best" algorithm (i.e., the one that optimally minimized both errors), our team leveraged National Grid's historical program data. This is possible because participation in a National Grid single-family or multifamily program is evidence that a given building is either single-family or multifamily. Accordingly, we assessed each iteration by comparing how closely that algorithm's classification results matched the single-family and multifamily program participation records.

According to National Grid's records, 24,699 and 65,572 unique accounts participated in National Grid's multifamily and single-family programs, respectively, between 2016 and 2020. These values include both the EnergyWise and Income-Eligible offerings for both sectors. Table 6-1 includes the frequencies of building classifications by actual participation (i.e., program tracking data) and the algorithm classification label.

Our most successful and final algorithm, outlined in Section 6.4, correctly identified 85% (20,937 of 24,699) of accounts in the historical multifamily participation data as multifamily. Another important perspective is that 89% (20,937 of 23,571) of the participant accounts we identified as multifamily were indeed multifamily.

As noted above, there is a trade-off when considering these two perspectives on accuracy (i.e., reducing one can increase the other). We selected this algorithm because it is consistent with the guidance provided by National Grid and optimally maximized these identification percentages and minimized the two types of errors described above. Specifically, our final algorithm had an 11% incidence of falsely identifying positives (2.634/23,571) and a 15%

incidence of falsely identifying negatives (3,762/24,699). Overall, the algorithm produced an underestimation of multifamily accounts by 5% [(23,571-24,699)/ 24,699].[30]

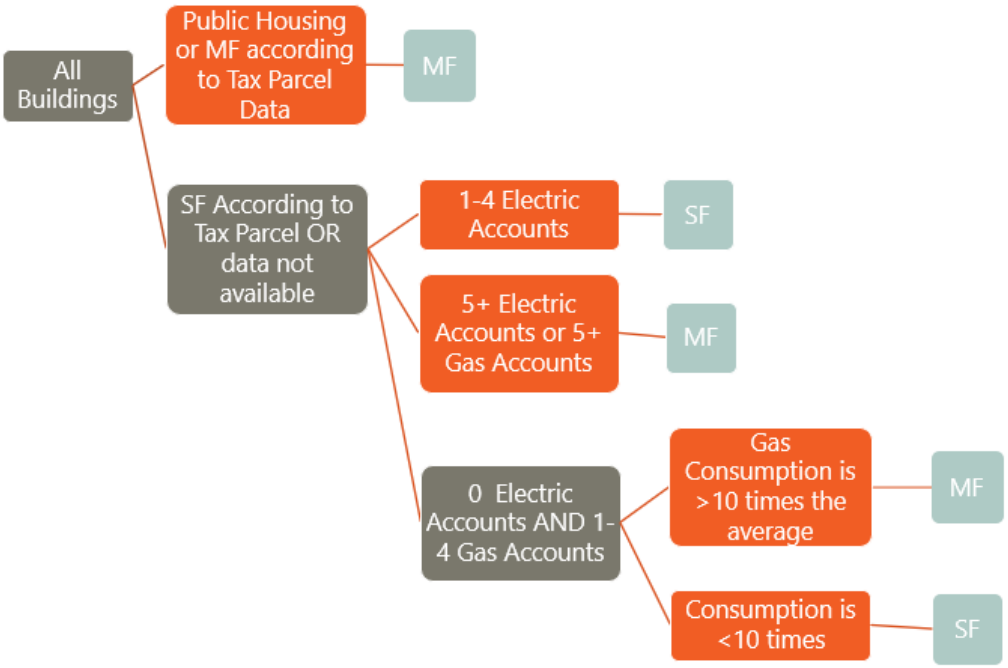
Table 6-1. Comparison of Algorithm Classification and Program Participation

		ALGORITHM		
		Single-Family	Multifamily	Program Total
PROGRAM	Single-Family	62,938	2,634	65,572
	Multifamily	3,762	20,937	24,699
	Algorithm Total	66,700	23,571	90,271

Figure 6-1 visually summarizes our final and most accurate multifamily identification algorithm. The algorithm prioritized the land use field in the tax parcel data to identify multifamily accounts.[31] When tax parcel data was not available or indicated single-family land use, the algorithm assessed the validity of the single-family land use code by considering the number of unique electric or gas accounts associated with the building, classifying any buildings with 1–4 electric accounts as single family and more than four electric or gas accounts as multifamily. Finally, for those accounts not identified as multifamily through other criteria, the algorithm considered total usage for accounts in buildings with gas consumption at least 10 times the average identified as multifamily.[32]

[30] The recent Massachusetts' Multifamily Buildings Census Study (https://ma-eeac.org/wp-content/uploads/RES43-Final-Report-2019-05-31.pdf) did not explicitly state an intention to balance false positives and false negatives in developing and determining their MF building classification approach. However, that study's report indicates their algorithm correctly identified historical MF participants as being MF 59% of the time (41% incidence rate of false negative), which is relatively like our identification rate of 78% (22% incidence rate of false negative).
[31] We prioritize land use over the number of calculated units to overcome two challenges that greatly impeded our ability to match how program participants had been labeled. We highlight both issues in Appendix K. In one case, separate units of a MF building have a different street number in their address and are seen as single-unit buildings. The other actual single-unit buildings that still fall under a single managed entity.
[32] There is a strong overlap in the distribution of gas consumption among previous participants that programs have labeled as single-family or MF. We selected the multiplier of 10 as a conservative method, suggesting that gas account is master-metered. The multiplier was not optimized for accuracy.

Figure 6-1. Final MF Identification Algorithm



6.4 Results

In total, this algorithm identified 24,012 unique multifamily buildings in Rhode Island making up approximate 7% of the residential buildings in the state.

Table 6-2 summarizes this information, as well as the number of National Grid electric and gas accounts by housing type. Approximately 19% of the electric accounts are in multifamily buildings. We see a lower percentage of gas accounts among multifamily buildings (16%).[33] This is likely because master metering is more common for gas. The previous NGRI Participation Study's estimates of multifamily accounts are very similar (19.1% for electric and 16.8% for gas accounts; percent of multifamily buildings is unavailable). NGRI indicated that electric accounts are rarely master metered; therefore, electric accounts provide us with the best estimate of the number of housing units.

[33] These results are comparable to the previous Rhode Island 2009-2015 Participation Study, which identified 19% of electric accounts and 16% of gas accounts as associated with multifamily buildings.

Table 6-2. Summary of Identified Residential Buildings and Active Accounts

	Single-Family (≤4 units)		Multifamily (5+ units)	
	Count	**Percent**	**Count**	**Percent**
Buildings	321,178	93%	24,012	7%
Electric accounts	360,328	81%	83,409	19%
Gas accounts	207,470	84%	40,649	16%

As part of this study's planning process, stakeholders requested that our team further classify multifamily buildings as one of two possibilities:

- An apartment building (managed by a single entity) or condominium (multiple owners of different housing units)
- Public or subsidized housing

Table 6-3 breaks down the identified multifamily buildings and units by these ownership types. Our team determined condominium or apartment for nonpublic housing buildings using the tax parcel land use code field (a First American code comparable to TXT_COUNTY_USE1_TRW).[34] We identified a building as public housing when the term "Housing" appeared in the customer's name field. Customers without land use code data and without "Housing" in the customer's name field appear in the table as "Unclassified buildings."[35] The previous NGRI Participation Study did not provide building-level data.

Table 6-3. Distribution of Multifamily Buildings by Ownership Type (Apartments, Condos, Public Housing)

Ownership Type	Number of Buildings
Apartment buildings	6,957
Condominium buildings	9,230
Unclassified buildings	7,238
Public housing	587
Total	**24,012**

[34] Land use codes indicating ownership (primarily "condominium," but also "townhouse," "cluster home," "cooperative," "row house," and "single-family") were labeled as condominiums. We considered land use codes indicating duplexes, triplexes, quadplexes, and apartments of different sizes as apartments.

[35] The unclassified buildings are likely distributed to have similar ratio of apartment to condominium as those known. Other indicators such as the NGRI provided premise type or the own/rent indicator are not as highly correlated with multifamily versus single-family participation and are a less reliable indicator.

Appendix A Glossary of Terms

Terms	Definitions
Account	An account represents a single electric or gas meter. This is the basic level of granularity provided by National Grid.
Account participation	Account level binary participation status, which indicates whether an account is a participant or nonparticipant of a program in which the account is eligible to participate.
Building	This is meant to identify a single building. Accounts roll up into buildings. A building is uniquely identified by the zip code and street address not including unit or apartment numbers (keybdg). Buildings contain 1 or more units. National Grid Rhode Island programs do track a BUILDING_ID, but this is not available for nonparticipants accounts.
Building participation	Building level binary participation status, which indicates whether any unit in the building is a participant.
Cumulative participation	Cumulative participation metrics estimate the number of unique participating accounts or buildings in any year during the evaluation period of interest.
Customer	This is the person tied to the account. In multiunit buildings, this may be the same person for all units. We do not conduct any analysis at the customer level.
Facility	This is a single property that may include multiple buildings. By definition, buildings rolled up into facilities. Some National Grid program data track a FACILITY_ID, but this is not available for any nonparticipant accounts.
Multifamily building	National Grid programs define a multifamily building as a building with five or more units.
Propensity to participate	A statistical scorecard that is used to predict program participation of customers. Propensity models are used to identify those most likely to respond to program offerings.
Savings-to-consumption (SC) ratio	Ratio of savings achieved and consumption of participants, which indicates the impact of participation. SC ratios are available at account level as well as building level.
Single-family building	National Grid programs define a single-family building as a building with 1-4 units.
Unit	An individual dwelling unit. We uniquely identify a unit using the zip code and street address including apartment or unit and apartment numbers. A unit contains 1 or more accounts.

Appendix B Description of Residential Energy Efficiency Programs Covered by This Study

Program	Program Description
EnergyWise Single-Family (EWSF)	Residents in single-family homes that do not qualify for income eligible services can enroll in this program to receive a home energy assessment, direct installation of measures such as lighting, showerheads, and power strips, and rebates or loans for other recommended measures. Program data included such measure types as lighting, weatherization, appliances and plug load, domestic hot water heaters, and thermostats and other controls.
Income Eligible Services - Single-Family (IESF)	Residents in in single-family homes that qualify for low-income rates can enroll in this program to receive no-cost home energy assessments, and direct installation of some measures. This program also provides weatherization services and some space and water heating system replacement at no charge. Program data included such measure types as lighting, weatherization, appliances and plug load, domestic hot water heaters, HVAC equipment, thermostats, and other controls.
EnergyWise Multifamily (EWSF)	Like the single-family programs, this program provides incentives for installation of weatherization, space and water heating and cooling systems, lighting, and appliances in multifamily homes. In the multifamily programs, the program is implemented in coordination with a property owner or manager. Program data included such measure types as lighting, weatherization, appliances and plug load, domestic hot water heaters, and thermostats and other controls.
Income Eligible Services – Multifamily (IEMF)	This program operates similarly to the EnergyWise program but serves buildings that are primarily made up of income eligible accounts or located in a building that in considered public housing. Program data included such measure types as lighting, weatherization, appliances and plug load, domestic hot water heaters, and thermostats and other controls.
Natural Gas Heating and Water Heating (NGWH)	Residents with natural gas space and water heating can participate in this program to receive rebates on equipment and controls that reduce gas use in the home. This includes rebates on gas water heaters and smart thermostats for gas heating systems. Program data included such measure types as thermostats, furnace upgrades, domestic hot water heaters, and ECM pumps.
ENERGY STAR HVAC (ESHVAC)	Residents can participate in this program to receive rebates on efficient electric heating and cooling equipment and controls including heat pump water heaters. The program also provides rebates for ECM pumps, central air conditioning, and smart thermostats. Program data included such measure types as thermostats and other controls, heat pumps, central air

Program	Program Description
	conditioning, domestic hot water heaters, ECM pumps, HVAC tune-up and downsizing.
ENERGY STAR Products (ESProducts)	Residents can enroll in this program to receive rebates on ENERGY STAR certified appliances and other products. Online marketplace and refrigerator recycling were also included when billing account number was available. Program data included such measure types as appliance recycling, HVAC equipment, appliances, and other plug loads.
Connected Solutions (CS)	Residents can enroll eligible thermostats, solar systems, and batteries and get paid to reduce electric use during periods of high electricity demand.
Residential New Construction (RNC)	RNC provides financial incentives and no-cost education, training and technical support to builders and homeowners to help single and multifamily new construction and major renovation projects meet high energy performance standards.

Source: National Grid RI Energy Savings Program Website, September 2021, https://www.nationalgridus.com/RI-Home/Energy-Saving-Programs/

Appendix C Program-Specific Annual Savings by Fuel by Year

Figure C-1. EWSF Program Annual Savings by Fuel by Year (in MMBTUs)

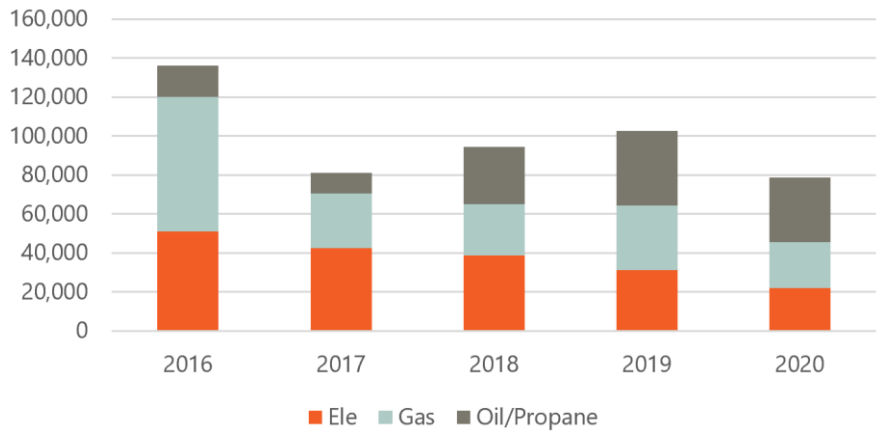


Figure C-2. IESF Program Annual Savings by Fuel by Year (in MMBTUs)

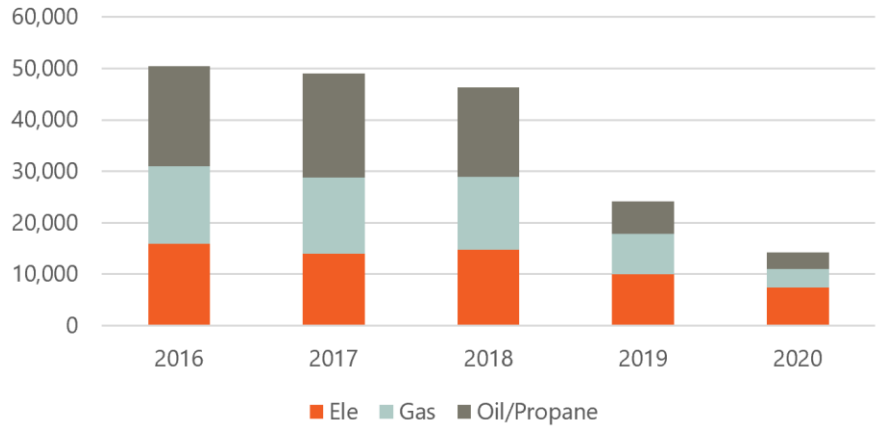


Figure C-3. EWMF Program Annual Savings by Fuel by Year (in MMBTUs)

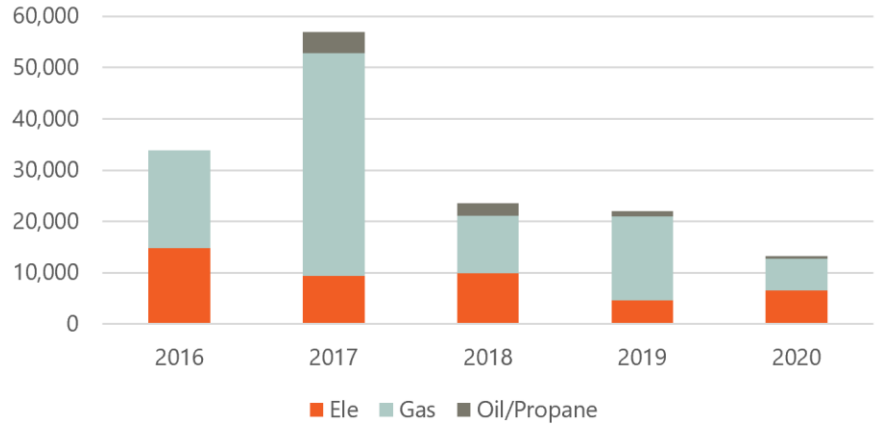


Figure C-4. IEMF Program Annual Savings by Fuel by Year (in MMBTUs)

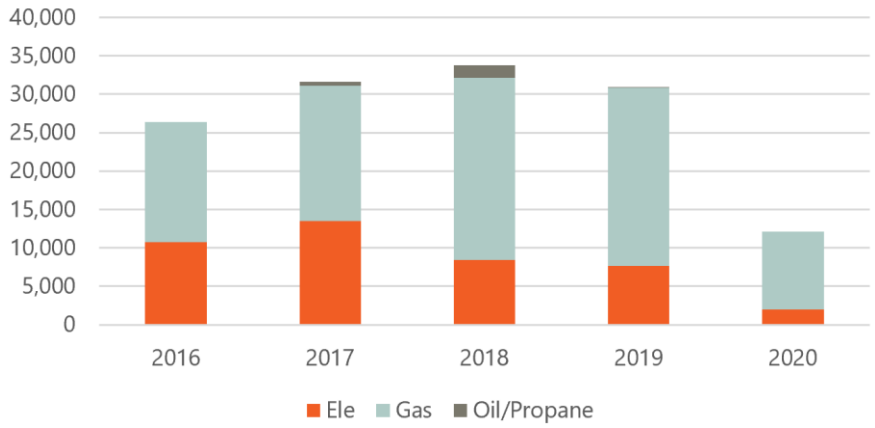


Figure C-5. ESHVAC & NGWH Program Annual Savings by Fuel by Year (in MMBTUs)

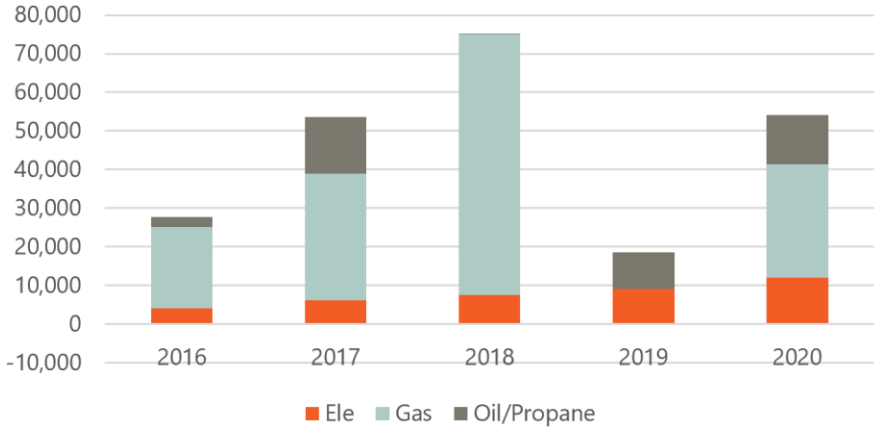


Figure C-6. ESProducts Program Annual Savings by Fuel by Year (in MMBTUs)

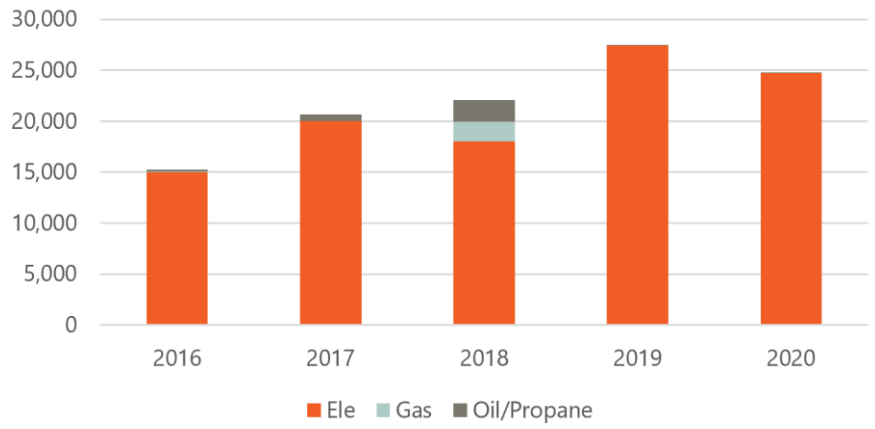


Figure C-7. RNC Program Annual Savings by Fuel by Year (in MMBTUs)

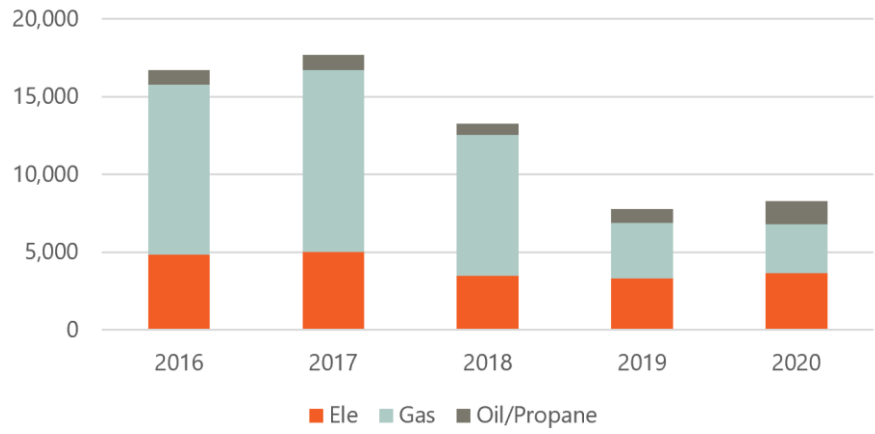


Figure C-8. HER Program Annual Savings by Fuel by Year (in MMBTUs)

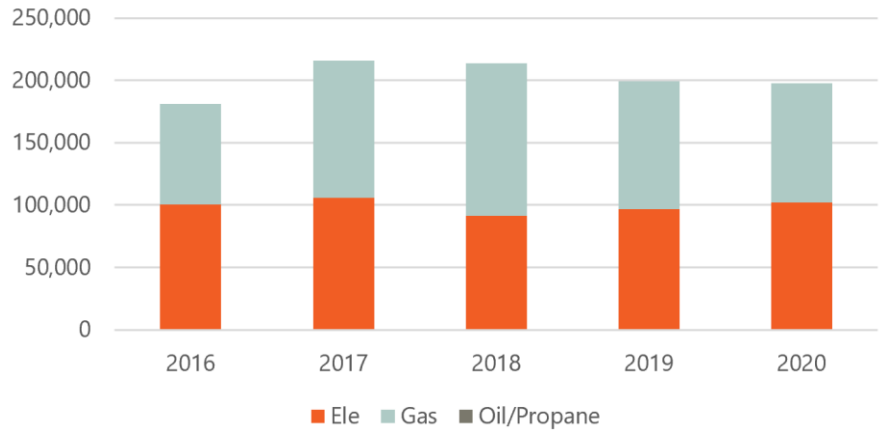
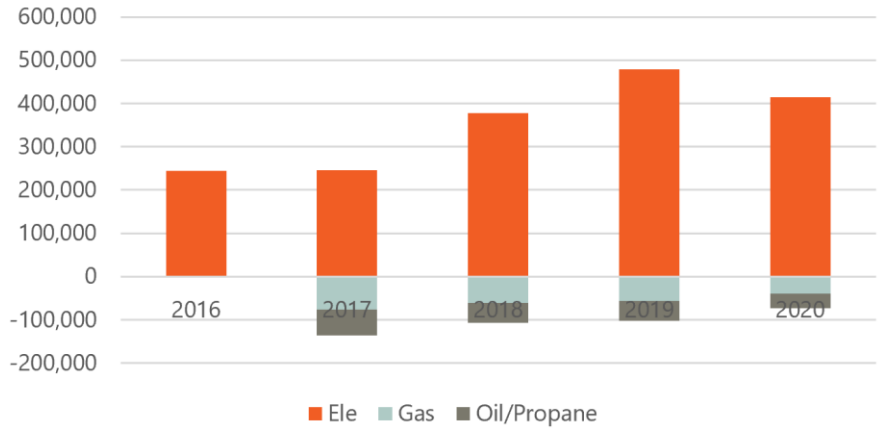


Figure C-9. ResLighting Program Annual Savings by Fuel by Year (in MMBTUs)



Appendix D Details of Program Eligibility Criteria

The table below provides a summary of the data fields used to determine account eligibility for each program.

Table D-1. Program Eligibility Criteria

Program	Eligibility Criteria
EnergyWise Single-Family	• MFSF_True = "SF" • flg_lowincome_acct = 0 • OR program participant between 2016 to 2020 more recently then IESF
Income Eligible Services - Single-Family	• MFSF_True= "SF" • flg_lowincome_acct = 1 • OR program participant between 2016 to 2020 more recently then EWSF • OR household income (D270_HOUSEHOLD_INCOME_MID) < 60% of AMI based on number of members in household (D263_HOUSEHOLD_MEMBER_COUNT)*
EnergyWise Multifamily	• MFSF_True= "MF" • flg_lowincome_acct = 0 AND sum(flg_low_income_acct)/# of accounts in keybdg <0.5 AND PublicHousing=0 • OR program participant between 2016-2020 more recently than IEMF
Income Eligible Services – Multifamily	• MFSF_True= "MF" • flg_lowincome_acct = 1 OR sum(flg_low_income_acct)/# of accounts in keybdg >=0.5 OR PublicHousing=1 • OR program participant between 2016 and 2020 more recently then EWMF • OR household income (D270_HOUSEHOLD_INCOME_MID) < 60% of AMI based on number of members in household (D263_HOUSEHOLD_MEMBER_COUNT)*
Natural Gas Heating and Water Heating	• elec_gas_flag = "GAS" • Primary_fuel_likely_gas = 1 • OR program participant between 2016 and 2020
ENERGY STAR HVAC	• elec_gas_flag = "ELEC" • OR program participant between 2016 and 2020
ENERGY STAR Products	• All accounts
Connected Solution	• elec_gas_flag = "ELEC"

* Based on FY2021 Rhode Island Income Limits for Low- and Moderate-Income Households, https://www.rihousing.com/wp-content/uploads/FY-21-HUD-Income-Limits.pdf

Appendix E National Grid Rhode Island's Year-End Report's Participation Counts

While our analysis is based on the active accounts only, National Grid's internal Year-End Report includes active and inactive accounts that participated. This appendix compares the Year-End Report's participant data and the raw program data for each program. The raw program data represent the analogous method of counting participants – including active and inactive accounts – as the Year-End Report. These counts are presented with gas and electric combined as the raw program data does not differentiate between gas and electric.

Before merging the raw program data to the active account list, Cadeo compared the raw program data and Year-End Report's participation counts and reached an agreement with National Grid that they are comparable.

For multifamily programs (EWMF and IEMF), raw program data is not comparable to National Grid's Year-End Report metrics because Cadeo received the program data to map facilities to active accounts only. CS and RNC were not included in the Year-End Report metrics and are therefore not included in the chart below.

Table E-1. Comparison of NGRI Year-End Report and Raw Program Participation Data (Electric and Gas Combined) Programs

Program	Sources	2016	2017	2018	2019	2020
EWSF	Raw Program Data	12,803	13,515	15,469	18,256	15,203
	Year-End Report	12,819	13,546	16,290	19,048	15,230
IESF	Raw Participant Count	3,264	3,440	4,234	4,492	3,944
	Year-End Report	3,738	3,774	4,465	4,685	4,142
EWMF	Raw Participant Count	8,582	8,211	9,953	10,850	7,508
	Year-End Report	14,961	14,491	16,355	18,230	14,504
IEMF	Raw Participant Count	2,025	2,138	1,316	1,686	1,228
	Year-End Report	3,120	2,571	1,742	1,757	1,179
NGWH	Raw Participant Count	1,652	2,934	3,104	3,842	3,280
	Year-End Report	1,652	2,949	3,113	3,846	3,282
ESHVAC	Raw Participant Count	1,976	3,014	3,261	6,377	6,723
	Year-End Report	1,978	3,023	3,269	6,298	6,745
ESProducts	Raw Participant Count	2,618	6,615	6,237	7,272	6,821
	Year-End Report	2,622	6,630	6,249	7,283	6,843

Appendix F Participation Metrics Calculation Details

F.1 Account Participation

This metric is the most straightforward due to its binary nature. For this metric, every active National Grid residential account is assigned a 1 or 0 for each year of the evaluation period (2016-2020) and cumulatively for each program:

- **1** – Identified as a participant in a given year when the program database showed that the account number installed a measure that year and the account is considered eligible for the program.[36] In alignment with National Grid's reporting standards, an account was counted in each year that a program measure was installed. For example, if one EWSF measure was installed for an account in July of 2017 and another EWSF measure was installed in January of 2019 for the same account, that account will be counted in both years
- **0** – Identified as a nonparticipant in a given year as the account does not appear within the program database for that year
- **Cumulative** - Each account is assigned a 1 for the evaluation period if they were identified as a participant in any year of the evaluation and a 0 if not. Therefore, cumulative metrics will estimate the number of unique participating accounts and will not count participating accounts more than once

F.2 Building Participation

Every residential building is also assigned a 1 or 0 for each year of the evaluation period (2016-2020) and cumulatively for each program:

- **1** – Identified as a participating building in a given year because at least one eligible account in the building installed a program measure in that year
- **0** – Identified as a nonparticipant because there are no participating accounts in the building in that year
- **Cumulative** – Each building is assigned a 1 if it was identified as a participating building in any year of the evaluation and a 0 if not. Therefore, cumulative metrics will estimate the number of unique participating buildings and will not count participating buildings more than once

[36]Note, this means that accounts will be counted in each year and program based on their eligibility as of March 2021. If that account participated in an income eligible program in 2016 but participated in EWSF in 2019, they will only be counted for participation in EWSF.

F.3 Account Savings-to-Consumption

To supplement the binary account participation metrics, the team assessed the impact of account participation by looking at the savings-to-consumption ratio (SC ratio). The SC ratio provided insight into the relationship between estimated program savings and average annual energy consumption of a subset of active participating accounts (excluding nonparticipants). This was calculated for each year of the evaluation period and cumulatively.

The team calculated estimated savings for each account and program per year by summing the estimated gross savings[37] associated with each program measure installed. The team calculated the average annual consumption for each account using the billing data associated with each account between 2016 and 2020.[38]

Since the SC ratio relied on complete and reliable consumption data, we could not calculate it for every participating account. Specifically, we excluded accounts from the analysis for the following reasons:

- Missing consumption data or missing savings data
- Negative savings or consumption data
- Outlier portfolio SC ratio values[39]

For all participating accounts with sufficient information, the team calculated the SC ratio in each year and each program by dividing the sum of estimated program savings in that year by all remaining accounts by the sum of the average annual consumption of those accounts. The SC ratios will appear as "0.08" for instance in the tables in the Result section, which indicates the estimated program savings represent 8% of the participants' average annual consumption.

To calculate the cumulative account SC ratio for the study period, the team divided the sum of program savings in all years of the evaluation period and divided by the sum of the average annual consumption[40] of those accounts, since savings will be relevant in every year after installation.

F.4 Building Savings-to-Consumption

To supplement the binary building participation metric, the team also calculated the SC ratio for a subset of participating buildings for each year and cumulatively.

[37] National Grid Rhode Island provided estimated gross savings data, not actual savings, for each measure installed at the account level
[38] The billing data for electric and gas accounts did not include a full years' worth of data for 2016. In order to include 2016 in the SC ratio analysis, the team used 2017 consumption data for 2016 and 2017.
[39] After analyzing the distribution of portfolio level SC ratios, the Cadeo team determined the lowest 3% and highest 3% of ratios were outliers. This includes ratios less than 0.003 and greater than 0.8.
[40] The average consumption is the average of 2017-2020, since 2016 was an incomplete year of billing data.

As with the account level SC ratio, the building level SC ratio relied on complete and reliable consumption data, so we could not calculate it for every participating building. Specifically, we excluded buildings from the analysis for the following reasons:

- Missing consumption data or missing savings data
- Negative savings or consumption data
- Buildings with a SC ratio of less than 0.003 or greater than 0.8

The team calculated the annual building level SC ratio for participating buildings by dividing the sum of program measure savings in each year for participating accounts in the building by the sum of consumption of all participating buildings in that year. Accounts were aggregated into buildings using the keybdg field. Cadeo used the building level consumption data provided by National Grid Rhode Island. We can expect that SC ratio at the building level to be smaller than at the account level because even a large MF building with just one participating account will have their entire annual consumption added into the denominator of the metric.

To calculate the cumulative building SC ratio for the study period, the team divided the sum of all program savings in all years of the evaluation period by the sum of the average annual consumption of all participating buildings.

Appendix G Program-Specific Participation Results

The following section summarizes participation for each program and year, split by electric and gas where relevant.

In the program specific sections, we present the "raw" program participation counts (i.e., all accounts that participated, regardless of whether they were still active in March 2021 when National Grid provided the comprehensive list of their residential customers that underpins much of our team's analysis) at the beginning of each program-specific subsection. These raw participation counts are comparable to and largely consistent with National Grid's previous Year-End reporting for each program.

However, because much of our analysis requires merging raw program data with additional data sources—most notably billing data to calculate the SC ratio—the rest of each program specific subsection focuses on participating accounts that were still active as of March 2021.

G.1 EnergyWise Single-Family

G.1.1 Raw Count of Participating Accounts

Table G-1 shows the number of unique gas and electric accounts in the EWSF program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-1. EWSF Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
12,803	13,515	15,469	18,256	15,203	**67,337**

G.1.2 Account and Building Eligibility

Next, Cadeo determined the total number of active accounts and buildings eligible to participate in EWSF. As shown in Table G-2, well over half of National Grid's accounts are eligible to participate in EWSF.

Table G-2. EWSF Account and Building Eligibility Error! Reference source not found.

	Residential Electric		Residential Gas	
	Count of Eligible	% of Total	Count of Eligible	% of Total
Accounts	267,224	60%	141,065	57%
Buildings	245,114	73%	128,942	66%

G.1.3 Annual and Cumulative Participation

Table G-3 compares historical EWSF participation of eligible active accounts and buildings against the total number of eligible active accounts and buildings. As shown in the table, 15% of electric accounts and 16% of gas accounts participated in EWSF between 2016 and 2020. Also, the percent of active eligible electric and gas accounts participating in EWSF increased every year during the study period except for 2020, which could be related to COVID-19. However, 2020 participation still outpaced previous years. The largest increase in participation occurred between 2018 and 2019.

Although there was an increase in the percent of active accounts and buildings participating in EWSF, the SC ratio of participating accounts and buildings trended downward over the five-year period. This is particularly true for electric accounts.

Table G-3. EWSF Annual & Cumulative Participation

Participation Metric		Annual					Cumulative	
		2016	2017	2018	2019	2020	2016-2020	2009-2020
Electric								
Account	Count	6,671	7,346	9,149	11,446	10,837	**40,200**	**56,187**
	% of eligible	2%	3%	3%	4%	4%	**15%**	**21%**
	SC Ratio	0.19	0.16	0.21	0.10	0.10	**0.17**	**NA**
Building	Count	6,554	7,190	8,896	11,011	10,341	**38,786**	**54,565**
	% of eligible	3%	3%	4%	4%	4%	**16%**	**22%**
	SC Ratio	0.17	0.15	0.20	0.09	0.10	**0.16**	**NA**
Gas								
Account	Count	2,224	2,465	2,818	3,670	2,941	**12,814**	**17,167**
	% of eligible	2%	2%	2%	3%	2%	**9%**	**12%**
	SC Ratio	0.18	0.18	0.09	0.09	0.09	**0.14**	**NA**
Building	Count	2,197	2,428	2,753	3,559	2,845	**12,487**	**16,822**
	% of eligible	2%	2%	2%	3%	2%	**10%**	**13%**
	SC Ratio	0.19	0.17	0.08	0.08	0.08	**0.13**	**NA**

The downward trend in the SC ratio is mathematically driven by declining average annual savings, per participating electric (-6%, average percent change) and gas (-17%, average percent

change) accounts. Table G-4 summarizes the average savings and consumption per accounts and buildings in EWSF.

Table G-4. Average Savings & Consumption per EWSF Participant by Year

	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Account Level						
Electric						
Avg Savings per Participant (kWh)	1,618	1,345	1,876	801	871	**-6%**
Avg Consumption per Participant (kWh)	8,652	8,319	8,889	8,024	8,397	**-1%**
Gas						
Avg Savings per Participant (therms)	182	158	84	80	78	**-17%**
Avg Consumption per Participant (therms)	995	889	960	913	826	**-4%**
Building Level						
Electric						
Avg Savings per Participant (kWh)	1,360	1,876	804	900	871	**-3%**
Avg Consumption per Participant (kWh)	9,497	8,883	9,410	8,781	9,398	**0%**
Gas						
Avg Savings per Participant (therms)	155	151	112	104	89	**-12%**
Avg Consumption per Participant (therms)	1,205	1,197	1,307	1,195	1,133	**-1%**

Figure G-1 shows the decline in average savings per EWSF participant by measure category between 2016 and 2020. The decline in average savings per EWSF electric participant is driven by a decline in the average savings per participant from lighting measures, which made up approximately 88% of the average savings per participant in the study period. Savings per participant form lighting measures fell by 44% between 2016 and 2020.

Figure G-1. Average Savings per EWSF Electric Participant by Measure Category

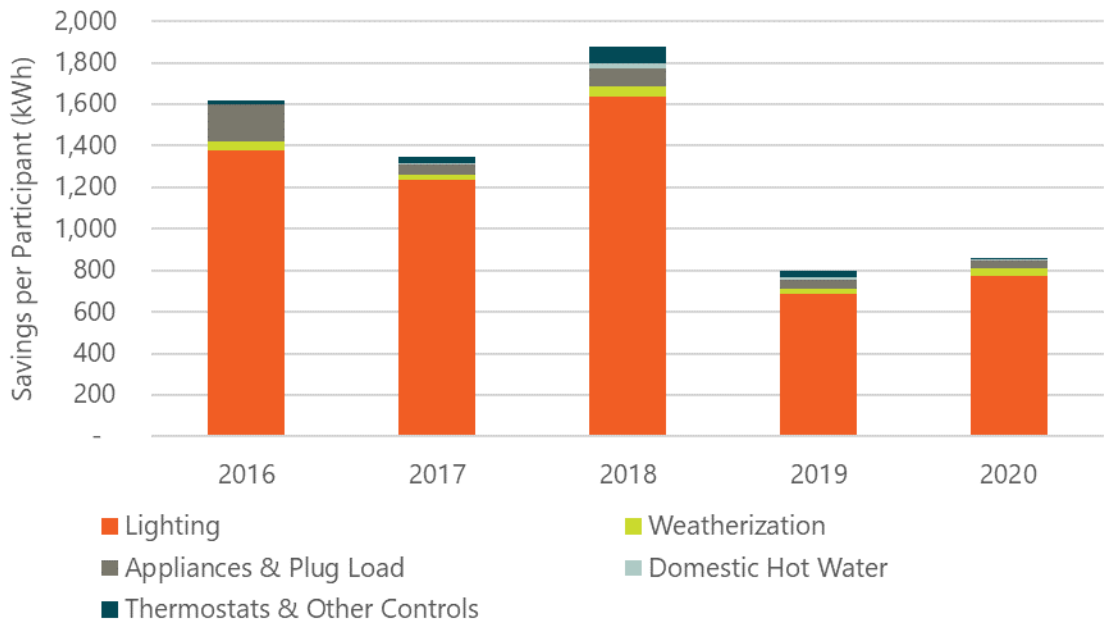
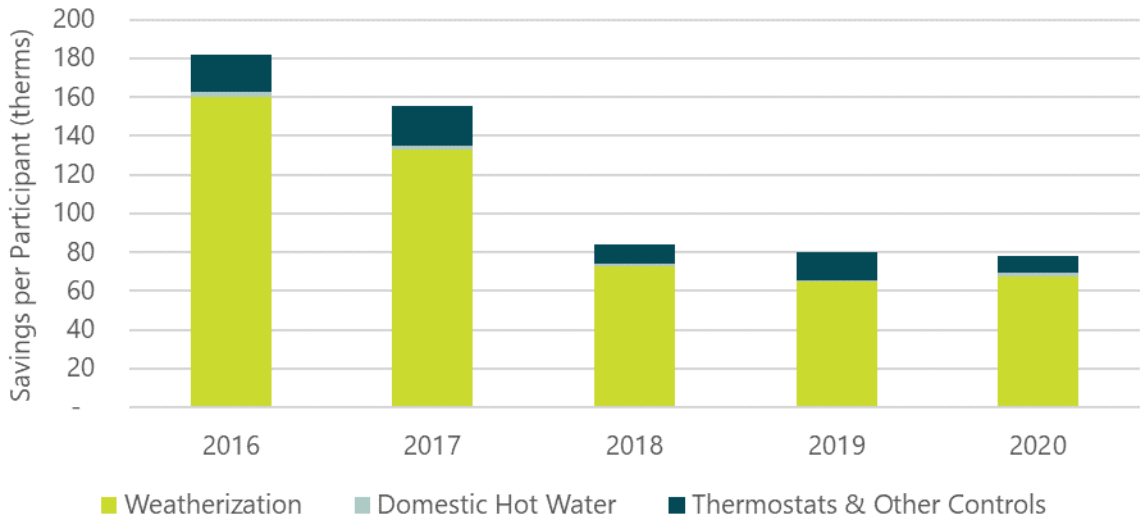


Figure G-2 illustrates the decline in average therm savings per EWSF gas participant by measure category. The decline was primarily driven by a drop in the average therm savings per participant from weatherization measures. Savings from weatherization measures made up approximately 78% of the average savings per participant in the study period. Savings from weatherization fell by 61% between 2016 and 2020.

Figure G-2. Average Savings per EWSF Gas Participant by Measure Category



G.2 Income Eligible Single-Family

G.2.1.1 Raw Count of Participating Accounts

Table G-5 shows the number of unique gas and electric accounts that appeared in IESF program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-5. IESF Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
3,264	3,440	4,234	4,492	3,944	**16,488**

G.2.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in IESF. As shown in Table G-6, over 20% of National Grid's accounts are eligible to participate in IESF. **Error! Reference source not found.**

Table G-6. IESF Account and Building Eligibility

	Residential Electric		Residential Gas	
	Count of Eligible	% of Total	Count of Eligible	% of Total
Accounts	93,104	21%	66,405	27%
Buildings	77,019	23%	54,458	28%

G.2.1.3 Annual and Cumulative Participation

The percent of eligible electric accounts and buildings participating in the program increased every year of the study period, while gas participation rates stayed relatively flat and declined in 2020. The SC ratio decreased over time for both electric and gas participating customers due to a decrease in average savings per participant over the study period.

The participation metrics for the IESF program for both electric and gas accounts is summarized in Table G-7.

Table G-7. IESF Annual and Cumulative Participation

Participation Metric		Annual					Cumulative	
		2016	2017	2018	2019	2020	2016-2020	2009-2020
Electric								
Account	Count	1,568	1,764	2,600	3,147	3,323	**10,193**	**13,144**
	% of eligible	2%	2%	3%	3%	4%	**11%**	**14%**
	SC Ratio	0.24	0.20	0.15	0.12	0.08	**0.17**	**NA**
Building	Count	1,533	1,720	2,542	3,039	3,207	**9,733**	**12,550**
	% of eligible	2%	2%	3%	4%	4%	**13%**	**16%**
	SC Ratio	0.14	0.13	0.10	0.07	0.05	**0.13**	**NA**
Gas								
Account	Count	372	482	478	522	256	**1,984**	**2,733**
	% of eligible	1%	1%	1%	1%	0.4%	**3%**	**4%**
	SC Ratio	0.24	0.24	0.23	0.13	0.14	**0.22**	**NA**
Building	Count	365	466	463	506	252	**1,922**	**2,655**
	% of eligible	1%	1%	1%	1%	0.5%	**4%**	**5%**
	SC Ratio	0.19	0.20	0.20	0.11	0.12	**0.18**	**NA**

The decline in SC Ratio is, in part, due to a decline in the average annual savings per IESF participant. Savings per participating account dropped by 22% for electric accounts and 17% per gas account on average during the period. Consumption per participant grew slightly. Similar trends occurred at the building level, though there was a slight decline in consumption per participating building. These results are summarized in Table G-8.

Table G-8. Average Savings and Consumption per IESF Participant by Year

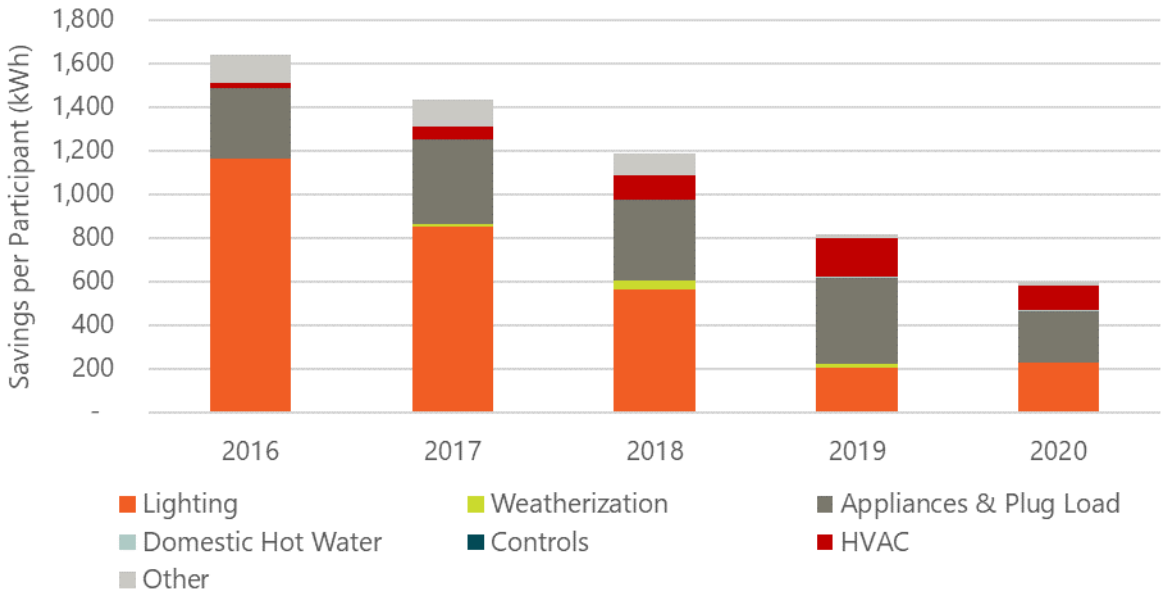
	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Account Level						
Electric						
Avg Savings per Participant (kWh)	1,636	1,423	1,147	802	597	**-22%**
Avg Consumption per Participant (kWh)	6,837	7,139	7,507	6,835	7,114	**1%**
Gas						
Avg Savings per Participant (therms)	182	158	84	80	78	**-17%**
Avg Consumption per Participant (therms)	804	867	920	909	828	**1%**
Building Level						
Electric						

	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Avg Savings per Participant (kWh)	1,681	1,464	1,169	816	616	**-22%**
Avg Consumption per Participant (kWh)	12,175	11,342	11,902	11,033	11,548	**-1%**
Gas						
Avg Savings per Participant (therms)	193	218	223	126	115	**-9%**
Avg Consumption per Participant (therms)	1,031	1,082	1,125	1,104	975	**-1%**

Figure G-3 and Figure G-4 provides more details about the decline of average savings per participant in the IESF program.

The decline in savings per electric participant is primarily the result of a decline in savings from lighting measures over time, which made up 53% of savings per participant. Between 2016 and 2020, lighting savings per participant fell by 81%.

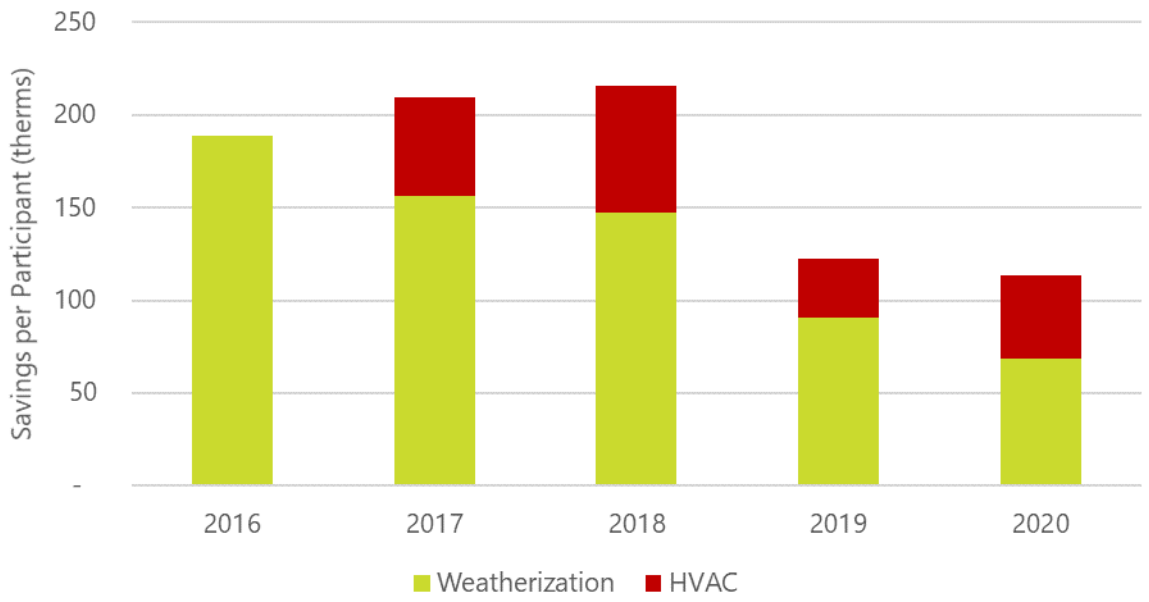
Figure G-3. Average Savings per IESF Electric Participant by Measure Category



The decline in savings per participant of gas participants in IESF is the result of declining savings from weatherization measures, which made up 77% of the savings per

participant in the study period, despite some growth in savings from HVAC measures. Between 2016 and 2020, savings from weatherization per gas participant fell by 64%.

Figure G-4. Average Savings per IESF Gas Participant by Measure Category



G.3 EnergyWise Multifamily

G.3.1.1 Raw Count of Participating accounts

Table G-9 shows the number of unique gas and electric accounts that mapped to the facility ID in EWMF program data in each year. As discussed in the methodology section, Cadeo received a key to map active accounts to facilities in EWMF program data. Unlike the other program data, these numbers are not comparable to National Grid's end of year report metrics because Cadeo received the data to map facilities to active accounts only.

Table G-9. EWMF Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
8,582	8,211	9,953	10,850	7,508	**28,684**

G.3.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in EWMF. As shown in Table G-10, over 10% of National Grid's accounts are eligible to participate in EWMF.

Table G-10. EWMF Account and Building Eligibility

	Residential Electric		Residential Gas	
	Count of Eligible	% of Total	Count of Eligible	% of Total
Accounts	50,835	12%	26,631	11%
Buildings	15,819	5%	12,539	6%

G.3.1.3 Annual and Cumulative Participation

Participation rates in the multifamily buildings were higher than in the single-family programs. The SC ratio decreased overtime at the account and building level for electric accounts. The SC ratio decreased less drastically for participating gas accounts and buildings.

The participation metrics for EWMF are summarized in Table G-11.

Table G-11. EWMF Annual and Cumulative Participation

Participation Metric		Annual					Cumulative	
		2016	2017	2018	2019	2020	2016-2020	2009-2020
Electric								
Account	Count	3,998	3,273	4,593	4,404	3,398	**13,558**	**13,947**
	% of eligible	8%	6%	9%	9%	7%	**27%**	**27%**
	SC Ratio	0.13	0.11	0.09	0.05	0.07	**0.13**	**NA**
Building	Count	1,950	1,142	2,060	1,928	1,255	**5,111**	**5,465**
	% of eligible	12%	7%	13%	12%	8%	**32%**	**35%**
	SC Ratio	0.07	0.07	0.05	0.02	0.05	**0.08**	**NA**
Gas								
Account	Count	1,943	2,026	1,913	2,614	1,610	**6,532**	**7,021**
	% of eligible	7%	8%	7%	10%	6%	**25%**	**26%**
	SC Ratio	0.08	0.07	0.07	0.07	0.05	**0.11**	**NA**
Building	Count	1,184	996	1,004	1,538	943	**3,586**	**4,066**
	% of eligible	9%	8%	8%	12%	8%	**29%**	**32%**
	SC Ratio	0.08	0.06	0.04	0.06	0.05	**0.09**	**NA**

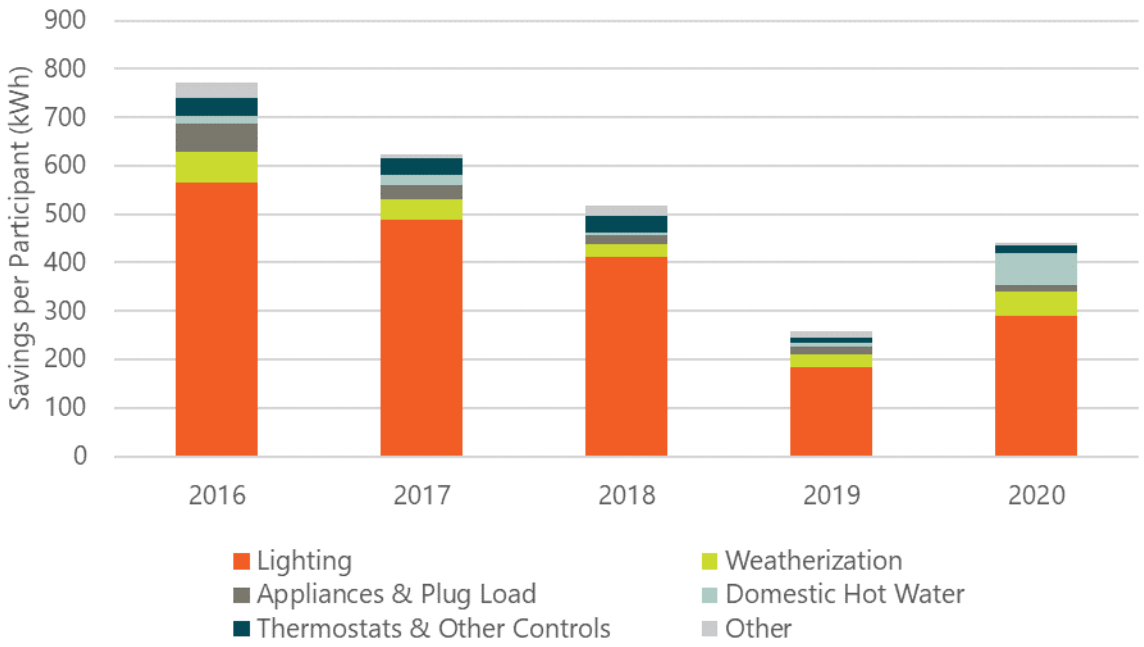
The slight decline in the SC ratio for electric accounts is, in part due to a decline in the average annual savings per EWMF participant and the increase in average consumption per participant. The average annual savings for gas accounts and building also decreased. Table G-12 summarizes the average annual savings and consumption per participant.

Table G-12. Average Savings and Consumption per EWMF Participant by Year

	2016	2017	2018	2019	2020	**Avg. % change**, 2016-2020
Account Level						
Electric						
Avg Savings per Participant (kWh)	773	625	517	258	443	**-4%**
Avg Consumption per Participant (kWh)	6,018	5,892	5,955	5,639	6,417	**0%**
Gas						
Avg Savings per Participant (therms)	60	41	47	49	33	**-11%**
Avg Consumption per Participant (therms)	778	622	707	664	617	**-5%**
Building Level						
Electric						
Avg Savings per Participant (kWh)	1,771	2,042	1,169	631	1,525	**17%**
Avg Consumption per Participant (kWh)	24,869	27,332	22,725	25,805	27,777	**4%**
Gas						
Avg Savings per Participant (therms)	102	93	67	78	59	**-11%**
Avg Consumption per Participant (therms)	1323	1,645	1,652	1,324	1,303	**1%**

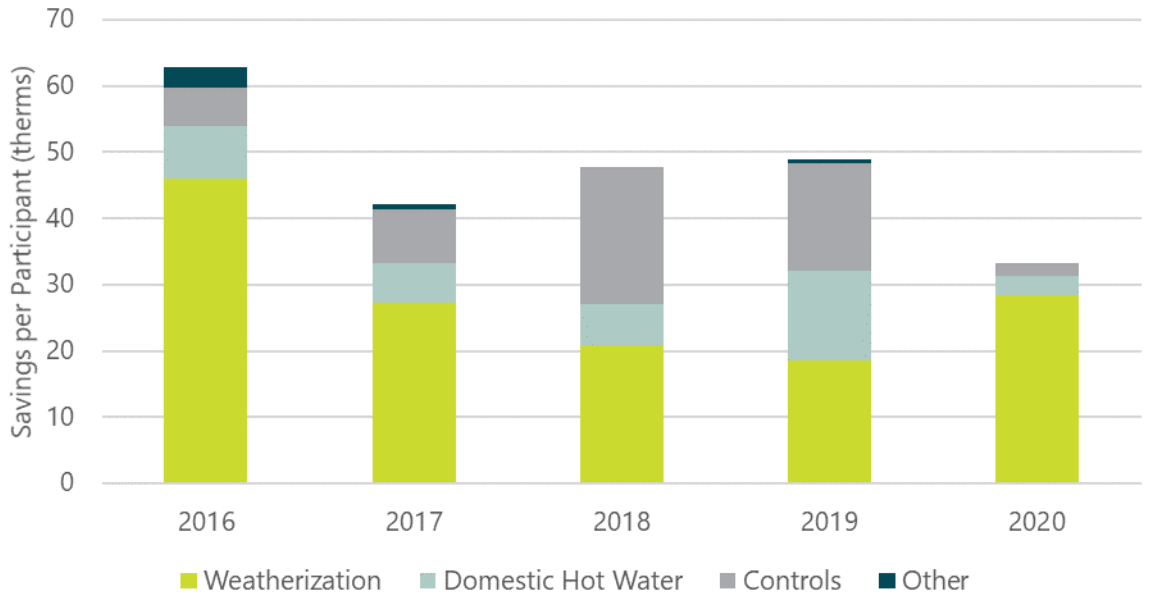
Figure G-5. and Figure G-6 provide more details into the declining savings per participant for EWMF. The decline in savings for electric participants is primarily due to the decline in savings from lighting measures, which made up 80% of all electric savings per participant in 2016 to 2020. Savings from lighting fell by 68% between 2016 and 2020.

Figure G-5. Average Savings per EWMF Electric Participant by Measure Category



As noted above, the decline in savings per gas participant is primarily due to a decline in gas-heated participants opting to weatherization their home, which resulted in a lower average savings for gas accounts. There was also a decline in savings from domestic hot water measures and control measures in 2020. Weatherization made up 60% of savings per participant between 2016 and 2020 Between 2016 and 2020, savings from weatherization fell by 38%.

Figure G-6. Average Savings per EWMF Gas Participant by Measure Category



G.4 Income Eligible Multifamily

G.4.1.1 Raw Count of Participating Accounts

Table G-13 shows the number of unique gas and electric accounts that mapped to the facility ID in the IEMF program data in each year. Cadeo received a key to map active accounts to facilities in the IEMF program data. Unlike the other program data, these numbers are not comparable to National Grid's Year-End Report metrics because Cadeo received the data to map facilities to active accounts only.

Table G-13. IEMF Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
2,025	2,138	1,316	1,686	1,228	**7,050**

G.4.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in IEMF. As shown in Table G-14, over 5% of National Grid's accounts are eligible to participate in IEMF.

Table G-14. IEMF Account and Building Eligibility

	Residential Electric		Residential Gas	
	Count of Eligible	% of Total	Count of Eligible	% of Total
Accounts	32,574	7%	14,018	6%

Buildings	8,681	3%	6,211	3%

G.4.1.3 Annual and Cumulative Participation

Participation rates in IEMF for electric and gas accounts and buildings fell slightly through the 2016 to 2020 period. The SC ratio also declined during the study period for both gas and electric accounts.

The participation metrics for IEMF are summarized in Table G-15.

Table G-15. IEMF Annual and Cumulative Participation

Participation Metric		Annual					Cumulative	
		2016	2017	2018	2019	2020	2016-2020	2009-2020
Electric								
Account	Count	1,005	1,267	532	816	824	**3,994**	**4,652**
	% of eligible	3%	4%	2%	3%	3%	**12%**	**14%**
	SC Ratio	0.19	0.21	0.21	0.10	0.09	**0.18**	**NA**
Location	Count	256	191	133	153	24	**731**	**860**
	% of eligible	3%	2%	2%	2%	0.3%	**8%**	**10%**
	SC Ratio	0.15	0.14	0.08	0.06	0.06	**0.11**	**NA**
Gas								
Account	Count	238	146	209	178	55	**615**	**788**
	% of eligible	2%	1%	1%	1%	0.4%	**4%**	**6%**
	SC Ratio	0.10	0.19	0.14	0.21	0.01	**0.33**	**NA**
Location	Count	185	130	156	159	54	**491**	**611**
	% of eligible	3%	2%	3%	3%	1%	**8%**	**10%**
	SC Ratio	0.09	0.15	0.14	0.17	0.02	**0.25**	**NA**

The SC ratio for electric accounts declined in part due to a significant decrease in the average annual savings per participant between 2016 and 2020. For gas accounts the SC ratio increased between 2016 and 2019 due to an increase in average annual savings per participant. Table G-16 summarizes the average annual savings and consumption per participating accounts and building.

Table G-16. Average Savings and Consumption per IEMF Participant by Year

	2016	2017	2018	2019	2020	**Avg. % change**, 2016-2020
Account Level						
Electric						
Avg Savings per Participant (kWh)	723	988	1,353	603	430	**-3%**
Avg Consumption per Participant (kWh)	3,818	4,607	6,343	6,069	4,956	**15%**
Gas						
Avg Savings per Participant (therms)	162	456	204	447	46	**39%**
Avg Consumption per Participant (therms)	1,684	2,418	1,477	2,141	3,295	**10%**
Building Level						
Electric						
Avg Savings per Participant (kWh)	1,681	1,464	1,169	816	616	**-22%**
Avg Consumption per Participant (kWh)	12,175	11,342	11,902	11,033	11,548	**-1%**
Gas						
Avg Savings per Participant (therms)	220	513	307	554	56	**21%**
Avg Consumption per Participant (therms)	2,459	3,434	2,265	3,165	3,240	**12%**

Figure G-7. and Figure G-8 provide more details about the declining savings per IEMF participant between 2016 and 2020.

The decline in savings per electric participant was primarily due to a decline in savings from lighting measures, which made up 71% of all savings per participant in the study period. Savings per participant from lighting measures fell by approximately 63% between 2016 and 2020. The category labeled "Other" consists of custom measures and TLC Kits.

Figure G-7. Average Savings per IEMF Electric Participant by Measure Category

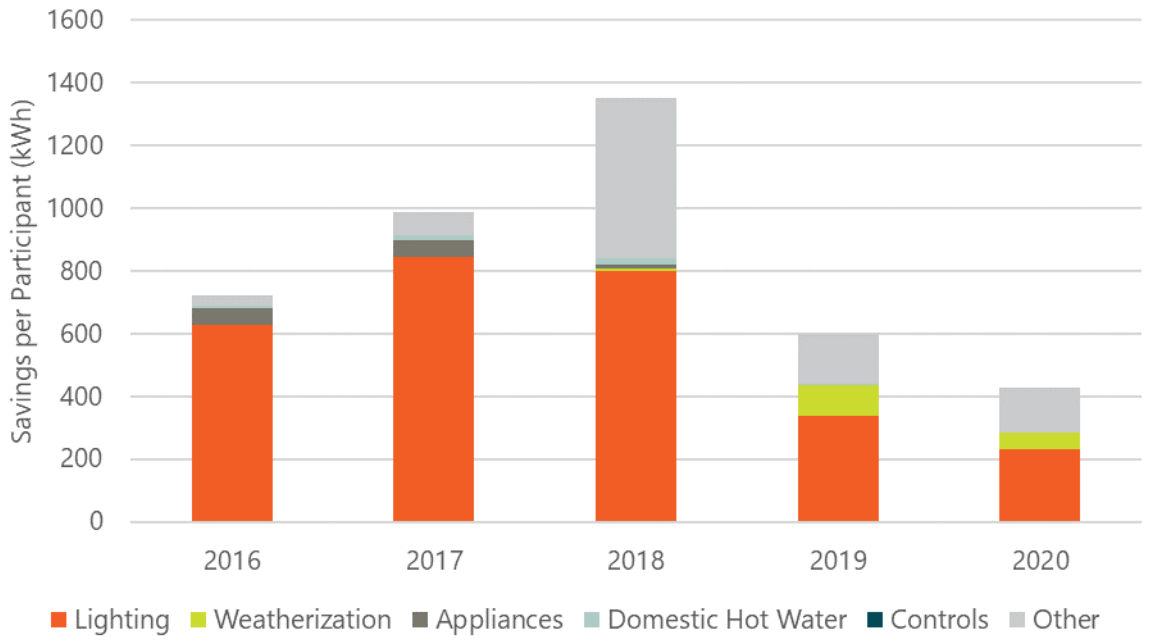
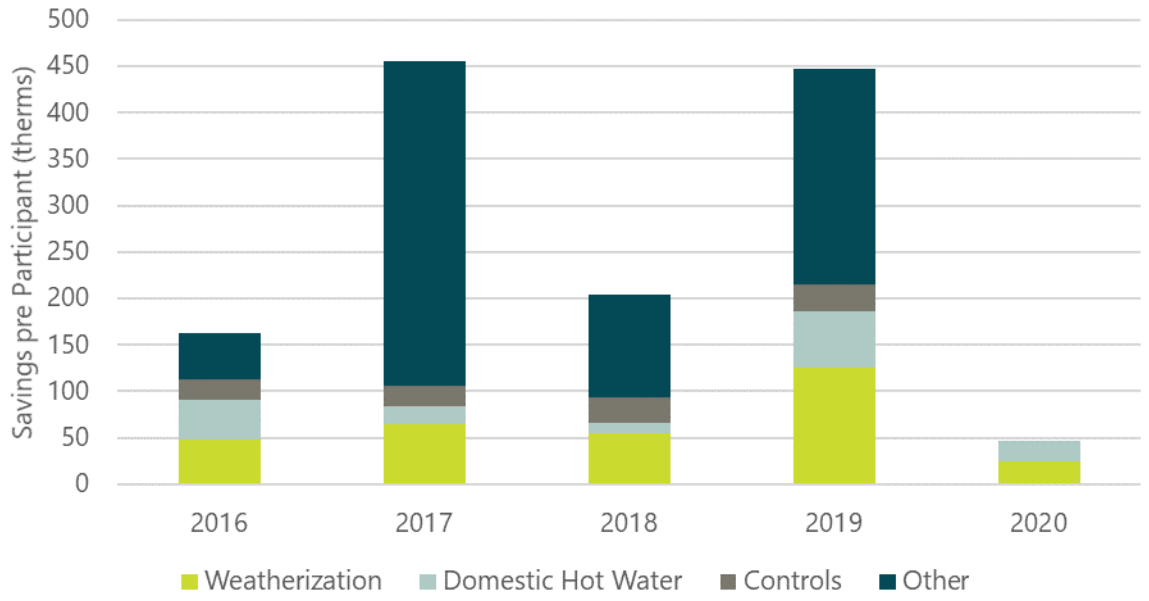


Figure G-8. Average Savings per IEMF Gas Participant by Measure Category



G.5 Natural Gas Water and Space Heating

G.5.1.1 Raw Count of Participating Accounts

Table G-17 shows the number of unique gas and electric accounts that appeared in the NGWH program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-17. NGWH Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
1,652	2,934	3,104	3,842	3,280	14,106

G.5.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in NGWH. As shown in Table G-18, all National Grid's gas accounts are eligible to participate in NGWH.

Table G-18. NGWH Account and Building Eligibility

	Residential Gas	
	Count of Eligible	% of Total
Accounts	248,119	100%
Buildings	195,302	100%

G.5.1.3 Annual and Cumulative Participation

The participation rate in NGWH increased slightly in every year except 2020. The SC ratio shows a slight decline but remained relatively steady between 2016 and 2020.

The participation metrics for NGWH are summarized in Table G-19.

Table G-19. NGWH Annual and Cumulative Participation

Participation Metric		Annual					Cumulative
		2016	2017	2018	2019	2020	2016-2020
Gas							
Account	Count	1,124	2,062	2,297	3,209	2,337	**10,395**
	% of eligible	0.5%	1%	1%	1%	1%	**4%**
	SC Ratio	0.13	0.13	0.12	0.12	0.13	**0.13**
Location	Count	1,107	2,043	2,261	3,142	2,268	**10,067**
	% of eligible	1%	1%	1%	2%	1%	**5%**
	SC Ratio	0.08	0.08	0.07	0.08	0.07	**0.10**

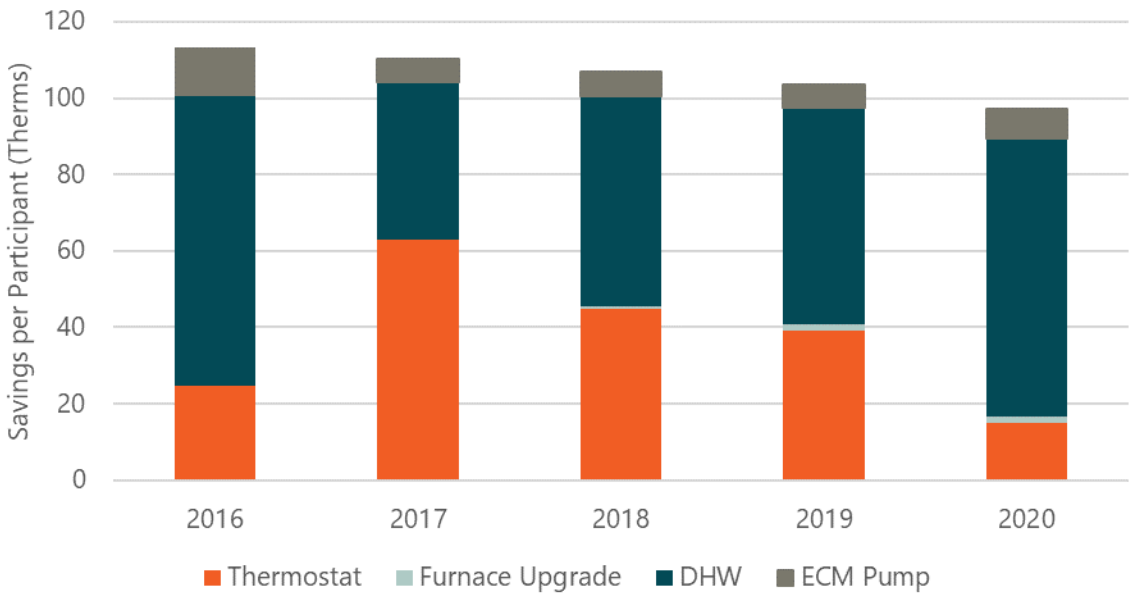
The SC ratio fell slightly due to a slight decline in the average saving per participant. The average annual savings and consumption per participant is summarized in Table G-20.

Table G-20. Average Savings and Consumption per NGWH Participant by Year

	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Account						
Avg Savings per Participant (therms)	114	110	107	104	97	**-4%**
Avg Consumption per Participant (therms)	854	850	889	871	766	**2%**
Building						
Avg Savings per Participating Building (therms)	117	112	109	106	100	**-4%**
Avg Consumption per Participating Building (therms)	1,430	1,325	1,475	1,317	1,345	**-4%**

Figure G-9 provides more details about the program savings per participant between 2016 and 2020. Savings from thermostat measures decreases over the study period while savings from domestic how water measures increased, particularly between 2017 and 2020. Between 2017 and 2020, savings form thermostats per NGWH participant decline by 76% while savings from domestic hot water measures increased by 30%.

Figure G-9. Average Savings per NGWH Participant by Measure Category



G.6 ENERGY STAR HVAC

G.6.1.1 Raw Count of Participating Accounts

Table G-21 shows the number of unique electric accounts that appeared in ESHVAC program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-21. ESHVAC Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
1,976	3,014	3,261	6,377	6,723	**19,989**

G.6.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in ESHVAC. As shown in Table G-22, all National Grid's electric accounts are eligible to participate in ESHVAC.

Table G-22. ESHVAC Account and Building Eligibility

	Residential Electric	
	Count of Eligible	% of Total
Accounts	443,737	100%
Buildings	336,408	100%

G.6.1.3 Annual and Cumulative Participation

Participation rates in ESHVAC increased every year in the study period. The SC ratio fell slightly between 2016 and 2020. The participation metrics for ESHVAC is summarized in Table G-23.

Table G-23. ESHVAC Annual and Cumulative Participation

Participation Metric		Annual					Cumulative
		2016	2017	2018	2019	2020	2016-2020
Electric							
Account	Count	1,504	2,342	2,673	5,459	5,901	**16,625**
	% of eligible	0.4%	1%	1%	1%	1%	**4%**
	SC Ratio	0.06	0.05	0.05	0.04	0.04	**0.05**
Building	Count	1,498	2,337	2,659	5,406	5,828	**16,359**
	% of eligible	0.4%	1%	1%	2%	2%	**5%**
	SC Ratio	0.03	0.03	0.02	0.03	0.02	**0.03**

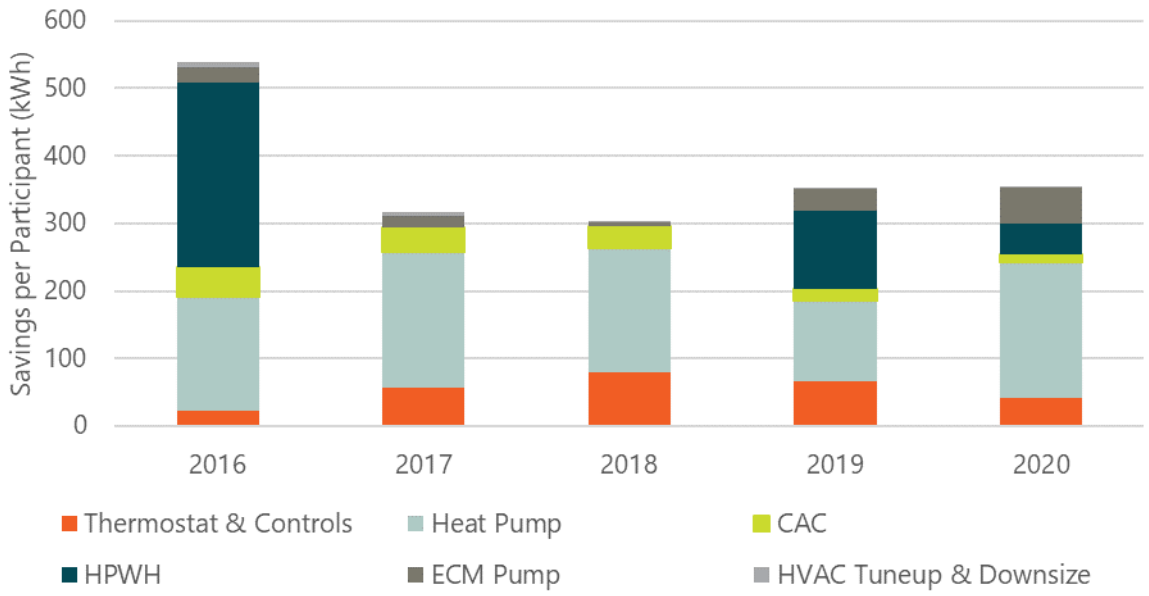
The SC ratio fell due, in part, to a decrease in average annual savings per participant. The change in average annual savings and consumption is summarized in Table G-24.

Table G-24. Average Savings and Consumption per ESHVAC Participant by Year

	2016	2017	2018	2019	2020	Avg % change, **2016-2020**
Account						
Avg Savings per Participant (kWh)	541	487	425	352	354	**-10%**
Avg Consumption per Participant (kWh)	9,462	9,012	9,302	8,479	8,644	**-1%**
Building						
Avg Savings per Participating Building (kWh)	552	482	427	354	368	**-9%**
Avg Consumption per Participating Building (kWh)	17,257	17,814	20,905	13,252	15,992	**-6%**

Figure G-10 provides more details about the change in average savings per participant between 2016 and 2020. Savings per person from heat pumps remained relatively steady between 2016 and 2020, while savings from central air conditioning (CAC) measures and heat pump water heaters (HPWH) measures fell by 74% and 83% per participant respectively.

Figure G-10. Average Savings per ESHVAC Participant by Measure Category



G.7 ENERGY STAR Products

G.7.1.1 Raw Count of Participating Accounts

Table G-25 shows the number of unique gas and electric accounts that appeared in ESProducts program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-25. ESProducts Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
2,618	6,615	6,237	7,272	6,821	**27,264**

G.7.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in ESProducts. As shown in Table G-26, all National Grid's electric accounts are eligible to participate in ESProducts.

Table G-26. ESProducts Account and Building Eligibility

	Residential Electric	
	Count of Eligible	% of Total
Accounts	443,737	100%
Buildings	336,408	100%

G.7.1.3 Annual and Cumulative Participation

Participation rates in ESProducts grew slightly during the study period. The SC ratio also grew between 2016 and 2020.

The participation metrics for ESProducts are summarized in Table G-27.

Table G-27. ESProducts Annual and Cumulative Participation

Participation Metric		Annual					Cumulative
		2016	2017	2018	2019	2020	2016-2020
Electric							
Account	Count	2,052	5,151	5,005	6,308	6,387	**22,808**
	% of eligible	0.3%	1%	1%	1%	1%	**5%**
	SC Ratio	0.02	0.06	0.05	0.06	0.08	**0.07**
Building	Count	2,044	5,129	4,983	6,261	6,310	**22,452**
	% of eligible	1%	2%	1%	2%	2%	**7%**
	SC Ratio	0.01	0.04	0.03	0.04	0.05	**0.05**

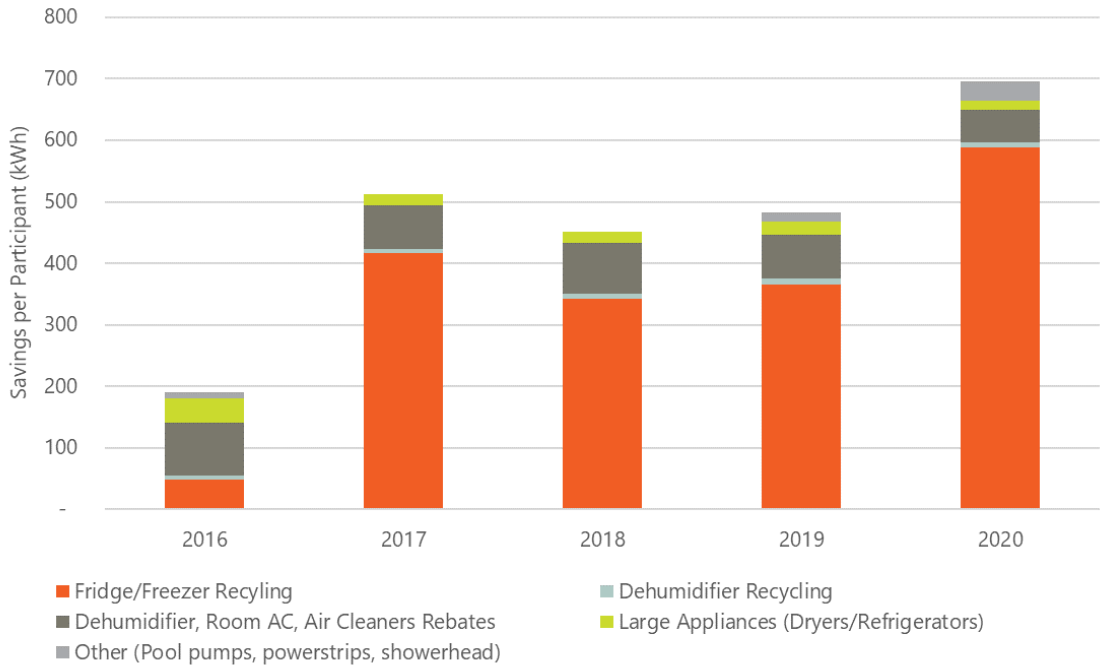
The SC ratio grew due to an increase in the average annual savings per participant. Table G-28 summarizes the average annual savings and consumption per ESProducts participant.

Table G-28. Average Savings and Consumption per ESProducts Participant by Year

	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Electric						
Avg Savings per Participant (kWh)	191	512	451	482	695	**52%**
Avg Consumption per Participant (kWh)	8,491	8,382	8,646	8,135	8,445	**0%**
Building						
Avg Savings per Participating Building (kWh)	193	511	456	490	705	**51%**
Avg Consumption per Participating Building (kWh)	20,202	12,524	13,323	13,829	14,944	**-6%**

Figure G-11 provides more details about the savings per ESProducts participants between 2016 and 2020. Savings per participant primarily grew due to the increase in savings from appliance recycling measures, which made up approximately 77% of the savings per participant in the study period. Appliance recycling measure savings per participant grew by 978% between 2016 and 2020.

Figure G-11. Average Savings per ESProducts Participant by Measure



This increase in savings from refrigerator and freezer recycling was primarily due to an increase in the number of refrigerators and freezers being recycled in the program per program participant. However, there was also an increase in the savings allocated to refrigerator and freezer recycling between 2016 and 2020. Table G-29 illustrates those changes.

Table G-29. Savings from Fridge/Freezer Recycling for ESProducts Participants

	2016	2017	2018	2019	2020	**Avg. % change,** 2016-2020
Average number of Recycling measures per participant	0.1	0.6	0.5	0.4	0.6	825%
Average kWh savings per Fridge/Freezer Recycled (kwh)	748.8	654.0	631.6	854.0	981.2	31%

G.8 Connected Solutions

G.8.1.1 Raw Count of Participating Accounts

Table G-30 shows the number of unique gas and electric accounts that appeared in the CS program data in each year before the team matched the data with the active account list and applied eligibility criteria.

Table G-30. CS Raw Count of Participating Accounts

Annual					Cumulative
2016	2017	2018	2019	2020	2016-2020
1,652	2,934	3,104	3,842	3,280	14,106

G.8.1.2 Account and Building Eligibility

Cadeo determined the total number of active accounts and buildings eligible to participate in CS. As shown in Table G-31**Error! Reference source not found.**, approximately 6% of National Grid's electric accounts are eligible to participate in CS.

Table G-31. CS Account and Building Eligibility

	Residential Electric	
	Count of Eligible	% of Total
Accounts	25,293	6%
Buildings	20,184	6%

G.8.1.3 Annual and Cumulative Participation

The CS program had the lowest number of participants of any program, which is not a surprise given the program's focus on demand response via controllable end uses. However, the program did see an increase in participation between 2016 and 2020. The team did not calculate SC ratio for this program because the program objective is demand reduction, not energy savings.

The participation metrics for the CS program are summarized in Table G-32.

Table G-32. CS Annual & Cumulative Participation

Participation Metric		Annual					Cumulative
		2016	2017	2018	2019	2020	2016-2020
Electric							
Account	Count	287	367	282	416	564	**1,917**
	% of eligible	1%	1%	1%	2%	2%	**8%**
Building	Count	287	367	282	413	557	**1,893**
	% of eligible	1%	2%	1%	2%	3%	**9%**

Appendix H Results of Program Sequence Analysis

Cadeo assessed the inter-program relationships of participation patterns within National Grid's residential portfolio.

Overall, we did not find that any specific sequences of program participation had a significantly higher rate of occurrence than is expected given the base rates for program participation.

As a baseline, we first examined the first program in which each account participated (the "entry program"). This includes accounts that only participated in one program as well as those that participated in more than one. The baseline establishes the relative rates of participation. When a customer participates in additional programs, we are interested in how the relative rates compare to the baseline. Table H-1 shows the frequency of accounts engaged in a single program. EWSF program was the program with the largest number of participating accounts with 80,546 accounts that only participated in EWSF, but had 14,434 accounts where EWSF was the first of multiple programs.

The three entry programs that had the highest rate of multiple program participation were the ESProducts, ESHVAC, and NGWH programs. However, extending the analysis to longer sequences shows us that the elevated rate of participation in multiple programs associated with these three programs is a biproduct of the popularity of the EWSF program and the fact that most participants in these programs are also eligible for EWSF.

Table H-1. Entry Program Analysis

Program	# of accounts	# of accounts in additional programs	% in additional programs
EWSF	80,507	14,434	15%
IESF	21,517	1,694	7%
EWMF	34,224	1,333	4%
IEMF	9,537	130	1%
ESProducts	18,448	4,374	19%
ESHVAC	13,432	3,206	19%
NGWH	7,885	2,522	24%

We placed all combinations of the entry and secondary program where at least 75 customers participated in at least that sequence of programs into Table H-2. For each entry program, the rank order of the secondary program (based on frequency) generally follows the rank order of entry programs from Table H-1.

Table H-2. Participation Sequences – Entry and Secondary Programs

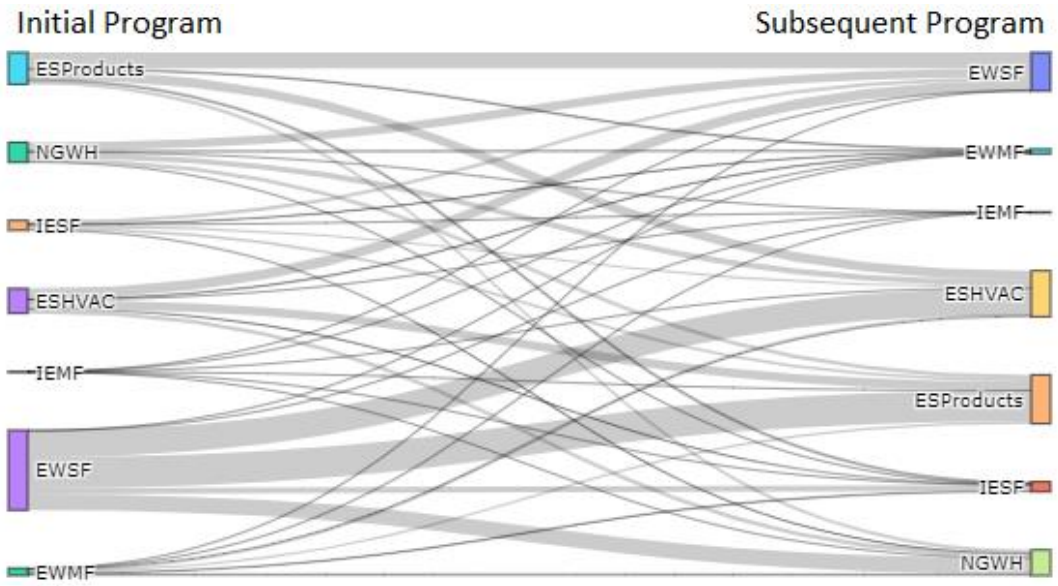
Entry Program	Secondary Program	# of accounts	# of accounts in additional programs	% in additional programs
EWSF	IESF	912	123	12%
EWSF	NGWH	2062	826	29%
EWSF	ESHVAC	3833	922	19%
EWSF	ESProducts	4512	1180	21%
IESF	ESHVAC	313	46	13%
IESF	ESProducts	410	91	18%
IESF	EWSF	502	98	16%
EWMF	IESF	153	31	17%
EWMF	NGWH	230	74	24%
EWMF	ESHVAC	262	71	21%
EWMF	ESProducts	345	78	18%
ESProducts	EWMF	170	11	6%
ESProducts	IESF	215	13	6%
ESProducts	NGWH	315	108	26%
ESProducts	ESHVAC	1087	182	14%
ESProducts	EWSF	1998	275	12%
ESHVAC	EWMF	121	4	3%
ESHVAC	NGWH	388	92	19%
ESHVAC	ESProducts	1023	111	10%
ESHVAC	EWSF	1263	120	9%
NGWH	EWMF	150	1	1%
NGWH	ESProducts	396	87	18%
NGWH	ESHVAC	567	120	18%
NGWH	EWSF	1003	80	7%

*Only sequences with 75 or more participants in the secondary program are shown.

provides a visualization of the sequence in which accounts tended to participate in National Grid's energy efficiency offerings shown in Table H-2. The table shows the details for a sequence of 2 programs, and the figure shows all pairings of participating in one program and then another regardless of whether it is the second, third, etc. program. The EWSF Program has the

most participants overall, regardless of whether the accounts started with the program, or started with another program.

Figure H-1. Sankey Participation Flow Diagram



Appendix I Correlation Analysis Results

Figure I-1. Correlation of Predictive Variables (Electric)

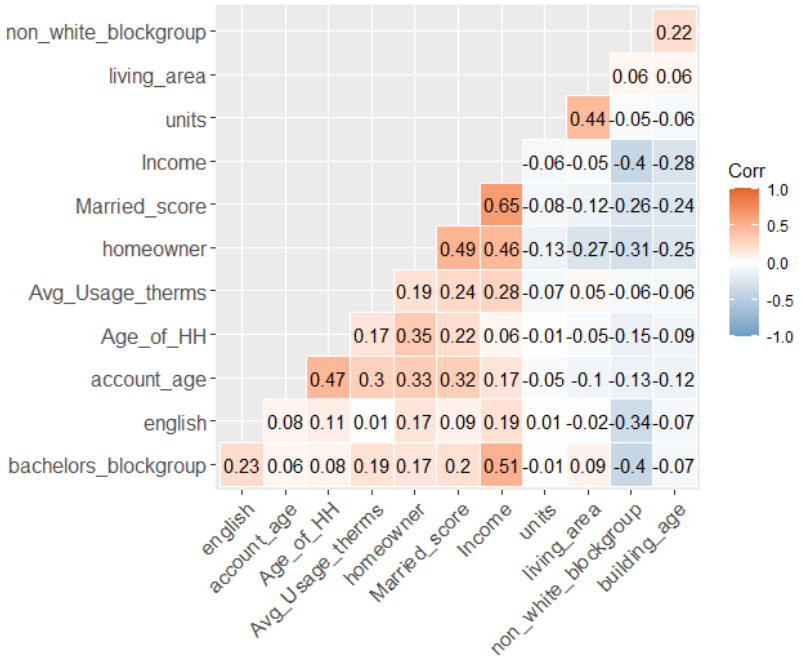
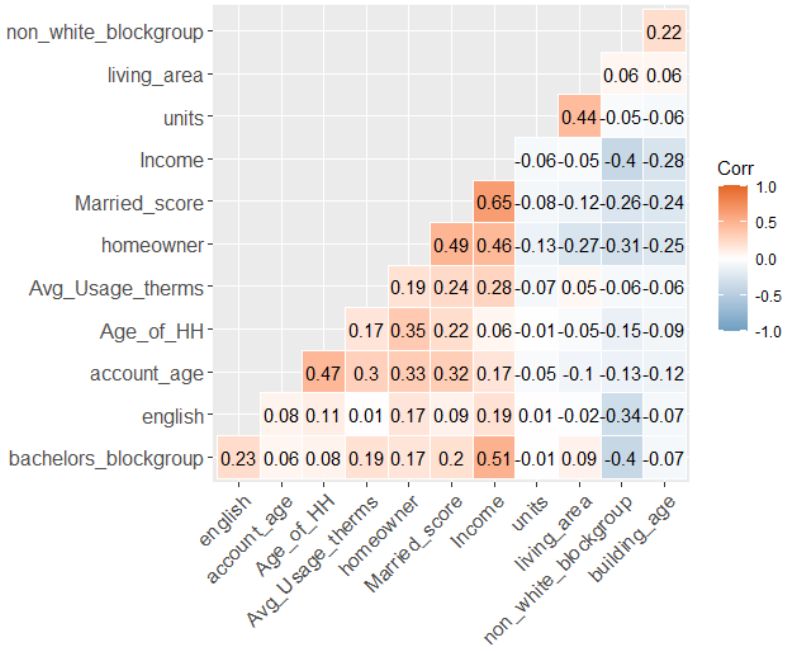


Figure I-2. Correlation of Predictive Variables (Gas)



Appendix J Model Coefficients

Table J-1. Participation

Variable	Value	Gas		Electric	
		Part rate	Coeff	Part rate	Coeff
Age_Head_of_Household	18 - 24	0.049303726	0	0.1194	0
Age_Head_of_Household	25 - 34	0.107413548	0.712067742	0.1927	0.4170188
Age_Head_of_Household	35 - 44	0.141575974	0.657053282	0.2463	0.3898074
Age_Head_of_Household	45 - 54	0.128412193	0.34449228	0.2304	0.1994525
Age_Head_of_Household	55 - 64	0.132386443	0.51172365	0.2347	0.3021746
Age_Head_of_Household	65 +	0.143273376	0.51932405	0.2469	0.4455141
Household_Income_Bucket	0 - 20	0.072260724	0	0.158	0
Household_Income_Bucket	20 - 40	0.094181113	0.310920849	0.1796	0.3253169
Household_Income_Bucket	40 - 60	0.115023474	0.48964963	0.2082	0.4159439
Household_Income_Bucket	60 - 80	0.138035126	0.691777944	0.2339	0.4374167
Household_Income_Bucket	80 - 100	0.149322614	1.413548811	0.2479	0.8150955
Household_Income_Bucket	100 - 120	0.16169066	0.09809666	0.265	0.3658819
Household_Income_Bucket	120 +	0.202347538	0.335932095	0.3076	0.7821004
Primary Language	English	0.139915964	0	0.23931	0
Primary Language	Non-English	0.100390382	-0.08495637	0.19911	-0.0230277
Home Ownership	Owner	0.151807959	-0.03463733	0.258	0.0380238
Home Ownership	Renter	0.070388587	0	0.147	0
units	1	0.257030352	0	0.2570304	0
units	2	0.135784101	-0.4506953	0.1357841	-0.2918899
units	3-4	0.086535165	-0.64060426	0.0865352	-0.5858314
units	5-6	0.063868893	-0.61404158	0.0638689	-0.9553856
units	7-10	0.109480813	0.037163189	0.1094808	-0.294569
units	11-19	0.163729809	-0.57262805	0.1637298	-0.0936941
units	20 - 49	0.298951049	-0.13960845	0.298951	0.2837641
units	50 +	0.327492832	0.556681927	0.3274928	0.8295631
year_built	< 1900	0.0802	0	0.1657	0
year_built	1900-1919	0.0689	-0.05187768	0.1438	0.0099389
year_built	1920-1939	0.09	-0.00861082	0.1855	0.113728
year_built	1940-1959	0.1385	0.118309072	0.2416	0.1218632
year_built	1960-1979	0.1514	0.180930264	0.2606	0.1736072
year_built	1980-1999	0.1938	0.445259539	0.2841	0.2408658
year_built	2000-2020	0.1813	0.365151941	0.2536	0.0676513
Age of Head of Household x Household Income	25 to 34 x $ 20 to $ 40		-0.08		-0.13
Age of Head of Household x Household Income	35 to 44 x $ 20 to $ 40		-0.17		-0.18
Age of Head of Household x Household Income	45 to 54 x $ 20 to $ 40		-0.1		-0.18

Variable	Value	Gas		Electric	
		Part rate	Coeff	Part rate	Coeff
Age of Head of Household x Household Income	55 to 64 x $ 20 to $ 40		-0.27		-0.26
Age of Head of Household x Household Income	65 + x $ 20 to $ 40		-0.22		-0.34
Age of Head of Household x Household Income	25 to 34 x $ 40 to $ 60		-0.1		-0.05
Age of Head of Household x Household Income	35 to 44 x $ 40 to $ 60		-0.32		-0.14
Age of Head of Household x Household Income	45 to 54 x $ 40 to $ 60		-0.18		-0.16
Age of Head of Household x Household Income	55 to 64 x $ 40 to $ 60		-0.4		-0.29
Age of Head of Household x Household Income	65 + x $ 40 to $ 60		-0.37		-0.41
Age of Head of Household x Household Income	25 to 34 x $ 60 to $ 80		-0.22		0
Age of Head of Household x Household Income	35 to 44 x $ 60 to $ 80		-0.39		-0.07
Age of Head of Household x Household Income	45 to 54 x $ 60 to $ 80		-0.29		-0.08
Age of Head of Household x Household Income	55 to 64 x $ 60 to $ 80		-0.48		-0.27
Age of Head of Household x Household Income	65 + x $ 60 to $ 80		-0.54		-0.38
Age of Head of Household x Household Income	25 to 34 x $ 81 to $100		-0.91		-0.33
Age of Head of Household x Household Income	35 to 44 x $ 81 to $100		-1.01		-0.49
Age of Head of Household x Household Income	45 to 54 x $ 81 to $100		-0.91		-0.46
Age of Head of Household x Household Income	55 to 64 x $ 81 to $100		-1.24		-0.65
Age of Head of Household x Household Income	65 + x $ 81 to $100		-1.19		-0.73
Age of Head of Household x Household Income	25 to 34 x $100 to $120		0.29		0.02
Age of Head of Household x Household Income	35 to 44 x $100 to $120		0.18		-0.04
Age of Head of Household x Household Income	45 to 54 x $100 to $120		0.21		-0.01
Age of Head of Household x Household Income	55 to 64 x $100 to $120		0.07		-0.22
Age of Head of Household x Household Income	65 + x $100 to $120		0.13		-0.27
Age of Head of Household x Household Income	25 to 34 x $120 or more		0.41		-0.11
Age of Head of Household x Household Income	35 to 44 x $120 or more		0.35		-0.17

Variable	Value	Gas		Electric	
		Part rate	Coeff	Part rate	Coeff
Age of Head of Household x Household Income	45 to 54 x $120 or more		0.36		-0.2
Age of Head of Household x Household Income	55 to 64 x $120 or more		0.04		-0.46
Age of Head of Household x Household Income	65 + x $120 or more		0.03		-0.6
Home Ownership x Age of Head of Household	Owner x 25 to 34		0.23		0.11
Home Ownership x Age of Head of Household	Owner x 35 to 44		0.13		0.05
Home Ownership x Age of Head of Household	Owner x 45 to 54		0.05		-0.03
Home Ownership x Age of Head of Household	Owner x 55 to 64		0.03		-0.04
Home Ownership x Age of Head of Household	Owner x 65 +		0.12		0.02
Home Ownership x Household Income	Owner x $100 to $120		0.24		0.38
Home Ownership x Household Income	Owner x $120 or more		0.1		0.37
Home Ownership x Household Income	Owner x $ 20 to $ 40		0.04		0.07
Home Ownership x Household Income	Owner x $ 40 to $ 60		0.1		0.14
Home Ownership x Household Income	Owner x $ 60 to $ 80		0.1		0.15
Home Ownership x Household Income	Owner x $ 81 to $100		0.03		0.2

Table J-2. SC Ratio

Variable	Value	SC_coef	
		Gas	Electric
Age_Head_of_Household	18 - 24	0	0
Age_Head_of_Household	25 - 34	-0.0072664	-0.0524295
Age_Head_of_Household	35 - 44	0.0003633	-0.0194519
Age_Head_of_Household	45 - 54	0.024736	-0.0210325
Age_Head_of_Household	55 - 64	0.054909	-0.0068137
Age_Head_of_Household	65 +	0.0005299	0.0287649
Household_Income_Bucket	0 - 20	0	0
Household_Income_Bucket	20 - 40	-0.0072664	-0.0031414
Household_Income_Bucket	40 - 60	0.0003633	-0.0301432
Household_Income_Bucket	60 - 80	0.024736	-0.0073765
Household_Income_Bucket	80 - 100	0.054909	-0.0571966
Household_Income_Bucket	100 - 120	0.0005299	0.0171826
Household_Income_Bucket	120 +	0.0316026	-0.0435778
Primary Language	English	0	0
Primary Language	Non-English	-0.0060804	0.0036609
Home Ownership	Owner	-0.019724	0.0043858

Variable	Value	SC_coef	
		Gas	Electric
Home Ownership	Renter	0	0
units	1	0	0
units	2	0.0083492	0.0012866
units	3-4	-0.0221961	-0.0012347
units	5-6	-0.013064	0.0066181
units	7-10	0.0774161	-0.0271132
units	11-19	0.0514084	0.0444038
units	20 - 49	-0.0663183	0.1037387
units	50 +	0.0485328	0.0111623
year_built	< 1900	0	0
year_built	1900-1919	0.0017209	-0.0003952
year_built	1920-1939	0.0105032	0.0113859
year_built	1940-1959	0.0298681	0.0026165
year_built	1960-1979	0.0261095	-0.003351
year_built	1980-1999	-0.01861	-0.0076537
year_built	2000-2020	-0.0141654	-0.0197538
Age of Head of Household x Household Income	25 to 34 x $ 20 to $ 40	0.031124	0.0317225
Age of Head of Household x Household Income	35 to 44 x $ 20 to $ 40	0.0297758	0.0148818
Age of Head of Household x Household Income	45 to 54 x $ 20 to $ 40	-0.0013949	0.0096178
Age of Head of Household x Household Income	55 to 64 x $ 20 to $ 40	-0.0108883	0.018065
Age of Head of Household x Household Income	65 + x $ 20 to $ 40	0.017429	-0.0021556
Age of Head of Household x Household Income	25 to 34 x $ 40 to $ 60	0.0023444	0.0558594
Age of Head of Household x Household Income	35 to 44 x $ 40 to $ 60	0.0219938	0.0257957
Age of Head of Household x Household Income	45 to 54 x $ 40 to $ 60	-0.0281446	0.0363146
Age of Head of Household x Household Income	55 to 64 x $ 40 to $ 60	-0.0657379	0.04417
Age of Head of Household x Household Income	65 + x $ 40 to $ 60	-0.0078113	0.0251221
Age of Head of Household x Household Income	25 to 34 x $ 60 to $ 80	0.0443701	0.0374521
Age of Head of Household x Household Income	35 to 44 x $ 60 to $ 80	0.0182305	-0.0068344
Age of Head of Household x Household Income	45 to 54 x $ 60 to $ 80	0.0171579	0.0076627
Age of Head of Household x Household Income	55 to 64 x $ 60 to $ 80	-0.0362995	0.0090558
Age of Head of Household x Household Income	65 + x $ 60 to $ 80	0.0060616	-0.0105471
Age of Head of Household x Household Income	25 to 34 x $ 81 to $100	-0.0333828	0.0760182
Age of Head of Household x Household Income	35 to 44 x $ 81 to $100	-0.0320743	0.0546727

Variable	Value	SC_coef	
		Gas	Electric
Age of Head of Household x Household Income	45 to 54 x $ 81 to $100	-0.0692495	0.0524138
Age of Head of Household x Household Income	55 to 64 x $ 81 to $100	-0.1066517	0.0550425
Age of Head of Household x Household Income	65 + x $ 81 to $100	-0.0541636	0.032929
Age of Head of Household x Household Income	25 to 34 x $100 to $120	0.0626879	0.0055247
Age of Head of Household x Household Income	35 to 44 x $100 to $120	0.0694859	-0.0222544
Age of Head of Household x Household Income	45 to 54 x $100 to $120	0.0166075	-0.0247842
Age of Head of Household x Household Income	55 to 64 x $100 to $120	-0.0071741	-0.0232037
Age of Head of Household x Household Income	65 + x $100 to $120	0.0434335	-0.045967
Age of Head of Household x Household Income	25 to 34 x $120 or more	-0.0349726	0.0889615
Age of Head of Household x Household Income	35 to 44 x $120 or more	-0.0387055	0.0438665
Age of Head of Household x Household Income	45 to 54 x $120 or more	-0.0796265	0.0475108
Age of Head of Household x Household Income	55 to 64 x $120 or more	-0.1129976	0.0454397
Age of Head of Household x Household Income	65 + x $120 or more	-0.0569929	0.0252126
Home Ownership x Age of Head of Household	Owner x 25 to 34	0.0060083	-0.0053661
Home Ownership x Age of Head of Household	Owner x 35 to 44	0.0133646	-0.0033153
Home Ownership x Age of Head of Household	Owner x 45 to 54	0.0083486	-0.0116729
Home Ownership x Age of Head of Household	Owner x 55 to 64	0.0104883	-0.0162504
Home Ownership x Age of Head of Household	Owner x 65 +	-0.0100494	-0.0140511
Home Ownership x Household Income	Owner x $100 to $120	0.0134188	0.000695
Home Ownership x Household Income	Owner x $120 or more	0.0080739	-0.0035486
Home Ownership x Household Income	Owner x $ 20 to $ 40	0.0025681	0.0028603
Home Ownership x Household Income	Owner x $ 40 to $ 60	0.0131541	0.0013242
Home Ownership x Household Income	Owner x $ 60 to $ 80	0.0105433	0.0009744
Home Ownership x Household Income	Owner x $ 81 to $100	0.0306395	0.0123358
Living Area	0 to 1,000 sq ft	0.0408967	0.0047403
Living Area	1,000 to 1,500 sq ft	0	0

Variable	Value	SC_coef	
		Gas	Electric
Living Area	1,500 to 2,000 sq ft	-0.0162395	0.0023791
Living Area	2,000 to 2,500 sq ft	-0.0261443	0.0068073
Living Area	2,500 to 3,000 sq ft	-0.038714	0.0120109
Living Area	3,000 to 4,000 sq ft	-0.0446919	0.0160629
Living Area	4,000 to 5,000 sq ft	-0.0506421	0.0132735
Living Area	5,000 to 6,000 sq ft	-0.086879	0.0147274
Living Area	6,000 sq ft +	0.0639725	0.0560796

Appendix K Examples of Condo Buildings and Algorithmic Classification

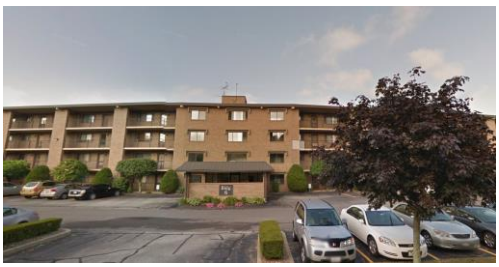
Multifamily buildings can have a variety of arrangements – many of which are difficult to differentiate visually and algorithmically from single-family buildings (as defined by National Grid). Any analysis of multifamily buildings will always have such ambiguities and influence the specific numbers reported. This appendix serves to highlight features of multifamily properties are often the reason behind a misclassification of single-family or treating each unit as a separate building. We do not have a specific identifier for buildings in our data. Keybdg is a generated field that is based on the address as provided and our best proxy for unique buildings. Several specific arrangements of multifamily buildings highlight limitations in the usage of keybdg to identify buildings. Other arrangements show limitations on making a classification only on building structure without considering HOAs or Co-op agreements.

One arrangement are buildings that include units side-by-side with a single attached wall. These buildings are often part of a larger property that has one or more similar buildings. Although the buildings are part of the same property, the data available to National Grid and provided to our team cannot always associate the units as part of the same building or property. Some such facilities give each building its own street number address using a separate unit or apartment number for each unit (see example 2). Others provide each unit with a separate address (see examples 3 & 4) even though units share a wall and roof structure. When looking at the data fields we have that provide a description of the dwelling type, this category of dwelling was classified as condos/apartments and sometimes as single-family dwellings.

The most difficult type of potential multifamily building to identify is the "detached" condo. These facilities often look like a neighborhood of single-family homes, duplexes, and/or triplexes (example 5). However, the facilities are identified as condos in real estate listings and may have some sort of homeowner's association responsible for the decisions necessary for program participation. It is also possible that the extent of the homeowner association is limited to other aspects of the building (e.g., paint color, mailbox type, landscaping) and individual condo owners would make independent decisions (more like a single-family program) regarding potential program improvements such as weatherization.

In Table K-1, examples 3-5 units are treated as separate buildings and counting the number of units in the building will always have them classified as single-family residences. The classifications provided in tax parcel data best capture the non-structural aspects of the building that align with program defined classification of multifamily. This is the primary data we use in our algorithmic classification.

Table K-1. Multifamily Condo Styles

#	keybdg	All Units have same keybdg	Google Image
1	02893 650 E GREENWICH AVE	Yes	
2	02893 117 SCENIC DR	Yes	
3	02893 18 PEPIN ST	No	
4	02911 32 SUNFLOWER CIR	No	
5	02818 30 FRYBROOK DR	No	

Appendix L Modeling Implementation Notes

Below are the specific model parameters used when fitting models. The first two were used when fitting LASSO regressions for classification for participation and for savings-to-consumption rations respectively.

sklearn.linear_model.**LogisticRegression**(*penalty='12', dual=False, tol=0.0001, C=10.0, fit_intercept=True, intercept_scaling=1, class_weight=None, random_state=None, solver= 'liblinear', max_iter=2000, multi_class='auto', verbose=0, warm_start=False, n_jobs=None, l1_ratio=None*)

sklearn.linear_model.**LassoCV**(**, eps=0.001, n_alphas=100, alphas=None, fit_intercept=True, normalize='deprecated', precompute='auto', max_iter=1000, tol=0.0001, copy_X=True, cv=None, verbose=False, n_jobs=None, positive=False, random_state=None, selection='cyclic'*)

*Sklearn.ensemble.**RandomForestClassifier**(n_estimators=100, *, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None, ccp_alpha=0.0, max_samples=None)*

*Sklearn.ensemble.**RandomForestRegressor**(n_estimators=100, *, criterion='gini', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, bootstrap=True, oob_score=False, n_jobs=None, random_state=None, verbose=0, warm_start=False, class_weight=None, ccp_alpha=0.0, max_samples=None)*