TECH CLEAN CALIFORNIA

Heat Pump HVAC Retrofit Cost Drivers

Impact of project and site features on the total installed cost of heat pump space heating retrofit projects in California single family homes

October 3, 2023

Dylan Sarkisian, Senior Data Scientist
Energy Solutions, dsarkisian@energy-solution.com

Desmond Kirwan, Senior Analyst
VEIC, dkirwan@veic.org

Sean Parker, Analyst
VEIC, sparker1@veic.org

Abigail Hotaling, Energy Data Analyst
VEIC, ahotaling@veic.org

Michael Fink
Energy Data Analyst
VEIC, mfink@veic.org

Energy Solutions veic

Table of Contents

Executive Summary

Electric heat pump HVAC equipment is used in less than 10 percent of California homes, but to meet our state's ambitious goals of installing six million heat pumps by 2030[1] and decarbonizing homes by 2045[2], heat pump HVAC must rapidly become the preferred option in the residential HVAC retrofit market. TECH Clean California, the state's flagship heat pump market transformation initiative, tests and scales strategies to spur heat pump adoption and put California on a path to carbon-free homes. A major barrier to heat pump adoption in residential retrofit markets is its higher purchasing cost compared to conventional gas alternatives. Meanwhile, myriad benefits like energy savings and grid value are not fully capturable at the time of installation.

By paying incentives for heat pump HVAC installations, TECH Clean California simultaneously helps customers overcome the first-cost barrier as well as collects detailed data on heat pump HVAC retrofits across the state. This data offers novel insight into which elements of heat pump HVAC retrofits affect cost the most. This report presents the TECH Clean California team's first attempt to quantify these cost drivers. Our results can help California homeowners, HVAC contractors, policymakers, agencies, other incentive program implementers, and a variety of other stakeholders make more informed investment decisions.

In this analysis, we use multivariate linear regression to measure the impact that 14 project- and site-related features ("covariates") have on total project cost. The model we built has limited predictive capacity, with an R2 of only 24 percent, but found statistically significant relationships between 12 covariates and total project cost. From these results, we drew conclusions including:

- As expected, project features that improve equipment performance also increase cost; e.g., higher SEER, performing a duct replacement, and performing Manual-D/Manual-J load calculations all increase total project cost. For many homes, this is a worthwhile investment.
- On average, performing an electrical panel upgrade alongside a heat pump HVAC retrofit increases total project cost by approximately $1,500.
- Disadvantaged community[3] status is not a cost driver.
- Installations in census tracts with old homes cost more, at a rate of $826 more per 10 years of the average age of owner-occupied homes in the census tract.
- Heat pump HVAC retrofits in homes with air conditioners cost approximately $1,000 less than those without air conditioners, on average.
- Projects in counties served by more TECH-enrolled contractors cost less; i.e., projects in counties served by 100 TECH-enrolled contractors cost $1,031 less, on average, than projects in counties served by 10 TECH-enrolled contractors.

Following this analysis, the TECH Clean California team plans to improve our modeling approach with new data types such as more specific data about the homes in which equipment was installed, the contractors performing the installation, and the installed brands. Following this, we intend to apply our cost driver analysis methods to heat pump water heaters, which are also eligible for TECH Clean California incentives. Finally, we plan to use meter data to measure savings for both heat pump HVAC and heat pump water heater installations, identify savings drivers, then compare cost drivers with savings drivers for both technologies.

---

1 "Governor Newsom Calls for Bold Actions to Move Faster Towards Climate Goals", July 2022. www.gov.ca.gov/2022/07/22/governor-newsom-calls-for-bold-actions-to-move-faster-toward-climate-goals

2 California Air Resources Board, 2022 Scoping Plan. ww2.arb.ca.gov/sites/default/files/2022-11/2022-sp.pdf

3 Disadvantaged Community status is defined for each California census tract by California EnviroScreen 4.0

Data Collection

The cost driver analysis uses data collected from TECH Clean California-funded heat pump installation projects performed from December 2021 through June 2023. This section describes how this data was collected from TECH Clean California enrolled contractors and assembled for use in the cost driver analysis. Herein, “installation,” “project,” and “claim” all refer to a heat pump installation that qualified for and received an incentive via the TECH Clean California Incentive Clearinghouse.

Incentive Overview

Beginning in December 2021, TECH Clean California has offered incentives to enrolled contractors who install qualifying heat pump heating, ventilation, and air-conditioning (“HVAC”) and heat pump water heater equipment. Contractors must hold the required license and sign a Trade Partner Participation Agreement to enroll. To receive a TECH Clean California incentive, an enrolled contractor must submit an incentive application via the TECH Clean California Incentive Clearinghouse[4] proving that a qualifying model was installed in a home eligible for TECH Clean California incentives and providing several other data points describing the installation.

Incentive Application Data Collection

Given the heat pump HVAC and heat pump water heater markets were both nascent upon the launch of TECH Clean California incentives, our team had to strike a balance between two conflicting motivations for incentive application data collection: (1) make the application process as simple as possible so contractors considered the time investment to perform a heat pump HVAC installation and complete the TECH Clean California incentive application worthwhile, and (2) collect as much data as possible on the incentive applications in order to maximize our understanding of heat pump HVAC retrofits. Given these constraints, our team set the following goals for the TECH Clean California incentive application:

1. Minimize contractor burden
2. Verify eligibility
3. Document important project features

Designing the TECH Clean California incentive application required careful compromise at every step. One fundamental compromise made in the TECH Clean California incentive application design was the decision to require contractors to report the total project cost but not require further disaggregation of the cost components. Contractors were instructed to “report total cost of the installation and related measures (material, labor, permitting, etc.) prior to all incentives being applied.” This decision was made for multiple reasons. First, it simplified the application process, aligning with minimizing contractor burden. Second, requiring disaggregation of total project cost into components would have helped delineate equipment costs, but it would also have effectively required contractors to disclose labor rates, overhead, and profit margin — which not only would have discouraged many contractors from participating but also, since TECH Clean California data is published online, would have created risk of a race to the bottom where contractors began competing on price. Third, and most relevant to this analysis, is that the most useful data to collect is not what individual contractors charge for project components but rather what all contractors, on average, charge. Thus, our team resolved to collect only

4 The TECH Clean California incentive clearinghouse is accessible only to enrolled contractors at catechincentives.com

total project cost and also to collect a variety of data points describing the project such that we could use statistical models to estimate the average cost of each component for the entire TECH Clean California claims dataset.

TECH Clean California incentives for heat pump HVAC and heat pump water heater installations replacing gas equipment — the vast majority of installations — were launched in December 2021 with baseline values of $3,000 and $3,100, respectively, in all eligible territories, with additional incentives offered for quality installation measures in select regions where partner programs layered incentives with TECH Clean California (see Table 1). Though our team set incentive levels with the intention of attracting a roughly equivalent number of heat pump HVAC and heat pump water heater incentive applications, the single family heat pump HVAC market moved more quickly to incorporate TECH Clean California incentives into contractor business models — especially in Southern California — and by May 2022 so much of the TECH Clean California incentive budget had been used that our team paused incentive applications throughout most of the state. Over 75 percent of the TECH Clean California single family incentive budget had been used by heat pump HVAC projects, and over 60 percent of these occurred in Southern California. After receiving additional funding through the California state 2022–2023 fiscal year budget, the TECH team re-launched TECH incentives for single family heat HVAC projects in April 2023 but has not yet re-launched incentives for single family heat pump water heater projects.[5]

Heat Pump HVAC Application Data

Given most TECH incentives paid-to-date are for heat pump HVAC installations in single family homes, our team has focused our first cost driver analysis on single family heat pump HVAC projects. Table 1 shows data types collected via the TECH Clean California single family heat pump HVAC incentive application claim form as well as those appended to incentive application data. Each claim represents a unique installation of a single model number in a home, though one or two units of that model number can be installed. Appendix A lists all data fields assembled by our team for heat pump HVAC applications. New data fields were included on the single family heat pump HVAC incentive application form when incentives re-launched in April 2023, but not enough projects have been paid since April 2023 for these new data fields to be a significant contributor to the cost driver analysis.

5 The latest available incentive types and remaining budgets can be tracked on the TECH public reporting website Incentives webpage, techcleanca.com/incentives

Table 1: Data Fields Collected on TECH Clean California Single Family Heat Pump HVAC Incentive Application and Appended Via Incentive Clearinghouse

Category	Questions asked on TECH incentive application	Data types appended to/created using data provided in application
Category	Questions asked on TECH Clean California incentive application	Data types appended to data provided in application
Contractor	Contact info, license #	Total number of TECH Clean California-funded projects performed by each contractor
Customer	Address, contact info	Eligibility, climate, census tract data – via geocoding of customer address
Home	Total floor area in square feet, electrical panel capacity pre- and post-installation in Amperes	
New Equipment	Model #, serial #, quantity of units installed	Specifications including cooling and heating capacity, SEER, EER, and HSPF, and equipment type — via QPL1
Prior Equipment	Furnace fuel type (Gas, Electricity, Other) Furnace model number Air Conditioning type (Room Unit, Central, None) Air conditioner model number	
Installation	Installation Start Date Installation End Date Furnace status after heat pump HVAC installation Quality Installation Measures: Electrical panel upgrade Duct replacement Duct sealing Smart thermostat installed Manual-D/Manual-J ASHRAE 221 Performance Report	Installation duration Number of TECH-funded projects performed by each contractor involving a QIM
Cost	Total Project Cost ($)	Normalized Cost ($ per ton of installed cooling capacity)

1 The Air Conditioning, Heating, and Refrigeration Institute ("AHRI") Qualified Product List was used to look up equipment specifications using the model number provided by the contractor in the incentive application.

All data gathered from paid TECH Clean California incentive applications is published monthly in an anonymous format on the public reporting website on the Download Data webpage at techcleanca.com/public-data/download-data.

Methodology

The Data Collection section explains that our team only collected the total project cost, not individual cost drivers, for each heat pump HVAC incentive application submitted by an enrolled contractor. However, we also collected a variety of other data points expected to impact the cost or performance of the installed equipment. This section describes our team's motivation, goals, and methodology for the cost driver analysis.

Motivation

The California 2019 Residential Appliance Saturation Survey (RASS) showed that heat pump HVAC equipment is used for space heating in less than 10 percent of homes, while gas furnace equipment continues to be used in over 80 percent of homes.[6] Newly constructed homes now typically use heat pumps for space and water heating, but the pace of new construction is not great enough for our state to reach its goals of installing six million heat pumps by 2030 and decarbonizing homes by 2045. To meet our state's ambitious decarbonization goals, heat pump HVAC must become competitive and widespread in the residential HVAC retrofit market.

A barrier to adoption of heat pump HVAC in residential retrofit markets is that both the equipment and installation is often more expensive than conventional alternatives like gas furnaces and air conditioners. By collecting detailed data on thousands of incentivized heat pump HVAC retrofit installations across the state, the TECH Clean California has a unique opportunity to learn which elements of heat pump HVAC retrofit projects affect cost and how much.

The first cost is not the only cost of owning a heat pump HVAC, and we recognize that many components of a heat pump HVAC installation that increase the upfront cost both improve performance and potentially reduce the owning and operating costs of the system, so could ultimately be cost-effective. To measure this, the TECH team is also performing meter-based electricity and gas savings analysis for TECH-funded heat pump HVAC installations. We plan to compare the savings versus cost impacts of heat pump HVAC project components to determine which components should be recommended or required by future incentive programs and equipment standards supporting heat pump HVAC adoption.

Beyond serving the strategic goals of the TECH team, this cost driver analysis is also a critical way to increase heat pump HVAC market transparency for our stakeholders. The lynchpin of the TECH Clean California market transformation model is that publishing data catalyzes market growth by increasing predictability and thus encouraging investment. This cost driver analysis aims to provide stakeholders with a deeper understanding of heat pump HVAC project costs, so more informed investments are made.

---

6 The 2019 California Residential Appliance Saturation Study can be accessed here: www.energy.ca.gov/data-reports/surveys/2019-residential-appliance-saturation-study

Goals

To guide the methodological choices the team used in the cost driver analysis, we set the following goals:

- Determine which project features contribute significantly to the total project cost of heat pump HVAC retrofit installation projects in single-family homes.
- Estimate the amount that each project feature contributes, on average, to a typical heat pump HVAC installation project.
- Measure how much of the variation in total project cost across all TECH-funded heat pump HVAC projects can be explained by interpretable relationships with project features.
- Create a baseline against which to measure future model results, especially cost prediction models.

Model Overview

Given the goals of the cost driver analysis, our team prioritized building a model with easily interpretable results. Our priority was more to enable statistical inference than prediction. Statistical inference is the process of using a sample to infer the properties of a population. Here, we use the cost drivers we measure for TECH-funded heat pump HVAC retrofits to infer the cost drivers for any comparable heat pump HVAC retrofit in California.

The TECH Clean California team chose to construct a multivariate linear regression model for the heat pump HVAC Cost Driver Analysis. We chose linear regression because of its easy interpretability[7], despite understanding that many project features likely do not have a linear relationship with total project cost. We applied transformations to project feature variables where possible to help account for this, but we still expect our model to present an overly simple picture of the population. Furthermore, given the primary goal of our model is inference rather than prediction, the TECH team decided to train our model on the entire available dataset rather than using a training set and test set to measure the model's ability to predict outcomes on "new" data. For a similar reason, we did not use variable selection methods like stepwise selection or ridge regression; our team began the analysis with a limited set of project features to choose from and a strong idea of which variables would have a significant influence on total project cost, so we prioritized building a model to understand the relationship between all project features and total project cost rather than finding the subset of features with the strongest predictive power. We discuss recommended modifications to the model to improve prediction under Next Steps.

The TECH team used Python implemented in a Jupyter notebook via Anaconda to build the multivariate linear regression model. We used this approach because Anaconda Jupyter are a free and widely used data science coding environment that allows extensive commenting in markdown to improve interpretability. We used Ordinary Least Squares multivariate linear regression via the Python statsmodels library "OLS" class.

---

7 Linear regression is often favored over nonlinear models for analyses with large intended audiences due to its inherent interpretability of results. The coefficients associated with each predictor indicate the direction and magnitude of the effect, allowing for straightforward interpretation. Nonlinear models, on the other hand, introduce complexity that hinders interpretability. Coefficients in nonlinear models do not directly translate into easily understandable explanations of how variables interact.

Data Overview

This section examines the variables analyzed in the Cost Driver Analysis and describes data cleaning and transformation used to prepare the dataset used to train the model. See Appendix A for a full list of data fields considered and selected for the Cost Driver Analysis.

Dependent Variable: Total Project Cost

The purpose of the Cost Driver Analysis is to measure how site and project features influence the total cost of installing a heat pump HVAC in an existing home, including equipment, labor, and related activities performed by the contractor. In short, we refer to this as total project cost. Total project cost is collected directly from the TECH Clean California Heat Pump HVAC incentive application in a data entry field titled "Invoice Total," completed by the contractor. On the application, contractors are instructed to "report the total cost of the heat pump installation and related measures (material, labor, permitting, etc.) prior to all incentives being applied."

Data is first cleaned and verified in the application processing system. Figure 1 shows that the distribution of observed total project cost for heat pump HVAC project receiving TECH Clean California incentives tapers off to a minimal fraction of projects beyond $40,000 and virtually none beyond $60,000. For this reason, any heat pump HVAC incentive application with a total project cost greater than $70,000 is flagged for review by the TECH Clean California application processing team. These applications are generally rejected, and those that are not have their total project cost scrubbed from the TECH Clean California project dataset because they are not comparable to what virtually all California homeowners should expect a heat pump HVAC installation to cost. Similarly, any application with a total project cost less than $3,000 has its total project cost removed from the TECH Clean California project dataset.

To prepare the total project cost data field for use in the Cost Driver Analysis, we aimed to make all projects as comparable as possible, controlling for variation in the sizes and dates of TECH Clean California-funded heat pump HVAC installations. We first normalized the total project cost field by the total rated cooling capacity of the installed unit(s). Heat pump HVAC equipment has both a rated cooling capacity and heating capacity, and the mandatory section 150.0 of the California Title 24, Part 6 building codes requires that both cooling load and heating load be used to size equipment. However, our team selected cooling capacity as the normalization factor because contractors typically prioritize sizing heat pump HVAC installations to meet the home's cooling needs. Our team also normalized total project costs for inflation by adjusting reported dollar values to July 2023 present value dollars using the Python "cpi" library, which adjusts prices using the Consumer Price Index. This is especially critical given the relatively high inflation observed in the CPI and most consumer products since TECH Clean California incentives became available in December 2021. This transformation yielded a total project cost field whose units are expressed as present value dollars per ton.

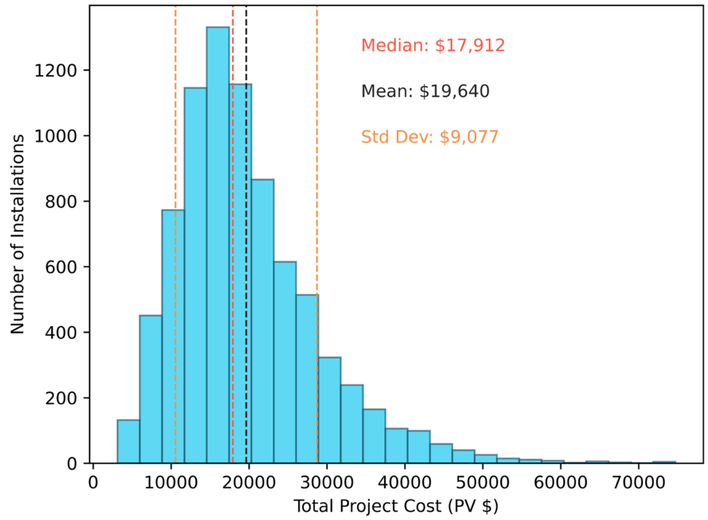


FIGURE 1: HISTOGRAM OF INFLATION-ADJUSTED TOTAL PROJECT COST OF TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC PROJECTS (N = 9,744)

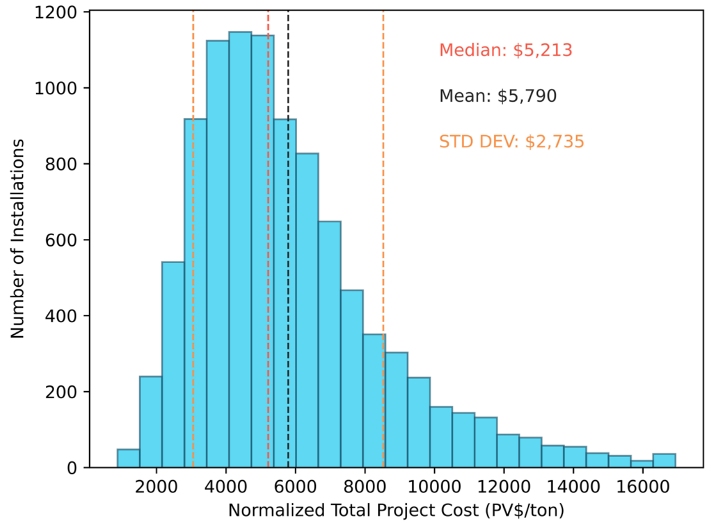


FIGURE 2: HISTOGRAM OF INFLATION-ADJUSTED COST OF TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC PROJECTS PER TON OF INSTALLED COOLING CAPACITY (N = 9,744)

Finally, to remove outliers, our team removed projects from the dataset whose normalized total project cost was in the top one percent of normalized total project costs observed in TECH Clean California-funded projects. Figure 1 shows the distribution of inflation-adjusted total project cost, while Figure 2 shows the same distribution now normalized by tons of cooling capacity.

Most TECH-incentivized heat pump HVAC projects had a cooling capacity of three tons, so one would expect the standard deviation of the distribution in Figure 2 to be three times less than that of Figure 1. However, Figure 2's standard deviation is 3.3 times than that of Figure 1. This, as well as the lower kurtosis in Figure 2, shows that normalization helps reduce variance in total project cost.

Independent Variables

The large variance in total project cost reported for TECH Clean California-funded heat pump HVAC installations drove our team to identify as many variables as possible that could contribute to these costs. The full list of variables considered is listed in Appendix A. We provide here a detailed account of how we transformed the selected variable into useful data fields for the HVAC Cost Driver Analysis.

EQUIPMENT TYPES

A variety of heat pump HVAC equipment types are eligible for TECH Clean California incentives, and all eligible models are listed in the AHRI QPL. When an incentive application is submitted to the TECH Clean California Incentive Clearinghouse, specifications about the model as well as key categorizations are derived from the AHRI QPL. Table 2 lists equipment type eligible for TECH Clean California Incentives by AHRI Material Category and Type code.

Table 2: Types of Heat Pump HVAC Equipment Eligible for TECH Clean California Incentives

Material Category	AHRI Type	Description	Equipment Type	Ducting Type
ushp1	HSP-A	Single-Package Heat Pump, Air Source	Packaged Unitary Equipment	Ducted
ushp	HRCU-A-C	Heat Pump with Remote Outdoor Unit, No Indoor Fan, Air Source	Split Unitary Equipment	Ducted
ushp	HRCU-A-CB	Split System Heat Pump with Remote Outdoor Unit, Air Source	Split Unitary Equipment	Ducted
ushp	HRCU-A-CB-O	Split System Heat Pump with Remote Outdoor Unit, Air Source, Free Delivery	Split Unitary Equipment	Ductless
ushp	SDHV-HRCU-A-CB	Small Duct High Velocity System Heat Pump with Remote Outdoor Unit, Air Source	Small Duct High Velocity	Ducted
vsmshp	HRCU-A-CB	Minisplit System Heat Pump with Remote Outdoor Unit, Air Source, Free Delivery	Minisplit	Ducted
vsmshp	HRCU-A-CB-O	Minisplit System Heat Pump with Remote Outdoor Unit, Air Source, Free Delivery	Minisplit	Ductless
vsmshp	HMSV-A-CB	Multisplit System Heat Pump, Air Source	Multisplit	Ducted
vsmshp	HMSV-A-CB-O	Multisplit System Heat Pump, Air Source, Free Delivery	Multisplit	Ductless
vsmshp	HMSR-A-CB	Multisplit System Heat Pump with Heat Recovery, Remote Outdoor Unit, Air Source	Multisplit	Ducted
vsmshp	HMSR-A-CB-O	Multisplit System Heat Pump with Heat Recovery, Remote Outdoor Unit, Air Source, Free Delivery	Multisplit	Ductless

1 The AHRI material category "ushp" includes all unitary package and split system heat pumps. These units can be ducted or ductless and typically do not have variable-speed compressors.

2 The AHRI material category "vsmshp" includes all variable-speed mini and multisplit heat pumps. These units can be ducted or ductless and always have variable-speed compressors.

The conventional terms listed in “Equipment Type” and “Ducted” columns of Table 2 are commonly used to describe HVAC equipment, so our team created a single variable merging these two terms to use in the HVAC Cost Driver Analysis. As shown in Figure 3, certain HVAC equipment categories were more popular with TECH Clean California participating contractors than others, with Ducted Split Unitary systems installed in almost two-thirds of single family homes. We removed projects from the HVAC Cost Driver Analysis dataset where ductless split unitary and small duct high velocity systems were installed because these represented only a fraction of a percent of all projects.

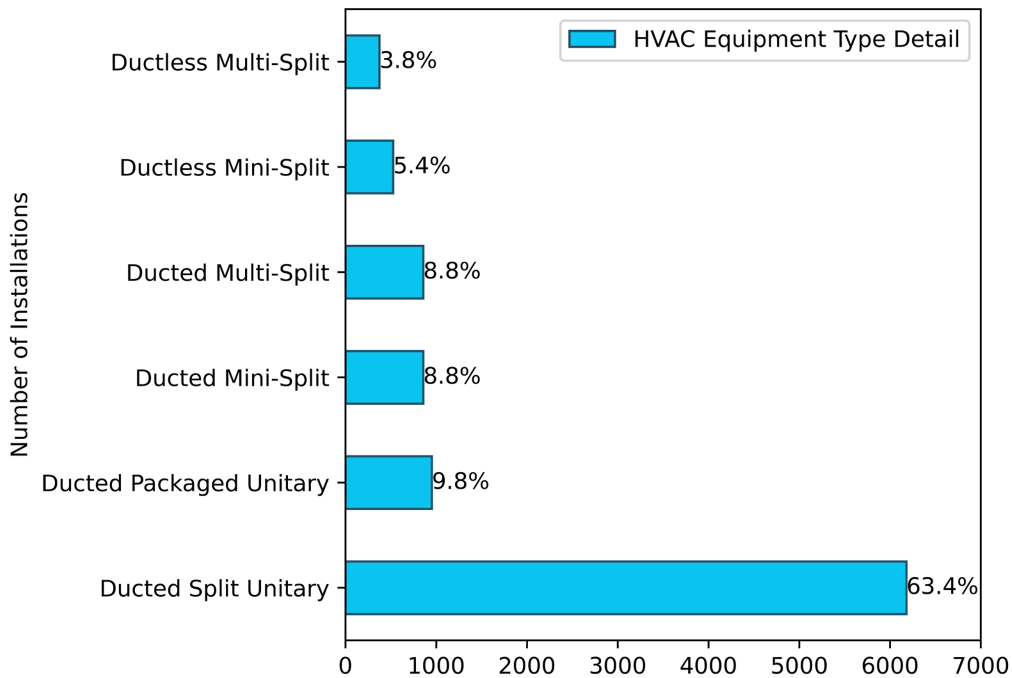


FIGURE 3: FRACTION OF TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC INSTALLATIONS BY EQUIPMENT TYPE (N = 9,744)

PROJECT FEATURES

The categorization of heat pump HVAC equipment types forms a scaffolding for the heat pump HVAC Cost Driver analysis because the most important distinction between projects is the type of equipment installed. However, many elements of a heat pump HVAC retrofit beyond the type of equipment installed impact the total project cost, especially given the immaturity of the heat pump HVAC market in California. The elements of heat pump HVAC retrofits that most strongly affect total project cost can be divided into project features and site features. Here we describe the project features included in the Heat Pump HVAC Cost Driver Analysis.

Table 3 lists each project feature data field included in the Heat Pump HVAC Cost Driver Analysis, the range of data observed in this field in the project dataset, and the transformations performed by the TECH team to prepare each field to use in the model.

Table 3: Project Features and Transformations Used in the Heat Pump HVAC Cost Driver Analysis

Category	Field	Data	Transformation
Installed Equipment Specifications	Seasonal Energy Efficiency Ratio ("SEER")	Ranges from 14.0 to 29.4, mean of 17.1. Distribution in Figure 4	Create SEER Minus Minimum in order to measure cost impact of per unit of SEER added over minimum
	Cooling Capacity	Ranges from 1.0 to 5.0, mean of 3.3. Distribution in Figure 5	Used to normalize Dependent Variable; not included as an Independent Variable
Replaced Equipment	Previous Air Conditioner Type	None: 48% Central: 50% Room Unit: 2%	Created "Air Conditioner Present" field by combining "Central" and "Room Unit" because "Room Unit" was too small a portion of projects
	Furnace Status After Installation	Decommissioned: 87% Setup to run in emergency scenarios only: 12% Setup to use the blower only: 1%	Created "Furnace Decommissioned field by combining "Left to run in emergencies only" and "Left to run the blower only" because these alone were too small a portion projects.
	Furnace Fuel Type	Natural Gas: 98% Electric Resistance: 1% Other: 1%	Not included as an Independent Variable because not enough variation
Quality Installation Measures	Electrical Panel Upgrade (T/F)	4%	N/A
	Duct Replacement (T/F)	15%	N/A
	Duct Sealing (T/F)	15%	Not included in the model due to high correlation with Duct Replacement – see Table 4
	Manual D/J Completed (T/F)	7%	N/A
	Performance Report Completed (T/F)	3%	Not included in the model due to low quantity of contractors that used this QIM.
	Smart Thermostat Installed (T/F)	44%	N/A
Installation Duration	Installation Duration (Days)	Ranges from 1 to 366, mean of 5.	Log transformation – because cost impact is expected to be greatest for low values
Contractor Participation	Number of TECH Heat Pump HVAC Projects Performed by the Contractor who Performed the Installation	Ranges from 1 to 406, mean of 89.	Log transformation – because cost impact is expected to be greatest for low values

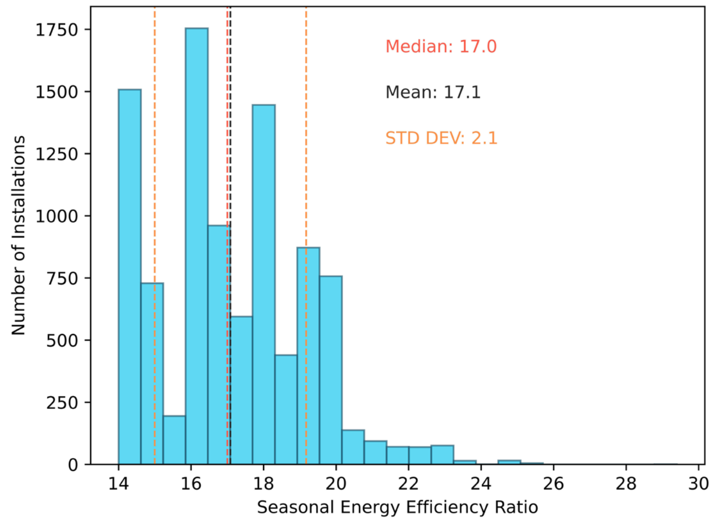


FIGURE 4: HISTOGRAM OF SEER OF TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC PROJECTS (N = 9,744)

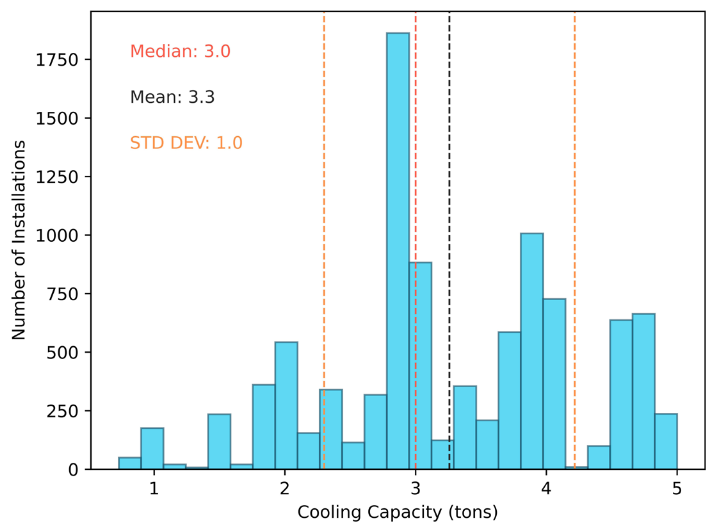


FIGURE 5: HISTOGRAM OF COOLING CAPACITY OF TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC PROJECTS (N = 9,744)

While preparing data to use in the heat pump HVAC cost driver analysis, our team measured correlations among project features. We suspected that many contractors had performed similar projects repeatedly, so certain project features would be highly correlated — especially the quality installation measures. Table 4 presents the frequency and correlation of quality installation measures. This illustrates why we chose to not include the duct sealing field, since it has an almost 50 percent correlation with duct replacement. Including two highly correlated variables in a linear regression model muddles the interpretability of the results. We selected duct replacement because we expected this to have a greater incremental cost than duct sealing, meaning it could help explain more of the observed variance of total project cost. The TECH Clean California team also measured multicollinearity of all variables using variance inflation factors, discussed in Multicollinearity.

Table 4: Quality Installation Measure Frequency and Correlations

Quality Installation Measure	Frequency of use in projects (%)	Correlation with: (%)				
		Panel Upgrade	Duct Replacement	Duct Sealing	Manual D/J	Smart Thermostat
Panel Upgrade	4	--[1]				
Duct Replacement	15	12	--			
Duct Sealing	15	7	46	--		
Manual D/J	7	6	16	16	--	
Smart Thermostat	44	9	28	30	20	
Performance Report	3	7	12	15	12	10

1 Variables always have 100% correlation with themselves, but we omitted this to help focus attention on the correlation of each unique combination of two distinct variables.

SITE FEATURES

In addition to project features, our team predicted that the location and features of the home in which equipment is installed significantly impact total project cost for a heat pump HVAC retrofit. We used a combination of data gathered from contractors via incentive applications, US Census data, and TECH Clean California contractor enrollment data to create data fields representing site features that we predicted would have the strongest influence on total project cost. Table 5 summarizes the selected site features and their transformation to prepare for use in the Cost Driver Analysis.

Table 5: Site Features and Transformations Used in Heat Pump HVAC Cost Driver Analysis

Category	Field	Data	Transformation
Home	Home Floor Area	Ranges from 500 to 10,000, mean of 2,033 sq ft. distribution shown in Figure 6.	Created "Home Floor Area Minus Minimum" field in order to measure impact of additional floor area above 500 sq ft on total project cost
Census Tract	Average Age of Owner-Occupied Housing in the Census Tract in which the Project Occurred	Ranges from 11 to 103 years, mean of 50 years. Distribution shown in Figure 7.	Created "Average age of Owner Occupied Housing Minus Minimum (10 Years)" field in order to measure impact of each additional decade of average home age in the Census Tract on total project cost, starting at 11 years old.
County	Number of Counties Served by the Contractor who Performed the Installation	Ranges from 1 to 15, mean of 3.	N/A
	Number of TECH Clean California Contractors Serving the County in which the Project Occurred	Ranges from 11 to 279, mean of 144.	Log transformation — because cost impact is expected to be greatest for low values

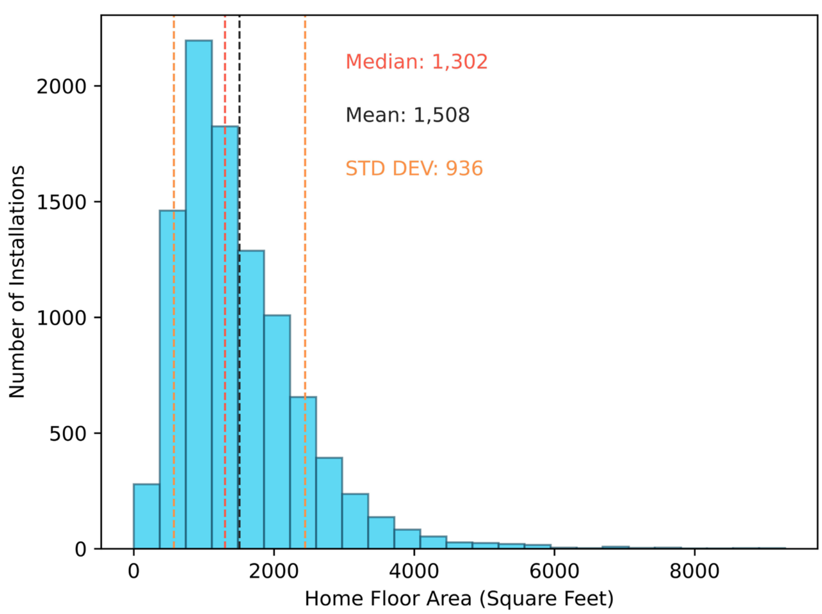


FIGURE 6: HISTOGRAM OF HOME FLOOR AREA FOR TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC INSTALLATIONS (N = 9,744)

AVERAGE AGE OF OWNER-OCCUPIED HOUSING

Housing vintage data was collected using U.S Census American Community Survey (ACS) data to provide some insights on the effects of housing age on the total costs of projects. ACS table S2504 Physical Housing Characteristics for Occupied Housing Units provides estimated raw counts and percentages of the number of houses within each census tract in California based on specified timespan bracket intervals in decades the homes were built. The table also breaks the data down between total, owner-occupied, and renter-occupied housing units. For this analysis, only owner-occupied housing units' data was analyzed. Using count and percentage estimates, an average weighted age was calculated for each census tract to be used in the regression model to estimate the building vintages for projects in the TECH Clean California project data.

The year 2023 was used as a baseline to calculate the weighted average ages of owner-occupied homes per census tract. The midpoint of each designated timespan bracket in the census data was used as a reference point to perform the weighted age calculation. For example, between the decade timespan between 2010 and 2020, the year 2015 was used as a single year reference. However, in some instances timespans were listed as a 20-year interval brackets instead. In each case, the midpoint year was used to create a singular year reference. For the 1939 or earlier bracket, the midpoint was assumed to be between the end year and start of the century. An exception was made for the average age of homes listed in the 2020 or later timespan bracket due to the shorter time of reference compared to all other brackets. Instead of using a midpoint year, it was determined an average age of three years would be used as an assumption instead. This three-year average was used to calculate the weighted age for each census tract in this timespan bracket.

The midpoint of each decade was subtracted from the baseline year to provide a singular estimated average age within each timespan bracket. For example, using the designated timespan listed in the previous example, the midpoint year 2015 yields an estimated average age of eight years compared to the baseline. These singular estimated ages for each timespan were then multiplied by the estimated percentage of homes for each census tract to yield a weighted age of each owner-occupied home within each timespan bracket for each census tract. Lastly, the weighted ages for each timespan bracket were added together for each census tract to provide a singular weighted age for each census tract. Figure 7 shows the distribution of the calculated census tract average age of owner-occupied housing for projects included in the heat pump HVAC cost drivers analysis.

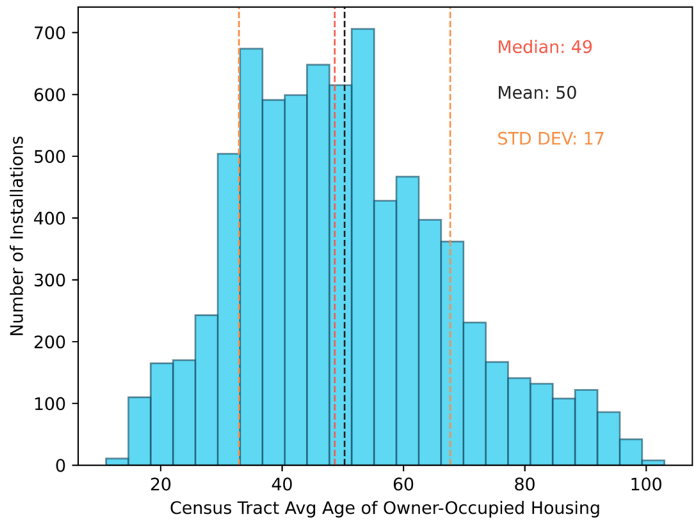


FIGURE 7: HISTOGRAM OF AVERAGE AGE OF OWNER-OCCUPIED HOUSING FOR TECH CLEAN CALIFORNIA-FUNDED HEAT PUMP HVAC PROJECTS (N = 9,744)

CONTRACTOR QUANTITY OF COUNTIES SERVED AND COUNTY QUANTITY OF CONTRACTORS SERVING

A key factor contributing to the cost of a product requiring a skilled and specialized installer like a heat pump HVAC system is the availability of installers in a geographic area. Data on service territories was collected from each TECH-enrolled contractor. Specifically, a list of each county served by each contractor was collected at the time of certification, and our team used this list to create a value describing the number of counties served for each contractor. This data was then mapped on to the project data, showing the number of counties served by the contractor who performed each installation.

In addition, we also generated the inverse variable: the number of contractors serving each county. Using similar methods, our team calculated the total number of TECH-enrolled contractors serving each county, then mapped this result onto the project data. This variable shows how many TECH-enrolled contractors serve the county in which each project occurred. Including both independent variables into the regression model allowed us to measure the impacts of contractor size and availability on total project cost.

Multicollinearity

To maximize the accuracy and interpretability of the Cost Driver Analysis model outputs, we measured the correlation between all planned independent variables, or "covariates," as well as multicollinearity. Highly correlated covariates and covariates with a high multicollinearity both reduce the meaningfulness of linear models' outputs. There is no universally accepted maximum tolerable correlation or multicollinearity, but our team aimed to avoid including any covariates with a correlation greater than 30 percent or a Variance Inflation Factor (VIF), a measure of multicollinearity, greater than 10.

We succeeded in not including any covariates with a greater than 30 percent correlation. However, as Table 6 shows, we chose to keep two covariates with a VIF greater than 10: SEER and Census Tract Average Age of Owner-Occupied Housing. Seeing that these two covariates had high VIF scores, we removed three other highly correlated covariates, but this did not reduce the VIF scores to below 10. We kept these covariates in the model because we expected them to have a strong influence on total project cost, and we could not find alternate data that we could use as a replacement for these covariates.

Table 6: Variance Inflation Factors for Quantitative Covariates Considered for the Cost Driver Analysis Before and After Removing Some Covariates to Reduce Multicollinearity

Covariate	Variance Inflation Factor	Variance Inflation Factor minus Highly Correlated Fields
Panel Upgrade (True/False)	1.1	1.1
Duct Replacement (True/False)	1.3	1.3
Performance Report Completed (True/False)	1.1	1.1
Smart Thermostat Installed (True/False)	2.1	2.1
HVAC Project Count Performed by Contractor	2.0	2.0
Installation Duration (days)	1.1	1.1
Seasonal Energy Efficiency Rating	21.7	17.7
Home Floor Area (sq ft)	6.8	5.6
County Quantity Contractors Serving	5.9	4.5
ACS Economics Median Household Income ($)	7.0	--
Disadvantaged Community1 (True/False)	1.3	--
Hard-to-Reach County2 (True/False)	1.6	--
Contractor Quantity of Counties Served	1.9	1.9
Census Tract Average Age of Owner-occupied Housing (years)	10.5	10.1

1 Disadvantaged Community is defined by California EnviroScreen 4.0

2 Hard-to-reach communities are defined in CPUC Resolution G-3497

A notable finding from the correlation analysis is that the number of TECH-enrolled contractors serving a county is strongly negatively (-52 percent) correlated with a county's hard-to-reach status as defined by CPUC in Resolution G-3497. A county's hard-to-reach status represents its urban characteristics, so this negative correlation shows that more TECH-enrolled contractors serve urban California counties in the San Francisco Bay Area, Greater Los Angeles Area, Greater Sacramento Area, and Greater San Diego Area. This verifies that the rural counties are in fact harder to reach, at least for TECH-enrolled heat pump HVAC contractors. Some correlation is expected due to a difference in population density, however.

A more surprising result of the correlation analysis is that SEER has a strong correlation with multiple site features, but little to none with project features, as shown in Figure 8. The strong correlation of SEER with hard-to-reach county status may be because the larger temperature variation in rural counties of California increases the cost-effectiveness of higher efficiency products compared to the mostly coastal and mild urban counties. However, if this correlation were purely weather-driven, the correlation between SEER and county quantity contractors serving would be less than the correlation between SEER and hard-to-reach county status. The higher correlation between SEER and county quantity contractors serving suggests that when more TECH-enrolled contractors serve a county, this decreases the average SEER of heat pump HVAC retrofits for more than just geographic reasons.

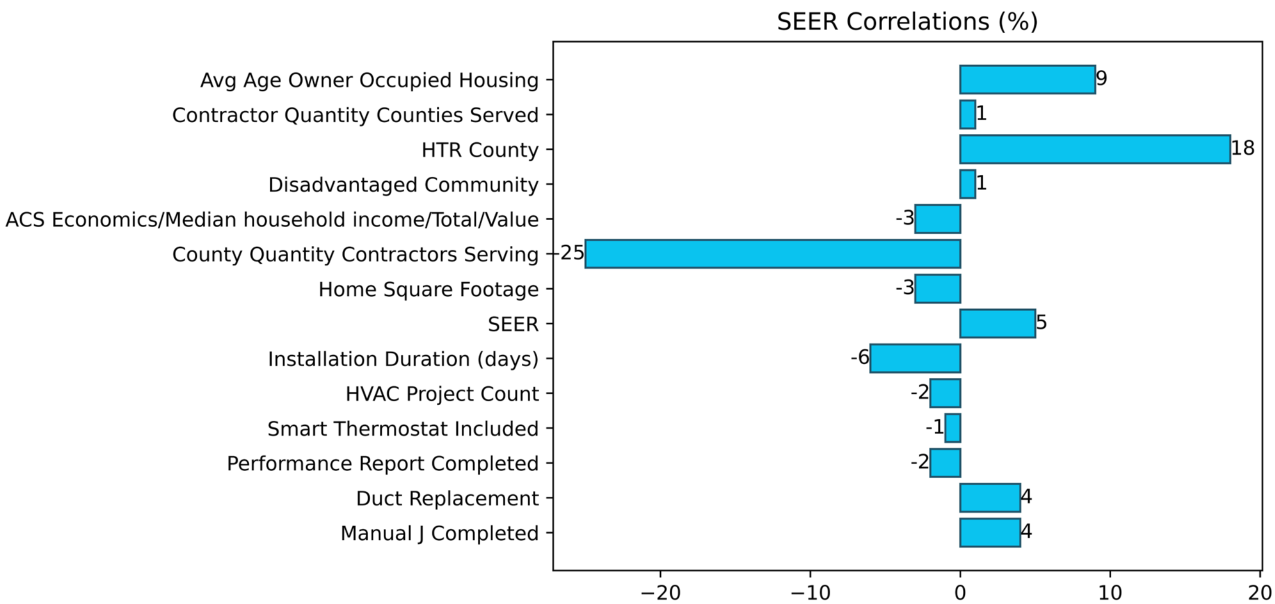


FIGURE 8: CORRELATIONS (%) BETWEEN SEER AND OTHER QUANTITATIVE COVARIATES INCLUDED IN THE COST DRIVER ANALYSIS

Results

This section presents outputs of an ordinary least squares multivariate linear regression model we developed for the Cost Driver Analysis. Discussion of results is divided into three parts. First, we review covariates that were found not to have a statistically significant linear relationship with total project cost. Next, we compare the baseline total project cost assigned to each equipment type, defined in the Equipment Types section. Finally, we present the incremental cost drivers: the average amount by which each of the Project Features and Site Features adds to or subtracts from the total project cost. Complete outputs of the regression model as reported by the Python statsmodels library "OLS" class object created for this analysis are presented in Model Outputs.

Model Fit

The R-squared and adjusted R-squared of the model were 0.245 and 0.243, respectively. This indicates that our team's OLS regression model can only explain approximately 24 percent of the observed variation in total project cost in the training dataset. However, most covariates were successfully assigned a statistically significant regression coefficient. Those that were not discussed in Insignificant Covariates.

The TECH Clean California team observed significant heteroskedasticity in the model results, so we adjusted the model's standard error estimates to make them more heteroskedasticity-robust using the "get_robustcov_results" method with the HC3 estimator.

Insignificant Covariates

Not every covariate included in the cost driver analysis was found to have a statistically significant relationship with total project cost. In a linear regression model, covariates whose regression coefficient has a P-value of five percent or greater are traditionally considered to not have a statistically significant relationship with the independent variable. For the heat pump HVAC cost driver analysis, this does not mean that "insignificant" covariates have no effect on total project cost; it only means that the model could not find a consistent linear relationship between the feature and total project cost. With that said, the following covariates were found to be statistically insignificant:

- Disadvantaged community[8] status of the census tract in which the project was installed.
- Installation of a smart thermostat.
- Decommissioning the furnace in the home during the installation.

Results presented below do not include the disadvantaged community covariate because it is highly correlated with other site features, so we did not include it in the model to avoid obfuscating results.

The level of statistical significance assigned to the covariates by the model is likely overestimated as a result of post-selection inference. However, our team only removed covariates from the model to minimize multicollinearity; we did not remove any covariates simply to avoid including covariates with low significance. We believe that removing covariates contributing to high multicollinearity improved the reliability of the model outputs enough to justify the risk of inflated significance due to post-selection inference.

8 As defined by California EnviroScreen 4.0

Baseline Equipment Costs

After the TECH Clean California team fit the OLS regression model to the training data, we translated model outputs into metrics applicable to a typical heat pump HVAC system installation with three tons of cooling capacity, or a "three-ton project" for short. We calculated average baseline three-ton project cost for each type of heat pump HVAC equipment, shown in blue in Figure 9, as well as an average incremental cost for each project and site feature, shown in Figure 10.

The baseline costs shown in blue in Figure 9 represent the average costs of a three-ton project for each heat pump HVAC equipment type assuming all other covariates in the model are minimized. (For Boolean covariates, minimization means the variable is set to false). The covariates that are minimized are listed in Cost Drivers. Baseline cost varies from $10,090 for ducted packaged unitary equipment to $17,736 for ducted minisplit equipment. The baseline cost estimate for all equipment types was statistically significant, with a standard error ranging from $377 for ducted split unitary equipment to $585 for ductless minisplit equipment.

Figure 9 also shows there is a larger difference between baseline cost and mean cost for certain equipment types than others. For example, the ductless multisplit equipment type has a lower baseline cost but a higher mean cost than the ductless minisplit type, indicating that features other than the equipment type in and of itself make ductless multisplit heat pump HVAC installations more expensive, on average, than ductless minisplit installations of the same size.

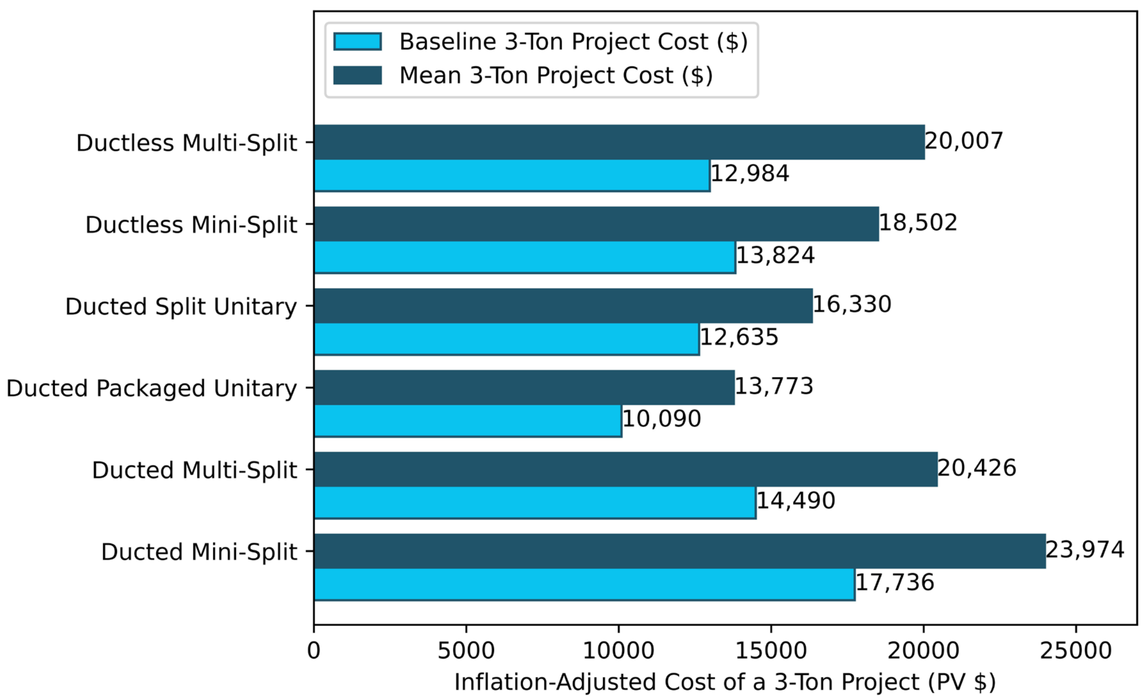


FIGURE 9: BASELINE COST AND MEAN COST BY EQUIPMENT CATEGORY

Cost Drivers

In addition to assigning a baseline cost to each heat pump HVAC equipment type, the TECH Clean California team estimated the average impact of each covariate on the total project cost of a typical three-ton project, shown in Figure 10. The covariates included in the model and their relationship with total project cost are explained in greater detail in Table 7. Both in Figure 10 and in Table 7, covariates are listed in descending order of statistical significance. In total, 12 of the 14 site and project features included in the model had a statistically significant linear relationship with total project cost. The final two features, smart thermostat installed and furnace decommissioned, are not statistically significant.

Given ducted split unitary systems were installed in over 60 percent of TECH Clean California single family heat pump HVAC projects, they represent the lion's share of the training dataset. Therefore, the results presented in Figure 10 and in Table 7 apply most broadly to ducted split unitary heat pump HVAC system installations in single family homes.

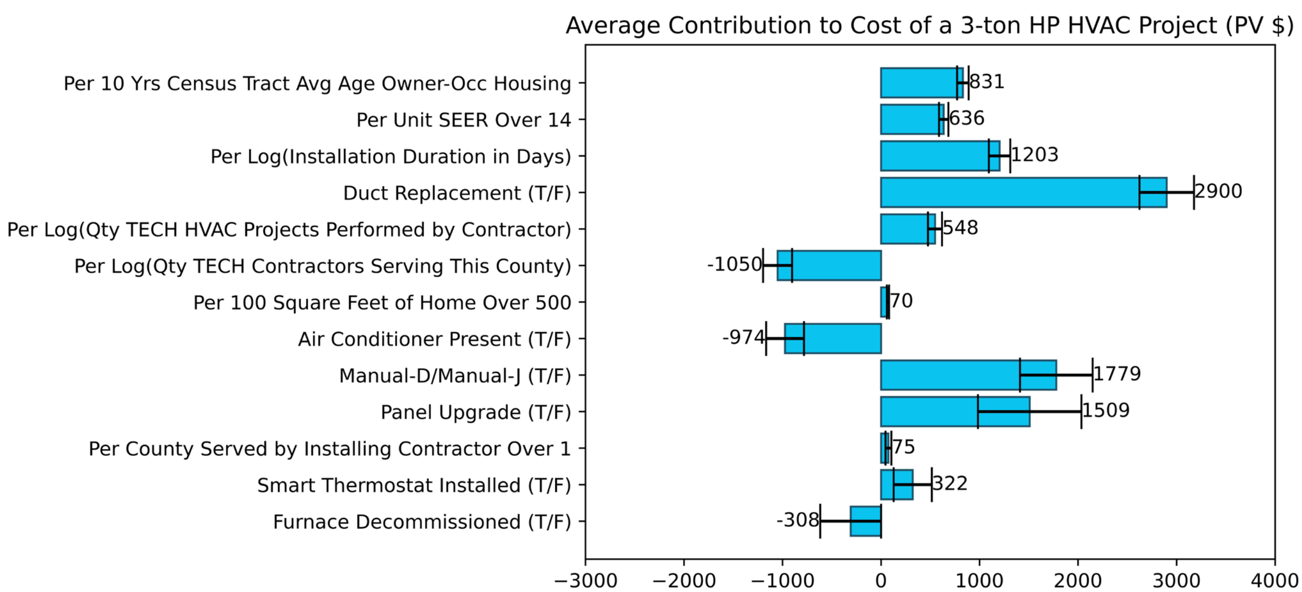


FIGURE 10: AVERAGE COST CONTRIBUTION OF PROJECT FEATURES TO A THREE-TON HEAT PUMP HVAC INSTALLATION (LABEL IS THE AVERAGE VALUE, ERROR BARS SHOW THE AVERAGE VALUE +/- THE STANDARD ERROR)

Table 7: Heat Pump HVAC Cost Driver Model Features in Descending Order of Statistical Significance

Field	Mean and Range	Average Impact on Total Project Cost of a three-ton Heat Pump HVAC installation
Average Age of Owner-Occupied Housing in the Census Tract in which the Project Occurred	Ranges from 11 to 103 years, mean of 50 years.	Adding 10 years to the average age of owner-occupied housing in a Census Tract, starting at 11 years, adds $826 (± $59) to Total Project Cost. Projects in Census Tracts with an average owner-occupied home age of 70 years are $3,221 more expensive on average than projects in Census Tracts with an average home age of 20 years.
Seasonal Energy Efficiency Ratio ("SEER")	Ranges from 14.0 to 29.4, mean of 17.1.	For each unit of SEER above 14 that the installed Heat Pump HVAC equipment is rated, the Total Project Cost increases by $637 (± $48).
Installation Duration (Days)	Ranges from 1 to 366, mean of 5.	Total Project Cost increases logarithmically with the installation duration. A two-day installation costs $363 (± $33) more than a one-day installation, while a 10-day installation costs $1,207 (± $109) more.
Duct Replacement (T/F)	True for 15% of projects	Performing a duct replacement alongside the installation increased the total project cost by $2,926 (± $277), on average.
Number of TECH-enrolled Contractors Serving the County in which the Project Occurred	Ranges from 11 to 279, mean of 144.	Total project cost decreases logarithmically with the total number of TECH-enrolled contractors serving the county. Projects in counties served by 100 TECH-enrolled contractors cost $1,031 (± $147) less than projects in counties served by just 10 contractors.
Number of TECH Clean California Heat Pump HVAC Projects Performed by the Contractor who Performed the Installation	Ranges from 1 to 406, mean of 89	Total project cost increases logarithmically with the total number of heat pump HVAC projects performed by the installing contractor. Projects by contractors with two TECH Clean California-funded installations cost $153 (± $22) more than those with just one. Projects performed by contractors with 50 TECH Clean California-funded installations cost $863 (± $122) more than those with just one.
Home Floor Area	Ranges from 500 to 10,000, mean of 2,033 sq ft	Total project cost increases by $69 (± $11) per 100 square feet of additional floor area in the home, while holding the cooling capacity of the installed system constant.
Air Conditioning Type	None: 48% Present: 52%	Presence of an air conditioner in the home prior to the installation reduces the total project cost by $972 (± $192).
Manual D/J Completed (T/F)	True for 7% of projects	Performing a Manual-D or Manual-J Load Calculation alongside the installation increased the total project cost by $1,847 (± $369).
Electrical Panel Upgrade (T/F)	True for 4% of projects	Performing a panel upgrade alongside the installation increased the total project cost by $1,567 (± $525).
Number of Counties Served by the Contractor who Performed the Installation	Ranges from 1 to 15, mean of 3	For each additional California county served by the installing contractor, total project cost increased by $72 (± $30).
Furnace Status After Installation	Decommissioned: 87% Commissioned to run in Emergencies or Blower Only: 13%	Not statistically significant
Smart Thermostat Installed (T/F)	True for 44% of projects	Not statistically significant

Conclusions

Model Performance

The TECH Clean California team's primary goal for the heat pump HVAC cost driver analysis was to estimate the average impact that common project features have on total project cost in an easily interpretable way. Because of this, we prioritized building a simple model to use for statistical inference rather than a model optimized for prediction. The linear regression model we created has a limited capacity to explain the observed variation in total project cost, with an adjusted R-squared of only 24 percent, making it an unreliable predictor of total project cost. Our team expects that the two primary flaws that limited the model's predictive capacity are: (a) covariates missing from the model that significantly impact total project cost, and (b) nonlinear relationships between covariates and total project cost.

However, the model found a statistically significant regression coefficient for the intercept as well as 12 out of 14 covariates, so it can help us understand how most of the covariates affect total project cost. Understanding these cost drivers is the first step to finding the most cost-effective pathway to electrify California homes.

Model Outputs

The lack of statistical significance for the "furnace status after installation" covariate and "smart thermostat installed" covariate confirmed the team's expectation. Both features typically require relatively inexpensive equipment, if any at all, and minimal additional labor, compared to the cost of the rest of the equipment and labor for the installation.

Furthermore, the sign and magnitude of the regression coefficients assigned to the following covariates aligned with the team's understanding of heat pump HVAC project costs: SEER ($637 in added cost per additional unit starting at 14), duct replacement ($2,926 added cost), Manual-D/Manual-J ($1,847 added cost), and panel upgrade ($1,567 added cost). These features either improve equipment performance or enable home electrification, and for many homes, this additional cost is a worthwhile investment.

Many model outputs, however, surprised or stood out to our team, including the following:

> **Disadvantaged community status is not a cost driver.** The disadvantaged community status of the census tract in which the project occurred did not have a statistically significant relationship with total project cost. This surprised our team given the number of salient demographic features that factor into a census tract's disadvantaged community status. However, some of these factors could counteract the effect that other factors have on total project cost. For example, many disadvantaged communities have a relatively high unemployment rate. This could reduce the availability of heat pump HVAC installers, driving costs up, but also reduce the cost of living, driving costs down. Because disadvantaged community status is highly correlated with other covariates that did significantly impact total project cost, such as the average age of owner-occupied housing in a census tract, we omitted disadvantaged community status from our model. We recommend that other teams estimating the cost of electrifying California homes omit disadvantaged community status in their models for the same reason.

Less expensive equipment types are popular, but the least expensive is not the most popular. Ducted split unitary systems were installed in over 60 percent of TECH Clean California single family heat pump HVAC projects and were one of the least expensive, though not the least expensive product type installed. Ducted packaged unitary systems were, on average, less expensive but were installed in only 10 percent of homes. Meanwhile, ducted minisplit and multisplit systems were the most expensive, on average, but were twice as popular as ductless minisplit and multisplit systems. This shows that the California heat pump HVAC market is not price-insensitive but is also not so price-sensitive that only the cheapest available heat pump HVAC systems are being installed.

Installations in census tracts with old homes cost thousands of dollars more. The average age of owner-occupied housing in the census tract had a strong positive relationship with total project cost, increasing it by $826 per 10 years of average age. While the team expected the age of an individual home to strongly impact total project cost, we were surprised by the magnitude of the impact and the fact that it was reflected in the average age of owner-occupied housing in the census tract. The model outputs suggest that installing a heat pump HVAC in a home in a census tract with an average home age of 60 years is likely to cost over $4,000 more than the equivalent heat pump HVAC installation in a census tract with an average home age of 10 years.

Heat pump HVAC retrofits in homes with air conditioners cost less. The presence of an air conditioner in the home prior to the heat pump HVAC retrofit was associated with a $972 reduction total project cost, on average. One caveat is that approximately ninety percent of homes with an air conditioner prior to the heat pump HVAC retrofit in the TECH Clean California database had a central air conditioner, so this is not a good representation of the impact of a room unit or other type of air conditioner on total project cost. We suspect the observed reduction in total project cost is partially because most ducted system installations were significantly cheaper than ductless systems, and customers without pre-existing central air conditioning likely opted more often for ductless systems.

Counties served by more TECH-enrolled contractors had lower costs. Projects in counties served by 100 TECH-enrolled contractors cost $1,031 less on average than projects in counties served by just 10 contractors. Through analysis of correlations among covariates considered for the model, the TECH Clean California team observed that urban counties—the San Francisco Bay Area, Greater Sacramento Area, Greater Los Angeles Area, and San Diego county—are currently served by a significantly higher quantity of TECH-enrolled contractors serving single-family homes than the state's rural counties. This reinforces the importance of TECH Clean California's commitment to serving hard-to-reach communities with incentive programs other than our statewide voluntary single family incentive program, whose data was used to train the model.

Next Steps

Integrate More Contractor and Branding Features

Two covariates expected to significantly impact total project cost, but not included in the heat pump HVAC cost driver analysis, were the installing contractor's company name and the brand of the installed equipment. We conducted an exploratory assessment of how contractor name and brand name would improve the predictive capacity of the model, the results of which are presented in Table 8. The results show that using either or both contractor name and brand name as covariates would improve the model's ability to explain variation in total project cost, though not drastically – only increasing R-squared by about six percent altogether – and do not add so much complexity to the model that they increase the akaike or bayesian information criteria ("AIC" and "BIC").

While these results are encouraging, they are not inherently useful for building an intuitive understanding of the heat pump HVAC market because the regression coefficients the model generates are specific to individual contractors and brands. However, they show that features of contractors and brands not already included in the cost driver analysis could improve the model's predictive capabilities.

Table 8: Model Results Including and Excluding Brand Index and Contractor Index

Model version	Number of covariates (including categorical options)	R2	Adjusted R2	AIC1 (x105)	BIC2 (x105)
Including only covariates listed in Data Overview	19	0.245	0.243	1.330	1.332
Including a categorical covariate for each of the 10 most common brands (used in >2% of projects)	30	0.266	0.263	1.328	1.330
Including a categorical covariate for each of the 19 contractors who performed >1% of projects	39	0.291	0.287	1.326	1.329
Including all of the above	50	0.305	0.300	1.325	1.328

1 Akaike Information Criterion

2 Bayesian Information Criterion

Build Inference Models Specific to Each Equipment Type

We expect that the amount that project features and site features impact total project cost varies significantly across different equipment types, especially differing between mini and multisplit equipment versus unitary equipment. Furthermore, some project features and site features are only applicable to certain equipment types. For example, we would not expect duct replacements to apply to ductless system installations, and the number of compressor speed options is only applicable to unitary equipment.

We chose to not make the cost driver analysis equipment-type-specific and instead included all equipment types in a single model because the heat pump HVAC project dataset was not large enough to divide by equipment type and still have a sufficient sample size for a robust statistical analysis of each partial dataset. However, more TECH Clean California incentives are paid every day for heat pump HVAC installations. Given this, we intend for future HP HVAC cost driver analyses to be equipment-type-specific.

Improve Total Project Cost Prediction

We expect that it is possible to develop a model optimized for prediction, rather than inference, that can explain a significantly larger percent of the variation in total project cost than the model discussed herein. To build a model optimized for predictive accuracy, our team recommends the following steps:

1. Address nonlinear relationships between covariates and total project cost by building a nonlinear model, such as a random forest regression model.
2. Address interactions among covariates by including interaction variables. In particular, variables representing the interaction of SEER and equipment type would be useful. Reduce model complexity and risk of overfitting by using ridge regression or stepwise selection to identify the best covariates.
3. Estimate the model's ability to make predictions on new data using cross-validation.

Compare Cost Drivers to Savings Drivers

Though there is value in understanding cost drivers and predicting total project cost in and of itself, the cost paid at the time of installation is only the beginning of a customer's experience with a heat pump HVAC system. Furthermore, many project and site features that increase project costs also improve performance, reduce operating costs, increase GHG emissions savings, or all the above. Thus, our team's ultimate goal is to create models similar to the heat pump HVAC cost driver analysis that use key savings and performance metrics as the independent variable instead of total project cost. With these, we can compare cost drivers to savings drivers to determine which features of heat pump HVAC projects are cost-effective, the extent and bounds of their cost-effectiveness, and create recommendations for which project features should be prioritized and deprioritized by customers, contractors, policymakers, and other decarbonization investors.

Appendix A — Data Fields Considered and Used

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
Project Index	Integer	A unique integer for each unique project	1	Calculated	No; too many categories
Contractor Index	Integer	A unique integer for each unique contractor	20	CSLB #	No; too many categories
Project Site Index	Integer	A unique integer for each unique customer address	500	Customer address	No; too many categories
Customer County	Text string	County in which the home where this project occurred resides	San Mateo County	Geocod.io	No; too many categories
Gas IOU Territory #1	Text (categorical)	Denotes in which investor-owned utility's natural gas service territory the project occurred	Pacific Gas and Electric	Incentive Clearinghouse	No; highly correlated with fields used
Gas IOU Territory #2	Text (categori-cal)	If the ZIP code in which the project occurred is shared by two IOU gas service territories, this field denotes the second IOU that serves this ZIP code	Southern California Gas	Incentive Clearinghouse	No; highly correlated with fields used
CA Climate Zone	Integer	Denotes in which of the 16 California Climate Zones, as defined by the CEC, the project occurred	16	Incentive Clearinghouse	No; too many categories
Disadvantaged Community	Boolean	Denotes whether a project occurred in a Census Tract labeled as a Disadvantaged Community per CalEnviroScreen 4.0	TRUE	Incentive Clearinghouse	Yes, but later removed due to low statistical significance
TECH Equity Community	Boolean	Denotes whether a project occurred in a home meeting the TECH Clean California team's definition of an Equity Community, defined in collaboration with CPUC in Q2 2023	TRUE	Incentive Clearinghouse	No, but similar field used
Program Name	Text	Name of the incentive program through which this incentive was provided	TECH Clean California - HPWH	Incentive Clearinghouse	No; only one applicable category
Program Income Qualification	Text (categorical)	Denotes if a program's focus is market-rate housing or low-income housing	Market Rate	Incentive Clearinghouse	No; only one applicable category
Product Group	Text (categorical)	Denotes if a program's focus is HP HVAC or HPWH technology	HPWH	Incentive Clearinghouse	Yes; used to filter initial dataset
Building Type	Text (categorical)	Type of residential building in which the project occurred: Single Family or Multifamily	Single-Family	Contractor	Yes; used to filter initial dataset
Home Square Footage	Float	Area of floorspace, in square feet, in the building in which the project occurred	1243	Contractor	Yes
Panel Capacity Pre-Install (Amps)	Float	Capacity of the electrical panel that will serve the installed heat pump, in Amperes, prior to the installation	125	Contractor	No; highly correlated with fields used

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
Panel Capacity Post Install (Amps)	Float	Capacity of the electrical panel that will serve the installed heat pump, in Amperes, following the installation	200	Contractor	No; highly correlated with fields used
Installed Equipment Description	Text	Written description of the installed equipment provided by the applicable QPL	Split System: Heat Pump with Remote Outdoor Unit-Air-Source	AHRI QPL	No, but a related field was used
Material Category	Text (categorical)	Identifier of the qualified product list used for the installed equipment	vsmshp	AHRI QPL	No, but a related field was used
Product Type	Text (categorical)	Category of heat pump product installed, as defined by AHRI	Mini/Multisplit	AHRI QPL	No, but a related field was used
Brand Index	Integer	A unique integer for each unique brand	5	AHRI QPL	No; too many categories
Model Production Status	Text (categorical)	Indicates whether the model of heat pump installed is still being manufactured	Discontinued	AHRI QPL	No; not a significant contributor to cost
AHRI Type	Text (categorical)	Model type, as defined by AHRI	HRCU-A-CB	AHRI QPL	No, but a related field was used
AHRI Descriptor	Alphanumeric	Additional description of the installed equipment provided by AHRI	10298040	AHRI QPL	No, but a related field was used
HVAC equipment type	Text (categorical)	Further detail on installed HVAC equipment, based on AHRI type. Options: packaged unitary equipment, split unitary equipment, minisplit, and multisplit.	Minisplit	AHRI QPL	Yes
Ducted	Boolean	Indicates if the installed HVAC equipment includes ducts	TRUE	AHRI QPL	Yes
Ducting Type	Text (categorical)	Detailed description of ducting for the installed HVAC equipment, provided by AHRI	Split System: Heat Pump with Remote Outdoor Unit-Air-Source	AHRI QPL	No; too many categories
Compressor Type	Text (categorical)	Description of the type of compressor used in the installed equipment	Variable Speed	AHRI QPL	No; highly correlated with used fields
Cooling Source	Text string	The type of heat source/sink used by the installed heat pump HVAC equipment	Air cooled	AHRI QPL	No; only one category
Cooling Capacity (tons)	Float	Rated capacity, not nominal. The amount of heat that the installed heat pump HVAC system can remove from the conditioned space in one hour	3.38	AHRI QPL	Yes; used to normalize dependent variable
Heating Capacity (tons)	Float	Rated capacity, not nominal. The amount of heat that the installed heat pump HVAC system can add to the conditioned space in one hour	1.83	AHRI QPL	No, but related field used

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
EER	Float	An indicator of cooling efficiency of the installed heat pump HVAC system. EER is the ratio of output cooling energy to input electrical energy at a given operating point. EER is normally calculated with a 95 °F outside temperature and an inside (return air) temperature of 80 °F and 50% relative humidity.	9.2	AHRI QPL	No, but related field used
SEER	Float	An indicator of cooling efficiency of the installed heat pump HVAC system. SEER is the ratio of output cooling energy to input electrical energy over an entire cooling season.	17	AHRI QPL	Yes
HSPF (Region IV)	Float	An indicator of heating efficiency of the installed heat pump HVAC system. HSPF is the ratio of output heating energy to input electrical energy over an entire heating season. HSPF is region-specific, and the region for the HSPF value provided is Region IV of the United States.	12.2	AHRI QPL	No, but related field used
EER2	Float	AHRI2 METRIC: An indicator of cooling efficiency of the installed heat pump HVAC system. EER is the ratio of output cooling energy to input electrical energy at a given operating point. EER is normally calculated with a 95 °F outside temperature and an inside (return air) temperature of 80 °F and 50% relative humidity.	9.2	AHRI QPL	No; too sparsely populated
SEER2	Float	AHRI2 METRIC: An indicator of cooling efficiency of the installed heat pump HVAC system. SEER is the ratio of output cooling energy to input electrical energy over an entire cooling season.	17	AHRI QPL	No; too sparsely populated
HSPF2 (Region IV)	Float	AHRI2 METRIC: An indicator of heating efficiency of the installed heat pump HVAC system. HSPF is the ratio of output heating energy to input electrical energy over an entire heating season. HSPF is region-specific, and the region for the HSPF value provided is Region IV of the United States.	12.2	AHRI QPL	No; too sparsely populated
Ex Ante annual electricity savings (kWh)	Float	Estimated annual electricity savings resulting from this installation based on CPUC Work Papers	-1000	Incentive Clearinghouse	No; not actionable interpretation
Ex Ante annual gas/ propane savings (Therms)	Float	Estimated annual gas savings resulting from this installation based on CPUC Work Papers	100	Incentive Clearinghouse	No; not actionable interpretation
Ex Ante annual GHG savings (Metric tons CO2e)	Float	Estimated annual greenhouse gas savings resulting from this installation based on CPUC Work Papers	10000	Incentive Clearinghouse	No; not actionable interpretation
Count Units Installed	Integer	The number of discrete heat pump units installed during this project	2	Contractor	Yes; used to normalize dependent variable
Total Project Cost ($)	Float	The total cost of equipment and installation for this project	10000	Contractor	No, but related field used
Total Project Cost per Unit ($)	Float	The total cost of equipment and installation for this project, divided by the number of units installed	5000	Incentive Clearinghouse	No, but related field used
HVAC Norm Cost ($/ton)	Float	Normalized cost of installation for HP HVAC projects, allowing easy comparison of projects	5,870.88	Incentive Clearinghouse	Yes; dependent variable

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
Incentive 1 ($)	Float	An incentive paid to the Contractor that performed this project	1100	Incentive Clearinghouse	No; incentive amount correlated with project features
Incentive 1 Funder	Text	The program that provided the incentive labeled “Incentive 1”	TECH Clean California	Incentive Clearinghouse	No; incentive amount correlated with project features
Incentive 2 ($)	Float	A second incentive paid to the Contractor that performed this project, layered with Incentive 1	2000	Incentive Clearinghouse	No; incentive amount correlated with project features
Incentive 2 Funder(s)	Text	The program(s) that provided the incentive labeled “Incentive 2”	BayREN Home +	Incentive Clearinghouse	No; incentive amount correlated with project features
Total Incentive Received by Contractor ($)	Float	The sum of all known incentives received by the contractor/ customer as a result of this project	3100	Incentive Clearinghouse	No; incentive amount correlated with project features
Installation Start Date	Date	The date on which the installation of all heat pump equipment and associated work began	4/30/2022	Contractor	No
Installation End Date	Date	The date on which the installation of all heat pump equipment and associated work ended	5/15/2022	Contractor	Yes; used to normalize dependent variable (inflation adjustment)
Installation Duration (days)	Integer	The duration of the installation of all heat pump equipment and associated work	2	Contractor	Yes
Permit Request	Boolean	Indicates whether a permit was requested for the project	TRUE	Contractor	No; almost all claims are TRUE
Panel Upgrade	Boolean	Indicates whether the contractor upgraded the amperage capacity of the home’s electrical panel as part of the installation	TRUE	Contractor	Yes
Type of Panel Upgrade	Text (categorical)	Describes any work performed to upgrade the home’s electrical panel alongside the heat pump installation; multiple options can be selected	Installation of smart load center	Contractor	No; too sparsely populated (added to incentive app form in April 2023)
Emergency Replacement	Boolean	Indicates whether the heat pump equipment was installed as the result of an emergency replacement	TRUE	Contractor	No; too sparsely populated (added to incentive app form in April 2023)
Emergency Backup Type	Text (categorical)	Lists any emergency backup power sources the home has in the event of a power outage	Battery	Contractor	No; too sparsely populated (added to incentive app form in April 2023)
Other Building Upgrades During Install	Text string	If applicable, a list of building upgrades performed in addition to the heat pump HVAC/HPWH installation that contributed to the total project cost	Closet expansion	Contractor	No; data not clean enough to use but work underway
Duct Replacement	Boolean	Indicates whether the contractor performed a duct replacement as part of this heat pump HVAC installation.	FALSE	Contractor	Yes

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
Duct Sealing	Boolean	Indicates whether the contractor performed duct sealing as part of this heat pump HVAC installation.	TRUE	Contractor	No; highly correlated with used fields
Smart Thermostat Included	Boolean	Indicates whether the contractor also installed a smart thermostat alongside the heat pump HVAC system.	FALSE	Contractor	Yes
Manual J Completed	Boolean	Indicates whether the contractor performed an ACCA Manual-J/ Manual-D load report as part of this installation.	TRUE	Contractor	Yes
Performance Report Completed	Boolean	Indicates whether the contractor conducted an ASHRAE 221-2020 Heating System Performance Ration (HSPr) or Cooling System Performance Ration (CSPr) report as part of this installation.	TRUE	Contractor	No; a very small quantity of contractors used this quality installation measure
Previous AC Type	Text	Describes the kind of air conditioning system in the customer's home prior to the heat pump HVAC installation	Central	Contractor	Yes
Previous AC Model Number	Alphanumeric	If available, the model number air conditioning system in the customer's home prior to the heat pump HVAC installation	AG94873483	Contractor	No; too many categories
Previous Furnace Fuel Type	Text (categorical)	The kind of fuel used by the furnace system in the customer's home prior to the heat pump HVAC installation	Gas	Contractor	Yes
Previous Furnace Model Number	Alphanumeric	If available, the model number of the furnace system in the customer's home prior to the heat pump HVAC installation	74A5000	Contractor	No; too many categories
Furnace Status After Install	Text (categorical)	Describes the way the furnace is set up to be used (if at all) after the heat pump HVAC system was installed	Fully Decommissioned	Contractor	Yes
Contractor Project Count	Integer	Number of projects awarded a TECH incentive performed by the contractor	84	Contractor	Yes
Contractor Quantity Counties Serving	Integer	Number of California counties served by the contractor who performed the installation	7	Contractor Enrollment Data	Yes
Contractor Duct Replacement Count	Integer	Number of projects awarded a TECH incentive featuring a duct replacement performed by the contractor	26	Incentive Clearinghouse	No; highly correlated with used fields
Contractor Duct Sealing Count	Integer	Number of projects awarded a TECH incentive featuring duct sealing performed by the contractor	18	Incentive Clearinghouse	No; highly correlated with used fields
Contractor Smart Thermostat Count	Integer	Number of projects awarded a TECH incentive featuring a smart thermostat performed by the contractor	10	Incentive Clearinghouse	No; highly correlated with used fields
Contractor Manual J Count	Integer	Number of projects awarded a TECH incentive featuring Manual-J/ Manual-D calculations performed by the contractor	20	Incentive Clearinghouse	No; highly correlated with used fields

Data Field name	Data type	Description	Example	Source	Used in Cost Driver Analysis?
Contractor Performance Report Count	Integer	Number of projects awarded a TECH incentive featuring an ASHRAE 221-2020 Heating System Performance Ration (HSPr) or Cooling System Performance Ration (CSPr) report performed by the contractor	0	Incentive Clearinghouse	No; highly correlated with used fields
ACS Economics/ Number of households/ Total/Value	Integer	Number of households in the Census Tract in which the installation occurred	511	Geocod.io	No; highly correlated with used fields
ACS Economics/Median household income/ Total/Value	Integer	Median household income in the Census Tract in which the installation occurred	105139	Geocod.io	No; highly correlated with used fields
HTR County	Boolean	Denotes if the building in which the equipment was installed was in a Hard-to-Reach County, as defined by CPUC	TRUE	Incentive Clearinghouse	No; highly correlated with used fields
County Quantity Contractors Serving	Integer	Number of TECH enrolled contractors serving the county in which the project occurred	100	Contractor Enrollment Data	Yes
Census Tract Average Age Owner Occupied Housing	Float	Average age of owner-occupied housing in the Census Tract in which the project occurred	55.5	US Census data	Yes

Appendix B — Model Outputs

OLS Regression Results

Dep. Variable:	Project_cost_per_ton_cooling	R-squared:	0.244
Model:	OLS	Adj. R-squared:	0.242
Method:	Least Squares	F-statistic:	119.1
Date:	Wed, 20 Sep 2023	Prob (F-statistic):	0.00
Time:	12:15:20	Log-Likelihood:	-66515.
No. Observations:	7222	AIC:	1.331e+05
Df Residuals:	7203	BIC:	1.332e+05
Df Model:	18		
Covariance Type:	HC3		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	5911.8994	302.398	19.550	0.000	5319.111	6504.688
Equipment_type[T.Ducted Multi-Split]	-1082.0314	154.365	-7.010	0.000	-1384.633	-779.430
Equipment_type[T.Ducted Packaged Unitary]	-2548.5463	136.378	-18.687	0.000	-2815.887	-2281.205
Equipment_type[T.Ducted Split Unitary]	-1700.1535	125.868	-13.507	0.000	-1946.891	-1453.416
Equipment_type[T.Ductless Mini-Split]	-1304.0427	195.206	-6.680	0.000	-1686.704	-921.381
Equipment_type[T.Ductless Multi-Split]	-1583.8437	190.059	-8.333	0.000	-1956.414	-1211.273
Previous_AC[T.Present]	-324.8023	64.094	-5.068	0.000	-450.445	-199.159
Furnace_status[T.Fully decommissioned]	-102.5963	103.163	-0.995	0.320	-304.825	99.633
SEER_minus_min	212.0241	16.101	13.168	0.000	180.461	243.587
Log_install_duration	400.8616	36.313	11.039	0.000	329.678	472.045
Panel_upgrade	502.9693	174.844	2.877	0.004	160.223	845.715
Duct_replacement	966.5043	92.214	10.481	0.000	785.737	1147.272
Smart_tstat	107.2643	64.178	1.671	0.095	-18.543	233.071
Manual_J	593.0859	123.000	4.822	0.000	351.970	834.202
Log_HVAC_Proj_Count	182.7781	23.689	7.716	0.000	136.340	229.216
Home_100_sqft_minus_min	23.2043	3.682	6.301	0.000	15.986	30.423
Log_county_qty_contractors_serving	-349.8607	48.963	-7.145	0.000	-445.842	-253.879
Contractor_qty_counties_served	24.9476	9.887	2.523	0.012	5.566	44.329
Avg_age_owner_occ_housing	276.9933	19.551	14.168	0.000	238.668	315.319

Omnibus:	1421.819	Durbin-Watson:	1.902
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2883.992
Skew:	1.173	Prob(JB):	0.00
Kurtosis:	5.021	Cond. No.	199.

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

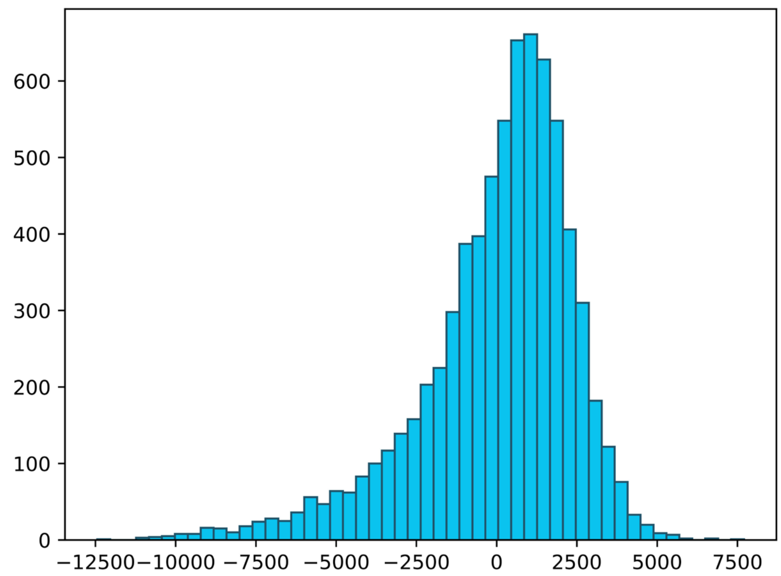
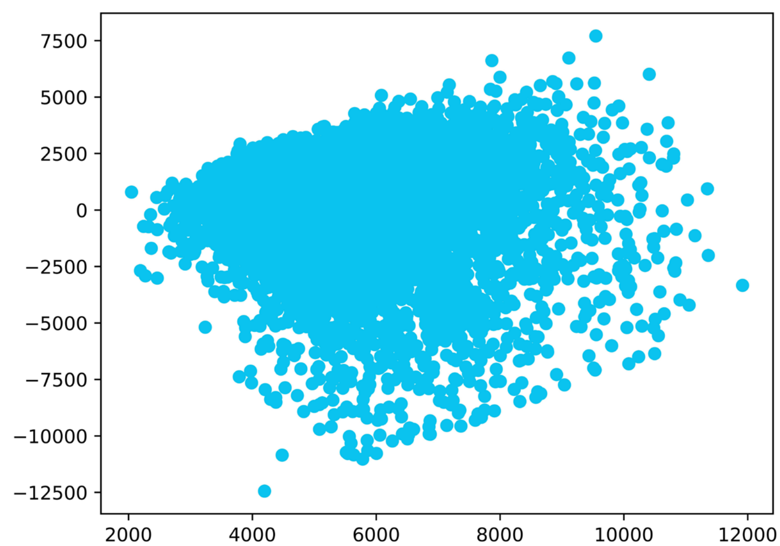


FIGURE 11: HISTOGRAM OF PREDICTIONS MINUS OBSERVATIONS OF TOTAL PROJECT COST, USED TO



MEASURE NORMALITY OF RESIDUALS

FIGURE 12: SCATTERPLOT OF PREDICTIONS VERSUS PREDICTIONS MINUS OBSERVATIONS, USED TO MEASURE HETEROSKEDASTICITY

TECH

CLEAN CALIFORNIA

California's award-winning heat pump program, TECH Clean California, has allocated $80.2 million in funds for heat pump water heater installations, designed to help accelerate the market for heat pump technology across the state through incentives, workforce training, and consumer education to create a pathway for achieving California's targets of six million heat pumps by 2030 and carbon-free, climate-ready homes by 2045.

TECH Clean California is funded by California ratepayers and taxpayers and administered by Southern California Edison Company under the auspices of the California Public Utilities Commission.

The TECH Clean California team is led by Energy Solutions and partners with Ardenna Energy, Association of Energy Affordability, Building Decarbonization Coalition, Electrify My Home, Frontier Energy, National Comfort Institute, Energy Outlet, Recurve Analytics, The Ortiz Group, Tre' Laine Associates, and VEIC.