Electric Vehicle Penetration Study Using Linear Discriminant Analysis

Pacific Gas & Electric Company
San Francisco, CA
Contributors: Jeff Swanson, Rick Aslin, Zeynep Yücel

Presented at June 2011 Center for Research in Regulated Industries (CRRI) Conference

Introduction

The mainstream availability of battery electric and plug-in hybrid electric vehicles (EVs), like the Nissan Leaf™ and Chevy Volt™, presents a potentially interesting challenge for an electric utility. An unplanned arrival of several EVs to a neighborhood could overwhelm the distribution system. This would be especially critical for areas with moderate climates where the distribution system does not have excess capacity for air conditioning.

In this study, we investigated the use of advanced statistical classification techniques, namely linear discriminant analysis (LDA), to identify the discriminating characteristics for the potential future EV customers. We then used the geographic information systems (GIS) technology to map areas where EV adoption is most likely to occur. The results from the statistical classification analysis along with the GIS maps would help assess the electric distribution system capacity needs to accommodate potential EV adoption in the service territory.

We have merged PG&E's customer information database with demographic information from the U.S. Census Bureau. Using this demographic data for neighborhoods where customers have or have had an electric vehicle, we developed a methodology to predict neighborhoods in which EVs are most likely to have the highest penetration. These areas of likeliest adoption were chosen using a linear discriminant analysis (LDA) model to find areas with similar demographic characteristics to those which already have an EV. One great benefit to this analytic method is that it is simple to update as more customers purchase EVs and the customer demographics shift through time.

An Overview of Linear Discriminant Analysis

Linear discriminant analysis (LDA) is one of the classification and grouping techniques in the area of multivariate statistics. It is used to identify distinct groups for a given data set, and to allocate new cases to previously defined groups. LDA produces a regression type linear equation called a discriminant function. This function is used to separate observations into the distinct groups in the population. Typically, the technique is applied to a test group of individuals for which descriptive values and population group subscription are known, such that the calculated discriminant function can then be used to classify individuals with unknown group subscription.

Below are a couple of examples of where classification and grouping type analyses are used[1]:

- ***Populations:*** Good and Poor Credit Risk.
  Discriminating Factors: Income, age, number of credit cards, family size.
- ***Populations:*** Purchasers of new products vs. laggards.
  Discriminating Factors: Education, income, family size, amount of previous brand switching.
- ***Populations:*** Successful vs. unsuccessful college students.
  Discriminating Factors: Entrance examination scores, high school grade-point average, number of high school activities.

The following illustrative example gives an intuitive demonstration of the method. Figure 1a shows two groups of a population graphed according to their values of two descriptive parameters X and Y. It is clear that values of X or Y on their own would not be adequate to classify individuals into the appropriate population groups.

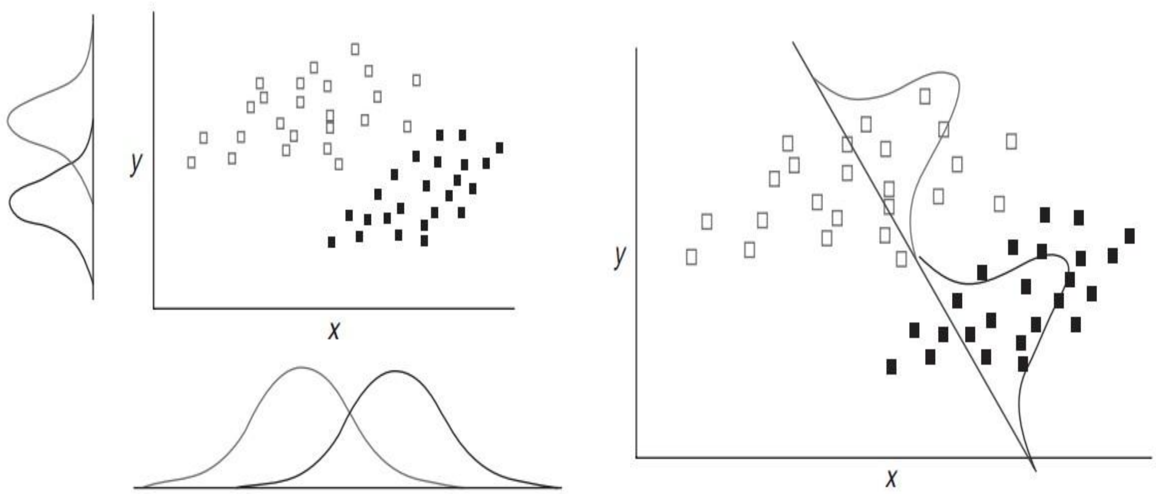

Figure 1a, 1b: Example distributions of population groups and parameters.[2]

There is a relationship between the characteristic parameters that can be leveraged to create a function that can accurately discriminate between groups (Figure 1b). In this case it is simply a linear combination of X and Y.

In this paper, the LDA technique is used to identify the discriminating factors for the EV customers, and identify customers who have similar attributes but do not own EV.

A Case Study – Understanding Plug-in Electric Vehicle Adoption in PG&E's Service Territory

[1] Johnson, Richard. Applied multivariate statistical analysis. Upper Saddle River, N.J: Pearson Prentice Hall, 2007.
[2] Burns, Robert. Business research methods and statistics using SPSS. London etc: SAGE, 2008.

The objective of this study is to demonstrate the use of LDA and GIS to analyze the customer information in assessing where likely to observe the future adoption of EVs in PG&E's service territory.

PG&E has a substantial customer information database which contains electric and gas monthly billing histories for each customer, rate schedule information, and service address latitude and longitude. Using the service address latitude and longitude, each customer can be placed into the appropriate census geography so that their billing history can be aligned with the area specific demographic information provided in the American Community Survey[3] database. The American Community Survey is conducted continually by the U.S. Census Bureau and database updates are released on an annual basis. The American Community Survey is not a census and therefore the demographic information pertaining to smaller geographic units is based on estimates derived over the past five years. The survey furnishes data about the age, sex, marital status, and race of an area's population, but more importantly it gives us some insight into their daily lives. The survey produces household estimates of income, education level, proportion that rent versus own, household and family size, the method and length of their daily commute, the average number of vehicles they own and many more. The census block group is the smallest geographic unit for which detailed data are available. PG&E's service territory contains 7,733 census block groups each containing an average of 720 households. PG&E has merged its customer information data with the American Community Survey data to create a database of average electric usage patterns and demographic information for each census block group in the service territory

PG&E has offered an EV specific electric rate schedule since 1995. This rate offers customers a real incentive to charge their EV during hours when demand for electricity is the lowest. Beyond helping out the customer, this rate schedule leads us directly to the group of the most proactive and earliest adopters of electric vehicles within PG&E's service territory. That information along with data from the 2008 American Community Survey from the U.S. Census Bureau allows us to generate a demographic snapshot of those pioneering EV owners.

Using the rate schedule information in PG&E's customer information database it is possible to find the census block groups that contain at least one of those pioneering EV owners. There are 269 census block groups that have or had a customer on the EV rate. That leaves 7,464 block groups with no present or past EV activity (of which PG&E is aware, electing to be served under the EV rate is completely voluntary). Summarizing the demographic variables for blocks groups with EV owners and for block groups without an EV owner allows us to see a demographic sketch of each group.

[3] http://www.census.gov/acs/www/

		Median Value
Avg Travel Time to Work	No EV	29 minutes
	EV	31 minutes
Avg Household Size	No EV	2.75 persons
	EV	2.63 persons
Avg Number of Vehicles per HH	No EV	1.8 vehicles
	EV	2 vehicles
Median Household Income	No EV	$ 59,701
	EV	$ 94,262
Percent Owner Occupied Housing	No EV	60%
	EV	74%
Percent Drive to Work	No EV	32%
	EV	38%

Table 1: Block Group Demographic Summary

Based on the information within Table 1, there are several characteristics that stand out for block groups containing EV customers. They tend to have slightly smaller household sizes, a higher average number of vehicles, much higher median household incomes, own their own homes and a larger proportion of them drive to work. Each of those differences in characteristics aligns with what would be expected of the prototypical, early-adopting EV owner.

The knowledge that census block groups with EV owners have a different demographic structure than census block groups without any EV owners makes it possible to use a multivariate statistical technique like linear discriminant analysis (LDA) to determine which census block groups are most likely to host new EVs as the vehicles become more commercially available.

PG&E conducted an LDA on a group of demographic variables using SAS 9.3. Starting with 25 different demographic variables the list was culled using a univariate t-test to test for differences in means of each demographic variable between the classes ($p \leq 0.05$). LDA produces a linear discriminant function with the following variables as the best indicators of class membership.

1) Average Family Size
2) Proportion of Households Occupied by their Owner
3) Average Number of Vehicles per Household
4) Median House Value
5) Per Capita Income
6) Median Household Income
7) Average Household Income
8) Proportion of the Population speaking English at Home
9) Proportion of the Population that has a 4-Year College Degree
10) Proportion of the Population that Drives to Work

The linear discriminant function with these variables best categorizes all census block groups based on their demographics.

Table 2 shows the relative success rate of the linear discriminant function in its ability to pick out the differences between the groups. We can see from the table that 71% of current EV census block groups and 73% of non-EV census block groups were correctly categorized by the function. However, since we are most interested in predicting the areas without a current EV that could see a high level of adoption when the vehicles become more widely available, we look to the 27% of non-EV block groups that were misclassified as EV block groups. Those block groups are the most demographically similar to the prototype EV block groups.

		Classified Into:	
		NO EV	EV
Actual:	NO EV	73%	27%
	EV	29%	71%

Table 2: Discriminant Analysis Resubstitution Summary

The linear discriminant function also calculates a classification coefficient for each census block group. The classification coefficient is a number between zero and one, and is used to indicate the level of probable EV adoption within the block group. If the classification coefficient is greater than 0.5 for a census block group, then that census block group is classified into the EV group. The classification coefficients for each census block were used in PG&E's GIS system to develop maps which illustrate the level of probable EV adoption for all areas of the service territory.

Figures 2, 3 and 4 show the probable level of Plug-In EV adoption for PG&E's Service Territory, the San Francisco Bay Area, and the city of Berkeley, respectively, displaying the classification coefficient for each census block group. Census block groups with a low coefficient are less probable areas of high density EV adoption. These areas are shown as the palest shade of blue. As the blue becomes increasingly saturated, the probability of higher levels of EV adoption increases. The areas with the highest level of probable adoption are highlighted in red. The areas that are white are not a part of PG&E's electricity distribution area. The areas with the highest levels of probable adoption tend to be located in the suburbs of cities and make a prominent ring around the San Francisco Bay, but there are areas elsewhere as well. Areas around Monterey and some suburbs of Sacramento also are prominently displayed in red. There are areas with slightly lower levels of probable adoption surrounding San Luis Obispo in Central California, in the Napa Valley and in several of the larger cities in the Central Valley like Bakersfield and Fresno.

As it can be seen from Figure 2, the PG&E Service Area includes most of northern and central California but the areas where EV adoption is most likely are concentrated in a relatively small geographic area near the San Francisco Bay. This is important information for system planners. However, at this stage of EV adoption, system level planning, which is primarily done to determine needed generation and bulk transmission infrastructure, is not expected to be immediately impacted by EV loads. While seeing the entire service territory is useful for system planning the primary concern of planners with respect to EVs is at regional, local and neighborhood levels. Figure 3 shows how EV adoption might be spread out in the San Francisco Bay Area. As it can be seen from the figures, even in regions which have high probabilities of adoption in general, there are significant differences in probability of adoption between local areas. This information is useful to local transmission planners who are responsible for projecting local area loads and insuring that sufficient infrastructure exists to meet those loads.

Figure 4 shows a map of the same data shown in Figure 3 with higher granularity. This map shows how adoption of EVs might be different among the many neighborhoods in the city of Berkeley located in the San Francisco Bay region. A major part of the city is the University of California at Berkeley. The university and all of the areas of student housing immediately surrounding it are shown in the center in the palest blue. That area is immediately adjacent to some of the areas with extremely high levels of probable adoption, areas with large proportions of single family housing and higher levels of income located in the hillier areas of Berkeley. EV adoption rates in specific neighborhoods are a significant concern for local planners because each EV requires a significant amount of capacity during charging. At the neighborhood level several electric vehicles connected to the same circuit could cause reliability problems. This is especially true of neighborhoods where the electric distribution infrastructure is not sized to support air conditioning load. The data shown on Figure 4 is exactly the data that distribution planners require to understand whether a neighborhood may need a capacity upgrade to support likely penetration of EVs in the near future to maintain the level of reliability that customers demand and to support EV adoption.

Ultimately, the data from this type of analysis can be used to cascade down projections of EV penetrations and associated loads that are done at a statewide or system level down to the circuit level where local planning is done and then aggregated up to feeder banks, busbars, substations and ultimately to the system level to cover all other planning functions.

Conclusion

In this paper, we demonstrated the use of linear discriminant analysis in assessing the probable level of Plug-In EV adoption for PG&E's Service Territory. We used PG&E's GIS system to display them on the maps. This information would help assess the electric distribution system capacity needs to accommodate potential EV adoption in the service territory.

The relatively simple framework of such analysis can be easily updated as the new generation of EVs become more widely available and infrastructure is installed to support their use. As new EV customers opt for the PG&E EV rate, they can be added to the prototype database and the analysis can be repeated. It would also be possible to choose only customers opting for the rate after this current generation of EVs became available to see if their demographic structure is different than that of the earliest adopters.

In the energy utility industry LDA methodology would have numerous applications. It could be used to target many demand side management programs such as demand response, energy efficiency, or customer owned renewable generation. It can be applied anywhere that a behavior (e.g., adoption of an EV, installation of rooftop solar, purchase of an energy efficient appliance etc.) can be mapped to a customer and that customer to a census block group for demographic representation. The methodology can be scaled to cover any geographic area from the city to the state level. It can be done at a regional level to compare the demographics of adopters across different areas. The methodology is relatively inexpensive to implement, intuitively appealing, and relatively easy to explain.

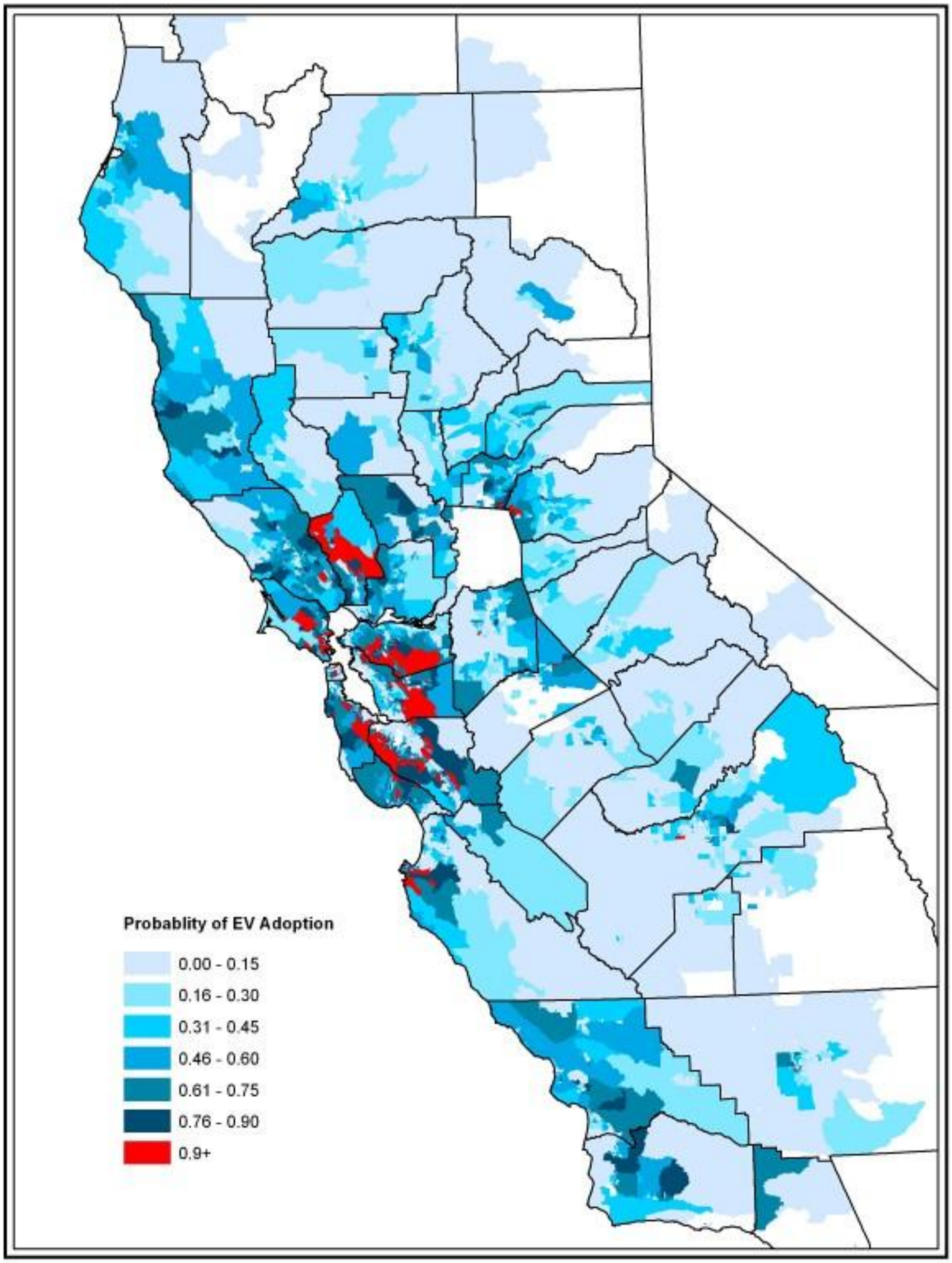

Figure 2: Probable level of Plug-In Electric Vehicle Adoption for PG&E's Service Territory.

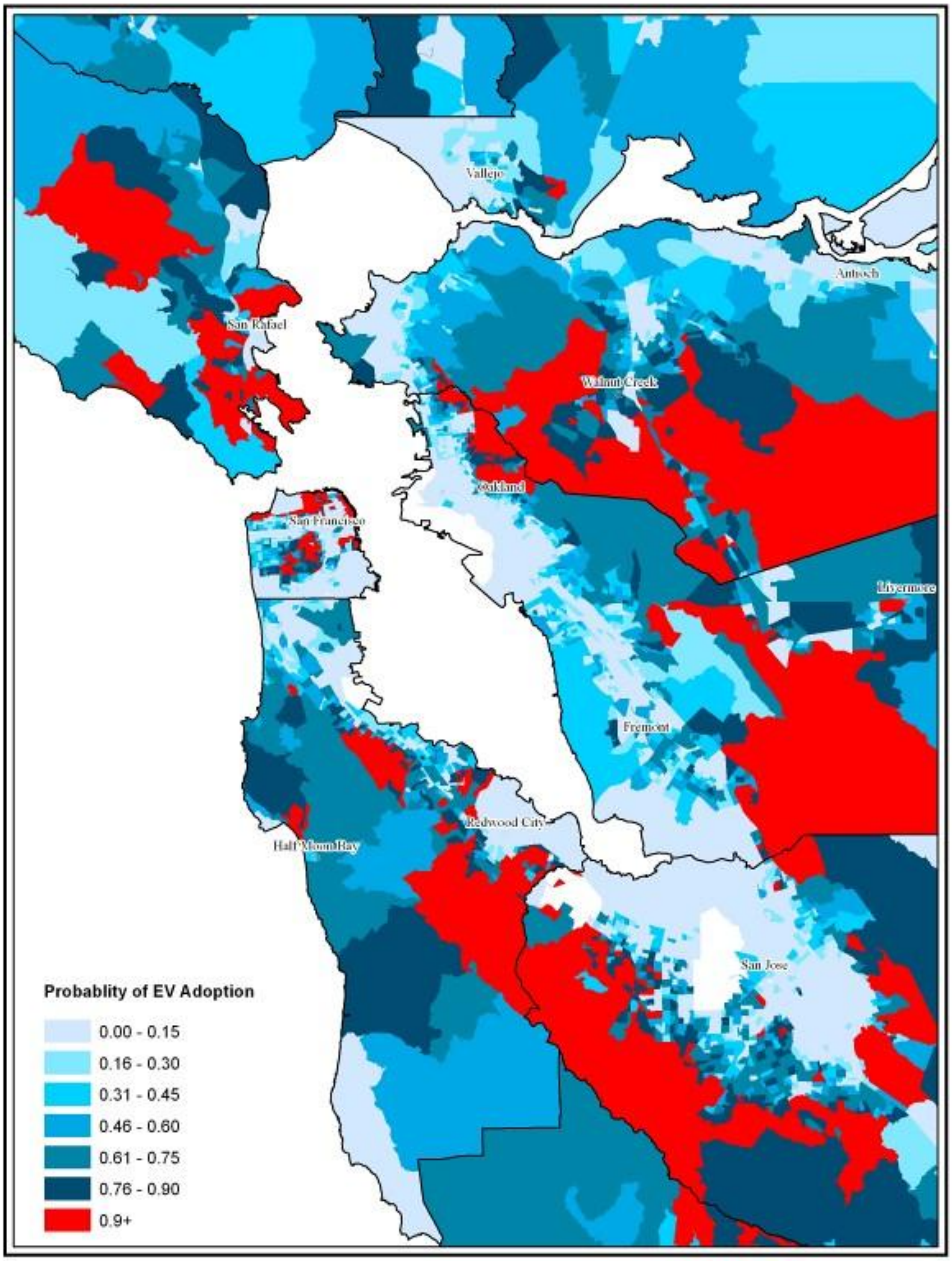

Figure 3: Probable level of Plug-In Electric Vehicle adoption for the San Francisco Bay Area.

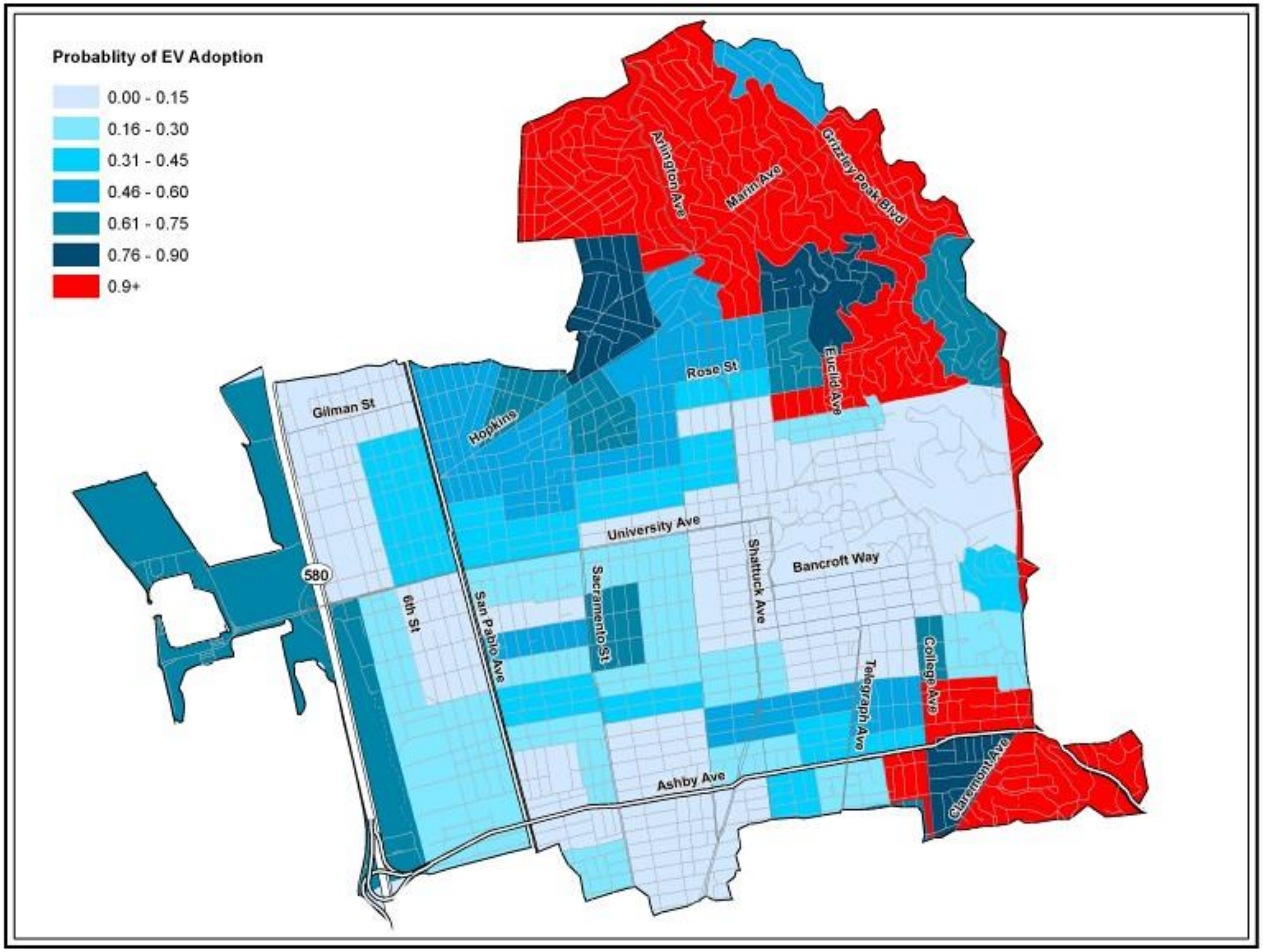

Figure 4: Probable level of Plug-In Electric Vehicle Adoption for Berkeley, California.