**TITLE: EV Propensity Analysis**  
POC: Yoon Soo Lee ([yoonsoo.lee@nationalgrid.com](mailto:yoonsoo.lee@nationalgrid.com)), ADA

OBJECTIVE:

The EV propensity analysis is geared to providing a robust EV adoption/purchase forecast of NGrid customers at the zip code level. The analysis uses a well-known classification method called random forest, which is a decision tree learning algorithm to find the correlation between attributes in making certain decisions in the future[[1]](#footnote-1). For our purposes, we will train the model based on the historical EV growth in vehicle[[2]](#footnote-2) numbers against the zip code level demographic customer data[[3]](#footnote-3).

A granular and location-based EV growth model is essential for the company in twofold: First, it will help NGrid to better determine potential sites for EVSE to not only meet the anticipated demand from EV growth, but also to instigate EV growth in target areas with lower EV density.

PROCESS:

The following section illustrates how to set up the data preparation and to run the propensity on the data for individual NGrid jurisdiction:

* In SQL, extract the customer data for specific NGrid jurisdiction from the ACXIOM database (by\_zip\_acxiom\_ny.sql). Normalize all relevant customer demographic data so that they can be quantified and used in an algorithmic function (Appendix A).
* In SQL, join the AXCIOM customer data with the EV sum and population data per zip code pulled from GIS (NY\_ZIP\_EV.xlsx) using ‘merge\_ciap\_gis\_ny.sql’ file.
* In SQL, download the Table ZIP\_AZXIOM\_DATA\_NY into the local machine, to be used as the training and test data for the propensity model.
* In Python, run the ‘ny\_propensity.py’ to initiate the propensity analysis, along with the ranking of zip codes based on the propensity. Ranking is given in an ascending order, where higher the ranking higher the probability that customers in that zip code will adopt/purchase an EV.

CONCLUSION:

According to the statistical regression produced by the random forest, only a handful of customer demographic data had the most significant correlation with the EV growth (Appendix B): Population density, education level, and financial status of customers seemed to be the strongest triggers toward EV adoption. This is synonymous with the popular literature on EV adoption. Using the top 5 attributes, a cost function is used to calculate the ranking for the zip codes, which is found in ***4.Propensity/Results/output.xlsx***.

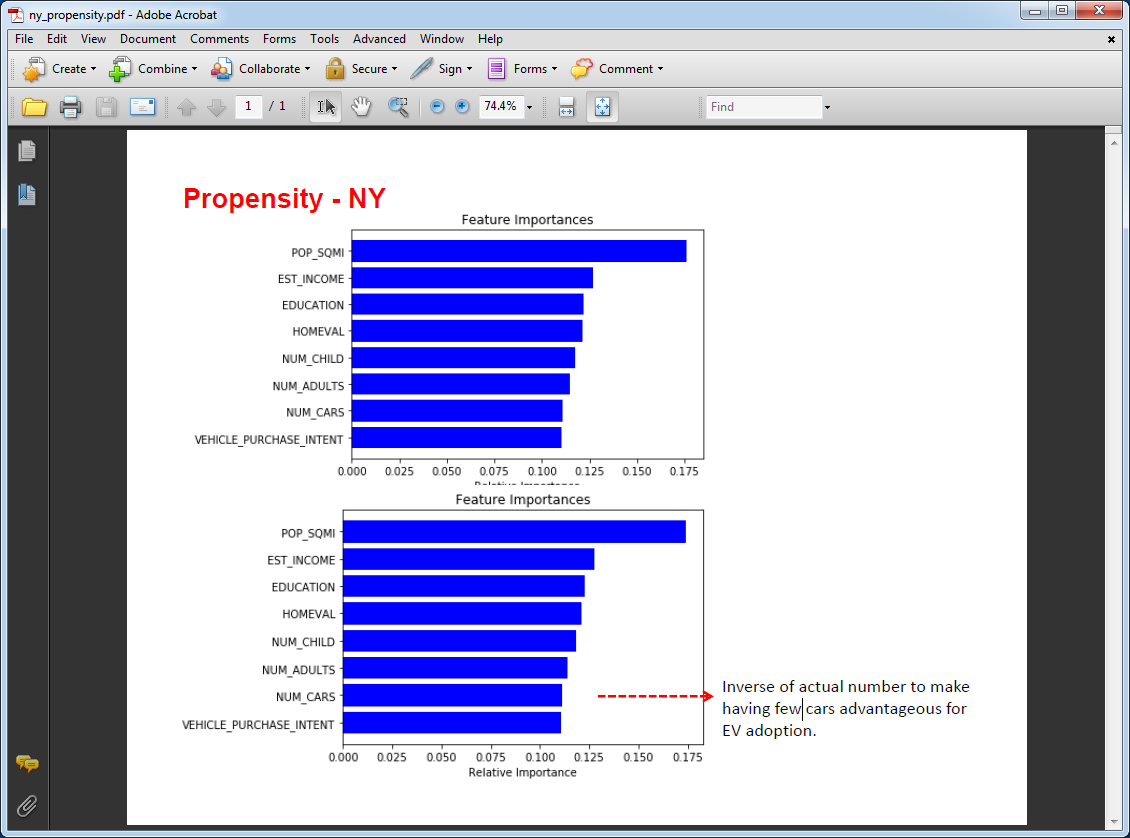
TO DO: Update all customer demographic data and the cost function for the ranking.

**APPENDIX A**

The following table itemizes the customer demographic data pulled from ACXIOM database for the propensity analysis.

|  |  |  |
| --- | --- | --- |
| ACXIOM\_Data | Content | Normalization |
| V8641 | Household income | 1 = Less than $15,000  2 = $15,000 - $19,999  3 = $20,000 - $29,999  4 = $30,000 - $39,999  5 = $40,000 - $49,999  6 = $50,000 - $74,999  7 = $75,000 - $99,999  8 = $100,000 - $124,999  9 = Greater than $124,999 |
| V8626 | Number of adults in household | Average. |
| V8602 | Number of children in household | Average. |
| V7475 | Vehicle purchase intent | 1 = ‘Yes’, 0 = ‘No’ |
| V9514 | Education level | 1 = Completed High School  1 = Attended Vocational/Technical  2 = Completed College  4 = Completed Grad School |
| V2100 | Ethnicity | 1 score per each person /1000 |
| V8647 | Number of existing cars | 1, 2, and 3 or more cars |
| V8560 | Existence of solar panels | 1 = ‘Yes’, 0 = ‘No’ |
| V8713 | Home value | Average/100000 |

**APPENDIX B**



*Mean cross validation score = 0.614288  
Cohen’s kappa score = 0.53781  
Accuracy Score = 0.58889*

1. https://en.wikipedia.org/wiki/Random\_forest [↑](#footnote-ref-1)
2. POLK database. [↑](#footnote-ref-2)
3. ACXIOM InfoBase Data. [↑](#footnote-ref-3)