Explain data

current variables and classification types

other potential variables

* Median, mean, avg. age of properties, proximity to highways, sq. footage, property of values,

Explain how the grid is

Preliminary Results that

Future steps:

* and predict how the usage of a charging station will be at a new location within the city.
* Get similar environmental data for charging stations outside of Harris county
* SVM and neural networks
* Regression resulrs

The objective of this draft is to summarize the steps that we have used to do a preliminary analysis of the effect of environmental variables on the demand of Electric Vehicle (EV) charging stations set up by NRG evGo in Houston city.

For our study, we obtained our data from 2 sources :

1. NRG evGo Charging data

1. **Harris County data**

The Harris County data includes real buildings data (available in 4 relevant text files) and polygon data (available in a single json format file).

**Harris County Real Buildings data:** This data was obtained from the Harris County Appraisal District website <http://pdata.hcad.org/download/index.html> . Since we are trying to analyze the effect of environmental variables on the demand of a charging station, the data that we were interested in, for our analysis was the Harris County real buildings data- in particular, the data about important variables like building type/use code, sq. ft. area, appraised property value, number of units in a building. This information is contained in four separate files- **Building\_res.txt**, **Builing\_other.txt**, **Extra\_features\_detail1.txt** and **Extra\_features\_detail2.txt**.

All these files have a unique HCAD (account) number for each building, which be used to merge Thus far, for our preliminary analysis, we have only been using the Building Type/Use code as a variable to study the demand of charging stations.

**Harris County Polygon Data**: The Harris County polygon data file (in json format) provides information regarding the geographical coordinates and physical address of a building associated with a particular HCAD number. The information about the HCAD number was especially important for us as it allowed us to map the features obtained from the Harris County Real Buildings Data to a geographical location.

1. **Classification of Building Type/Use codes:** There are in all around 180 building type (impv\_type) codes in the files. For the purpose of our study, we grouped all these building type codes into 19 different categories, namely: Residential buildings, Offices, Industrial buildings, Warehouses, Shopping centers, Restaurants, Recreation centers, Movie halls/Theaters, Social/Cultural halls, Medical Center, Service Stations, Banks, Transport Terminals, Emergency stations, Libraries, Religious places, Car dealerships, Public Utility Infrastructure companies and Amusement Parks . These categories were made in order to study the effect of similar kind of buildings together. However, taking into account the categories made so far, we agree that there is still some scope of improvement in the classification scheme which we have developed.
2. **Dividing Harris County into square grid:** For analyzing the effect of environmental variables, we divided the Harris County into a grid consisting of squares of sizes 10,000 x 10,000 sq. ft each and then aggregated information about different features like total number of residential buildings/offices/shopping centers/restaurants etc. for each square in the grid.

By mapping the coordinates of each charging station, we located the grid square to which the charging station belonged. We then considered the aggregated features obtained in the previous step for that particular grid square (in which the charging station lay) to do a regression analysis to identify the influence of those features on the charging station demand.

We note here that dividing the city into a square grid and extracting features for each square in the grid is especially useful in achieving our final objective, which is - to be able to predict the demand of a charging station if placed at any new location in the city. Since it is very likely that the charging station within a square will not be located in the centre of a grid square, therefore, in general, because of asymmetry, we cannot assume that the features in the grid square will represent that charging station perfectly.

Therefore, as a next step, we propose to first map the location of a charging station and then consider the aggregated features in a circular of region of about 2 mile radius for regression analysis.

1. **Regression analysis:** For our regression analysis, we considered the response variable to be the number of monthly transactions for a charging station in month t, where t was assumed to be the time period from the first transaction. We used the NRG eVgo DC Charging stations’ transactions data to extract the values of the response number of monthly for the set of independent environmental variables obtained from step 4.

Results from regression

1. **Next steps**

Some of the next steps on which we plan to work on are:

1. Use the results obtained from regression analysis to predict how the demand of a charging station will be at a new location within the city.
2. Some of the charging stations were found to be outside the domain of Harris County and hence, values of environmental variables for those charging stations could not be determined.
3. As mentioned before, some modifications could also be made to the classification scheme of building types and new results could be obtained accordingly.
4. Other forecasting methods based on Support Vector Regression and Neural Networks could be tried out next.