Haverhill Deliverable 0

Team members: Juntao Wang, Yanshan Song, Yifei Fang, Tian Ding

1. Motivation/background of the project

The City of Haverhill aims to convert their 311 customer service data (QAlert) to best assist the city government in preparing for first responses to citizen requests. The city would like to map the data by wards and precincts to best predict and prepare for requests. The city would like to merge their existing databases with precinct and ward information in the GIS format with their 311 customer service data (QAlert) to gain further insights on customers. They would like to map this data and have us merge their databases together.

2. Goals, hypothesis, outcomes

CDBG Area and Wards/Precincts:

a. Extract CDBG geometry data from Hav_CDBG_Area_WGS84.json and Precincts/Wards geometry data from Hav_Precincts_Wards_WGS84.json.
b. Plot the detailed area on a map in the order of CDBG, Precinct, and Ward.
c. Analyze the QAlert data in the haverhill-requests.csv and sort the services that are being requested by frequency per ward.
d. Use different colors to mark different kinds of services per ward on the map.

Refuse Routes:

a. Extract data from HAV_Refuse_Routes.json.
b. Classify the data into weekdays of BlueWeek and RedWeek, or sometimes there will be a site for Every week.
c. According to the division above, mark them on the map. The blue week will be colored blue, and Red week will be colored red. As for the MERCANTILE or other areas, they will be colored green.

3. Non-goals, out of scope topics

a. Learn how to use some python packages like Scikit and Plotly.

b. Try to use some methods like regression, classification and clustering analysis in data science.
c. Now the refuse routes are just marked as different areas but without exact routes. In the future, we can design routes for the compartment trucks to make them collect the refuse as quickly as possible.

4. End result/product -> definition of done

a. Successfully map classifications for existing data and merge databases.
b. Use the data that is given by the client to plot all images. We may need to plot about 10 or more results on the map.

5. Open questions, uncertainties

a. What kinds of services are being requested per ward/precinct along with their frequency?
b. What kinds of services are being requested by the Community Development Block grant (CBG) area along with their frequency?
c. What kinds of services are being requested per tax parcel area along with their frequency?
d. Should we plot a single result for every different ward/precinct or we plot them together and mark them together?