TEAM 1:
AYUSH SHIRSAT
SARTHAK ARORA
JAINAM SHAH

DRAFT ANALYSIS REPORT

Summary

City of Haverhill is interested in improving and refining their decision-making process from insights born out of data analysis and visualization. The 311-system set up receives non-emergency complaints such as Highway street improvements, street light repairs, etc. All relevant details regarding the issue are stored in QAlert. QAlert is a citizen request management solution that has been deployed by the City of Haverhill. The data stored in QAlert serves as a good repository to analyze and visualize. Having access to such data will help us find the underlying patterns and points of interest. The data was made available to us in a csv format.

Relevant fields of data

Field Name	Field Description
Request ID	The unique identifier for the request
Create Date	The date the request was entered into the system
Status	The status of the request
Close Date	The date on which the request was executed
Request Type	The category of the Requests
Department	The department under which the request is assigned
Longitude	The longitude coordinate of the Request origin
Latitude	The latitude coordinate of the Request origin
Origin	The source from which the Request was made

Table 1: Fields of data

Pre-Processing

- Some columns that we felt were not needed were dropped.
- Some columns have both and ID and name, only the ID's were retained as we can always map them to their respective names using dictionary.
- Dates were present in string and were converted to datetime objects with the appropriate format which enabled us to calculate request completion time.
- To plot latitude and longitude all 0 entries were dropped.
- To find completion time only "closed" entries were taken into consideration as others did not have a completion date.

Data visualizations

The dataset consists of 70,060 requests which are categorized as "Closed", "In Progress" or "Open". The distribution of such requests is shown in Figure 1. Closed requests help determine the time taken for a request to be completed. The trends of "closed" requests will determine the approximate time required for "In Progress" requests to be completed.

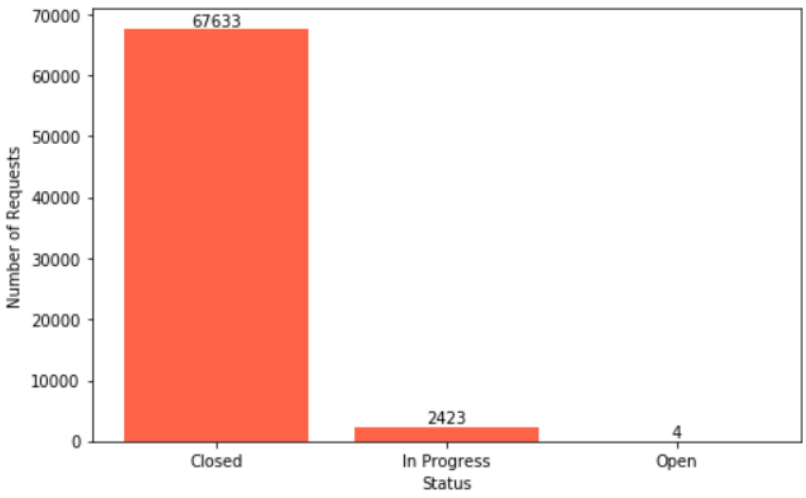

Fig. 1: Status of Requests

Figure 2 demonstrates the origin of requests and the platform people used to report a complaint.

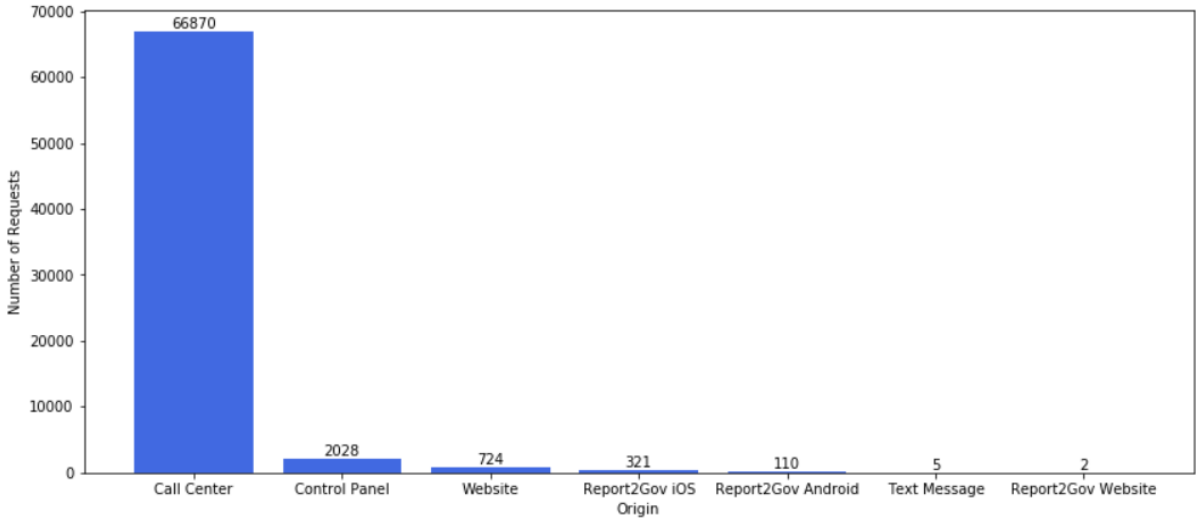

Fig. 2: Origin of Requests

Figure 3 is a plot of all requests based on their geographical location. All points are mapped as per their respective latitude-longitude coordinates. There are 18,080 non-zero coordinates that are mapped. Figure 4 demonstrates these points on actual map of Haverhill.

Note that Folium package is used to plot these points on actual map. Only 1000 samples of data are plotted cause of performance issues. However, one can notice some trends on either side of the Merrimack river where most people reside. The density of requests is higher in this region.

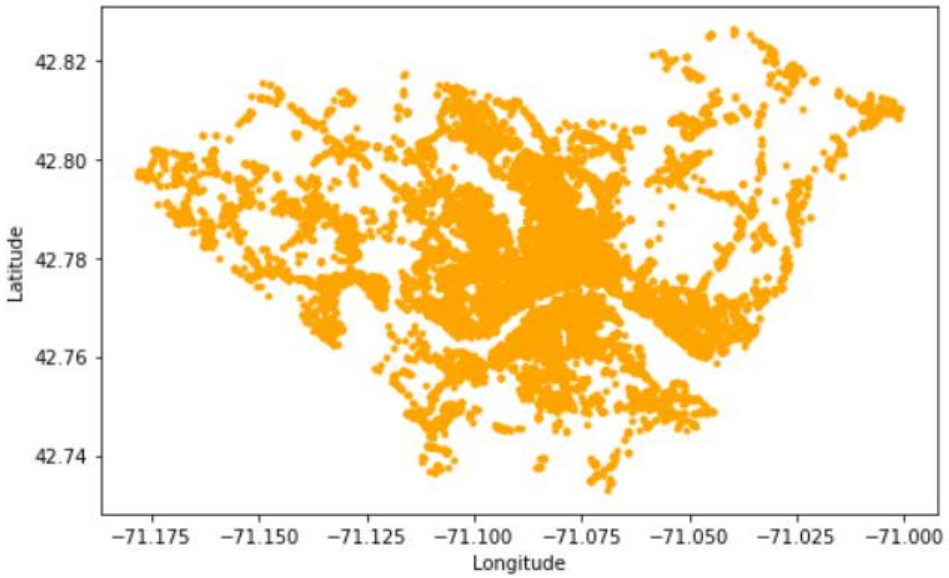

Fig. 3: Distribution of all requests

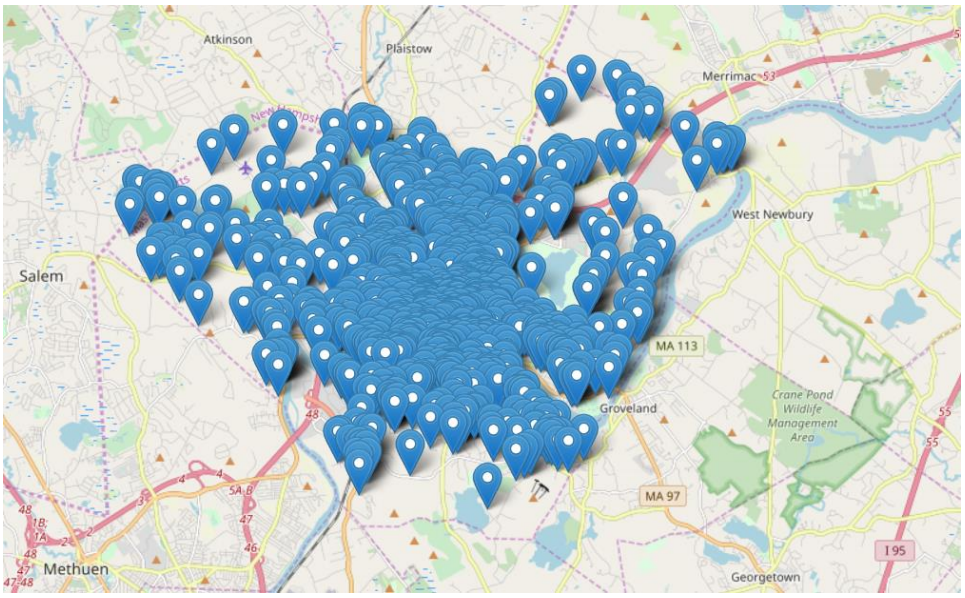

Fig. 4: Distribution of requests (1000 samples) on map

Figure 5 shows the count of requests department wise. Most requests are handled by 311 call center and Highway. In future we might consider grouping departments with less counts as one category.

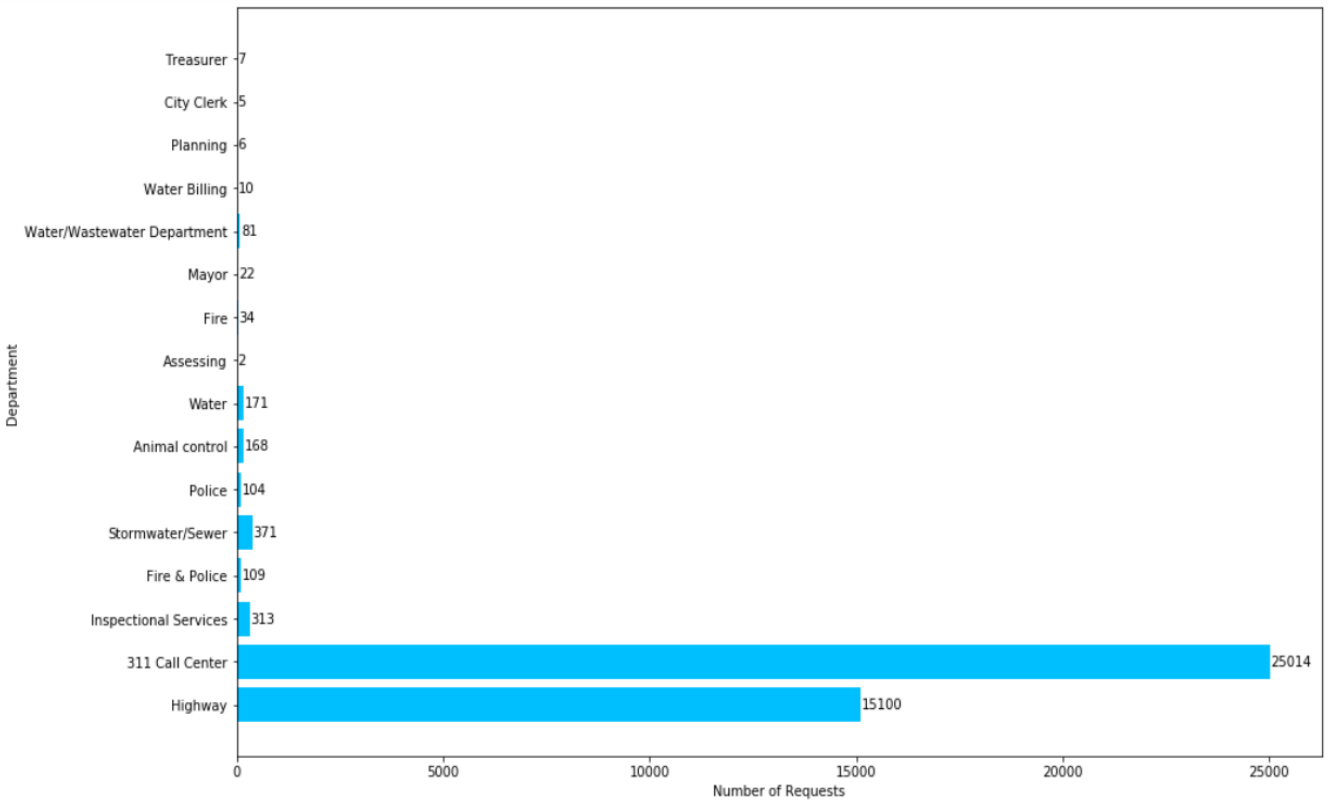

Fig. 5: Department requests

Figure 6 shows the number of requests as per season. This will be useful in predicting if a certain type of request is seasonal.

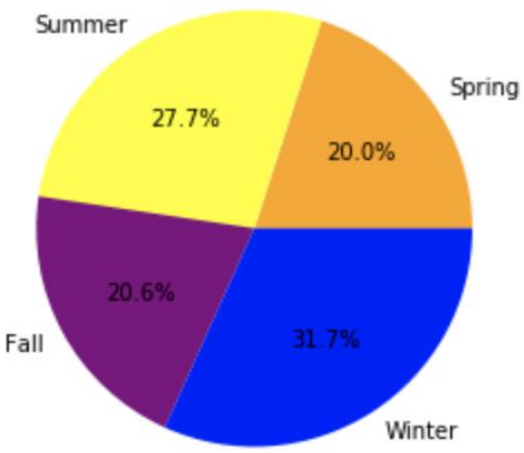

Fig. 6: Percentage of requests based on season

Spring(Apr,May), Summer(Jun, Jul, Aug),
Fall(Sep, Oct, Nov), Winter(Dec, Jan, Feb, Mar)

Figure 7 shows the count of requests based on its request type. Mostly information related requests have a higher count.

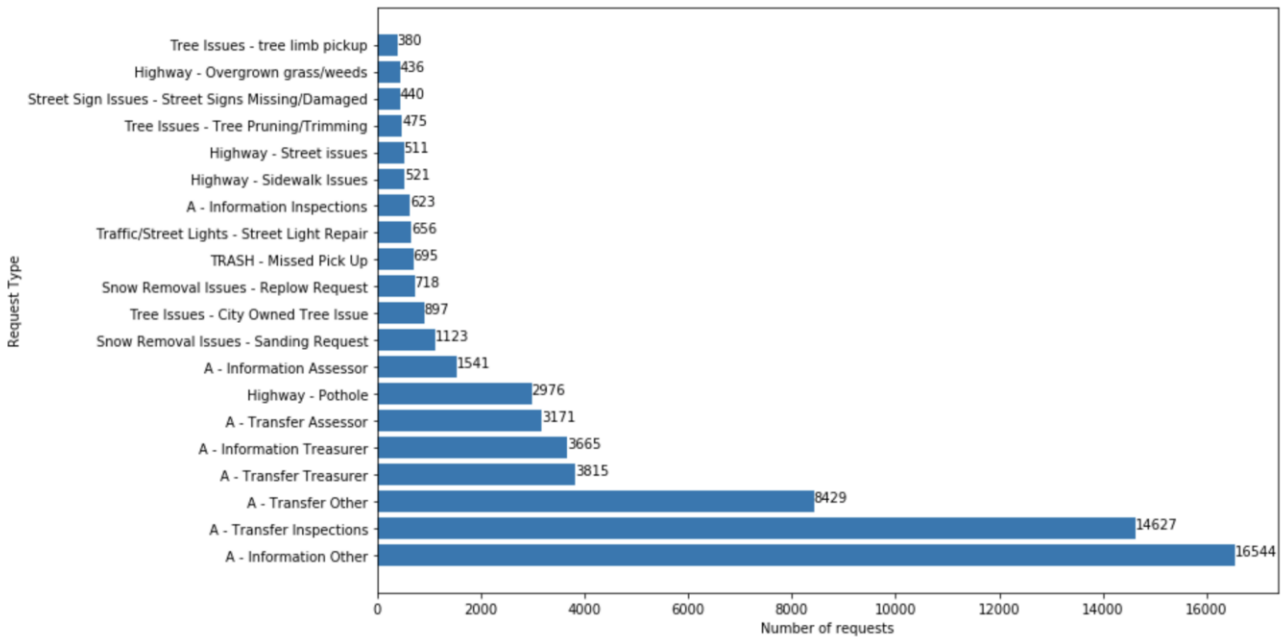

Fig. 7: Requests by type (top 20)

Table 2 shows the average time taken by the top 20 Types of Requests to complete. This gives us a rough idea as to how different types of request vary in their average times and can also act as a good indicator to take action on requests that should not take long but are.

Request Type	Avg Time taken
A - Information Other	0 days 00:59:55.383220
A - Transfer Inspections	0 days 00:05:03.129828
A - Transfer Other	0 days 00:09:54.960256
A - Transfer Treasurer	0 days 00:00:00
A - Information Treasurer	0 days 00:00:00
A - Transfer Assessor	0 days 00:05:28.647114
Highway - Pothole	1 days 21:52:23.306451
A - Information Assessor	0 days 00:12:00.700843
Snow Removal Issues - Sanding Request	0 days 23:19:59.038290
Tree Issues - City Owned Tree Issue	2 days 14:11:37.056856
Snow Removal Issues - Replow Request	4 days 13:51:17.715877
TRASH - Missed Pick Up	0 days 05:30:48.086330
Traffic/Street Lights - Street Light Repair	2 days 19:27:45.640243
A - Information Inspections	0 days 00:00:00
Highway - Sidewalk Issues	1 days 14:29:51.708253

Highway - Street issues	1 days 05:10:04.579256
Tree Issues - Tree Pruning/Trimming	0 days 18:52:14.400000
Street Sign Issues - Street Signs Missing/Damaged	6 days 02:39:33.818181
Highway - Overgrown grass/weeds	1 days 09:07:54.633027
Tree Issues - tree limb pickup	0 days 00:00:00

Table 2: Average time per request type (top 20 types)

Table 3 depicts the average time taken by each department to close the request.This table gives us an overview as to how long each type of problems takes on an average which is helpful in improving the process to rectify it.

Department	Average time taken
311 Call Center	0 days 00:42:58.433677
Animal control	0 days 08:47:21.785714
Assessing	0 days 03:13:00
City Clerk	0 days 00:00:00
Fire	17 days 13:58:45.882352
Fire & Police	0 days 20:45:45.137614
Highway	4 days 01:07:21.806622
Inspectional Services	4 days 02:32:46.198083

Mayor	14 days 11:41:21.818181
Planning	0 days 00:00:00
Police	0 days 06:12:15.576923
Stormwater/Sewer	1 days 15:27:13.908355
Treasurer	0 days 00:00:00
Water	15 days 01:42:55.087719
Water Billing	0 days 00:09:12
Water/Wastewater Department	0 days 17:48:46.666666

Table 3: Average completion time per department

Plan

Most 311 requests as expected, come from places that are densely populated. When number of requests were plotted with season it was observed that most requests came during winter and summer. Apart from information requests the most popular requests are related to highway potholes, snow or trees. Snow issues are consistent with the high number of requests during winter.

Some of these issues can be grouped together in future which will help us discover a pattern. These trends would be trained by a regression model which would map the type of request with the season or even month. The idea behind using regression machine learning model is so that we get an estimate of time. In future when requests come, these requests can be trained by the same model which would make predictions more robust.

Most requests are handled by 311 call center and Highway department. The average time taken by these departments is noted in a table. While averaging is a decent enough estimate it is not very accurate to outliers, example: if a task takes on an average 1 day however, if the same task was reported to take 1 week on an occasion, the average would be shifted significantly and predictions would be wrong. Once, again we propose to use a regression algorithm with regularization to mitigate this issue and make robust predictions.

The type of request will also be mapped with its latitude longitude to see trends if there is any recurring issue in a particular area.

We plan to look into Weather APIs to correlate type of requests based on weather and draw insights from there. Looking at a particular section of requests, such as only requests during winter months and observe any trends we can draw from there. Looking at such groups of data could help us engineer more features for our analysis.