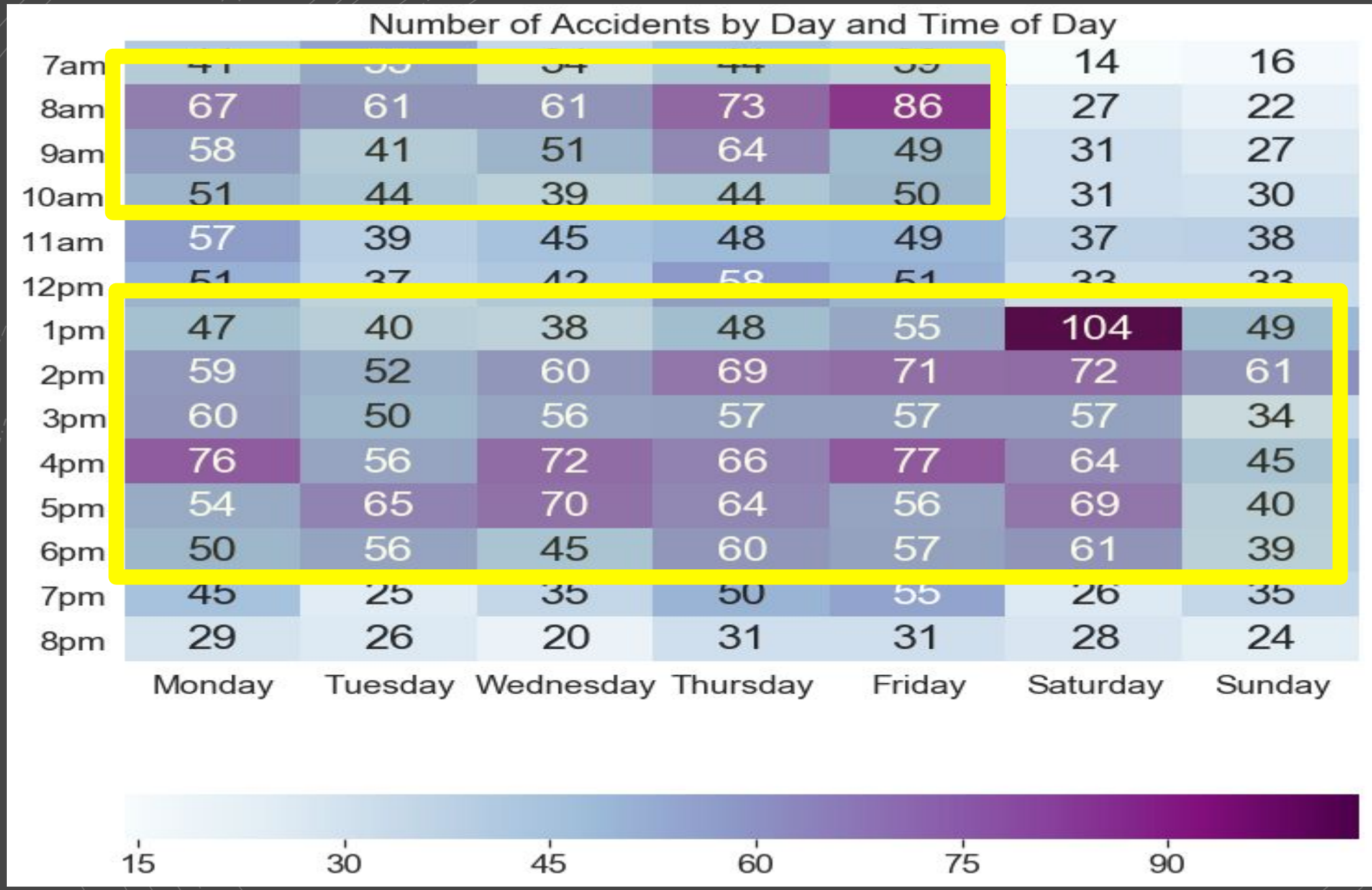
Mapping Evacuation Routes Using Live Traffic Information

Amy Ontiveros-Bear, Benjamin Haile, & Ruchika Sah

Problem Statement

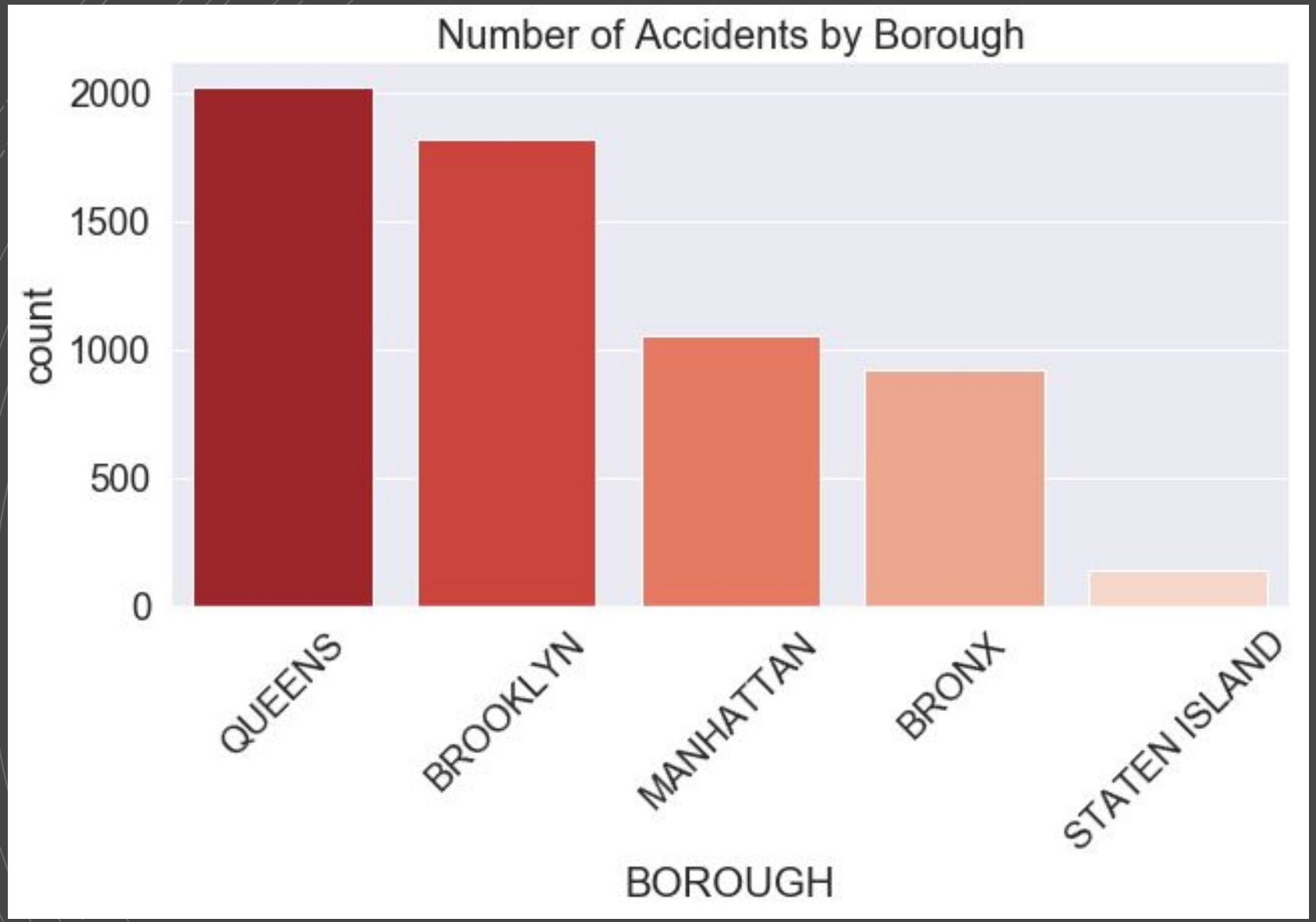
- In this project, we were tasked with optimizing the best evacuation route from Central Park to Prospect Park. The ideal output should map the fastest route given live and historical traffic data. Since Tom Tom essentially does this already, we wanted to validate its efficiency. We will test our theory by utilizing DBSCAN to cluster historically dense traffic regions. We then compare the live routing mechanism of Tom Tom to our cluster output. The pseudo-metric used will be silhouette score.

Exploratory Data Analysis and Outside Research


Number of Accidents by Day and Time of Day
7am 14 16
8am 67 61 61 73 86 27 22
9am 58 41 51 64 49 31 27
10am 51 44 39 44 50 31 30
11am 57 39 45 48 49 37 38
12pm
1pm 47 40 38 48 55 104 49
2pm 59 52 60 69 71 72 61
3pm 60 50 56 57 57 57 34
4pm 76 56 72 66 77 64 45
5pm 54 65 70 64 56 69 40
6pm 50 56 45 60 57 61 39
7pm 45 25 35 50 55 26 35
8pm 29 26 20 31 31 28 24
Monday Tuesday Wednesday Thursday Friday Saturday Sunday
15 30 45 60 75 90

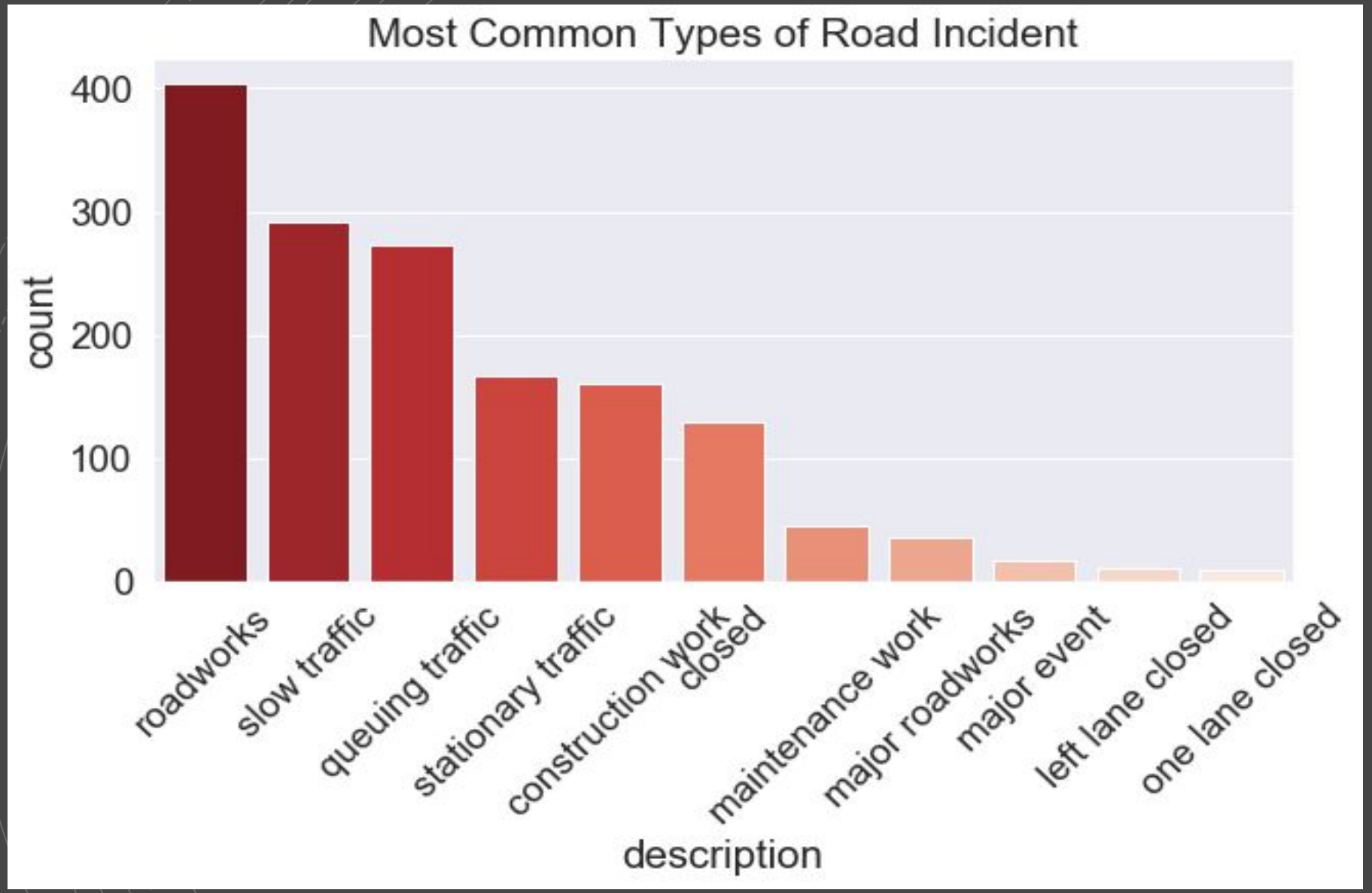
Observations

- Monday-Friday: 8-10 AM, 4-6 PM, consistent with typical rush hours
- Afternoon becomes denser from Thursday-Saturday
- Saturday 1 PM and Thursday 8 AM are especially high


Number of Accidents by Borough
2000
1500
1000
500
0
count
QUEENS
BROOKLYN
MANHATTAN
BRONX
STATEN ISLAND
BOROUGH

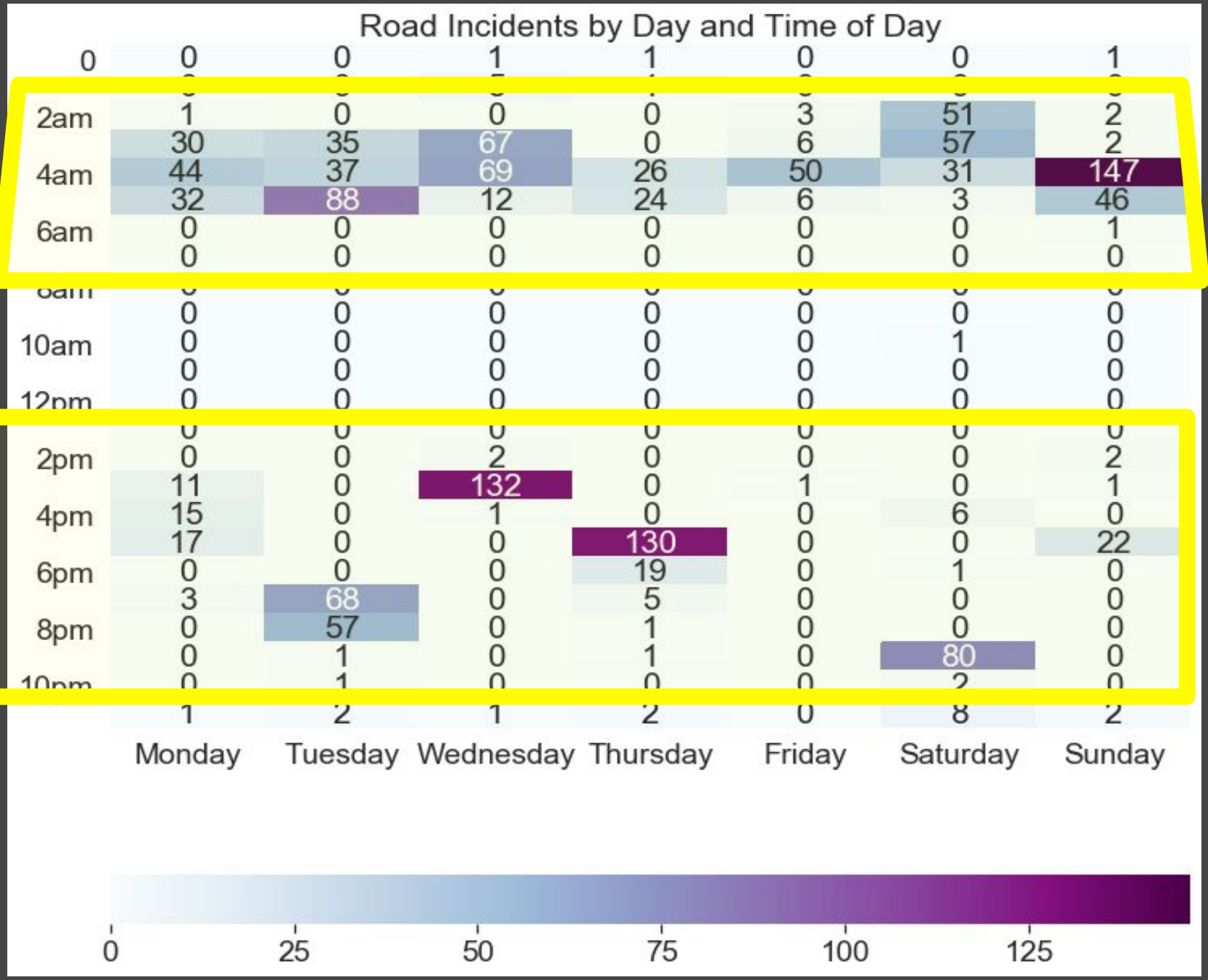
Observations

- Higher count in Brooklyn than Manhattan
- Congestion from Brooklyn-Queens Expressway
- Contradicted by our modeling


Most Common Types of Road Incident
count
400
300
200
100
0
roadworks
slow traffic
queuing traffic
stationary traffic
construction work closed
maintenance work
major roadworks
major event
left lane closed
one lane closed
description

Observations

- Road works, slow traffic, and gridlock are the most common
- Accidents are a lesser cause of closure
- Consider construction and congestion


Road Incidents by Day and Time of Day
0
2am
4am
6am
10am
12pm
2pm
4pm
6pm
8pm
10pm
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
0
25
50
75
100
125

Observation

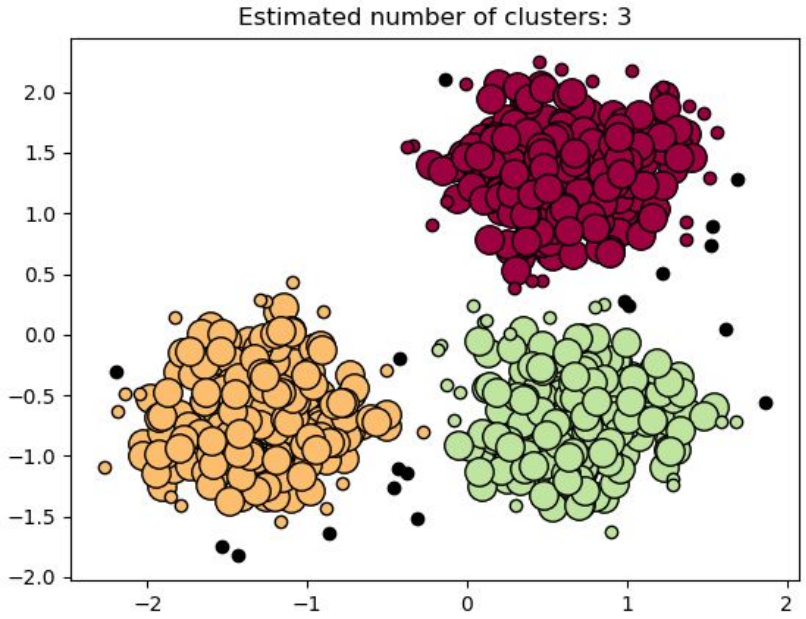
- Monday-Thursday 3-8 PM
- All week 2-5 AM
- Highest: Sunday 4 AM, Wednesday 3 PM, Thursday 5 PM

Modeling

DBSCAN

DBSCAN Clustering

The most used models for unsupervised clustering are the KMeans and the DBSCAN. KMeans requires you to specify the number of clusters in advance and it allocates each observation into a cluster with the nearest mean. We chose the DBSCAN because it is a density-based clustering algorithm that groups points that are close to each other according to a radial distance measurement and minimum number of points. It can find clusters with amorphous shapes and mark outliers as points in low-density regions.



Why did we use DBScan?

We wanted to use a model that clustered historically traffic-dense areas in order to easily visualize evacuation routes and validate the efficiency of Tom Tom's live routing API.

DBSCAN is an UNSUPERVISED clustering algorithm, so there is no true way to evaluate it. The mechanism for clustering is that the neighborhood of a given radius must contain at least a minimum number of points. Therefore, the parameters of the model are the radius and the number of points included in the cluster. For our modeling, we utilized the coordinates of traffic incidents.

Model Evaluation

Being that there is no way to really evaluate the model, we still used pseudo-metrics for our evaluation purposes.

- Inertia is the sum of squared error for each cluster. Therefore the smaller the inertia, the denser the cluster.
- The Silhouette Score is from -1 to 1 and shows the proximity and density of clusters.

We chose the Silhouette Score for our model because it showed us the relationship between our clusters.

Model Evaluation

Percentages of our data in each cluster:

- Cluster 0 : 73.3%
- Cluster 1: 1.4%
- Cluster 2: 2.2%
- Cluster 3: 9.3%
- Cluster 4: 5.8%
- Cluster -1: 7.3%

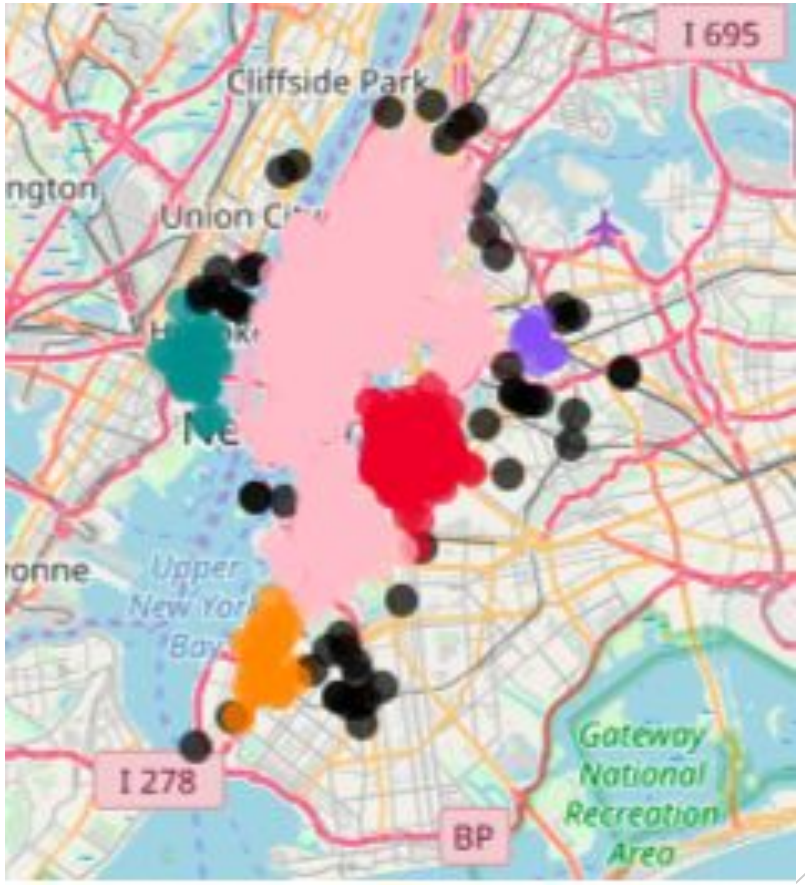
Silhouette Score : 0.40

Being that this is a pseudo-metric and we are utilizing DBSCAN for visual purposes only, we are satisfied with this score.

We mapped our Cluster coordinates. As you can see that the biggest cluster (cluster 0) includes most of Manhattan with the smaller clusters in the outer boroughs. Our outliers are the black dots (cluster = -1), they reflect the fact that those coordinates are further away from the city and have fewer traffic incidents.

Legend:

Cluster -1 = Black
Cluster 0 =Pink
Cluster 1 = Dark Cyan
Cluster 2 = Dark Orange
Cluster 3 = Crimson
Cluster 4 = Blue



Tom Tom For Developers

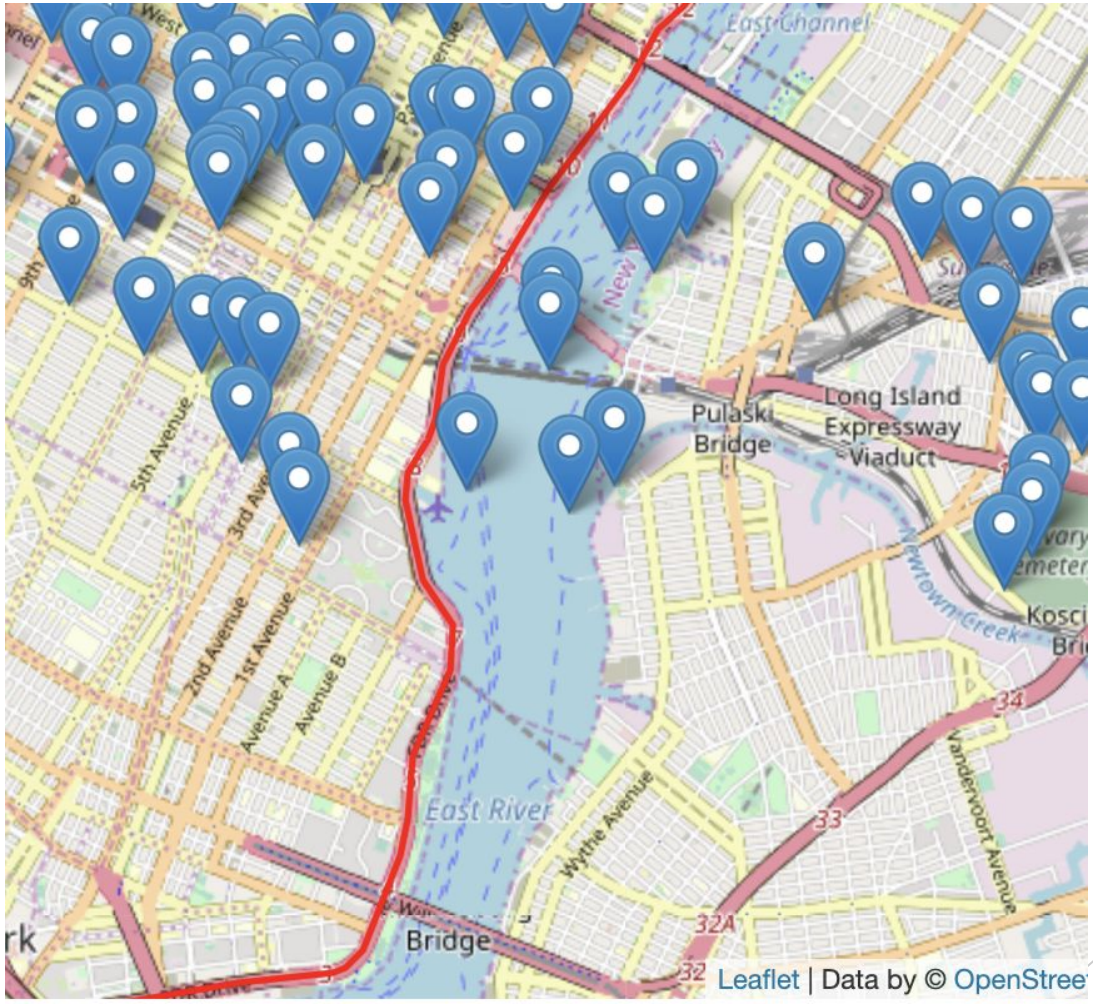
- TomTom is an API that provides real time maps and traffic data, with global coverage.
- We utilized its live routing and traffic incident features for our project.

Tom Tom Advantages

- Can be utilized globally .
- Comes in multiple languages.
- Pulling live data from the API is free to all!
- It is more up to date than government websites.
- The routing system is already trained on historical traffic data, we will prove this in a second.

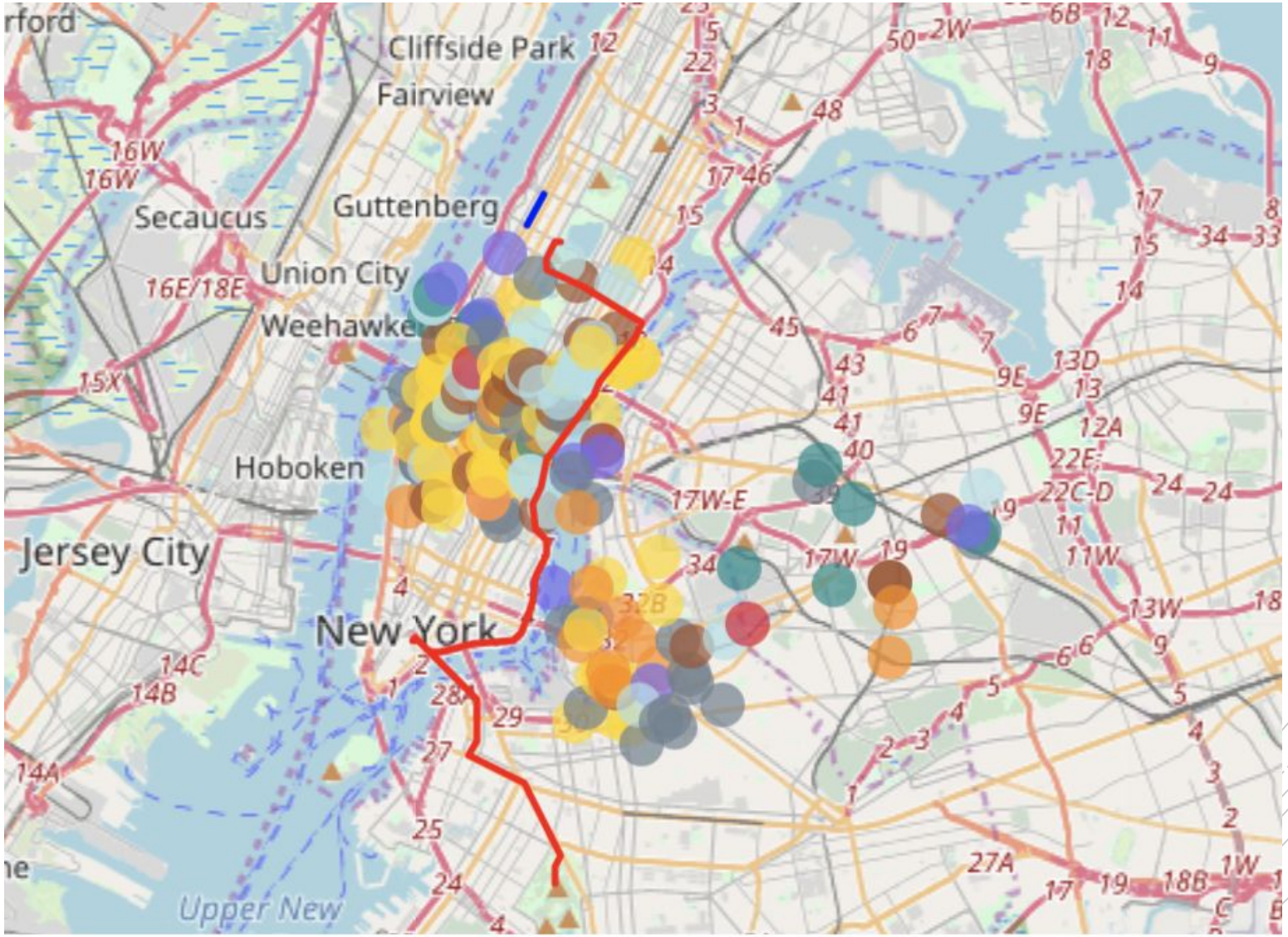
Folium Mapping

- I chose to use Folium because it had the best interactive map interface for python and is highly customizable.



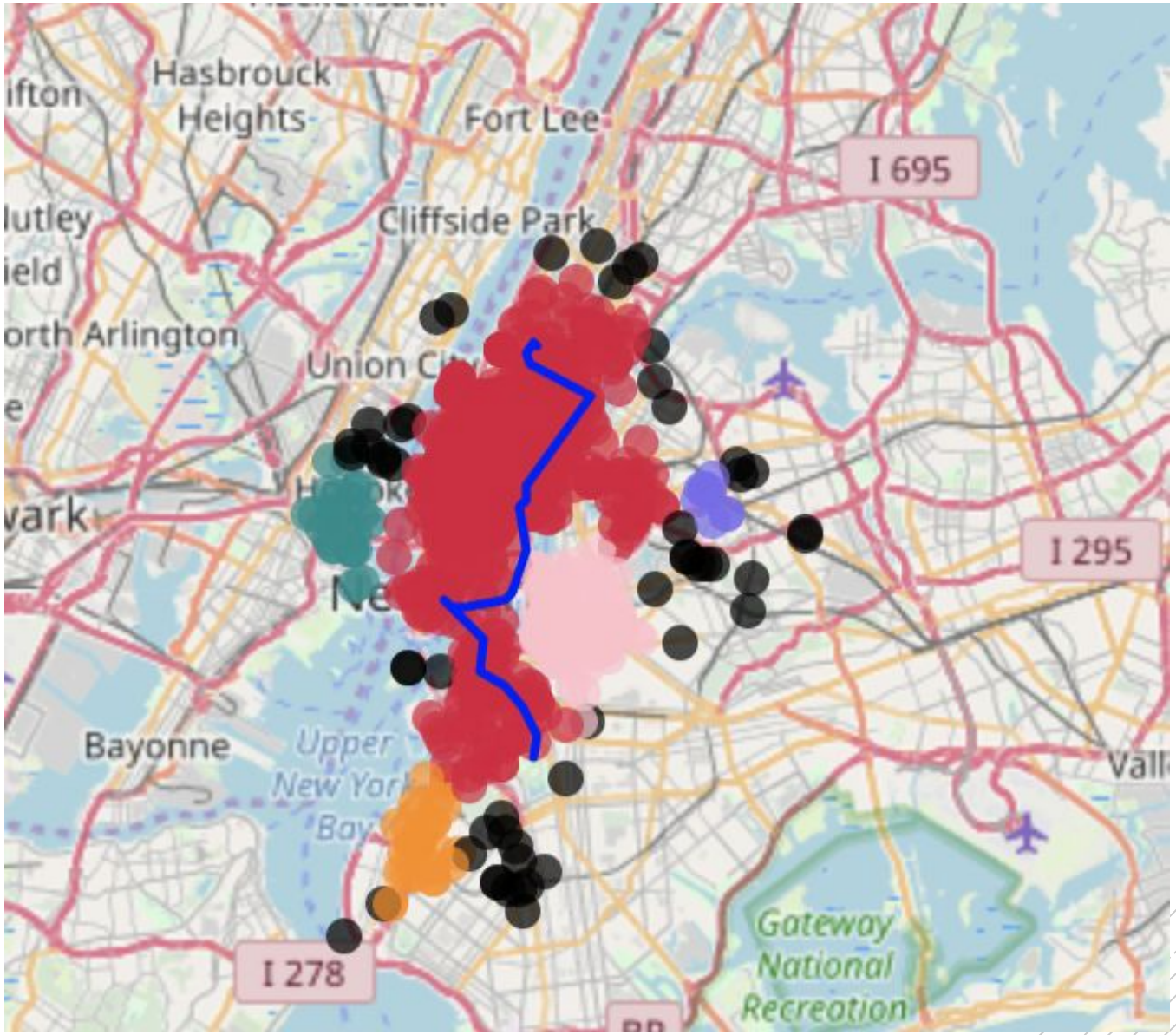
Folium Mapping Live Traffic & Routing

Traffic Incidents are colored coded according to type.



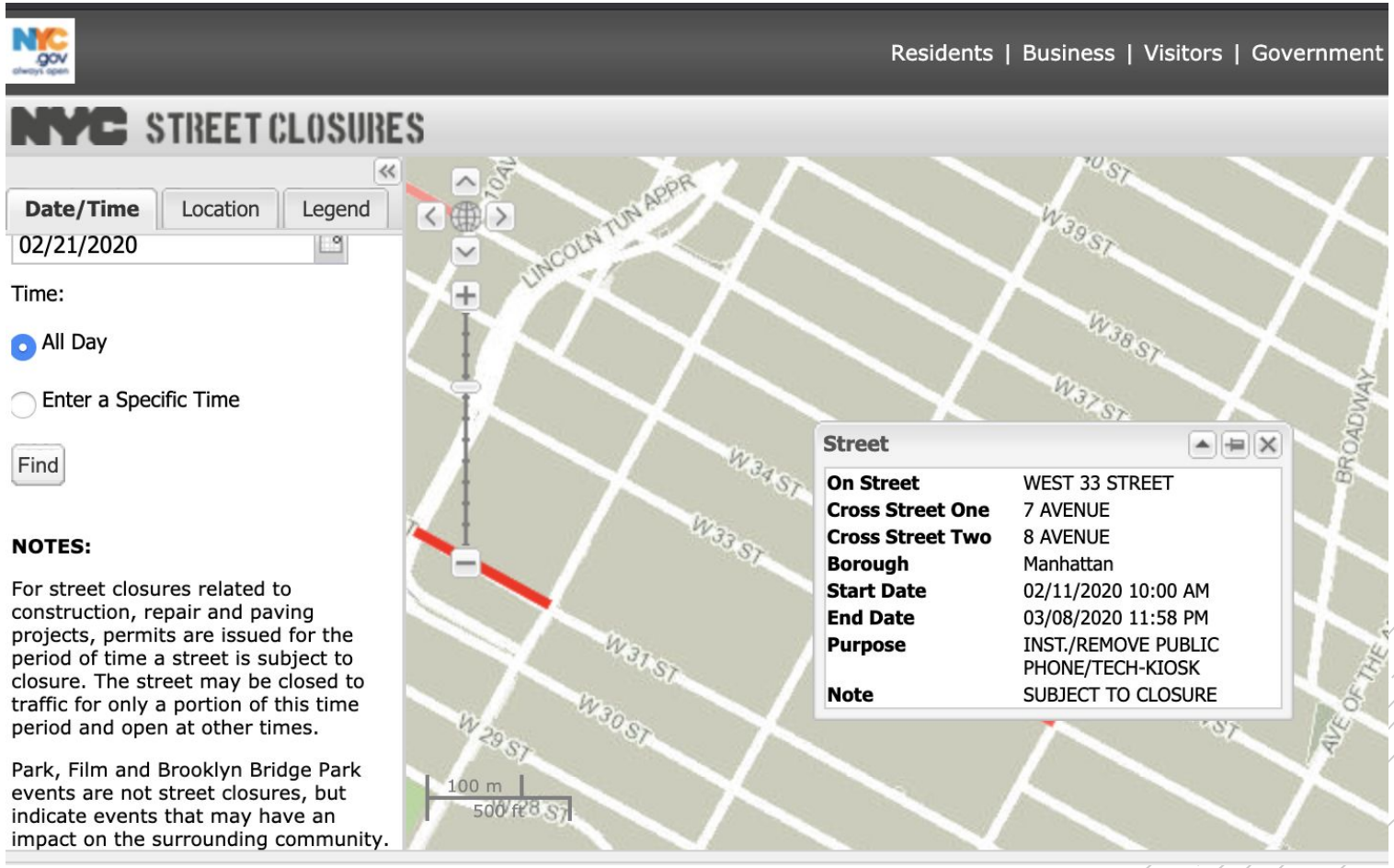
Clusters are color coded

Folium
Mapping
Historical Traffic
Clusters & Live
Routing



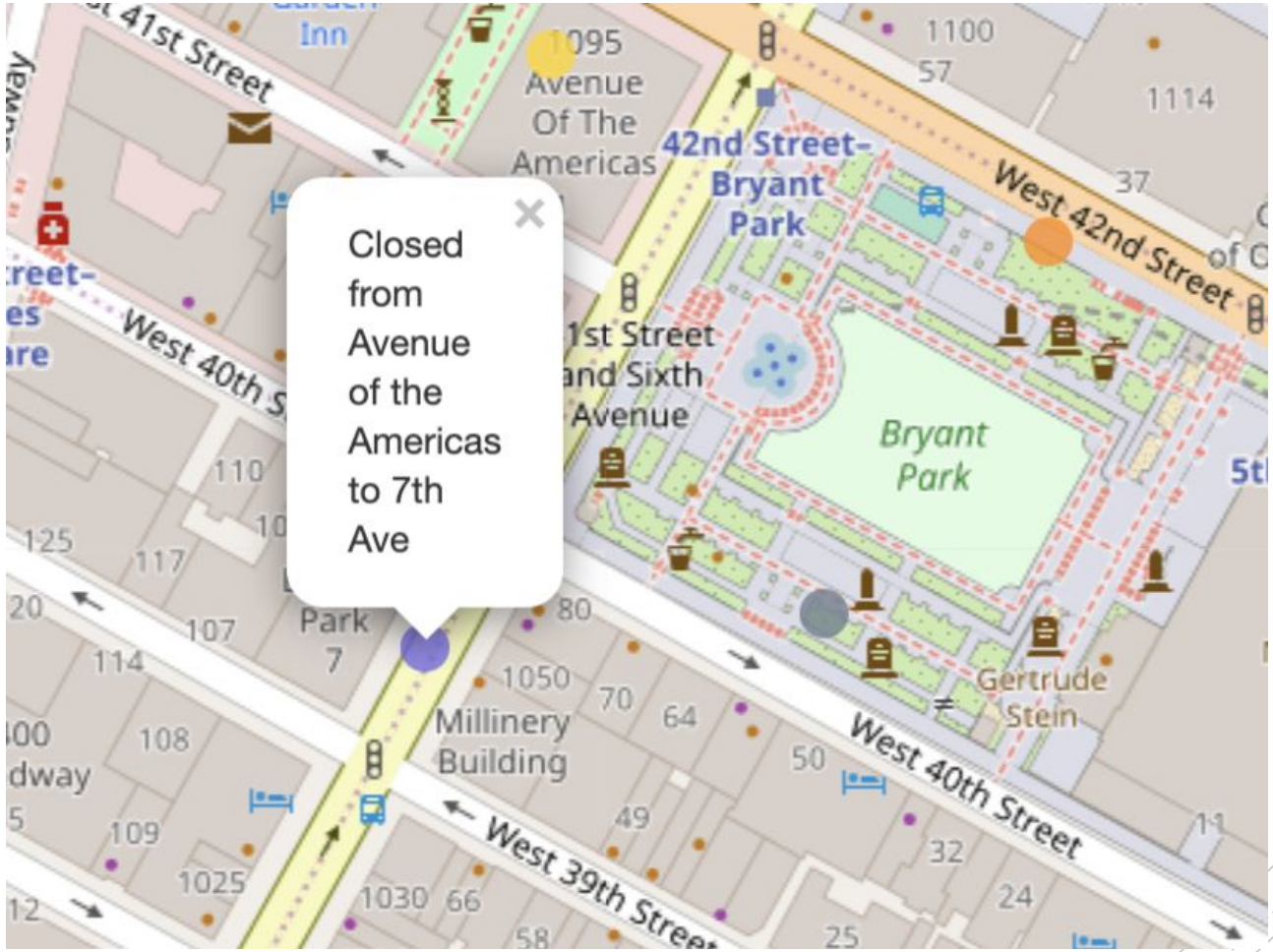
Proving Tom Tom's Efficiency & Accuracy

Government road closure data is not up to date, only gives vague time frames when it may or may not be closed.



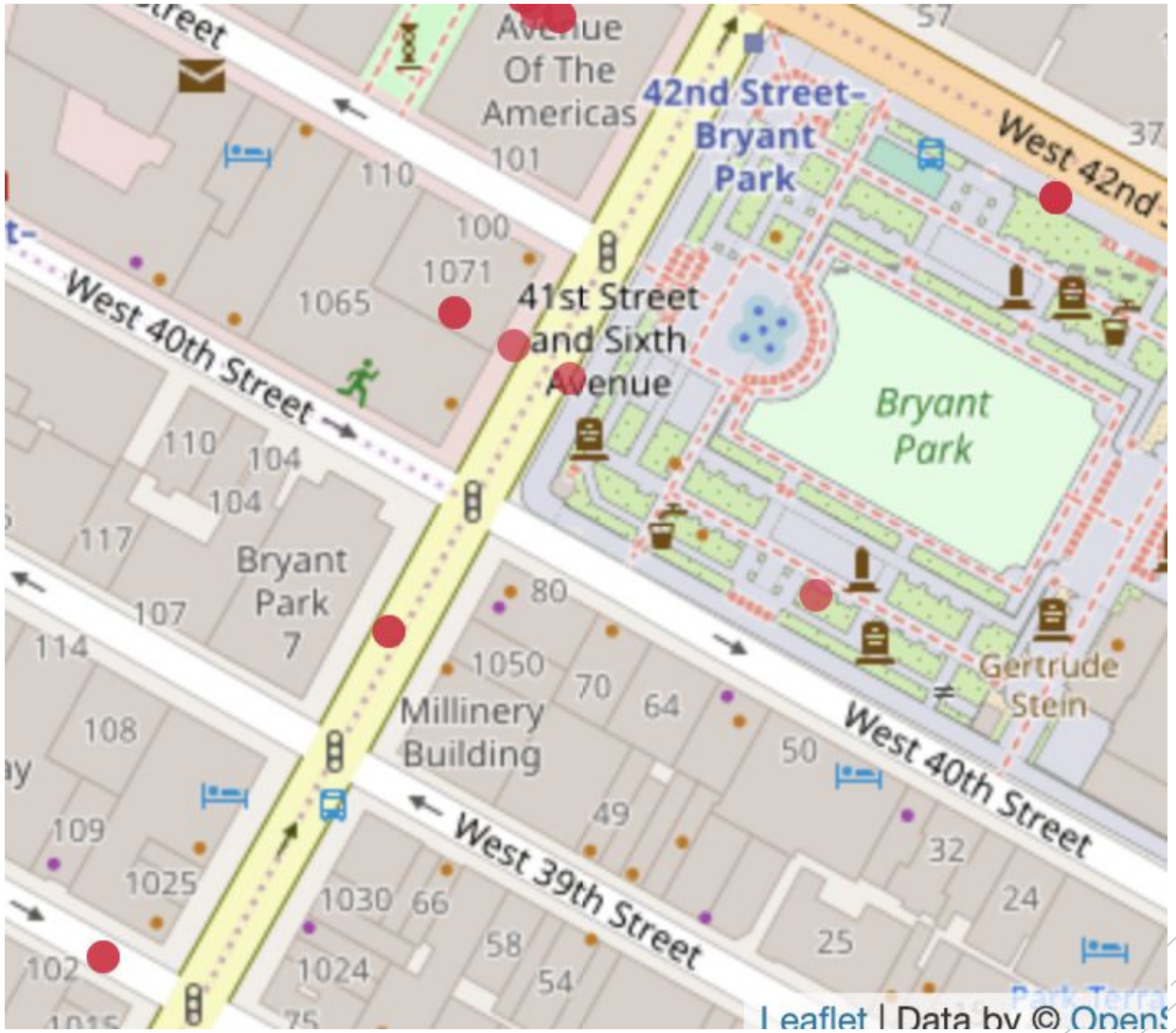
Proving Tom Tom's Efficiency & Accuracy

Live Traffic Incidents at Bryant Park



Proving Tom Tom's Efficiency & Accuracy

Historical Traffic Incidents at Bryant Park



Conclusions

- The DBSCAN clustering algorithm does a decent job of clustering the traffic incident areas. Since it is an unsupervised model and it is being utilized for visual purposes only, a Silhouette score of 0.40 will suffice. Therefore, we are satisfied with our final model.
- TomTom API is a reliable source for live traffic data. It is able to be utilized globally.

Next Steps

- Geocoding to use physical addresses
- Overlay traffic polylines
- Additional research on causes of traffic-dense hours
- Apply to other regions of the world