ITM - 891 Project Presentation

Presented by Vamsi Vivek Teja(adibhat1@msu.edu)

What are we dealing with?

- Main Problem Statement:
    - Utilizing **Telematics** Traffic Data to understand congestion/wait time at signals and predict potential wait time at a signal
- Data Source:
    - Kaggle Data Set
- Cities involved:
    - Philadelphia
    - Boston
    - Chicago
    - Atlanta

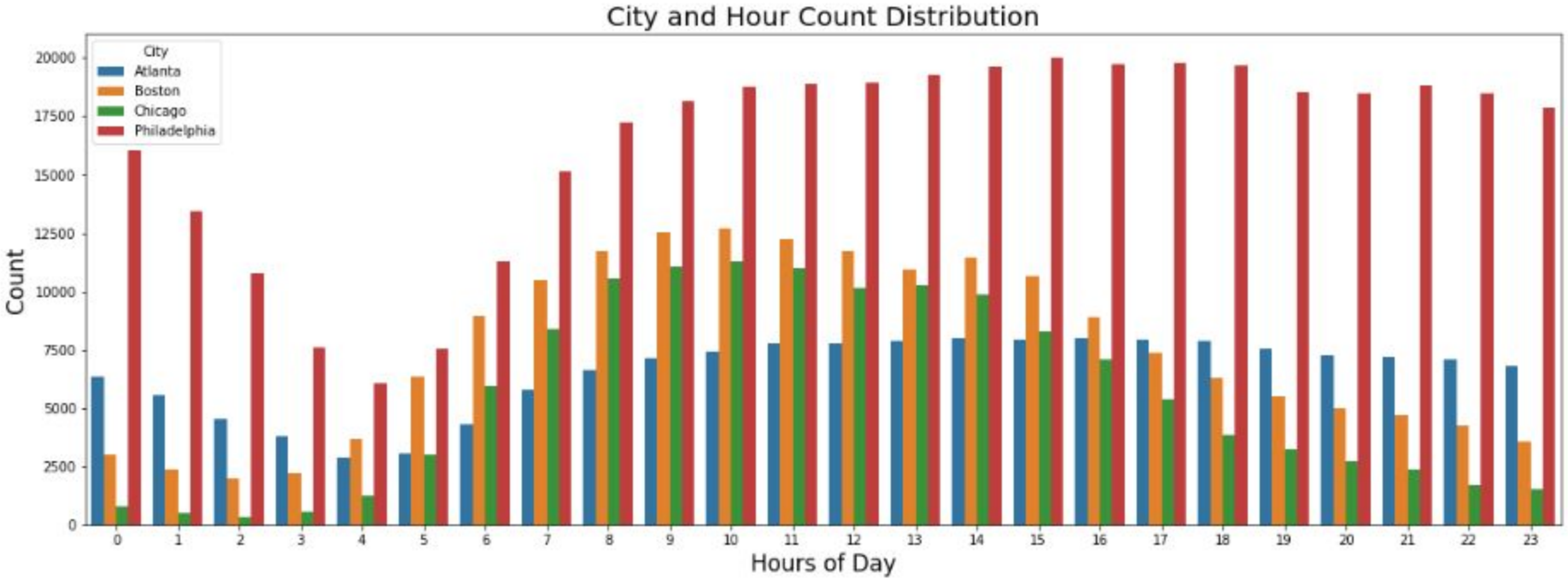
Exploring the Data

- Quantitative Metrics:
  - Total Time Stopped
  - Time from First Stop
- Categorical Metrics:
  - IntersectionID
  - City
  - Hour, Month & Weekend Flag
  - Path taken by the telematics device
  - Entry & Exit Direction
  - Latitude and Longitude of the Intersection

Dataset Shape: (856387, 28)

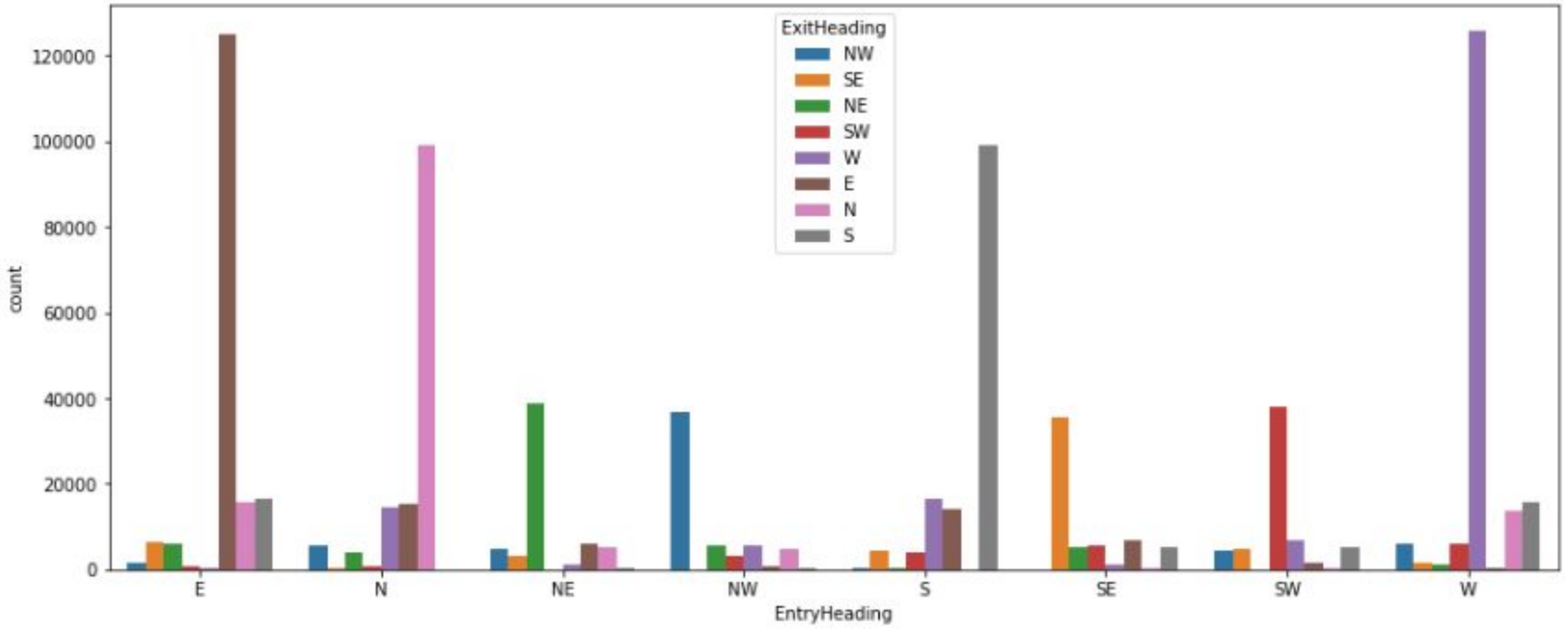
	Name	dtypes	Missing	Uniques	First Value	Second Value
0	RowId	int64	0	856387	1921357	1921358
1	IntersectionId	int64	0	2559	0	0
2	Latitude	float64	0	4799	33.7917	33.7917
3	Longitude	float64	0	4804	-84.43	-84.43
4	EntryStreetName	object	8148	1723	Marietta Boulevard Northwest	Marietta Boulevard Northwest
5	ExitStreetName	object	6287	1703	Marietta Boulevard Northwest	Marietta Boulevard Northwest
6	EntryHeading	object	0	8	NW	SE
7	ExitHeading	object	0	8	NW	SE
8	Hour	int64	0	24	0	0
9	Weekend	int64	0	2	0	0
10	Month	int64	0	9	6	6
11	Path	object	0	15075	Marietta Boulevard Northwest_NW_Marietta Boule...	Marietta Boulevard Northwest_SE_Marietta Boule...
12	TotalTimeStopped_p20	float64	0	171	0	0
13	TotalTimeStopped_p40	float64	0	238	0	0
14	TotalTimeStopped_p50	float64	0	262	0	0
15	TotalTimeStopped_p60	float64	0	306	0	0
16	TotalTimeStopped_p80	float64	0	403	0	0
17	TimeFromFirstStop_p20	float64	0	244	0	0
18	TimeFromFirstStop_p40	float64	0	316	0	0
19	TimeFromFirstStop_p50	float64	0	336	0	0
20	TimeFromFirstStop_p60	float64	0	353	0	0
21	TimeFromFirstStop_p80	float64	0	355	0	0
22	DistanceToFirstStop_p20	float64	0	3631	0	0
23	DistanceToFirstStop_p40	float64	0	6415	0	0
24	DistanceToFirstStop_p50	float64	0	7751	0	0
25	DistanceToFirstStop_p60	float64	0	9826	0	0
26	DistanceToFirstStop_p80	float64	0	13689	0	0
27	City	object	0	4	Atlanta	Atlanta

Exploratory Data Analysis and Insights



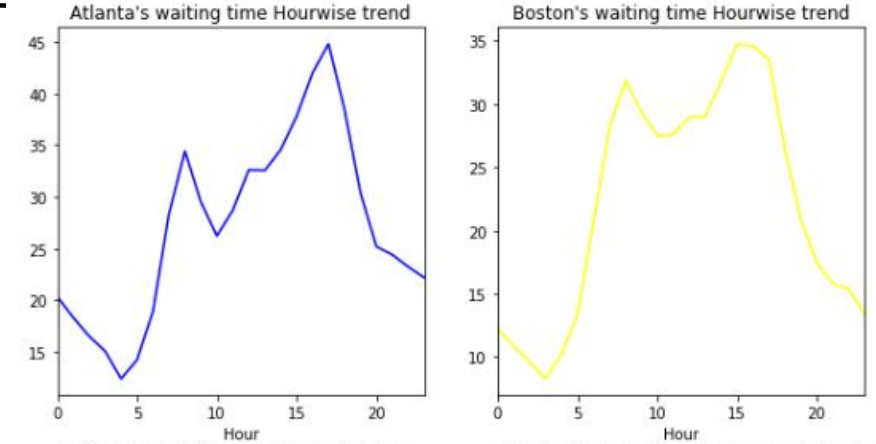
- Philadelphia has a lot of data points(Almost 40 k and 46% of the data)
- Atlanta's Evening Peak remained consistent

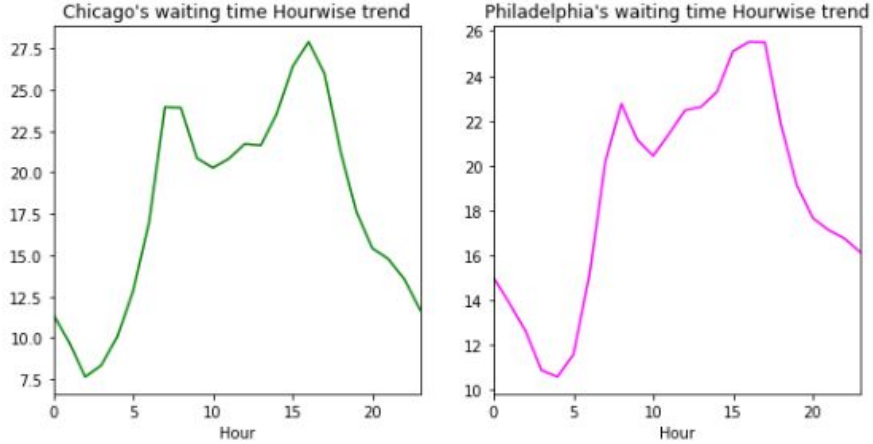
Exploratory Data Analysis and Insights - Entry and Exit Split

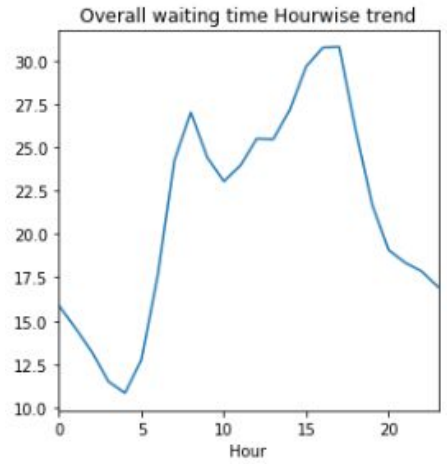


- Traffic moving in a straight direction is more in volume than turns - North, East, West, South have both as Entry and Exit in the 4 tall towers

Exploratory Data Analysis and Insights - City wise Overall Waiting T

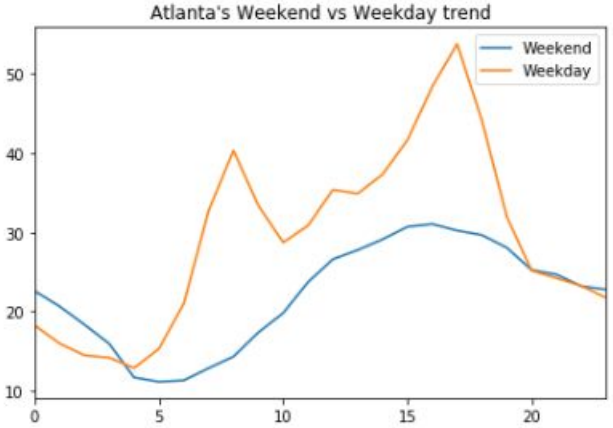


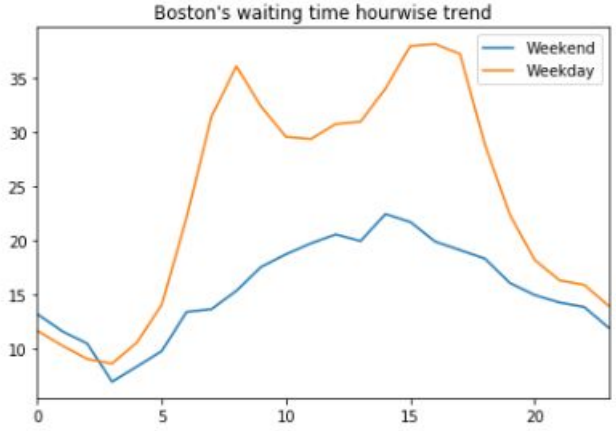


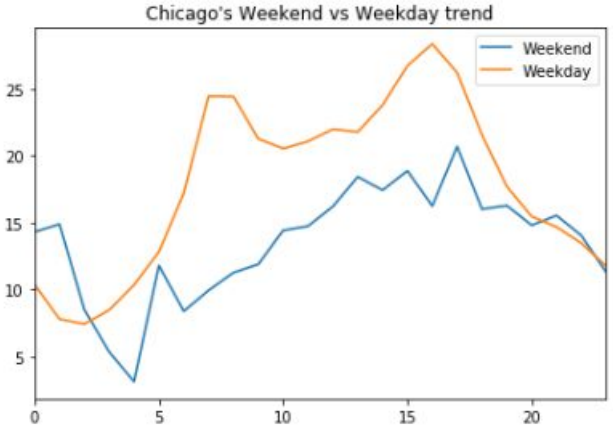


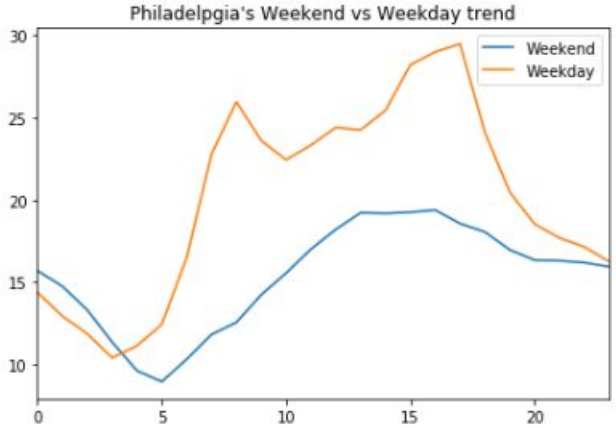
- Similar Patterns are seen for the Waiting Time
- Atlanta's Waiting Time dip at 10 and peak at 17 seem very sharp when compared to other cities
- When compared to overall Avg. Waiting time:
  - Atalanta and Boston seem to have a greater avg. waiting
  - Philadelphia and Chicago seem to have a lesser avg. waiting

Exploratory Data Analysis and Insights - Weekend/Weekday Trends









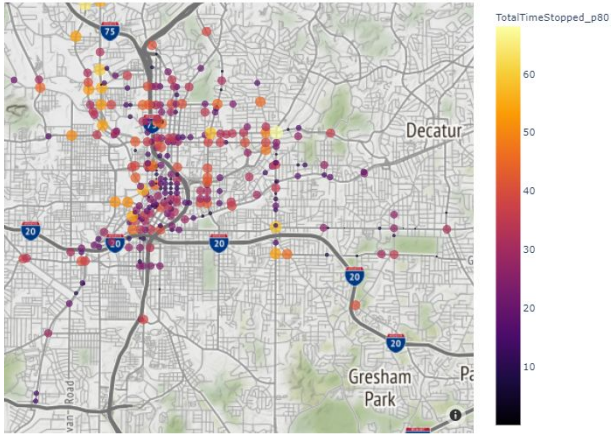
- Weekday Waiting time considerably higher than Weekend across all 4 cities

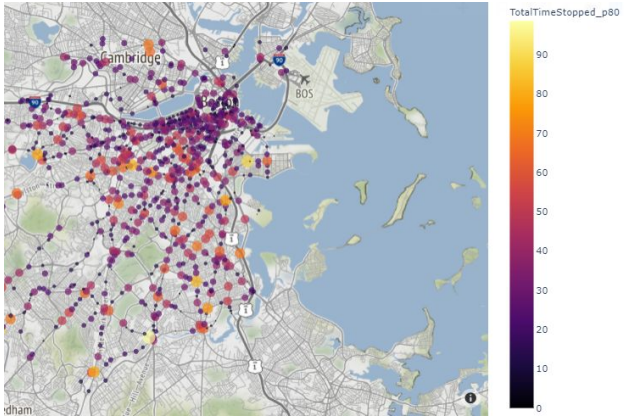
- The disparity is wider in Boston and Philadelphia than Atlanta and Chicago

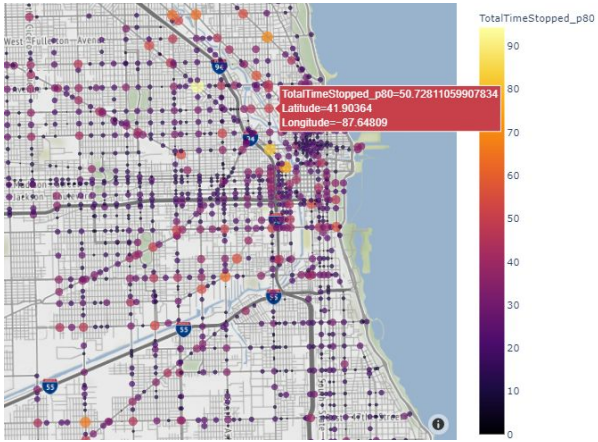
- Afternoon Lean evident during the weekday

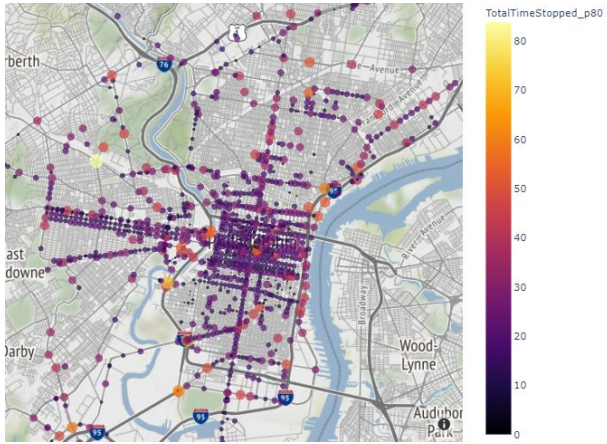
- Weekend trends peak late in the day but the peak continues during the day until late evening

Heat Maps



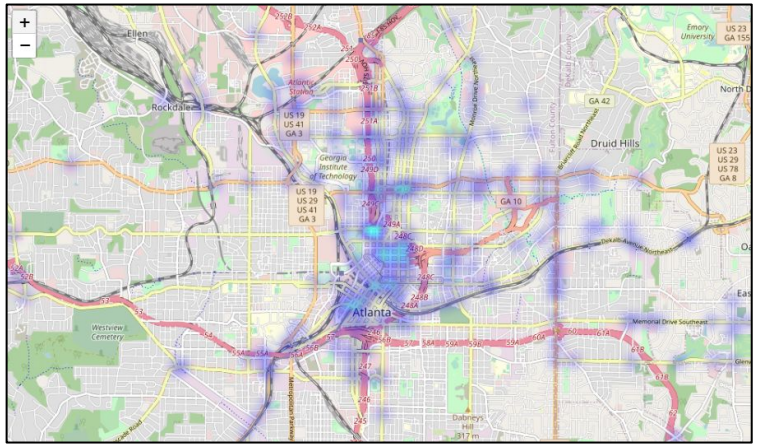


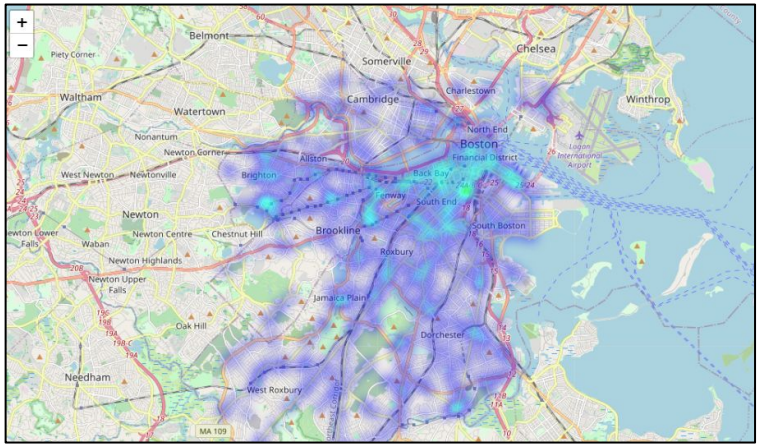


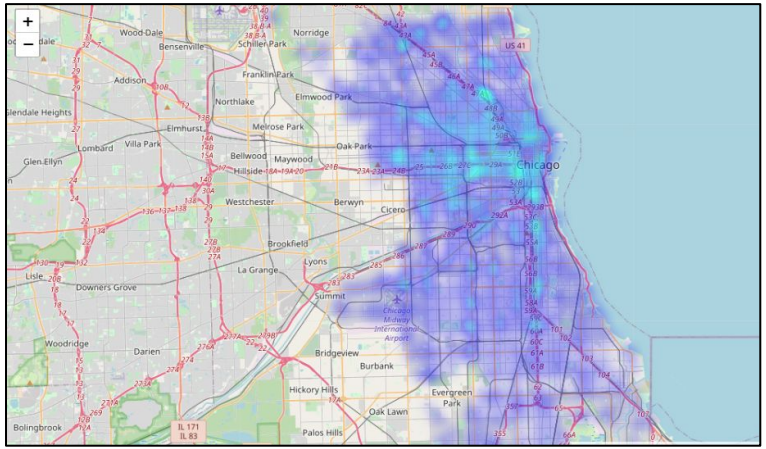


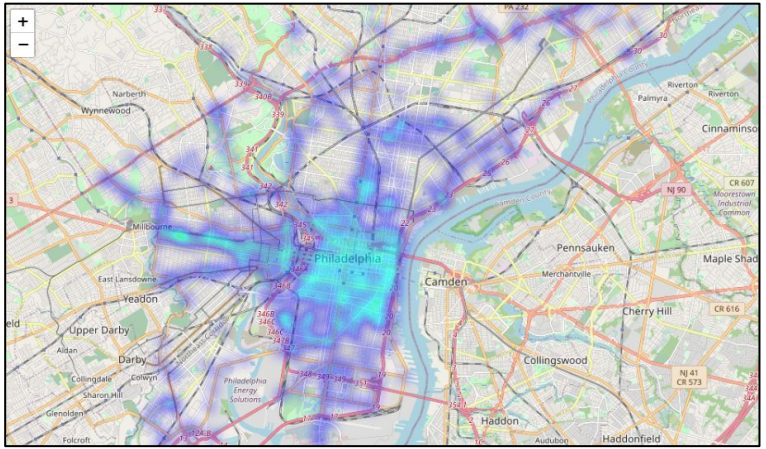
- High Wait Time near the Peach Tree Center for Atlanta
- High Wait Time near the West Ohio Street for Boston
- Highest Waiting time near 30 th Street for Philadelphia
- The variance in Atlanta sees less with a lot of yellow dots
- Philadelphia has a lot of junctions close by with relatively smaller time stopped
- The order of the maps is Atlanta, Boston Chicago, Philadelphia

More Heat Maps

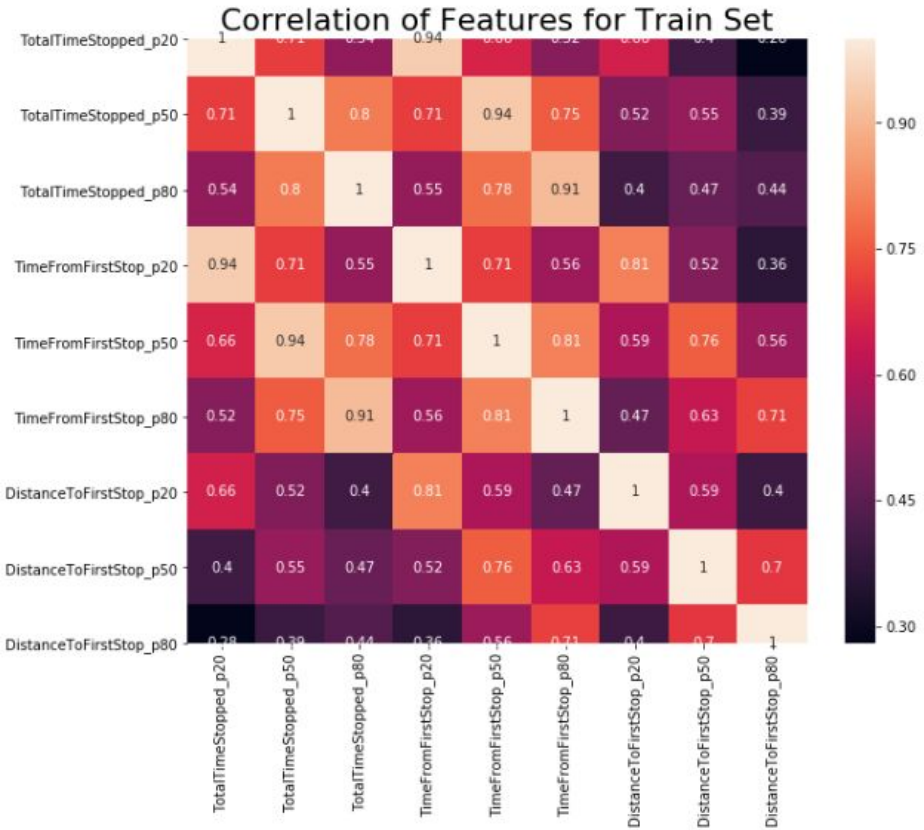


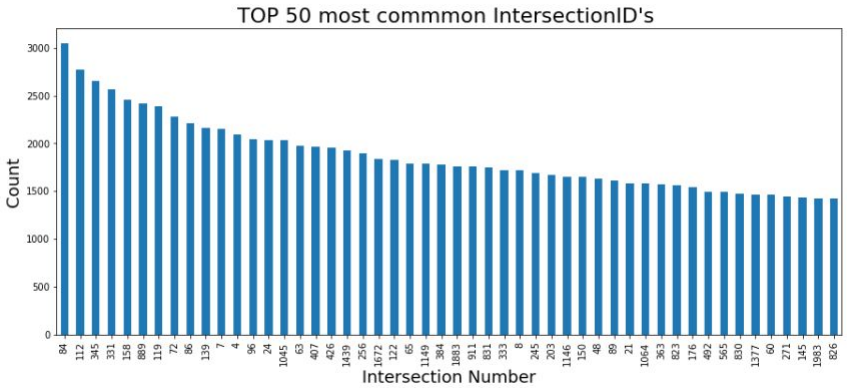






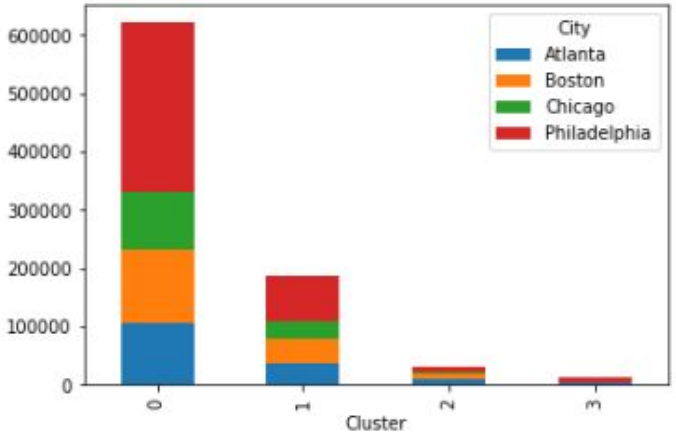
Correlation and Top Junctions

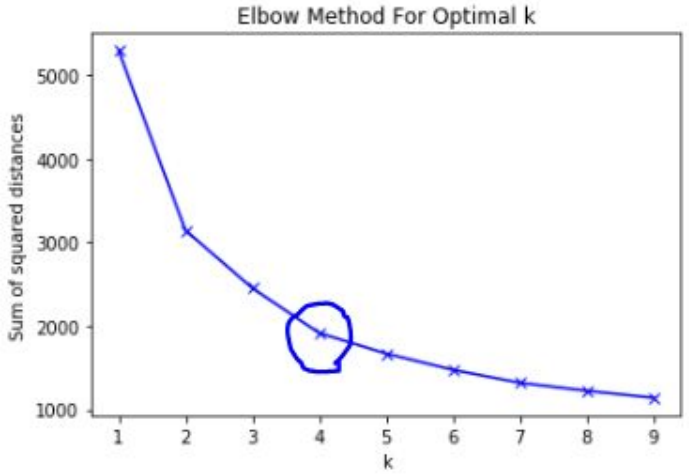


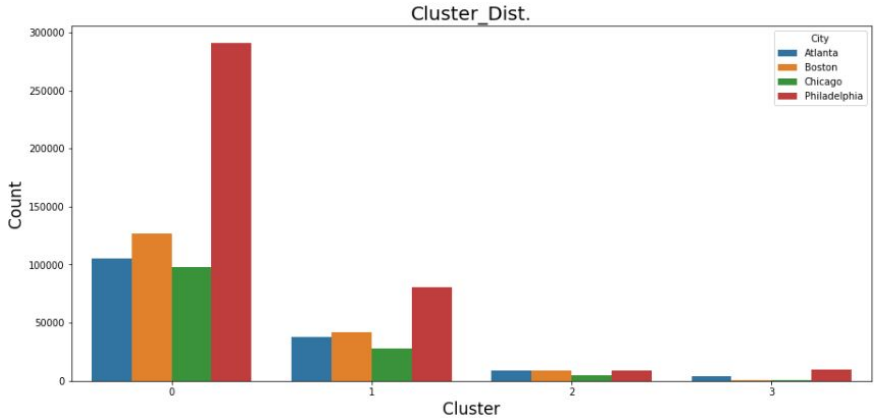


- Multicollinearity is seen here in some variables
- Does removing correlated variables improve the model? Seen in the following slides..

Clustering

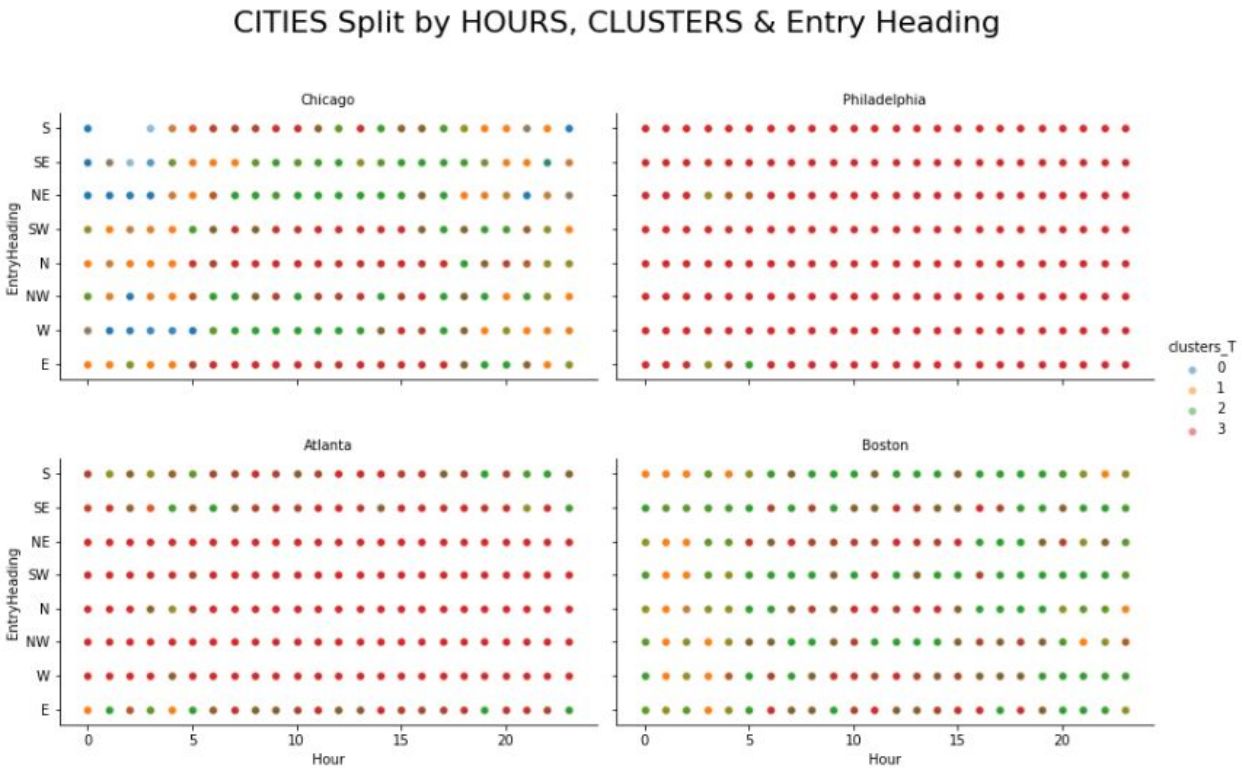






- TotalTimeStopped and DistanceToFirstStop are used for Clustering
- Elbow method signified 4 clusters to be optimal

Clustering



- TotalTimeStopped and DistanceToFirstStop are used for Clustering
- Elbow method signified 4 clusters to be optimal

Modeling

Dependent Variable:

- Total_Time_Stopped at an intersection

Predictors:

- Entry & Exit direction (Dummies)
- Weekend Flag
- Total Time Stopped
- Distance to First stop

Model Used:

Random Forest Regressor

```
labels1 = np.array(b_train1['TotalTimeStopped_p80'])
features1= b_train1.drop('TotalTimeStopped_p80', axis = 1)
feature_list1 = list(features1.columns)
features1 = np.array(features1)

test_labels1 = np.array(b_test1['TotalTimeStopped_p80'])
test_features1= b_test1.drop('TotalTimeStopped_p80', axis = 1)


rf2 = RandomForestRegressor(n_estimators = 10, random_state = 42)
rf2.fit(features1, labels1)
test_labels1 = np.array(b_test1['TotalTimeStopped_p80'])
test_features1= b_test1.drop('TotalTimeStopped_p80', axis = 1)
predictions2 = rf2.predict(test_features1)
errors2 = abs(predictions2 - test_labels1)
print('Mean Absolute Error:', round(np.mean(errors2), 2), 'degrees.')
RMSE2 = np.sqrt(((predictions2 - test_labels1) ** 2).mean())
RMSE2
```

```
Mean Absolute Error: 1.13 degrees.

2.9600191415592496
```

Modeling

City	Root Mean Squared Error (in Secs.)
Atlanta	3.048489
Boston	2.960019
Chicago	2.447817
Philadelphia	2.419955

Multicollinearity:

Removing some highly correlated variables increased the RMSE and decreased the output of the Random Forest

Sample Random Forest

```
'RandomForest': [RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
                       max_features='auto', max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=10,
                       n_jobs=None, oob_score=False, random_state=42, verbose=0,
                       warm_start=False),
```

Questions?

Thank You!