

CSC5800 - Class Project Report

Vehicular Queue Length Classification from High Resolution Stop Bar Detection Data

Rajratna Patil and Tony Geara

Abstract:

This project uses high resolution (HiRes) data obtained from a real-life City of Detroit Intersection and attempts to predict the classification under which the vehicle queue for that approach falls (No Queue, Low Queue, or High Queue). This helps the decision makers in understanding the performance of the signal and potentially adjust traffic signal timing (cycle and phase times) in order to improve signal performance. The high resolution data is collected from a single source 360° fisheye camera providing high-accuracy stop bar pulse and presence data paired with high resolution traffic signal telemetry (phase status and timings).

An extensive data extraction procedure was conducted through APIs, data was converted into usable formats, sorted and tabulated, and finally processed under multiple detailed tables in order to make reflective of the conditions it was measuring. Outliers were eliminated, and a final per-cycle amalgamation of a series of different data features were combined into a single data set for statistical evaluation and model development. The Final data set consisted of 21 features and one classifier.

This project attempted to be a first step in that direction by processing a one-day long set of data for an approach at an intersection and deriving the statistical relationships between the chosen attributes. Additionally, various prediction classification models were test to provide a comparison (Logistic Regression, Decision Tree, Linear Discriminant Analysis, K Nearest Neighbors, Gaussian Naïve Bayes, SVM, and NN -Multilayer Perceptron).

Based on the statistical analyses performed, we reached interesting inferences regarding intersection queue formation and the potential implication on traffic signal performance. For instance, some of the investigated analyses: highly correlated continuous input Attributes, distribution pattern for attributes, models that predict our objective accurately, the decision boundaries, and the Receiver operating characteristic curve (ROC) for two class-problem.

Based on the classification prediction models developed we managed to predict the three classes of Queue Status – No Queue (NQ=1), Light Queue (LQ=2), Full Queue (FQ=3) with varying degrees of success for a balanced accuracy. Out of the tested methods, **K nearest neighbor (KNN) seems to be outperforming the other methods** (even the Neural Network based Multi-Level Perceptron). However, due to a ramification of an uneven class distribution (with much lower instances of FQ compared with the other two classes), potential data accuracy issues, review of analysis methods and configurations, and the need for a larger data pool, the results were not as strongly along the entire data range. Prediction of the FQ class specifically fell short in all methods and needs further investigation. However, for the NQ and LQ, the model seems to be highly accurate.

I. INTRODUCTION

Understanding traffic signal performance (TSP) has been one of traffic engineering's most challenging problems. In order to do so, traffic engineers need to have a number of detectors set up at various locations of the intersection in order to completely and accurately translate real-life conditions into quantifiable and useful metrics. Data is then converted into Traffic Signal Performance Measures (TSPM) to directly enhance the operational performance and to indirectly improve upon the safety of a traffic signal. The TSPMs could be used either for real-time traffic operations or for after-the-fact evaluation or assessment of the performance of a traffic signal. Traffic operations is usually assessed by the common Level-of-Service ranking based on intersection or approach delay. However, in recent years Purdue University, Iowa State, and Utah DOT amongst others have been developing and innovating with automating the process of those TSPM in order to make them practical and useful. In addition to the basic performance measures of Approach Volume and turning movement counts, there's been a number of more advanced Automated Traffic Signal Performance Measures (ATSPM) that have been developed such as the Purdue Phase Determination, Split Monitoring, Pedestrian Delay, Preemption, Split Failures, Approach Delay, and Arrivals on Red. However, there are a few others that are a little more involved that haven't been as popular such as Yellow-and-Red actuation, Headway, and Speed.

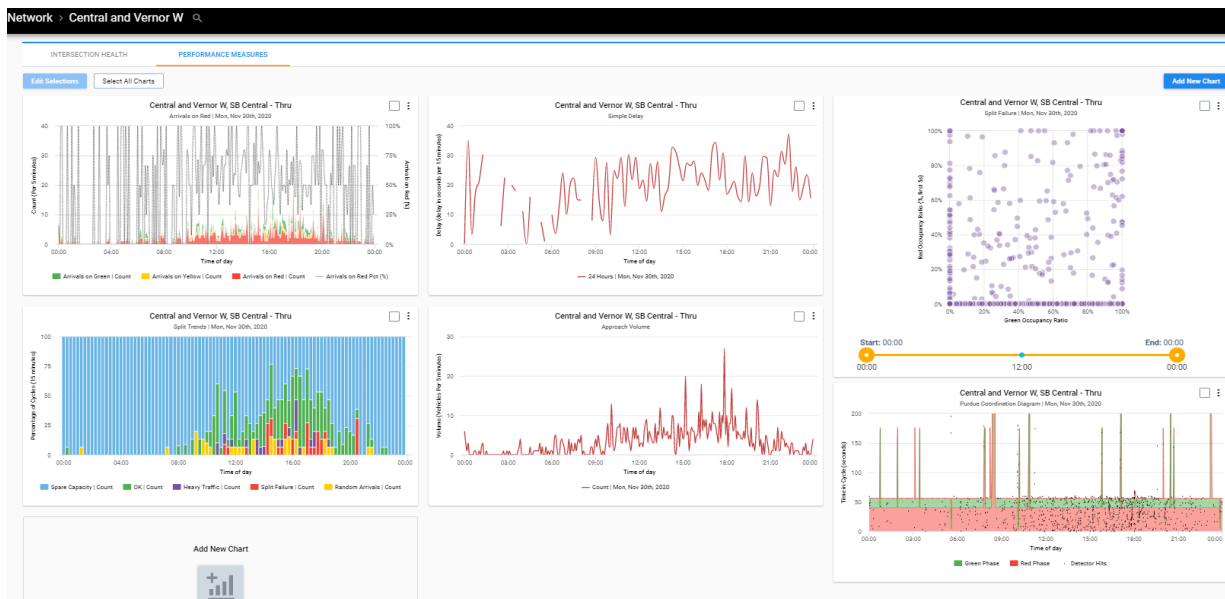


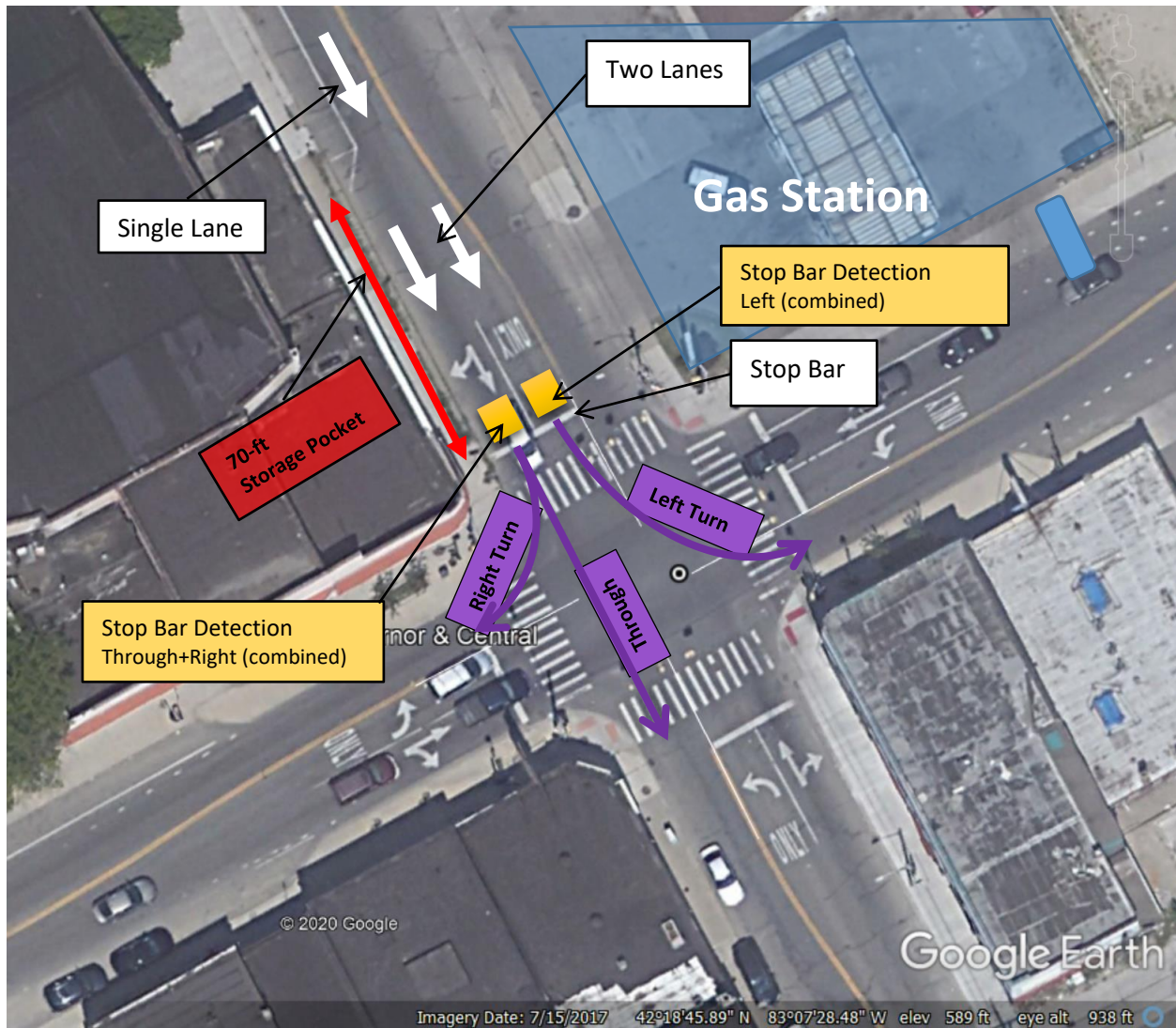
Figure: Miovision provided ATSPM's (Central & Vernor)

As powerful as those measures are, they come at a very hefty resource investment and cost for installation, operations, and maintenance in order to achieve their accuracy levels. In order to predict all those TSPM, at a minimum a set of two detectors are needed per approach and then connected to a data collection device in the cabinet. Many agencies may not have the financial means to initiate such a solution regardless of the benefit it may bring. However, stop-bar detection alone is more attainable and cheaper for smaller agencies. Those could be in the form of in-pavement detection pucks, loops, or single source 360° fisheye cameras and they provide high-accuracy stop bar pulse and presence data. This project aims at using high resolution (HiRes) stop bar detection data alone to try identifying features with the highest correlation to queue and which could potentially predict the queue with the highest accuracy (without the hefty investment in infrastructure).

a. Chosen location, configuration, and timing

The intersection location of Central Ave and Vernor Hwy in the southwest side of Detroit was chosen as the test location. This intersection was chosen for its typical 4-leg configuration and two-phase signal timing. In order to simplify the evaluation development process, one approach was selected: Southbound Central Avenue approach. This approach

The 60-second pre-timed cycle with a 40/20 Red/Green split. i.e.: for the southbound approach, every 60 second cycle consisted of 40 seconds of red and 20 seconds of green. This pre-timed cycle information makes it easier to identify missing information and timing errors.



The lane configurations of this approach consist of one approach lane that branches out into two lanes about 70 feet from the stop bar. In the most ideal of situations, this allows for approximately a storage pocket capacity of two (2) vehicles for the left turn lane (permissive movement) and five (5) vehicles for the through+right turn lane. In order for the approach to be considered blocked (maximum queue), there would need to be between 5 to 7 vehicles in the storage pockets. That is dependent on which movement lane the cars are stacking up in.

To add more complexity and potential error to the configuration, there is a gas station in the northeast quadrant of the intersection which has vehicles accessing it from Central Avenue and potentially disrupting typical flow and operation of the signal.

b. Detection and Data Accuracy

For this project a single source 360° fisheye camera providing high-accuracy stop bar pulse and presence data was used paired with high resolution traffic signal telemetry (phase status and timings). Additionally, the Camera capabilities were used to get accurate vehicle type classification (car, heavy, pedestrian, or bicycle) and Turning Movement Count (TMC) data in order to be used for the analyses and for the actual queue calculations used in training the model. The team attempted to use an advance detection box for this purpose however, the camera was exceeding its detection ranges and the data was inaccurate and unusable.

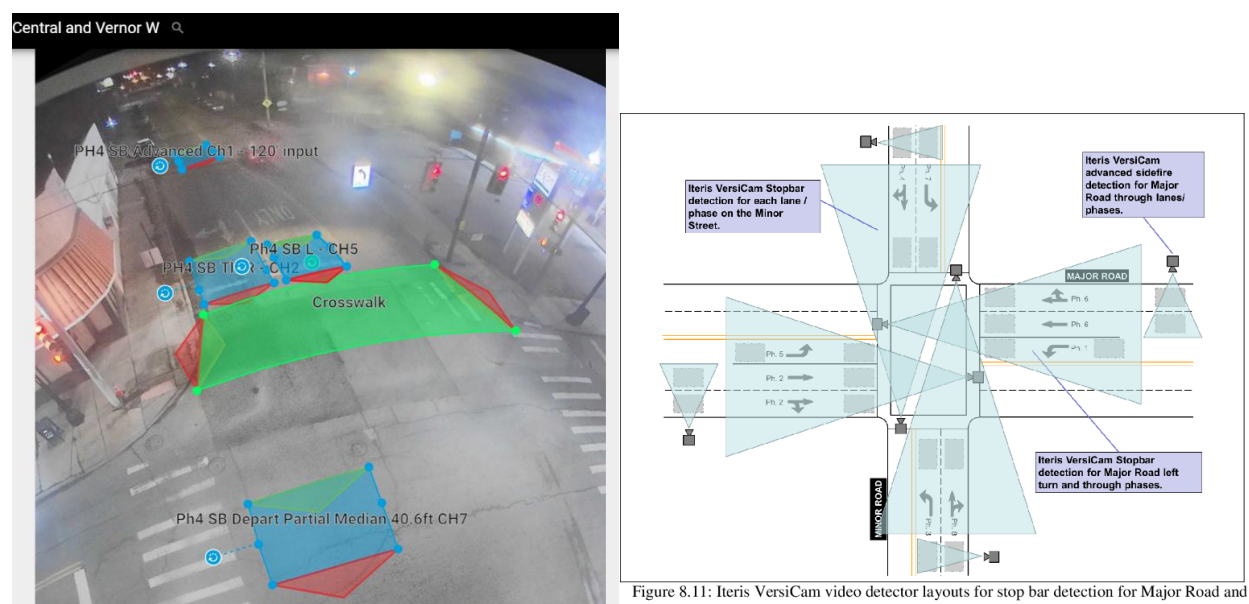


Figure: Single Camera Detection Box Configuration (left) vs. full lane-specific detection configuration (right)

Limited stop-bar detection, allows the traffic engineer to get a sense of what's happening at the stop-bar of the signal. When paired with traffic signal telemetry (signal phase and time) data, a better replication of the signal is possible. And just like with forensic evidence, the pieces of the puzzle are put together to try to come up with the intended queue prediction. The concern here is the limited view of the intersection that this approach starts with and the consequential limitation in accuracy. In order to offset those it was necessary to interpret the raw high resolution detector data and pair it with the high resolution signal telemetry to better tell the story of the arrival and departure of the vehicles. C

c. Problem statement (Research Question)

- How can we leverage existing HiRes stop bar video detection in the City of Detroit to predict performance (queueing TG1) at Urban intersections?
- If so, how accurately can that be done relative to existing state of the practice metrics?

- iii. Are there other existing data that can be leveraged to improve the accuracy of the performance metrics[TG2]?

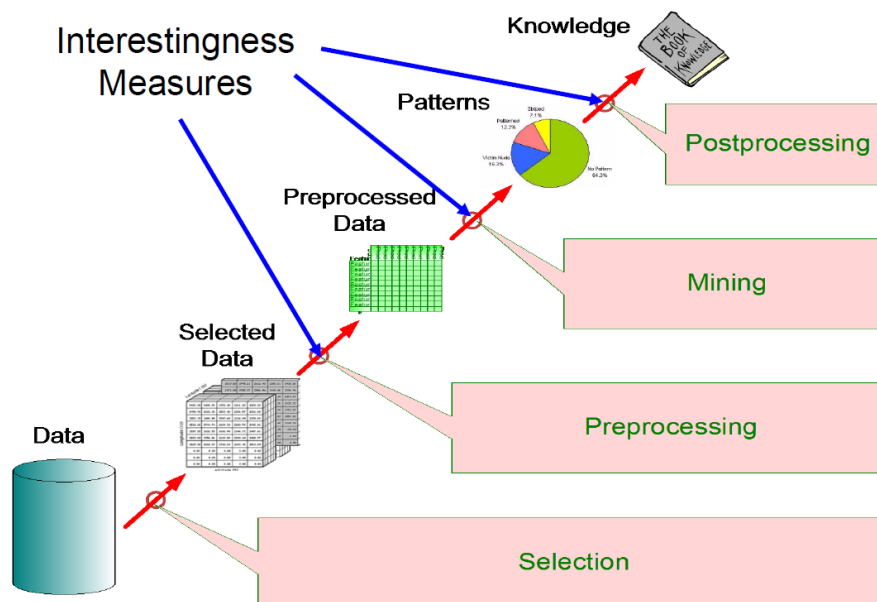
d. Objective statement

- i. Collect and Visualize existing High Resolution Signal Controller data at an intersections in the city of Detroit.
- ii. Attempt to create a Methodology to define the performance measure logic using stop bar detection
- iii. Compare performance measures using advanced detection (real world) with performance measures using stop bar detection
- iv. Allow method to be able to Scale and measure the performance difference between advanced and stop bar logic.

II. METHODOLOGY

The entire data development process is described below in bullet points up to the data analysis and modeling section. The entire process was very cumbersome as it accounted for a full effort to select the data, extract it, preprocess it, process it to make it usable and relevant, and organize the tables with usable attributes. The data set was preprocessed using excel, however, it is recommended for this process to be automated using python in the future.

Application of Interestingness Measure



Source: Steven F. Ashby Center for Applied Scientific Computing Month DD, 1997 (Lecture Notes – Introduction to Data Mining)

- Data Extracted from Miovision API –
 - November 30th was chosen
 - Day within the last month with highest southbound movement and overall signal volume for the day and per hour to account for worst case scenario.
 - High Resolution (HiRes) Data Extracted:
 - Stop-Bar Detection data
 - Advance Pulse-Box Detection data
 - Turning Movement Counts data (TMC)
 - Aggregated 1-minute Data Extracted”
 - Turning Movement Counts
- Account for Timestamp time in extraction (based on UTC time for the HiRes data and local time for the TMC) – HiRes had to be converted to local time.
- Extracted HiRes files were a series of 1-hr groups in JSON format that needed conversion and joining
- JSON converted to .txt files then worked with as .csv
- Timestamp converted to regular time with millisecond precision (functional time accuracy is to 100 millisecond [i.e.: 0.1 seconds])
- Data was checked for errors and gaps (missing data)
- A queue estimation model was built from advanced and stop-bar detection boxes (paired with the traffic signal status [red or green]). However, error was high.
 - Alternatively, a queue estimation model was used from the HiRes TMC data.
- Visualization:
 - HiRes Stop Bar Detection data was graphed for visual inspection of the arrivals on Red and Green.
- Recreated Stop-Bar Detection, signal status (and time left on red or green), headway (closeness of vehicles), speed, and historical TMC (some of those items were not accurate enough and were not used in the model).
- Model – Queue prediction (classification) discussed next section

III. ANALYSIS (entire workbook can be found on GitHub)

<https://github.com/rajratnapatil9/TRAFFIC-QUEUE-PREDICTION/blob/main/Queue%20Prediction.ipynb>

DATA DIMENSION: 1343 x 19 (final set)

a. Schema:

List of Attributes and classifier are listed below with a short description and an example. Outliers and erroneous data points were eliminated from the original data set based on missing and off cycles. That reduced the data points from 1440 to 1343 over two steps.

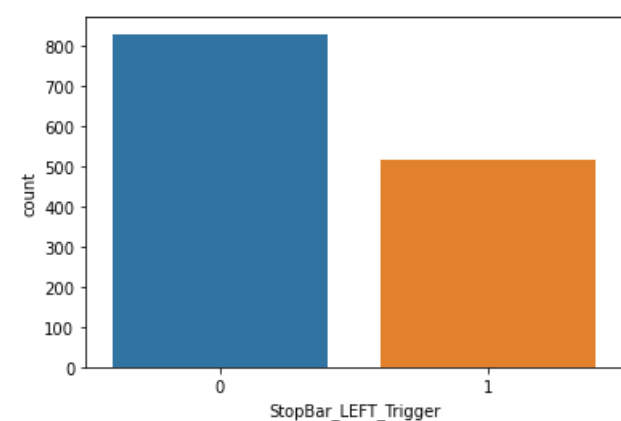
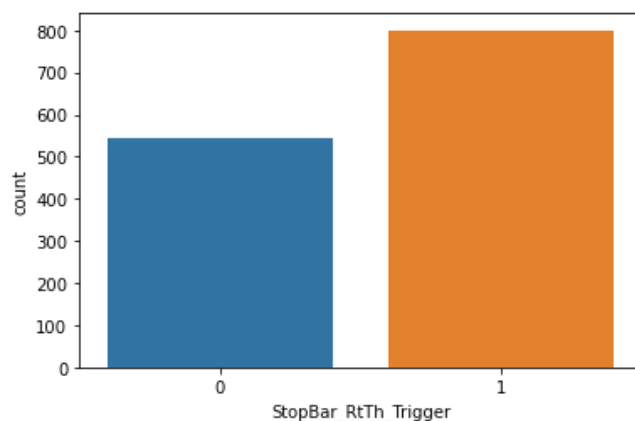
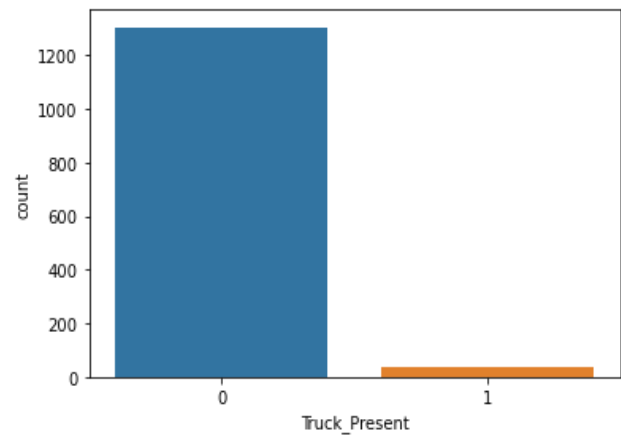
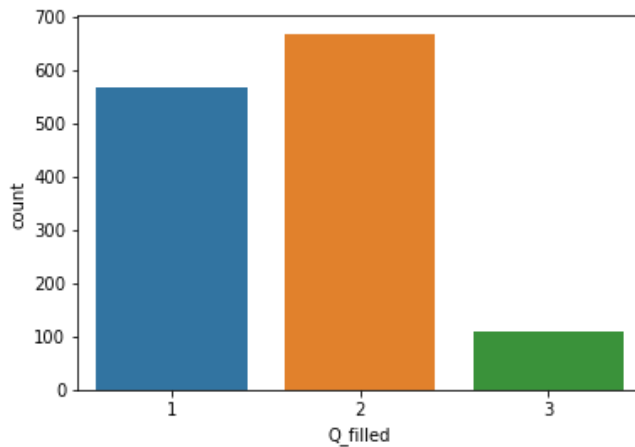
#	Attributes	Description	Example
1	Hour	Hour during which the data point occurred	10
2	minutes	Minute during which the data point occurred	20
3	H_M	Hour+Minute (fraction) during which the data point occurred	10.33
4	Cycl_Length	Cycle length in Seconds (to the 1/10th of a second)	59.97
5	SGN_Red_Time	Red Time in Seconds (to the 1/10th of a second)	40.55
6	SGN_Green_Time	Green Time in Seconds (to the 1/10th of a second)	19.42
7	StopBar_LEFT_Trigger	Trigger of detector box at stop bar during the cycle for the Left Turn lane	1
8	Freq. AOR by Cycle	Number of Vehicles Arriving on Red during Cycle (Left Turn)	1
9	Freq. AOG by Cycle	Number of Vehicles Arriving on Green during Cycle (Left Turn)	0
10	Freq. DOR by Cycle	Number of Vehicles Departing on Red during Cycle (Left Turn)	0
11	Freq. DOG by Cycle	Number of Vehicles Departing on Green during Cycle (Left Turn)	1
12	StopBar_RtTh_Trigger	Trigger of detector box at stop bar during the cycle for the Through+Right Turn lane	1
13	RTOR Depart (assumption)	Number of Vehicles Turning Right on Red (RTOR) during Cycle (Th+Rt lane)	0
14	Freq. AOR by Cycle	Number of Vehicles Arriving on Red during Cycle (Th+Rt lane)	1
15	Freq. AOG by Cycle	Number of Vehicles Arriving on Green during Cycle (Th+Rt lane)	0
16	Freq. DOR by Cycle	Number of Vehicles Departing on Red during Cycle (Th+Rt lane)	0
17	Freq. DOG by Cycle	Number of Vehicles Departing on Green during Cycle (Th+Rt lane)	1
18	Truck_Present	Presence of trucks during cycle	0
19	Q_filled	Classification: NQ, LQ, FQ (see provided table)	LQ

b. Distribution of Variables: NQ=1 , LQ=2, FQ= 3

Queue Classification Category	NQ	LQ	FQ
Category Numerical Code	1	2	3
Category Values	0	1-5	6 and above
Description	No cars during cycle	1 to 5 vehicles queued during cycle	6 or more vehicles queued during cycle
frequency	570	677	111

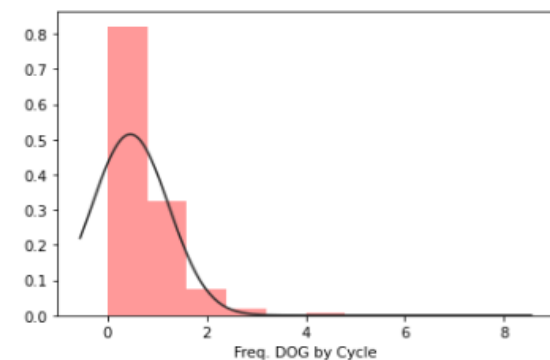
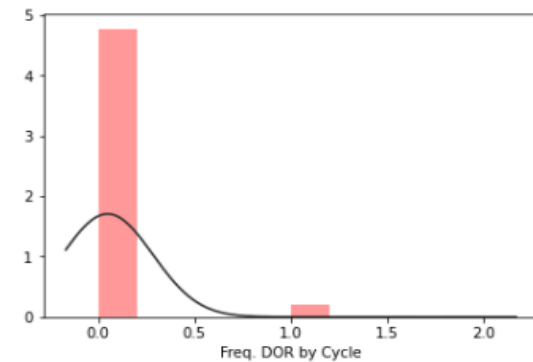
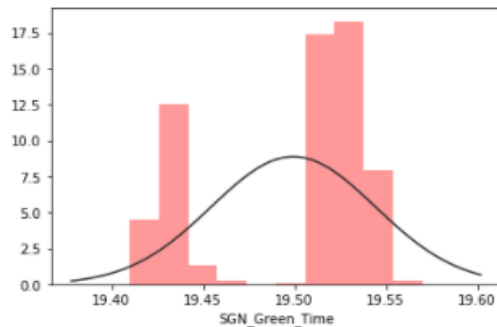
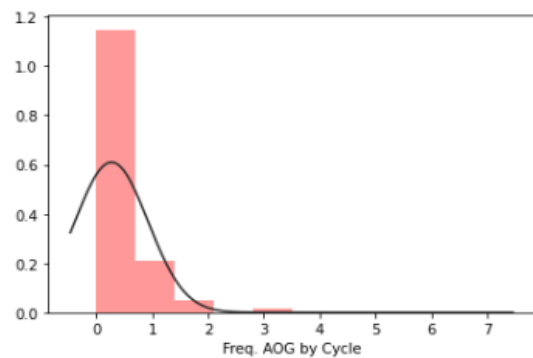
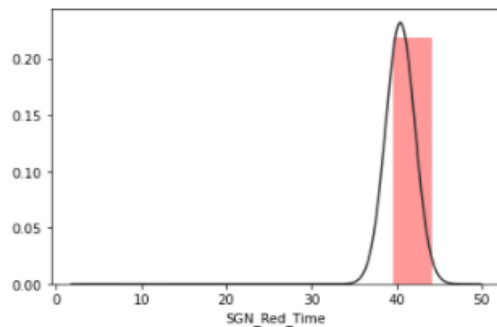
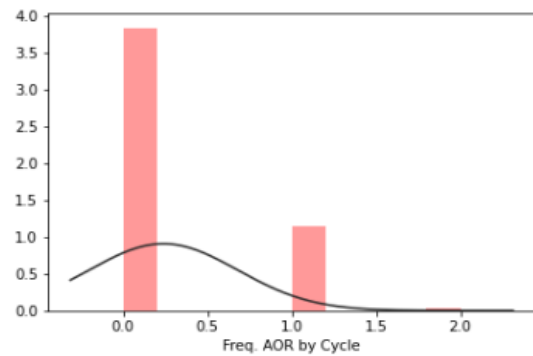
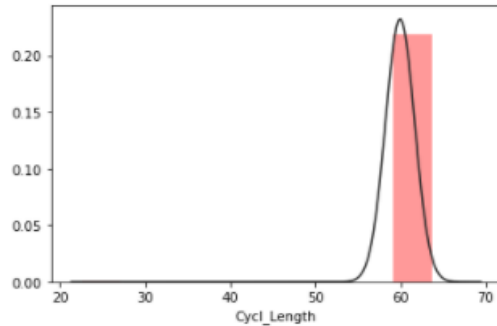
The derived actual queue lengths were categorized into three classes: NQ, LQ, and FQ (described above). Some basic descriptive statistics were performed to attempt to identify the class range, however, it was more crucial to follow a more practical approach that would yield useful results. The frequency of FQ is lower than LQ and NQ for our dataset. But we can't avoid this, as we need this information to accurately predict to improve traffic signal systems. NQ was used for when there are no queues present. FQ is used when the queue is expected to be full (backed up), and the LQ is the level incorporating everything in between. And because this is a sample data set, we didn't have a big data representation for the FQ level. However, with more days and locations to review in the future, it is expected that this category would be better represented.

MaxQ_NotNeeded	
Mean	1.865243
Standard Error	0.058582
Median	1
Mode	0
Standard Deviation	2.158826
Sample Variance	4.66053
Kurtosis	0.039275
Skewness	1.010565
Range	8
Minimum	0
Maximum	8
Sum	2533
Count	1358
Confidence Level(95.0%)	0.114922



c. Histograms of select Independent Variables:

The following represent a set of histograms for a select number of independent variables.



Uniformly Distributed (expected consistency)

Cycl_length

SGN_Red_Time

SGN_Green_Time

Non-uniformly Distributed (more random in nature)

Freq- AOR, AOG, DOR, DOG

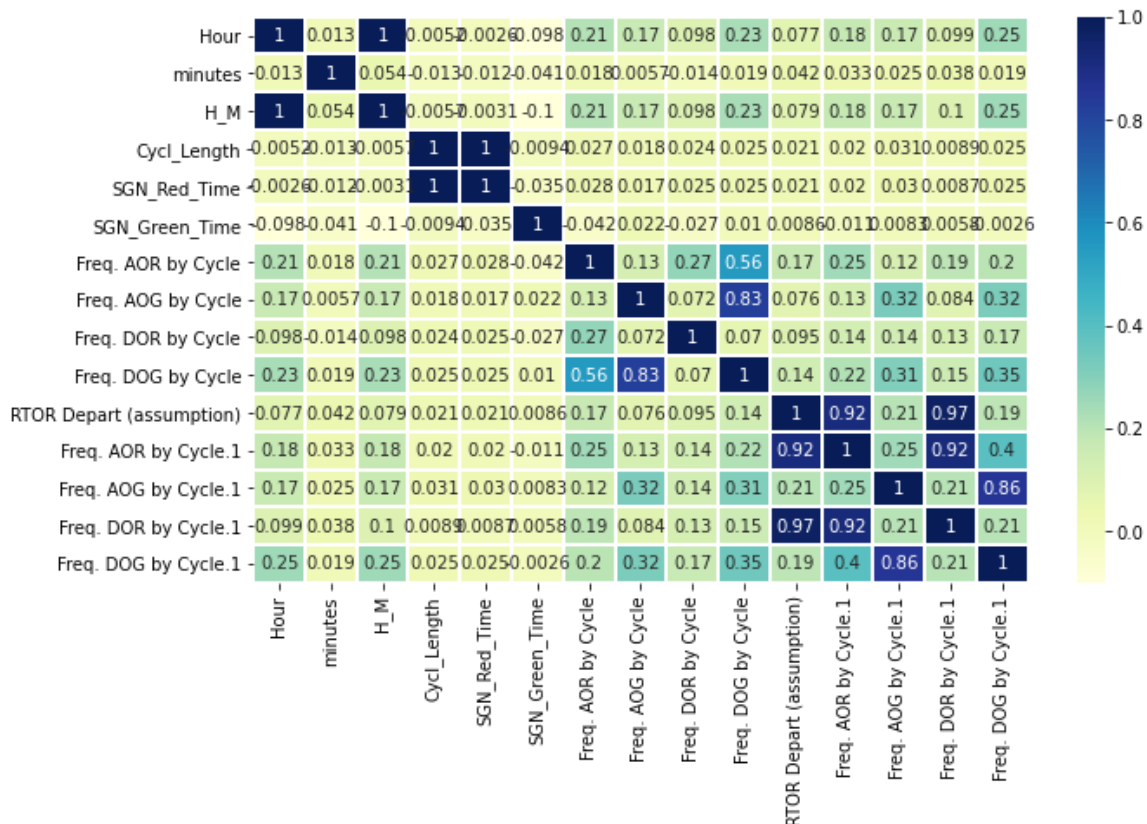
d. Correlations Analysis

According to the correlation coefficient matrix, there were six highly correlated attributes that stood out between. The most significant of which is: RTOR Depart (assumption) : Frequency DOR by Cycle.1 (0.97). This correlation is justified when the practical explanation of the Attribute is delved into. In this case the RTOR was derived from any DOR that occurs after 3 seconds. Additionally, RTOR seems to be occurring in correlation with vehicles arriving on red and departing on red (DOR).

Additionally, the data is indicating that most vehicles are arriving on green and departing on green for both lanes (that is a good value but could be improved with coordination with signals upstream). However, the final relationship is showing a high correlation between vehicles arriving on red (AOR) and departing on green (DOG) which is not a good indication of signal performance.

ATTRIBUTES	CORRELATION
RTOR Depart (assumption) : Frequency DOR by Cycle.1	0.97
RTOR Depart (assumption) : Frequency AOR by Cycle.1	0.92
Freq. AOR by Cycle.1 : Frequency DOR by Cycle.1	0.92
Freq. DOG by Cycle.1 : Freq. AOG by Cycle.1	0.86
Freq. DOG by Cycle : Freq. AOG by Cycle	0.83
Freq. DOG by Cycle: Freq. AOR by Cycle	0.56

The below correlation coefficient matrix represents the full relationship between the various continuous attributes. Categorical variables are not included.



e. MODEL BUILDING:

For Model Building, as our data is non-linear, we used these classification methods to search for accurate model. Seven models were investigated. The accuracy between the training and testing data sets for the following seven models is shown in the following table.

1. Logistic Regression

5. Gaussian Naïve Bayes

2. Decision Tree

6. SVM

3. Linear Discriminant Analysis

7. NN -Multilayer Perceptron

4. K nearest Neighbors

MODEL	TRAINING ACCURACY	TESTING ACCURACY
1. Logistic Regression	0.82	0.83
2. Decision Tree	1	0.75
3. Linear Discriminant Analysis	0.81	0.82
4. K Nearest Neighbors	0.85	0.82
5. Gaussian Naïve Bayes	0.52	0.52
6. SVM	0.85	0.85
7. NN -Multilayer Perceptron	0.85	0.85

f. Comparing the top performing classifiers

If we compare only the testing accuracies, then we have the following Top-3 Classifiers:

- **SVM**
- **NN-Perceptron**
- **KNN-K-Nearest Neighbors**

Inferences:-

- If we compare other measures like confusion metrics, F-measure, Precision and Recall, then we see that KNN wins
- Also if we closely observe, KNN gives us a balanced classification accuracy, for instance it predicts class FQ=3 more accurately than other two classifiers.

```

print('Confusion Matrix and Precision-Recall for SVM')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred3 = svm.predict(X_test)
print(confusion_matrix(y_test, pred3))
print(classification_report(y_test, pred3))

```

```

Confusion Matrix and Precision-Recall for SVM
[[160  8  0]
 [ 22 180  1]
 [  2  27  3]]
              precision    recall  f1-score   support

     1         0.87        0.95        0.91         168
     2         0.84        0.89        0.86         203
     3         0.75        0.09        0.17          32

 accuracy          0.85         0.85         0.85         403
 macro avg         0.82        0.64        0.65         403
 weighted avg         0.84        0.85        0.83         403

```

```

#MLP
print('Confusion Matrix and Precision-Recall for MLP')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred = clf.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))

```

```

Confusion Matrix and Precision-Recall for MLP
[[160  8  0]
 [ 22 179  2]
 [  0  28  4]]
              precision    recall  f1-score   support

     1         0.88        0.95        0.91         168
     2         0.83        0.88        0.86         203
     3         0.67        0.12        0.21          32

 accuracy          0.85         0.85         0.85         403
 macro avg         0.79        0.65        0.66         403
 weighted avg         0.84        0.85        0.83         403

```

```

#KNN
print('Confusion Matrix and Precision-Recall for KNN')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred2 = knn.predict(X_test)
print(confusion_matrix(y_test, pred2))
print(classification_report(y_test, pred2))

```

```

Confusion Matrix and Precision-Recall for KNN
[[150  18  0]
 [ 21 174  8]
 [  0  24  8]]
              precision    recall  f1-score   support

     1         0.88        0.89        0.88         168
     2         0.81        0.86        0.83         203
     3         0.50        0.25        0.33          32

 accuracy          0.82         0.82         0.82         403
 macro avg         0.73        0.67        0.68         403
 weighted avg         0.81        0.82        0.81         403

```

	Correctly classified SVM	Correctly classified MLP	Correctly classified KNN
NQ=1	160	160	150
LQ=2	180	179	174
FQ=3	3	4	8

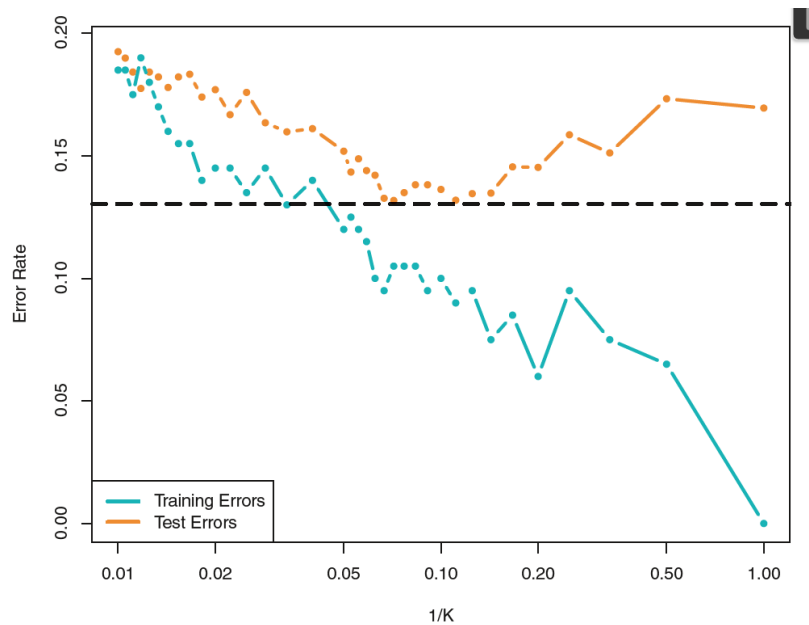
As we can see, although Accuracy of other models are equivalent for NQ=1 and LQ=2, Class FQ=3 is lagging behind with only a fraction of the data points correctly classified. Out of the three methods, FQ is more accurately classified under KNN (8 out of 32) which is a very important factor for Traffic signal regularization.

- **K Nearest Neighbors ****

We used $k=5$ because we felt it optimized the confusion matrix better for our test results

Theoretically: At $K=5$, most case scenario observes a min Training and testing error rates. At $K=5$, $1/K = 0.2$

PG 42 of Textbook - An Introduction to Statistical Learning with Applications in R



Practically:

After plotting the graph we observed that at $K=5$, both Training and testing errors are minimum.

```

neighbors = np.arange(1,20)
train_accuracy = np.empty(len(neighbors))
test_accuracy = np.empty(len(neighbors))

for i,k in enumerate(neighbors):
    #Setup a knn classifier with k neighbors
    knn = KNeighborsClassifier(n_neighbors=k)

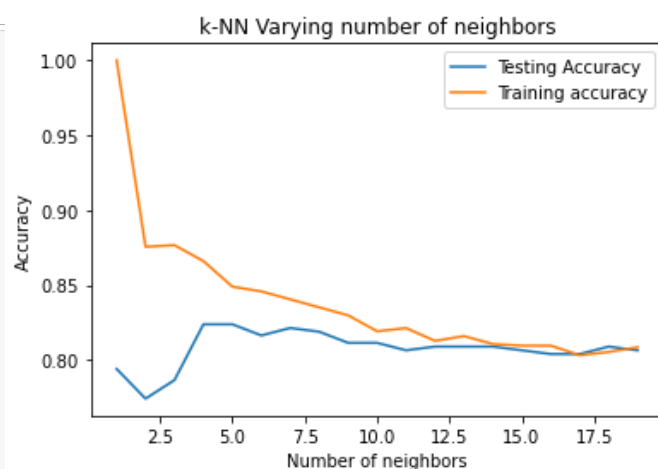
    #Fit the model
    knn.fit(X_train, y_train)

    #Compute accuracy on the training set
    train_accuracy[i] = knn.score(X_train, y_train)

    #Compute accuracy on the test set
    test_accuracy[i] = knn.score(X_test, y_test)

plt.title('k-NN Varying number of neighbors')
plt.plot(neighbors, test_accuracy, label='Testing Accuracy')
plt.plot(neighbors, train_accuracy, label='Training accuracy')
plt.legend()
plt.xlabel('Number of neighbors')
plt.ylabel('Accuracy')
plt.show()

```



- **ROC-CURVE**

From ROC curve, we see for LQ=2 the KNN model is a very good model (above 50% chance) but for FQ=3, the KNN model current is a bad model (below 50%). But it is much better and balanced than any other model tried so far.

In this case, the penalization of one class prediction vs the other is subject to the objective and the cost function associated with it. This signifies that both of them are penalizing each other and tradeoff between these need to be decided based on the objective of the analysis. For our Analysis, we want both of these to be predicted accurately.

```
# 2-class classification

from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score

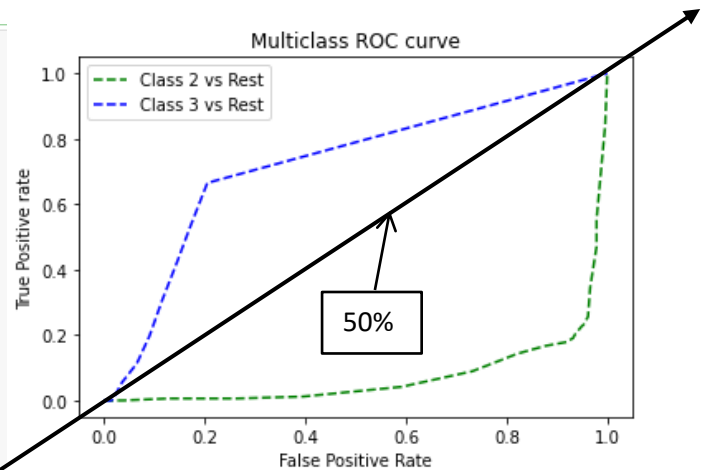
pred = knn.predict(X_test)
pred_prob = knn.predict_proba(X_test)

# roc curve for classes
fpr = {}
tpr = {}
thresh = {}

n_class = 3

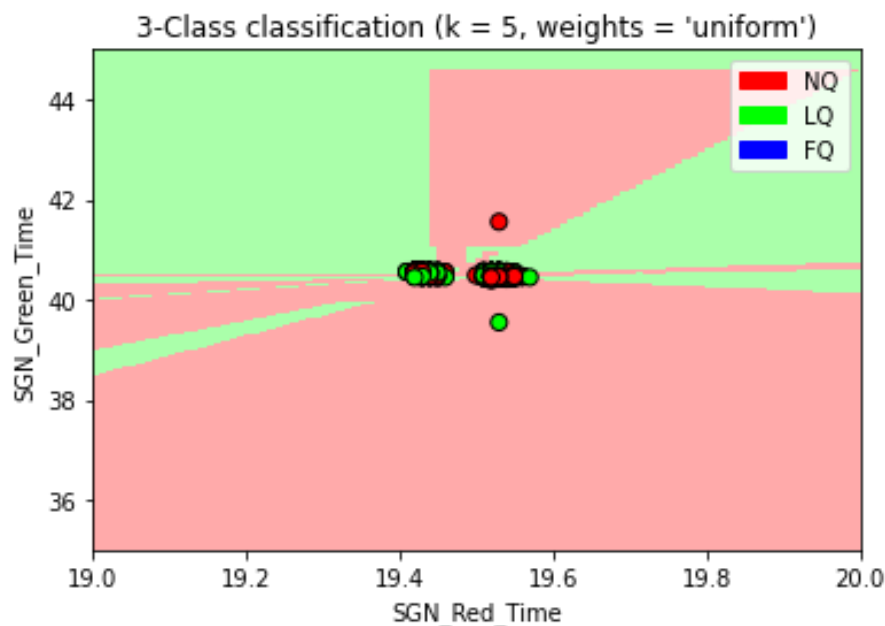
for i in range(1,n_class):
    fpr[i], tpr[i], thresh[i] = roc_curve(y_test, pred_prob[:,i], pos_label=i)

# plotting
plt.plot(fpr[0], tpr[0], linestyle='--',color='orange', label='Class 1 vs Rest')
plt.plot(fpr[1], tpr[1], linestyle='--',color='green', label='Class 2 vs Rest')
plt.plot(fpr[2], tpr[2], linestyle='--',color='blue', label='Class 3 vs Rest')
plt.title('Multiclass ROC curve')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive rate')
plt.legend(loc='best')
plt.savefig('Multiclass ROC',dpi=300);
```



- **Decision Boundary**

For NQ=1, LQ=2, FQ=3, as we can see, a clear boundary for FQ is missing which needs to be improvised.



V. CONCLUSION

The issue of predicting queue classification solely from stop bar detection high resolution data is a hurdle that has not been addressed or resolved through research and industry due to the complexity it requires and the large set of data required to come to a reasonable selection process, method, and result interpretation. This project attempted to be a first step in that direction by processing a one-day long set of data for an approach at an intersection and deriving the statistical relationships between the chosen attributes. Additionally, various prediction classification models were test to provide a comparison (Logistic Regression, Decision Tree, Linear Discriminant Analysis, K Nearest Neighbors, Gaussian Naïve Bayes, SVM, and NN -Multilayer Perceptron).

Out of those methods, **K nearest neighbor (KNN) seems to be outperforming the other methods** (even the Neural Network based Multi-Level Perceptron). However, due to a ramification of an uneven class distribution (with much lower instances of FQ compared with the other two classes), potential data accuracy issues, review of analysis methods and configurations, and the need for a larger data pool, the results were not as strongly along the entire data range. Prediction of the FQ class specifically fell short in all methods and needs further investigation. However, for the NQ and LQ, the model seems to be highly accurate.

VI. REFERENCES

The following references were used throughout this report:

Model Analysis: <https://github.com/rajratnapatil9/TRAFFIC-QUEUE-PREDICTION/blob/main/Queue%20Prediction.ipynb>

PG 42 of Textbook - An Introduction to Statistical Learning with Applications in R
<http://faculty.marshall.usc.edu/gareth-james/ISL/>

Steven F. Ashby Center for Applied Scientific Computing Month DD, 1997 (Lecture Notes – Introduction to Data Mining)

Rahman, R., Hasan, S. Real-time signal queue length prediction using long short-term memory neural network. *Neural Comput & Applic* (2020). <https://doi.org/10.1007/s00521-020-05196-9>

Solving A Simple Classification Problem with Python — Fruits Lovers' Edition
<https://towardsdatascience.com/solving-a-simple-classification-problem-with-python-fruits-lovers-edition-d20ab6b071d2>

K-Nearest Neighbors, Boxplot, Standard Score, Feature Scaling, Curse of Dimensionality, Missing Values, Confusion Matrix, Classification Report, ROC-Curve, AUROC <https://limebit-jupyter-notebooks.s3.eu-central-1.amazonaws.com/jupyter-KNN-und-SVM.html>

AUC-ROC Curve in Machine Learning Clearly Explained
<https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/>


KNN for Classification using Scikit-learn <https://www.kaggle.com/amolbhivarkar/knn-for-classification-using-scikit-learn>

ROC Curve explained using a COVID-19 hypothetical example: Binary & Multi-Class Classification Tutorial
<https://towardsdatascience.com/roc-curve-explained-using-a-covid-19-hypothetical-example-binary-multi-class-classification-bab188ea869c>

🔗 main ▾

...

[TRAFFIC-QUEUE-PREDICTION](#) / Queue Prediction.ipynb

 **rajratnapatil9** Add files via upload 🕒 History

👤 1 contributor

<> 📄

RawBlame

🖥️ ✎️ 🗑️

1659 lines (1659 sloc) | 412 KB

```
In [2]: import numpy as np
import csv
import matplotlib.pyplot as plt
import pandas as pd
```

```
In [117]: data = pd.read_csv("Data_CSV.csv")
data.head()
```

Out[117]:

	Hour	minutes	H_M	Cycl_Length	SGN_Red_Time	SGN_Green_Time	StopBar_LEFT_Trig
0	0	0	0.00	59.97	40.45	19.52	0
1	0	1	0.02	59.98	40.46	19.52	1
2	0	2	0.03	60.00	40.46	19.54	1
3	0	3	0.05	59.97	40.44	19.52	1
4	0	4	0.07	59.97	40.45	19.52	1

```
In [3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1343 entries, 0 to 1342
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Hour                                  1343 non-null   int64
1   minutes                              1343 non-null   int64
2   H_M                                  1343 non-null   float64
3   Cycl_Length                          1343 non-null   float64
4   SGN_Red_Time                         1343 non-null   float64
5   SGN_Green_Time                       1343 non-null   float64
6   StopBar_LEFT_Trigger                 1343 non-null   int64
7   Freq. AOR by Cycle                   1343 non-null   int64
8   Freq. AOG by Cycle                   1343 non-null   int64
9   Freq. DOR by Cycle                   1343 non-null   int64
10  Freq. DOG by Cycle                   1343 non-null   int64
11  StopBar_RtTh_Trigger                 1343 non-null   int64
12  RTOR Depart (assumption)             1343 non-null   int64
13  Freq. AOR by Cycle.1                 1343 non-null   int64
14  Freq. AOG by Cycle.1                 1343 non-null   int64
15  Freq. DOR by Cycle.1                 1343 non-null   int64
16  Freq. DOG by Cycle.1                 1343 non-null   int64
17  Truck_Present                       1343 non-null   int64
18  Q_filled                             1343 non-null   int64
dtypes: float64(4), int64(15)
memory usage: 199.5 KB
```

```
In [118]: data['StopBar_LEFT_Trigger'] = data['StopBar_LEFT_Trigger'].astype('category')
data['StopBar_RtTh_Trigger']=data['StopBar_RtTh_Trigger'].astype('category')
data['Truck_Present']=data['Truck_Present'].astype('category')
data['Q_filled']=data['Q_filled'].astype('category')
data.info()
```



```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1343 entries, 0 to 1342
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Hour                                  1343 non-null   int64
1   minutes                             1343 non-null   int64
2   H_M                                  1343 non-null   float64
3   Cycl_Length                          1343 non-null   float64
4   SGN_Red_Time                         1343 non-null   float64
5   SGN_Green_Time                       1343 non-null   float64
6   StopBar_LEFT_Trigger                 1343 non-null   category
7   Freq. AOR by Cycle                   1343 non-null   int64
8   Freq. AOG by Cycle                   1343 non-null   int64
9   Freq. DOR by Cycle                   1343 non-null   int64
10  Freq. DOG by Cycle                   1343 non-null   int64
11  StopBar_RtTh_Trigger                 1343 non-null   category
12  RTOR Depart (assumption)             1343 non-null   int64
13  Freq. AOR by Cycle.1                 1343 non-null   int64
14  Freq. AOG by Cycle.1                 1343 non-null   int64
15  Freq. DOR by Cycle.1                 1343 non-null   int64
16  Freq. DOG by Cycle.1                 1343 non-null   int64
17  Truck_Present                       1343 non-null   category
18  Q_filled                             1343 non-null   category
dtypes: category(4), float64(4), int64(11)
memory usage: 163.1 KB
```

```
In [321]: (data['Q_filled']==1).sum()
```

```
Out[321]: 567
```

```
In [322]: (data['Q_filled']==2).sum()
```

```
Out[322]: 668
```

```
In [323]: (data['Q_filled']==3).sum()
```

```
Out[323]: 108
```

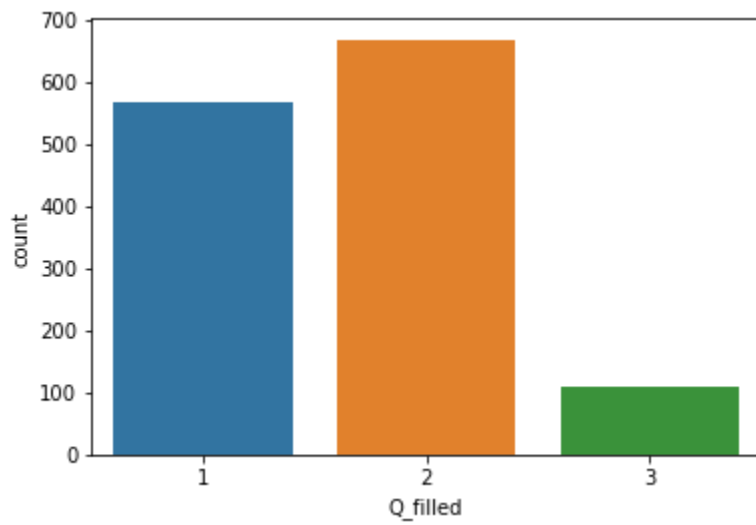
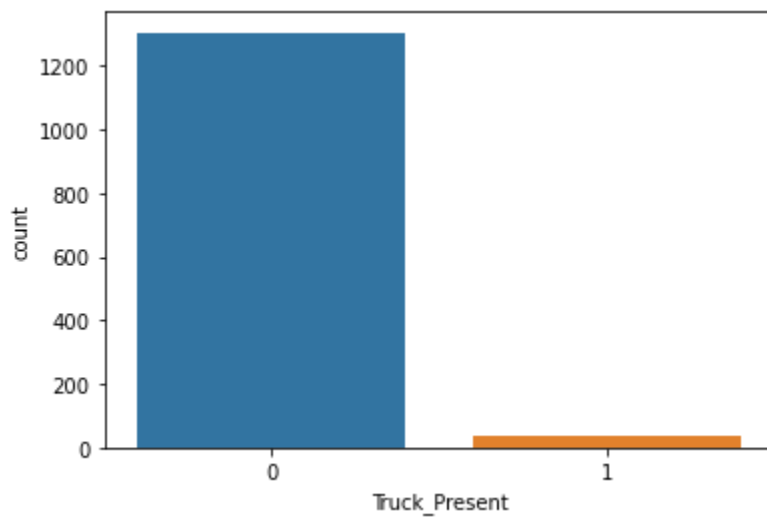
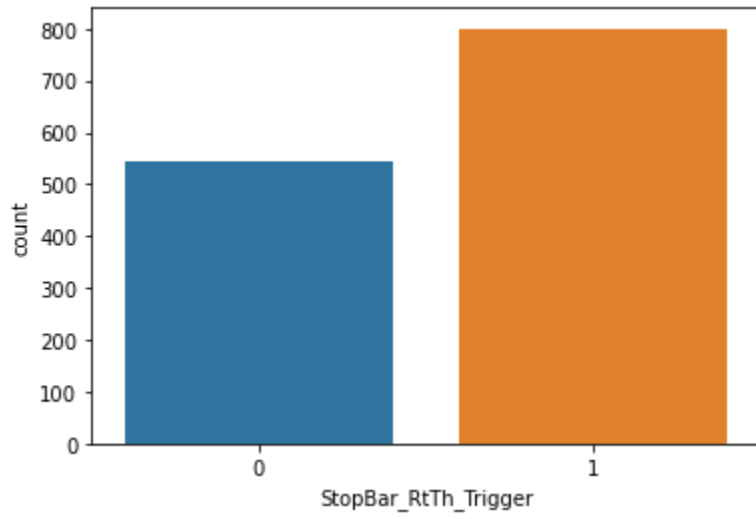
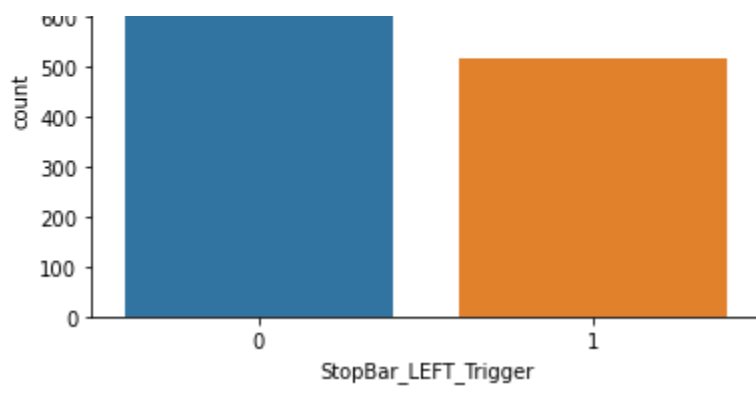
```
In [5]: import seaborn as sns
sns.countplot(data['StopBar_LEFT_Trigger'],label="Count of NT/T")
plt.show()

import seaborn as sns
sns.countplot(data['StopBar_RtTh_Trigger'],label="Count of NT/T")
plt.show()

import seaborn as sns
sns.countplot(data['Truck_Present'],label="Count of NT/T")
plt.show()

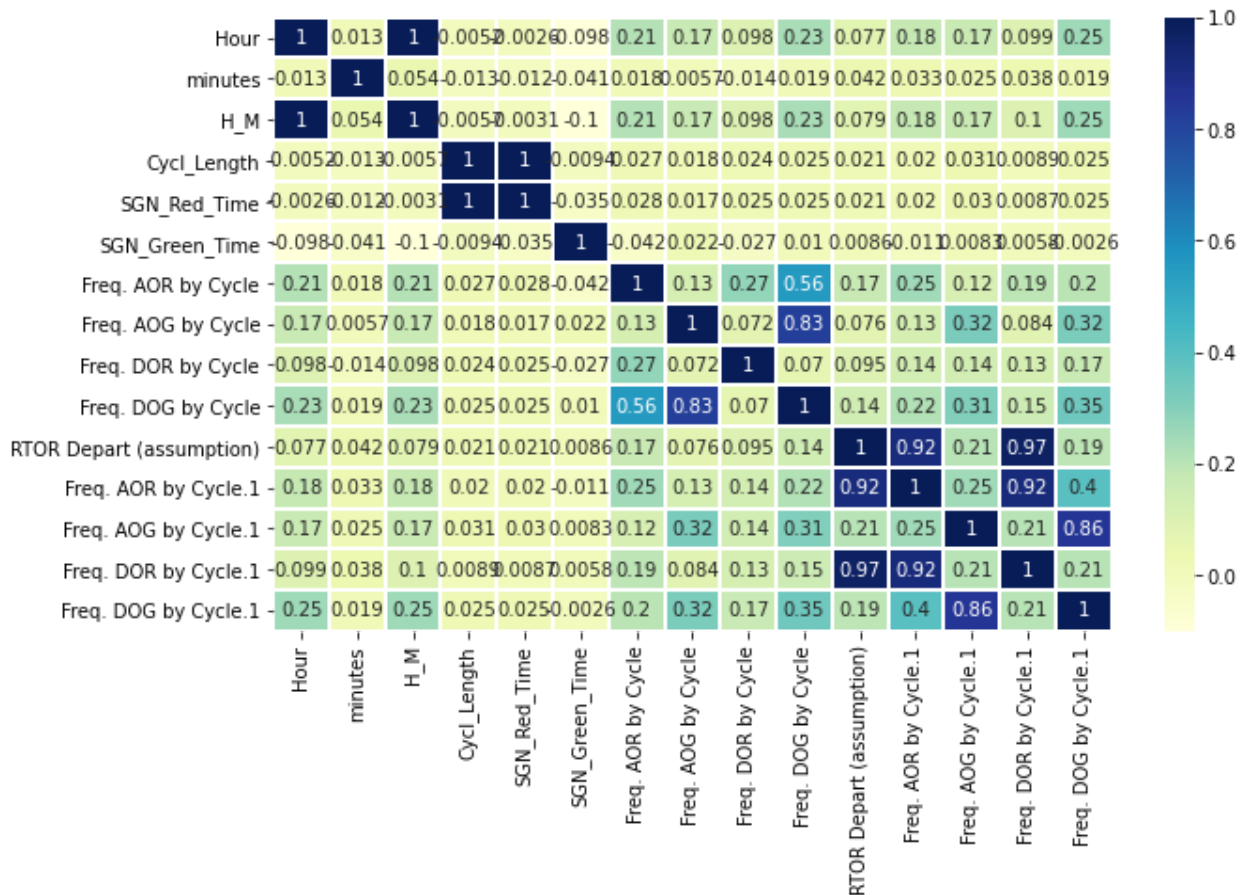
sns.countplot(data['Q_filled'],label="Count of NT/T")
plt.show()
```





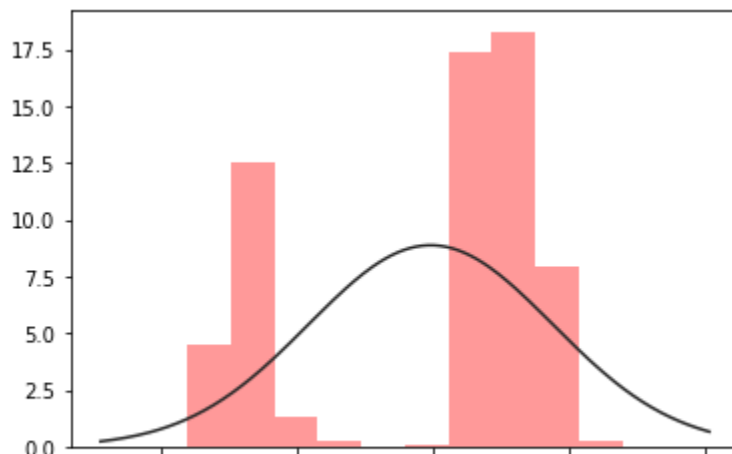
Correlation Plot for Continuous Attributes

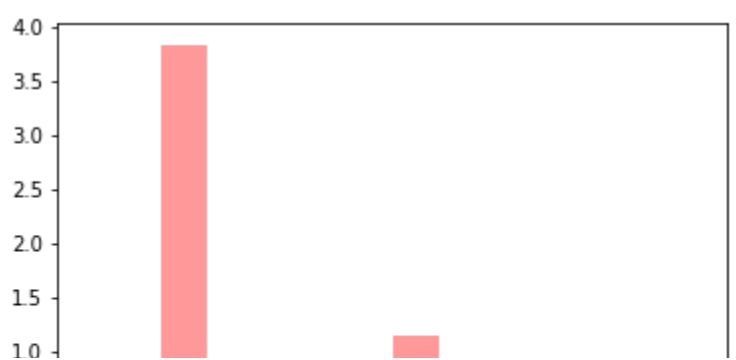
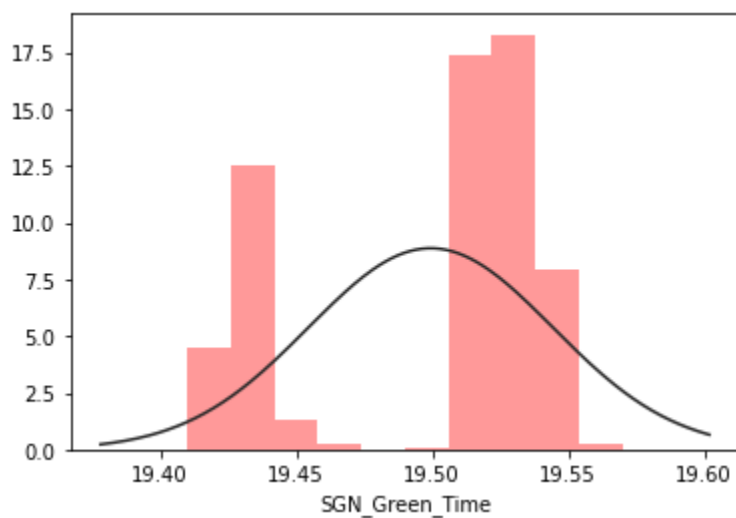
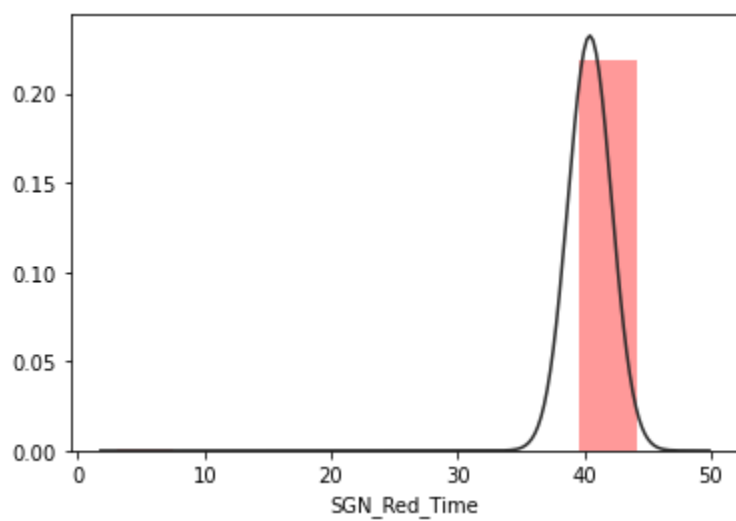
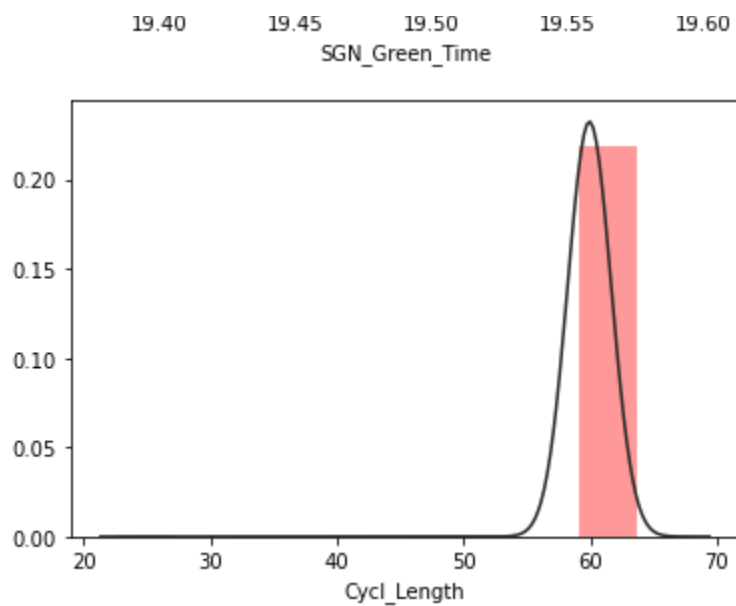
```
In [324]: plt.figure(figsize=(10,6))
sns.heatmap(data.corr(), annot=True, linecolor = 'white', linewidths = 1, cmap="YlGnBu")
plt.show()
```

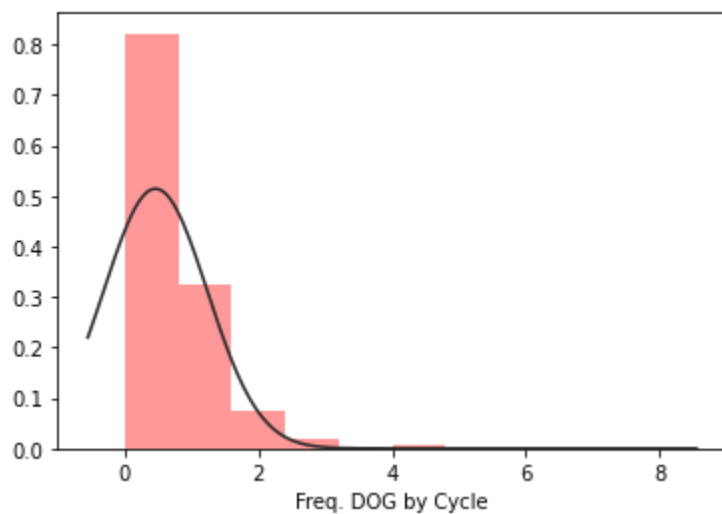
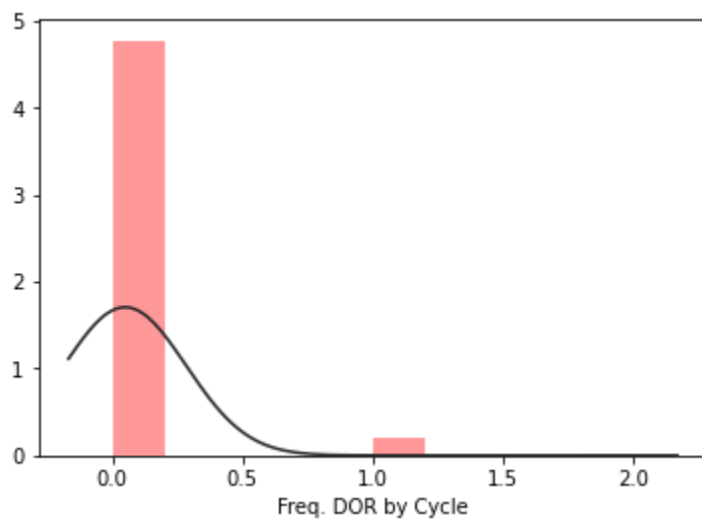
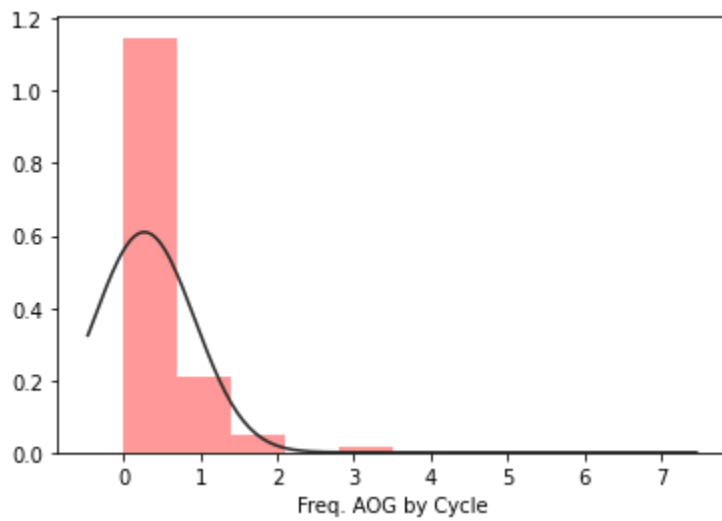
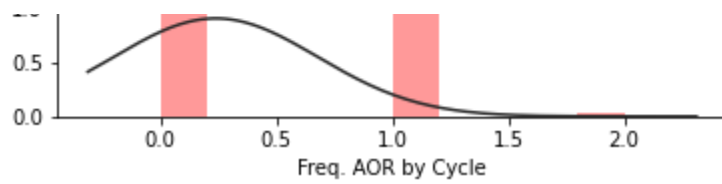


```
In [316]: sns.distplot(data.SGN_Green_Time, color = 'red', bins=10,fit=norm, kde=False)
plt.show()
```

```
sns.distplot(data.Cycl_Length, color = 'red', bins=10,fit=norm, kde=False)
plt.show()
sns.distplot(data.SGN_Red_Time, color = 'red', bins=10,fit=norm, kde=False)
plt.show()
sns.distplot(data.SGN_Green_Time, color = 'red', bins=10,fit=norm, kde=False)
plt.show()
```



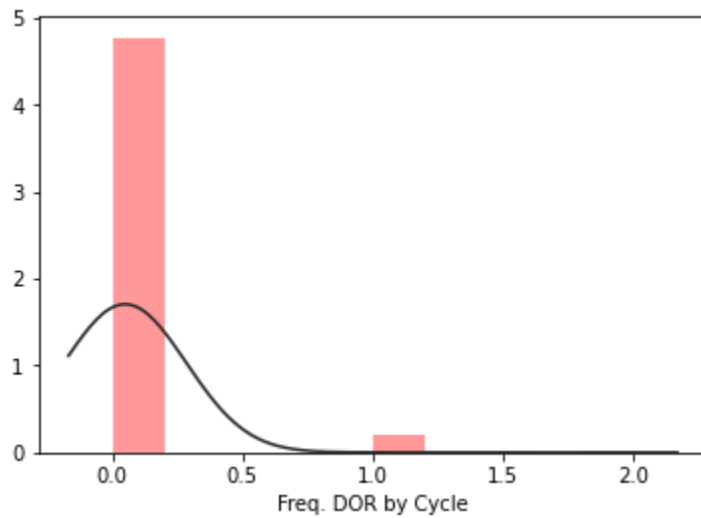
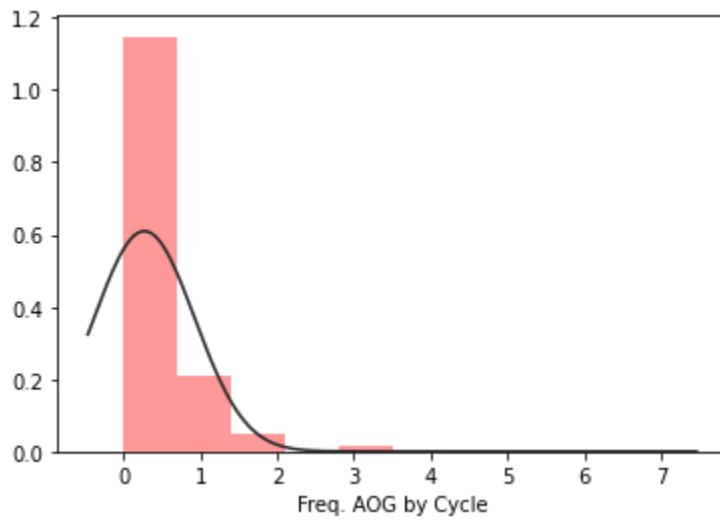
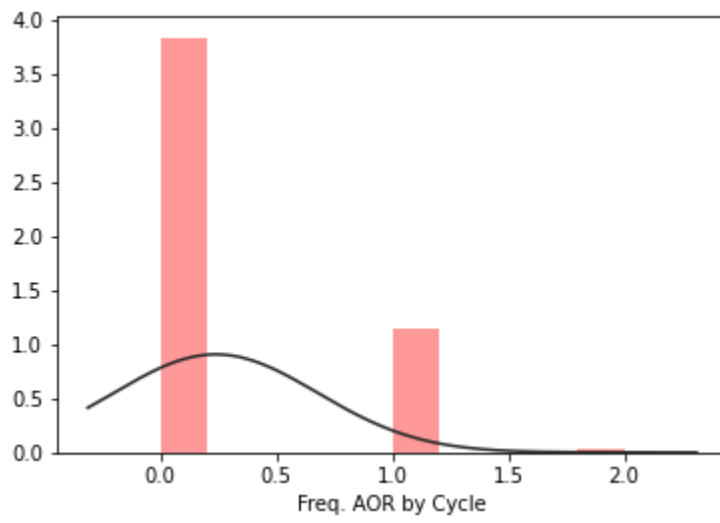


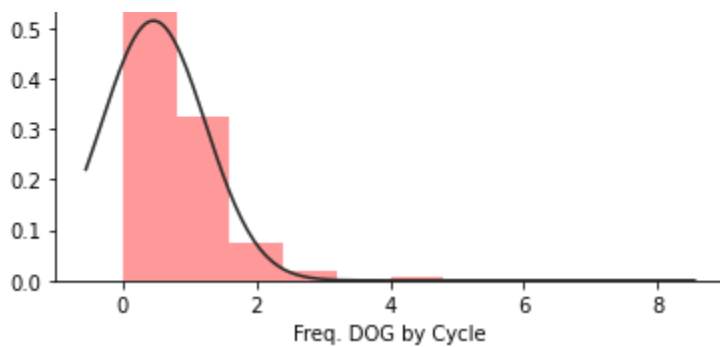


```
In [317]: sns.distplot(data['Freq. AOR by Cycle'], color = 'red', bins=10,fit=norm, kde=False)
plt.show()
sns.distplot(data['Freq. AOG by Cycle'], color = 'red', bins=10,fit=norm, kde=False)
plt.show()
```



```
sns.distplot(data['Freq. DOR by Cycle'], color = 'red', bins=10,fit=norm,kde=False)
plt.show()
sns.distplot(data['Freq. DOG by Cycle'], color = 'red', bins=10,fit=norm, kde=False)
plt.show()
```





1. Splitting data into continuous and categorical Attributes

2. Scaling and Normalizing COntinuous Attributes

3. Combining them into single dataset

```
In [119]: data.catX= data[['StopBar_LEFT_Trigger', 'StopBar_RtTh_Trigger' , 'Truck_Present'
, 'Q_filled']]
data.continuous = data[['Hour','minutes','H_M', 'Cycl_Length','SGN_Red_Time','SGN
_Green_Time', 'Freq. AOR by Cycle','Freq. AOG by Cycle' , 'Freq. DOR by Cycle','Fr
eq. DOG by Cycle', 'RTOR Depart (assumption)', 'Freq. AOR by Cycle.1','Freq. AOG
by Cycle.1', 'Freq. DOR by Cycle.1','Freq. DOG by Cycle.1']]
data.catX.head()
data.continuous.head()
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data.continuous = scaler.fit_transform(data.continuous)
data.continuous = scaler.transform(data.continuous)
data.continuous = pd.DataFrame(data=data.continuous, columns=['Hour','minutes','H
_M', 'Cycl_Length','SGN_Red_Time','SGN_Green_Time', 'Freq. AOR by Cycle','Freq. A
OG by Cycle' , 'Freq. DOR by Cycle','Freq. DOG by Cycle', 'RTOR Depart (assumptio
n)', 'Freq. AOR by Cycle.1','Freq. AOG by Cycle.1', 'Freq. DOR by Cycle.1','Freq.
DOG by Cycle.1'])
data.continuous.head()

data.norm=(data.continuous-data.continuous.min())/(data.continuous.max()-data.con
tinuous.min())
data.norm.head(2)

data1 = pd.concat([data.norm, data.catX],axis=1)
data1.head(2)
```

<ipython-input-119-270207f77d5e>:1: UserWarning: Pandas doesn't allow columns to be created via a new attribute name - see <https://pandas.pydata.org/pandas-docs/stable/indexing.html#attribute-access>

```
data.catX= data[['StopBar_LEFT_Trigger', 'StopBar_RtTh_Trigger' , 'Truck_Presen
t', 'Q_filled']]
```

<ipython-input-119-270207f77d5e>:2: UserWarning: Pandas doesn't allow columns to be created via a new attribute name - see <https://pandas.pydata.org/pandas-docs/stable/indexing.html#attribute-access>

```
data.continuous = data[['Hour','minutes','H_M', 'Cycl_Length','SGN_Red_Time','S
GN_Green_Time', 'Freq. AOR by Cycle','Freq. AOG by Cycle' , 'Freq. DOR by Cycl
e', 'Freq. DOG by Cycle', 'RTOR Depart (assumption)', 'Freq. AOR by Cycle.1','Fre
```

```
q. AOG by Cycle.1', 'Freq. DOR by Cycle.1', 'Freq. DOG by Cycle.1']]
```

```
<ipython-input-119-270207f77d5e>:12: UserWarning: Pandas doesn't allow columns to  
be created via a new attribute name - see https://pandas.pydata.org/pandas-docs/s  
table/indexing.html#attribute-access
```

```
data.norm=(data.continuous-data.continuous.min())/(data.continuous.max()-data.c  
ontinuous.min())
```

Out[119]:

	Hour	minutes	H_M	Cycl_Length	SGN_Red_Time	SGN_Green_Time	Freq. AOR by Cycle	Freq. AOG by Cycle
0	0.0	0.000000	0.000000	0.820153	0.820333	0.6875	0.0	0.00000
1	0.0	0.016949	0.000834	0.820372	0.820552	0.6875	0.0	0.14285

Splitting data into Training and Testing sets.

```
In [139]: from sklearn.model_selection import train_test_split
```

```
X= data1.iloc[0:,0:18]
```

```
y= data1["Q_filled"]
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random  
state=0)
```

```
X_train.head()
```

Out[139]:

	Hour	minutes	H_M	Cycl_Length	SGN_Red_Time	SGN_Green_Time	Freq. AOR by Cycle	F A b C
467	0.347826	0.118644	0.338757	0.820591	0.822743	0.1250	0.0	0
477	0.347826	0.593220	0.357947	0.820591	0.820771	0.7500	0.0	0
472	0.347826	0.254237	0.344180	0.820372	0.820552	0.6875	0.0	0
503	0.391304	0.050847	0.377555	0.820372	0.820552	0.6875	0.0	0
97	0.043478	0.677966	0.069670	0.820153	0.822524	0.1250	0.0	0

1. Logistic Regression

```
In [49]: from sklearn.linear_model import LogisticRegression
```

```
logreg = LogisticRegression()
```

```
logreg.fit(X_train, y_train)
```

```
print('Accuracy of Logistic regression classifier on training set: {:.2f}'  
.format(logreg.score(X_train, y_train)))
```

```
print('Accuracy of Logistic regression classifier on test set: {:.2f}'  
.format(logreg.score(X_test, y_test)))
```

Accuracy of Logistic regression classifier on training set: 0.82

Accuracy of Logistic regression classifier on test set: 0.83

2. Decision Tree

2. Decision Tree

```
In [50]: from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier().fit(X_train, y_train)
print('Accuracy of Decision Tree classifier on training set: {:.2f}'
      .format(clf.score(X_train, y_train)))
print('Accuracy of Decision Tree classifier on test set: {:.2f}'
      .format(clf.score(X_test, y_test)))
```

Accuracy of Decision Tree classifier on training set: 1.00
Accuracy of Decision Tree classifier on test set: 0.75

3. K-Nearest Neighbors

```
In [212]: from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)
print('Accuracy of K-NN classifier on training set: {:.2f}'
      .format(knn.score(X_train, y_train)))
print('Accuracy of K-NN classifier on test set: {:.2f}'
      .format(knn.score(X_test, y_test)))
```

Accuracy of K-NN classifier on training set: 0.85
Accuracy of K-NN classifier on test set: 0.82

4. Linear Discriminant Analysis

```
In [52]: from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis()
lda.fit(X_train, y_train)
print('Accuracy of LDA classifier on training set: {:.2f}'
      .format(lda.score(X_train, y_train)))
print('Accuracy of LDA classifier on test set: {:.2f}'
      .format(lda.score(X_test, y_test)))
```

Accuracy of LDA classifier on training set: 0.81
Accuracy of LDA classifier on test set: 0.82

5. Gaussian Naive Bayes

```
In [53]: from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
gnb.fit(X_train, y_train)
print('Accuracy of GNB classifier on training set: {:.2f}'
      .format(gnb.score(X_train, y_train)))
print('Accuracy of GNB classifier on test set: {:.2f}'
      .format(gnb.score(X_test, y_test)))
```

Accuracy of GNB classifier on training set: 0.52
Accuracy of GNB classifier on test set: 0.52

6. Support Vector Machine

```
In [219]: from sklearn.svm import SVC
svm = SVC(C=1, kernel='poly')
svm.fit(X_train, y_train)
print('Accuracy of SVM classifier on training set: {:.2f}'
      .format(svm.score(X_train, y_train)))
print('Accuracy of SVM classifier on test set: {:.2f}'
      .format(svm.score(X_test, y_test)))
```

Accuracy of SVM classifier on training set: 0.85
Accuracy of SVM classifier on test set: 0.85

7. NN-Multilayer Perception

```
In [189]: from sklearn.neural_network import MLPClassifier
clf=MLPClassifier(activation='relu', alpha=1e-05, batch_size='auto',
beta_1=0.9, beta_2=0.999, early_stopping=False,
epsilon=1e-08, hidden_layer_sizes=(5,2),
learning_rate='constant', learning_rate_init=0.001,
max_iter=2000, momentum=0.9, n_iter_no_change=10,
nesterovs_momentum=True, power_t=0.9, random_state=1,
shuffle=True, solver='lbfgs', tol=0.0001,
validation_fraction=0.1, verbose=False, warm_start=False)

clf2=clf.fit(X_train, y_train)
clf.fit(X_train, y_train)
print('Accuracy of MLP classifier on training set: {:.2f}'
      .format(clf.score(X_train, y_train)))
print('Accuracy of MLP classifier on test set: {:.2f}'
      .format(clf.score(X_test, y_test)))
```

Accuracy of MLP classifier on training set: 0.85
Accuracy of MLP classifier on test set: 0.85

WINNER MODEL -- MLP VS KNN VS SVM

MLP

```
In [218]: #MLP
print('Confusion Matrix and Precision-Recall for MLP')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred = clf.predict(X_test)
print(confusion_matrix(y_test, pred))
print(classification_report(y_test, pred))
```

Confusion Matrix and Precision-Recall for MLP

```
[[160  8  0]
 [ 22 179  2]
 [  0  28  4]]
      precision    recall  f1-score   support

     1       0.88       0.95       0.91        168
     2       0.83       0.88       0.86        203
     3       0.67       0.12       0.21         32
```


accuracy			0.85	403
macro avg	0.79	0.65	0.66	403
weighted avg	0.84	0.85	0.83	403

KNN

```
In [217]: #KNN
print('Confusion Matrix and Precision-Recall for KNN')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred2 = knn.predict(X_test)
print(confusion_matrix(y_test, pred2))
print(classification_report(y_test, pred2))
```

Confusion Matrix and Precision-Recall for KNN

```
[[150  18   0]
 [ 21 174   8]
 [   0  24   8]]

      precision    recall  f1-score   support

     1         0.88        0.89        0.88         168
     2         0.81        0.86        0.83         203
     3         0.50        0.25        0.33          32

 accuracy          0.82          403
 macro avg         0.73          403
 weighted avg      0.81          403
```

SVM

```
In [221]: #SVM
print('Confusion Matrix and Precision-Recall for SVM')
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
pred3 = svm.predict(X_test)
print(confusion_matrix(y_test, pred3))
print(classification_report(y_test, pred3))
```

Confusion Matrix and Precision-Recall for SVM

```
[[160   8   0]
 [ 22 180   1]
 [   2  27   3]]

      precision    recall  f1-score   support

     1         0.87        0.95        0.91         168
     2         0.84        0.89        0.86         203
     3         0.75        0.09        0.17          32

 accuracy          0.85          403
 macro avg         0.82          403
 weighted avg      0.84          403
```

Comparing other scores for KNN vs MLP vs SVM

KNN wins as it balances the prediction for all three classes more correctly.

Drawing Decision Boundary

```
In [223]: import matplotlib.cm as cm
from matplotlib.colors import ListedColormap, BoundaryNorm
import matplotlib.patches as mpatches
import matplotlib.patches as mpatches

#X = fruits[['mass', 'width', 'height', 'color_score']]
#y = fruits['fruit_label']
#X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
#X1= X_train.to_numpy(X_train)
#y1=y_train.to_numpy(y_train)

def plot_Queue_knn(X_train, y_train, n_neighbors, weights):
    #XX1= pd.DataFrame(X_train)
    #XXX= XX1.iloc[0:,0:18]
    #yy1= pd.DataFrame(y_train)
    #yyy= yy1.iloc[0:19]
    #XXX= XXX.to_numpy()
    #yyy= yyy.to_numpy()

    XXX= data[['SGN_Green_Time','SGN_Red_Time']]
    #XXX= X_train.iloc[0:,3:4]
    yyy= data['Q_filled']

    #X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
    X_mat = XXX.to_numpy()
    y_mat = yyy.to_numpy()

    # Create color maps
    cmap_light = ListedColormap(['#FFAAAA', '#AAFFAA', '#AAAAFF'])
    cmap_bold = ListedColormap(['#FF0000', '#00FF00', '#0000FF'])

    knn2 = KNeighborsClassifier(n_neighbors, weights=weights)
    knn2.fit(X_mat, y_mat)

    # Plot the decision boundary by assigning a color in the color map
    # to each mesh point.

    mesh_step_size = .01 # step size in the mesh
    plot_symbol_size = 50

    x_min, x_max = X_mat[:, 0].min() - 1, X_mat[:, 0].max() + 1
    y_min, y_max = X_mat[:, 1].min() - 1, X_mat[:, 1].max() + 1
    xx, yy = np.meshgrid(np.arange(x_min, x_max, mesh_step_size),
                          np.arange(y_min, y_max, mesh_step_size))
    Z = knn2.predict(np.c_[xx.ravel(), yy.ravel()])

    # Put the result into a color plot
```

```

Z = Z.reshape(xx.shape)
plt.figure()
plt.pcolormesh(xx, yy, Z, cmap=cmap_light)

# Plot training points
plt.scatter(X_mat[:, 0], X_mat[:, 1], s=plot_symbol_size, c=y, cmap=cmap_bold
, edgecolor = 'black')
#plt.xlim(xx.min(), xx.max())
plt.xlim(19, 20)
plt.ylim(35, 45)

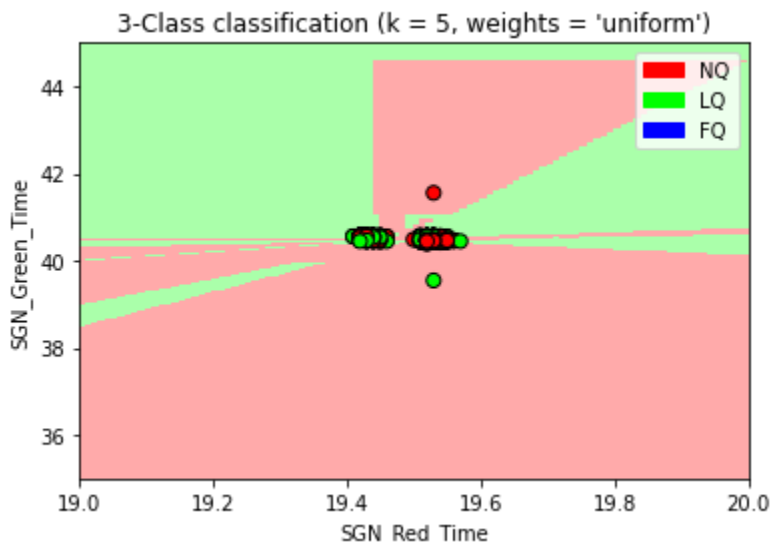
patch0 = mpatches.Patch(color='#FF0000', label='NQ')
patch1 = mpatches.Patch(color='#00FF00', label='LQ')
patch2 = mpatches.Patch(color='#0000FF', label='FQ')
#patch3 = mpatches.Patch(color='#AFAFAF', label='Lemon')
plt.legend(handles=[patch0, patch1, patch2])

plt.xlabel('SGN_Red_Time')
plt.ylabel('SGN_Green_Time')
plt.title("3-Class classification (k = %i, weights = '%s')"\
          % (n_neighbors, weights))

plt.show()

plot_Queue_knn(X_train,y_train, 5, 'uniform')

```



Drawing Nearest Neighbours VS Accuracy

We find k=5 to be optimum for modelling

```

In [264]: neighbors = np.arange(1,20)
train_accuracy = np.empty(len(neighbors))
test_accuracy = np.empty(len(neighbors))

for i,k in enumerate(neighbors):
    #Setup a knn classifier with k neighbors
    knn = KNeighborsClassifier(n_neighbors=k)

    #Fit the model
    knn.fit(X_train, y_train)

```

```

#Compute accuracy on the training set
train_accuracy[i] = knn.score(X_train, y_train)

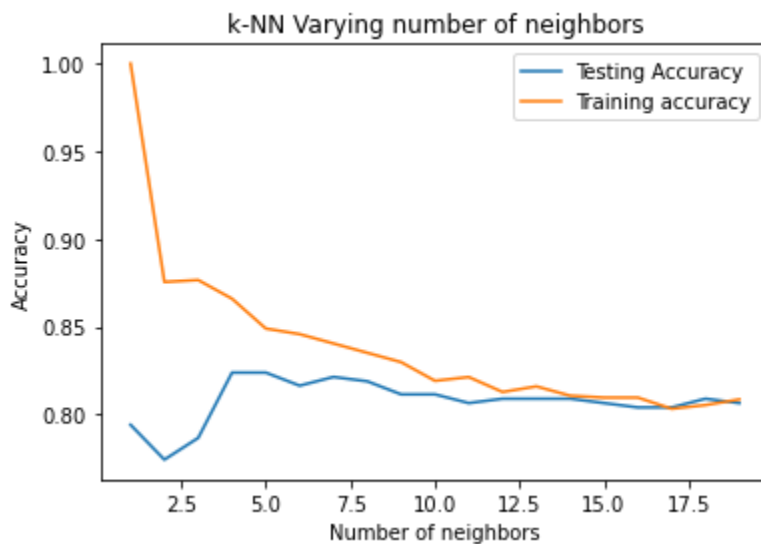
#Compute accuracy on the test set
test_accuracy[i] = knn.score(X_test, y_test)

```

```

plt.title('k-NN Varying number of neighbors')
plt.plot(neighbors, test_accuracy, label='Testing Accuracy')
plt.plot(neighbors, train_accuracy, label='Training accuracy')
plt.legend()
plt.xlabel('Number of neighbors')
plt.ylabel('Accuracy')
plt.show()

```



Plotting ROC curve for LQ=2 and FQ=3

In [291]: *# 2-class classification*

```

from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score

pred = knn.predict(X_test)
pred_prob = knn.predict_proba(X_test)

# roc curve for classes
fpr = {}
tpr = {}
thresh = {}

n_class = 3

```

PCD Generation and Other Data Mining Tool (adopted for and developed for CSCS800)

API Key

3N6gV0u8LpL3
bYnKMv9HwY

Intersection ID b3f45f76-a6ef-400c-b8a2-47dd64169b8d

Location	Central	and	Vernor
Direction of Study Approach	SB		Central

Lane Configuration Th-Left + Right

Phase	4
Description	Analysis for Directed Study & Project for CSCS800

Date of Analysis 23-Nov

Analysis by TG

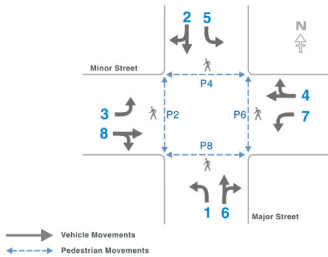
Event Codes of Interest	Begin of Green	Begin of Red	Detector on	Detector Off
	1	10	82	81

Parameters of Interest

- 4 SB Phase
- 1 Detection Channe Advanced Pulse
- 2 Detection Channe Stop-bar RT+TH Presence
- 5 Detection Channe Stop-bar LT Presence
- 7 Detection Channe Through Departure Presence

Assumptions and Configurations

- Red or Green
- 30 mph ~ 45 ft/sec
- ~150 ft advance pulse Detector
- No residual queue assumed



Notes from Dec-3-2020

- check for: should a Protected left turn be required?
- LT Permit?
- We also have TMC for opposing traffic (allowing us to look at gap and ability to turn left)

Performance Measure

- Data driven approach
- Cross product HCM method vs. Actual data driven method.

Sensitivity Analysis

is it with in a percentage? (~10%)

- Maybe try it out on Jefferson with higher dat and more coordination
- Issue with Trucks
- more peds

Elevator Pitch and Parameter Summary Schematic

Training Set

- We need to fill in the blanks (Assume Missing Data) for the cycles highlighted in yellow and marked as Missing
- or identify and ignore the outliers (just as spikes in TMC/Queue when some data is missing)
- Input attributes (features) are marked in green "use" on top
- Ignore all Columns with a balck "No" or "Do not Use"
- Neural Network

Test Set to follow

Do we need a validation set as well?

Assuming no residual queue for now

In the future we can add that based on Headway below a certain threshold and TMC counts about a certain value

Big questions for later:

- Distinguish between Arrival on Green (AOG) and Pass of Green (POG)
- POG is when the vehicle is already in the queue but passes on green
- AOG indicates that it arrived and crossed on Green

When the Vehicle arrives on Green or Red at the Channel box is very important to the decision on the Queue.

for Ch-2 if there is much DOR, then the length of presense for the last AOR is important to identify vehicle status

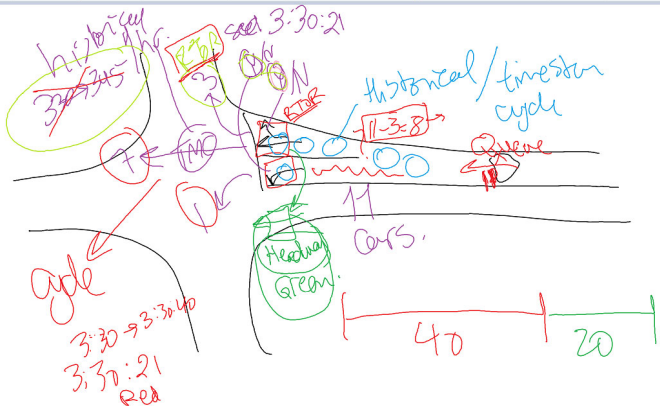
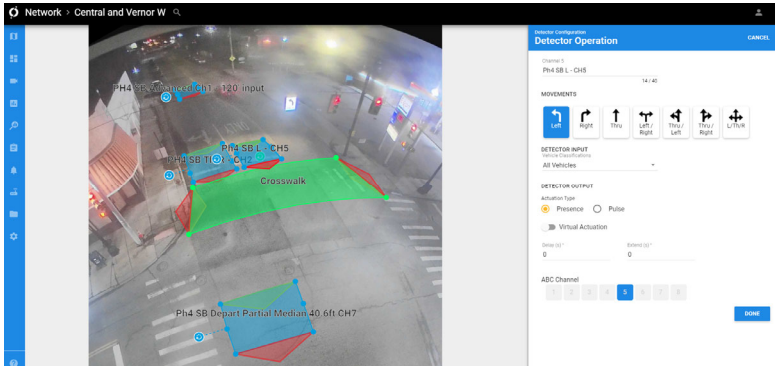
There's a very credible correlation between the RTOR achieved from the TMC and that identified from movement on RED at CH2 detector that need to be investigated (LATER)

GET /intersections/{intersectionId}/hiresdata Intersection Telemetry in High Resolution Data Logger Output Format

Returns telemetry data from the intersection in the HiRes Data Format, which is documented [here](#). TrafficLink supports the following event codes:

- 0 (Phase On)
- 1 (Phase Begin Green)
- 7 (Phase Green Termination)
- 8 (Phase Begin Yellow Clearance)
- 9 (Phase End Yellow Clearance)
- 10 (Phase Begin Red Clearance)
- 11 (Phase End Red Clearance)
- 12 (Phase Inactive)
- 21 (Pedestrian Begin Walk)
- 22 (Pedestrian Begin Clearance)
- 23 (Pedestrian Begin Solid Don't Walk)
- 61 (Overlap Begin Green)
- 62 (Overlap Begin Yellow)
- 64 (Overlap Begin Red Clearance)
- 65 (Overlap Off)
- 81 (Detector On)
- 82 (Detector Off)
- 89 (PedDetector On)
- 90 (PedDetector Off)

Requests must be for one hour or less.



Training Set

- We need to fill in the blanks (Assume Missing Data) for the cycles highlighted in yellow and marked as Missing
- or identify and ignore the outliers (just as spikes in TMC/Queue when some data is missing)
- Input attributes (features) are marked in green "use" on top
- Ignore all Columns with a balck "No" or "Do not Use"
- Neural Network

Test Set to follow

Do we need a validation set as well?

Assuming no residual queue for now

In the future we can add that based on Headway below a certain threshold and TMC counts about a certain value

Big questions for later:

- Distinguish between Arrival on Green (AOG) and Pass of Green (POG)
- POG is when the vehicle is already in the queue but passes on green
- AOG indicates that it arrived and crossed on Green

When the Vehicle arrives on Green or Red at the Channel box is very important to the decision on the Queue.

for Ch-2 if there is much DOR, then the length of presense for the last AOR is important to identify vehicle status

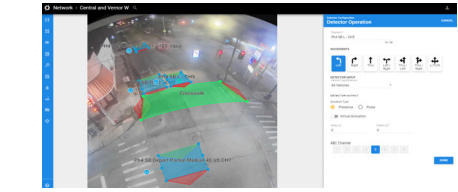
There's a very credible correlation between the RTOR achieved from the TMC and that identified from movement on RED at CH2 detector that need to be investigated (LATER)

[illegible]

Stop-bar Vehicles in LT Queue (Approx.)
Based on Arrivals

Stop-bar Vehicles in LT Queue (Approx.)
Based on Departures

Legend: 2013 LT Queue (Red), 2013 LT Queue (Green), 2013 LT Queue (Blue), 2013 LT Queue (Grey)



Red at Green **Detector** **30 mph ~ 45 Kph**

ArduPilot (Laser) **OpenStreetMap**
Geo-JSON **Geo-JSON**
Geo-JSON **Geo-JSON**
Geo-JSON **Geo-JSON**

Stage
 Plotting high-res data for Event Code ID and ID & Parameter 1
 2 Channel 2 and 3 are brought together, the other two are
 brought back - this are signposting for road-length upstream

Stop-bar Vehicles in IT Queue (Approx.)
Based on Arrivals

100% 100% 100% 100% 100% 100% 100% 100% 100%

Stop-bar Vehicles in IT Queue (Approx.)
Based on Departures

100% 100% 100% 100% 100% 100% 100% 100% 100%

Network - Control and Vector W

Time	eventParam (7 - Advance)	eventCode (81 - Detector off)	Time	eventParam (2 - Th+RT off)	eventCode (81 - Detector off)	Match	Chan	difference	speed (ft/sec)	Speed (mi/hr)
15:59:10.2	7	81	16:00:10.1	2	81	#N/A	#N/A		0.0	0.0
16:00:02.7	7	81	16:00:15.0	2	81	#N/A	#N/A		0.0	0.0
16:00:11.3	7	81	16:01:07.3	2	81	16:00:10.1	00:00:01.2		42.2	28.7
16:00:17.5	7	81	16:01:15.7	2	81	16:00:15.0	00:00:02.4		20.6	14.0
16:00:32.7	7	81	16:02:09.9	2	81	16:00:15.0	00:00:17.6		2.8	1.9
16:00:35.1	7	81	16:02:22.8	2	81	16:00:15.0	00:00:20.1		2.5	1.7
16:00:42.8	7	81	16:03:14.5	2	81	16:00:15.0	00:00:27.8		1.8	1.2
16:01:10.3	7	81	16:03:18.4	2	81	16:01:07.3	00:00:03.0		16.8	11.5
16:01:13.4	7	81	16:04:00.1	2	81	16:01:07.3	00:00:06.1		8.2	5.6
16:01:15.9	7	81	16:04:06.5	2	81	16:01:15.7	00:00:00.2		24.9	16.6
16:01:56.7	7	81	16:04:08.3	2	81	16:01:15.7	00:00:41.1		1.2	0.8
16:02:19.9	7	81	16:04:29.7	2	81	16:02:09.9	00:00:10.0		5.0	3.4
16:02:24.7	7	81	16:04:51.8	2	81	16:02:22.8	00:00:01.9		25.7	17.5
16:02:37.2	7	81	16:05:13.2	2	81	16:02:22.8	00:00:14.4		3.5	2.4
16:02:43.2	7	81	16:06:10.6	2	81	16:02:22.8	00:00:20.4		2.4	1.7
16:03:10.8	7	81	16:06:39.5	2	81	16:02:22.8	00:00:48.0		1.0	0.7
16:03:13.7	7	81	16:07:04.8	2	81	16:02:22.8	00:00:50.9		1.0	0.7
16:03:16.3	7	81	16:07:22.1	2	81	16:03:14.5	00:00:01.8		27.5	18.7
16:03:24.7	7	81	16:08:16.5	2	81	16:03:18.4	00:00:06.3		7.9	5.4
16:03:30.5	7	81	16:08:22.1	2	81	16:03:18.4	00:00:12.1		4.1	2.8
16:04:01.7	7	81	16:08:36.3	2	81	16:04:00.1	00:00:01.6		30.9	21.1
16:04:11.0	7	81	16:09:21.2	2	81	16:04:08.3	00:00:12.7		18.8	12.8
16:05:09.4	7	81	16:10:16.8	2	81	16:04:51.8	00:00:17.7		2.8	1.9
16:05:12.3	7	81	16:10:21.5	2	81	16:04:51.8	00:00:20.5		2.4	1.7
16:05:15.9	7	81	16:10:22.3	2	81	16:05:13.2	00:00:02.7		18.6	12.7
16:05:47.4	7	81	16:10:26.6	2	81	16:05:13.2	00:00:34.2		1.5	1.0
16:05:55.6	7	81	16:11:10.9	2	81	16:05:13.2	00:00:42.4		1.2	0.8
16:06:54.9	7	81	16:11:14.3	2	81	16:06:37.5	00:00:17.4		2.9	2.0
16:07:23.9	7	81	16:11:16.9	2	81	16:07:22.1	00:00:01.8		27.5	18.8
16:07:30.1	7	81	16:12:12.0	2	81	16:07:22.1	00:00:07.9		6.3	4.3
16:07:54.2	7	81	16:12:16.8	2	81	16:07:22.1	00:00:32.1		1.6	1.1
16:08:23.9	7	81	16:14:13.0	2	81	16:08:21.2	00:00:03.1		27.5	18.8
16:09:17.5	7	81	16:14:38.2	2	81	16:08:36.3	00:00:41.3		1.2	0.8
16:09:24.1	7	81	16:15:18.5	2	81	16:09:21.2	00:00:03.0		16.8	11.5
16:09:34.0	7	81	16:15:22.4	2	81	16:09:21.2	00:00:12.9		3.9	2.6
16:10:18.6	7	81	16:16:06.6	2	81	16:10:16.8	00:00:01.8		27.2	18.5
16:10:35.5	7	81	16:16:54.9	2	81	16:10:26.6	00:00:06.9		7.2	4.9
16:11:09.6	7	81	16:17:12.2	2	81	16:10:26.6	00:00:43.0		1.2	0.8
16:11:16.7	7	81	16:17:22.8	2	81	16:11:14.3	00:00:02.4		20.8	14.2
16:11:18.7	7	81	16:17:27.3	2	81	16:11:16.9	00:00:01.8		27.2	18.6
16:11:31.4	7	81	16:17:58.0	2	81	16:11:16.9	00:00:14.5		3.5	2.4
16:11:35.0	7	81	16:18:06.8	2	81	16:11:16.9	00:00:18.1		2.8	1.9
16:12:09.8	7	81	16:18:12.3	2	81	16:11:16.9	00:00:52.9		0.9	0.6
16:12:11.7	7	81	16:19:19.2	2	81	16:11:16.9	00:00:48.4		0.9	0.6
16:12:14.5	7	81	16:19:27.4	2	81	16:12:12.0	00:00:02.5		19.8	13.5
16:12:18.7	7	81	16:20:17.7	2	81	16:12:16.8	00:00:01.9		26.3	18.0
16:13:51.3	7	81	16:20:22.1	2	81	16:12:16.8	00:01:34.5		0.5	0.4
16:14:13.7	7	81	16:21:19.9	2	81	16:14:13.0	00:00:00.8		66.1	45.1
16:14:31.4	7	81	16:21:23.3	2	81	16:14:13.0	00:00:18.5		2.7	1.8
16:15:09.8	7	81	16:22:10.5	2	81	16:14:38.3	00:00:31.5		1.6	1.1
16:15:12.1	7	81	16:22:15.1	2	81	16:14:38.3	00:00:33.8		1.5	1.0
16:15:14.0	7	81	16:22:18.9	2	81	16:14:38.3	00:00:35.7		1.4	1.0
16:15:17.2	7	81	16:23:10.6	2	81	16:14:38.3	00:00:39.0		1.3	0.9
16:15:15.0	7	81	16:23:16.3	2	81	16:15:22.4	00:00:28.6		1.7	1.2
16:16:09.5	7	81	16:23:37.6	2	81	16:16:06.6	00:00:02.8		17.6	12.0
16:17:14.2	7	81	16:23:37.8	2	81	16:16:54.9	00:00:19.3		2.6	1.8
16:17:17.3	7	81	16:23:38.5	2	81	16:17:15.2	00:00:02.0		24.8	16.9
16:17:25.5	7	81	16:23:41.3	2	81	16:17:22.8	00:00:02.7		18.4	12.5
16:17:34.3	7	81	16:24:06.2	2	81	16:17:27.3	00:00:07.0		7.2	4.9
16:17:41.1	7	81	16:24:55.3	2	81	16:17:27.3	00:00:13.8		3.6	2.5
16:17:57.5	7	81	16:25:00.7	2	81	16:17:27.3	00:00:30.2		1.7	1.1
16:18:00.7	7	81	16:25:06.4	2	81	16:17:58.0	00:00:02.2		22.2	15.2
16:18:10.3	7	81	16:25:51.9	2	81	16:18:06.8	00:00:03.6		14.0	9.5
16:18:14.6	7	81	16:25:52.9	2	81	16:18:12.3	00:00:02.3		21.6	14.7
16:18:35.9	7	81	16:27:25.2	2	81	16:18:12.3	00:00:23.6		2.1	1.4
16:18:38.4	7	81	16:28:27.2	2	81	16:18:12.3	00:00:26.2		1.9	1.3
16:18:59.1	7	81	16:30:16.6	2	81	16:18:12.3	00:00:46.8		1.1	0.7
16:19:09.9	7	81	16:31:12.3	2	81	16:18:12.3	00:00:57.7		0.9	0.6
16:19:13.9	7	81	16:31:18.8	2	81	16:18:12.3	00:01:01.6		0.8	0.6
16:19:21.9	7	81	16:31:22.4	2	81	16:19:21.9	00:00:00.0		0.0	0.0
16:19:24.5	7	81	16:31:43.0	2	81	16:19:21.9	00:00:02.6		19.2	13.1
16:19:36.0	7	81	16:33:07.0	2	81	16:19:27.4	00:00:08.5		5.9	4.0
16:19:38.6	7	81	16:33:29.1	2	81	16:19:27.4	00:00:11.2		4.5	3.0
16:20:10.3	7	81	16:33:30.1	2	81	16:19:27.4	00:00:42.8		1.2	0.8
16:20:17.4	7	81	16:33:53.4	2	81	16:19:27.4	00:00:49.9		1.0	0.7
16:20:36.1	7	81	16:34:17.9	2	81	16:20:22.1	00:00:14.0		3.6	2.4
16:20:56.2	7	81	16:35:05.3	2	81	16:20:22.1	00:00:34.2		1.5	1.0
16:20:59.4	7	81	16:36:14.1	2	81	16:20:22.1	00:00:37.3		1.3	0.9
16:21:14.7	7	81	16:36:20.5	2	81	16:20:22.1	00:00:52.6		1.0	0.6
16:21:15.9	7	81	16:37:16.9	2	81	16:20:22.1	00:00:53.9		0.9	0.6
16:21:19.2	7	81	16:37:42.1	2	81	16:20:22.1	00:00:57.2		0.9	0.6
16:22:08.4	7	81	16:37:45.7	2	81	16:21:23.3	00:00:45.1		1.1	0.8
16:22:11.4	7	81	16:37:48.2	2	81	16:22:10.6	00:00:00.8		65.9	44.9
16:22:14.8	7	81	16:37:48.8	2	81	16:22:10.6	00:00:04.2		11.9	8.1
16:22:17.1	7	81	16:37:53.1	2	81	16:22:15.1	00:00:02.0		25.3	17.2
16:22:21.0	7	81	16:37:55.4	2	81	16:22:18.9	00:00:02.1		23.8	16.2
16:22:57.4	7	81	16:37:57.0	2	81	16:22:18.9	00:00:38.5		1.3	0.9
16:23:13.7	7	81	16:38:06.0	2	81	16:23:10.6	00:00:03.1		16.3	11.1
16:23:31.1	7	81	16:38:06.9	2	81	16:23:16.3	00:00:14.9		3.4	2.3
16:23:43.7	7	81	16:38:14.3	2	81	16:23:41.3	00:00:02.0		25.1	17.1
16:24:12.3	7	81	16:38:21.6	2	81	16:24:06.2	00:00:06.1		8.2	5.6
16:24:14.8	7	81	16:39:11.7	2	81	16:24:06.2	00:00:08.5		5.9	4.0
16:24:36.9	7	81	16:39:14.7	2	81	16:24:06.2	00:00:30.7		1.6	1.1
16:24:37.8	7	81	16:40:17.5	2	81	16:24:06.2	00:00:11.6		1.6	1.1
16:24:40.3	7	81	16:40:22.0	2	81	16:24:06.2	00:00:34.1		1.5	1.0
16:25:09.2	7	81	16:40:45.3	2	81	16:25:06.4	00:00:02.8		17.7	12.1
16:25:31.4	7	81	16:41:22.3	2	81	16:25:06.4	00:00:25.0		2.0	1.4
16:25:35.2	7	81	16:41:51.1	2	81	16:25:06.4	00:00:28.8		1.7	1.2
16:25:49.2	7	81	16:42:11.5	2	81	16:25:06.4	00:00:42.8		1.2	0.8
16:26:37.2	7	81	16:42:12.1	2	81	16:25:52.9	00:00:44.3		1.1	0.8
16:27:12.4	7	81	16:42:17.1	2	81	16:25:52.9	00:01:19.5		0.6	0.4
16:27:15.2	7	81	16:43:15.2	2	81	16:25:52.9	00:01:22.3		0.6	0.4
16:27:20.4	7	81	16:43:50.6	2	81	16:25:52.9	00:01:27.4		0.6	0.4
16:27:31.9	7	81	16:44:12.8	2	81	16:27:25.5	00:00:06.4		7.8	5.3
16:27:36.4	7	81	16:44:16.8	2	81	16:27:25.5	00:00:10.9		4.6	3.1
16:28:10.7	7	81	16:46:29.5	2	81	16:27:25.5	00:00:45.2		1.1	0.8
16:28:22.6	7	81	16:47:07.4	2	81	16:27:25.5	00:00:57.1		0.9	0.6
16:28:31.0	7	81	16:47:11.4	2	81	16:28:22.7	00:00:08.3		6.0	4.1

[illegible]