CCTV Cardinal Direction Priority and ITS Data Inventory

Zixuan Zhou – Data Analysis and System Configuration Intern with ITS Team – Syracuse University

Introduction

Main Projects:

CCTV Cardinal Direction Priority Analysis: Analyzing the Traffic Event Data and Traffic Congestion Data, defining priorities of current Cameras, and making suggestions of where new Cameras should be added.

ITS Database Design and Implementation: Collecting and organizing Camera information, designing data inventory, and implementing database in SQL.

Data Description and Preprocessing

Data Source:

1. Traffic Event Data, available on DDOT SQL database.
2. District Mobility Data, available on INRIX and American Community Survey (ACS).
3. WMATA Automatic Passenger Count (APC)

Traffic Event Data Cleaning:

1. Checking missing value
2. Removing outliers
3. Dropping useless columns
4. Renaming Columns
5. Changing data format for Latitude and Longitude.

Data Cleaning	Before	After
Row	49	25
Column	27308	26292

Traffic Event Data Transformation:

1. Adding two columns, Year and Month by abstracting data from Date Creation Column.
2. Categorizing the Time Period into 6 shifts, AM Early (4:00 - 6:00), AM Peak (6:00 - 9:00), Mid Day (9:00 - 15:00), PM Peak (15:00 - 19:00), Early Night (19:00 - 23:00), and Late Night (23:00 - 4:00).
3. Grouping the Event Types into 11 categories, Disabled Vehicle, Motor/Bike, Other, Other Vehicle Issues, Pedestrian Event, Special Event, Traffic Control, Traffic Signal, Trouble Access, Vehicle Collision, and Weather.

District Mobility Data Transformation:

1. Merged all line files and turned lines into points.
2. Merged all points.

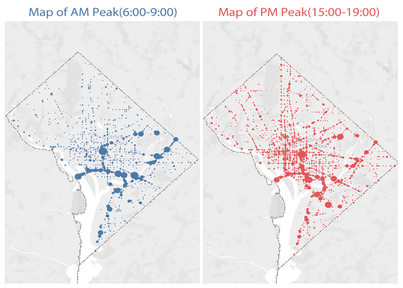
Data Exploration

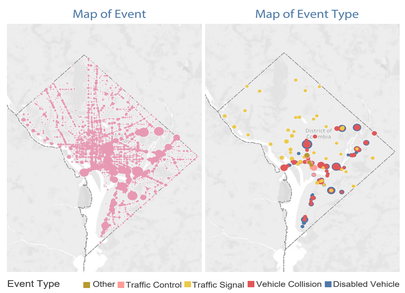
Traffic Event Data

Number of Traffic Events changing from November 2015 to July 2019

Traffic Signal, Disabled Vehicle, and Vehicle Collision are the top 3 Event Types.

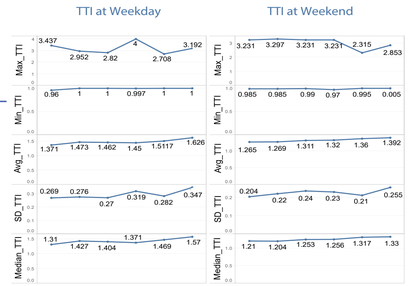
Even though the bar chart shows that more traffic events occur in Mid Day and Early Night, it's because Mid Day and Early Night have large time range. In fact, AM Peak and PM Peak have higher traffic event rates per hour than other time periods.





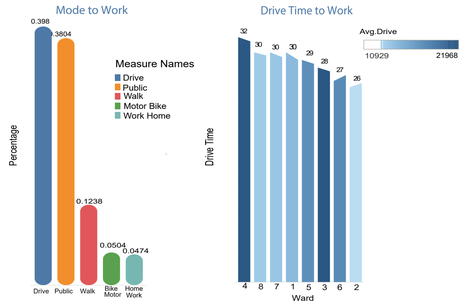
In the map of traffic event distribution, the area with higher events can be roughly seen.

Travel Time to Work



The Travel Time Index (TTI) is the ratio of travel time in the peak period to the travel time at free-flow conditions. A value of 1.35 indicates a 20 minutes free-flow trip takes 27 minutes in the peak.

Mode to Work

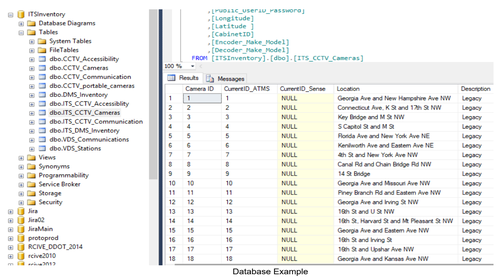


Database Design and Implementation

Data Description: Data files contain information of cameras.

Design and Implementation:

1. Set up Primary key for main tables, CCTV_Cameras, CCTV_Communication, and CCTV_Accessibility.
2. Took each data file as a table and made ERD to link each table.
3. Exported camera data from SQL database and matched the common features between SQL data and other data files by using data analysis tools such as R, SQL, and Teableau.
4. Organized data into three designed main tables, CCTV_Cameras, CCTV_Communication, and CCTV_Accessibility.
5. Wrote SQL query to build database construction.
6. Imported data into SQL database.



Database Example

CCTV Cardinal Direction Priority Project

Methodology: Clustering the Camera Location and DC Major Road by calculating the number of Traffic Events and average TTI of Traffic Congestion points closest with them based on the distance between Event Location and Camera Location, Event Location and Road Location, Congestion Location and Camera Location, Congestion Location and Road Location.

Kernel Density for Event Points with Camera Location, Event Points with Roads, Congestion Points with Camera Location, Congestion Points with Roads.

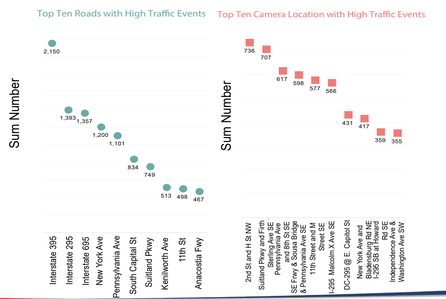
Merging Kernel Density of Event and Congestion.

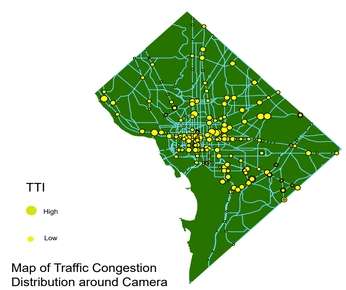
Tips for Kernel Density:
The Kernel Density tool calculates the density of features in a neighborhood around those features. It can be calculated for both point and line features.

$$Density = \frac{1}{(radius)^2} \sum_{i=1}^{n} \frac{3}{\pi} \cdot pop_i \left(1 - \left(\frac{dist_i}{(radius)}\right)^2\right)^2$$

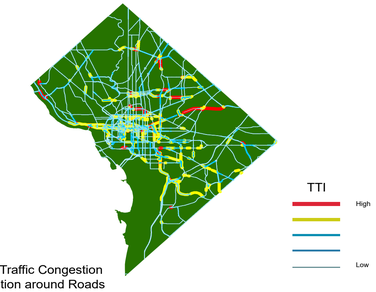
For $dist_i < radius$

Result

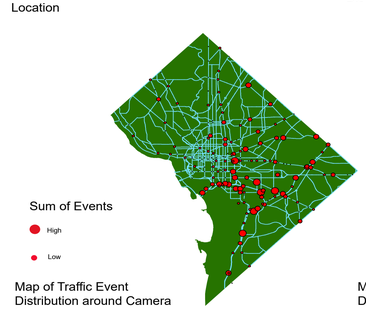




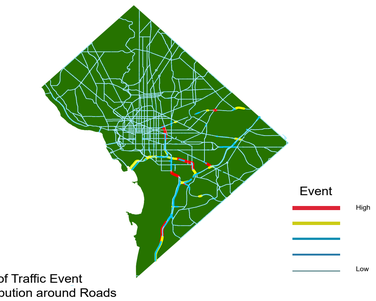
Map of Traffic Congestion Distribution around Camera Location



Map of Traffic Congestion Distribution around Roads

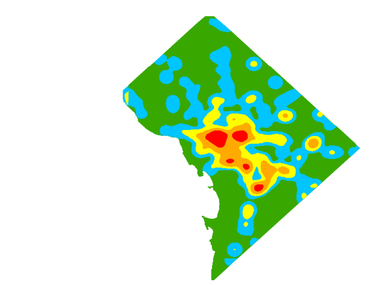


Map of Traffic Event Distribution around Camera Location



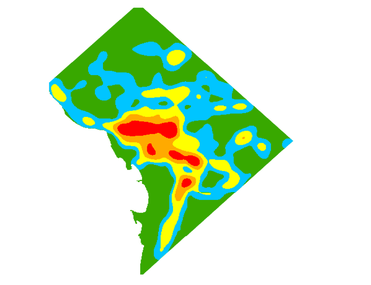
Map of Traffic Event Distribution around Roads

Kernel Density of Traffic Event and Congestion



CCTV_Cardinal Direction Priority
TTI
Level 1
Level 2
Level 3
Level 4
Level 5

Kernel Density of Merged Traffic Event and Congestion with Camera Location(Feature 1)



CCTV_Cardinal Direction Priority_Suggestion
TTI
Level 1
Level 2
Level 3
Level 4
Level 5

Kernel Density of Merged Traffic Event and Congestion with Roads (Feature 2)

Conclusion:

Feature 1 shows the current camera location in 5 levels.

Feature 2 shows the areas where cameras are suggested to be installed.

Future Work

Data Inventory Project

1. Updating information to database.
2. Building Interface.

CCTV Cardinal Direction Priority

1. Updating District Mobility Data and using latest data to do analysis.
2. Predicting Traffic Event in the future by machine learning.
   a. Reselecting tables and adding more features from SQL database.
   b. Using PCA to find top components.
   c. Two models maybe used, Linear Regression and Random Forest.
   d. Model Comparison by splitting into three dataset, 60% training, 30% validating, and 10% testing.
   e. Comparing prediction performance by checking the RMSE and R^2.

Reference:

1. https://www.census.gov/programs-surveys/acs/
2. https://www.census.gov/programs-surveys/geography.html
3. https://www.inrix.com/products/apis/
4. https://www.bts.gov/content/travel-time-index
5. https://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm
6. https://opendata.dc.gov/datasets/3573febb6405f4bb0bfdcedd521ac3122_59