

1.6 million UK traffic accidents

Visualise and analyse traffic demographics

Alger Remirata
Data Scientist

The Rationale

Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. Patterns involved in dangerous crashes could be detected if we develop a prediction model that automatically classifies the type of injury severity of various traffic accidents. These behavioral and roadway patterns are useful in the development of traffic safety control policy.

Road accidents are never a happy issue to discuss. It not only has severe consequences for those involved, it also affects the lives of many others like friends and family. With more vehicles on the road than ever before, its important to understand them in greater detail, and possibly 'predict' the locations and consequences of these accidents. Government agencies in the UK have been collecting data about the accidents that were reported since the year 2005. The data includes generic and specific details about the vehicles, driver, number of passengers and number of casualties.

Goal

Key Business Problems:

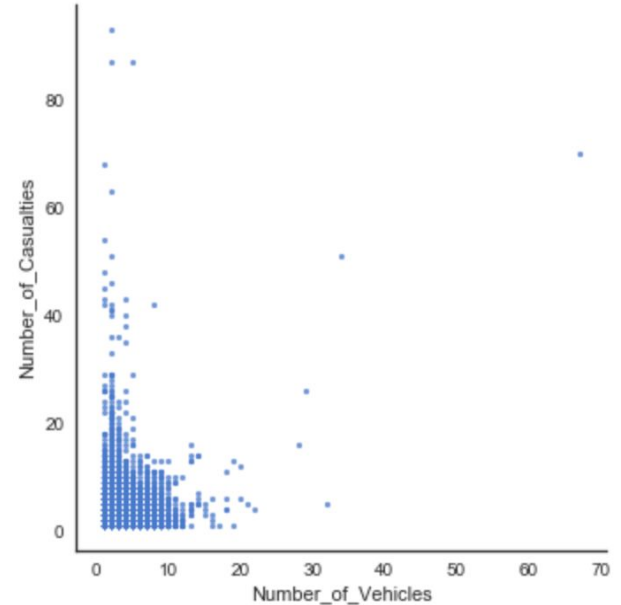
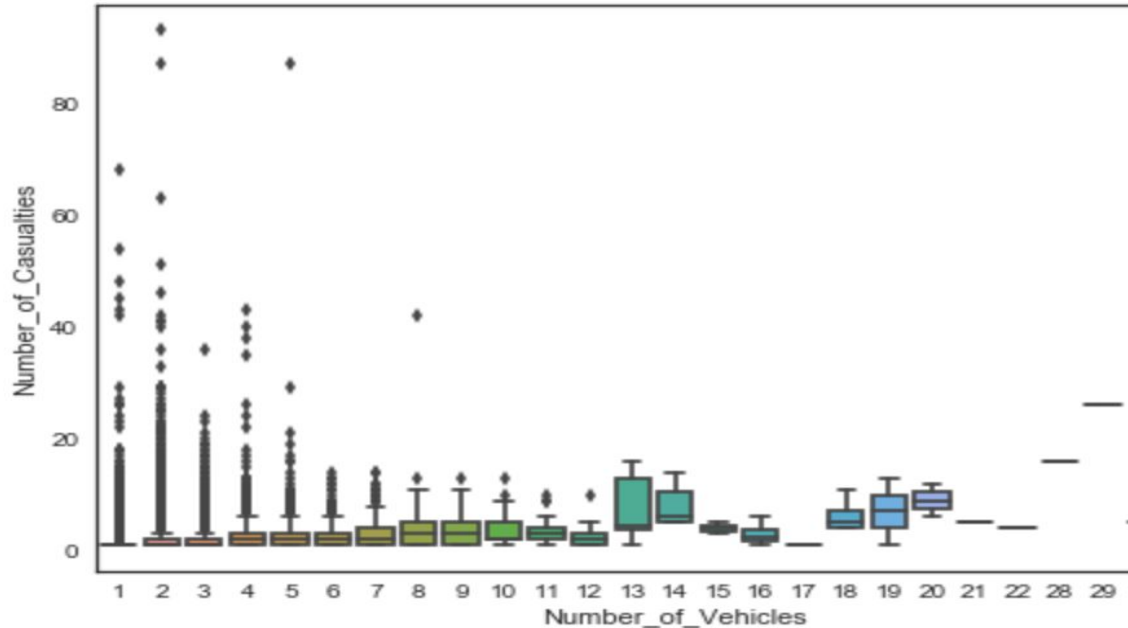
1. Can we identify and quantify associations between number of casualties and other variables in a road accident?
2. Explore whether it is possible to predict accident hot-spots based on the data.

Possible Stakeholders:

1. Policy-making agencies like the government
2. Engineers and researchers in the automobile industry
3. Freight transportation company and other companies providing international courier, parcel and express mail services.

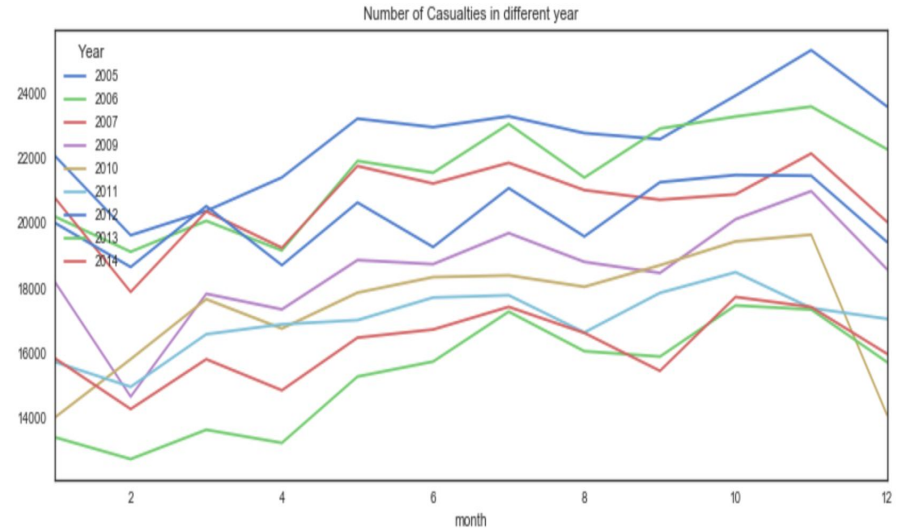
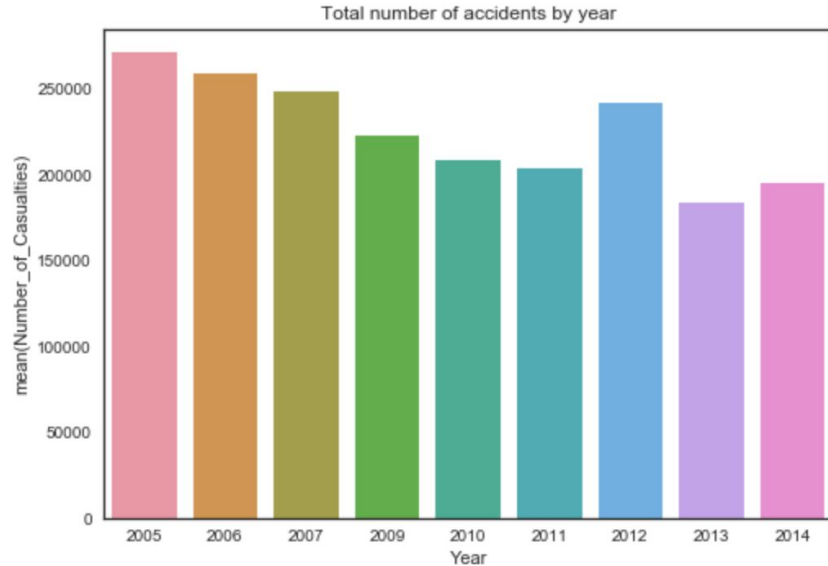
Exploratory Data Analysis

Question: If we have more vehicles involved in the accident, do we have more casualties?



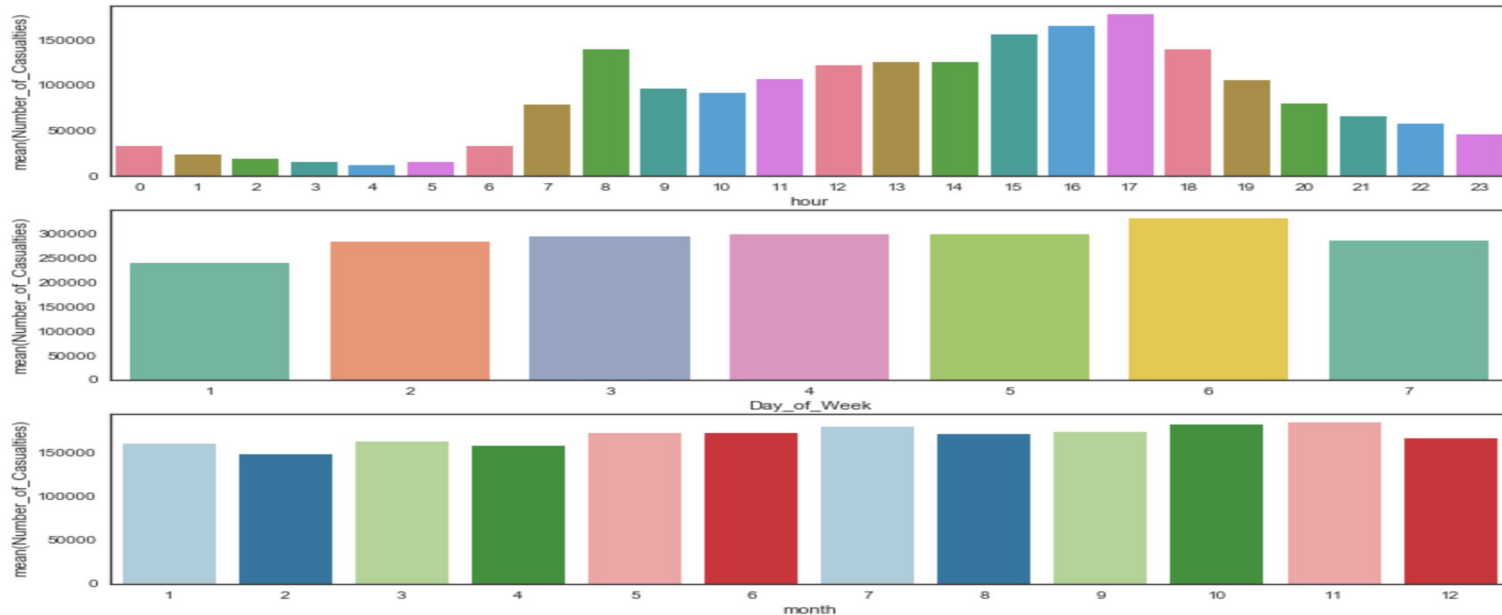
Exploratory Data Analysis

Question: In which year we have more casualties in road accidents? Does the number of casualties increase or decrease?



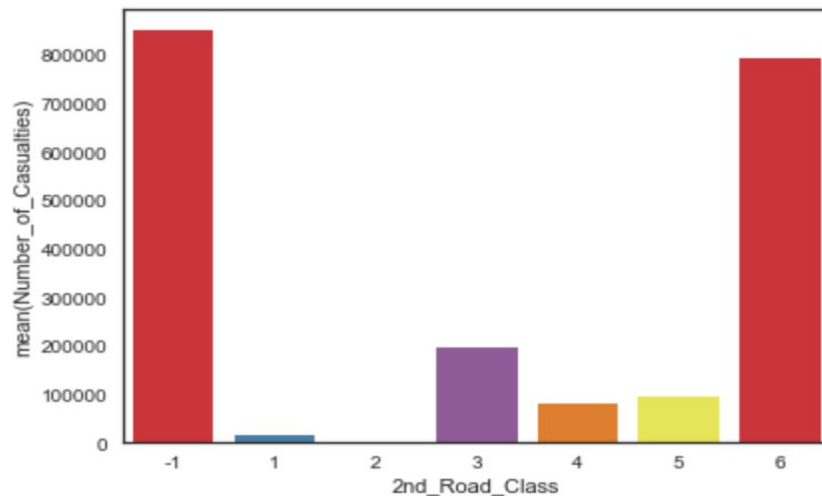
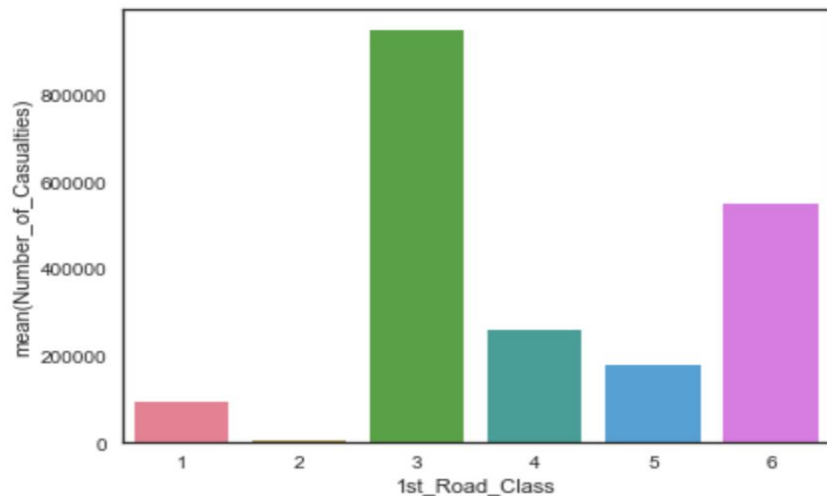
Exploratory Data Analysis

Question: What time of the day we have more accidents happening? What day of the week? What month can we see an increase in the number of road accidents happening?



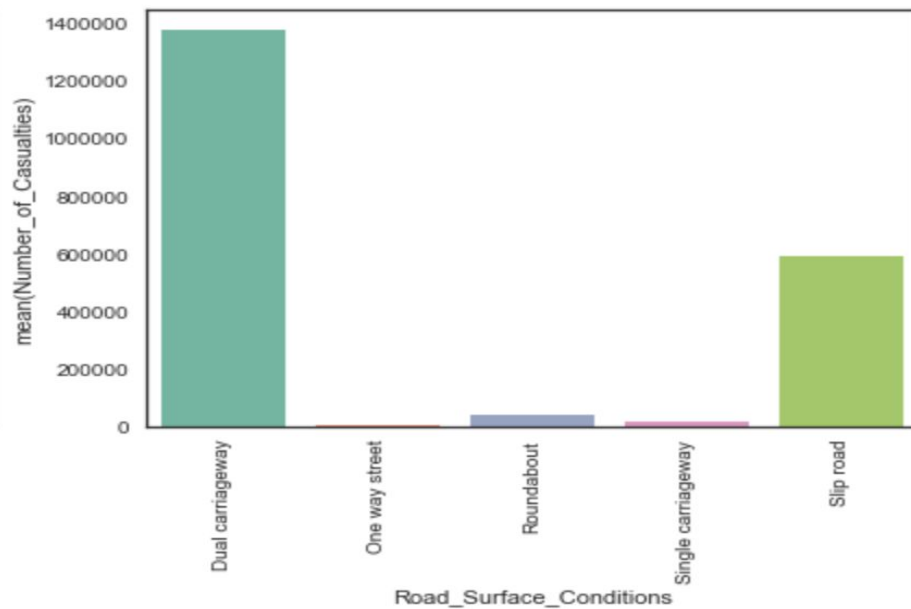
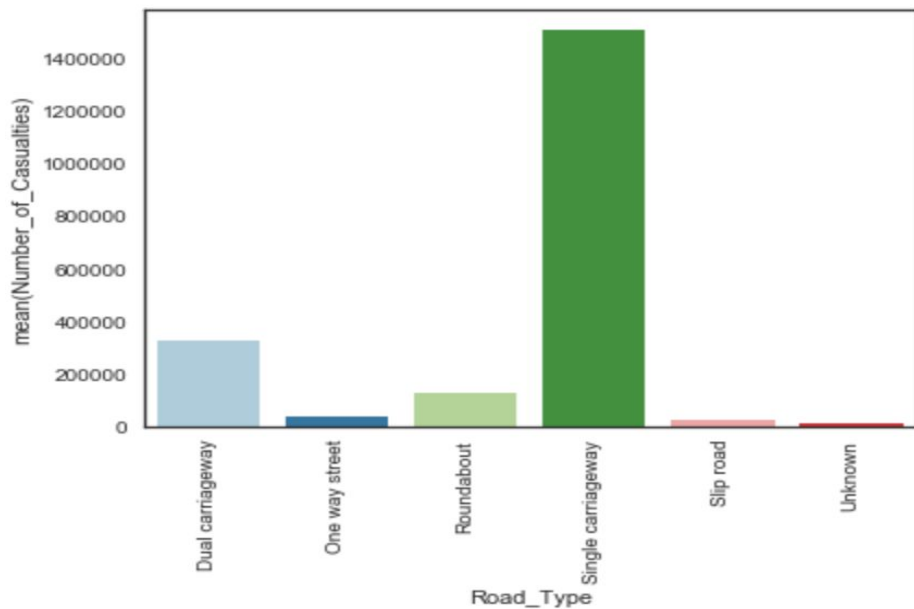
Exploratory Data Analysis

Question: Is there specific class of roads accidents happen more frequently?



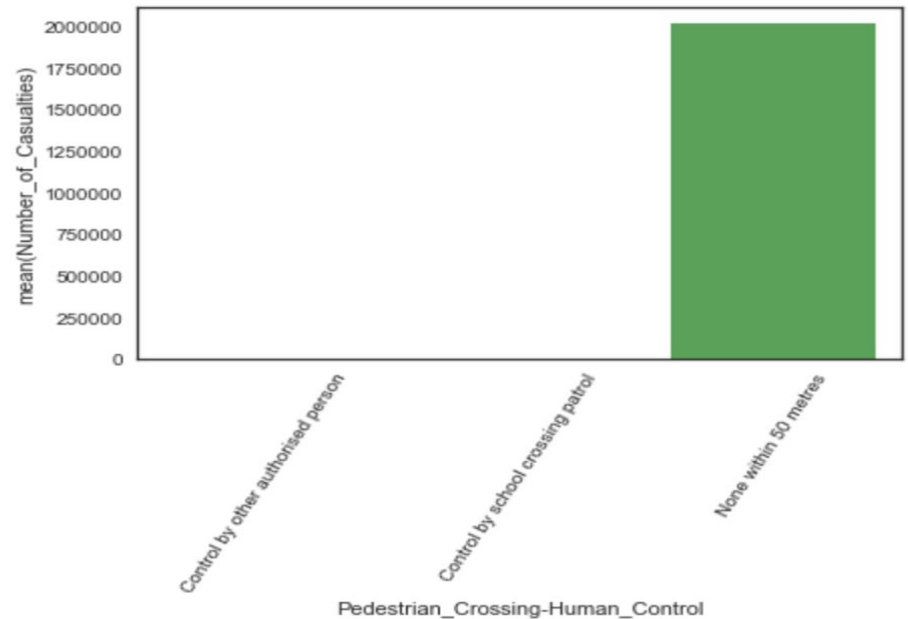
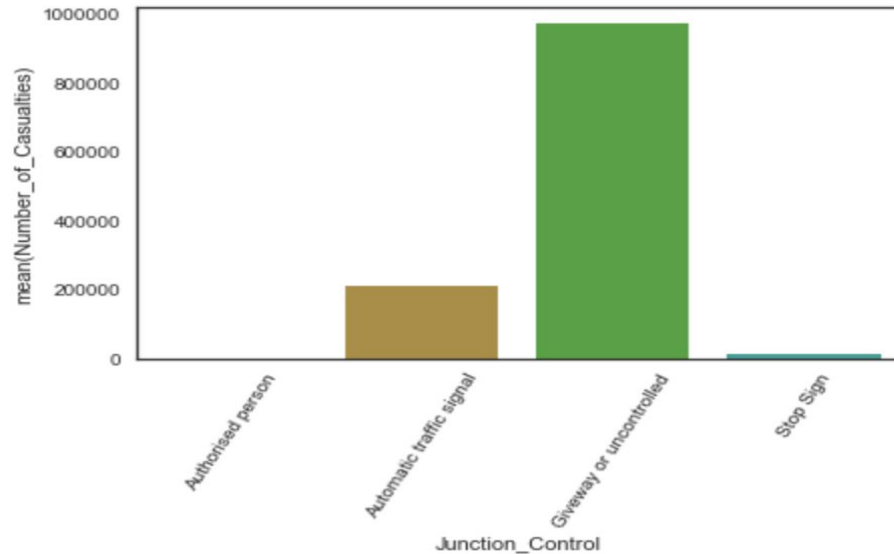
Exploratory Data Analysis

Question: What type and conditions of roads accidents frequently happen?



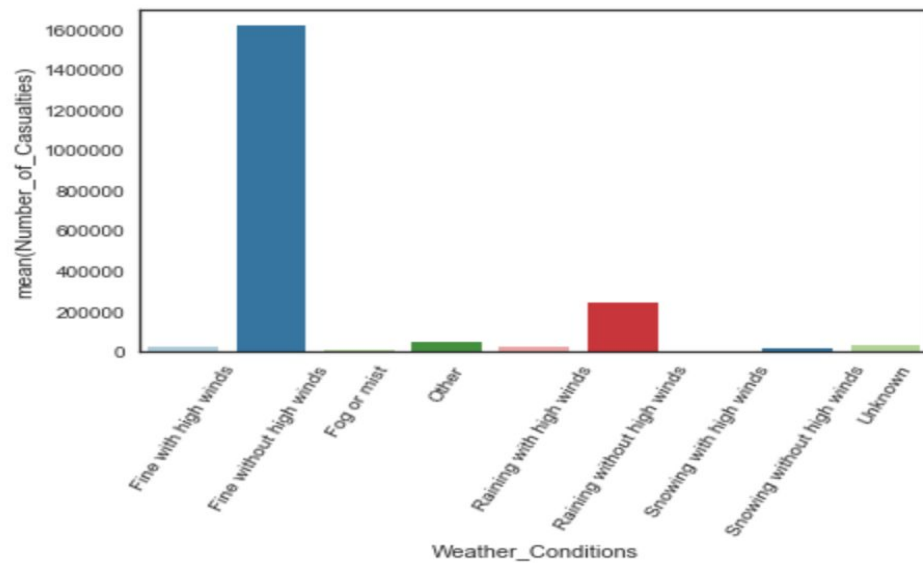
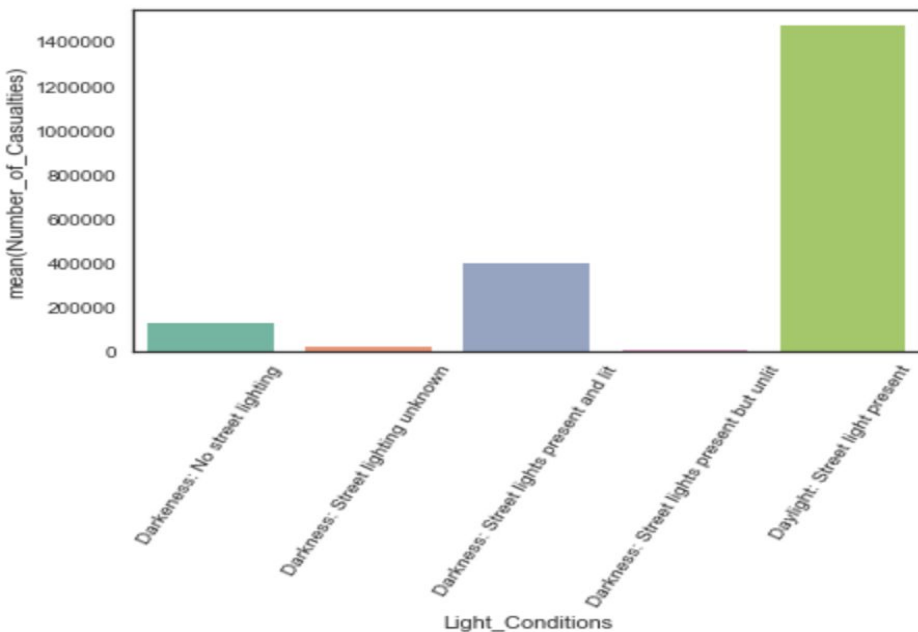
Exploratory Data Analysis

Question: In what junction control and pedestrian crossing human control most accidents happen?



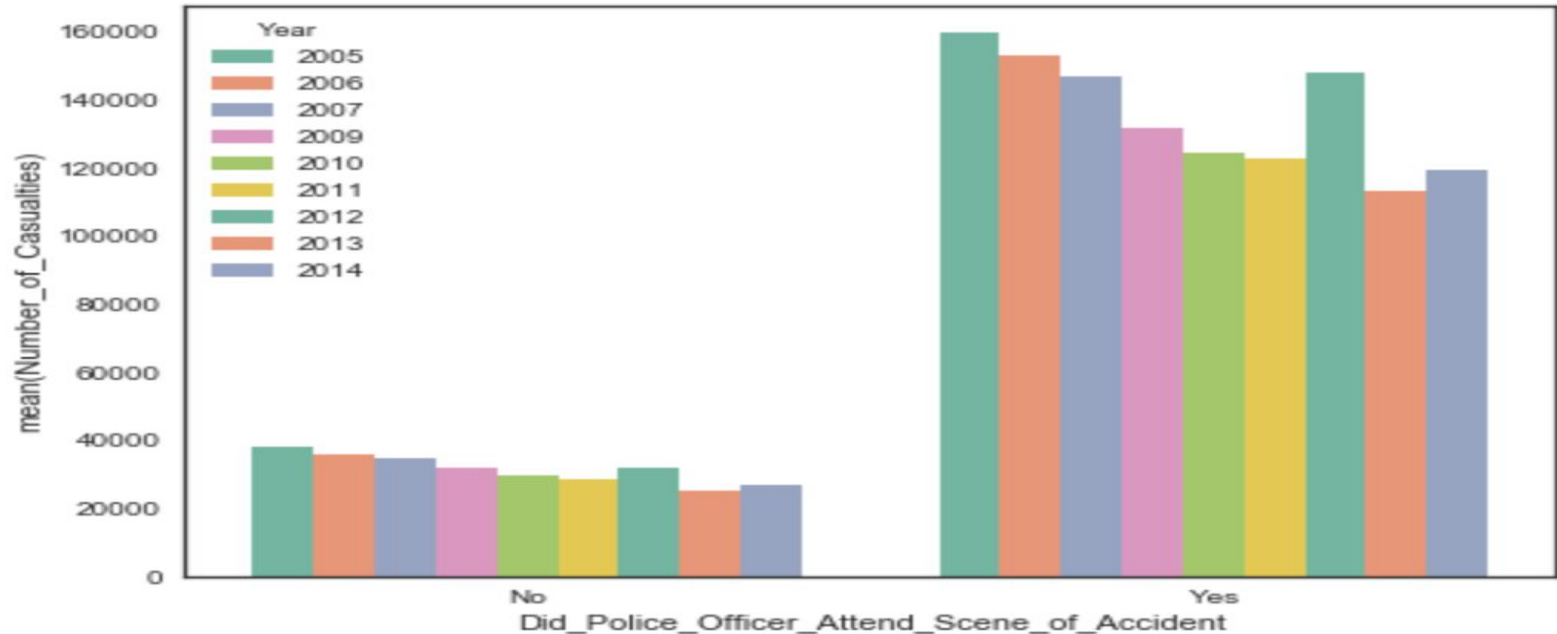
Exploratory Data Analysis

Question: In what light and weather conditions most accidents happen?



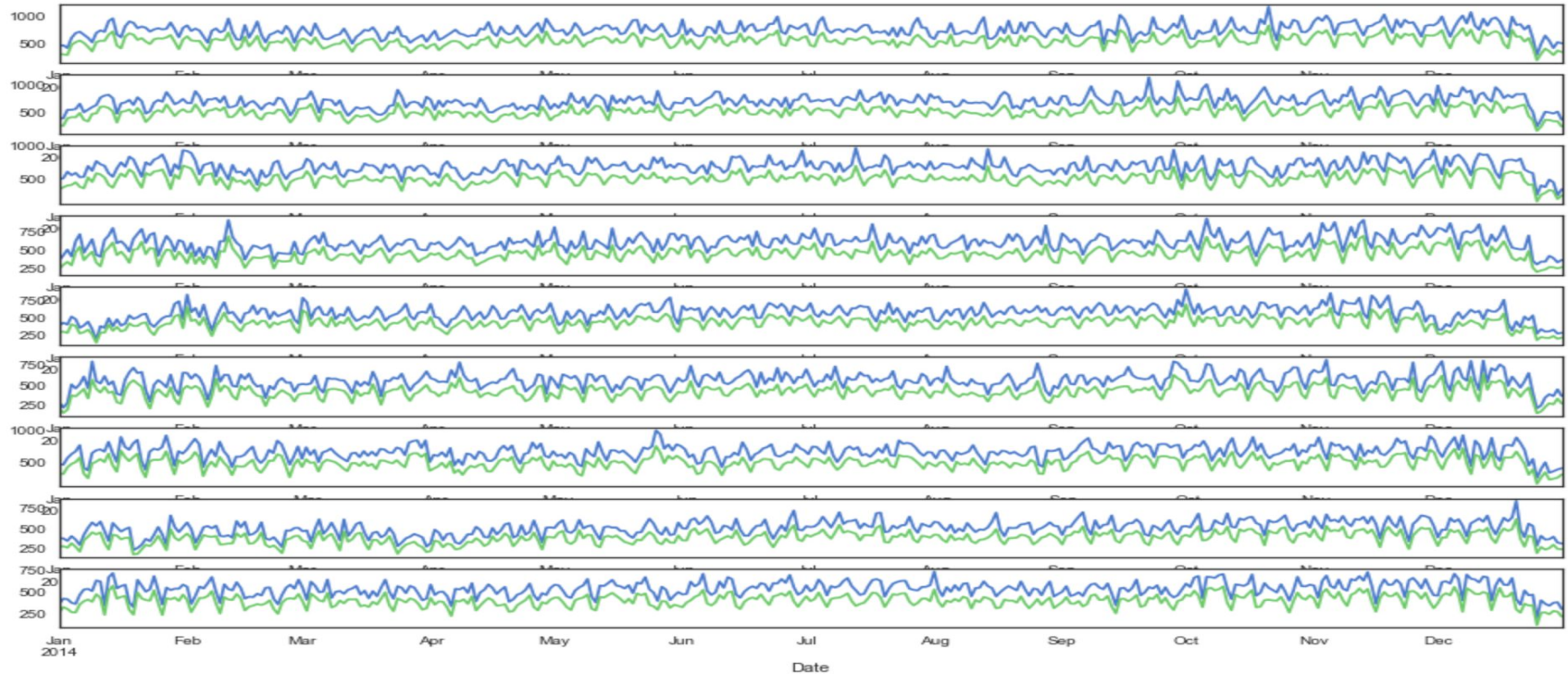
Exploratory Data Analysis

Question: How many accidents in every year police officers attended the scene of accident?



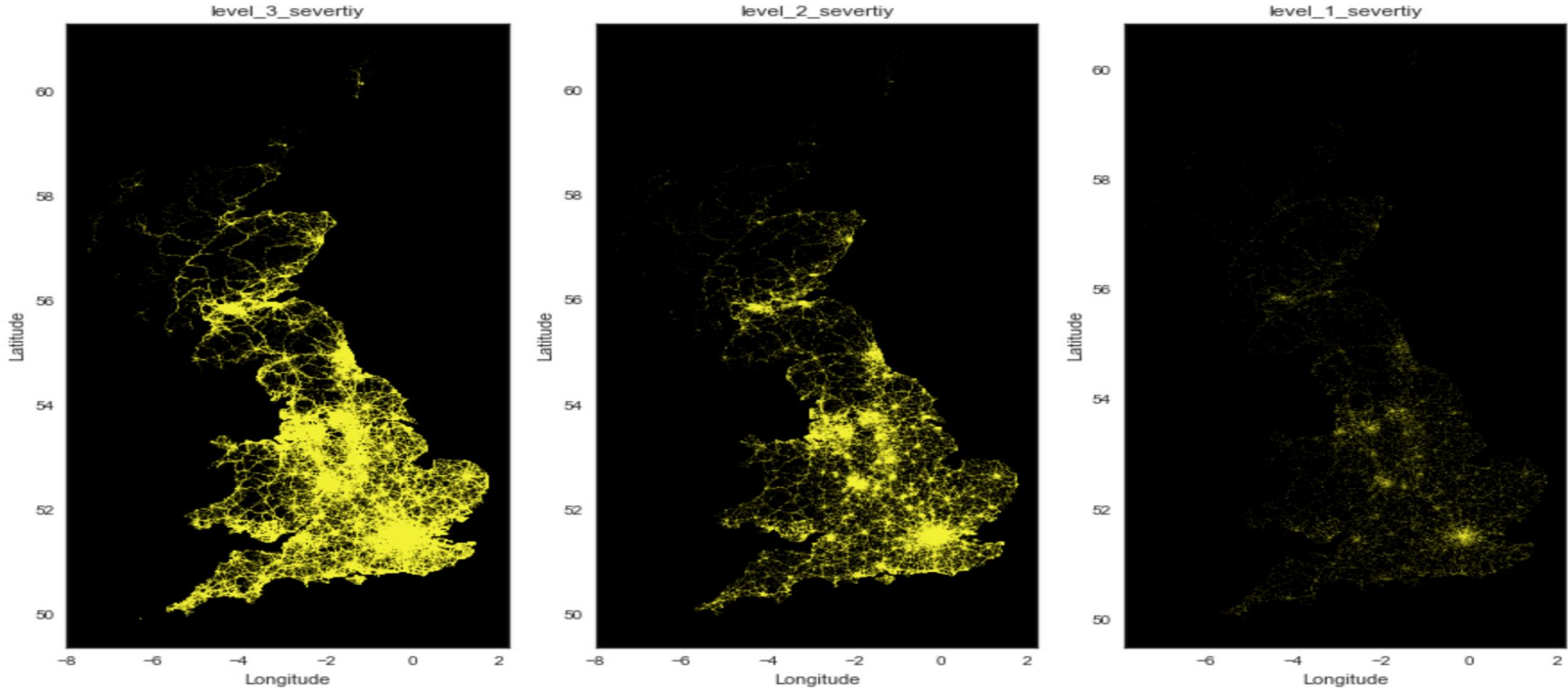
Exploratory Data Analysis

Question: Can we see pattern on the number of casualties happening every day in a yearly cycle?

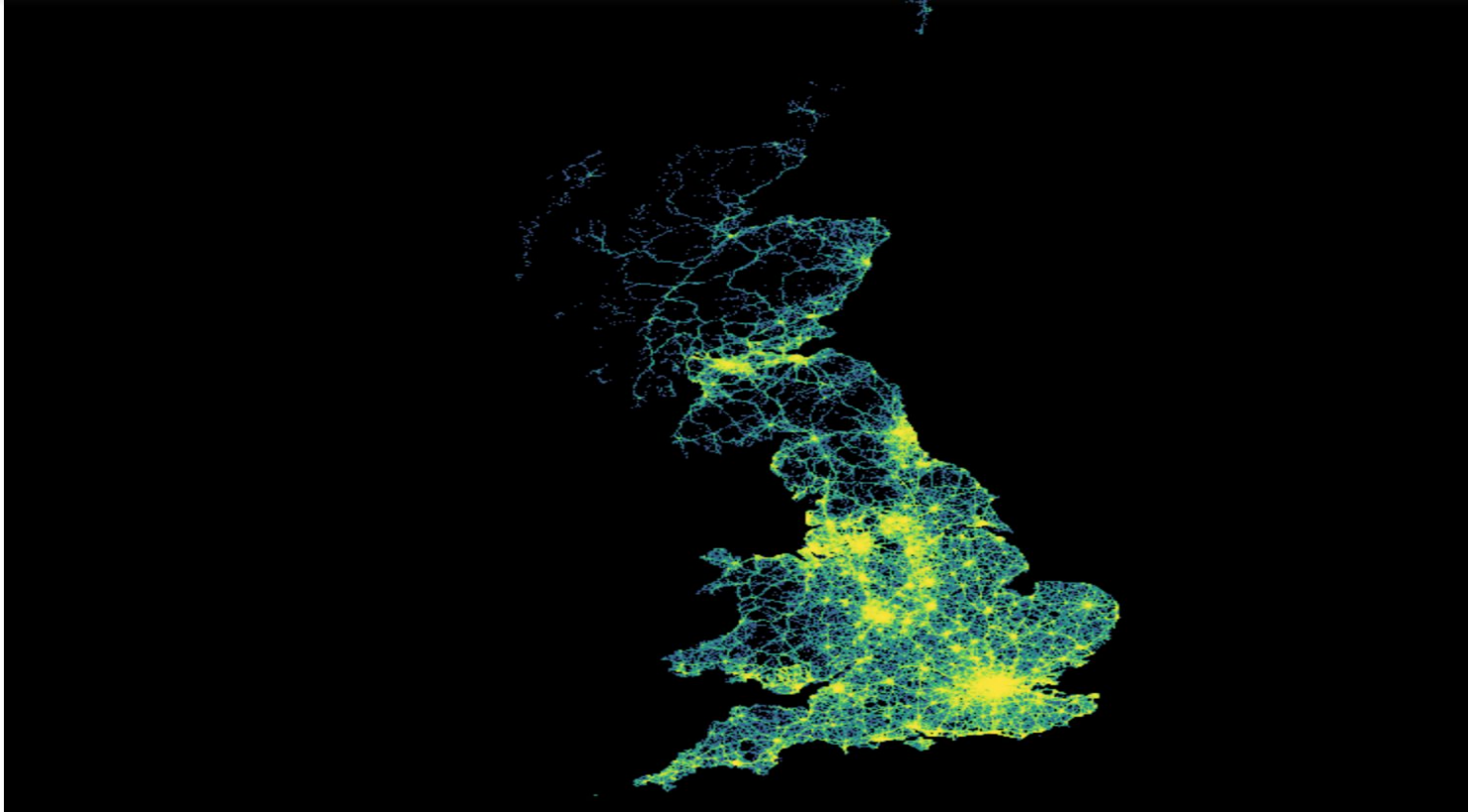


Exploratory Data Analysis

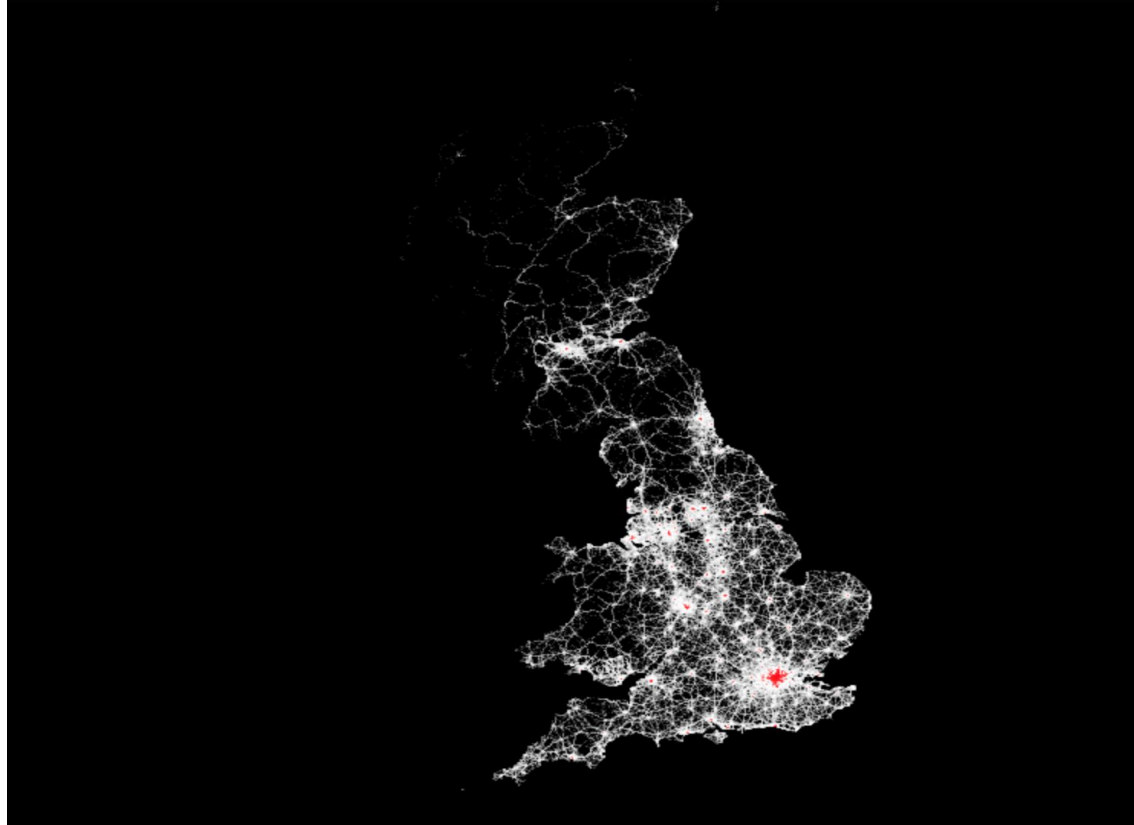
Question: What is the spread of different kinds of accidents across the nation?



Traffic Accident Density



Top 1 Traffic Accidents



Observation: we can clearly see the top 1% of accidents reported occur in these major cities:

England: London, Birmingham, Liverpool, Manchester, Leeds, Newcastle, Norwich, and Bristol

Wales: Cardiff

Scotland: Edinburgh, Glasgow, Aberdeen

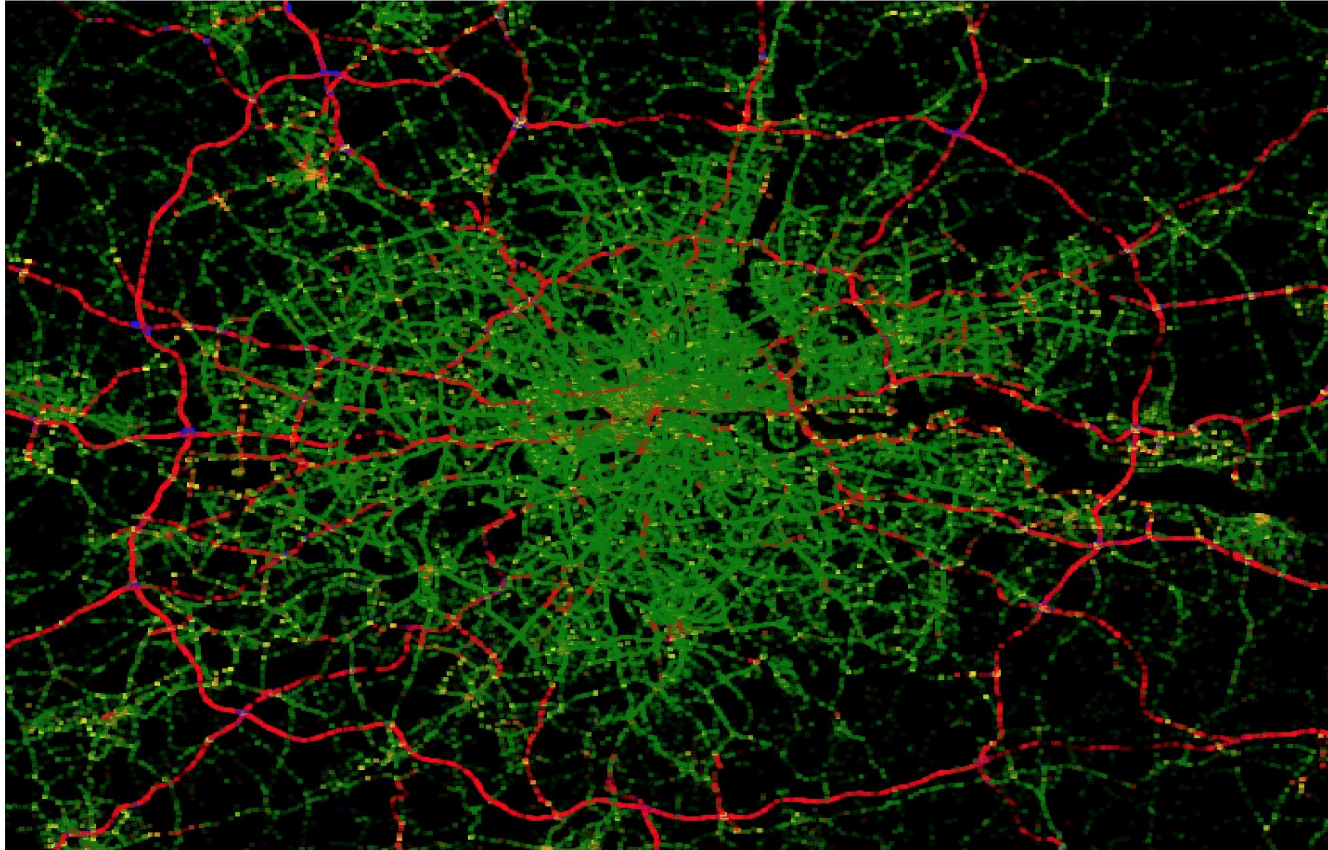
Traffic Accident Density In London Area



Traffic Accident Density In Manchester Area



Traffic Accident Density Based on Road Type In London (Interactive)



'Dual
carriageway': 'red'

'One way
street': 'orange',

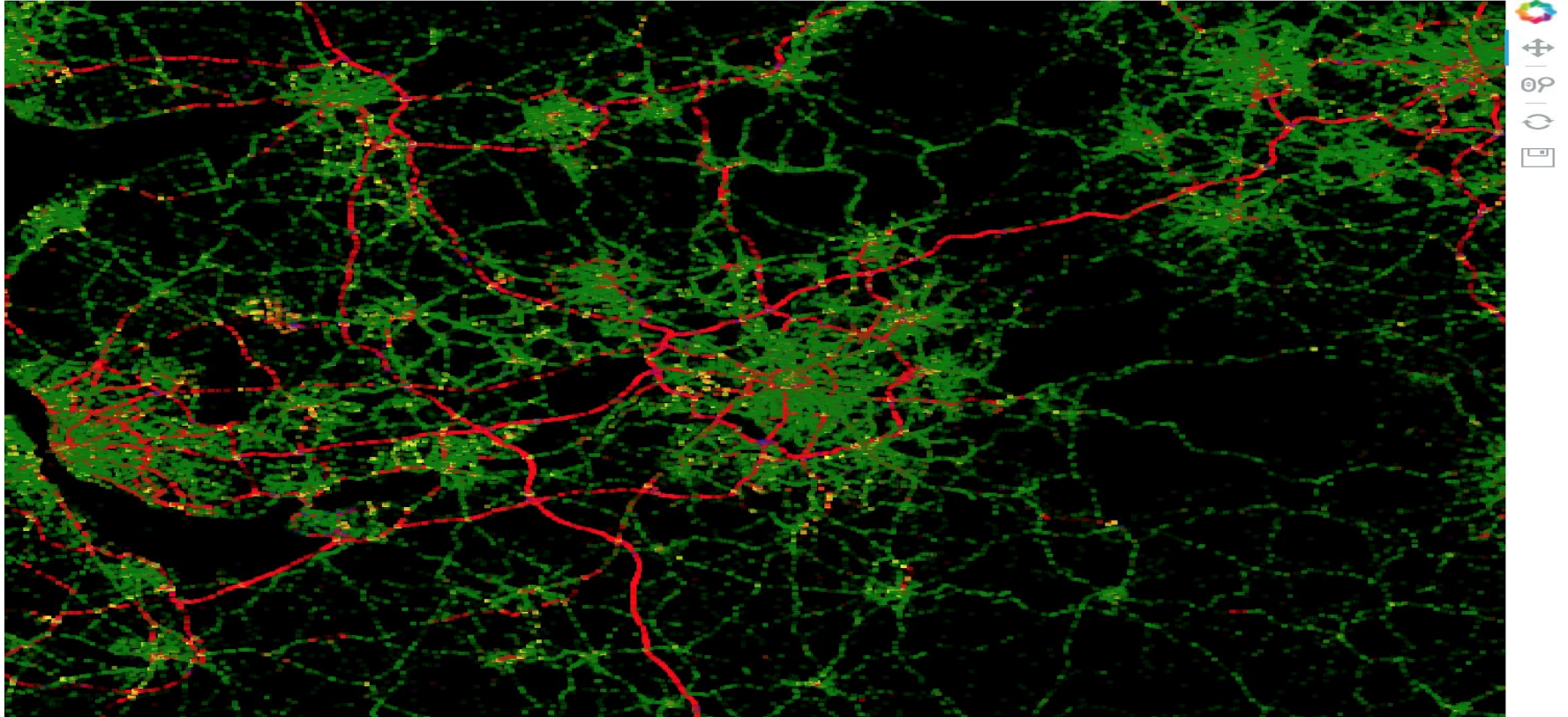
'Roundabout': 'yellow',

'Single
carriageway': 'green'

'Slip road': 'blue'

'Unknown': 'purple'

Traffic Accident Density Based on Road Type In Manchester (Interactive)



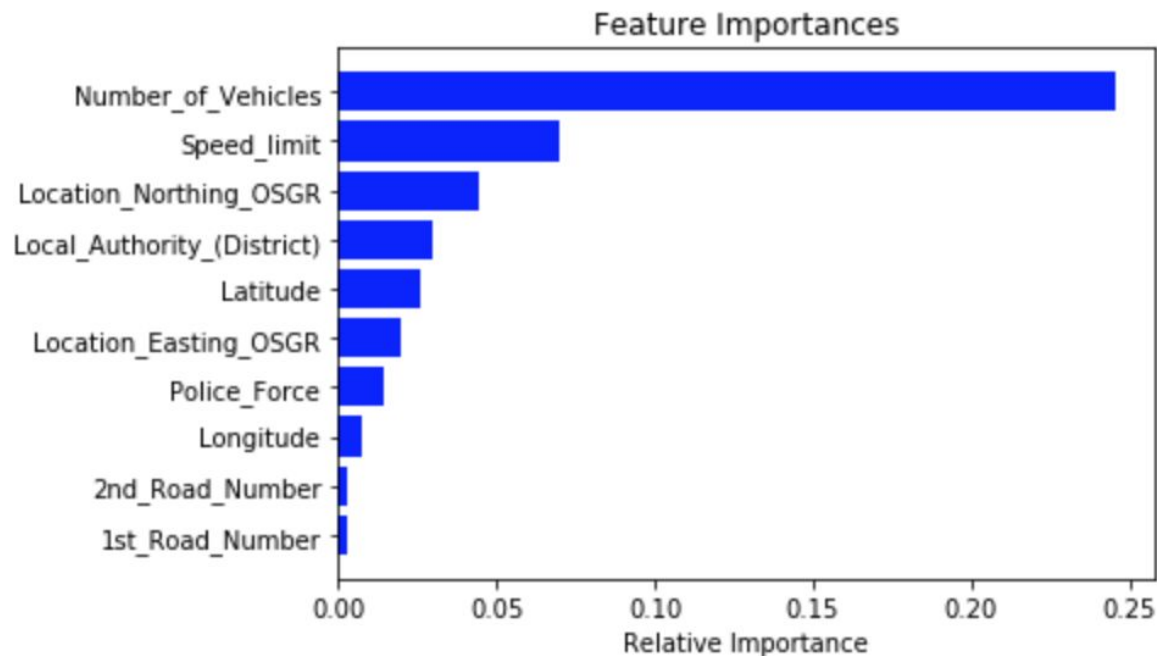
Predicting Number of Casualties

- Steps:**
- 1. Remove 'Junction_Detail' column because these are all None.**
 - 2. Binarize all categorical columns.**
 - 3. For easy modeling, drop rows with Null values in any of its columns. Consider interpolation on dealing with this data again.**
 - 4. Feature Set:**

['Location_Easting_OSGR', 'Location_Northing_OSGR', 'Longitude', 'Latitude', 'Police_Force', 'Number_of_Vehicles',
'Local_Authority_(District)', '1st_Road_Number', 'Speed_limit', '2nd_Road_Number', 'Year', 'month', 'day_of_month', 'day_of_year', 'hour',
'Local_Authority_(Highway)', 'XXXXXXX', 'Road_Type_Dual', 'XXXX', 'Junction_Control', 'XXXXX',
'Pedestrian_Crossing-Human_Control', 'XXXX', 'Pedestrian_Crossing-Physical_Facilities', 'XXXX',
'Light_Conditions', 'XXXX', 'Weather_Conditions', 'XXXX', 'Road_Surface_Conditions', 'XXXX', 'Did_Police_Officer_Attend_Scene_of_Accident_N
o', 'Did_Police_Officer_Attend_Scene_of_Accident_Yes', 'Urban_or_Rural_Area_1', 'Urban_or_Rural_Area_2', 'Accident_Severity_X',
'Accident_Severity_2', 'Accident_Severity_3', 'Day_of_Week_X', '1st_Road_Class', 'XXXX', '2nd_Road_Class', 'XXXX']

Predicting Number of Casualties

Regression Model	RMSE
Linear Regression	0.72
XGBoost	0.72



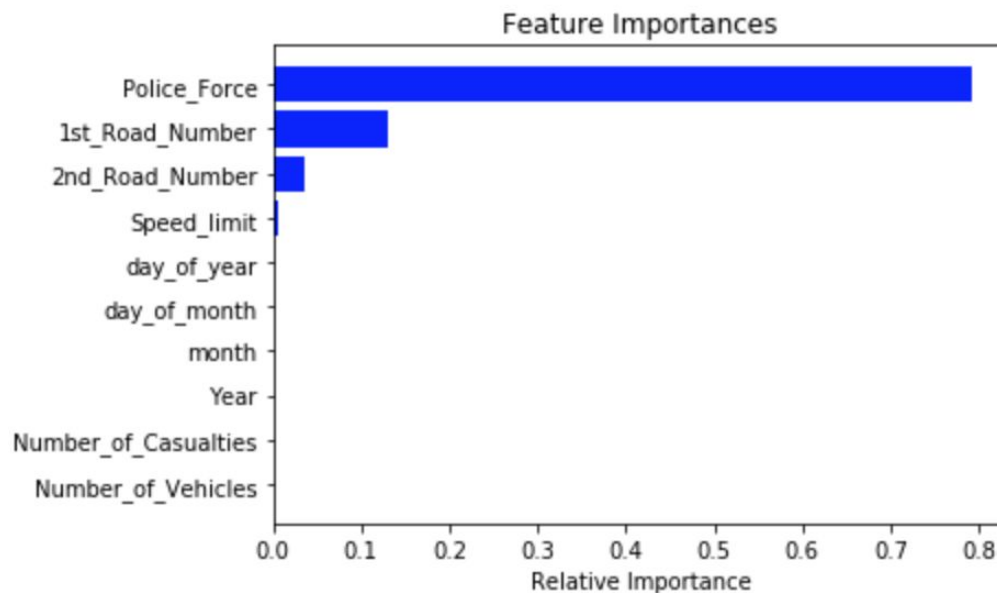
Predicting Latitude of the Accident

- Steps:**
- 1. Remove 'Junction_Detail' column because these are all None.**
 - 2. Binarize all categorical columns.**
 - 3. For easy modeling, drop rows with Null values in any of its columns. Consider interpolation on dealing with this data again.**
 - 4. Remove all features having some details on the location of the accident.**
 - 4. Feature Set:**

['Police_Force', 'Number_of_Vehicles', 'Number_of_Casualties', '1st_Road_Number', 'Speed_limit', '2nd_Road_Number', 'Year', 'month', 'day_of_month', 'day_of_year', 'hour', 'Road_Type_XXXX', 'Junction_Control_XXXX', 'Pedestrian_Crossing-Human_Control_XXXX', 'Pedestrian_Crossing-Physical_Facilities_XXXX', 'Light_Conditions_Darkness: XXXX', 'Weather_Conditions_XXXX', 'Road_Surface_Conditions_XXXX', 'Did_Police_Officer_Attend_Scene_of_Accident_No', 'Did_Police_Officer_Attend_Scene_of_Accident_Yes', 'Urban_or_Rural_Area_1', 'Urban_or_Rural_Area_2', 'Accident_Severity_1', 'Accident_Severity_2', 'Accident_Severity_3', 'Day_of_Week_X', '1st_Road_Class_X', '2nd_Road_Class_X']

Predicting Latitude of the Accident

Regression Model	RMSE
Linear Regression	0.95
XGBoost	0.15



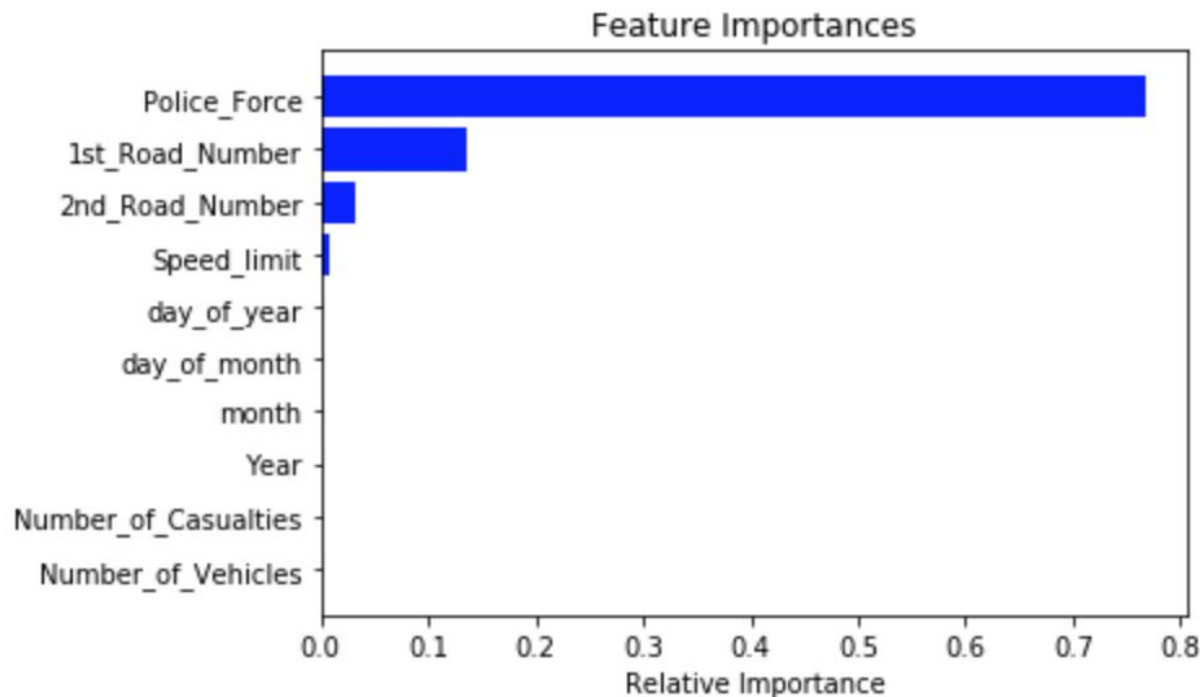
Predicting Longitude of the Accident

- Steps:**
- 1. Remove 'Junction_Detail' column because these are all None.**
 - 2. Binarize all categorical columns.**
 - 3. For easy modeling, drop rows with Null values in any of its columns. Consider interpolation on dealing with this data again.**
 - 4. Remove all features having some details on the location of the accident.**
 - 4. Feature Set:**

['Police_Force', 'Number_of_Vehicles', 'Number_of_Casualties', '1st_Road_Number', 'Speed_limit', '2nd_Road_Number', 'Year', 'month', 'day_of_month', 'day_of_year', 'hour', 'Road_Type_XXXX', 'Junction_Control_XXXX', 'Pedestrian_Crossing-Human_Control_XXXX', 'Pedestrian_Crossing-Physical_Facilities_XXXX', 'Light_Conditions_Darkness: XXXX', 'Weather_Conditions_XXXX', 'Road_Surface_Conditions_XXXX', 'Did_Police_Officer_Attend_Scene_of_Accident_No', 'Did_Police_Officer_Attend_Scene_of_Accident_Yes', 'Urban_or_Rural_Area_1', 'Urban_or_Rural_Area_2', 'Accident_Severity_1', 'Accident_Severity_2', 'Accident_Severity_3', 'Day_of_Week_X', '1st_Road_Class_X', '2nd_Road_Class_X']

Predicting Longitude of the Accident

Regression Model	RMSE
Linear Regression	1.19
XGBoost	0.25



Recommendations