Capstone Project
Car Accident Severity

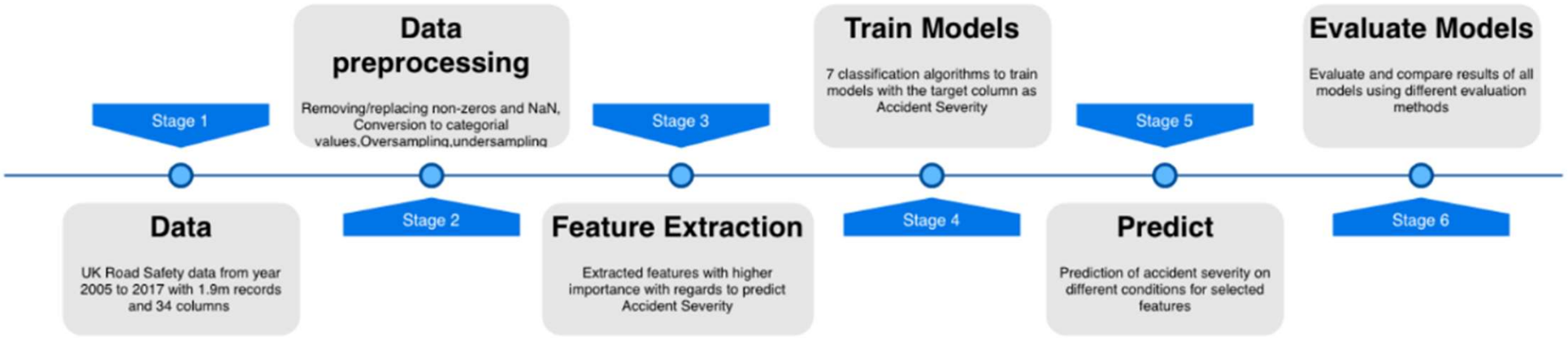
Predicting Car Accident Severity plays a crucial factor for various Government Departments/Authorities like Police, R&B and Transport to take proactive precautionary measures.

- Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole
- 93% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately 60% of the world's vehicles
- Hence the prediction aim for sustainable development, has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2021

Data Acquisition and Preprocessing

- Accident Dataset come from “Kaggle” public data platform, where they have been published
- The dataset contains 8849 records and 17 columns in Accident train and 1549 records and 17 columns in Accident_test
- Collision Severity attribute has three severity.
  - 3: Slight injury collisions is 89.23%, 2: Serious injury collision is 9.52% and 1: Fatal injury collision is 1.25%
- Negatively correlated features are selected to be dropped
- Data set is imputed by replacing NAN and missing values with most frequent values or central tendency

Methodology Flow Diagram

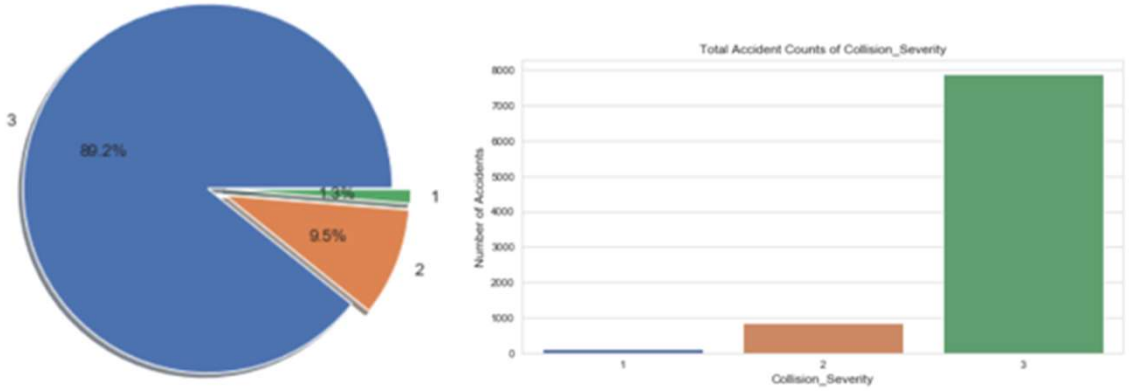


Exploratory Data Analysis

- Examining the Collision Severity

****Collision_Severity****

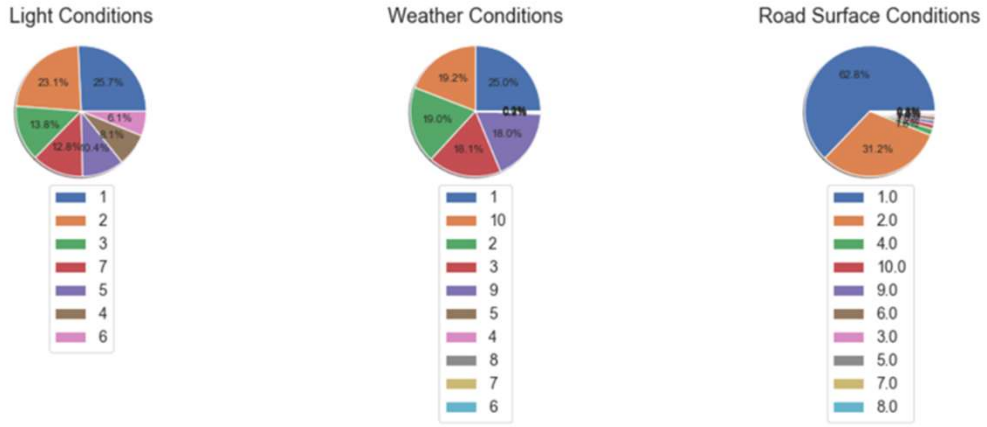
1: Fatal injury collision, 2: Serious injury collision and 3: Slight injury collision



Accident occurrences of Collision Severity

Exploratory Data Analysis

- Examining the Environment Conditions



Accident Occurrence of Environmental Conditions

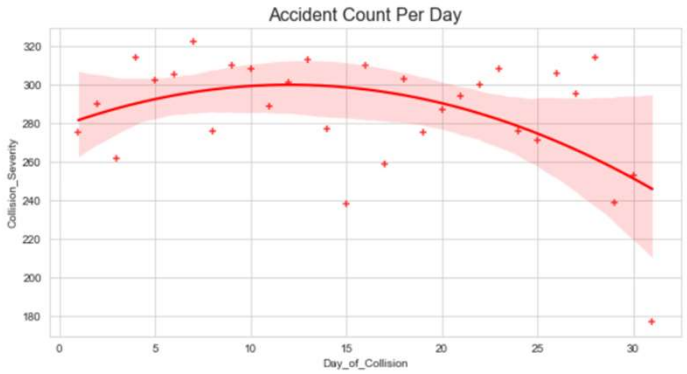
****Lighting Conditions**** 1 : Daylight : street lights present,2 : Daylight : no street lighting,3 : Daylight : street lighting unknown,4 : Darkness : street lights present and lit,5 : Darkness : street lights present but unlit,6 : Darkness : no street lighting and 7 : Darkness : street lighting unknown

****Weather Conditions**** 1 : Fine without high winds,2 : Raining without high winds,3 : Snowing without high winds,4 : Fine with high winds,5 : Raining with high winds,6 : Snowing with high winds,7 : Fog or mist - if hazard,8 : Strong sun (glaring),9 : Other and 10 : Unknown

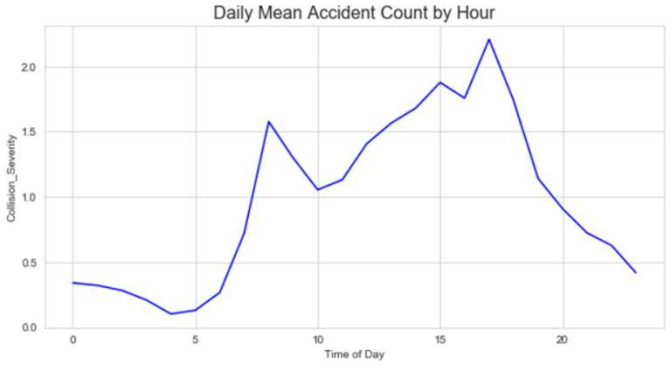
****Road_Surface_Conditions**** 1 : Dry, 2 : Wet / damp,3 : Snow,4 : Frost / ice,5 : Flood,6 : Oil,7 : Mud,8 : Leaves,9 : Slippery (after dry spell) and 10 : Other

Exploratory Data Analysis

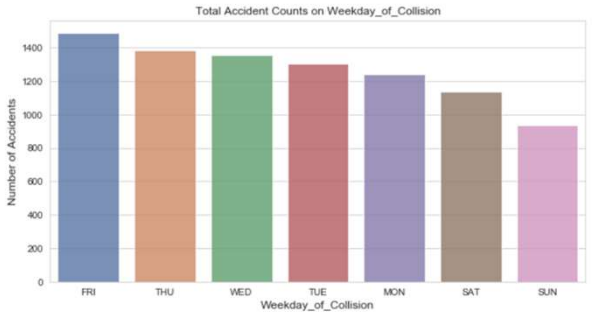
- Accident Distribution across Time and Cities



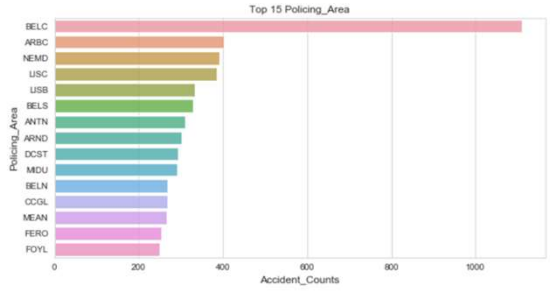
Day wise Distribution of Accidents throughout the year



Hourly Distribution of Accidents throughout the day



Accidents Occurrences of Weekday

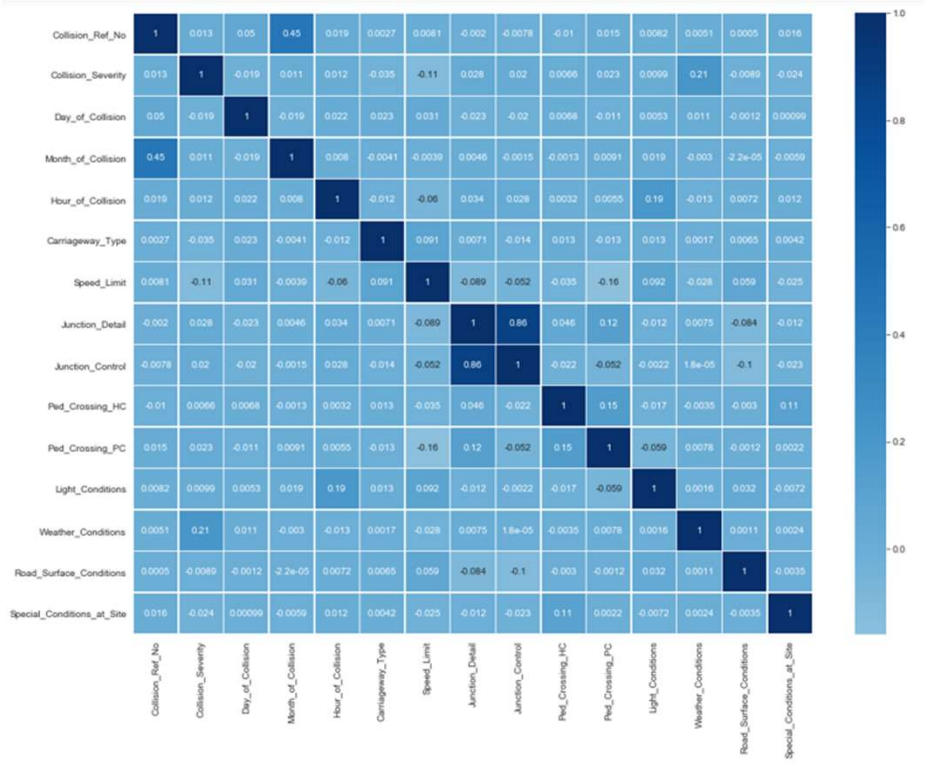


Accidents occurrences in Top 15 Cities

- An outlier in late December
- Peaks during the morning and evening rush hours
- High on Friday
- Top1 City is "BELC: Belfast City"

Exploratory Data Analysis

- Correlation Matrix



Correlation Matrix

"Collision severity has +ve correlation with Weather conditions -ve correlation with Speed limit"

Machine Learning Models

- XGB Classifier
- Random Forest Classifier
- Support Vector Machine
- Logistic Regression
- Gaussian Naive Bayes

Model	Score	Count
XGBoost	88.64	{3: 1664, 2: 104, 1: 2}
Support Vector Machines	88.59	{3: 1770}
Logistic Regression	88.59	{3: 1770}
Random Forest	88.42	{3: 1740, 2: 29, 1: 1}
Naive Bayes	21.41	{1: 1365, 2: 43, 3: 362}

Report Table

Conclusion

- Objectives of the investigation were met
- Road traffic injuries can be prevented with proactive measures
- More scope to investigated further, such as how junction layout or vehicle type relate to collision rates in different conditions.