Capstone Project Week 3

Table of contents

1. Introduction / Business Problem

On September 11, 2001, the twin towers of the World Trade Center were destroyed and history recorded that over 3,000 people were killed. Not many people know that about the same number of people die every day on roads world wide1. This figure does not include at least the 30,000 others injured or disabled.¶

This accumulates to over 1 million people killed and between 20–50 million injured or crippled in road accidents each year2. It is obvious that road traffic injuries are a major public health problem globally. In fact, as projected by the WHO that road traffic disability-adjusted life years (DALYs) loss will move from being the ninth leading cause of DALYs in 1999 to the third leading cause by year 2020.

Traffic accidents incur immense losses to individuals, families and the country by being the cause of so many untimely deaths, debilitating injuries, damage to properties and loss in productivity. In addition to economic losses traffic accidents have a social component in that victims and/or their families are often beset with grief, hardship and even a degraded quality of life.

Key facts from World Health Organization

- Approximately 1.35 million people die each year as a result of road traffic crashes.
- Road traffic crashes cost most countries 3% of their gross domestic product.
- More than half of all road traffic deaths are among vulnerable road users: pedestrians, cyclists, and motorcyclists.

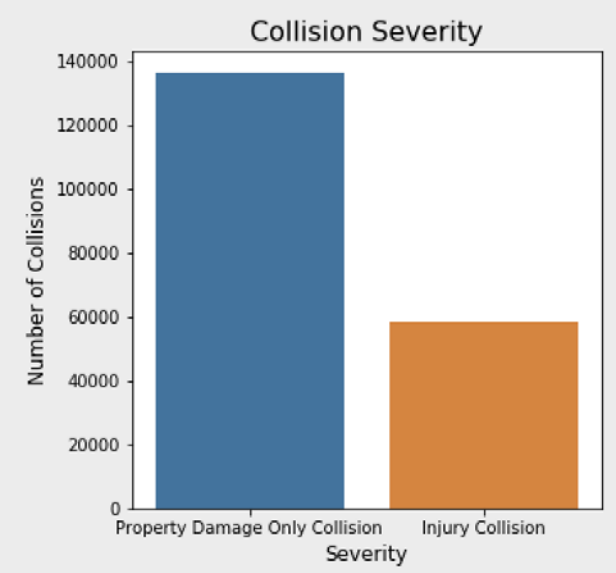
The project objective is to study and understand the important variable accidents occur which would help us (Government / People) to reduce the number of deaths and injuries from road traffic crashes in the future

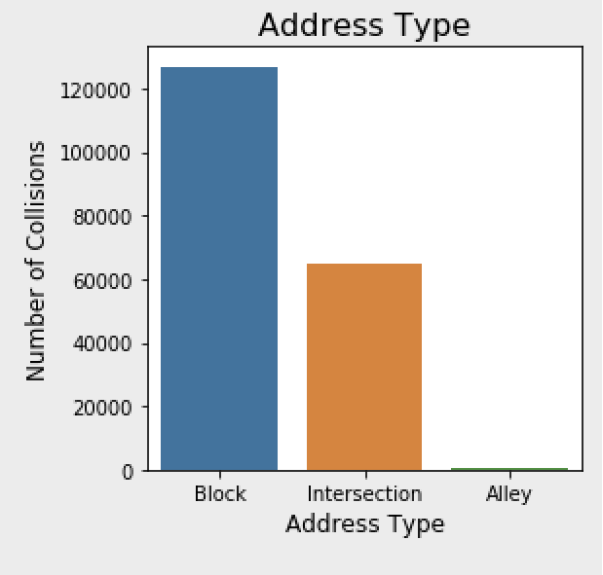
2. Data

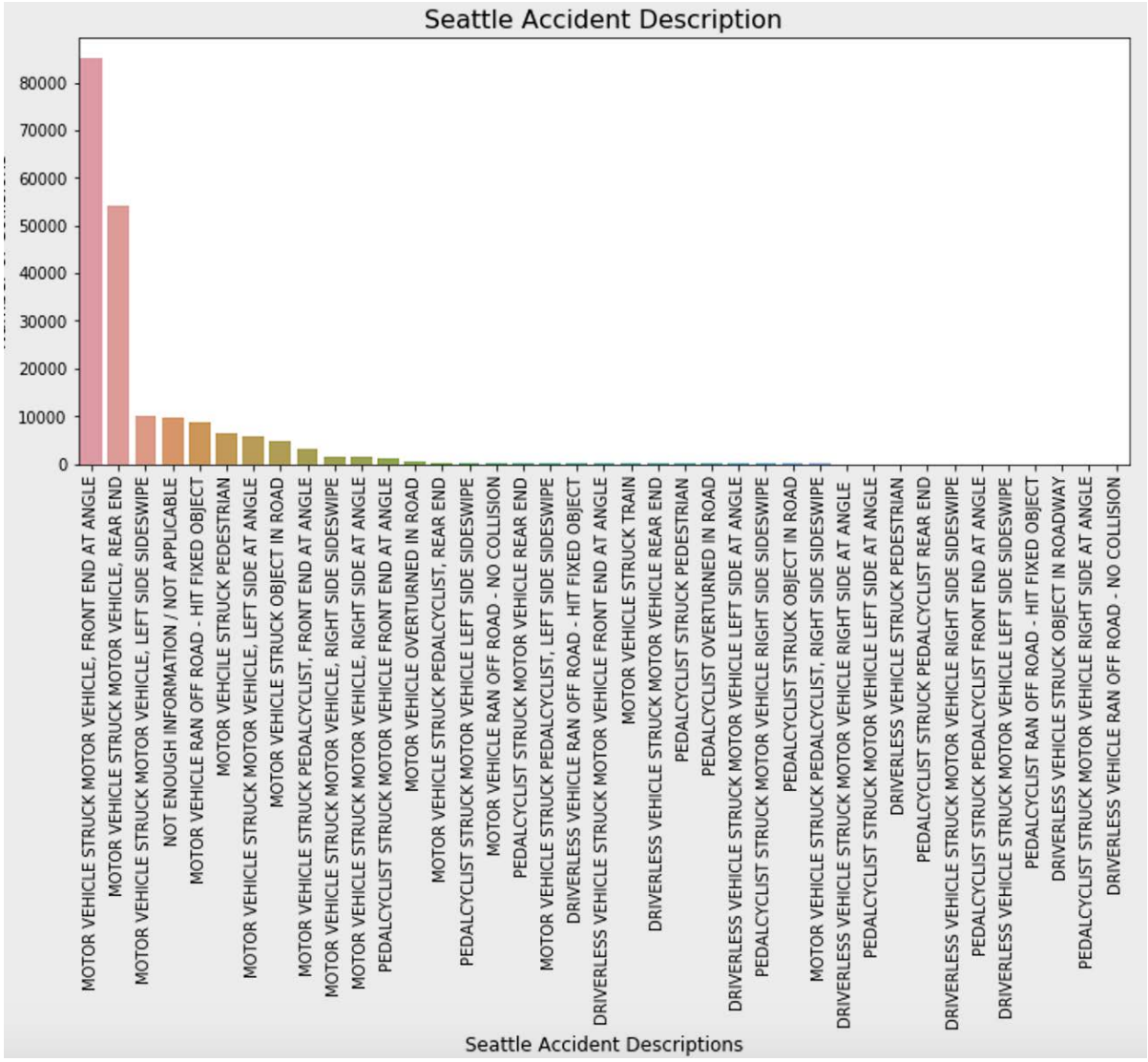
The data used is the Seattle's Department of Transportation and recorded by Traffic Records.

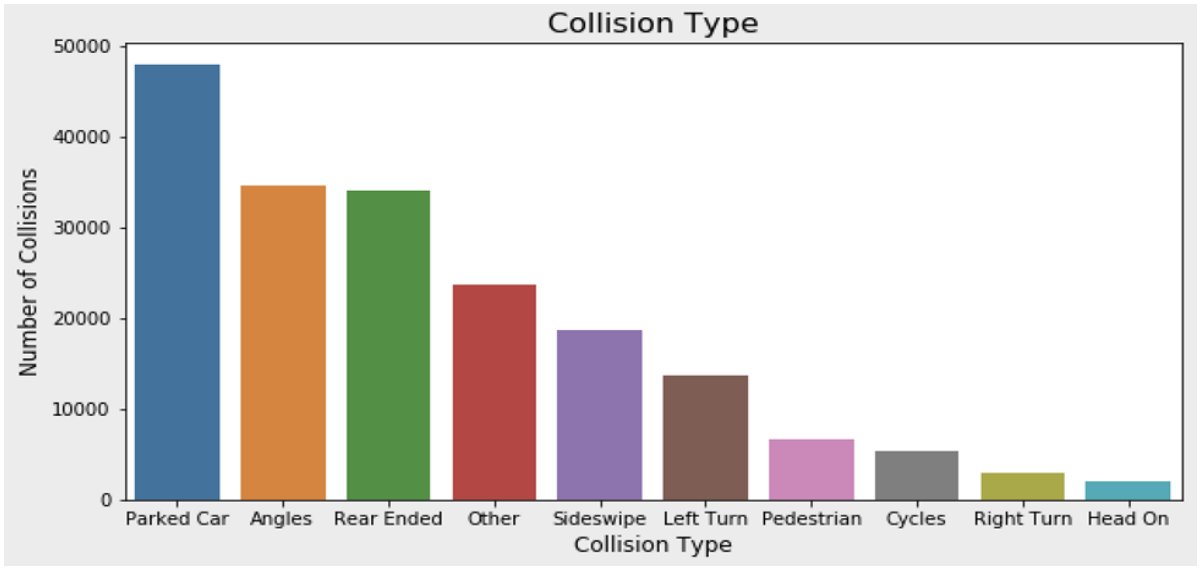
The data used is storing and displaying data related to positions on Earth's surface from 2004 to 2020. The data contains various features such as location, the severity of the collision, number of vehicles/cyclists/pedestrians involved, date/time of incident, weather, road conditions and more.

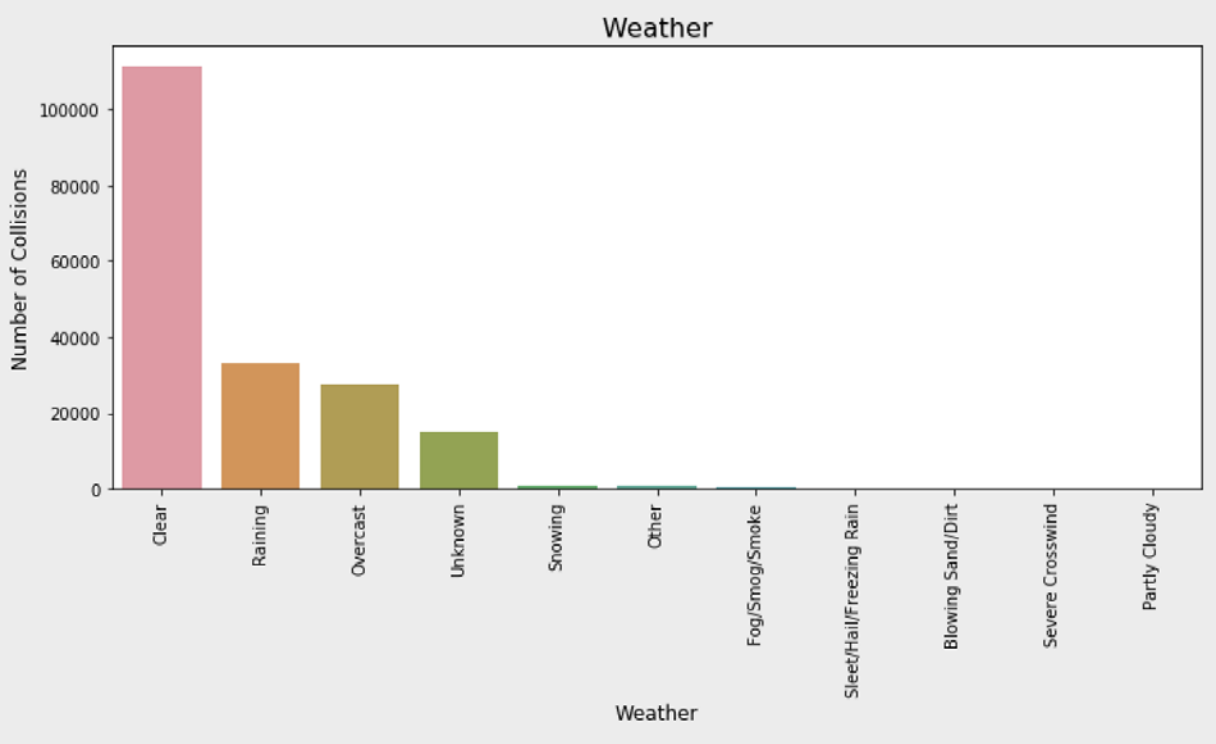
This study look at major environment and use variable only related to ADDRTYPE, COLLISIONTYPE, WEATHER, ROADCOND, LIGHTCOND and JUNCTIONTYPE.

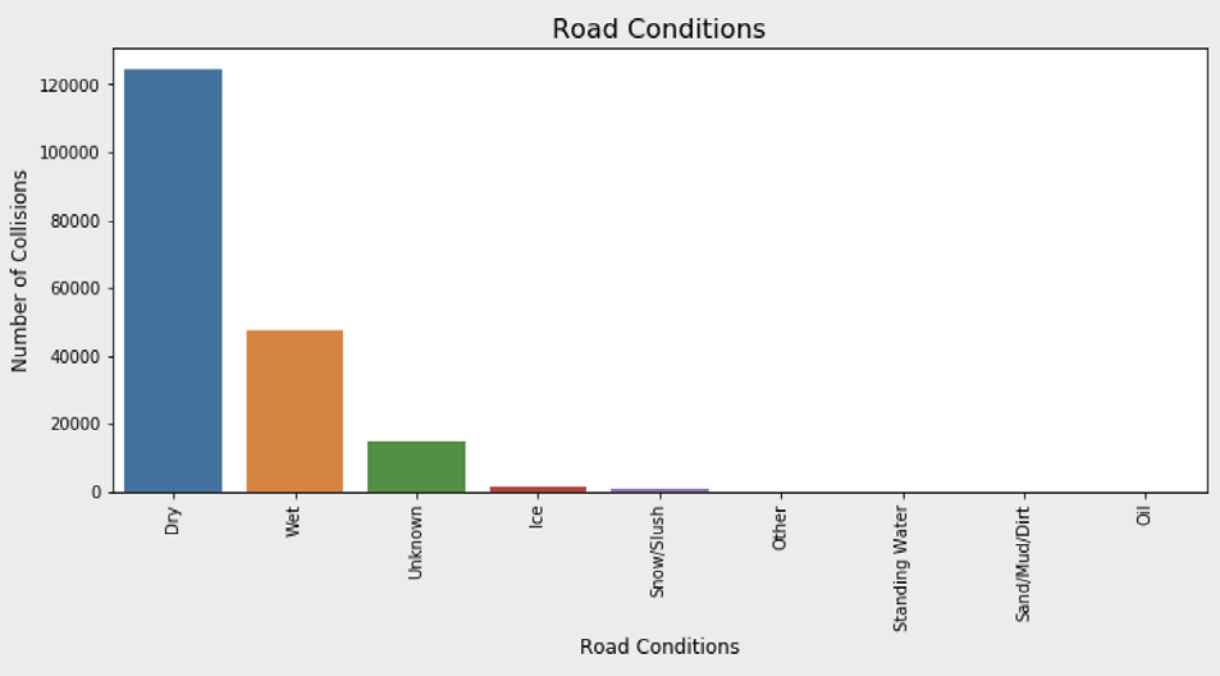


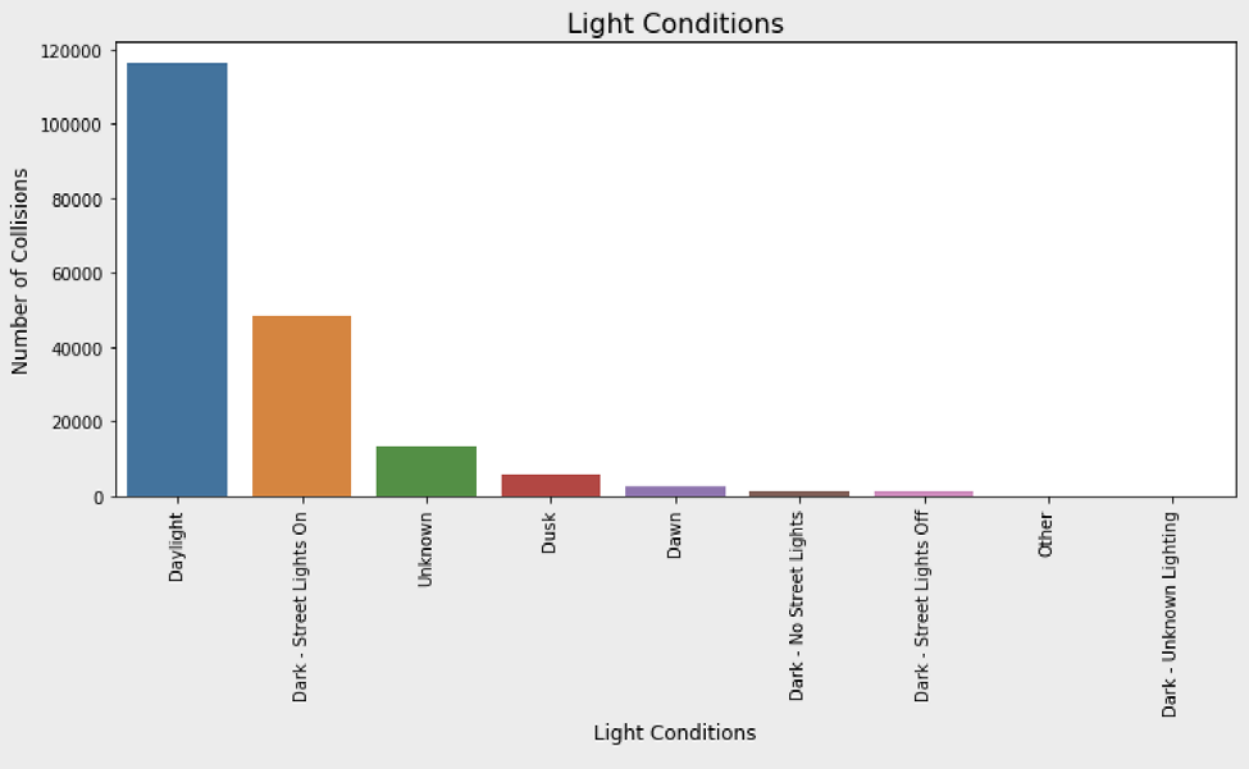


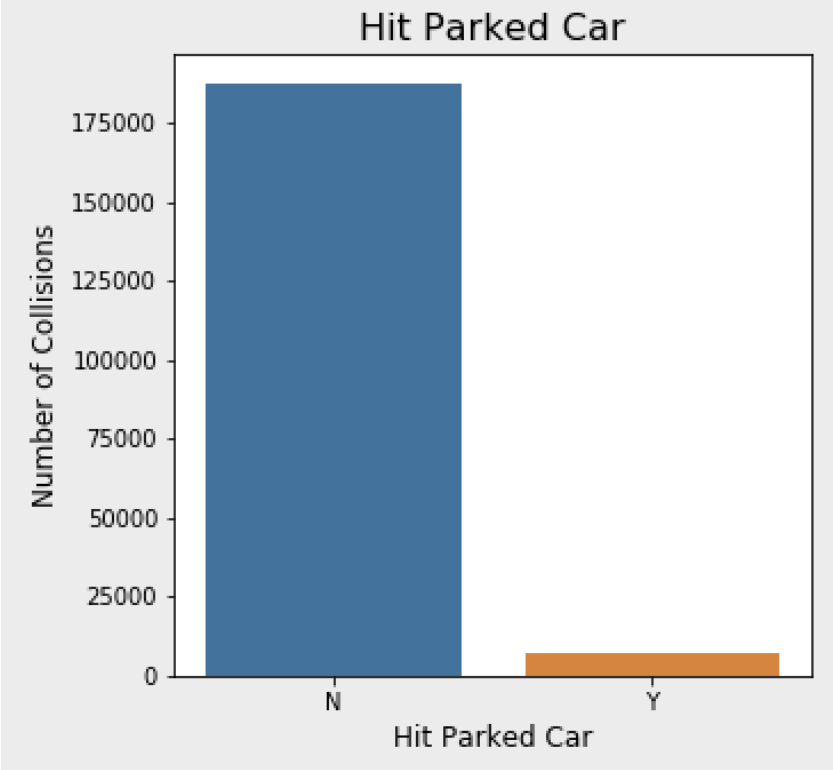

Seattle Accident Description
80000
70000
60000
50000
40000
30000
20000
10000
0
MOTOR VEHICLE STRUCK MOTOR VEHICLE, FRONT END AT ANGLE
MOTOR VEHICLE STRUCK MOTOR VEHICLE, REAR END
MOTOR VEHICLE STRUCK MOTOR VEHICLE, LEFT SIDE SIDESWIPE
NOT ENOUGH INFORMATION / NOT APPLICABLE
MOTOR VEHICLE RAN OFF ROAD - HIT FIXED OBJECT
MOTOR VEHICLE STRUCK PEDESTRIAN
MOTOR VEHICLE STRUCK MOTOR VEHICLE, LEFT SIDE AT ANGLE
MOTOR VEHICLE STRUCK OBJECT IN ROAD
MOTOR VEHICLE STRUCK PEDALCYCLIST, FRONT END AT ANGLE
MOTOR VEHICLE STRUCK MOTOR VEHICLE, RIGHT SIDE SIDESWIPE
MOTOR VEHICLE STRUCK MOTOR VEHICLE, RIGHT SIDE AT ANGLE
PEDALCYCLIST STRUCK MOTOR VEHICLE FRONT END AT ANGLE
MOTOR VEHICLE OVERTURNED IN ROAD
MOTOR VEHICLE STRUCK PEDALCYCLIST, REAR END
PEDALCYCLIST STRUCK MOTOR VEHICLE LEFT SIDE SIDESWIPE
MOTOR VEHICLE RAN OFF ROAD - NO COLLISION
PEDALCYCLIST STRUCK MOTOR VEHICLE REAR END
MOTOR VEHICLE STRUCK PEDALCYCLIST, LEFT SIDE SIDESWIPE
DRIVERLESS VEHICLE RAN OFF ROAD - HIT FIXED OBJECT
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE FRONT END AT ANGLE
MOTOR VEHICLE STRUCK TRAIN
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE REAR END
PEDALCYCLIST STRUCK PEDESTRIAN
PEDALCYCLIST OVERTURNED IN ROAD
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE LEFT SIDE AT ANGLE
PEDALCYCLIST STRUCK MOTOR VEHICLE RIGHT SIDE SIDESWIPE
PEDALCYCLIST STRUCK OBJECT IN ROAD
MOTOR VEHICLE STRUCK PEDALCYCLIST, RIGHT SIDE SIDESWIPE
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE RIGHT SIDE AT ANGLE
PEDALCYCLIST STRUCK MOTOR VEHICLE LEFT SIDE AT ANGLE
DRIVERLESS VEHICLE STRUCK PEDESTRIAN
PEDALCYCLIST STRUCK PEDALCYCLIST REAR END
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE RIGHT SIDE SIDESWIPE
PEDALCYCLIST STRUCK PEDALCYCLIST FRONT END AT ANGLE
DRIVERLESS VEHICLE STRUCK MOTOR VEHICLE LEFT SIDE SIDESWIPE
PEDALCYCLIST RAN OFF ROAD - HIT FIXED OBJECT
DRIVERLESS VEHICLE STRUCK OBJECT IN ROADWAY
PEDALCYCLIST STRUCK MOTOR VEHICLE RIGHT SIDE AT ANGLE
DRIVERLESS VEHICLE RAN OFF ROAD - NO COLLISION
Seattle Accident Descriptions


Collision Type
50000
40000
30000
20000
10000
0
Number of Collisions
Parked Car
Angles
Rear Ended
Other
Sideswipe
Left Turn
Pedestrian
Cycles
Right Turn
Head On
Collision Type


Weather
Number of Collisions
0
20000
40000
60000
80000
100000
Clear
Raining
Overcast
Unknown
Snowing
Other
Fog/Smog/Smoke
Sleet/Hail/Freezing Rain
Blowing Sand/Dirt
Severe Crosswind
Partly Cloudy
Weather


Road Conditions
Number of Collisions
0
20000
40000
60000
80000
100000
120000
Dry
Wet
Unknown
Ice
Snow/Slush
Other
Standing Water
Sand/Mud/Dirt
Oil
Road Conditions


Light Conditions
Number of Collisions
120000
100000
80000
60000
40000
20000
0
Daylight
Dark - Street Lights On
Unknown
Dusk
Dawn
Dark - No Street Lights
Dark - Street Lights Off
Other
Dark - Unknown Lighting
Light Conditions


Hit Parked Car
Number of Collisions
175000
150000
125000
100000
75000
50000
25000
0
N
Y
Hit Parked Car

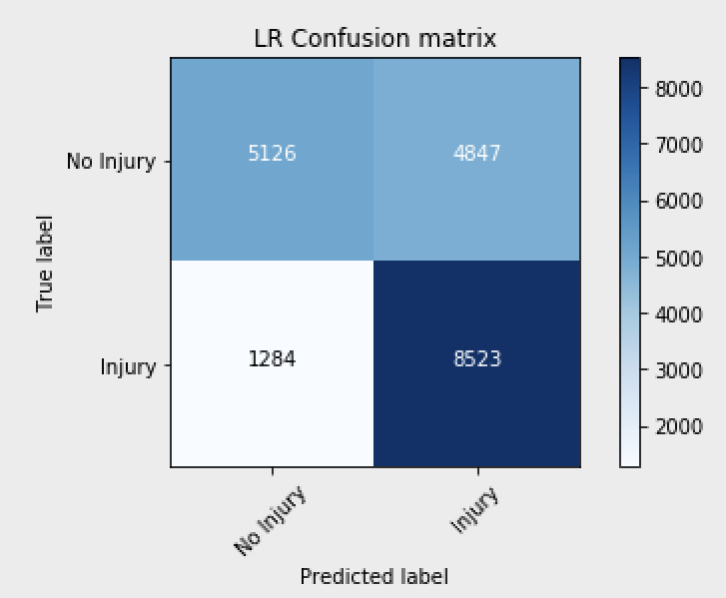
3. Methodology

Data preparation
Created subset of collision environment data. (ADDRTYPE, COLLISIONTYPE, WEATHER, ROADCOND, LIGHTCOND and JUNCTIONTYPE)
Insert value to missing data.
Clear or change data if necessary.

Modeling
Run Logistic Regression model
Run K-Nerrest model
Run Decision Tree model
Run Random Forest Classifier model
Run Support Vector Machine Classifier medel

Logistic Regression

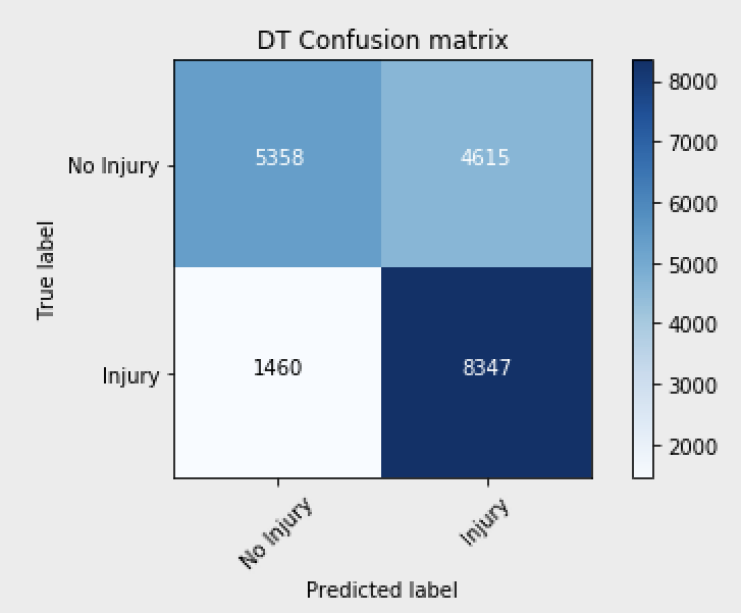
Logistic Regression is a variation of Linear Regression, useful when the observed dependent variable, y, is categorical. It produces a formula that predicts the probability of the class label as a function of the independent variables. After Run Logistic Regression, we found Confusion Matrix as follow :



Decision Tree

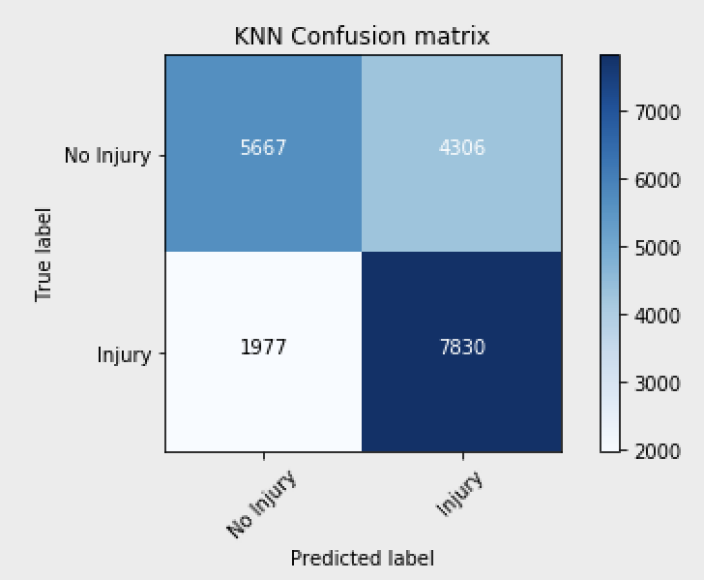
A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one

way to display an algorithm that only contains conditional control statements. After run Dicision Tree, we found Confusion Matrix as Follow:



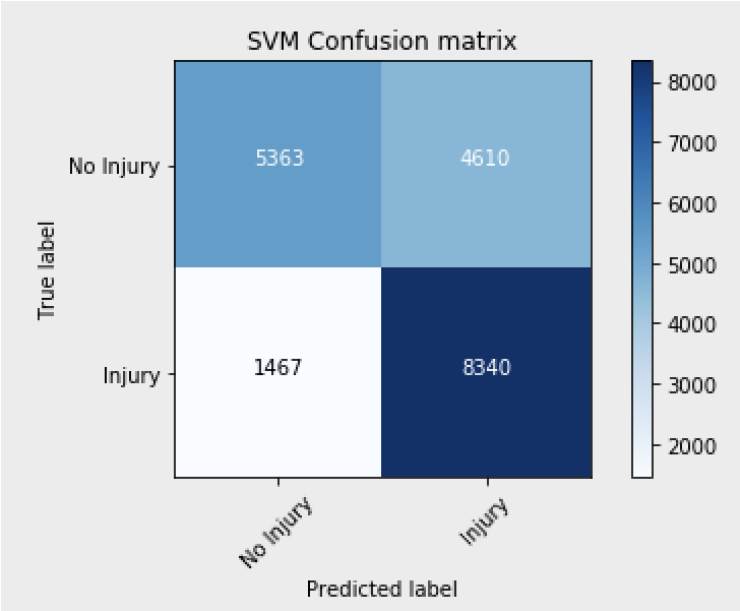
K-Nearest Neighbors (KNN)

K-Nearest Neighbors is an algorithm for supervised learning. Where the data is 'trained' with data points corresponding to their classification. Once a point is to be predicted, it takes into account the 'K' nearest points to it to determine it's classification. After run K-Nearest Neighbors, we found Confusion Matrix as follow:



Support Vector Machine (SVM)

SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data is transformed in such a way that the separator could be drawn as a hyperplane. After run Support Vector Machin, we found Confusion Matrix as follow:



4. Results

After run Logistic Regression, Decision Tree, K-Nearest Neighbors and Support Vector Machine models, we found that immaterial of average Jaccard and F1 Score between each model. Each score over 0.68.

Algorithm	Jaccard	F1-score
Logistic Regression	0.6900	0.6802
Decision Tree	0.6929	0.6853
K-Nearest Neighbors	0.6882	0.6876
Support Vector Machine	0.6900	0.6802

5. Discussion

Important features of collisions is Park Car. However Jaccard and F1-Score are not over 0.7. In my opinoin, the study features in this model don’t cover every factor of collisions. We Know that the collisions factor are not only the environment. Car Performance, Driver ability and habit also the maim factor of collisions.

6. Conclusion

This study can't cover every factor of collisions and need more study. however this study should aware the goverment and people to find out the strategy to reduce the number of collisions. such as:

* Alert people in area or environment that more opportunity to have collisions and accident.
* Set car performant standard.
* Check or test ability of drivers.
* change dangerous environment.
* Set emergency team to clear the dangerous factor.