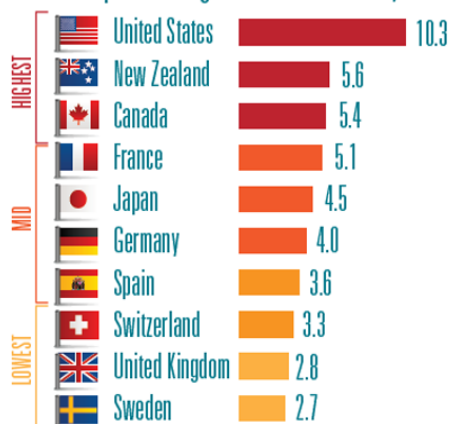


Indroduction :

In recent year ther is certain increase in accident. This cause serious damage in property , parts of body and even life.So to understand and find the real cause ,we have to work on certain data and find pattern to resolve the problem or to minimise it.By this project our project is to increase the chances of safty and will be used for various industries. This analysis has multiple applications like driving suggestion ,carefull analysis of fatal accident, depending on the weather and road conditions on any given day.This also contains their employers, insurance firms, emergency and health care personal. This is the standard data published by Seattle's police.It would be great if real-time conditions can be provided to estimate the trip safeness.

Road traffic deaths in the US and other high-income countries.

Motor vehicle crash deaths in 10 comparison high-income countries, 2013



Deaths per 100,000 people
SOURCE: WHO Global Status Report on Road Safety, 2015.

Countries with the highest and lowest reductions in crash deaths, 2000-2013

Spain (highest)

↘ 75%

AVERAGE

of 19 high-

income countries

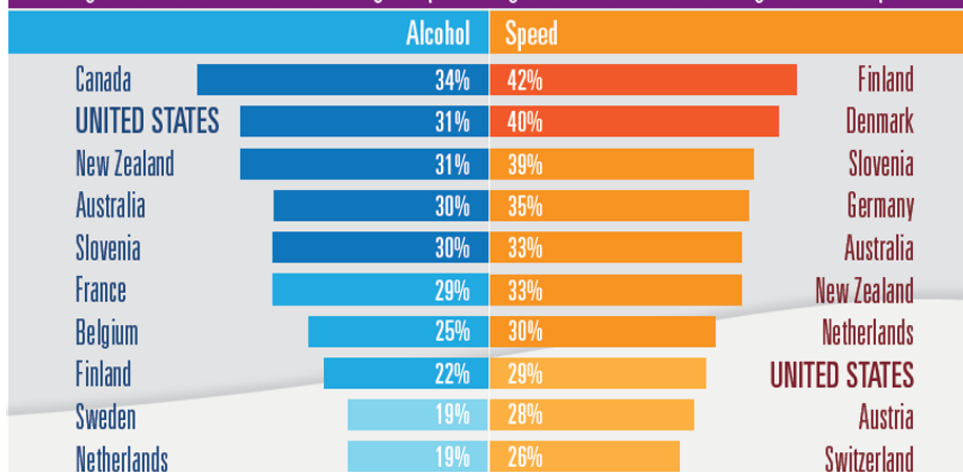
↘ 56%

United States (lowest)

↘ 31%

Deaths per 100,000 people
SOURCE: International Road Traffic and Accident Database (IRTAD) Road Safety Annual Report, 2015.

High-income countries with the highest percentage of crash deaths involving alcohol or speed



Note: 19 countries reported the percentage of deaths involving alcohol and 15 countries reported the percentage of deaths related to speeding.
SOURCES: WHO Global Status Report on Road Safety, 2015. IRTAD Road Safety Annual Report, 2015.

By this figure many countries are in crisis of accident ,so we have to keep in mind this figure and make our robust solution . This project may also use in self driving cars in decision making to avoid accident and make the passenger safe . This accident is increase day by day and common people not concern ,so as company to provide any product company have to make product safe ,and safe product will get more cost as well.

Dataset :

This is an standard data set published by the Seattle Police Department, with over 194673 observations with 37 attributes collected over the last 15 years. By this huge data we have to make a effective model to prevent future accident and reduce severity, so it can be use by people for getting security and also use by companies to build a reilable system .

Attributes : SEVERITYCODE , X , Y , OBJECTID , INCKEY , COLDETKEY , REPORTNO , STATUS , ADDRTYPE , INTKEY , LOCATION , EXCEPTRSNCODE , EXCEPTRSNDESC , SEVERITYCODE.1 , SEVERITYDESC , COLLISIONTYPE , PERSONCOUNT , PEDCOUNT , PEDCYLCOUNT , VEHCOUNT , INCDATE , INCDTTM , JUNCTIONTYPE , SDOT_COLCODE , SDOT_COLDESC , INATTENTIONIND , UNDERINFL , WEATHER , ROADCOND , LIGHTCOND , PEDROWNOUTGRNT , SDOTCOLNUM , SPEEDING , ST_COLCODE , ST_COLDESC , SEGLANEKEY , CROSSWALKKEY , HITPARKEDCAR

Methodology :

In this project python is used for easily avialblity of functionality, coding is performed on IBM watson jupyter notebook. In python data analysis is easy to perform and python also contain sufficient libery for data tranformation like Pandas, Numpy, Matplotlib, and Seaborn .The data was mostly categorical so I stuck to graphical representation to see correlation between various variables.

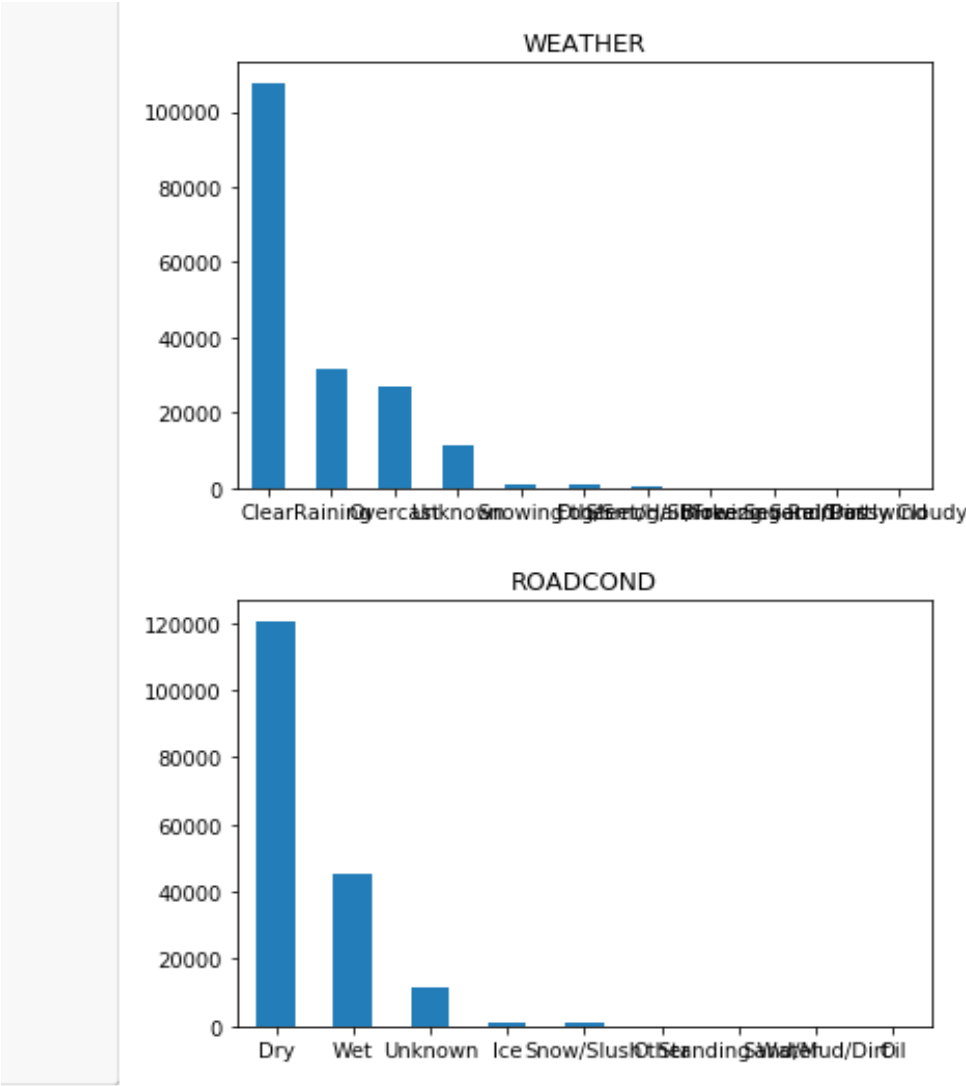
Process have to be followed for proper prediction severity are :

- **Problem Understanding :** First understand the problem in user and business aspect. Check if there is already solution of this problem ,if yes then identify the modifications.Otherwise try to find goal and how to reach that goal at least cost.
- **Data Collection :** Collect the data from standard source or collect data youself with least missing values.
- **Data Visualization :** Visualise data with matplotlib or seaborn liberay . Try to form correlation with scatter plot , pair plot and heat map ,it will give which feature is important and which have to remove .It also help to identify the outliers.
- **Data Transformation :** This process involve make the data to satisfies for the mathematical model which have filling missing values, normalised data , try to reduce or remove outlier , remove the independent features.
- **Data Modeling :** This project target variable is in labeled formed , so supervised learning is used. We have to apply supervised algorithm and get more accuracy.Different algorithm is used in this project random forest , xg boost , svm .
- **Data Evaluation :** After apply above algorithm we have to check accuracy and have to maximise the accuracy by adjusting paramenter and checking for algorithm. this process done by flscore , confusion matrix , by creating distribution plot ,etc

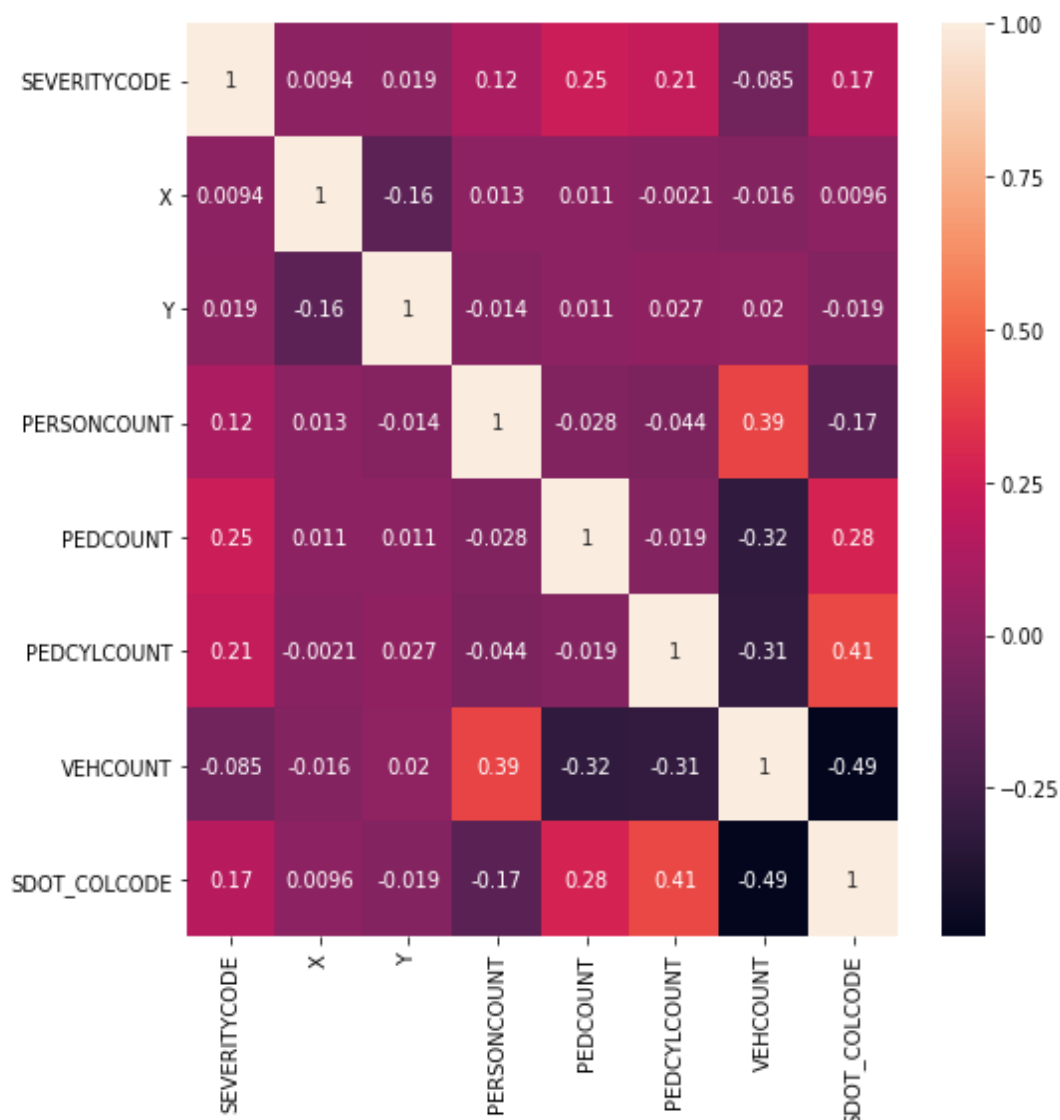
Data Visualization :

Some analysis should be made on data :

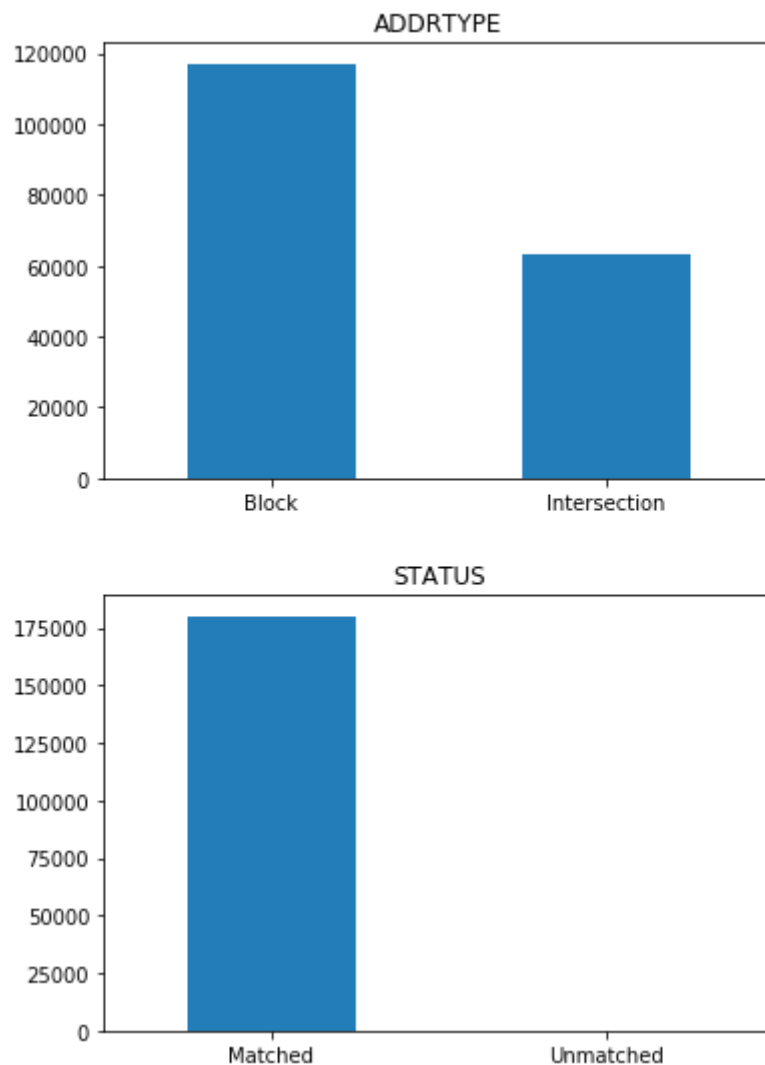
- 1. Weather and Road conditions :



2. Correlations among feature :



3. Status and Address :



Results :

By the graph analysis and after filling missing values we are ready to apply the ML model , These models are use for prediction new events ,while trains with previously happened event that is why ML models are appropriate .I am using random forest because it take several decision and came to right conclusion . Random forest is taken from sklearn library with 100 estimators.

In figure all library are used for random forest are represented . X is independent variable and Y is depended variable. Whole dataset is split into test and train ,so should be used for training .

```
y=pd.DataFrame()
y=x['SEVERITYCODE']
x=x.drop(['SEVERITYCODE'], axis=1)
```

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score,train_test_split
from sklearn.svm import SVC
from sklearn.pipeline import make_pipeline
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score,mean_absolute_error
```

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=0)
```

Implementation of ML algorithms :

```
tree=RandomForestClassifier(n_estimators=100)
tree.fit(x_train, y_train)
y_pred = tree.predict(x_test)
```

Conclusion :

Although this analysis has given us some good insight, there needs to be a closer inspection of certain other variables. It seems like a lot of these accidents are minor and avoidable. Having said that there is still a considerable amount of loss of property and these findings can be helpful to the Seattle PD in enforcing some new measures to prevent future accidents.

```
for i in range(9):
    print(y_pred[i],end=' ')
```

```
2 1 1 1 2 1 1 1 1
```