



# CAR ACCIDENT SEVERITY

Capstone Project

# BACKGROUND

High number of road traffic accidents

**Over 3000 people dying on the world's roads every day**

Can Data Science help to predict the Severity of Car Accidents?

**Development of Machine Learning Algorithms**

Identification of the trend and patterns in the data

Benefiting from an automated mining process

The capacity present for the improvement

## ❑ Data Understanding and Preparation

### ❑ Data Source

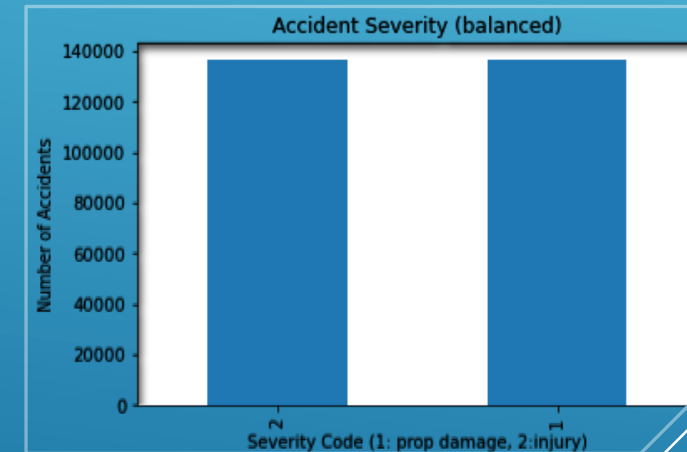
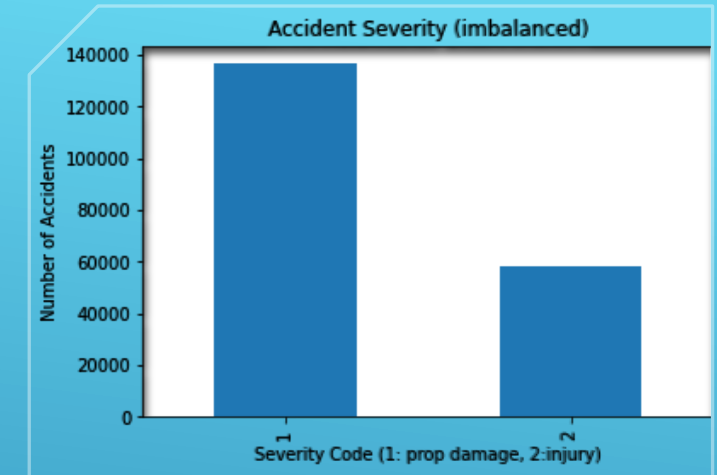
- ❑ Provided by the SDOT Traffic Management Division Traffic Roads Group, Seattle (2004\_Present)

### ❑ Data Balancing: Using resampling technique

### ❑ Data Cleaning

- ❑ Identify and handling of missing values
  - ❑ For numeric values by mean
  - ❑ For categorical values by the frequency
- ❑ Conversion of categorical values to numeric values

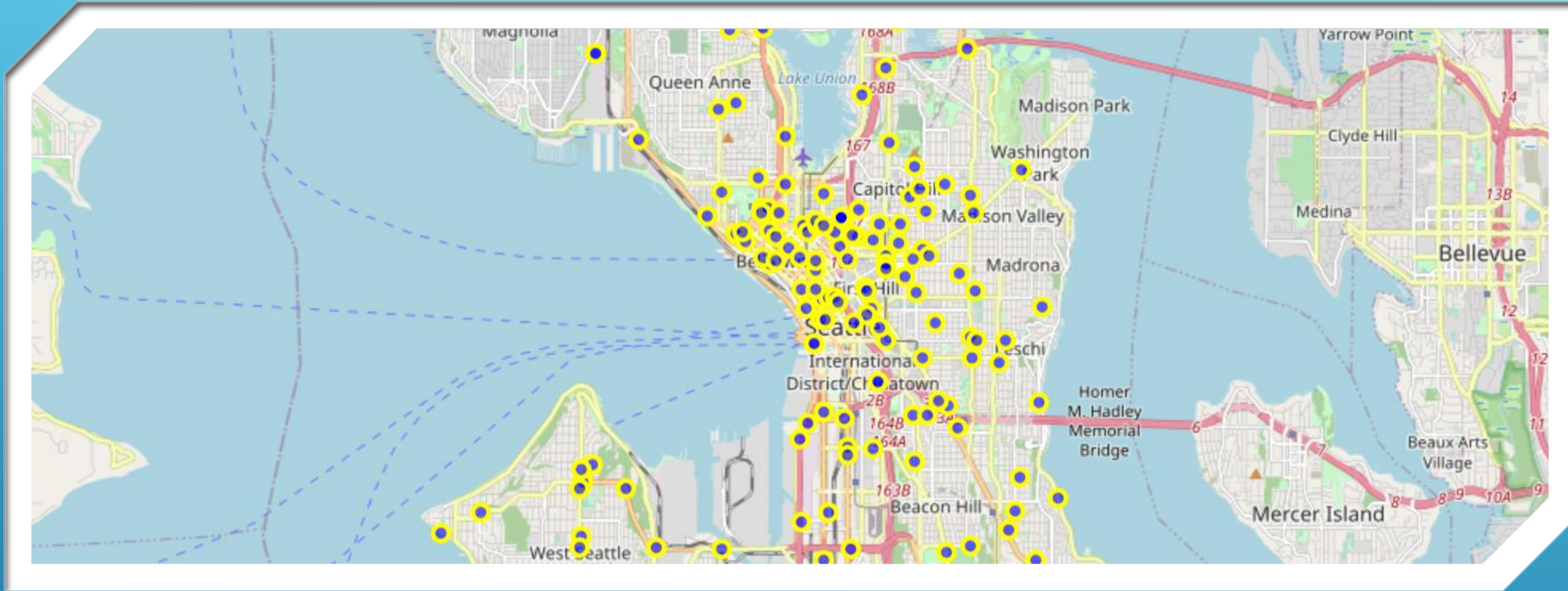
# DATA



	FEATURE	DESCRIPTION
1	LONGITUDE	Longitude
2	LATITUDE	Latitude
3	PERSONCOUNT	Total number of people involved in the collision
4	VEHCOUNT	The number of vehicles involved in the collision
5	JUNCTIONTYPE	Category of junction at which collision took place
6	INATTENTIONIND	Whether or not collision was due to inattention
7	WEATHER	A description of the weather conditions during the time of the collision
8	ROADCOND	The condition of the road during the collision
9	LIGHTCOND	The light conditions during the collision
10	SPEEDING	Whether or not speeding was a factor in the collision

- ❑ Dropping irrelevant information including:
  - ❑ The inter-organizational codes
  - ❑ Redundant descriptive values
  - ❑ Repeating values

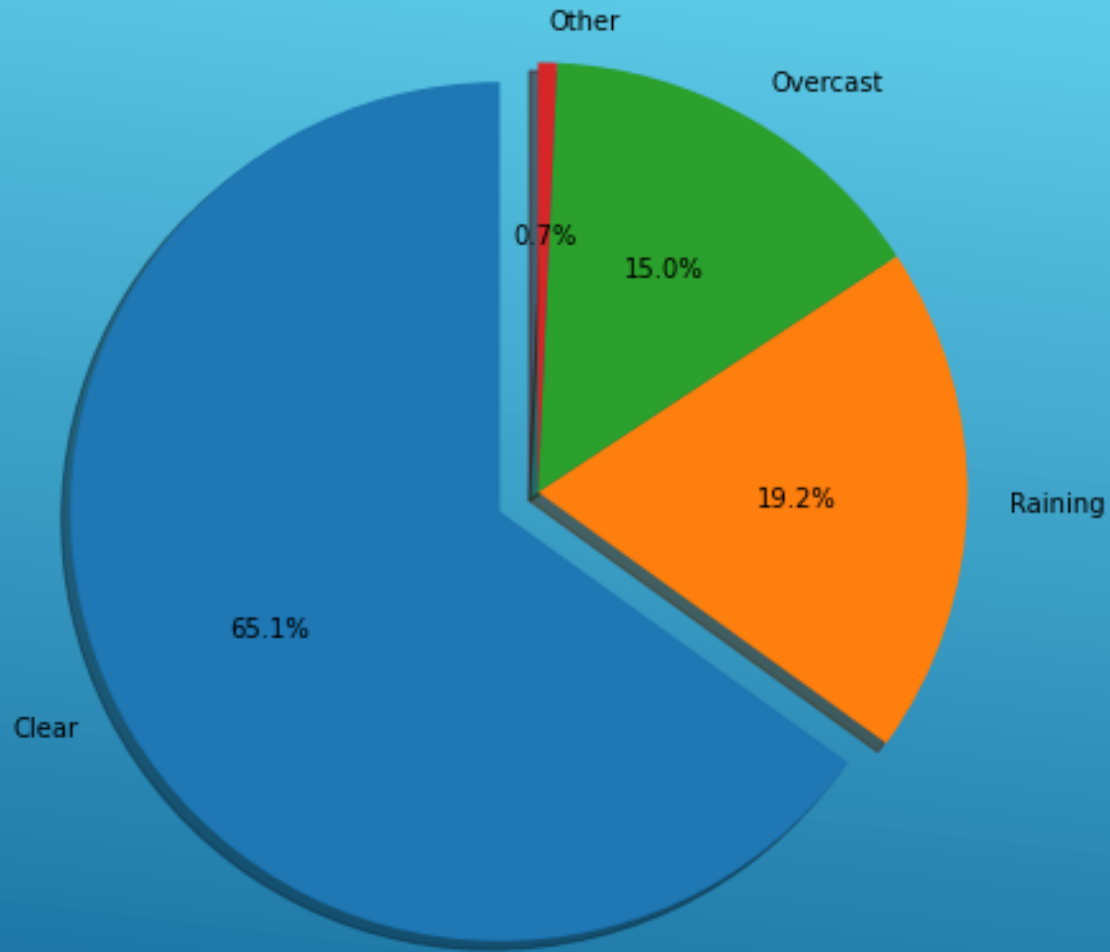
# FEATURE SELECTION



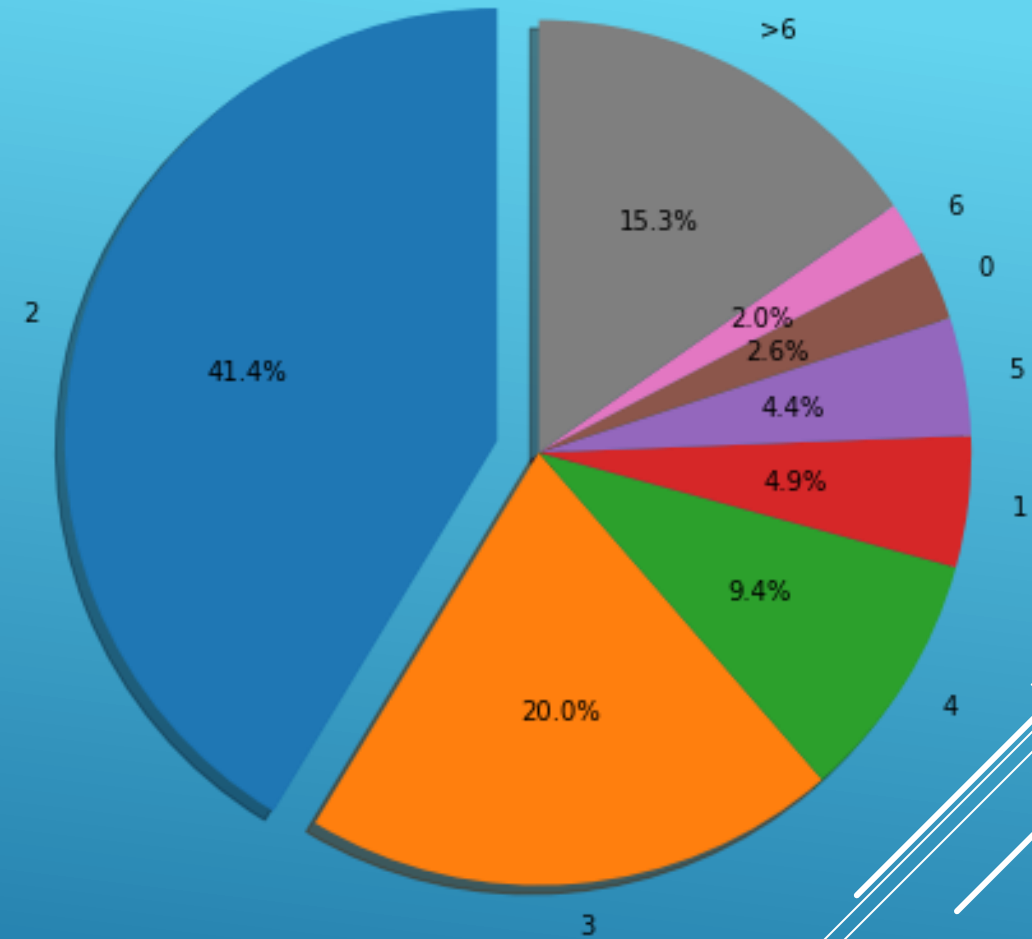
The map with markers of the accident locations in Seattle (300 data points)

# EXPLANATORY DATA ANALYSIS\_1

Effect of Weather Conditions on the Severity with Injury

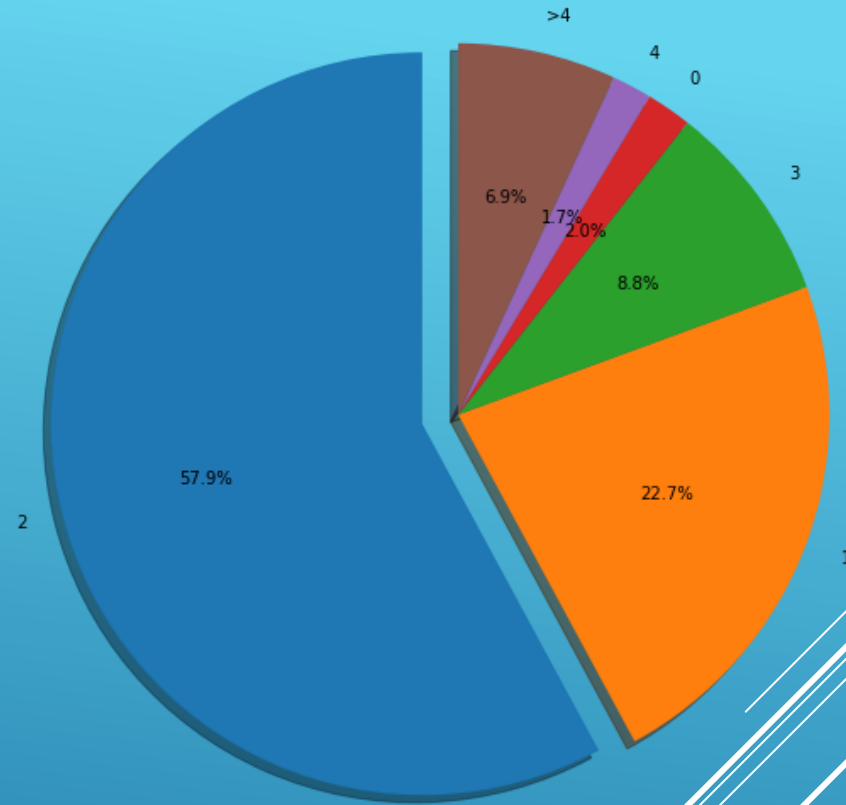


Effect of Person count on the Severity with Injury

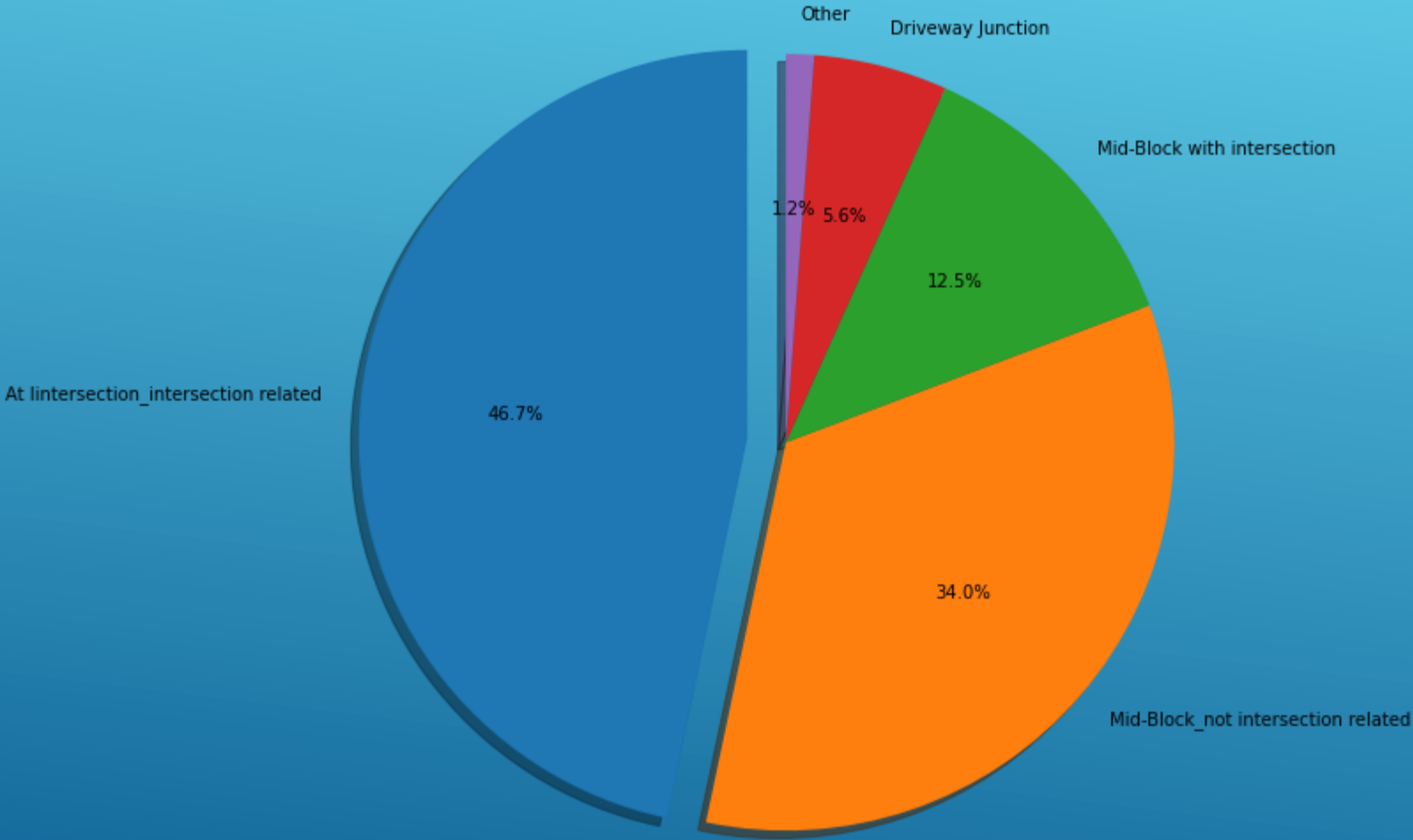


EXPLANATORY DATA ANALYSIS\_2

Effect of Vehicle count on the Severity with Injury



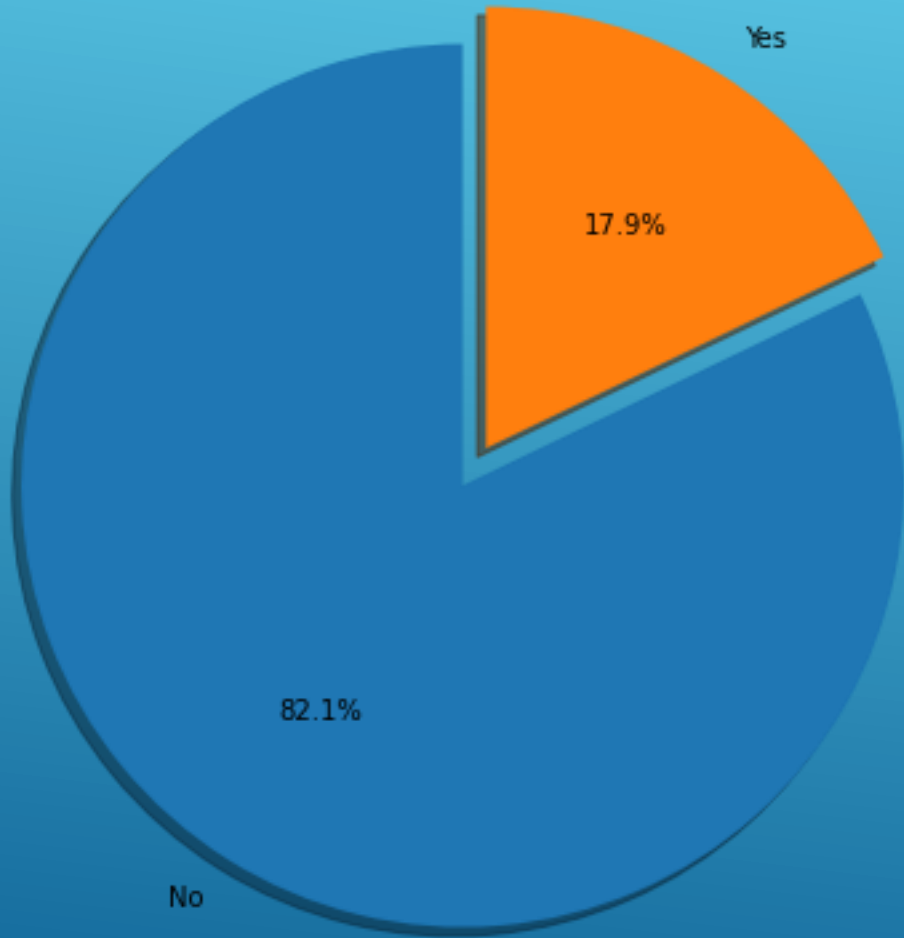
Effect of Junction type on the Severity with Injury



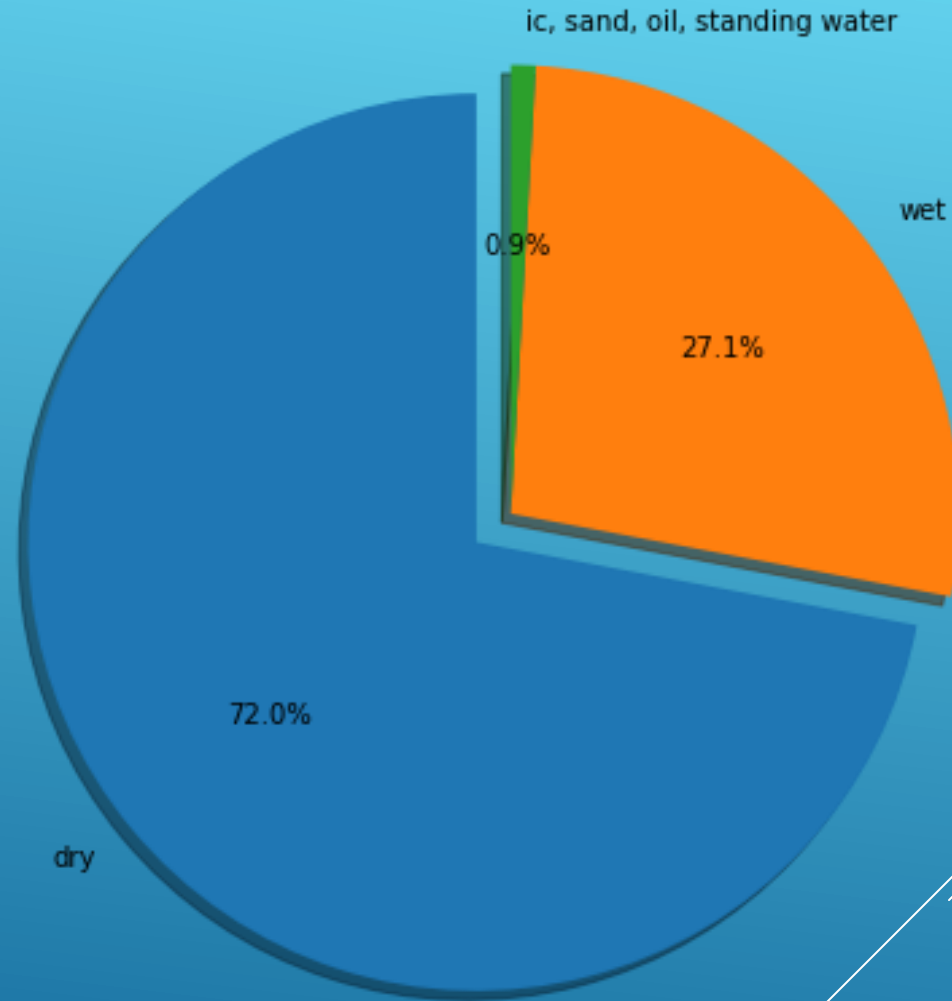
EXPLANATORY DATA ANALYSIS\_3



Effect of Inattention on the Severity with Injury



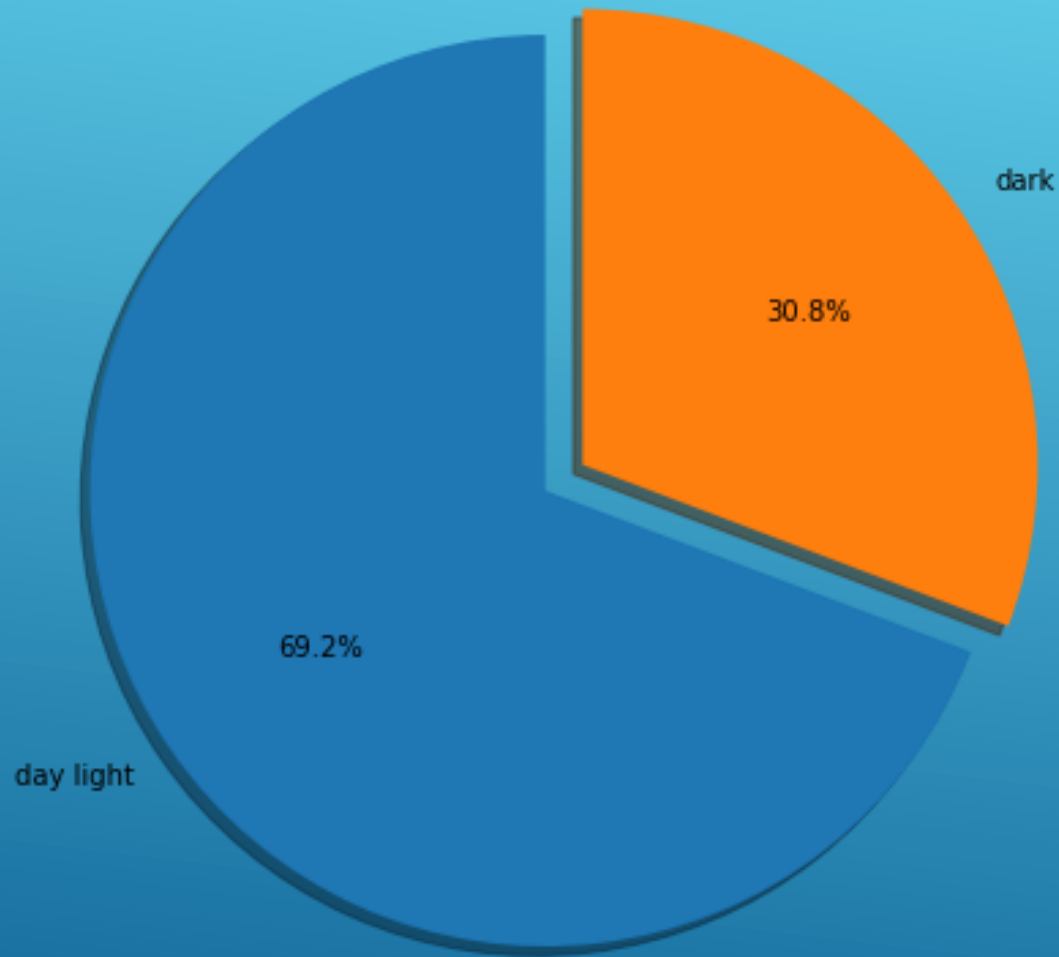
Effect of Road condition on the Severity with Injury



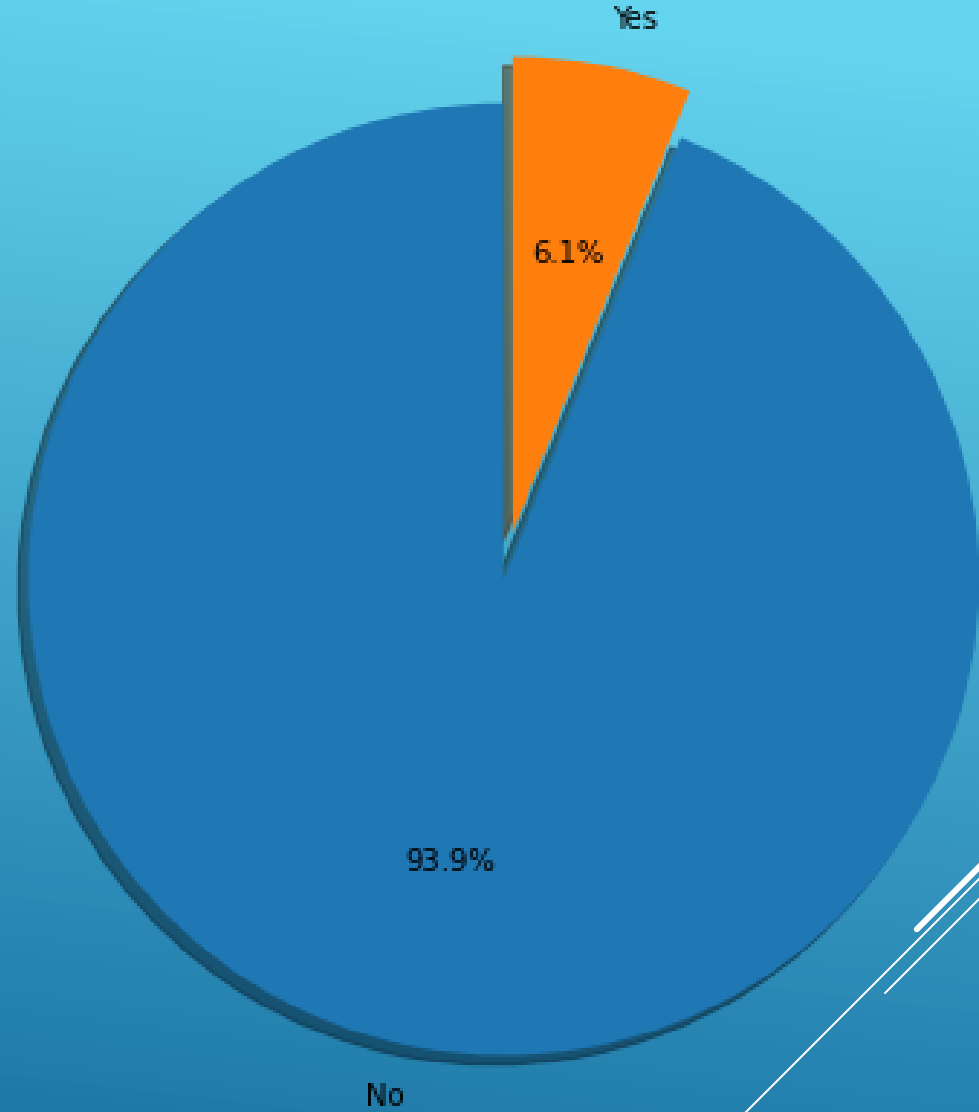
EXPLANATORY DATA ANALYSIS\_4



Effect of Light condition on the Severity with Injury



Effect of Speeding condition on the Severity with Injury



EXPLANATORY DATA ANALYSIS\_5

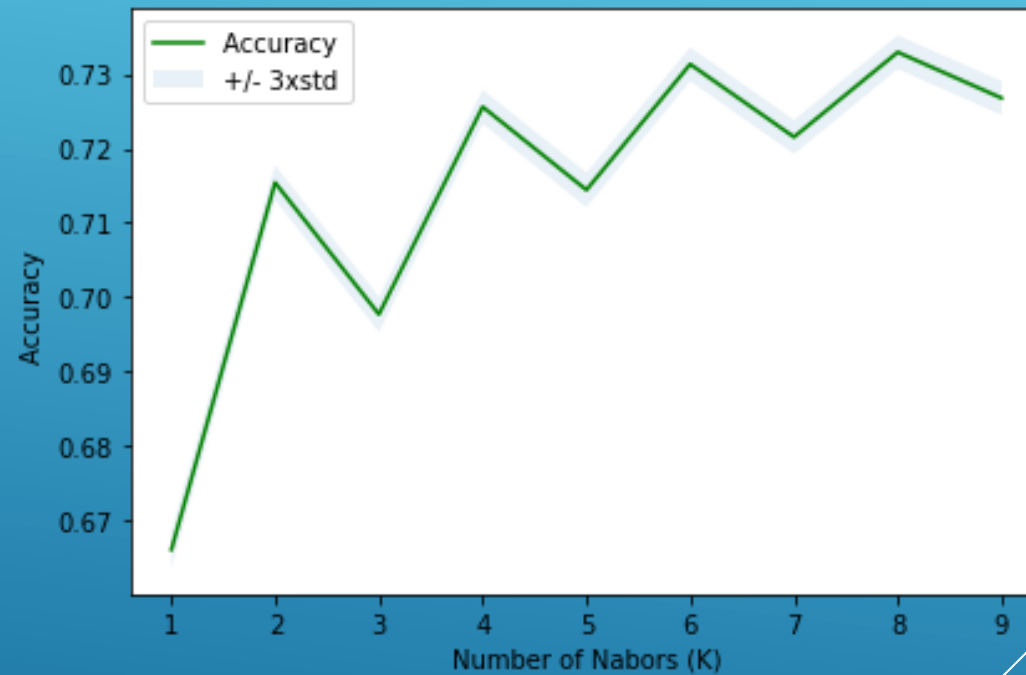
- Normalization of the data
  - Using preprocessing standard scalar
- Splitting into train-test sets
  - 80% training set
  - 20% test set

Best K = 8

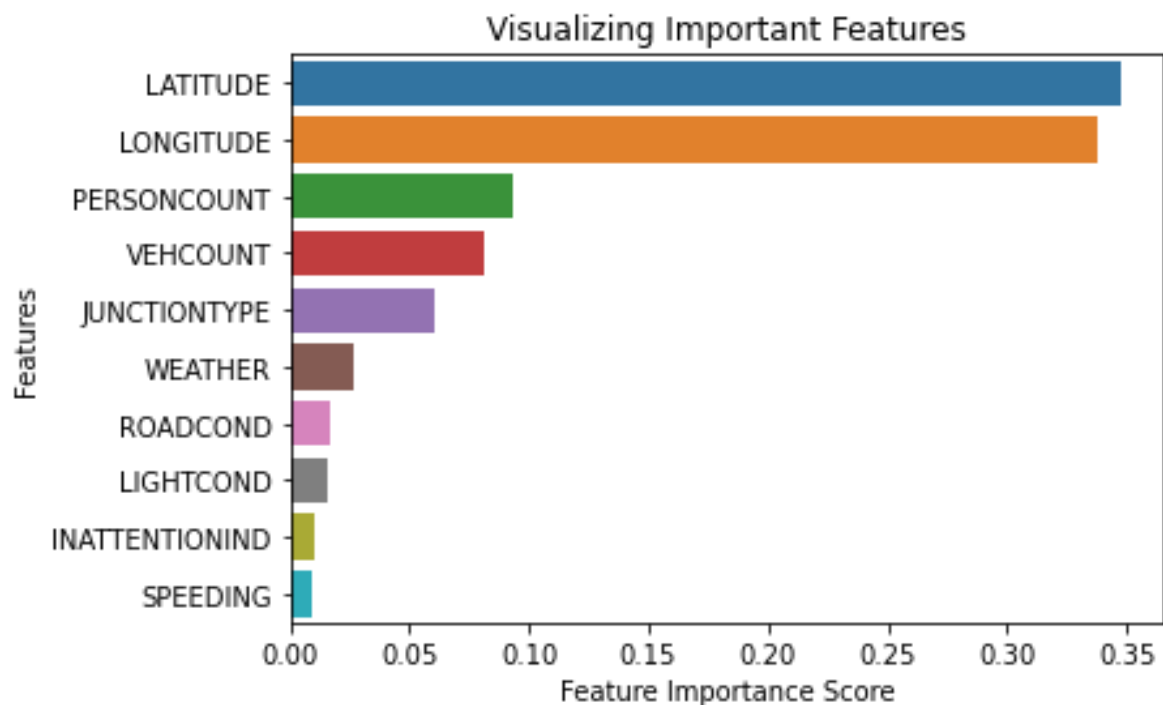
Obtained KNN accuracy = 0.73

Modeling using K-Nearest Neighbors

KNN classifier



# MODELING, EVALUATION AND TESTING\_1



## ❑ Other applied classification techniques

### ❑ Decision Tree

❑ Obtained Accuracy = 0.74

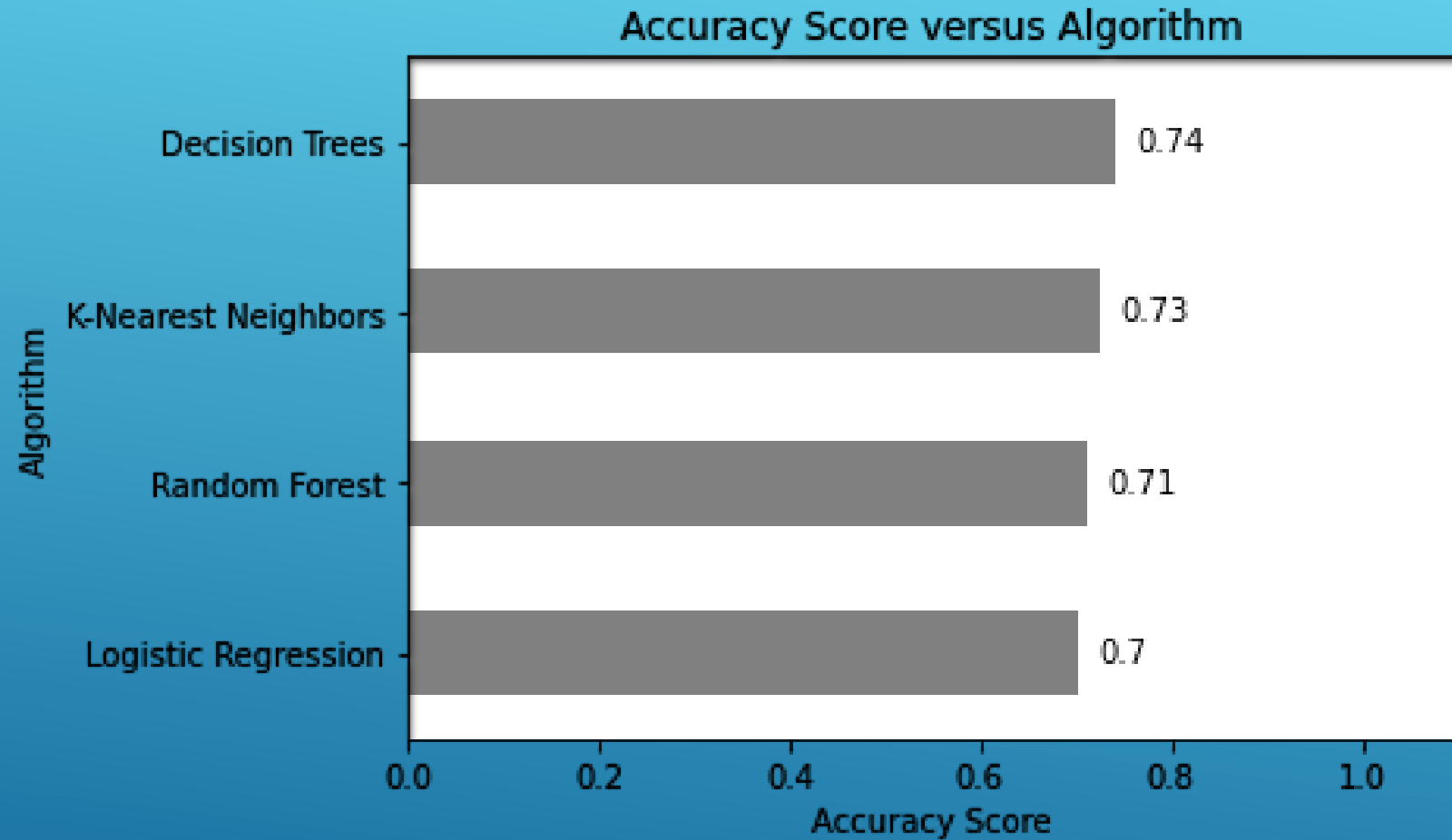
### ❑ Logistic Regression

❑ Obtained Accuracy = 0.7

### ❑ Random Forest

❑ Obtained Accuracy = 0.71

# MODELING, EVALUATION AND TESTING\_2



- Decision Tree
  - Better accuracy score
  - More compatibility with the unscaled data
  - Less sensitivity to the missing values

MODELING, EVALUATION AND TESTING\_3



The dataset has been balanced and further prepared for the predictive modeling analysis.



Four classification algorithms are employed

K-Nearest Neighbors  
Decision Tree  
Logistic Regression  
Random Forest



Evaluation and testing of the created models

Decision Tree has shown better accuracy and more computability for the given dataset.



Idea for future work

Development of the decision Tree to improve its accuracy  
Addition of more data to the dataset to compensate for the missing values  
Taking other factors such as the age of car drivers into account for the prediction of accident severity

# CONCLUSION AND OUTLOOK