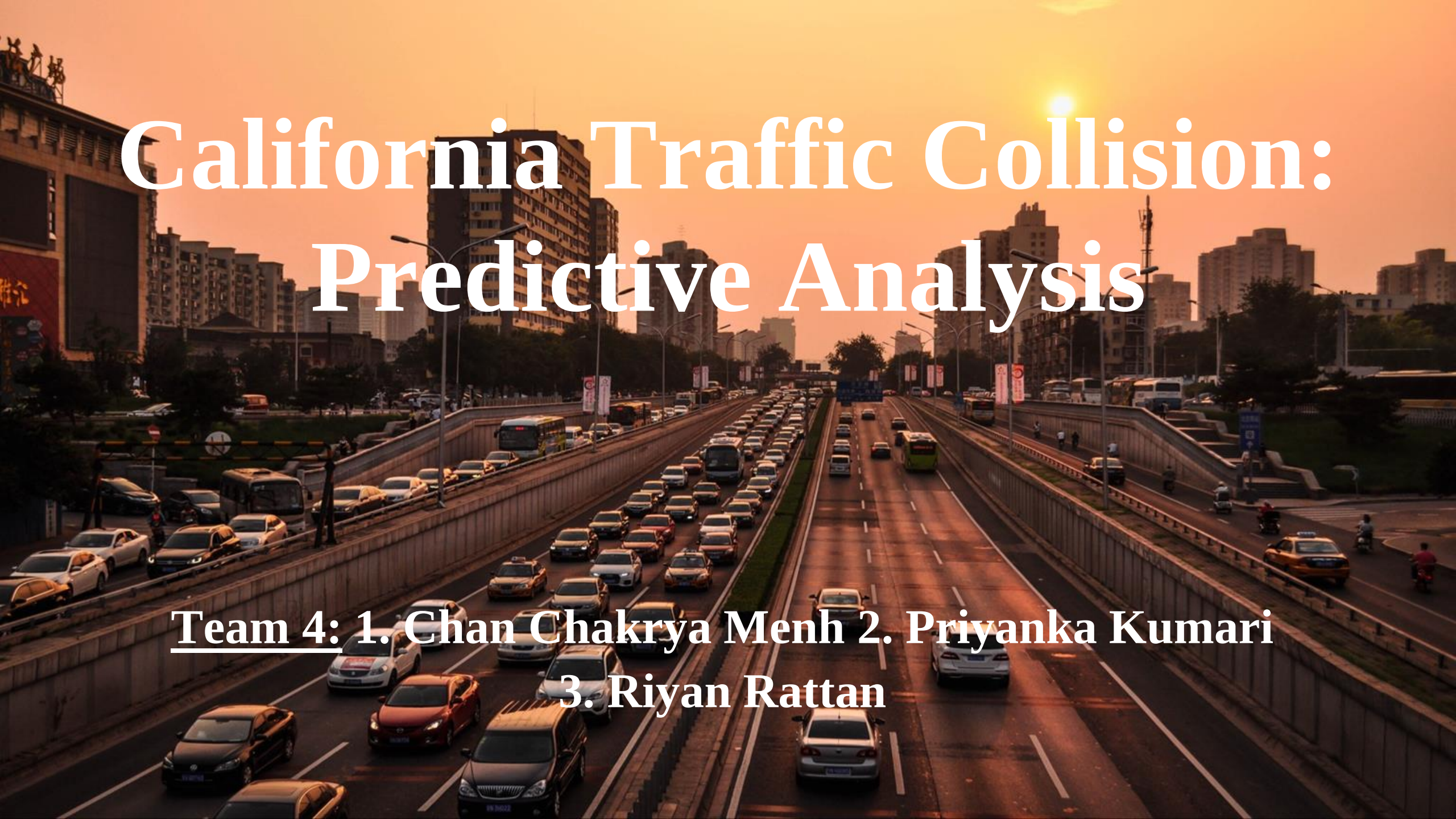


California Traffic Collision: Predictive Analysis

Team 4: 1. Chan Chakrya Menh 2. Priyanka Kumari 3. Riyan Rattan

I. Introduction

Background:

What is Traffic Collision?

A traffic collision or crash occurs when a vehicle collides with another vehicle, pedestrian, road barrier, or a stationary obstacle such as a tree or a utility pole. It may result in injury, death, vehicle damage, possession damage which causes death and disability, and financial burden.

The National Highway Traffic Safety Administration (NHTSA) disclosed its early estimation of traffic fatalities for 2021. NHTSA projects an estimated 42,915 individuals died in motor vehicle traffic crashes last year, a 10.5% expansion from the 38,824 fatalities in 2020. The projection is the highest fatalities since 2005 and the most significant annual percentage increase in the Fatality Analysis Reporting System history.

Motive/Goals:

- This project is to study California traffic collision by using dataset in 2019 to do predictive analysis. This prediction will be achieved by utilizing Orange for machine learning to develop a model with training dataset and eventually with the test dataset.

II. Research Question

- How can we use machine learning to detect the type of collision?
- What type of collision that has more fatalities and injury?
- Which month people was kill and get injure the most from traffic collision?
- What kind of weather has more fatalities and injury?

III. Data Preparation

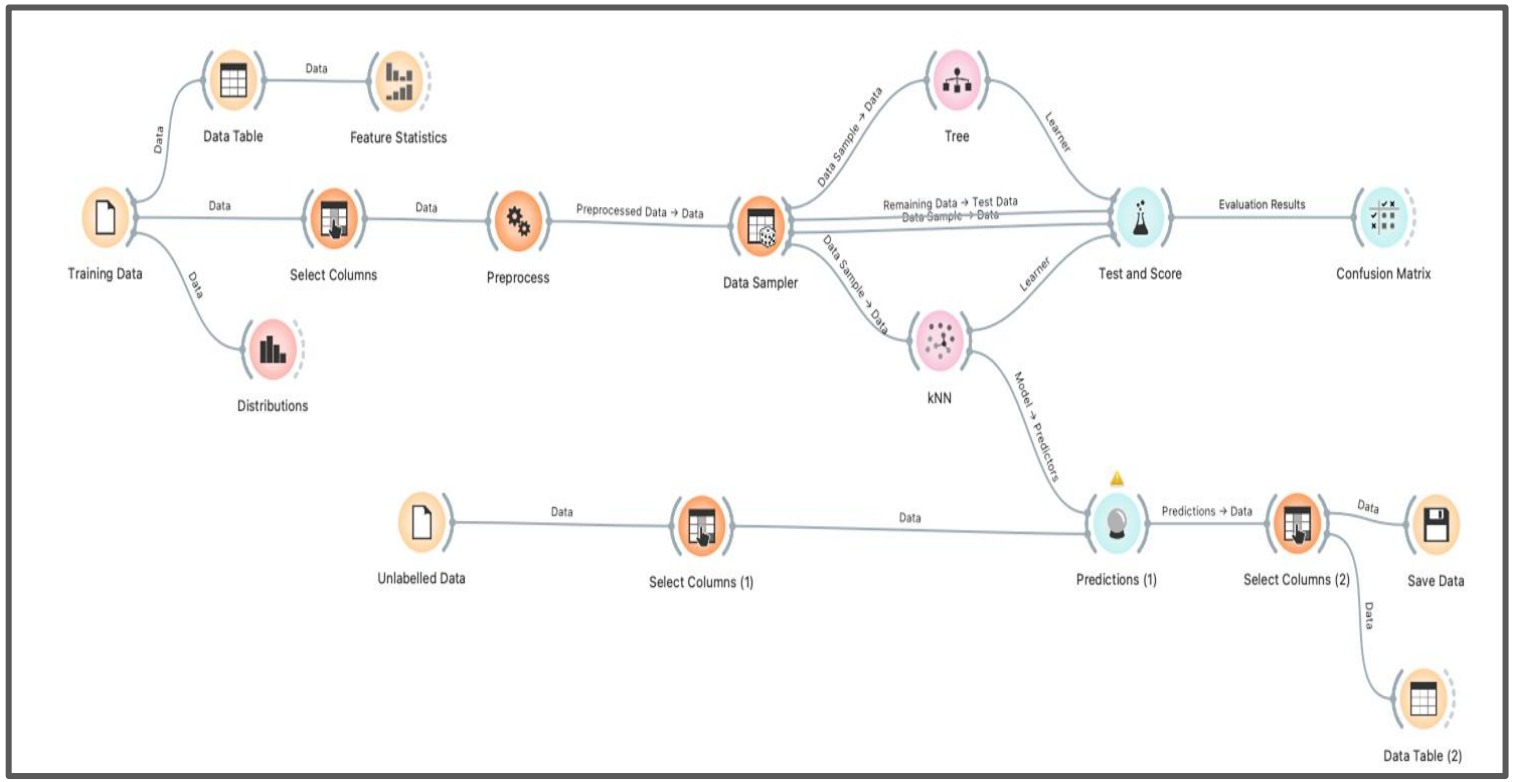
- Cleaning the dataset in csv format
  - Remove irrelevant columns, which eliminated missing values
  - Limit the sample size up to 100K, for faster processing for sake of learning
- Import Data to Orange to check data attributes, statistics, and distribution
- Set Type of Collision as the target in the train data
- Run test score, prediction, and confusion metric
- Use Tableau for data visualization
- Set up second pipeline for test data by removing weather and type of collision to explore possibilities as gender, type of vehicles, etc. in the finalized data model

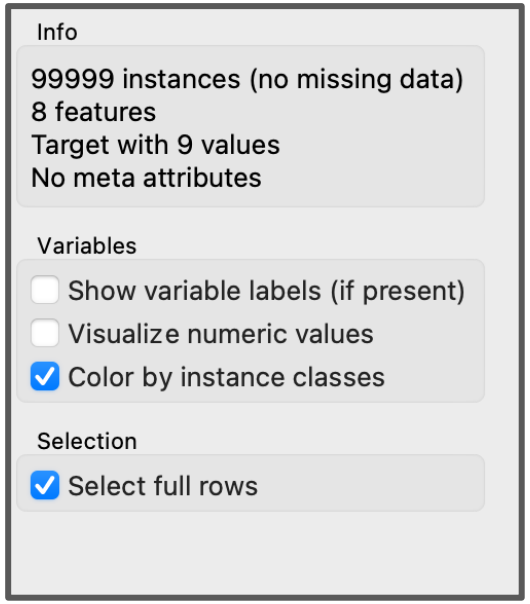
IV. Data Analysis

- Create training and test data set
- Target response to traffic collision is quantitative method
- Better data models for training the data set:
  - kNN
  - Decision Tree

Orange:

Create two pipeline for train and test data.





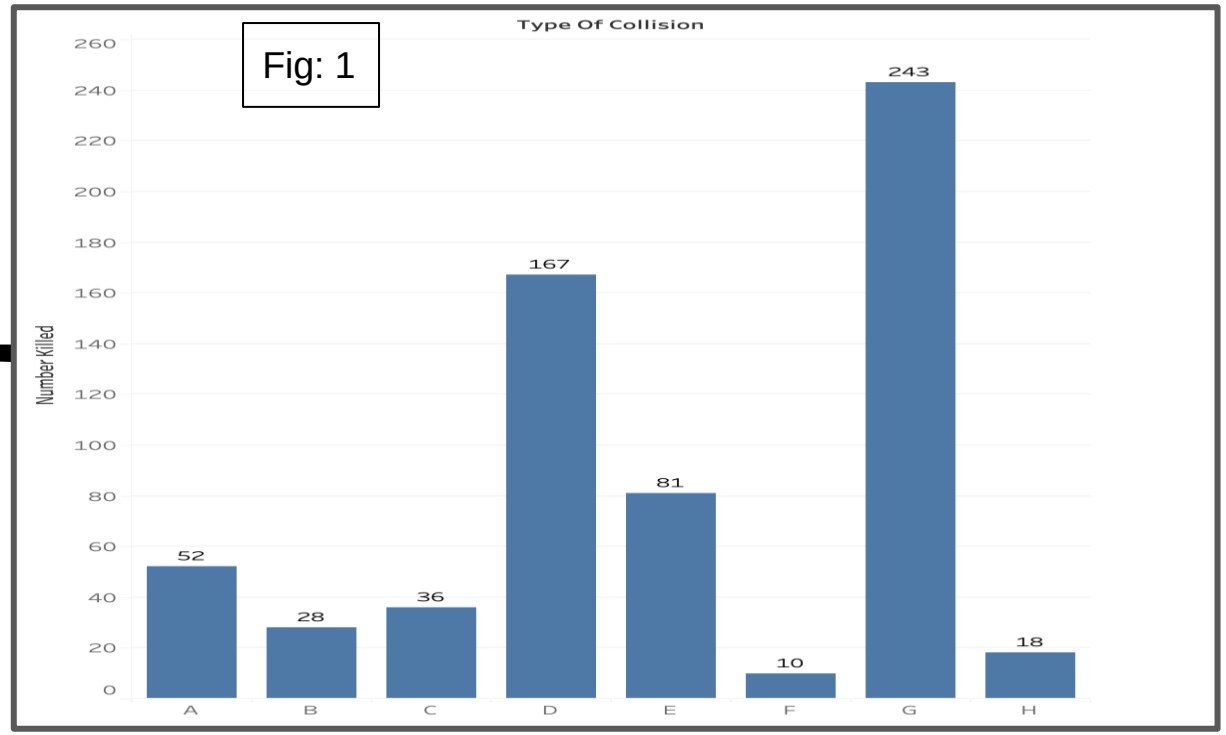
V. Data visualization : Tableau

Fig1: Number Killed Vs Type of Collision.

G type has most fatalities, and second is D type.

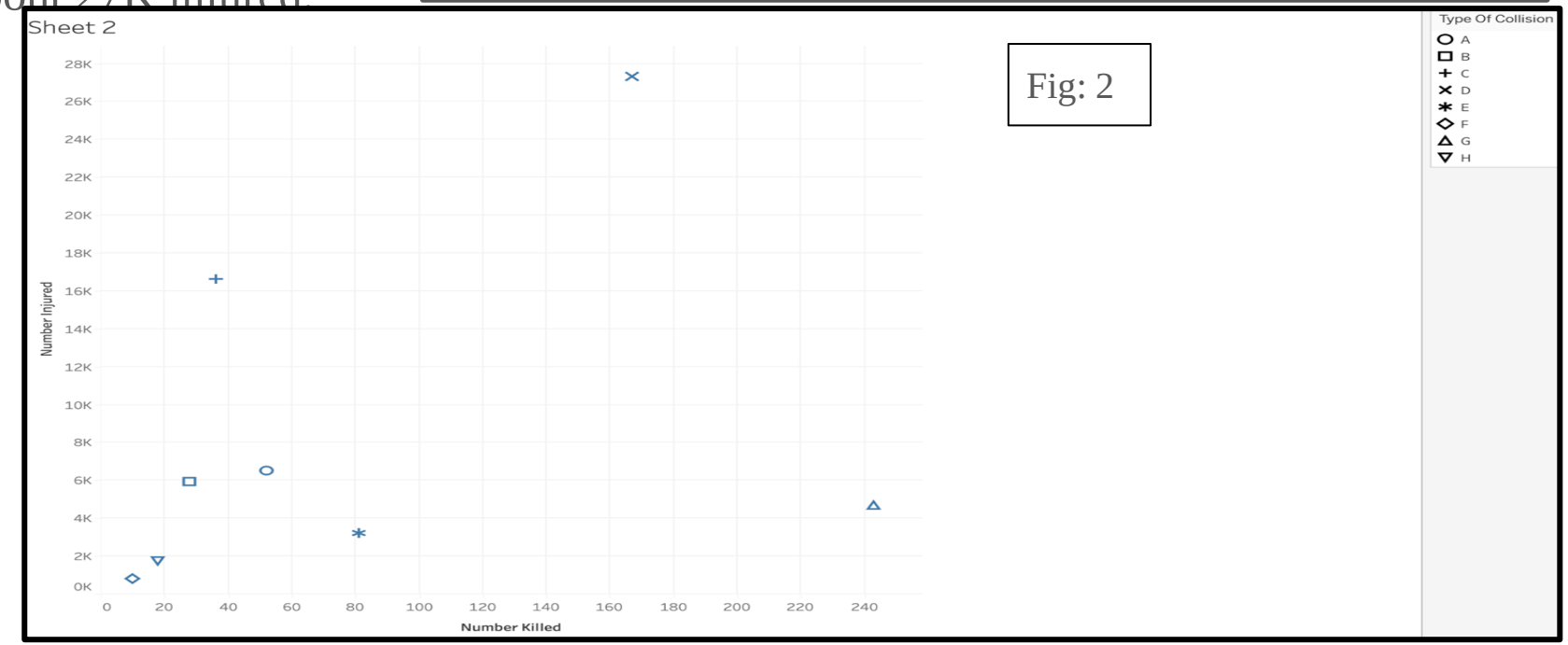
Fig2: Number Killed Vs Number of Injured,

D type has about 167 fatalities and about 27K injured.



Type of Collision

A - Head-On
B - Sideswipe
C - Rear End
D - Broadside
E - Hit Object
F - Overturned
G - Vehicle/Pedestrian
H - Other

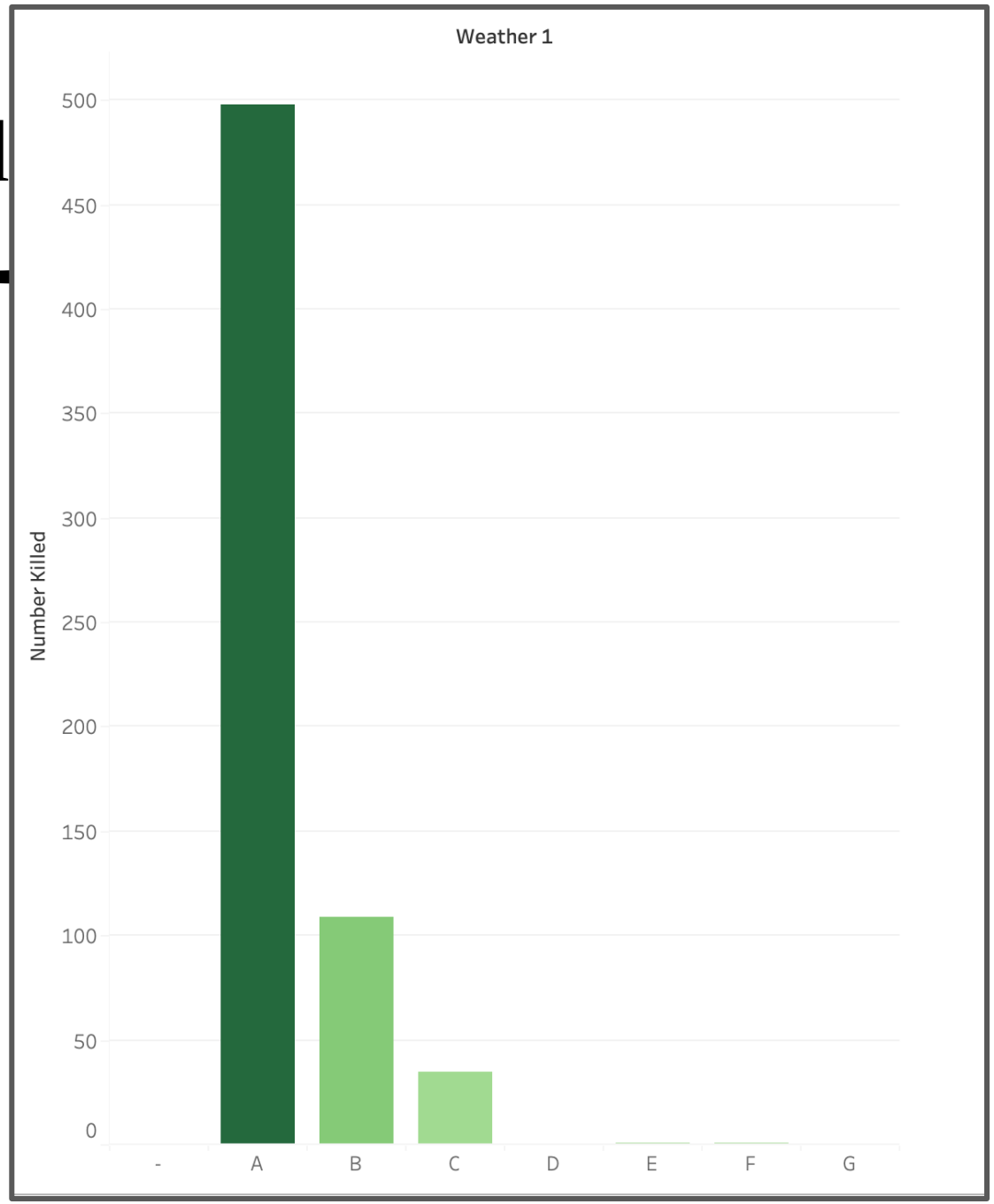


Weather Vs Number Killed

This graph shows that most people died when the Weather is clear.

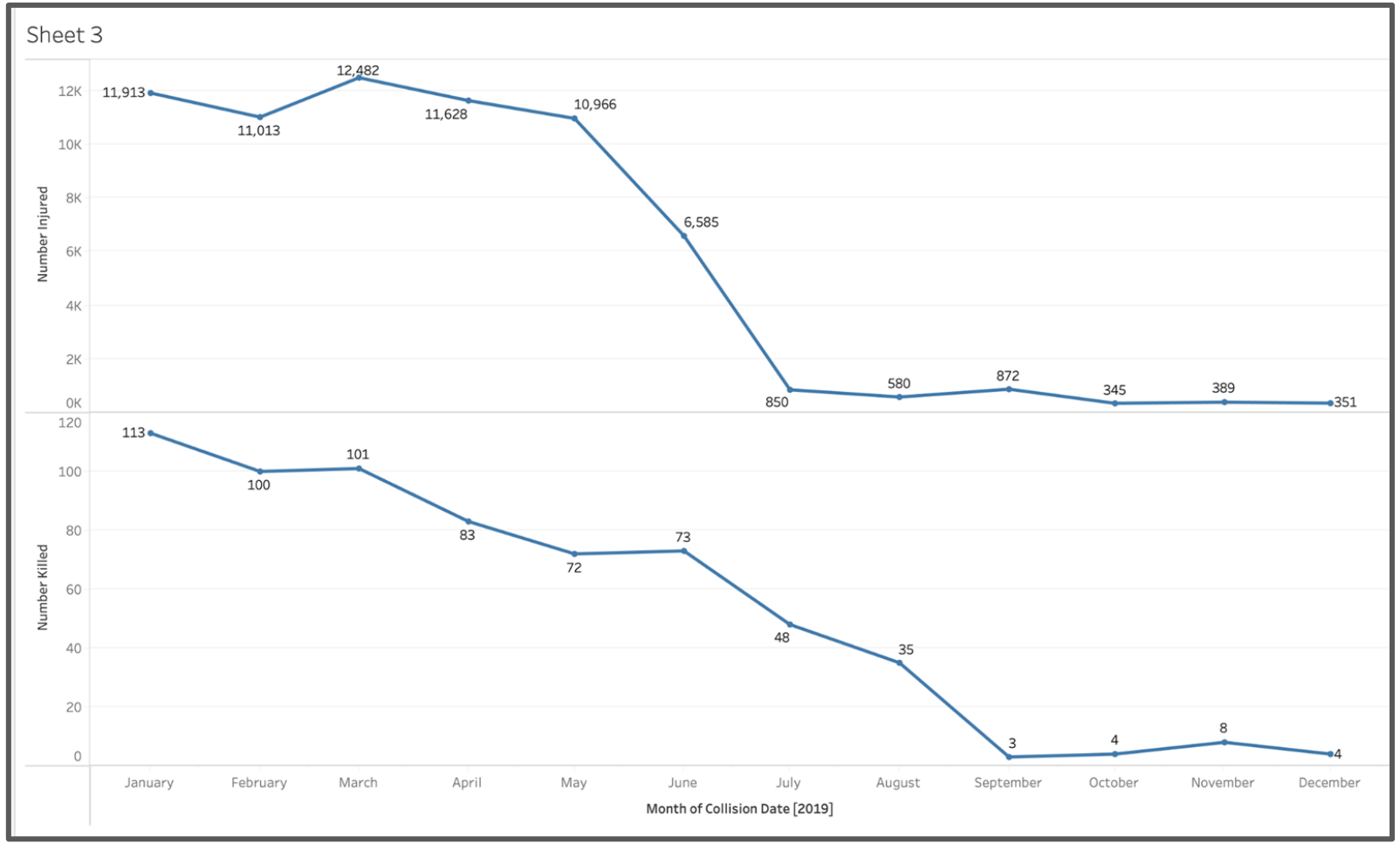
Weather 1:

A - Clear
B - Cloudy
C - Raining
D - Snowing
E - Fog
F - Other
G - Wind

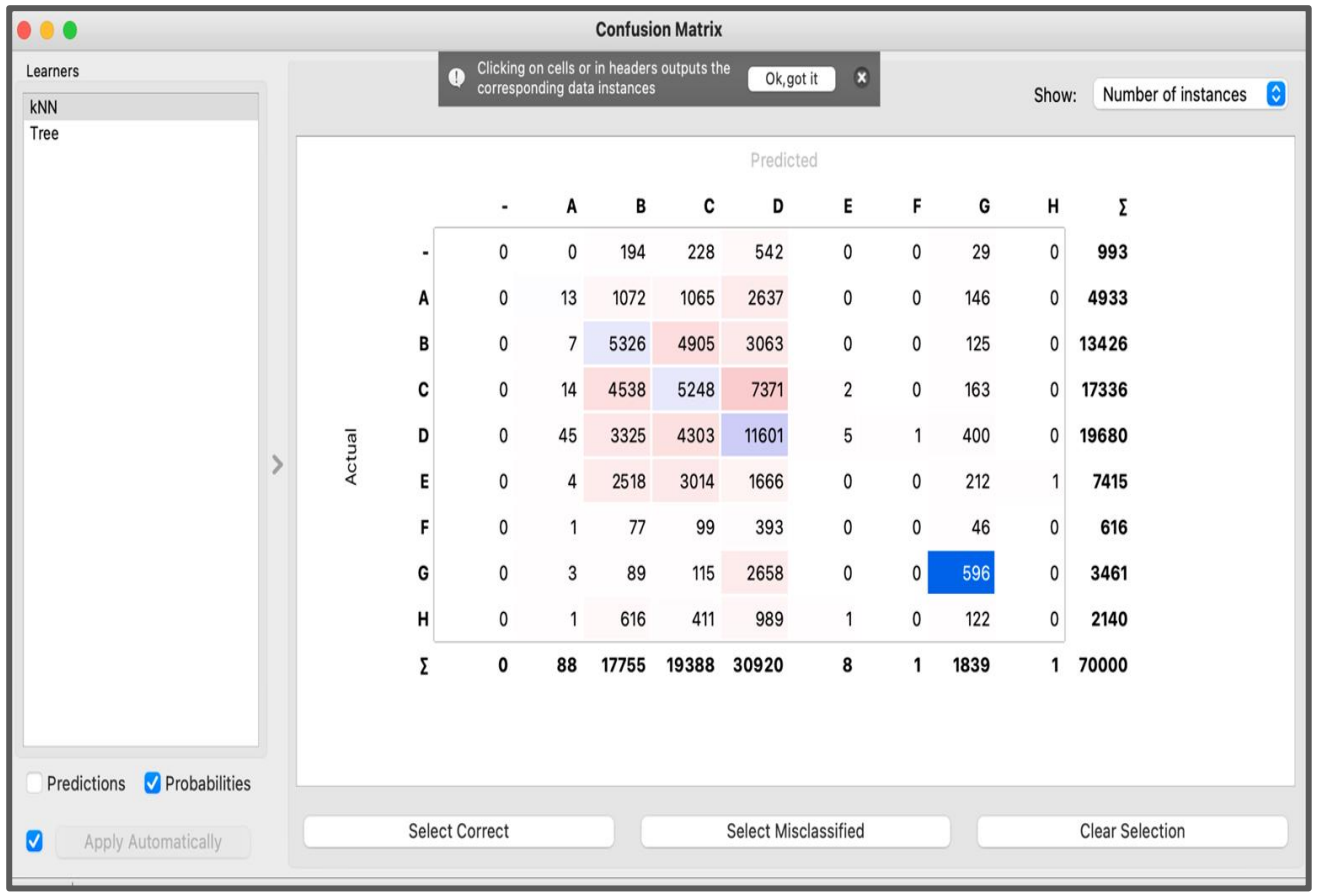


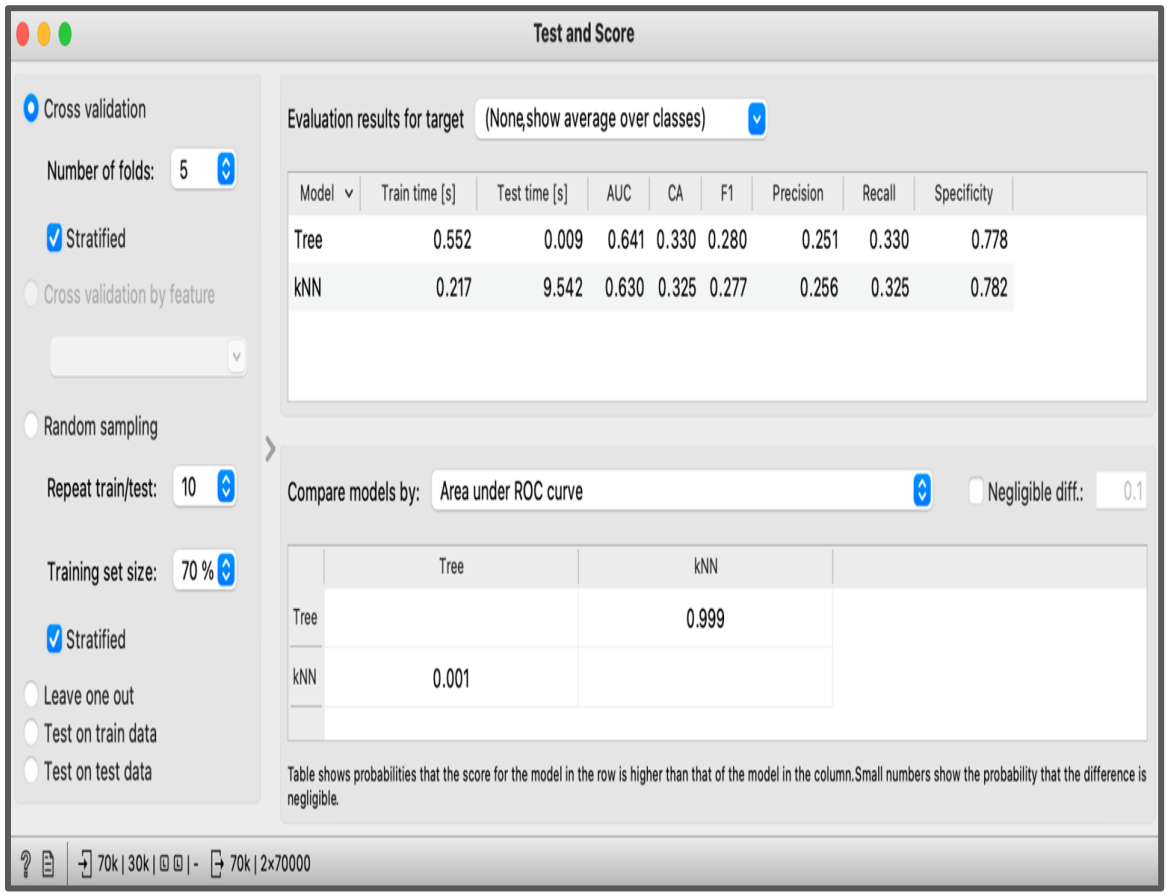
Number Killed Vs Number Injured by Month

In comparison, number of injured on monthly bases is more than number of killed.



VI. Model Evaluation (Train Dataset)





VI. Project Action Plan

No.	Description	Start Date	End Date	Status
1.	Data Cleaning	July 17	July 24	15%
2.	Data Analyzing	July 25	July 31	45%
3.	Data Modeling	Aug 1	Aug 7	65%
4.	Data Visualization	Aug 8	Aug 14	75%
5.	Data Report Submission	Aug 15	Aug 24	100%

VII. Future Plan

- Explore more on data and features from data set, like gender and age to the test data
- Add more visualization
- Working on Modeling and improve accuracy
- Make final slide presentation with all findings