Predicting Injury in a Vehicle - Crash

Table of Contents

01 Problem Statement

02 Target Audience

03 Brief Summary

04 Data Collection

05 Data Cleaning

06 EDA

07 Modeling

08 Conclusions & Recommendations

09 Model Deployment

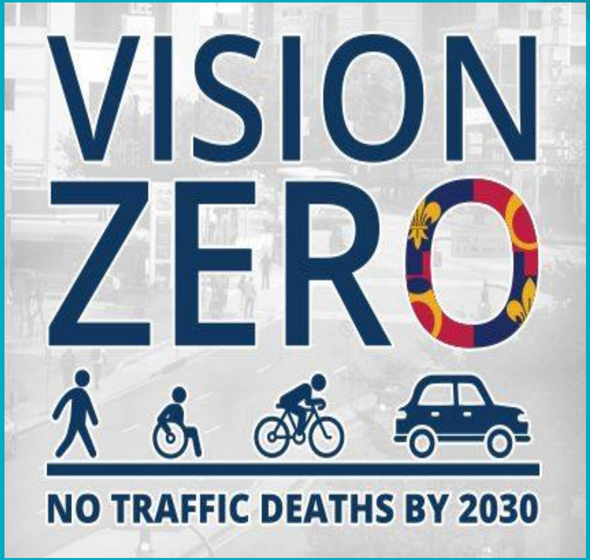
01. Problem Statement:

"Predicting if a crash involves an injury or not, and what factors lead to more crashes."

02. Target Audience

Automobile insurance companies like Geico, Inshur, etc.

Arizona Department of State



Programs like Vision Zero.

03. Brief Summary:

Automobile Insurance Companies:

- According to The Street, liability insurance fraud is staging an accident or injury in order to deliberately file a claim against someone's insurance.
- The FBI estimates that the total cost of insurance fraud (excluding health insurance) is more than $40 billion per year. Insurance fraud costs the average U.S. family between $400 and $700 per year.

ADOT:

- Arizona Department of Transport, responsible for maintaining & constructing Arizona's highway infrastructure.

Vision Zero:

- **Vision Zero** is a multi-national road traffic project that aims to achieve a highway system with no fatalities or serious injuries involving road traffic (Wiki).

04. Data Collection:

- The dataset is collected from DATA.gov .
- It consist of around 40k observations of vehicle-crash in Tempe, Arizona.
- The set contains 35 features like age of drivers, drug/alcohol use, collision manner, weather, surface condition, etc.

0.5 Data Cleaning

1st

- Filled null values

2nd

- Renamed columns

3rd

- Set ‘Datetime’ as the index to organize dataset

4rth

- Dropped useless columns

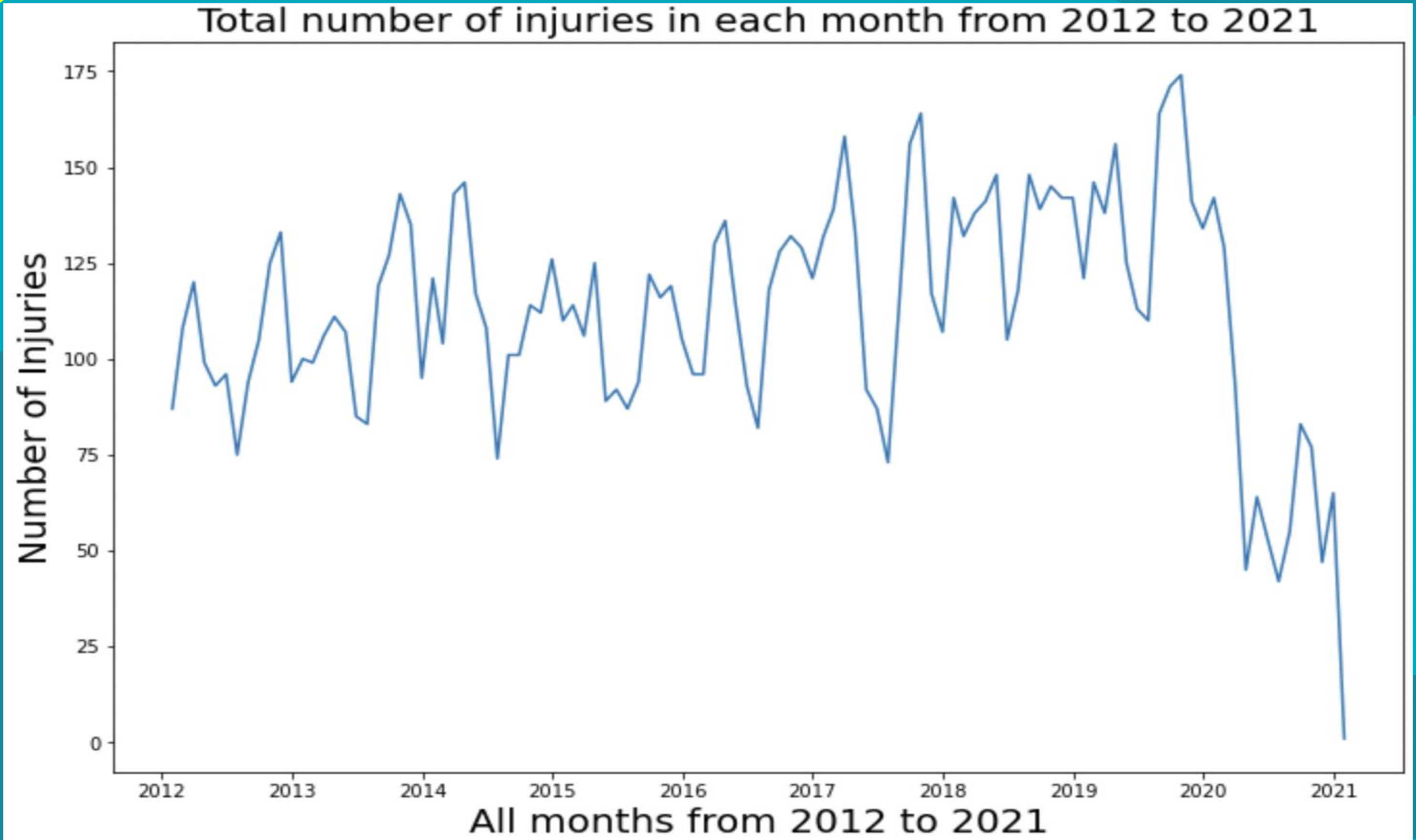
5th

- Used .map on many features to decrease unique values

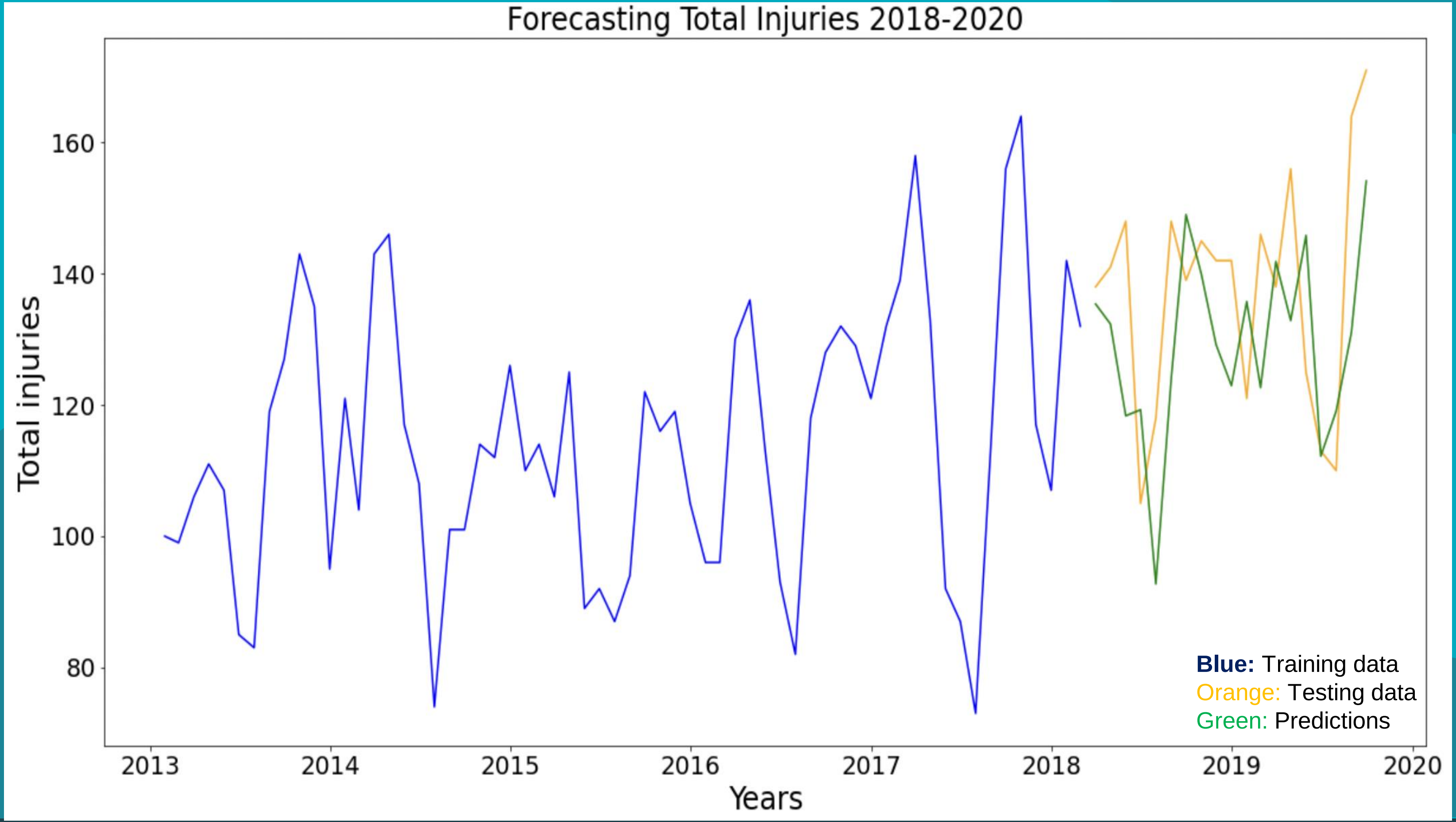
6th

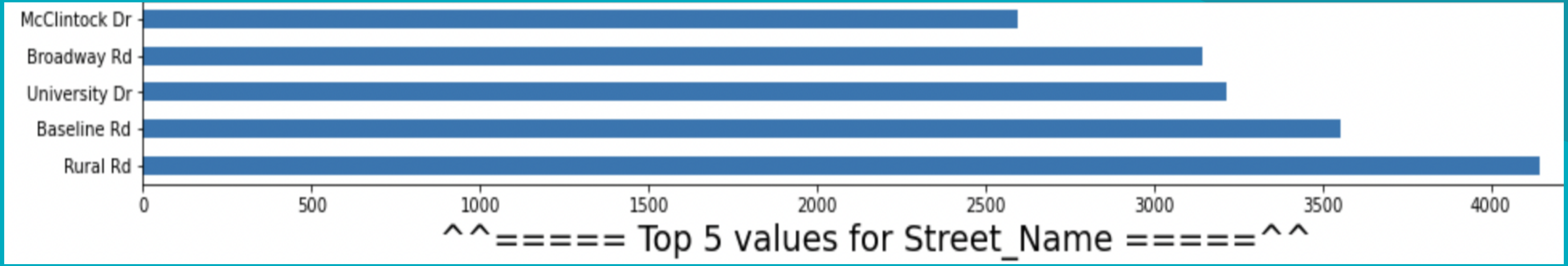
- Used dummy variables

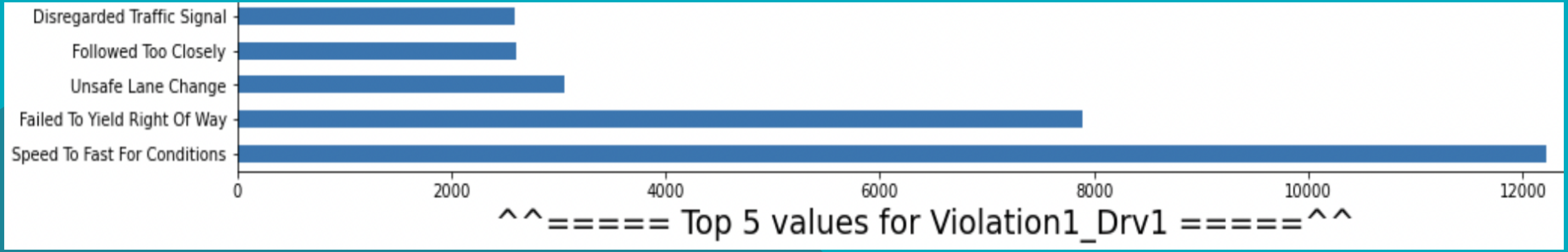
06. EDA:

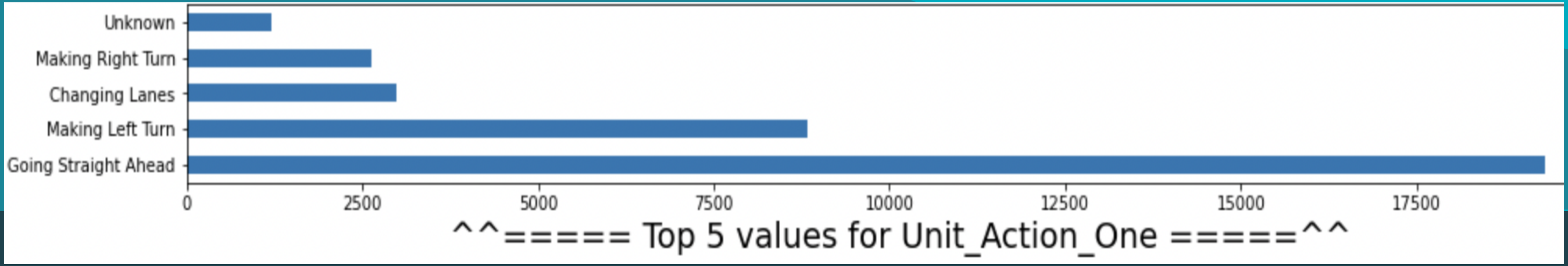


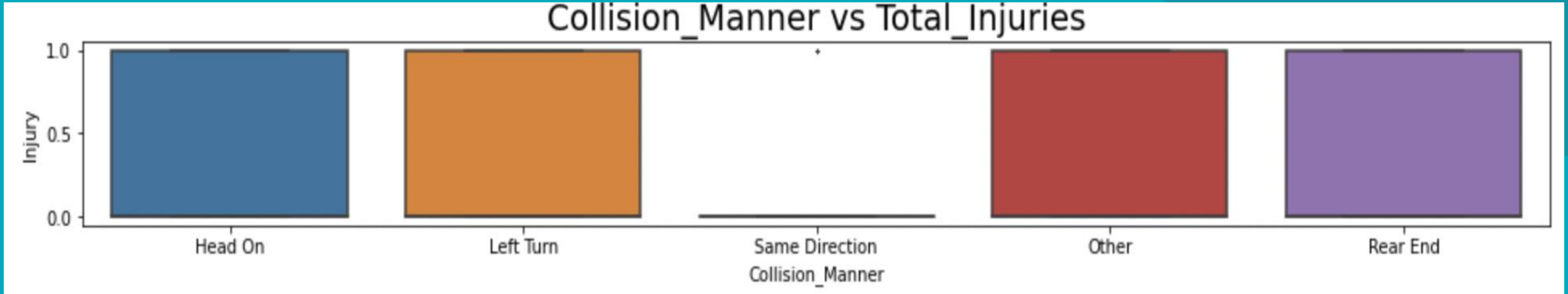
Total number of injuries in each month from 2012 to 2021
Number of Injuries
175
150
125
100
75
50
25
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
All months from 2012 to 2021

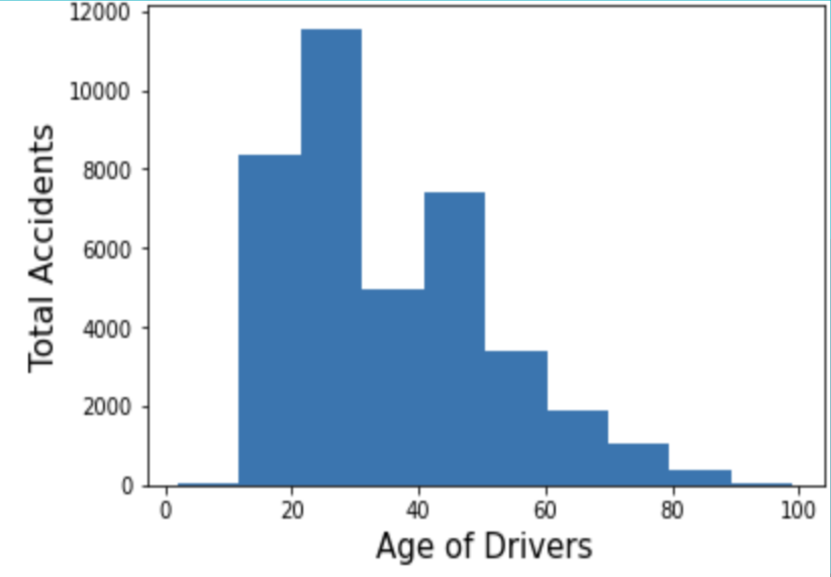

Forecasting Total Injuries 2018-2020
160
140
120
100
80
Total injuries
2013
2014
2015
2016
2017
2018
2019
2020
Years
Blue: Training data
Orange: Testing data
Green: Predictions

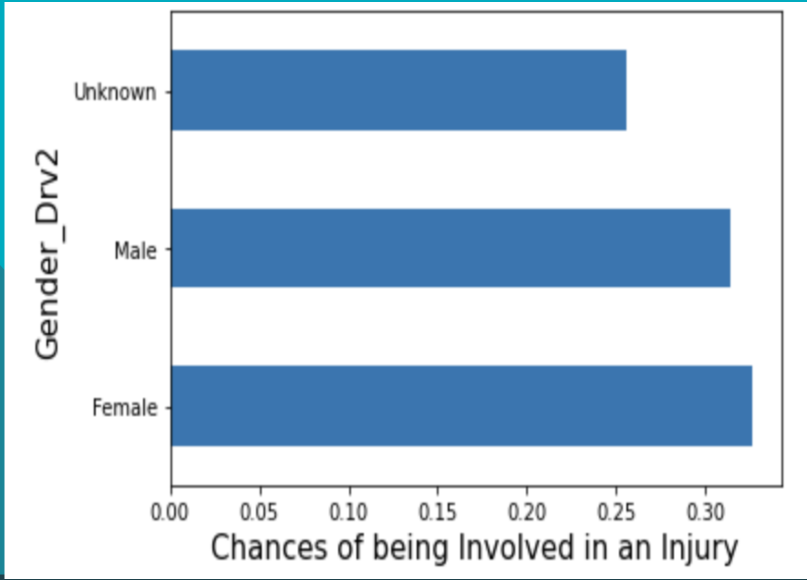

McClintock Dr
Broadway Rd
University Dr
Baseline Rd
Rural Rd
0
500
1000
1500
2000
2500
3000
3500
4000
^^===== Top 5 values for Street_Name =====^^


Disregarded Traffic Signal
Followed Too Closely
Unsafe Lane Change
Failed To Yield Right Of Way
Speed To Fast For Conditions
0
2000
4000
6000
8000
10000
12000
^^===== Top 5 values for Violation1_Drv1 =====^^


Unknown
Making Right Turn
Changing Lanes
Making Left Turn
Going Straight Ahead
0
2500
5000
7500
10000
12500
15000
17500
^^===== Top 5 values for Unit_Action_One =====^^


Collision_Manner vs Total_Injuries
Injury
1.0
0.5
0.0
Head On
Left Turn
Same Direction
Other
Rear End
Collision_Manner


Total Accidents
12000
10000
8000
6000
4000
2000
0
0
20
40
60
80
100
Age of Drivers


Gender_Drv2
Unknown
Male
Female
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Chances of being Involved in an Injury

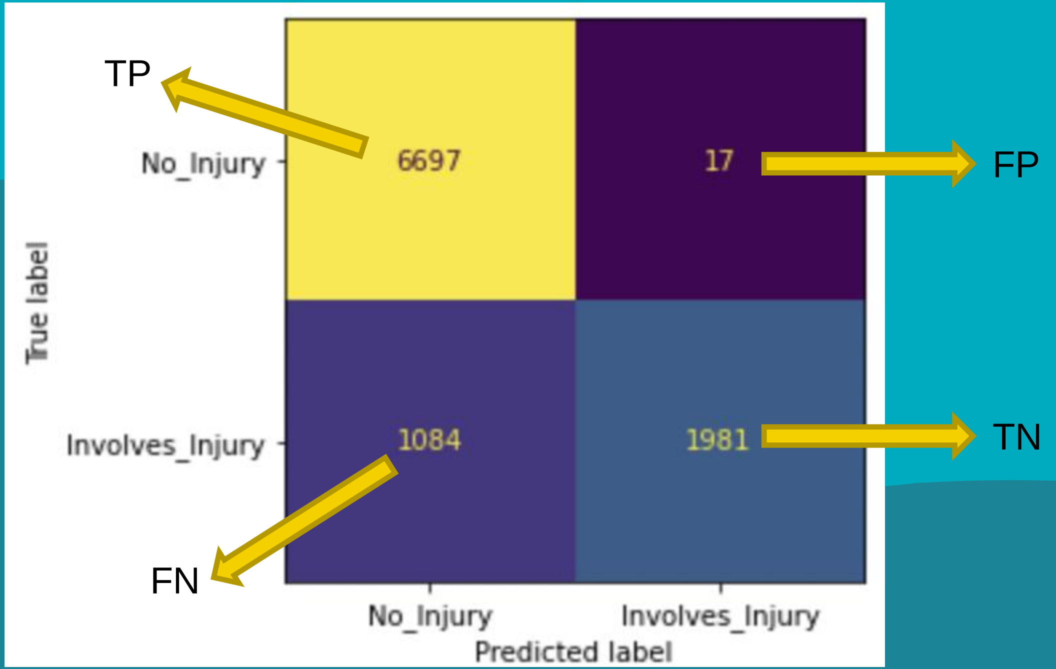
07. Modeling

	Model	Specificity Score
1	Logistic (GridSearch)	99.66%
2	Logistic (PCA)	99.61%
3	KNN (GridSearch)	98.36%
4	SVM	99.69%
5	Decision Tree	99.11%
6	Bagging Classifier	98.88%
7	Random Forest	98.85%
8	Extra Tree	99.05%
9	Naive Bayes (Bernoulli)	99.74%
10	**ADA Boosting**	**99.75%**
11	Neural Netork (NN)	99.67%

Best Score

Baseline Score = 69%

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}}$$



08.a Conclusions:

- EDA graphs.
- Top positive correlated features to Injury: Alcohol_Use, Drug_Use, Injury_Severity.
- Top negative correlated features: Age_Drv, Surface condition, weather.
- Most of the collisions were vehicle to vehicle, very few crash reports on the vehicle to pedestrian/bicycle.
- Best Model: Ada Boosting (highest specificity score 99.75%).

08.b Recommendations:

- Higher fines/points on traffic violations (Alcohol, Drug).
- Installing video cameras at the top 5 streets and intersections where most accidents happened.
- Free driving lessons at any time.
- Requiring people to renew license every year and compulsory taking lessons to renew it.
- Our ADA Boosting can be used by automobile insurance companies to check for false injury claims.
- With all the above steps, we can help the Arizona Department of Transportation (ADOT) minimize crashes and contribute to the Vision Zero initiative.

09. Streamlit App

THANK YOU!

ANY QUESTIONS?