CHICAGO CAR CRASHES PROJECT

AUTHOR: CHRISTINE MUKIRI KIRIMI

CONTENTS

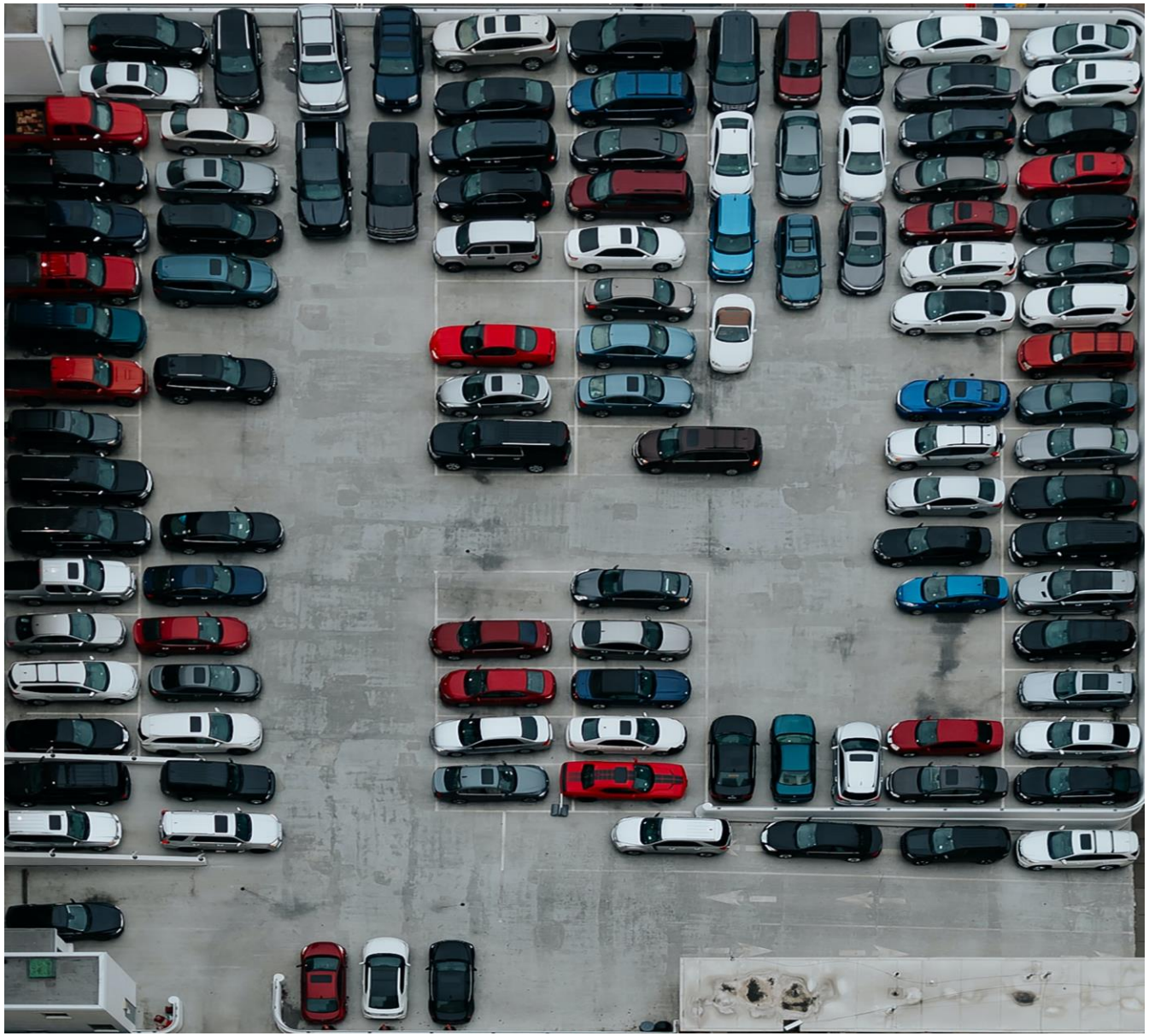
Summary of my contents for the Chicago Car Crashes Dataset.

- OVERVIEW
- DATA UNDERSTANDING
- EDA
- MODELLING
- EVALUATION
- CONCLUSION
- RECOMMENDATIONS

Project Introduction

OVERVIEW

OVERVIEW



My purpose in this project was to design a predictive model that would be able to analyze the primary causes for car crashes. Being able to design a model that can accurately predict the main cause for an incident occurring will allow the city to effectively act against it. If we know the cause for an incident, a city can then plan appropriately as to what measures should be taken to prevent them from happening again. In this project, I will be looking at car crash data from the city of Chicago in order to build a predictive model for their causes as well as identify a number of trends from the incidents.

DATA
UNDERSTANDING

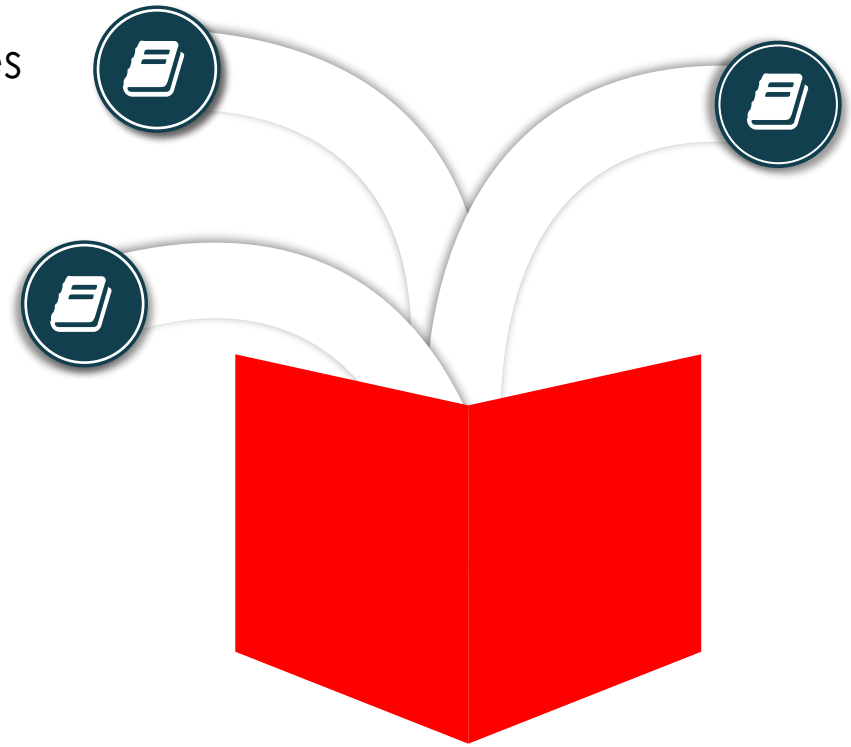
DETAILS

I worked with the following datasets;

Traffic_Crashes_Crashes

Traffic_Crashes_People

Traffic_Crashes_Vehicles



DATA UNDERSTANDING

The Traffic Crashes data comes from the Chicago Data Portal, an open data source maintained by the city of Chicago. The dataset all traffic crashes that were reported by the police within the city limits. Linked to the crash dataset are two datasets corresponding to Vehicles and Persons in the crash. Each crash incident has a unique crash record ID and report number associated with it, which allows for cross-referencing on the dashboards provided for the datasets.

The Crashes dataset contains a number of details related to the incident, such as location/time information, conditions of the road and traffic safety device functionality. The most important detail available is the primary contributory cause for the crash as determined by the office who reported the crash.

DETAILS

I did the analyses guided by the following research questions;

QUESTION 1

What were the primary contributory cause of a car accident, given information about the car, the people in the car, the road conditions.

QUESTION 3

want to answer the question in which year did the most accidents happen?

QUESTION 2

What is the distribution of accidents per age?

MODELLING

MODELLING

In my prediction I used the following models

- LOGISTIC REGRESSION
- DECISION TREES
- K NEAREST NEIGHBORS

MODELLING

TARGET VARIABLE : PRIMARY CAUSES OF ACCIDENTS

We are building a model to predict primary cause of accidents:

IMPROPER/AGGRESSIVE DRIVING	95058
UNABLE TO DETERMINE	79529
RECKLESS BEHAVIOR	55134
EXTERNAL FACTORS/ OTHER	21926

The above is a summary of the main primary cause of the accidents, it is showing the number of data points it represents

MODELLING

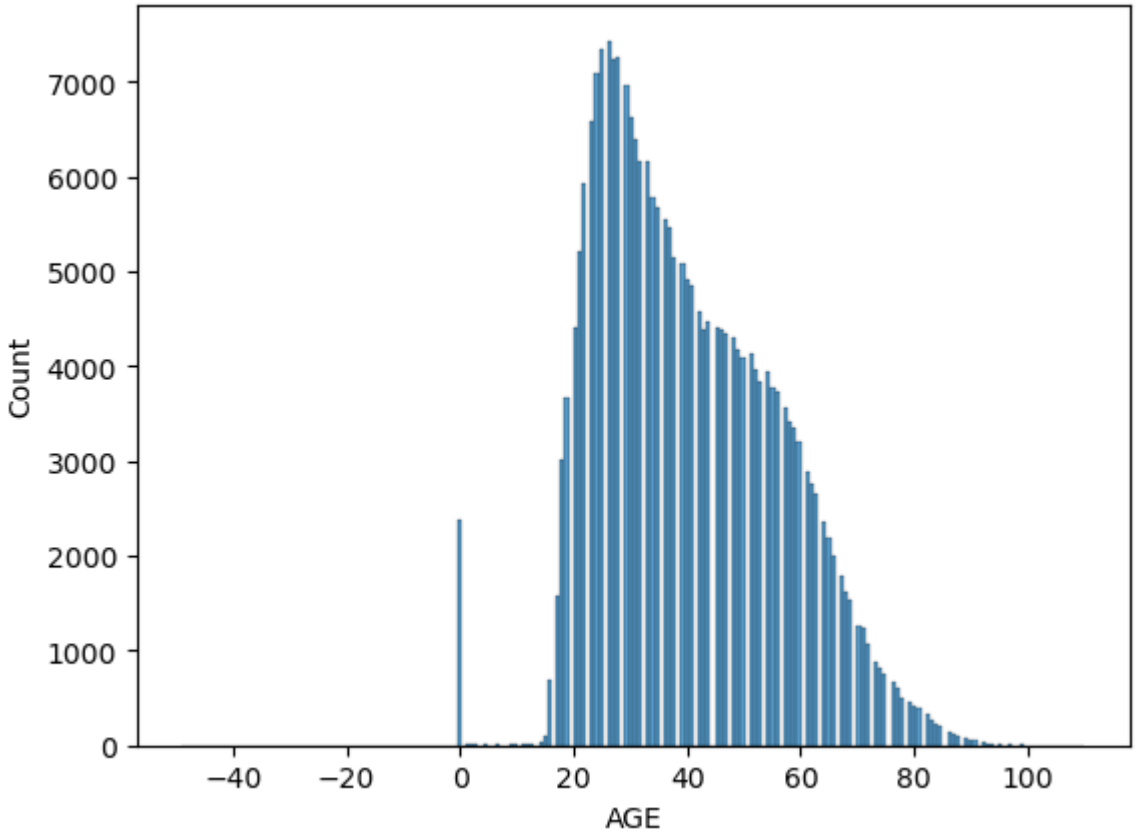
PREDICTOR VARIABLES:

The variables that were used to predict the outcome are the following:

- Sex
- Airbag deployed
- Driver Action
- Physical Condition
- Device Condition
- Weather Condition
- Lighting Condition

EDA FINDINGS

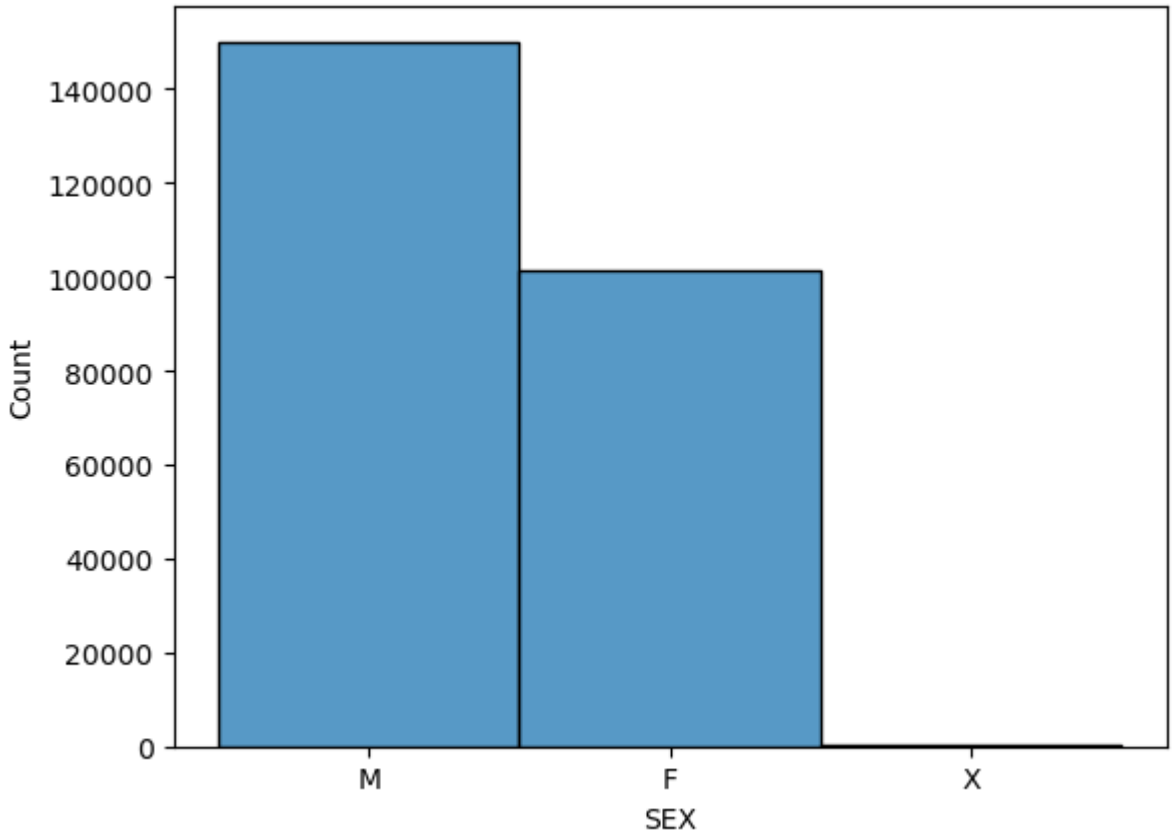
What is the distribution of accidents per age?



The people between the ages of 25 and 35 caused the most accidents. After the age of 40 the number of accidents reduced.

EDA FINDINGS

What is the distribution of Accidents per sex?

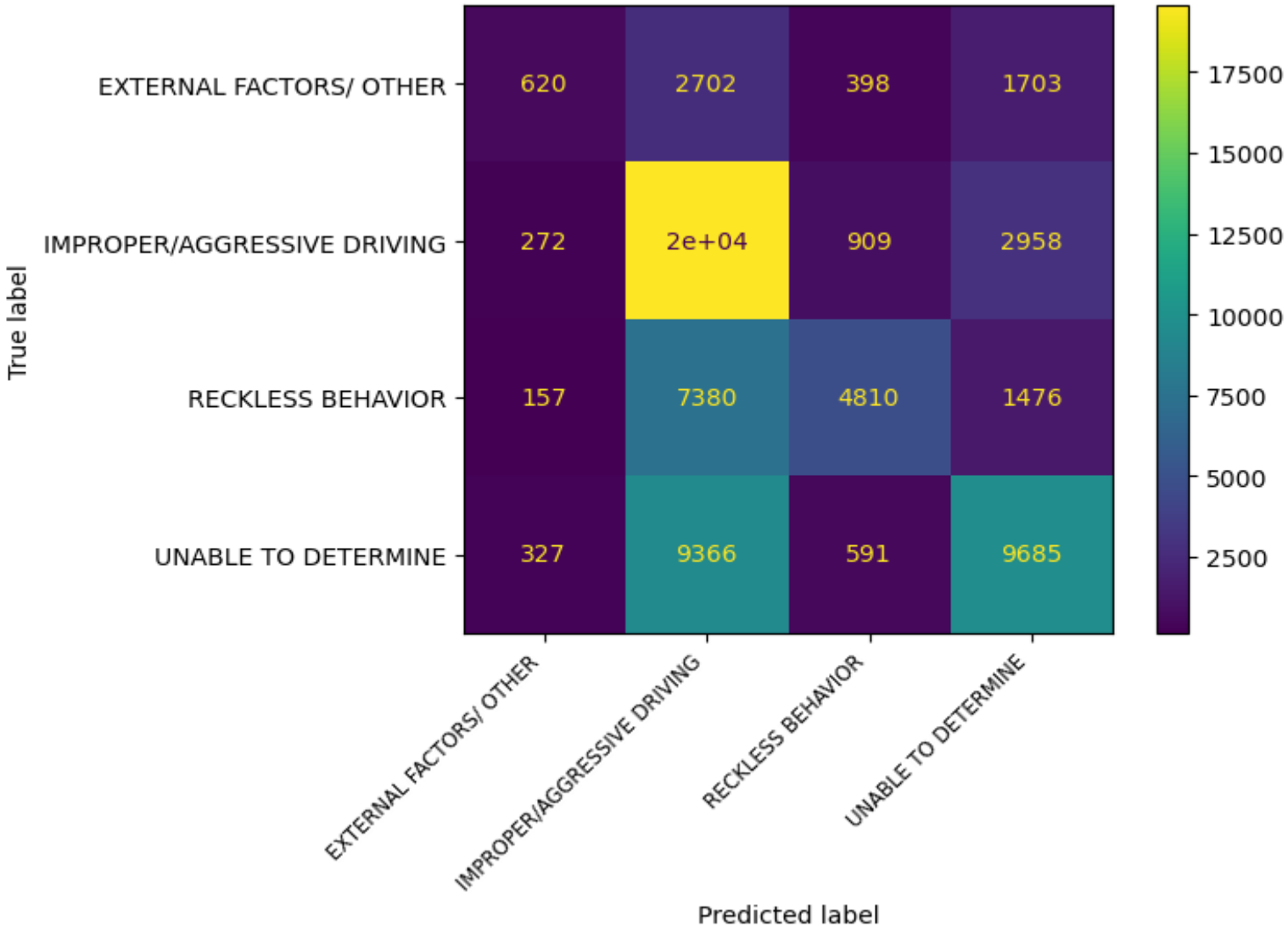


MALE

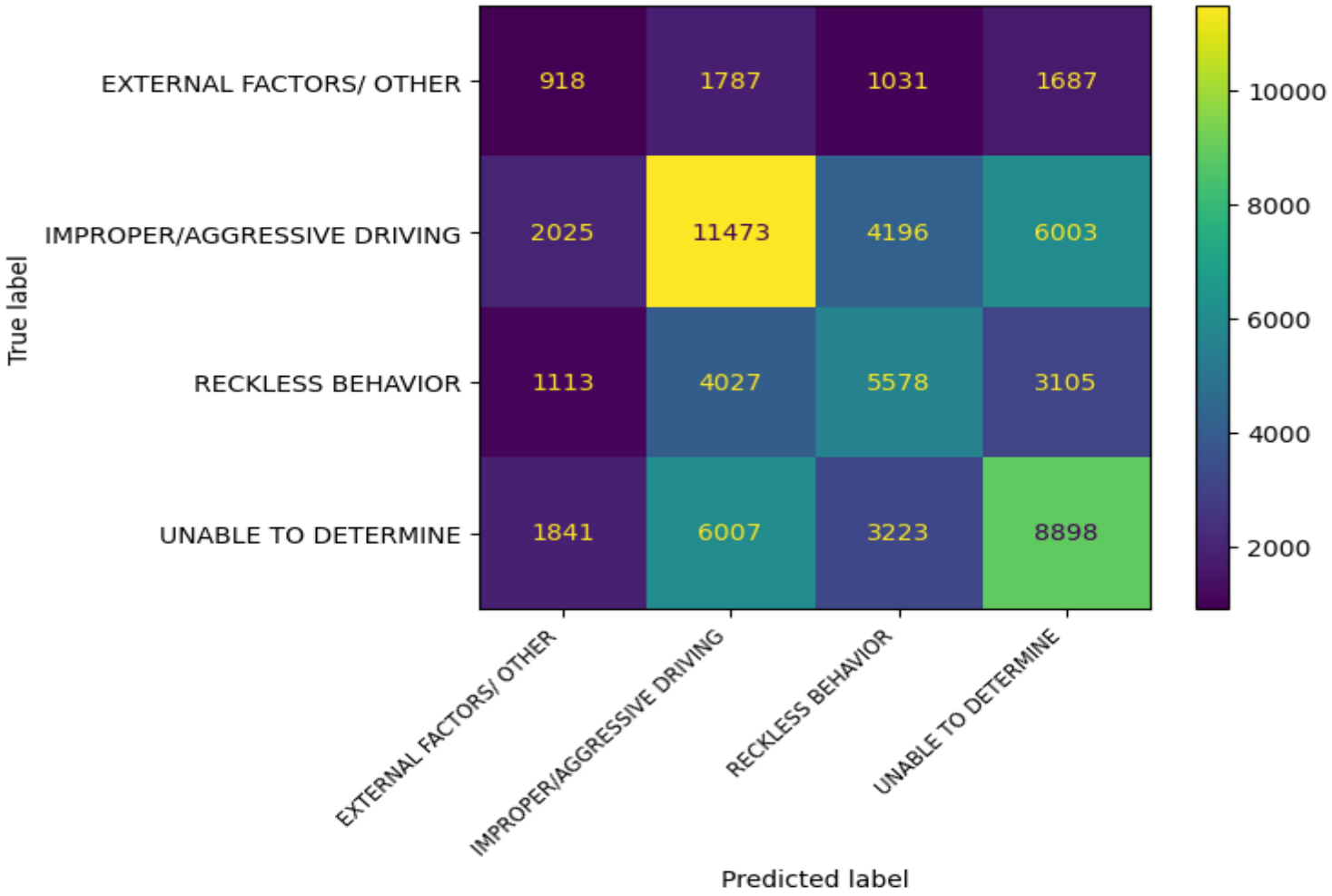
More men caused accidents compared to females. Hence gender is a factor contributing to the cause of accidents.

FEMALE

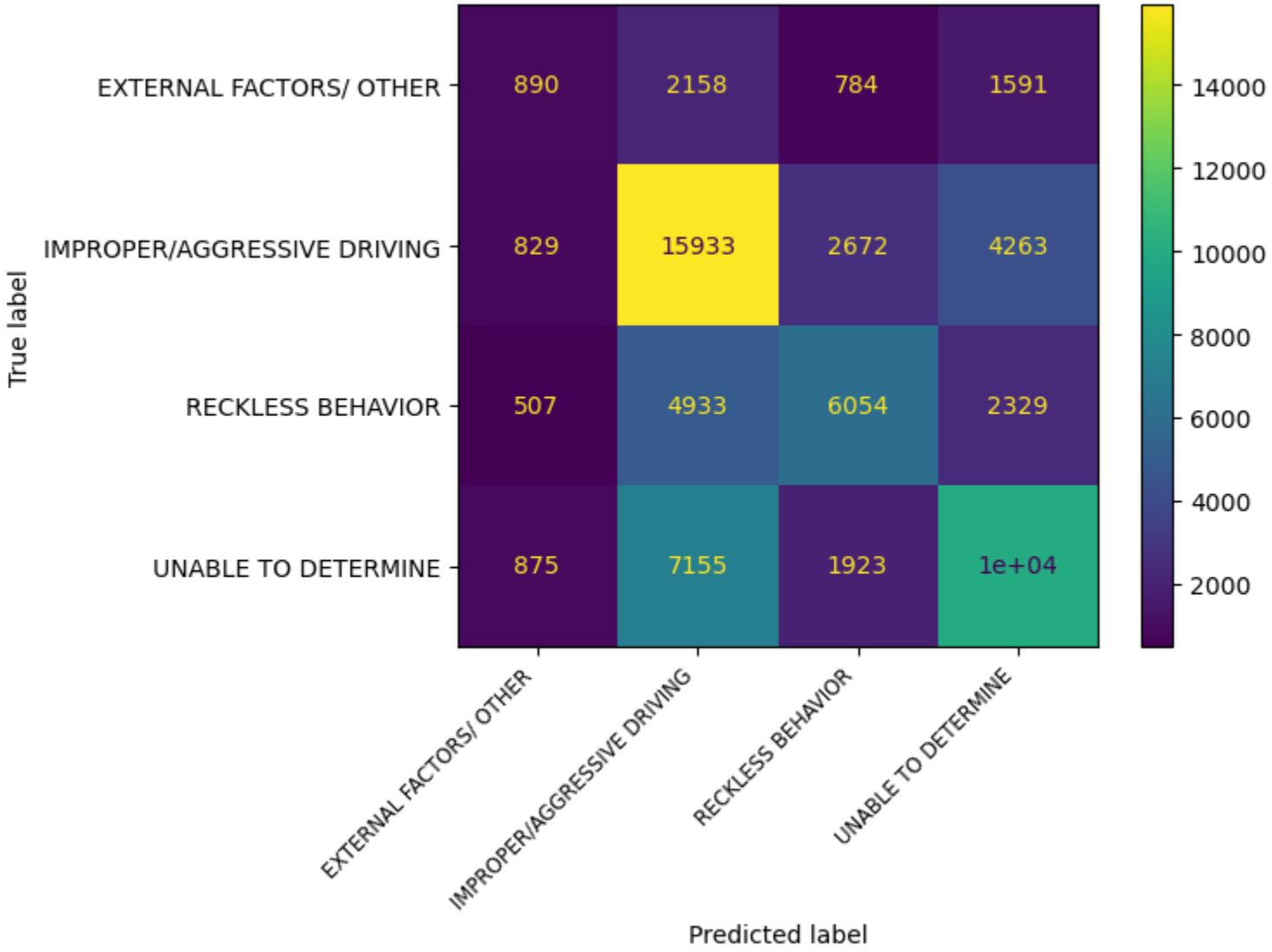
LOGISTIC REGRESSION MATRIX



DECISION TREES MATRIX



K NEAREST NEIGHBORS MATRIX



CONCLUSION

CONCLUSION

After Analysis of this dataset we have come to a conclusion that:

a) Most accidents happened in the year 2015

b) ILLNOIS State had the most car accidents

c) There is a high correlation between weather condition and primary cause of accidents

d) K-NEAREST NEIGHBORS is the model we will use to evaluate our model because it performed better than other models.

e) Males caused more accidents compared to Females.

RECOMMENDATIONS

RECOMMENDATIONS

After Analysis of this dataset we have come up with the following recommendations:

a) Educate the masses on road safety measures.

b) Encourage drivers to take actions that prevent accidents.

c) People should be discouraged from driving during bad weather conditions

d) People should be discouraged from using faulty vehicles.

e) Encourage the male gender and the young people to be careful in their driving.

THANK YOU !