"Visualizing, Modeling, and Forecasting Crashes in Wake County, North Carolina"

Cortlyn Morris, Kelly Wenzlof, Ayan Gulzar

7/14/2021

Contents

Abstract

Studies have shown that road traffic crashes are one of the leading causes of non-natural death for U.S citizens. There are many different ways that road crashes occur, and several various factors lead to it. It is essential to analyze the most common elements of road crashes. In this report, we will model and forecast crashes in Wake County, North Carolina. To accomplish this goal, we will use the North Carolina Wake County crashes dataset from 2015 to March 2021. We were able to observe which factors have a higher cause of crashes in Wake County. We will address the above research question by using histograms of common factors leading to crashes and Time Series Forecasting crashes in Wake County.

Introduction

Car crashes are something that occurs nearly every day. On average, there are 16,438 car crashes per day in the U.S. When it comes to car accidents, several factors correlate. However, it is crucial to understand some of the most frequent causes of crashes. Analyzing the most common reasons can help notify other drivers on the road so that there can be a decrease in crashes.

When it comes to identifying car crashes, different states in the U.S experience crashes at different rates. Road accidents do not only harm the driver of the vehicle. It also affects the passenger and the pedestrians. Most people assume that the driver gets injured the most in a road accident. Passengers and pedestrians are still dying at high rates.

From the beginning of 2015, information on car crashes has been listed in the Wake County car accident database. This data includes the driver's age, gender, vehicle type, and many more critical car accident factors. Our database is based on what police officers have recorded at the scene of the accident. Most of the data found in this database are true, but there is a small percentage of error with some of the data. Our goal is to sort out essential factors from the database to help forecast crashes in Wake County.

In this report, we will be visualizing, modeling and forecasting car crashes in Wake County. Our research questions are 1) What are the main factors that contribute to a crash? 2) Did the law change to obtaining a driver's license due to the COVID pandemic impact crash frequency? 3) What are the conditions that lead to worse outcomes in terms of crash injuries?4) Can we forecast the daily number of crashes in Wake County? 5) Can we predict traffic hotspots in Wake County?.We assembled a dataset from the Wake County crashes database from 2015 to March 2021. Although some of the input in the database is missing, this data will help provide some information about crashes in Wake County. This will also help us forecast car crashes after 2021.

Background

When it comes to car crashes, people assume that the driver was either speeding or not paying attention. It is essential to understand that there are other factors to consider when a car accident occurs. Since everything has slowly started to open after the COVID-19 pandemic, there has been an increase in car crashes. Studying the different factors of impacts can help decrease the number of accidents.

Data, Method, and Analysis

Data

Method

Exploratory Data Analysis

Identifying Hot Spots of Crashes in Wake County, North Carolina

Time Series Visualization

Time Series Forecasting

Statistical Analyses

Modeling Injury

Results

Conclusion

Discussion

References

Acknowledgements

The authors acknowledge financial support from the NSF HBCU-UP ACE: NSF HRD-1719498 and Research Experiences for Undergraduate Students in Data Science Analytics, National Security Agency H98230-18-1-0097.

Appendix