

Analysis of Traffic Accidents and Violations in Chicago

Abhay Rajendra Dixit, Pranjal Pandey, Ravikiran Jois Yedur Prabhakar

1. INTRODUCTION

Traffic violations are a common and inevitable problem in every city. Traffic violations include jumping redlights, overspeeding, overtaking from the wrong side, driving on wrong lanes etc. Often, these violations are directly or indirectly linked to accidents. Accidents are one of the main contributors to global mortality rate. According to the World Health Organization (WHO), approximately 1.35 million people globally die every year as a result of road traffic crashes[4]. Yet, unfortunately, there is not enough effort made to address this issue. The focus is usually on making stricter traffic rules and not on carrying out a root cause analysis of the problem at hand. Most of the existing analysis is based on survey data obtained from self-reported accidents and violations which are usually prone to errors. Hence, it is important that we study the relationship between traffic violations and car crashes at greater depth to gain meaningful insights which might help in taking preventive measures to avoid accidents.

In this project, we study the relationship between red-light violations in particular and the accidents happening in a particular city. For this purpose, we will be picking two transportation datasets namely, Red Light Camera Violations[2] and Traffic crashes[3] of the city of Chicago obtained from the City of Chicago portal[1]. The chosen datasets consist of 550K (approx.) rows and 10 columns of Red Light Camera violations and 400K (approx.) rows and 49 columns of Traffic Crashes. They contain attributes like address, date, time, the traffic control device at the location of crash and the number of red light violations on the respective date, to name a few. The datasets consist of information spanning over the period from 2014 to June 2020. These datasets have different types and formats of attributes like address and street number, missing values, redundant values etc., which would require thorough pre-processing and preparation before it is loaded into our data mining algorithm.

We will be considering various attributes related to violations and accidents such as date, time and location in the aforementioned datasets to achieve our goal. We would be making use of MongoDB as the document-based model as it would be better than a relational model for data analysis because it would be more efficient in handling complex mathematical calculations. Our deliverables include source code to load the data to MongoDB, filter and clean and process the data, mine meaningful information from the data and a detailed report describing the procedures and techniques

involved in implementation of data cleaning and mining algorithms used in this project.

The paper is organised as follows. We are going to talk about the motivation for taking up this topic. This is followed by the design section. We then move on to the implementation section that provides details about the techniques used in cleaning and analysing data. This section is followed by the inferences from the analysis, current status and future work.

2. REFERENCES

- [1] The chicago data portal website.
<https://data.cityofchicago.org>.
- [2] Red light camera violations.
<https://data.cityofchicago.org/Transportation/Red-Light-Camera-Violations/spqx-js37>. Chicago Data Portal.
- [3] Traffic crashes - crashes.
<https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>. Chicago Data Portal.
- [4] W. Pietrasik. Road traffic injuries.
<https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.