DEV THAKKAR-210968148

04/07/23

Project Report

# Title

Accident Locations on Indian Roads

# Abstract

One major requirement of everyday technology is travelling. In a country of 1.4 billion people managing road travelling becomes excessively difficult leading to some inevitable mishaps and accidents. The best way to prevent the inevitable is to reduce it. Thus, the project aims to visualize and analyze the patterns in which we observe the most frequent accident locations in the entire country. These locations that observe frequent accident frequency are called Accident Black Spots and these locations account for the greatest number of deaths and casualties. With a focus on enhancing road safety and reducing the number of accidents, this study leverages geospatial data and statistical analysis techniques to identify accident-prone areas, understand the contributing factors, and propose targeted interventions.

# Introduction

Road accidents pose a significant threat to public safety and have severe socio-economic consequences. In India, the rising number of accidents and their associated casualties have become a matter of great concern. To address this issue effectively, it is essential to have a comprehensive understanding of accident locations, contributing factors, and patterns across the country. Seeing this urgent need and requirement for a deeper study, this project interested me among the others. Seeing the scope of further insights from the available data and the lack of any useful results generated by the government and the road transport ministry motivated me to look into this further. This project also gives me the opportunity to work on real world government data with the potential to have a society impact. A brief introduction to the technical aspects of the project is mentioned below.

Upon research, the data available on government websites is mostly in the form of Portable Document Format (PDF) hence we first need to convert to CSV (UTF-8) encoded format. The data consists mainly of named Black Spots of your specified area which is then preprocessed via certain python scripts to convert into viable query names to be Geocoded via Maps API. Once the geocodes have been obtained we need to filter out the outliers based on the pre-known state boundaries. We also need to remove any miscoded areas on the desired map based on some other name in the country depending on the language. This manual filtering is tedious and can be done for only a small segment of the data and hence the chosen area (state) is Maharashtra. After getting the final Geocodes the final list is appended and then the file is used as a Delimited Layer in QGIS. This shape file is used for visualization and insights.

# Related Work

## Hospitals and Police Stations

Knowing the accident location is not enough if you cannot provide or obtain immediate help from any nearby emergency sources. Hence an additional python script includes the creation of dataset of nearby hospitals for each accident location using the Places API. Further a dataset for all the nearby police stations and government help centers is included separately.

## Additional Analysis

From the dataset available one was found consisting of attributes that allow us to delve deeper into the factors responsible for the accident and the timing for the same. From this we can find the time in which the roads are the least safe and find suitable correlation patterns with other factors that affect the accidents. It is supported by additional visuals and charts included in the python file.

# Data Sources

The following sites have been used as the main source for the data for this project. Other than the mentioned websites further data can be found for each state at the available links on the main page of the morth website. Maharashtra has been chosen as the region of choice for specific analysis.

* <https://morth.nic.in/about-road-safety>
* <https://data.gov.in/>
* <https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents>
* <https://morth.nic.in/road-accident-in-india>
* <https://morth.nic.in/sites/default/files/Maharashtra.pdf>

# Work

After getting the data from the website we first convert it into a Microsoft CSV (UTF-8) encoded file otherwise no manipulation can be performed as the available format is un-editable.

The workflow is roughly divided into the following steps:

## Preprocessing

Once the data is acquired in a suitable format, we have a few issues right off the bat. Firstly, we need to remove any unnecessary formatting that was present in the PDF file and had crept into the CSV document. Having done this, following were the steps taken on the data:

1. Check for NaN values and null values.
2. Remove any regional language names and replace them with English translations if required to the visualization.
3. Remove the additional headers.
4. Preserve the original columns and keep the data as close to the original as possible.

## Geocoding

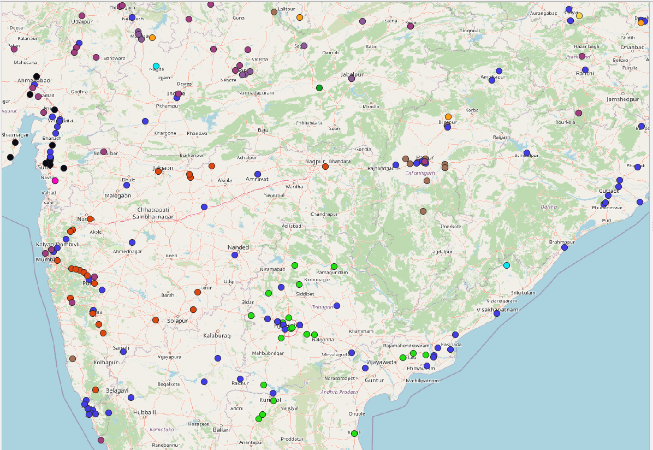
After getting the final list of places (preprocessed) we move on to write a function that firstly converts the name into a viable query name as in some cases the query cannot contain escape characters or spaces and such symbols. Thus, we remove those and replace them with the appropriate character, and we set a substitute array as well just in case the exact location fails to yield the geocode, the substitute array will give the district name or the nearest police station name.

A screenshot of a computer program

Description automatically generated

* As you can see from the image shown, we first input in the two arrays which are chosen attributes from the original dataset.
* The request is then made from the array and then the results if exist are converted into json.
* The lat and lng field is appended into the dataset.
* Since there exist points that have been miscoded, they are replaced and filtered for a small area of choice. (As the process is not possible for a huge scale)

## QGIS

* The consolidated CSV file is added as a delimited layer for each state and then styled accordingly to make it easily visible.
* The X field is set as the longitude and the Y field is set as the latitude and the number of deaths or the number of accidents can be set as the 3rd dimension.
* Additional insights have been generated for the state of Maharashtra mentioned below.

# Additional Insights (Maharashtra)

The following insights were observed along with some proposed solutions while looking at the mapping on QGIS: (graphs shown are included in the python file)

* Maharashtra has had one of the most, if not the greatest number of accidents since 2014.
* Most of the accidents that occur are prevalent in the Mumbai to Pune NH48 highway.
* Major accident locations located in Interior Maharashtra and Southern part of the state have no hospitals nearby to provide immediate first aid.
* Upon API inspection we found that all accident locations near central Maharashtra must be relied on in Solapur District which has no ‘OPERATIONAL’ super specialty hospitals. Hence either building one in Solapur or building emergency care centers near those areas is imperative.
* Police Stations have been well distributed even in the rural areas and most of them are operational.
* Most number of accidents occurred in clear weather conditions contrary to expectations.
* A screenshot of a graph

  Description automatically generatedMost of the accidents happen during 18-21 hundred hours which is not surprising as it gets dark in most parts, however the second most dangerous segment is 15-18 hundred hours which is almost equal if not greater than the previous segment. This is alarming as it is broad daytime, yet accidents are rampant. Also, most of the accidents occur in May which is during Summer. Most of the accidents occurred on surfaced metaled roads under good road conditions. Hence all the possibilities point to one major conclusion that over speeding could be the major cause. Another possibility is that on long stretches of highway drivers are experiencing heat strokes and dehydration during these hours, during these months.
* Solutions could include speeding signs and extra speed breakers on long stretches.
* Installing speed cameras

# References

* Identification and Analysis of Accident Blackspots Using GIS Prof. Leni Stephen1, Gladwin G Kelakom2, Jency Maria Sojan3, K.S Sreelakshmi 4, Vishnu N.B5
* Apparao. G, P. Mallikarjunareddy, Dr. SSSV GopalaRaju.” Identification Of Accident Black Spots For National Highway Using GIS”, International journal of scientific & technology volume- 2, pp154-157