Exploratory Data Analytics- Real Time Traffic incidences
By Manzi Patrick

Instructions

1. Go through the dataset and explain what you find, a brief explanation on what you think about the data and what you can possibly analyze about the data.

2. Explain which tools you will use to analyze the data and why the chosen tools.

3. Load the dataset to your preferred tools as explained above and perform your data wrangling, please supply code or steps used to wrangle the data.

4. Please supply code or methods used to perform your analytics.

5. Provide findings and visualizations of your analysis.

Report

1.A Introduction

This dataset entitled "Real time traffic incident" contains information regarding traffic reported issues from 2017 to 2022 where incidences reported ranges from vehicle collision, traffic hazards and many more. The dataset contains 9 attributes with 257651 records.

### 1.B Possible analysis that can be made about data

Dashboard depicting summary statistics from the data.
Annual, monthly and hourly time series of traffic incidences.
Areas/location with high traffic accidents.
Categories of traffic incidences that occurs mostly.
Safe places where minimal incidences were reported.

2.A Methodology

In this assessment, python programming language was used as it is believed to have powerful tools to conduct data analysis that goes beyond simple data analytics. A variety of techniques for data preprocessing was applied for tracing and handling missing data. For data cleaning, I employed diverse functions such as datatime, tools like pandas and NumPy. Similarly Different functions like info and value count were used to detect missing values however the missing values was not deleted or replaced since it wouldn't contribute anything to analytics. Most of visualization were achieved by python libraries named matplotlib and seaborn albeit few graphs such as pie chart were built in MS excel and the dashboard that display summary statistics was designed using adobe XD.

3. Exploratory data analytics – results - findings

Dashboard depicting summary of traffic incidences along with annual time series.
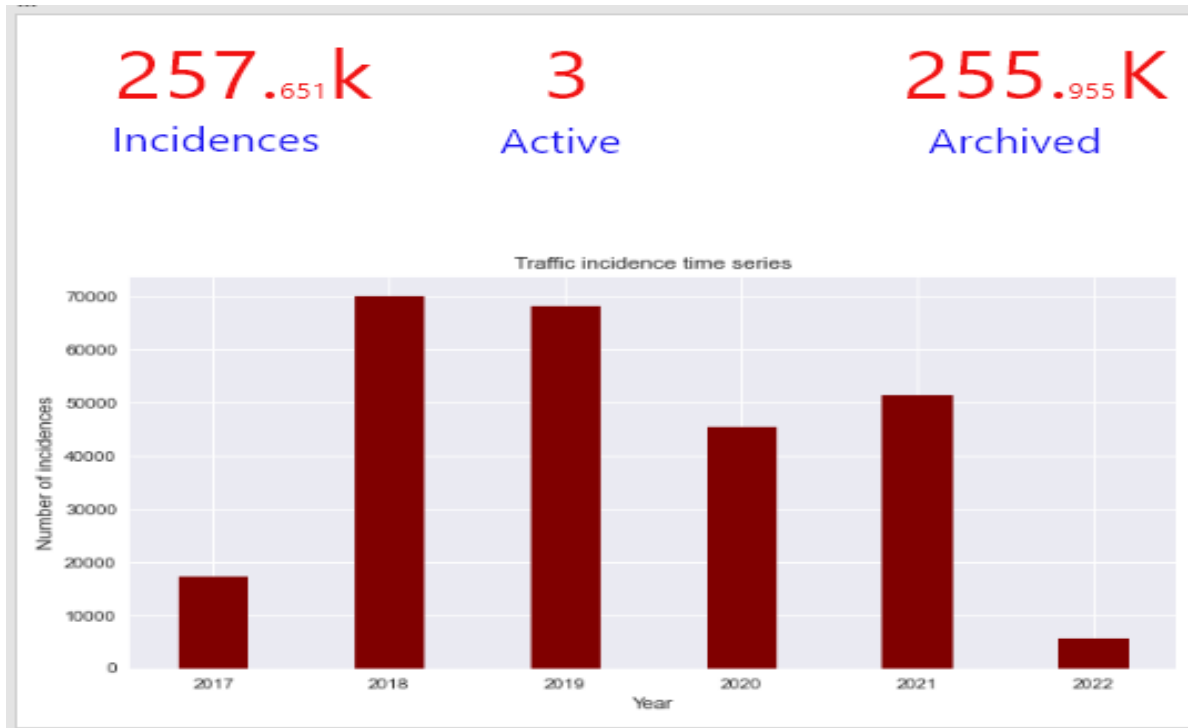
This figure depicts total number of incidences, active and archived ones. It also shows that 2017 had minimal traffic incidences compared to 2018 and 2019. It also shows that in 2020 traffic incidences reduced, this might be due to several factors such as lockdown that restricted massive movements. There was slight increase in 2021 probably because several nationalities stopped lockdowns and resumed almost normal operations. Albeit it seems that 2022's traffic incidences reduced but that is not the case, instead it was caused by the dataset that only includes the months of January and February of 2022.

Top 5 places with many accidents
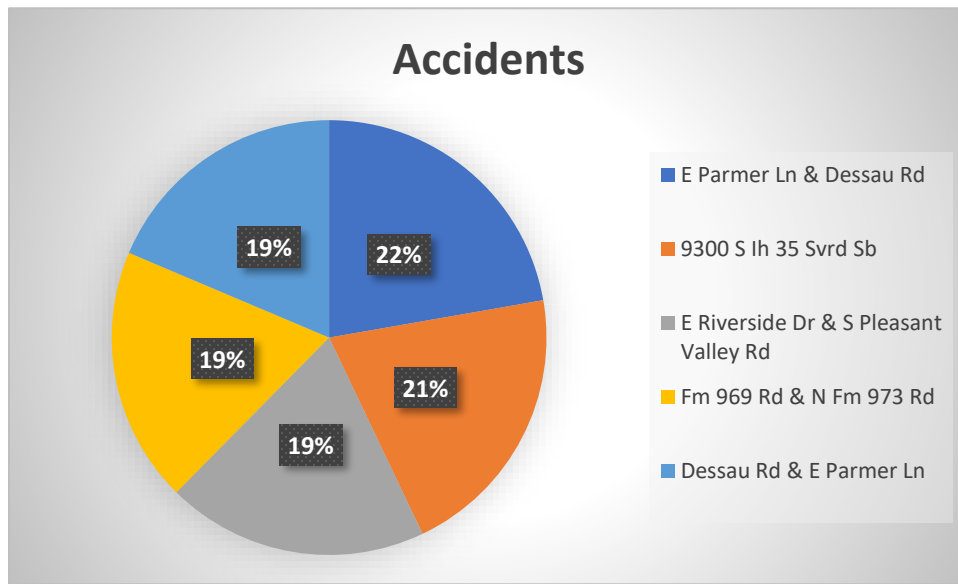
This pie chart shows the percentage of top five locations with many reported incidences, it is apparent that E Parmer Ln & Dessau Rd is the area where many accidents are likely to happen.

Places with less traffic accidents

Address	7200-7324 Red Pebble Rd	11100 - 11106 SANTA CRUZ DR	13301 Mizzen St	4206-4320 Greystone Dr	900-1008 S SH 130 SVRD SB
Counts	1	1	1	1	1

## Traffic incidents that happen mostly and their frequency
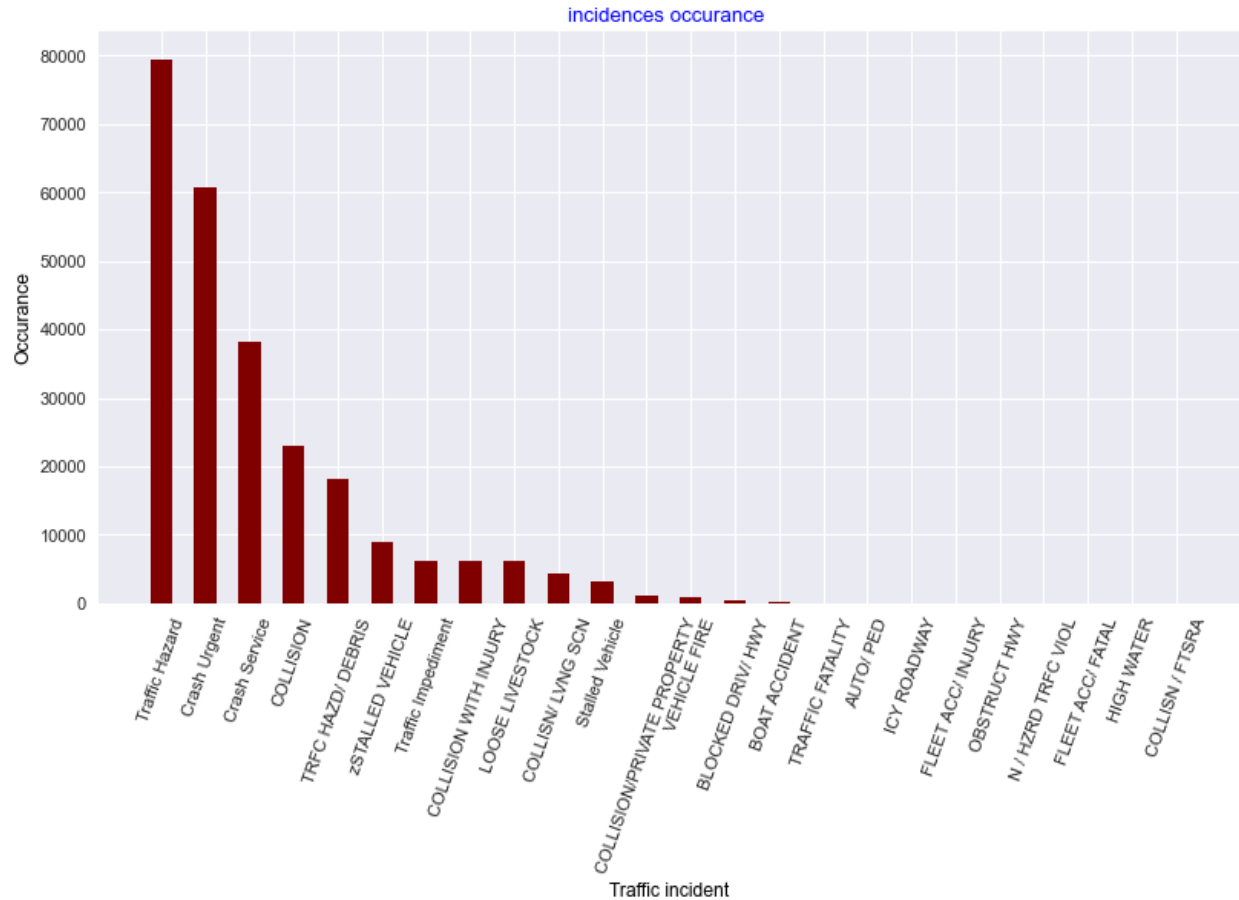
This graph above depicts frequency of real time traffic incidences from 2017 till the beginning of 2022. It shows that traffic hazard incident is the one that mostly occurred whereas collision/FTSRA is the one that happened less.

Monthly time series

Monthly time series of traffic accidents
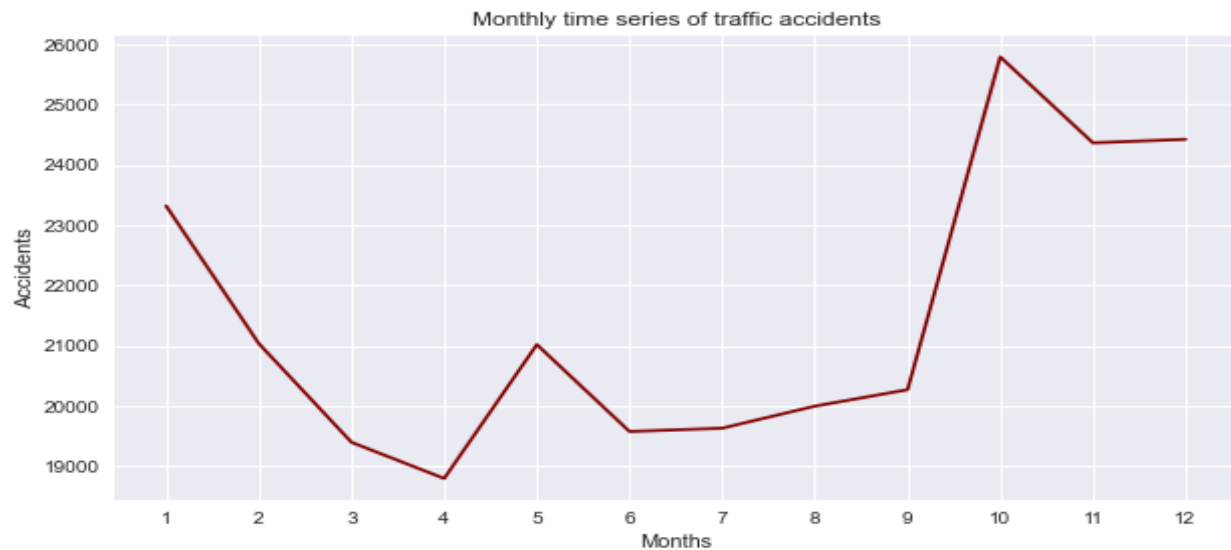
This is a time series graph that illustrates how traffic incidences varies monthly. It shows that many accidents take place in October whereas less incidences take place around April. This can mean that there are so many accidents at the beginning and end of year compared to middle months.

Hourly Time series.

Hourly time series of traffic accidents
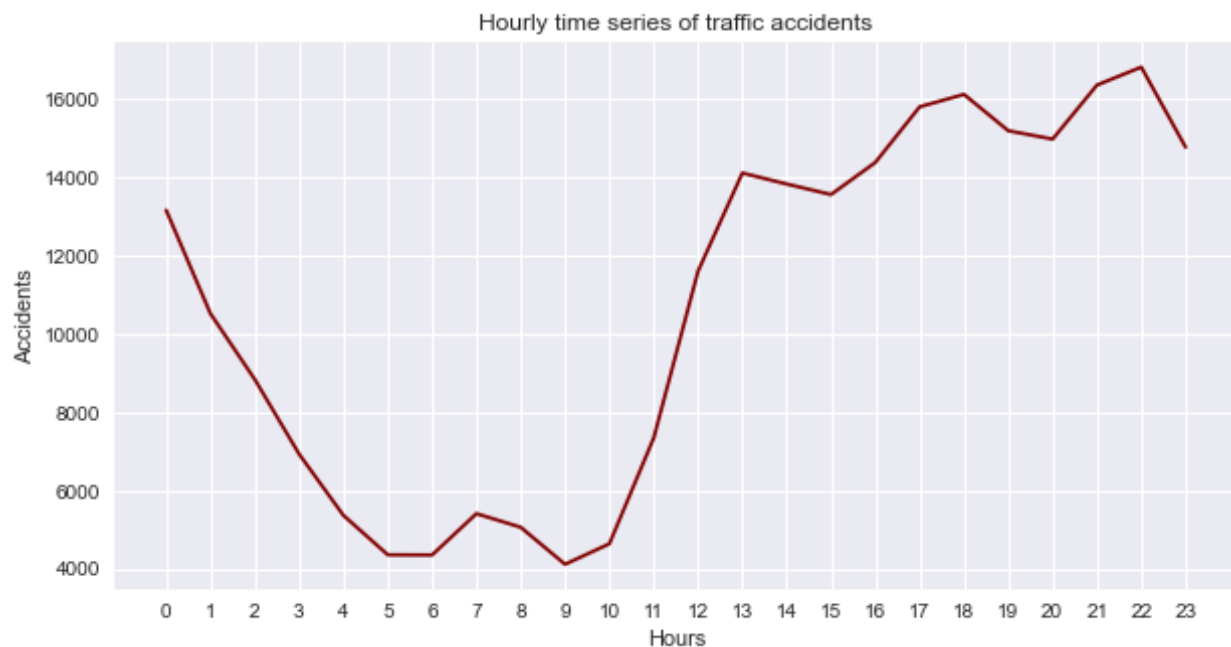
The above hourly time series graph shows occurrence of traffic incidences on hourly basis. It shows that the pick is 21h while the hour with minimal accidents is 9 am. The graph generally shows that most of traffic incidences take place afternoon.

limitations

Due to limited time allocated for assessment, graphs are not interactive.

Dashboard is not designed as it is supposed to let's say using python- Dash -plotly framework or PowerBI.

Modeling was not performed (no machine learning models for forecasting incidences that were built)

No automation of analytics tasks/not sure if needed at this stage.

Insight might not be as perfect as they would be when given enough time.

4. Conclusion

Many economies are expanding rapidly. As a result, cities in most countries are growing at an unsustainable rate in terms of population and size. This means that as the city's population and size grow, so will **road congestion** in terms of the number of people and vehicles on the roads. This implies that the **number of accidents** in cities will rise as well. Now, this exploratory data analytics aims to address the expected increase in accidents by finding and communicating insights about traffic incidences in real time hence **enabling governments to make decisions on interventions to implement that would contribute to reduction in traffic incidences**. This analysis was made possible through use of python tool, MS excel and Adobe XD.