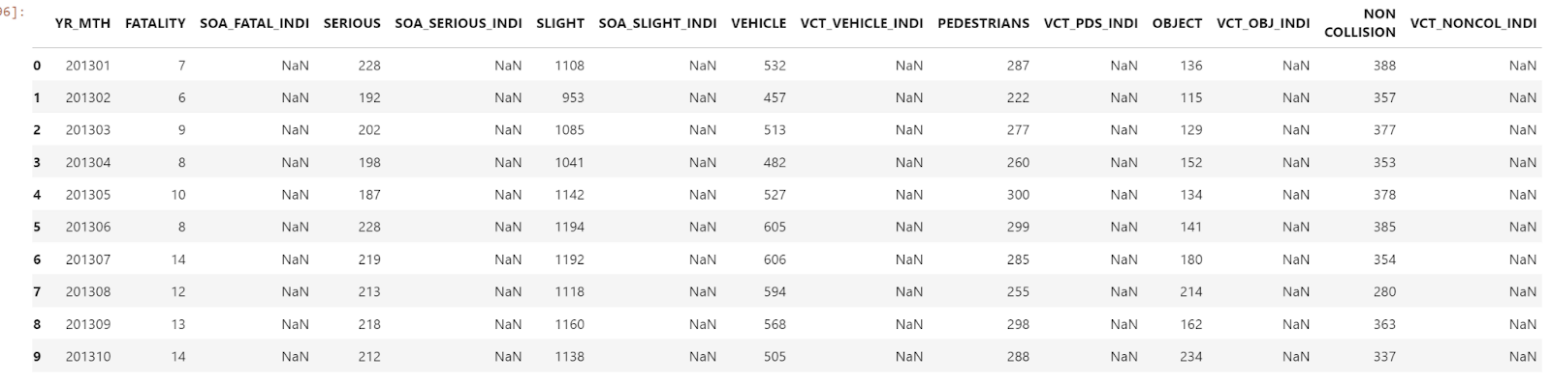
***The road traffic accident prediction in Lion Rock Tunnel***

Objectives: The goal is to anticipate future road traffic accidents in Hong Kong's Lion Rock Tunnel using past data. This includes recognizing trends and patterns, strengthening safety measures and raising awareness through road safety education. The goal of forecasting accidents is to show the prediction of road traffic accident at Lion Rock Tunnel from 2020 to 2047

Data preprocessing:

Dataset used: [Section 7 : Road Traffic Accident Statistics - Table 7.1 - Road Traffic Accidents by Severity and Type of Accident Collision(English)](https://data.gov.hk/en-data/dataset/hk-td-tis_17-monthly-traffic-and-transport-digest-csv/resource/9a13e02d-3f92-46ad-9a62-73b8854011fc) (data.gov.hk)

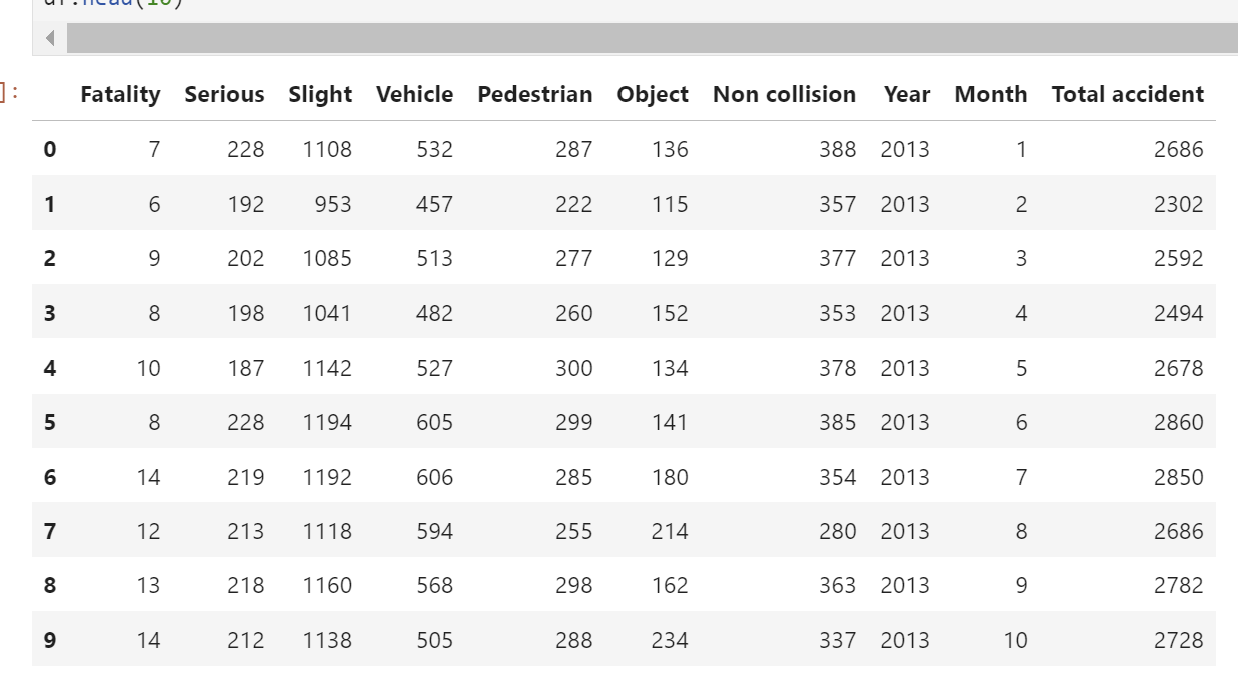
The dataset shows types of traffic accidents cause from January 2013 to January 2024



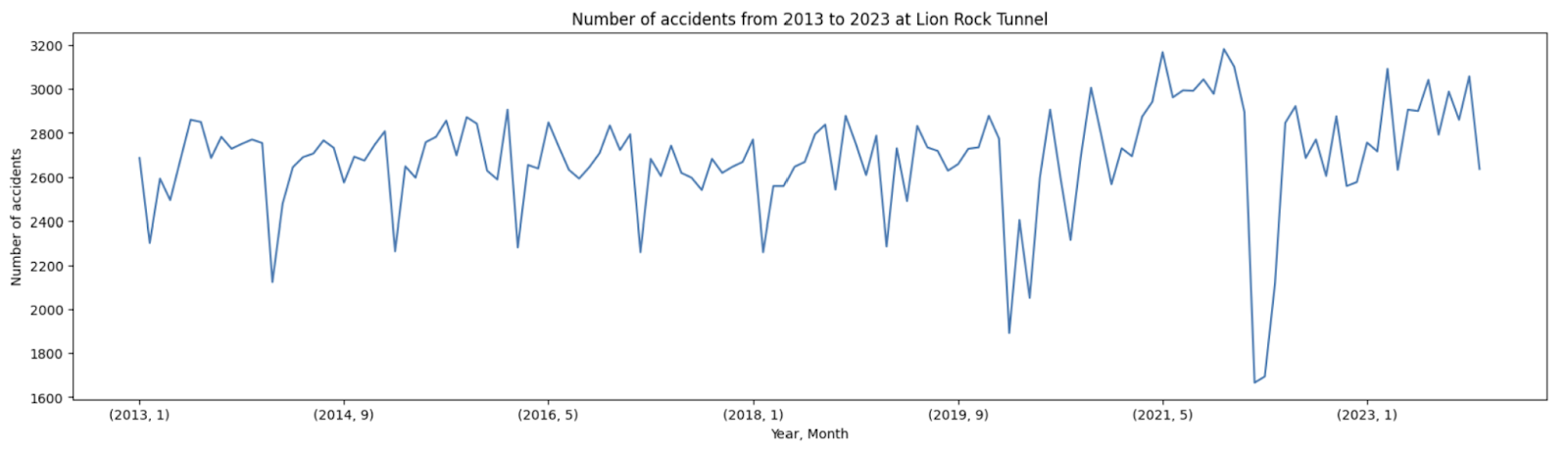
Data cleaning:

This task only uses a dataset from 2013 to 2023 because of the lack of 2024 data.

In detail, for the data cleaning, a number of columns with null value (NaN) are removed. The column “YR\_MTH” is deleted and replaced by two columns named “Year” and “Month”. These two new columns are extracted from the column named “YR\_MTH”. Next, sum up total accidents of 7 types of accidents for each row which reflect total accidents that happen each month of each year. The final dataset includes the number of accident occurrences for each cause including: Fatality, Serious, Slight, Vehicle, Pedestrian, Object, Non collision and total accident happens with an aim to predict traffic accidents in next years.

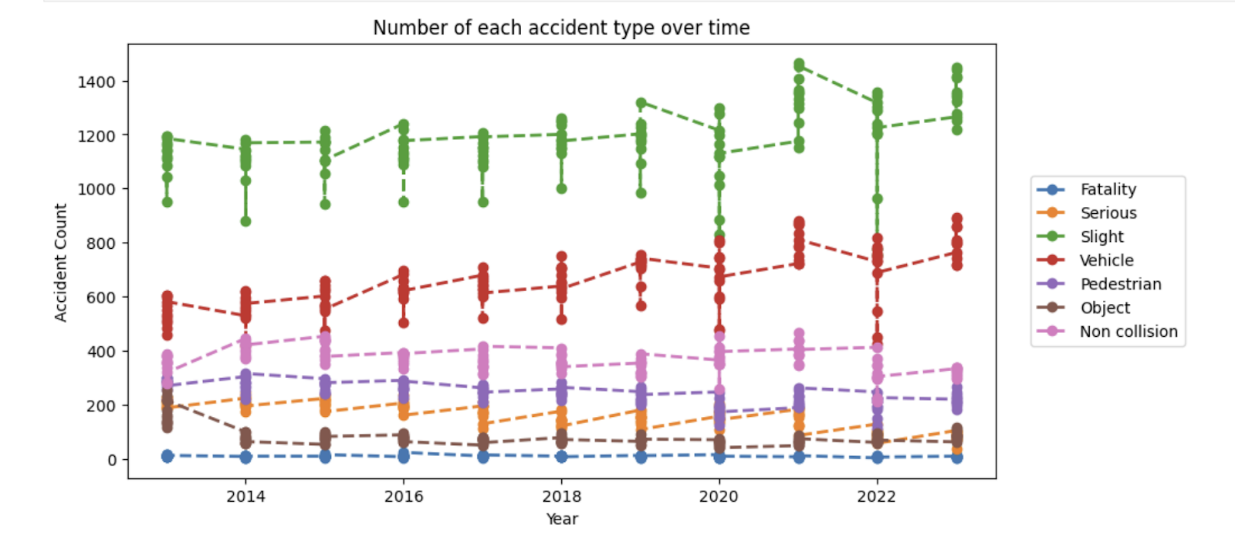


Data visualization:



  Graph 1: Number of accidents from 2013 to 2023 at Lion Rock Tunnel

Based on the information provided, it appears that Graph 1 illustrates the fluctuation in road traffic accidents in Lion Rock Tunnel, Hong Kong. The graph displays a consistent pattern of around 2800 accidents per month. Moreover, the highest number of accidents, approximately 32,000, occurred towards the end of 2021. On the other hand, the lowest number of accidents, less than 1500, was recorded at the beginning of 2023. This data suggests that there is an upward trend in accidents during the recent period and potentially in the future.



  Graph 2: Number of each accident type over time (from 2013 to 2023)

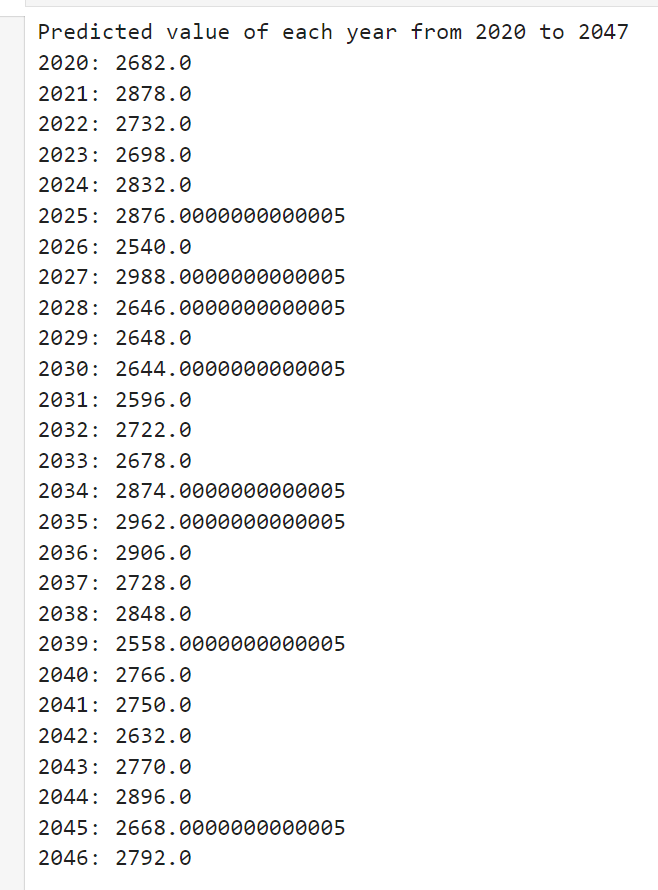
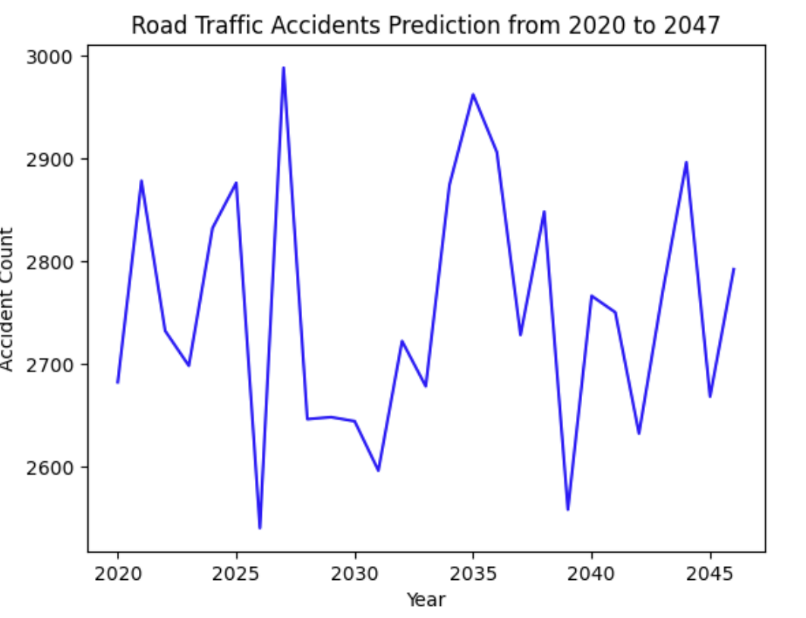
The graph shows that the most common form of accident in the current decade is slight accidents, which are indicated by the green line. Slight accidents have the largest number of occurrences over the time period depicted in the graph. The next most common form is vehicle accidents, as shown by the red line. Lastly, non-collision accidents, as illustrated by the pink line, are another form of accident that occurs frequently. Non-collision accidents have a continuous number of occurrences, however they are often lower than minor and vehicular accidents. The blue line represents fatality accidents, which are the least common form of accident. Fatality accidents have the fewest occurrences over the given time period, showing that they are a relatively rare event.

Data analysis:

Linear Regression will be applied to predict data with features (X) as accident types and target variable (Y) as total accident then training, testing sets and linear regression model are created and fitted. Lastly, predictions are made on the test data to show predicted values as below.

Reason for Selecting Linear Regression for Data Prediction (coding shown in Python)

|  |  |
| --- | --- |
| MAE:  1.3473995581821158e-13 | The MAE calculates the average absolute difference between expected and actual values. In this situation, the MAE is around 1.35e-13. It implies that the model functions well and exhibits a high degree of accuracy. |
| MSE:  6.127263796689094e-26 | Similarly, MSE calculates the average squared difference between expected and actual values. In this situation, the MSE is about 6.13e-26; a lower MSE suggests that the model's predictions are quite near to the actual values, with little variation. |
| RMSE: 2.475331047898259e-13 | The square root of the MSE, RMSE, represents the average magnitude of prediction mistakes. In your situation, the RMSE is roughly 2.48e-13. A decreased RMSE, like a lower MSE, shows that the model's predictions are very accurate and have few errors. |

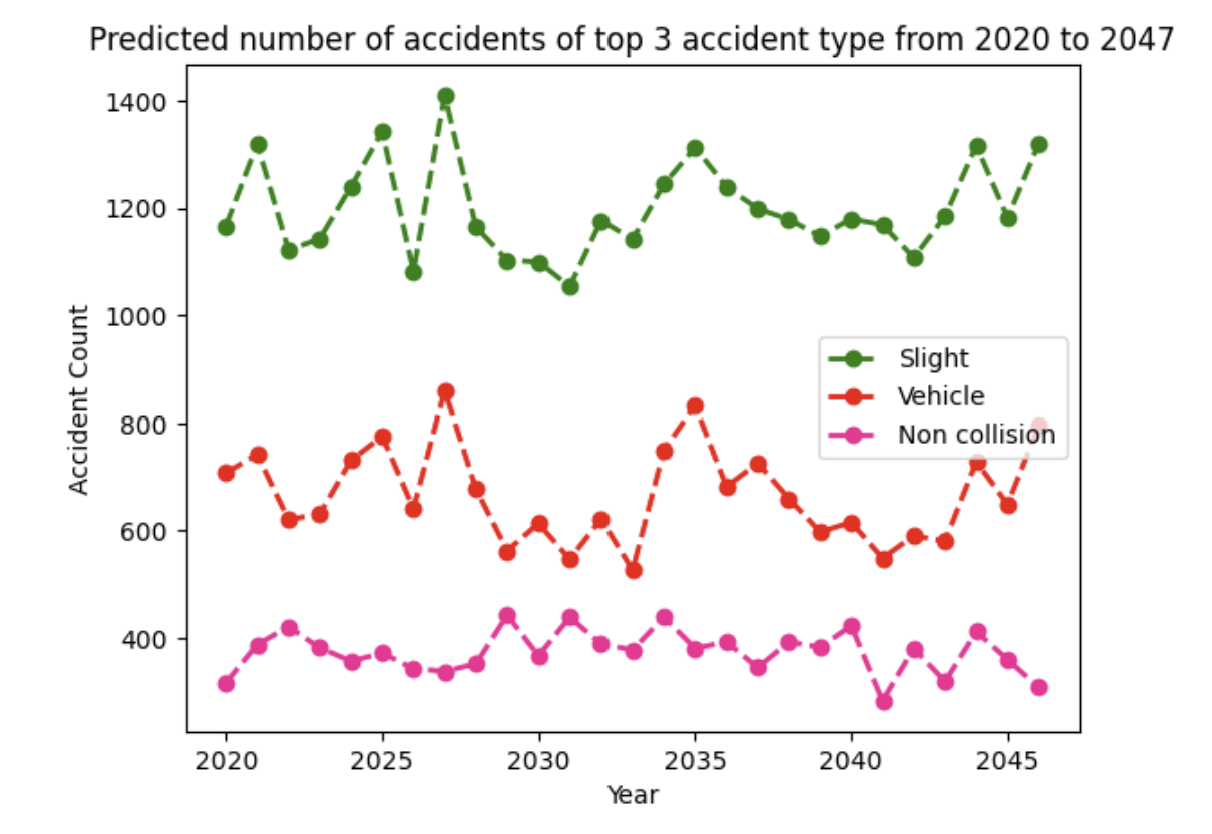


Graph 3: Road traffic accidents prediction from 2020 to 2047

**Results**: The study of the anticipated values using the Linear Regression model demonstrates a steady increase trend in the number of road traffic accidents over time. The graph shows an upward trend that peaks in 2041. The prediction shows that there will likely be an increase in traffic accidents in the future.

Next, we will predict the number of accidents that are able to happen based on the top 3 accident types including slight accident, vehicle accident and non collision accident as shown in data visualization part (Graph 2).

Linear Regression will be applied to predict data with features (X) as accident types and target variable (Y) as slight accident, vehicle accident and non collision accident respectively then training, testing sets and linear regression model are created and fitted. Lastly, predictions are made on the test data to show predicted values as below.



Graph 4:  Predicted number of accidents of top 3 accident types from 2020 to 2047

**Result**: From the graph, we can see the number of slight accident types (green line) and non collision type (pink line) will keep 1200 accidents and 400 accidents respectively as compared to its figures in Graph 2. In contrast, the predicted number of vehicle accidents will increase approximately 700 accidents, more than 200 accidents as compared to its figures in Graph 2. This suggests that vehicle accidents are likely to become more common and pose a higher risk.