

Research Question:

Which Roads needs Traffic Signs in  
Victoria?

# Structure

- Objective
- Why is it worth tackling and who would care?
- Dataset
- Visualisations
- Wrangling Plan
- Risk and Challenges

# Objective

- Investigate which roads in Victoria is needing traffic signs based on accident and traffic volume statistics.
- Help us understand the current livability state of Victoria as we find out the current state of roads in Victoria
- To an extent, improve the livability in victoria as in the right hands , it can improve state of roads through out Victoria as it may reduce traffic and reduce accidents

# Why is it worth tackling and who would care?

- This research will be most useful to traffic management department such as VicRoads to make decision on improving traffic
  - It will narrow down the location where traffic lights are needed
  - Therefore, it will possibly help narrow down the amount of traffic light simulation needed to be run

# Dataset

- Traffic Volume in victoria

- Type: csv
- Size: 14,675 rows
- Info: Average Annual daily traffic for all different type of vehicles in each road
- Organization: VicRoads
- Link: [https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/5512df2ff41e4941bacf868053dbfba9\\_0](https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/5512df2ff41e4941bacf868053dbfba9_0)
- Metadata: <http://data.vicroads.vic.gov.au/metadata/evol.html>

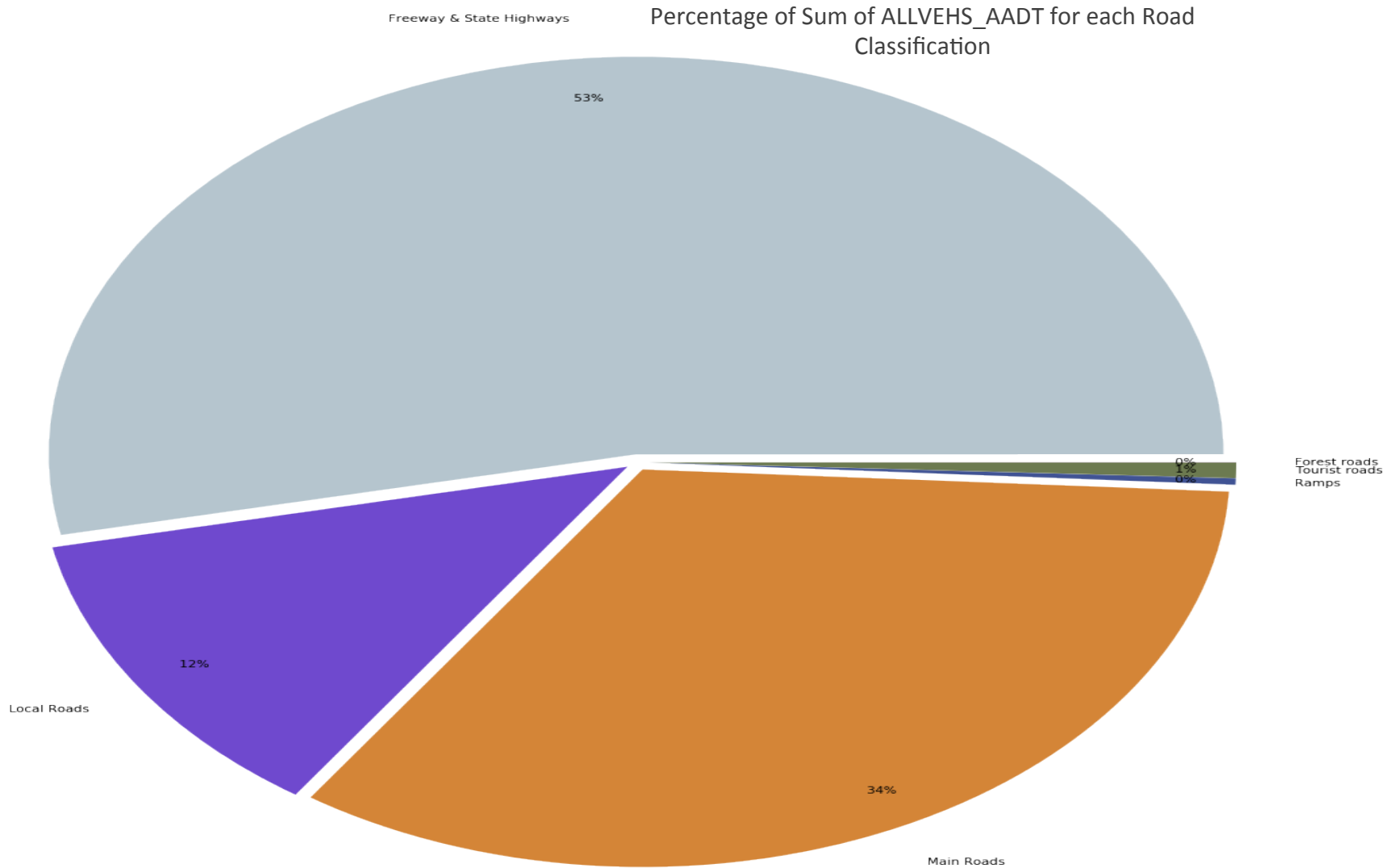
- Crash Statistics

- Type: csv
- Size: 176,938 rows
- Info: Location of accidents
- Organization: DataVic
- Link: [https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/1f3cb954526b471596dbffa30e56bb32\\_0](https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/1f3cb954526b471596dbffa30e56bb32_0)
- Metadata: <http://data.vicroads.vic.gov.au/metadata/Crash%20Stats%20-%20Data%20Extract%20-%20Open%20Data.html>

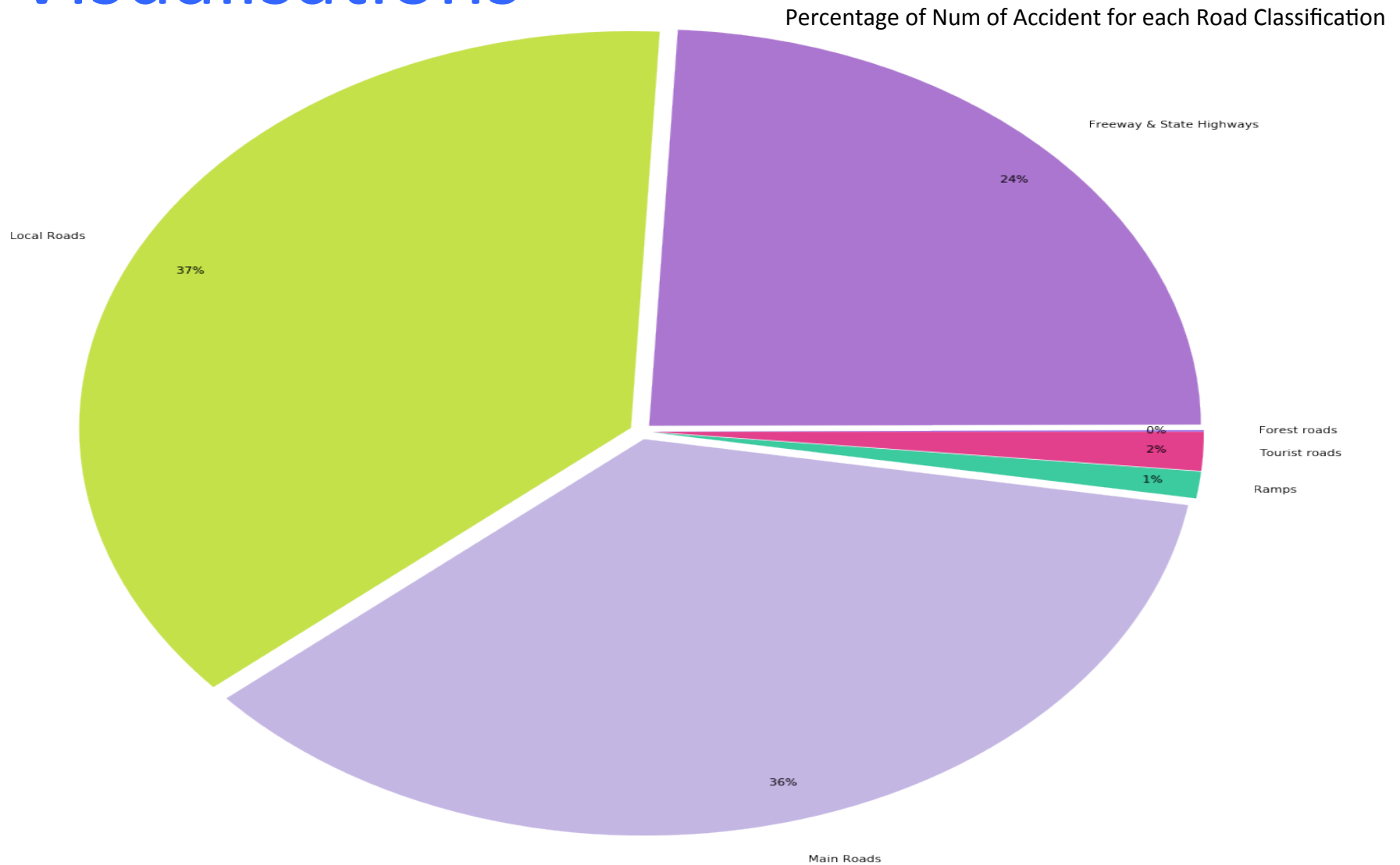
- Traffic Lights in victoria

- Type: csv
- Size: 4,582 rows
- Info: Shows the location of traffic signals across Victoria
- Organization: VicRoads
- Link: [https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/1f3cb954526b471596dbffa30e56bb32\\_0](https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/1f3cb954526b471596dbffa30e56bb32_0)
- Metadata: [http://data.vicroads.vic.gov.au/metadata/traffic\\_lights.html](http://data.vicroads.vic.gov.au/metadata/traffic_lights.html)

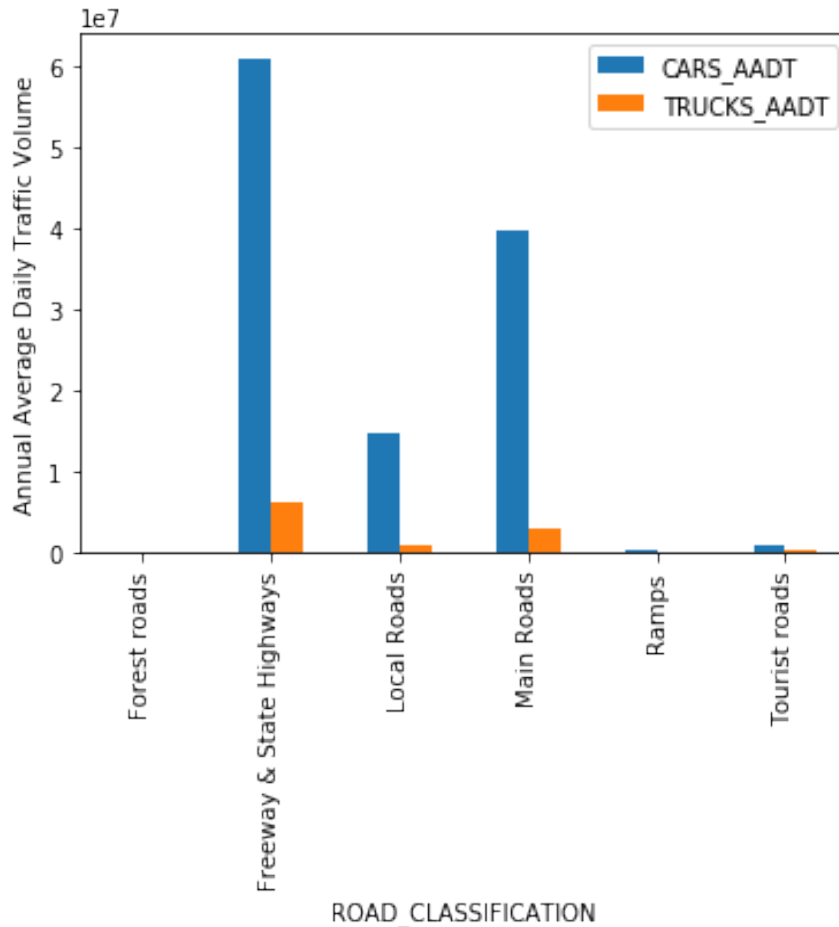
# Visualisations



# Visualisations



# Visualisations



--- Top 10 Road Accident ---

- 1: PRINCES 4438 accidents
- 2: HIGH 2730 accidents
- 3: NEPEAN 2107 accidents
- 4: SPRINGVALE 1447 accidents
- 5: SOUTH GIPPSLAND 1409 accidents
- 6: MONASH 1398 accidents
- 7: SYDNEY 1381 accidents
- 8: BURWOOD 1339 accidents
- 9: DANDENONG 1248 accidents
- 10: MAROONDAH 1172 accidents



# Wrangling Plan

- Data cleaning
  - Case deletion on accident location dataset
    - Accident location dataset has small records where the roads is not specified
    - Meaning the accident may not exactly happen on the road, thus removing would be the best option
- Integration
  - Join Traffic Volume and Accident Location dataset through blocking of road number and road name
  - Using group by method on both data set
  - Create Percentage Likely Accident feature for every blocking
  - This is better than raw data as we can now do analysis and prediction on this data

# Wrangling Plan

- Correlation on Number of accidents and Annual Average Daily traffic
  - By using both mutual information and Pearson correlation on ALLVEHS\_AADT and Percentage Likely Accident both discretized with equal length method using appropriate bin
  - Aiming to know if both factors actually has correlation
  - Expecting Heat maps of pearson correlation to be fairly clustered but not mostly clustered as accident may not be caused by amount of vehicles
  - Line graphs of ALLVEHS\_AADT and Percentage Likely Accident to be more towards increasing straight line as number of bins goes up
- Make a model using train and test split methodology on existing traffic lights
  - Set a data frame where the traffic signs location with Percentage Likely Accident and average annual daily traffic of vehicles
  - Run the train and split function with possibly 80% as training set
  - Run K-nearest neighbor on new roads to show the accuracy of the conditions of new roads to be placed with traffic light to current roads with traffic lights
  - Expecting list of roads with sorted from high to low of accuracy

# Risk and Challenges

- Finding up to date open data
- Political and budgeting decision
  - As receiving audience has the final decision, it may trump over this research
- Not all factors are considered
  - E.g: Parking, reputation