Objective:

* This research will identify which roads in Victoria is currently needing traffic signs based on accident and volume statistics
* This question will help us understand the livability as it find out about current state of roads
* To an extent in fact this research question can improve livability as it may reduces traffic and accidents in the roads

Why and who:

* This research will target audience traffic management department such as VicRoads to aid them in decision making of placing new traffic signs
* It will narrows down the hotspot locations where traffic lights is needed
* Therefore reducing amount of traffic light simulations needed for them to work

Dataset:

* First data set would be traffic volume in victoria
* This data contains the average annual daily traffic , AADT for short, for every road in victoria which is needed to know the traffic conditions in victoria
* Second data set would be crash statistics
* This data set has a lot of files relating to accident but we are interested with ACCIDENT LOCATION file where it stores the location of every accident
* We want this dataset to know how likely accident is going to happen in that road
* Third relevant dataset would be traffic lights in victoria
* This data set contains the information about current traffic signs that is in victoria
* Through this dataset we know where are the currents traffic signs locations

Visualising:

* This would be the visualization of my dataset
* First one represents sum of all vehicles average annual daily traffic for every classification of roads
* We can see that free way has the majority with main roads (53%) and local roads and main roads has quite large with 12 and 34 percent respectively. Tourist roads, ramps and forest roads has very little proportion, thus explaining why we don’t really see traffic lights there
* Next one represents percentage number of accident for every classification of roads
* Proportion is rather even between local main and freeway
* But similarly we see that forest tourist and ramp roads has very little proportion, further enforcing why we rarely saw traffic signs on those roads
* Next graph would be how cars and trucks affect the AADT
* And top 10 roads with most accident

Wrangling plan:

* My plan was firstly to implement data cleaning method of case deletion
* Since accident location data has a small record of roads that is not specified
* This means that accident may not exactly happen in that road
* Thus removing would be the best option
* Next I would intergrate my dataset traffic Volume and Accident Location through blocking of road number and road name
* Using group by method for both datasets
* I want to also create a new feature of percentage likely accident will happen for every block
* This is definetly better as we have room to do analysis, prediction on this data
* Next was to find correlation in number of accidents and AADT to know if there is actually a correlation
* Using mutual information discretized with equal length method and pearson correlation
* Expecting heatmap graph where it is fairly clustered but not mostly clustered as accident may not caused just by traffic volume
* Also expecting line graph to be more of a straight line with positive gradient as number of bins goes up
* Lastly I’m planning to do model by using train and test split on existing traffic light data.
* I would first need to join traffic light with other dataset to know the conditions of the traffic sign roads
* Run train and split function with possibly 80% training 20% test set
* Then run K-nearest neighbor to show accuracy of conditions of new roads to current roads
* Expecting list of roads with sorted from high to low of accuracy

Risk and Challenges

* Challenge of this research would be finding appropriate open data which are also up to date. Or at least the same date time.
* Another challenge would be political and budgeting decision
* Final decision goes to receiving audience, this will always trump evidence based research
* Another risk is that not all factors can be considered
* There are subjective options such as reputation of the road and available parking on the road may affect the dataset

Thus concludes my pitch, thankyou

Who: VicRoads (is a [statutory corporation](https://en.wikipedia.org/wiki/Statutory_corporation) which is the road and traffic authority in the [state](https://en.wikipedia.org/wiki/State_(Australia)) of [Victoria](https://en.wikipedia.org/wiki/Victoria_(Australia)), [Australia](https://en.wikipedia.org/wiki/Australia)) responsible for maintenance and construction of the arterial road network, as well as [driver licensing](https://en.wikipedia.org/wiki/Driver%27s_licence_in_Australia#Victoria) and [vehicle registration](https://en.wikipedia.org/wiki/Vehicle_registration_plates_of_Australia#Victoria).

specifically for traffic management department.

Wrangle:

Plan on slice location on current traffic signs then compare to other dataset