**Project 1 Proposal**

**Team number:** 4

**Members:** Marcus, Tamer, Ivana, Sanuli

**Project title:** Victorian Road Accidents Analysis

**Project description (broad)**

* Analyse causes, locations and timing of road accidents in Victoria and whether weather conditions have an impact on their frequency/severity during the years 2014-2019.

**Research question to answer (specific)**

* What do you plan to show with your data analysis?
  + Graphs of road accident type/location/timing, road accident frequency vs various weather conditions, road accident severity vs various weather conditions
* What research questions do you plan to answer? That is, what are your hypotheses?
  + Are their areas of Victoria where accident frequency is higher?
  + Is accident frequency higher on weekdays or weekends?
  + Does weather (rainfall, temperature etc) have a significant effect on accident frequency or severity (fatality)?
* What do you expect to see as a result of your analysis?
  + Expect **higher** accident frequency in Melbourne CBD compared to other areas (more cars on road) – Sanuli (histogram/bar chart, t-test or anova)
  + Expect **higher** frequency of road accidents during weekdays (more cars on the road). – Marcus (boxplot, t-test or anova)
  + **Increased** fatality/accident frequency during wet weather – Ivana (regression, scatter, correlation, two-sample)
  + **Increased** fatality/accident frequency during cold weather – Tamer (regression, scatter, correlation, two-sample)

**Datasets required**

* What data sets will you access?
  + VicRoads Open Data – Road Crashes for 5 Years (<https://vicroadsopendata-vicroadsmaps.opendata.arcgis.com/datasets/vicroadsmaps::road-crashes-for-five-years-victoria/about>)
  + Bureau Of Meteorology (BOM) weather data sets (<http://www.bom.gov.au/climate/data/?ref=ftr>)
  + Google API
* How will you access the data? (e.g. via an API or downloading the data)
  + Download CSV files from VicRoads and BOM
  + Visualise locations using Google API
* What format will the data be in?
  + CSV files
* What problems might arise given your choices above?
  + Data may not be clean, the datasets might not come with descriptions of attributes/columns
  + A limitation is we can’t pinpoint the exact weather at exact time (daily totals/max’s/averages provided by BOM)
  + Limitation that BOM API for current weather only, no historical data. Will need to use select locations (weather stations) for weather analysis using CSV files.
* Is your data sufficiently large?
  + Yes, VicRoads Road Crashes contains 77513 rows
  + Yes, BOM data is daily across the required years (2014-2019)
* Is your data messy or relatively clean?
  + Relatively clean – except for NaN to reconcile.

**Rough breakdown of tasks**

* Who will be doing what?
* What tasks can you divide between team members?
  + Clean data sets as a group (so all agree/understand data sets)
  + Split graphs/questions
  + Repo: Marcus + All
  + Initial Data analysis: Ivana + All
  + Code Review : Tamer + All
  + Presentation : Sanuli + All