**Data Analytics Research Project**

**Have Covid-19 cases in each Victorian suburb changed their popularity for homeowners in Victoria?**

Brainstorm - draft 1 - Tuesday 29th September

1. **The aim and questions**
2. Aim

What is it that we are trying to find out?

Using historical data for Victorian property sales, can we predict which suburb would gain the most popularity in the post covid environment

**The aim is to predict the Victorian suburb that would gain the most popularity (among home buyers?) in the post covid time**

Significance: Inform decisions to prepare for future plannings.

1. Questions (3-4 questions)

What are the questions we ask to find out the aim?

**I. Suburb preference**

1. What are the average house prices of houses in different suburbs?
2. What are the average houses in different suburbs (number of rooms, types of house, number of bathrooms, number of parkings, land area)
3. What are the average zones in different suburbs (schools, hospitals, shops, public transport)
4. What are the crime rates in different suburbs and how do they affect suburb popularity?
5. What are the geographical characteristics and how do that affect suburb popularity?
6. What are the areas with the most celebrities' houses and how do that affect suburb popularity?

II. Occupants

1. What are the demographics that buy houses in different areas?
2. What are the most common occupations of occupants living in different suburbs?
3. What are the lifestyles of occupants living in different suburbs?
4. Which suburbs that professionals concentrate in and how do covid affect suburb popularity?

Popularity = price & numbers of buys

1. **The data**
2. Type of data

What data do we need to answer the questions?

Victoria

* Data on the sales of homes (how many properties sold in each area, price for each property) [pre covid and now]
* Data on the houses (rooms, bathrooms, garage)
* Data on the zones (hospitals, schools, etc.)
* Data on the occupants (occupation, lifestyle, demographics, income)
* Data on the crime rates
* Data on the demographics
* Data on the celebrities
* Data of covid cases in suburbs

1. Database

Which database with accessible API can we get the data from?

1. API

Is the API:

* 1. Free?
  2. Easy enough to navigate?

1. **The analysis**
2. What type of analysis will we conduct to answer the research question?

Mean, mode, median, standard deviation, variance

Regression analysis

Something else?

1. What are the limitations of the analysis? What does the analysis tell and not tell us?