**Data Analytics Research Project**

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

1. **Thursday 1st October**
2. *Find all dataset* ☑
3. *Split the tasks between the team members* ☑
4. *Create the repo and hook it with our computers* ☑
5. **Clean datasets, Create a Jupyter Notebook, in the notebook**
   1. **Use Pandas to clean and format data set(s)**
   2. **Add postcodes and suburbs columns to all datasets**
   3. **If not completed, finish cleaning & formatting dataset at home**
   4. **By Saturday, bring in cleaned DataFrame in Jupyter Notebook**
6. **Research aim and questions**

***Aim and main question***

The aim of this project is to determine whether the number of Covid-19 cases in each Victorian suburb changed their popularity among homeowners.

The main question we are seeking to answer is:

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

***Working questions***

1. **What is the popularity of each suburb in Victoria before the Covid-19 pandemic happened (before February 2020)?** 
   1. Measured by changes to housing price in each suburb
   2. Measured by numbers of housing purchase in each suburb
   3. Measure the popularity difference between metropolitan and regional Victoria
2. **What are the factors that contribute to this popularity?**
   1. Does the crime rate in each suburb contribute to its popularity?
   2. Does the family-friendly zonings (schools, hospitals) in each suburb contribute to its popularity?
3. **What is the popularity of each suburb in Victoria after the Covid-19 pandemic happened (after February 2020)?**
4. **What are the factors that contribute to this popularity?**
   1. Does the number of Covid-19 cases in each suburb contribute to its popularity?

\*\*\* All dataset will be organised by suburbs

1. **Datasets**
2. **To measure the popularity of each suburb before and after Covid-19 - Matt**

<https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020>

1. **To measure the factors that contribute to the popularity before Covid-19**
   1. Crime rate - Kenny

<https://www.crimestatistics.vic.gov.au/crime-statistics/latest-crime-data-by-area>

* 1. Family-friendly zonings

School - Jeremy

<https://discover.data.vic.gov.au/dataset/hospital-locations-spatial>

Hospital - Anna

<https://discover.data.vic.gov.au/dataset/school-locations-time-series>

1. **To measure the factors that contribute to the popularity after Covid-19**
2. Covid-19 cases in each suburb

Victoria active cases by postcode

<https://docs.google.com/spreadsheets/d/e/2PACX-1vTwXSqlP56q78lZKxc092o6UuIyi7VqOIQj6RM4QmlVPgtJZfbgzv0a3X7wQQkhNu8MFolhVwMy4VnF/pub?gid=0&single=true&output=csv>

1. Lockdown-friendly zonings

Google Maps API

1. **Methodology and analysis**
2. **To measure the popularity of each suburb before Covid-19**

Summary statistics (Mean, median, mode)

1. **To measure the popularity of each suburb after Covid-19**

Summary statistics (Mean, median, mode)

1. **To measure the factors that contribute to the popularity before Covid-19**
2. **Crime rate**

Regression analysis between suburb popularity and crime rate

1. **Family-friendly zonings**

Regression analysis between suburb popularity and zone infrastructures

1. **To measure the factors that contribute to the popularity after Covid-19**
2. **Covid-19 cases in each suburb**

Regression analysis between suburb popularity and Covid cases

\*\* Keep monthly sales normalised

\*\* Find recent real estate data

1. **Project timeline**
2. **Thursday 1st October**
3. *Find all dataset* ☑
4. *Split the tasks between the team members* ☑
5. *Create the repo and hook it with our computers* ☑
6. **Clean datasets, Create a Jupyter Notebook, in the notebook**
   1. **Use Pandas to clean and format data set(s)**
   2. **Add postcodes and suburbs columns to all datasets**
   3. **If not completed, finish cleaning & formatting dataset at home**
   4. **By Saturday, bring in cleaned DataFrame in Jupyter Notebook**
7. **Saturday 3rd October**
8. In Jupyter Notebook, write # the steps to analyse the data
   1. Add graphs when possible
   2. Make sure the steps are coherent and actually answer the question
9. Start & finish analysis
   1. If not completed, finish analysis at home
   2. By Tuesday, bring in a completed analysis in Jupyter Notebook
10. **Tuesday 6th October**
11. Bring in polished & completed Jupyter Notebook
    1. Make sure PNG images are pretty and saved
12. Discuss findings with each other
13. Make tweaks
    1. To the research plan
    2. If needed, make tweaks to the analysis and redo analysis
14. **Thursday 8th October**
15. Compile pretty PNG images
16. Write a summary of our major findings. Include:
    1. A heading for each question
    2. A short description of findings and relevant plots
    3. Any limitations to our findings
17. Do a powerpoint to present in class
    1. Split the presentation for each member
18. Write a speech & practice presenting

Brainstorm - draft 2 - Thursday 1st October

1. **Research aim and questions**

***Aim and main question***

The aim of this project is to determine whether the number of Covid-19 cases in each Victorian suburb changed their popularity among homeowners.

The main question we are seeking to answer is:

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

***Working questions***

1. **What is the popularity of each suburb in Victoria before the Covid-19 pandemic happened (before February 2020)?**
   1. Measured by changes to housing price in each suburb
   2. Measured by numbers of housing purchase in each suburb
   3. Measure the popularity difference between metropolitan and regional Victoria
2. **What are the factors that contribute to this popularity?**
   1. Does the crime rate in each suburb contribute to its popularity?
   2. Does the family-friendly zonings (schools, hospitals) in each suburb contribute to its popularity?

BONUS: Find dataset for question below if we want more later

* 1. Does the proximity to the CBD in each suburb contribute to its popularity?
  2. Does the celebrity homeowners in each suburb contribute to its popularity?

1. **What is the popularity of each suburb in Victoria after the Covid-19 pandemic happened (after February 2020)?**
2. **What are the factors that contribute to this popularity?**
   1. Does the number of Covid-19 cases in each suburb contribute to its popularity?

BONUS: Find dataset for question below if we want more later

* 1. Does the lockdown-friendly zonings (parks, rivers, nature) in each suburb contribute to its popularity?
  2. Does the health-friendly zoning (hospitals, medical clinics, testing place) in each suburb contribute to its popularity?

1. **Datasets**
2. **To measure the popularity of each suburb before and after Covid-19**

<https://www.kaggle.com/htagholdings/aus-real-estate-sales-march-2019-to-april-2020>

1. **To measure the factors that contribute to the popularity before Covid-19**
   1. Crime rate

<https://www.crimestatistics.vic.gov.au/crime-statistics/latest-crime-data-by-area>

* 1. Family-friendly zonings

<https://discover.data.vic.gov.au/dataset/hospital-locations-spatial>

<https://discover.data.vic.gov.au/dataset/school-locations-time-series>

BONUS: Find dataset for question below if we want more later

* 1. Celebrity homeownership
  2. Proximity to the CBD

1. **To measure the factors that contribute to the popularity after Covid-19**
2. Covid-19 cases in each suburb

Victoria active cases by postcode

<https://docs.google.com/spreadsheets/d/e/2PACX-1vTwXSqlP56q78lZKxc092o6UuIyi7VqOIQj6RM4QmlVPgtJZfbgzv0a3X7wQQkhNu8MFolhVwMy4VnF/pub?gid=0&single=true&output=csv>

1. Lockdown-friendly zonings

Google Maps API

BONUS: Find dataset for question below if we want more later

1. Health-friendly zonings
2. **Methodology and analysis**
3. **To measure the popularity of each suburb before Covid-19**

Summary statistics (Mean, median, mode)

1. **To measure the popularity of each suburb after Covid-19**

Summary statistics (Mean, median, mode)

1. **To measure the factors that contribute to the popularity before Covid-19**
2. **Crime rate**

Regression analysis between suburb popularity and crime rate

1. **Family-friendly zonings**

Regression analysis between suburb popularity and zone infrastructures

1. **Celebrity homeownership**

Regression analysis between suburb popularity and celebrity homeownership

1. **To measure the factors that contribute to the popularity after Covid-19**
2. **Covid-19 cases in each suburb**

Regression analysis between suburb popularity and Covid cases

1. **Lockdown-friendly zonings**

Regression analysis between suburb popularity and zone infrastructures

\*\* Keep monthly sales normalised

\*\* Find recent real estate data

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?