**Springboard Capstone Project Proposal**

**Topic: Real Estate Sale Price Predictor for Melbourne (AUS)**

Detailed below is the Springboard Capstone Project Proposal for the project idea ‘Real Estate Sale Price Predictor for Melbourne (AUS)’

**Industry:** Real Estate, Mortgage Lending

**Clients:** Buyers, Investors, Property Owners

**Other Potential Stakeholders:** Bankers/Mortgage Lenders, Real Estate Agents

**Dataset name:** Melbourne Housing Market – housing clearance data from Jan 2016

**Link:** <https://www.kaggle.com/anthonypino/melbourne-housing-market>

**Data Type:** CSV

**Data Size** (FULL dataset): 34858 rows, 21 columns; (LESS dataset): 63024 rows, 13 columns

### **OVERVIEW**

The Springboard Capstone Project that I have picked is ‘Real Estate Sale Price Predictor for Melbourne (AUS)’. I am a resident of Melbourne, Australia working in a top National bank, that has a huge focus on Mortgage Lending. Also, being a home owner and investor, I am keen to follow and see the trends in the clearance rates of Melbourne Home Market over the last few years and subsequently predict the price of a given property in a particular suburb.

Domain.com.au is one of the leading online real estate websites widely used by real estate agents and home buyers alike. This dataset from Kaggle was scraped from publicly available results posted every week from Domain.com.au during the period 2016-2018.

Melbourne housing market has seen some tremendous and exponential growth over the years and experts predict the city is experiencing a housing bubble which is expected to burst soon. There are trends to indicate the real estate market in Melbourne has reached its peak, and is starting to flatten or even crash in the near future.

This project attempts to predict the sale price of a property given its address based on various factors like the location/suburb, distance from facilities like hospital, school, train station etc.

**THE CLIENT**

The client in scope for this project is Home Buyers, both first time and investors, who would be able to use this tool to predict the sale price of the property they are interested in, which would help them in their decision making of where to buy their next property. Alternatively, this could also be used by the home owners/sellers to determine what is the sale price their property is likely to fetch

**PROBLEM DESCRIPTION**

The Melbourne House Price Predictor tool attempts to assist the user (home buyer/seller) by predicting the sale price of a property in a particular Melbourne suburb that based on the following parameters:

* Type of House (House/Apartment/Townhouse/Unit)
* Number of rooms
* Land size
* Building area

In the process, we attempt to answer the following questions for the stakeholder:

1. Given a suburb/locality in Melbourne, what is the recent sale price of properties over the last two years
2. Compare the predicted sale price of similar characteristic real estate properties (e.g. same size/same number of rooms/same distance from essential facilities like school, hospital etc.) across various suburbs in Melbourne.
3. Identify the most expensive and least expensive suburbs in Melbourne
4. Identify any evidence/trend indicating the relative change in the property prices in relation to the distance from the CBD (city center), i.e. suburbs closer to the CBD are more expensive than the suburbs farther away from the city.
5. How is the property price influenced by its proximity to:
6. Schools
7. Hospital/Medical Center
8. Train Station/Bus Stand

Does the predictor model behave differently when these above parameters (distance to the school/hospital/public transport) are added into the model?

**DATASET**

**Link:** <https://www.kaggle.com/anthonypino/melbourne-housing-market>

**Data Type:** CSV

**Data Size** (FULL dataset): 34858 rows, 21 columns; (LESS dataset): 63024 rows, 13 columns

The columns in the dataset are explained below:

**Suburb:** Suburb

**Address:** Address

**Rooms:** Number of rooms

**Price:** Price in Australian dollars

**Method:** S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.

**Type:** br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.

**SellerG:** Real Estate Agent

**Date:** Date sold

**Distance:** Distance from CBD (Central Business District; city center) in Kilometres

**Regionname:** General Region (West, North West, North, North east ...etc)

**Propertycount:** Number of properties that exist in the suburb.

**Bedroom2:** Scraped # of Bedrooms (from different source)

**Bathroom:** Number of Bathrooms

**Car:** Number of carspots

**Landsize:** Land Size in Metres

**BuildingArea:** Building Size in Metres

**YearBuilt:** Year the house was built

**CouncilArea:** Governing council for the area

**Lattitude:** Self explanatory

**Longtitude:** Self explanatory

**APPROACH**

1. Additional Data Capture – Apart from the existing data set which provides the sale prices of the real estate properties sold in Melbourne AU from 2016-2018, this project requires additional data sourced from internet, e.g. location of the following essential facilities in each suburb – school, hospital, train station/bus stand
2. Data Cleanup/Data Wrangling – This would include cleanup of the dataset to include all the additional data captured in step 1. Also, as part of data clean-up, the distance column of the data set needs to be revalidated by calculating the distance of the given address from the city center.
3. Data Presentation – Data Presentation would involve presenting the existing data set in a visual format.
4. Data Analysis and Prediction – This would include Model building and Machine Learning to build a model to predict the sale price of the property.
5. Data Visualization – This would involve presenting the results of the predictor tool in a visual format.

**DELIVERABLES**

The deliverables for this project will include:

* The project report writeup
* A presentation and associated slide-deck
* Jupyter Notebooks containing all code used for data exploration, transformation, model-building, and evaluation