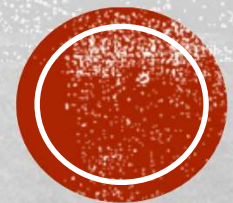


MINI PROJECT 1: MELBOURNE HOUSING PRICES

Lillian Leong

Jun 2022



AGENDA



Data Processing

Missing Values, Imputing,
Replace Values, Outliers
handling



Visualisation

Seaborn libraries
Correlation Heatmap,
Boxplots, BarCharts,
Countplots



Insights

Highest Correlated
Feature to Price | Region
Analysis | Room Feature
Analysis



BUSINESS PROBLEM

We, a property investment firm, would like to know what are the residential property to invest in Melbourne, Australia so that we can better advice our potential customers.

We attempt to answer the following questions:

- What is the price distribution in different parts of Melbourne?
- What are the property feature will affect the price?
Example – number of rooms, bathrooms, housing type?



METADATA OF DATASET

- Melbourne housing clearance data from Jan 2016
<https://www.kaggle.com/datasets/anthonypino/melbourne-housing-market>
- `melbourne_data_df.shape` → (13580, 21)
- `melbourne_data_df.columns` →
Index(['Suburb', 'Address', 'Rooms', 'Type', 'Price', 'Method', 'SellerG', 'Date', 'Distance', 'Postcode', 'Bedroom2', 'Bathroom', 'Car', 'Landsize', 'BuildingArea', 'YearBuilt', 'CouncilArea', 'Latitude', 'Longitude', 'Regionname', 'Propertycount'],



DATA CLEANING

- Missing Values:
 - Mean Value imputation fill in the missing values in the Car, BuildingArea, and YearBuilt columns
 - Mode value imputation to fill in the missing values in the SellerG and CouncilArea column
- Categorical data transformation for “Type”
 - `melbourne_data_df.replace({'h': 'House', 'u': 'Unit', 't': 'Townhouse'}`
- Outliers Handling
 - Minimum Year=1196, CouncilArea=“Unavailable”



EDA

- Correlation Heatmap



- By using the heatmap to corroborate with the table of correlation between price and the various features, it is evident the features with the highest correlation with sale price are (defined as values with a correlation of around 0.5 in this case): Rooms, Bathroom



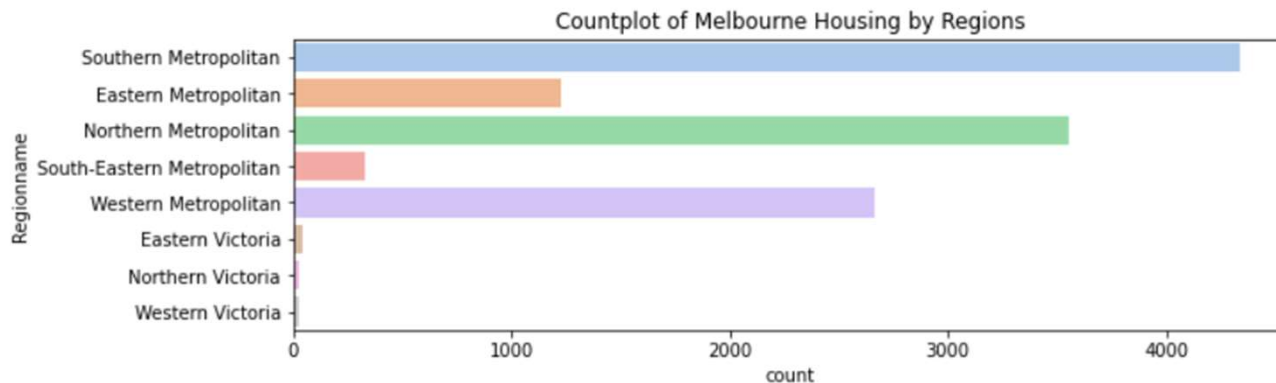
DATA STATISTICAL SUMMARY

	min	mean	max	median	count
Regionname					
Southern Metropolitan	131000	1338721	8000000	1206000	4339
Eastern Metropolitan	288000	1087686	4000000	986750	1228
Northern Metropolitan	145000	896672	4525000	805000	3554
South-Eastern Metropolitan	266000	901530	9000000	805000	325
Western Metropolitan	85000	864937	3900000	795500	2668
Eastern Victoria	400000	688204	1085000	653000	44
Northern Victoria	330000	559750	898000	533000	26
Western Victoria	280500	382910	550000	385000	25

- Overview of Price column using `df.describe()` and `.groupby()`
 - The Price column is a continuous variable with values ranging from 85,000 to 9,000,000.
 - The average sale price of all the houses in our dataset is 1,063,659



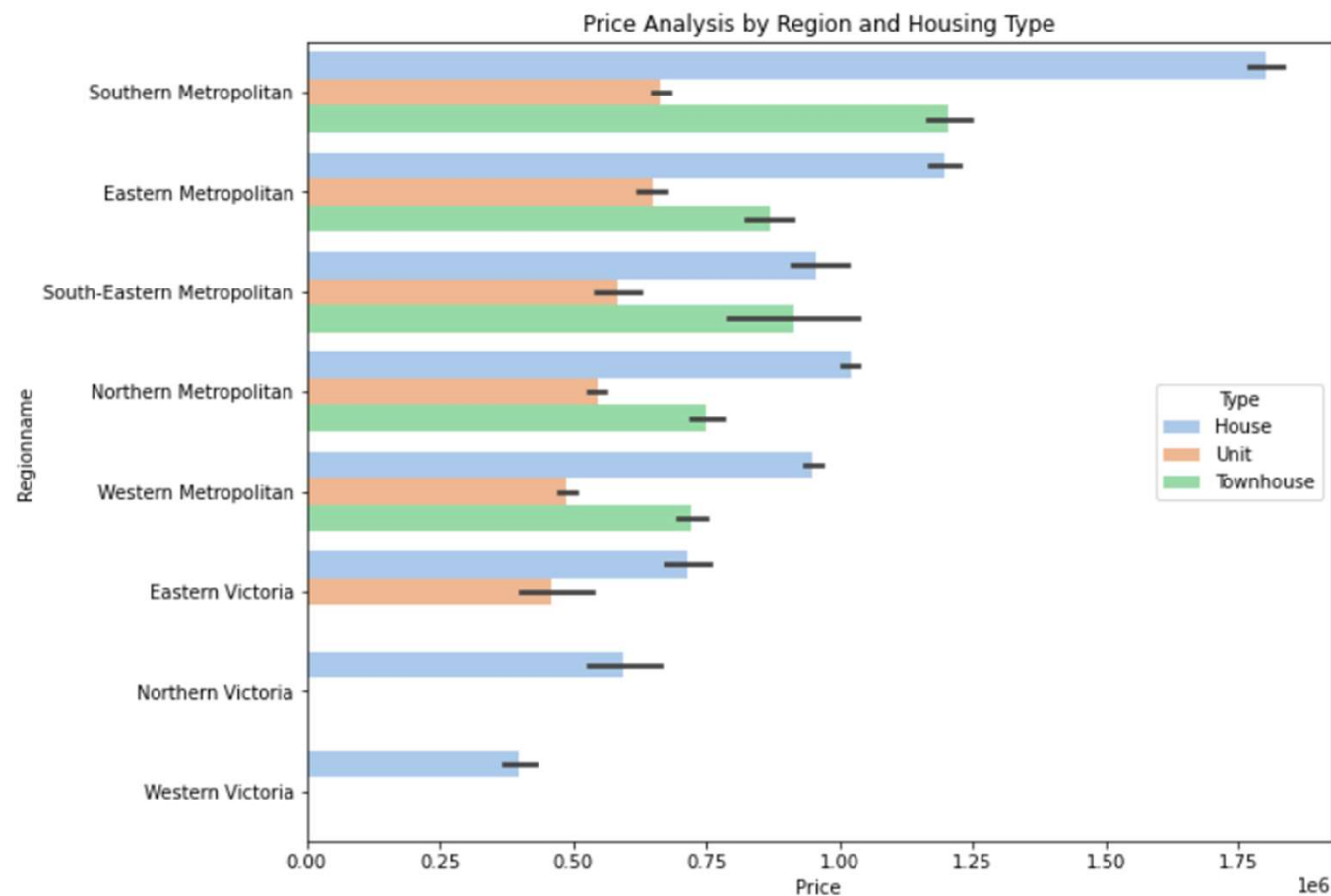
PRICE ANALYSIS BY REGIONAL FEATURES



- The boxplot illustrated that Southern Metropolitan has the highest price IQR and even the median is also highest across the regions.
- There is an outlier in South Eastern Metropolitan which requires further investigation.
- Countplot also showed that South Metropolitan has the highest counts in the dataset which suggest highest activity in the real estate industry in Melbourne.



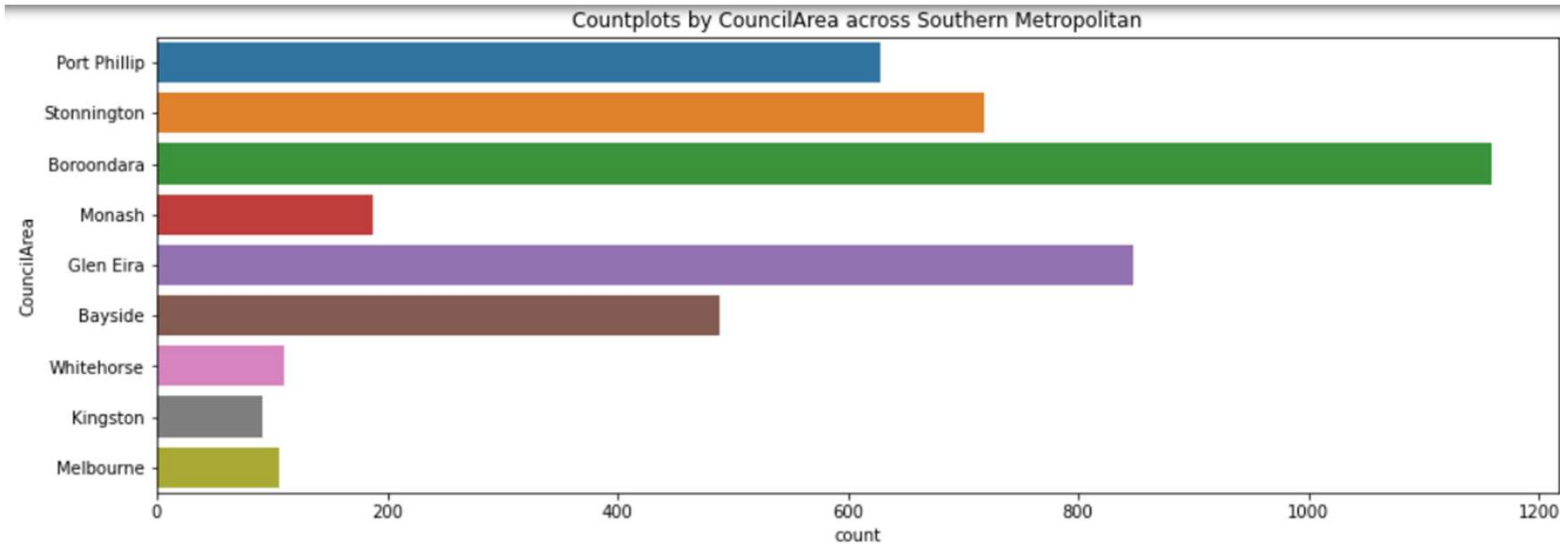
PRICE ANALYSIS BY REGION AND TYPE



- Prices for all House type is the highest in South Metropolitan, as high as 1.75M for House type
- Prices for all House type is the lowest in Western Metropolitan, approx. 1M for House type
- Data seems to suggest that Eastern Victoria might be better developed (in construction wise) within Victoria region. There is only House type in Northern & Western Victoria.



PRICE ANALYSIS BY COUNCILAREA

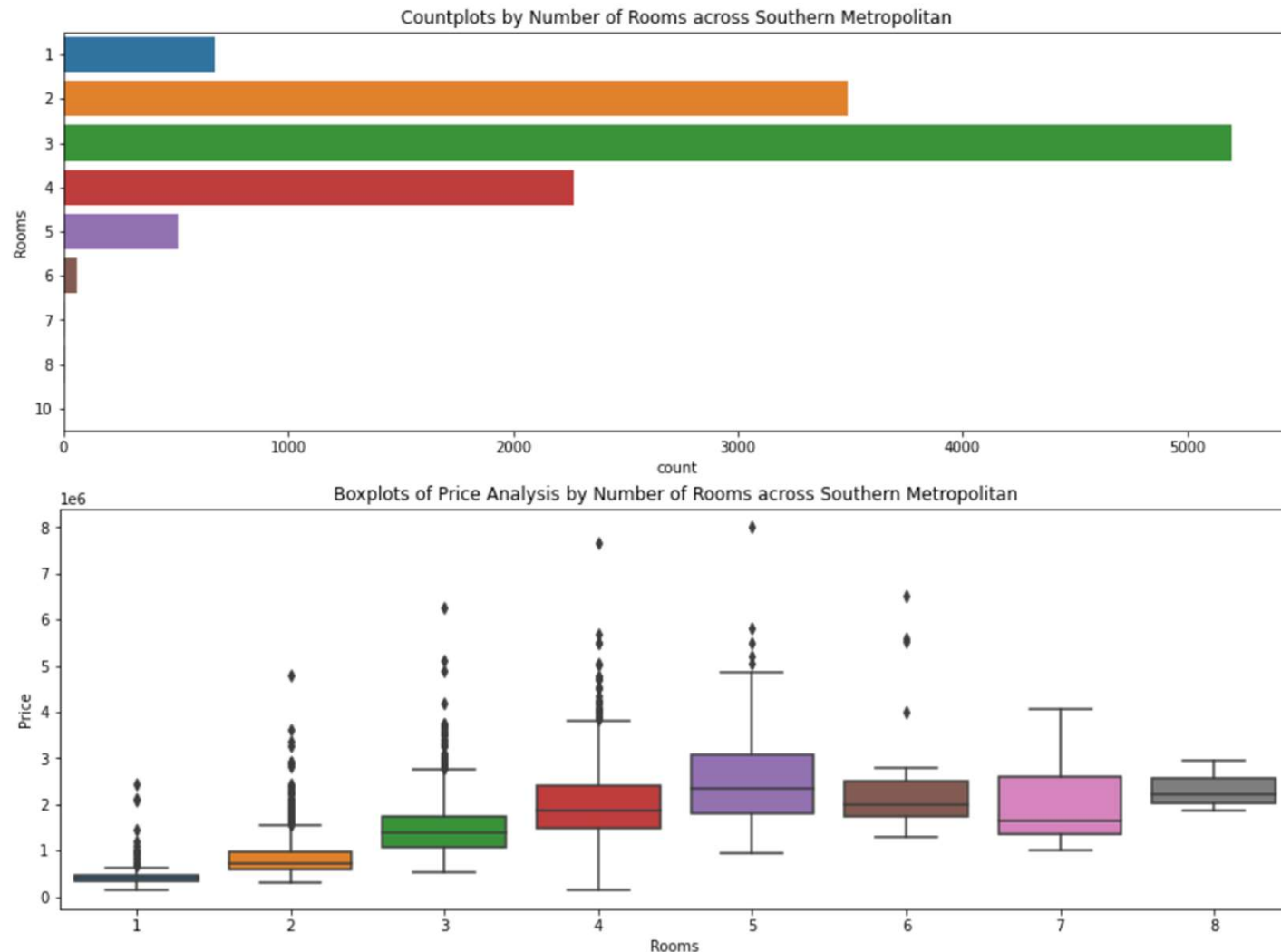


	count	mean	min	median	max
CouncilArea					
Boroondara	1159	1646740	160000	1615000	8000000
Bayside	489	1652168	290000	1550000	5800000
Whitehorse	111	1297237	550000	1253500	2630000
Monash	187	1091499	305000	1060000	2525000
Glen Eira	848	1069279	131000	1040500	3000000
Stonnington	719	1293382	240000	1020000	6250000
Kingston	92	990965	240000	993500	1740000
Port Phillip	628	1144346	275000	920000	5600000
Melbourne	106	920908	260800	720000	3000000

- The highest frequency is from Boroondara which has a median price around 1.6M
- The lowest frequency is from Kingston which has a median price less than 1M



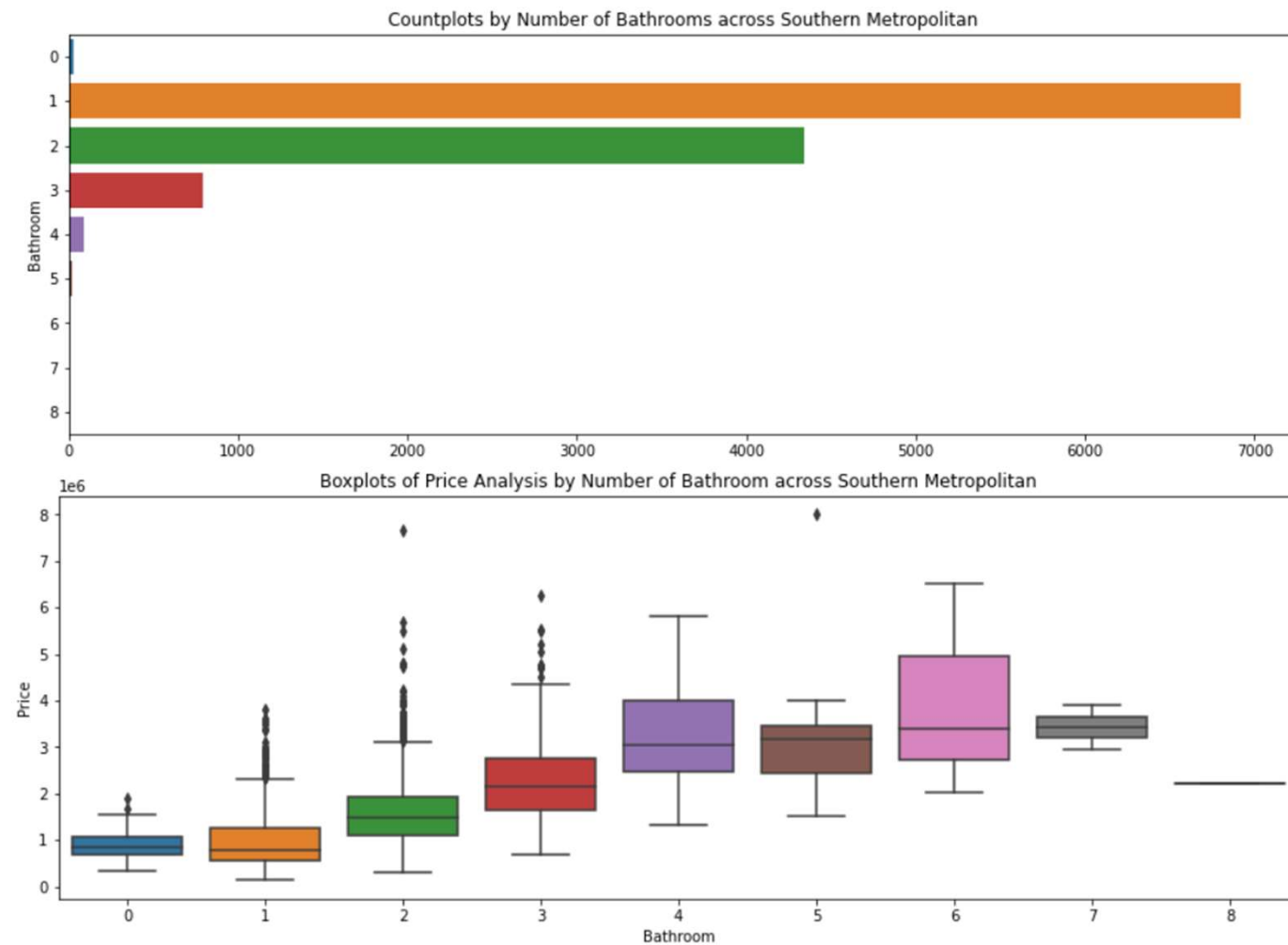
PRICE ANALYSIS BY ROOM FEATURES



- The highest frequency is from 3 room which has a median price of less than 1.5M.
- The lowest frequency is from 5 room which has a median price around 2.35M
- Properties with 7 rooms has a high proportion in the upper quartile beyond the median price.



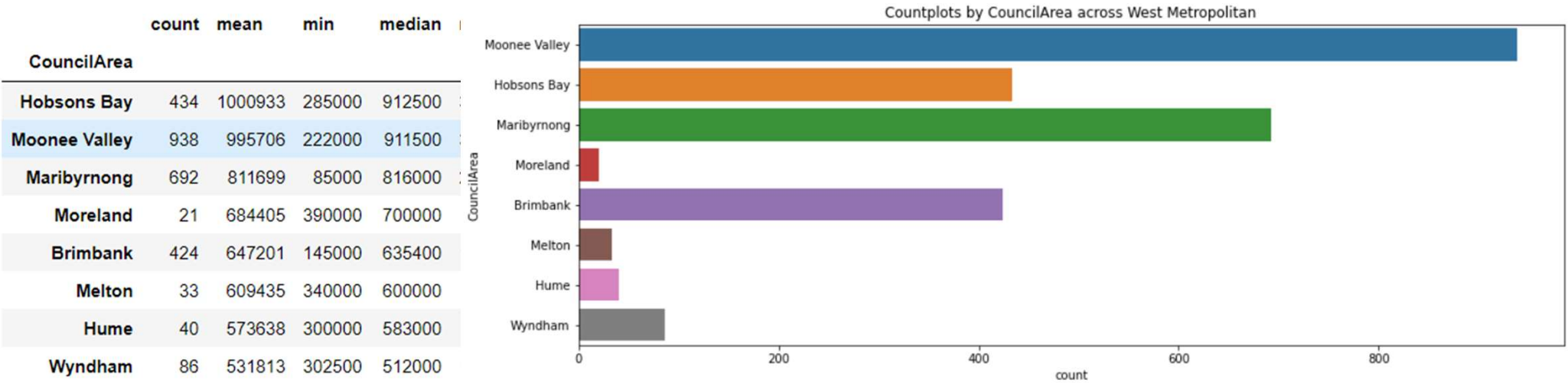
PRICE ANALYSIS BY BATHROOM FEATURES



- The highest frequency is from 1 bath which has a median price of less than 1.0M.
- The lowest frequency is from 5 room which has a median price around 3.2M
- Properties with 7 rooms has a high proportion in the upper quartile beyond the median price.



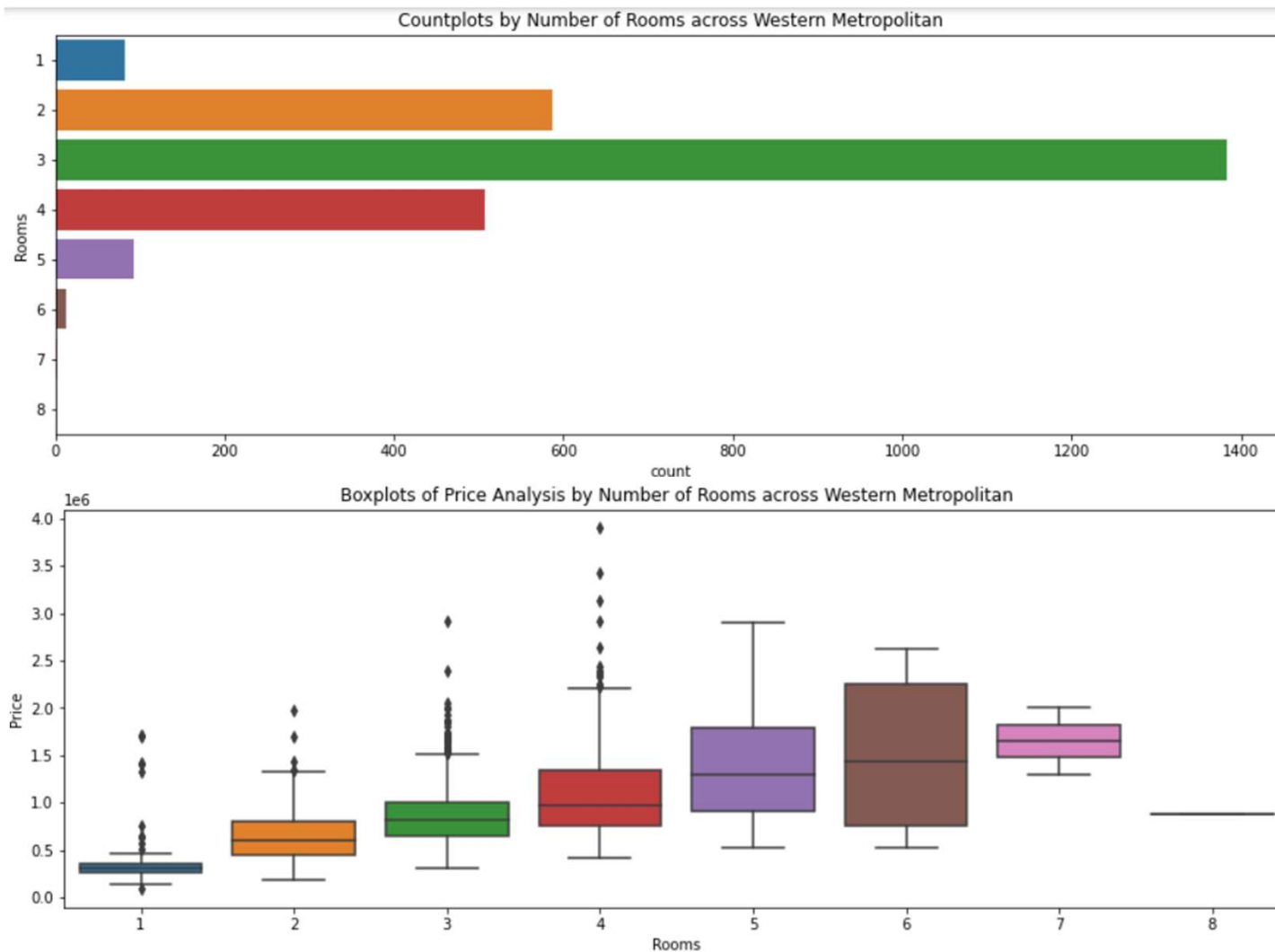
PRICE ANALYSIS BY COUNCILAREA



- The highest frequency is from Moonee Valley which has a median price around 900K
- The lowest frequency is from Moreland which has a median price around 700K



PRICE ANALYSIS BY ROOM FEATURES

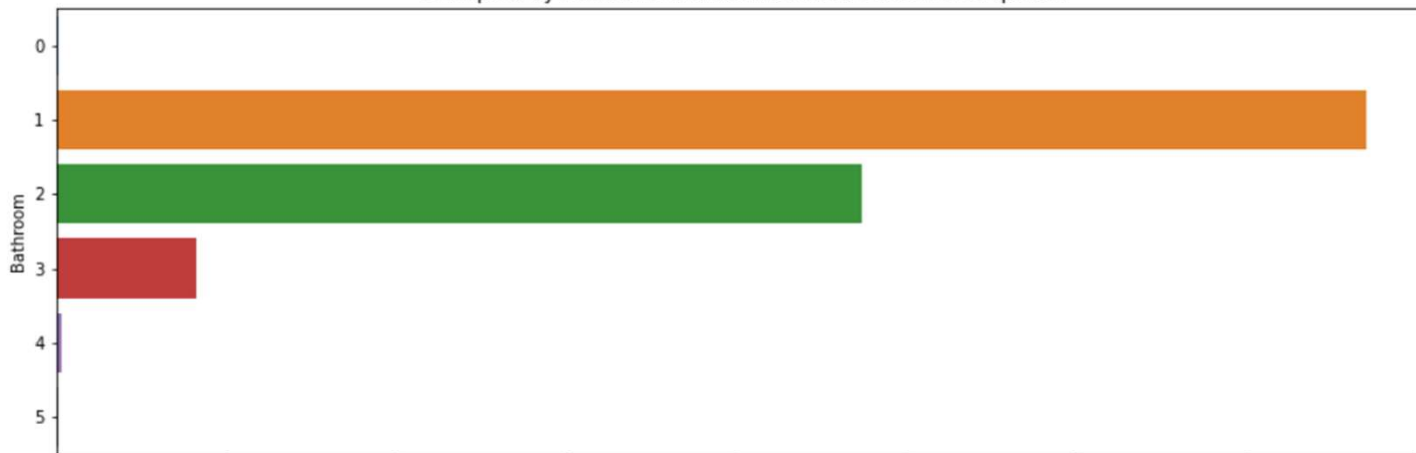


- The highest frequency is from 3 room which has a median price of less than 815K
- The lowest frequency is from 7 room which has a median price around 1.65M
- Properties with 6 rooms has a high IQR from its median price of 1.43M

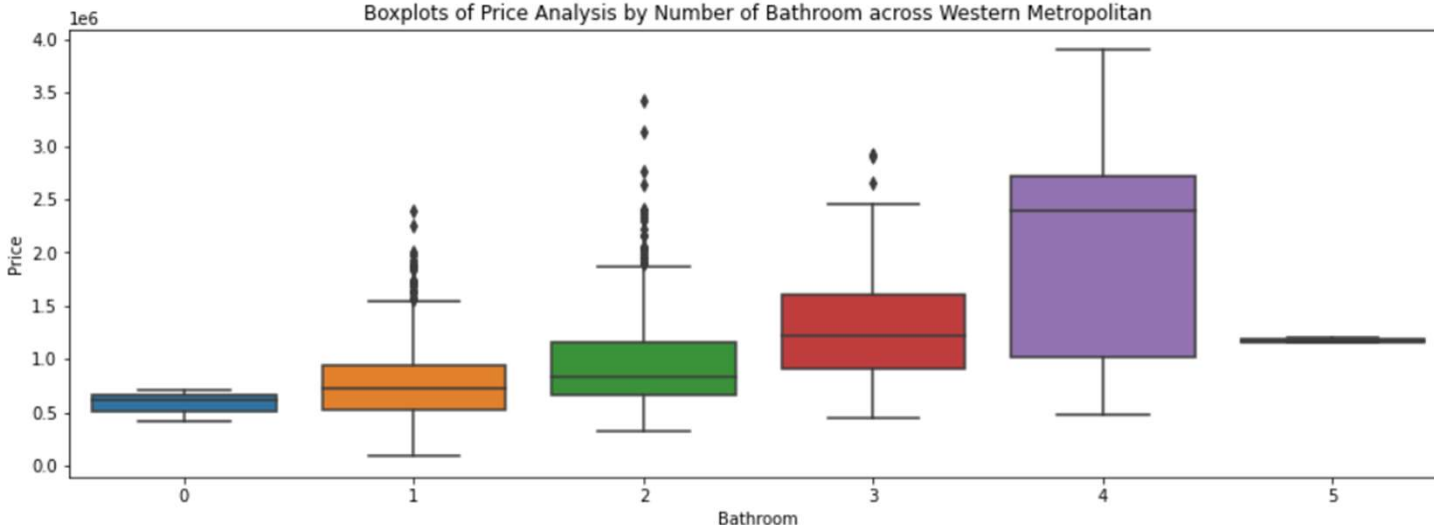


PRICE ANALYSIS BY BATHROOM FEATURES

Countplots by Number of Bathrooms across Western Metropolitan



Boxplots of Price Analysis by Number of Bathroom across Western Metropolitan



- The highest frequency is from 1 bath which has a median price of 725K.
- The lowest frequency is from 5 bath which has a median price around 1.2M
- Properties with 4 bath has a high IQR from its median price of 2.4M



KEY TAKEAWAYS

	South Metropolitan	West Metropolitan
Overall investment budget	Mean: \$1.3M to \$1.4M	Mean: \$1M to 1.1M
Popular Council Area	Stonnington and Port Phillip will be around \$1.1M to \$1.3M. Boroondara & Bayside requires a higher budget of \$1.6M to \$1.65M	Moonee Valley is within the mean and Maribyrnong is around 800K
Features: No of Rooms	With median around 1.5M, the property could probably be around 3 rooms With median around 2.4M, the property could probably be around 5 rooms	With median around 1.1M, the property could probably be around 4 rooms. With median around 1.3M, the property could probably be around 5 rooms
Features: No of Baths	With median around 1.5M, the property could only have 2 bathrooms. Property with 3 room + 3 bathrooms has a median price of 1.73M Property with 5 room + 3 bedrooms has a median price of 2.4M	With median around 1.1M, the property could only have 3 bathrooms. Property with 4 room + 3 bathrooms has a median price of 1.2M Property with 5 room + 3 bedrooms has a median price of 1.5M



AREAS TO LOOK AT

- Multi years data collection to perform time series analysis
- Perform similar analysis for Victoria Area as well

