

MELBOURNE

HOUSING PRICES

Predicting Housing Prices in Melbourne, Australia

Presented by Usman Shaikh

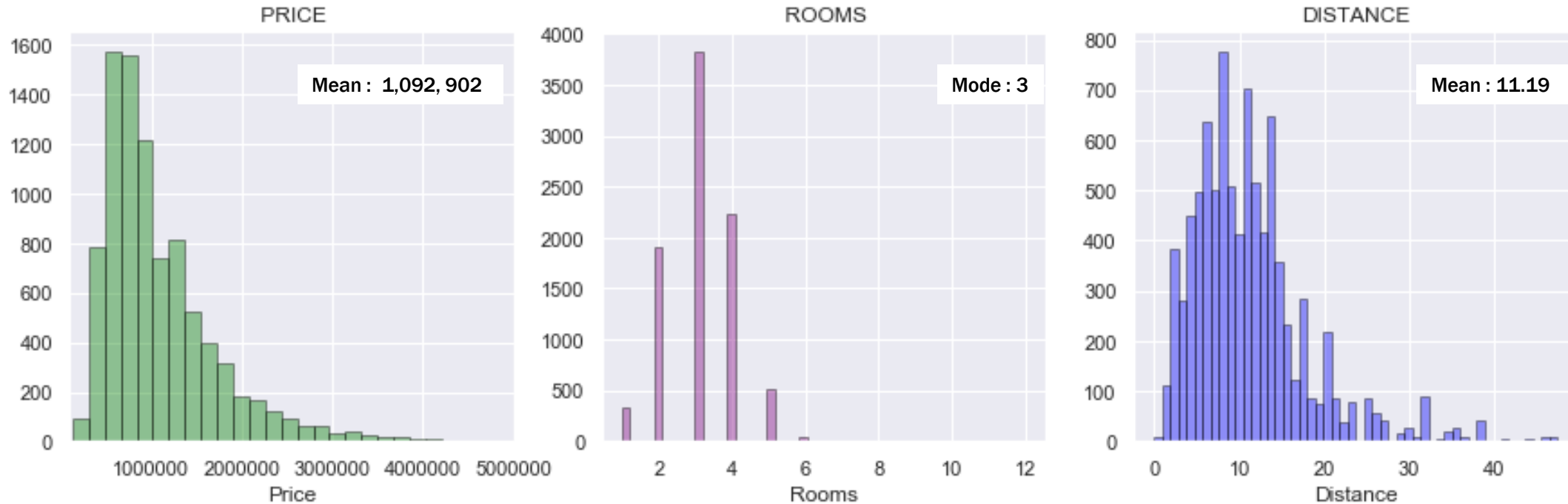


Dataset: Melbourne housing clearance data from Jan 2016

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

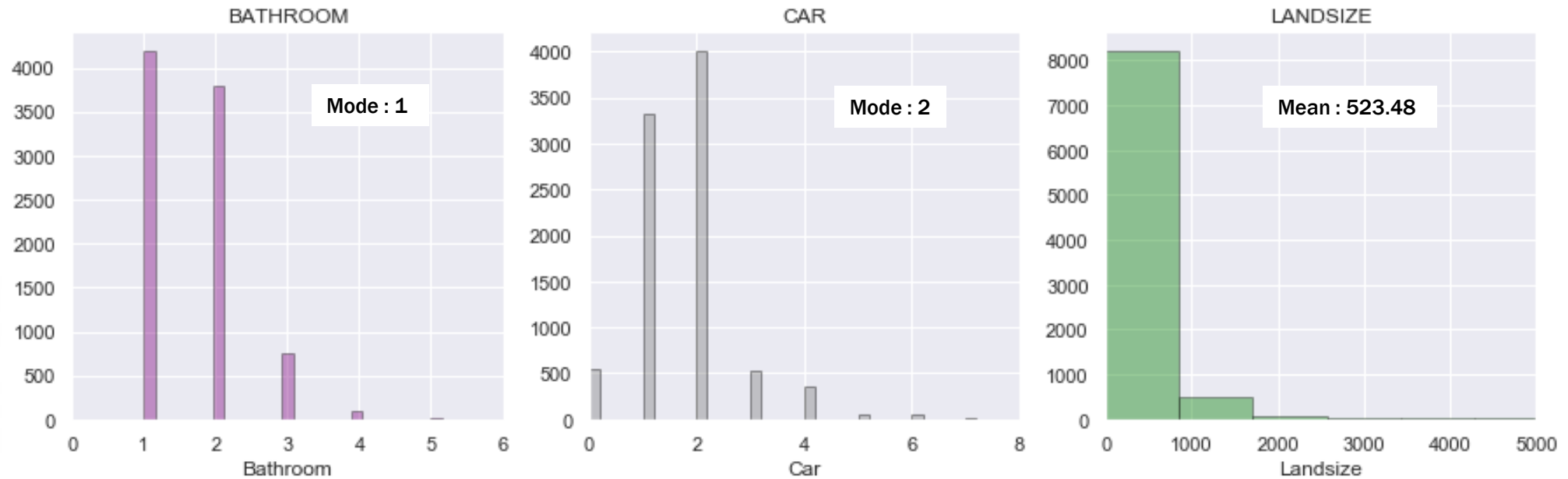


Exploratory Analysis: Understanding the Dataset – I



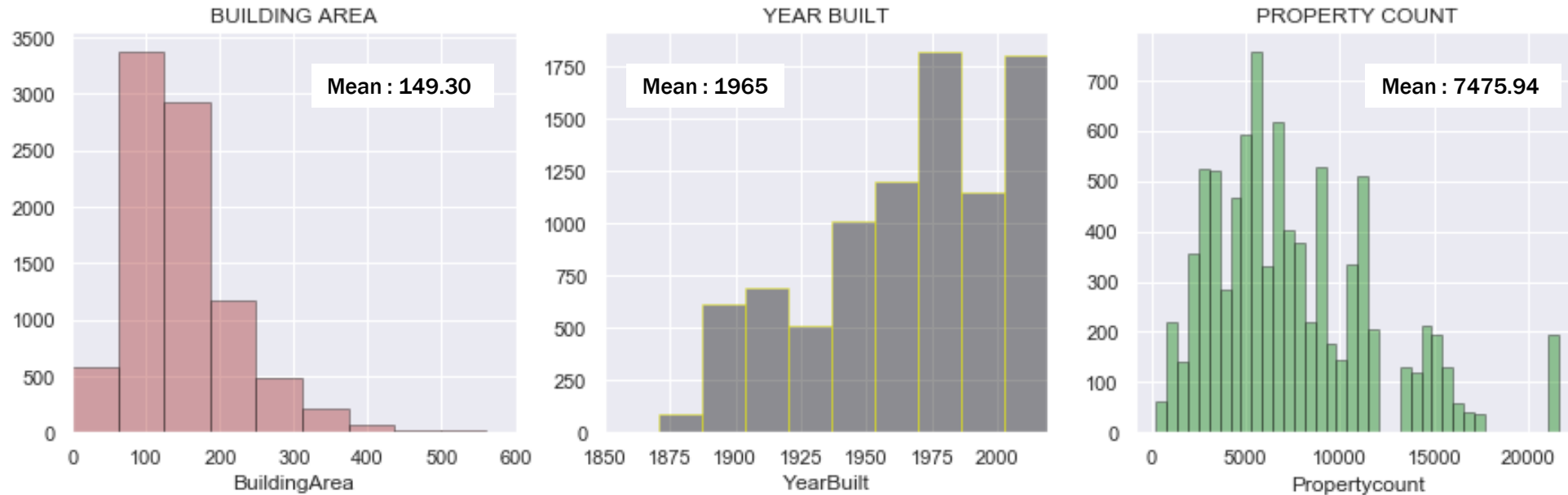
- ❑ Bulk of the housing prices in Melbourne Australia fall between 0 to 1.5 Million Dollars with a mean value of \$1,092,902.
- ❑ Most places have 3 rooms and are located at a mean distance of 11.19 Kilometers from the Central Business District.
- ❑ Most properties are located within a range of 18 Kilometers from Central Business District (CBD).

Exploratory Analysis: Understanding the Dataset – II



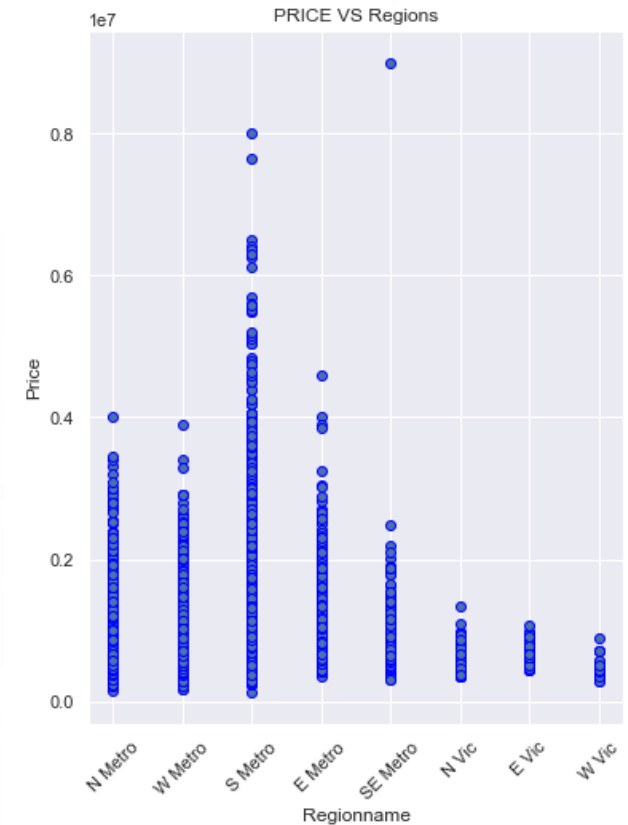
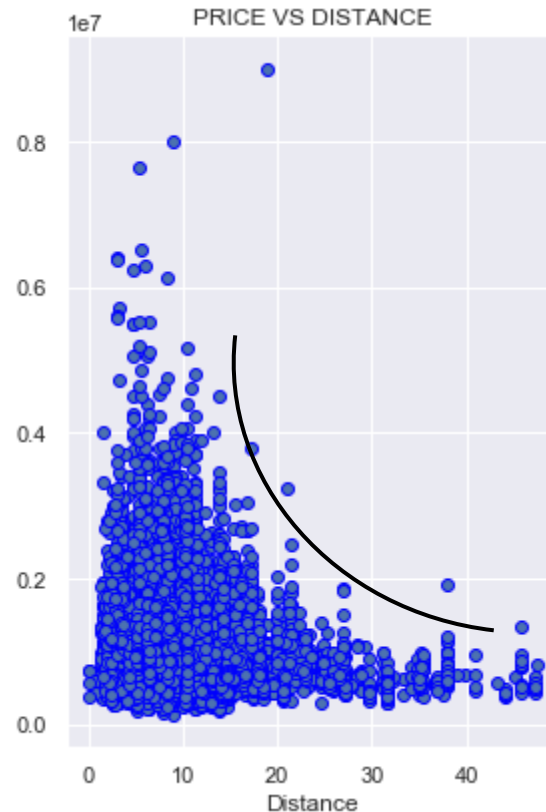
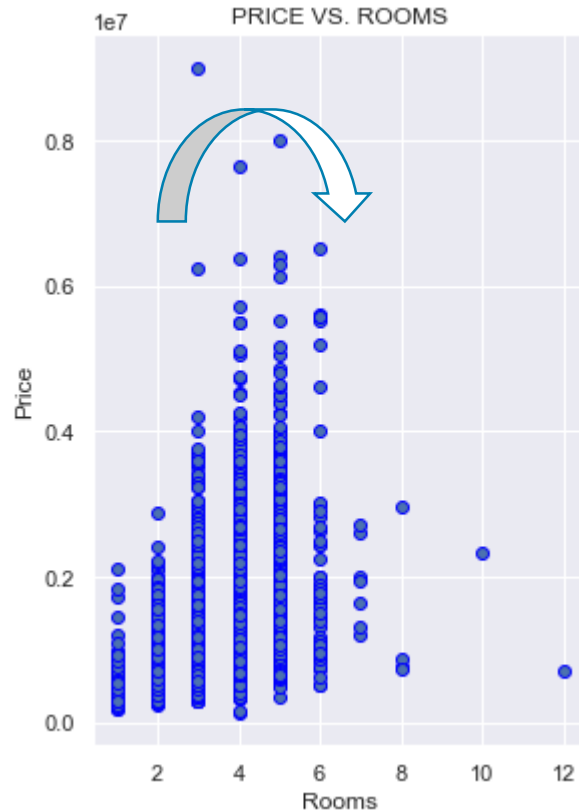
- ❑ The common number of Bathrooms per property appears to be 1 followed by 2 car spaces.
- ❑ The Land size of most properties fall between 0 to 2,500 Square Meters, with over 90% of the values falling under 1,000 Square Meters.

Exploratory Analysis: Understanding the Dataset – III



- ❑ The building area for most of the properties fall between 50 and 200 Square Meters with a mean value of 149.30
- ❑ The housing construction shows a gradual rising trend over the years with a mean value year of 1965.
- ❑ The housing market witnessed increased amount of construction in mid seventy's , late 90's followed by 2000's.

Relationship between Price and other Factors – I



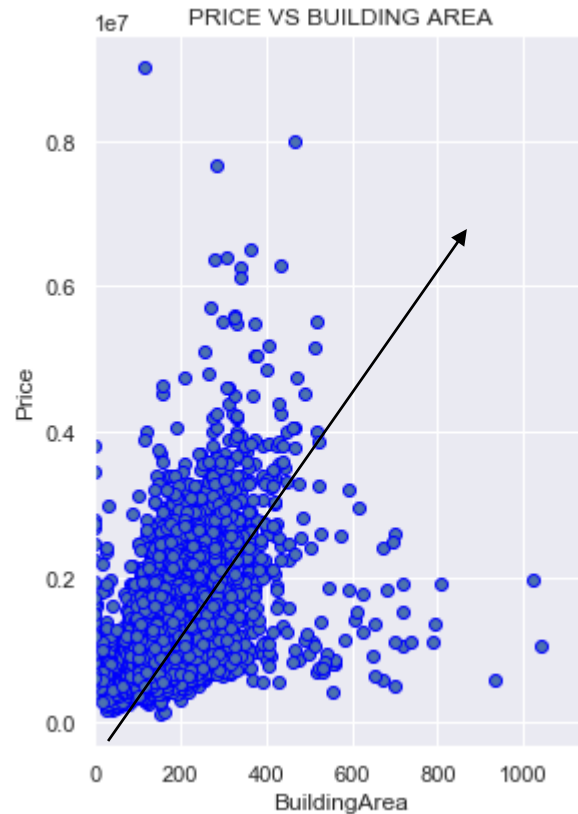
- ❑ The housing prices appear to increase with the number of rooms but starts to fall after 5 rooms
- ❑ Likewise the price appears to rise with fall in distance with highest values for properties located in a 5 to 6 Km radius from CBD.
- ❑ Region wise South Metro is closest to the Central Business District and appears to be most expensive followed by East Metro.

Relationship between Price and other Factors – II



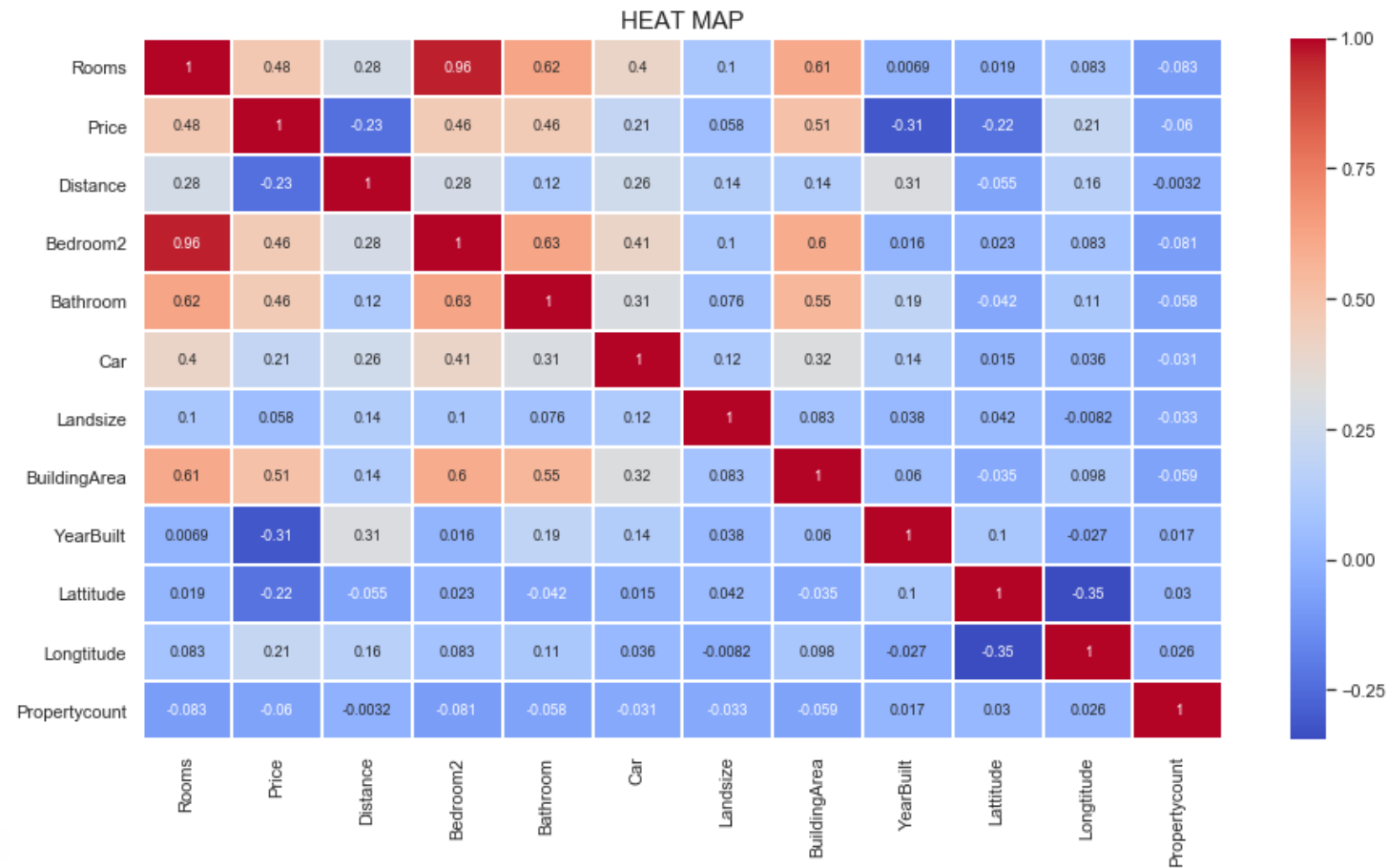
- ❑ The housing prices appear to increase with the number of Bathrooms however price starts to fall after a number of 3 Bathrooms.
- ❑ Majority of the land size falls under 1500 Square Meters and the price is highest around 1000 Square Meters.
- ❑ The price is highest for 4 car spaces however it starts to fall once it reaches four car spaces

Relationship between Price and other Factors - III



- ❑ There appears to be a correlation between Price and Building Area, 0.51 to be precise on a scale of -1 to 1
- ❑ The priciest properties seem to be built in early 1900's, late 90's and 2000's , as population in those years appear more dense.
- ❑ Price analysis show that homes with 3 bedrooms, 2 bathrooms, 2 car spaces, land size of 1000 Sqm, building area of 300 Sqm built in early 1900's, late 90's and 2000's appear to be of higher price.
- ❑ Similarly as distance increases from the Central Business District land size and building area increase with decrease in price.

Correlation HEAT MAP – Understanding influence of factors on each other and Price



- ❑ Rooms and Bedroom2 appear highly correlated **0.96** followed by Bathroom and Bedroom2 **0.63**
- ❑ Rooms and Bathroom: **0.62**, Building Area and Rooms:**0.61**, followed by Building Area and Bathroom: **0.55** and **0.51** for Building Area and Price.
- ❑ *Building Area and Bedroom2 not included in Machine Learning Model*

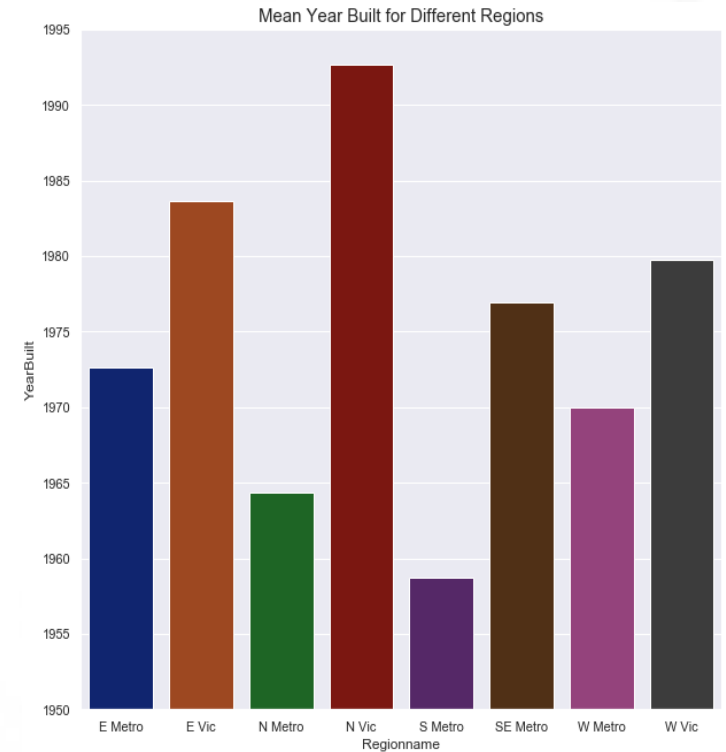
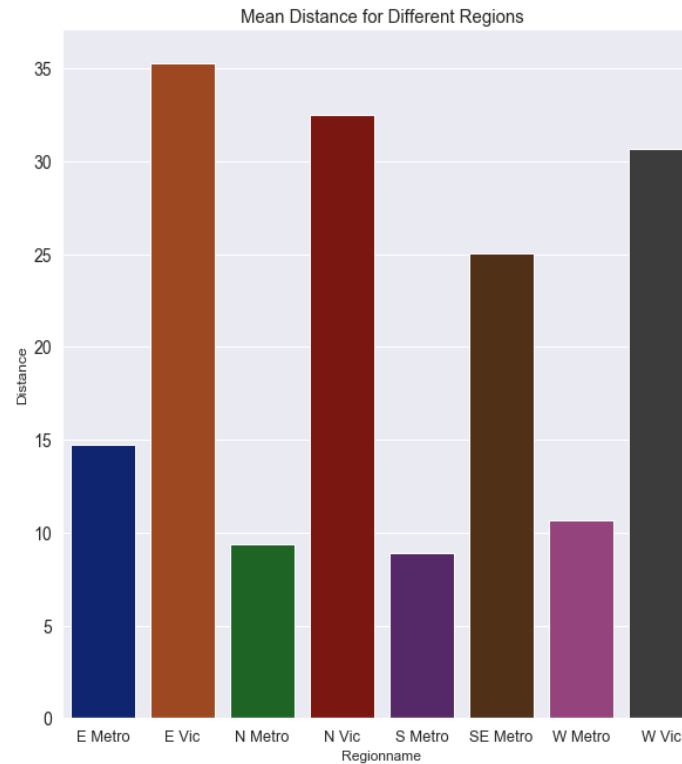
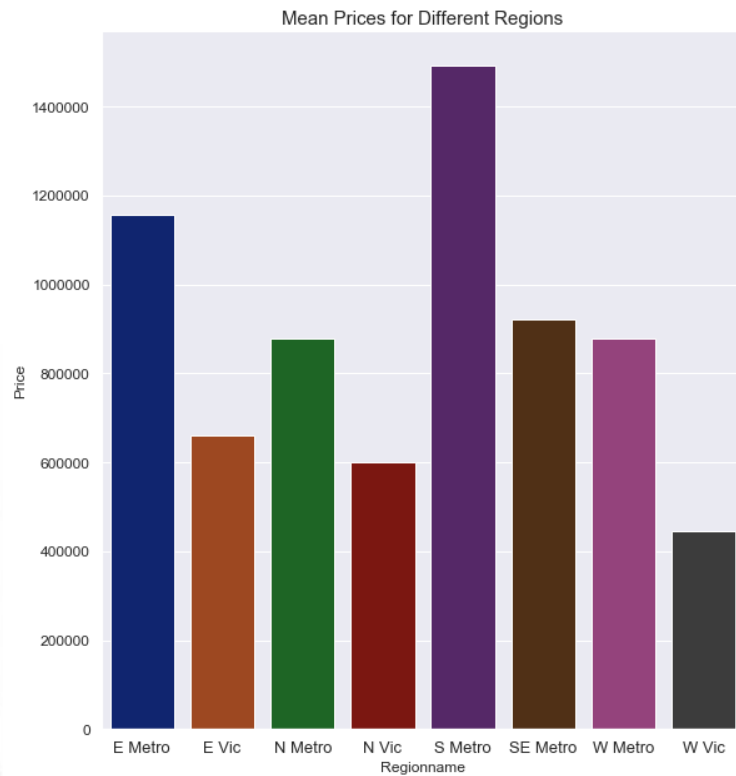
Analysis by Regions

To simplify analysis and visualizations city regions have been assigned abbreviated names as follows.

1. Northern Metropolitan: **N Metro**
2. Western Metropolitan: **W Metro**
3. Southern Metropolitan: **S Metro**
4. Eastern Metropolitan: **E Metro**
5. South-Eastern Metropolitan: **SE Metro**
6. Northern Victoria: **N Vic**
7. Eastern Victoria: **E Vic**
8. Western Victoria: **W Vic**

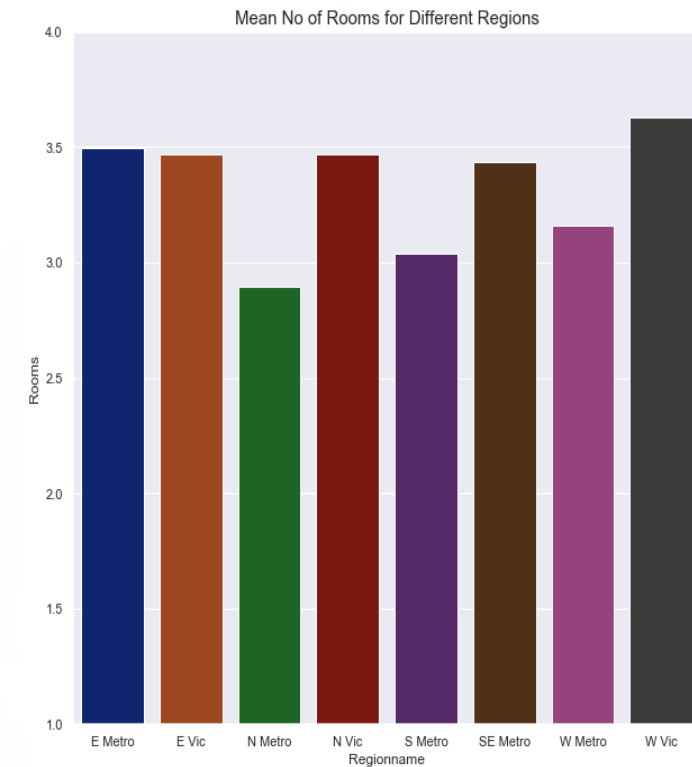
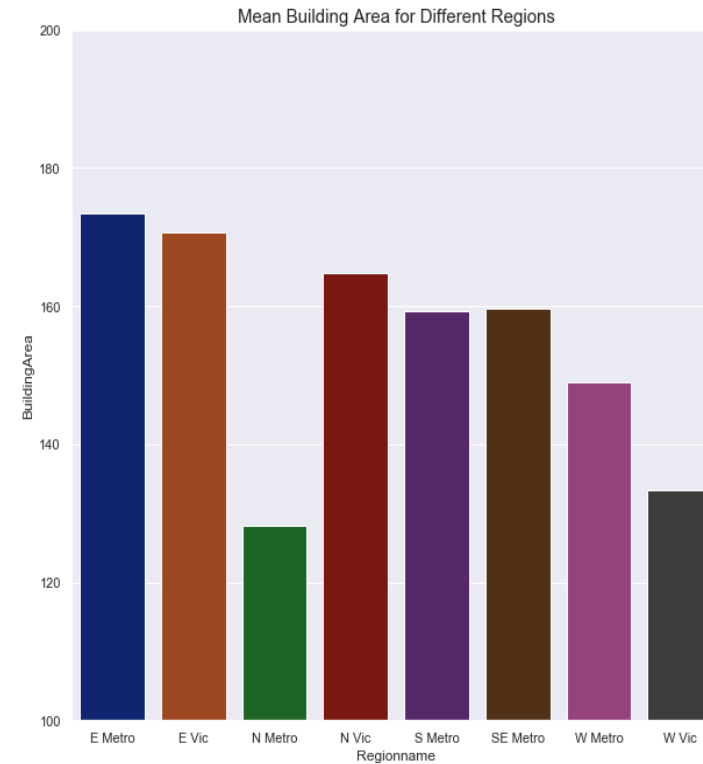
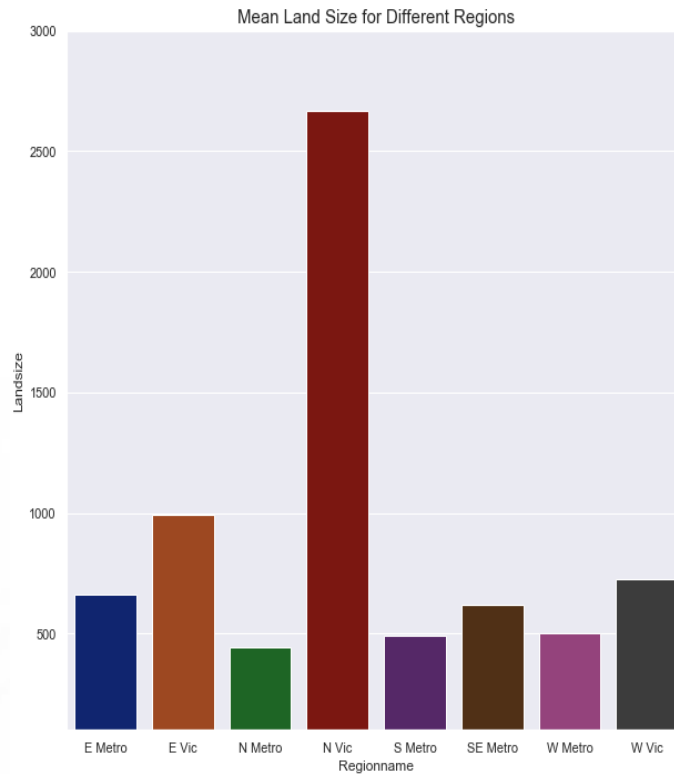


Mean Values – Analysis by Regions – I



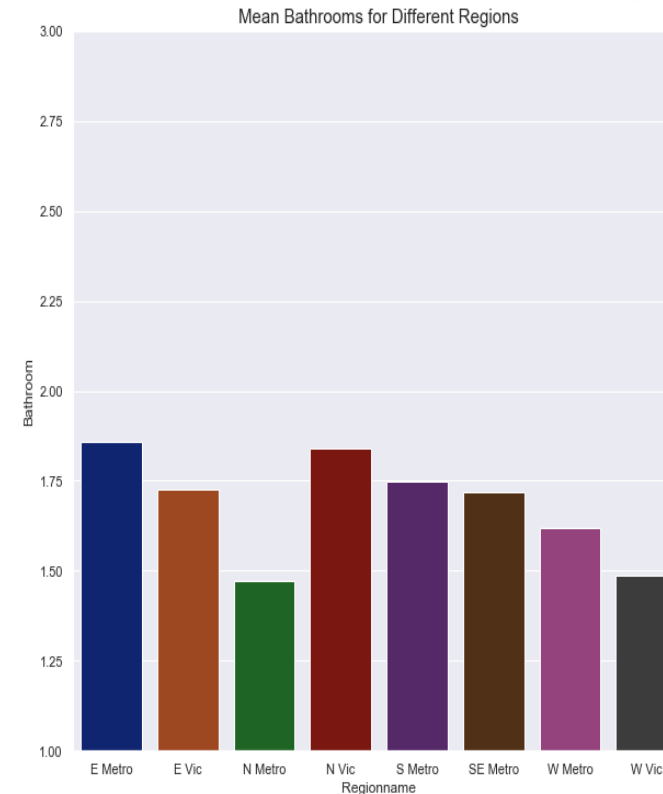
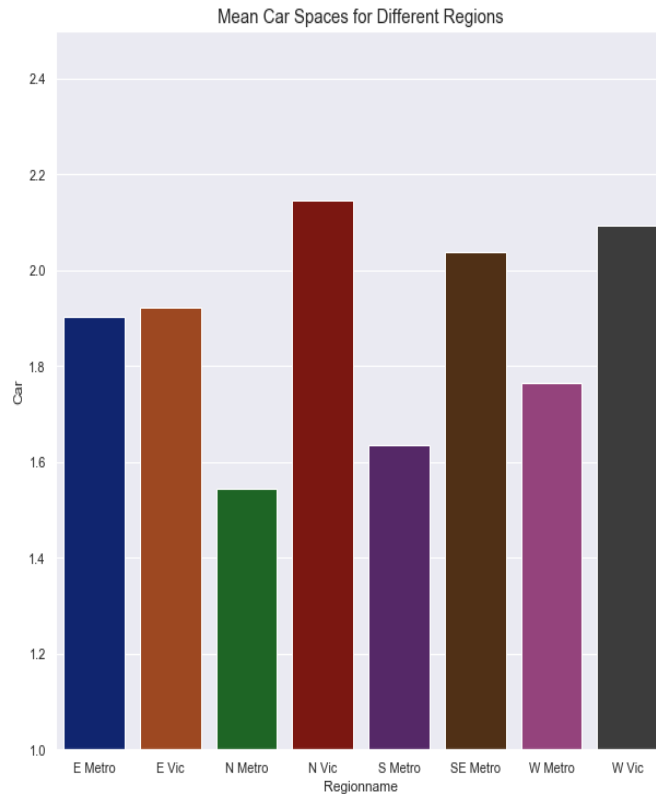
- ❑ The regions closest to CBD including N Metro, S Metro, E Metro appear to have a higher price with 50's, 70's and 90's construction.
- ❑ Where as regions with higher distance appear to have newer construction and lower price for example North Victoria.

Mean Values – Analysis by Regions – II



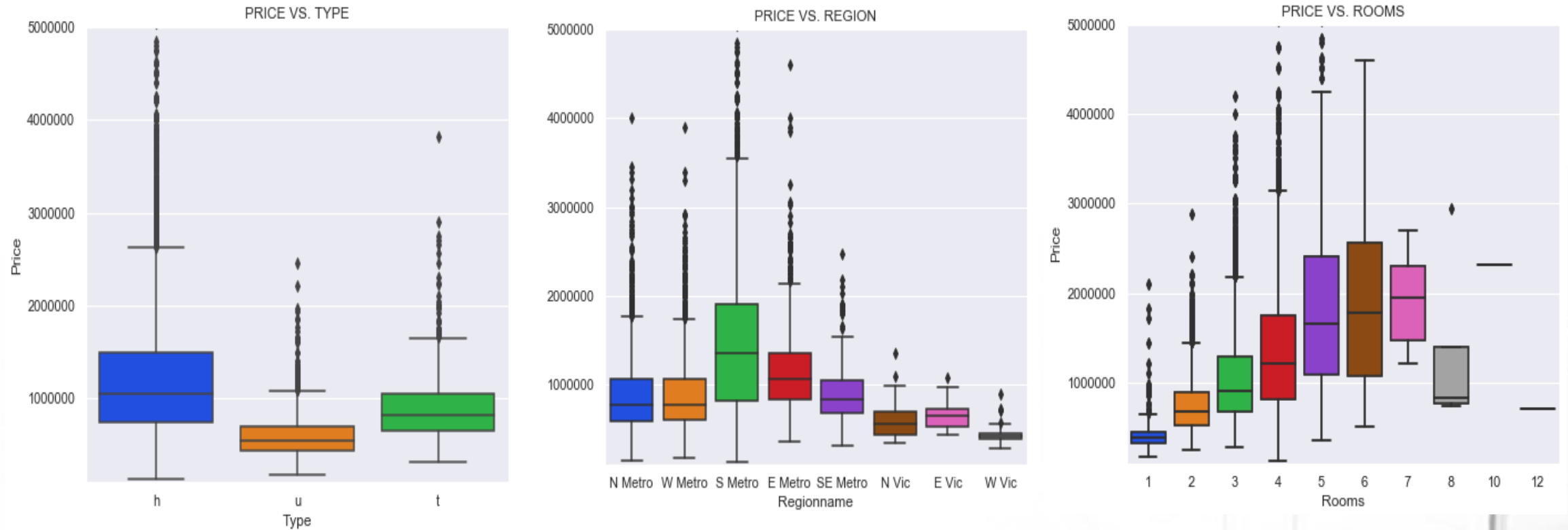
- ❑ Regions closer to the CBD appear to have less land size and building but higher no of rooms (Apartment buildings)
- ❑ Where as regions such as North and East Victoria which are located at a distance have higher land size and building area with higher number of rooms

Mean Values – Analysis by Regions – III



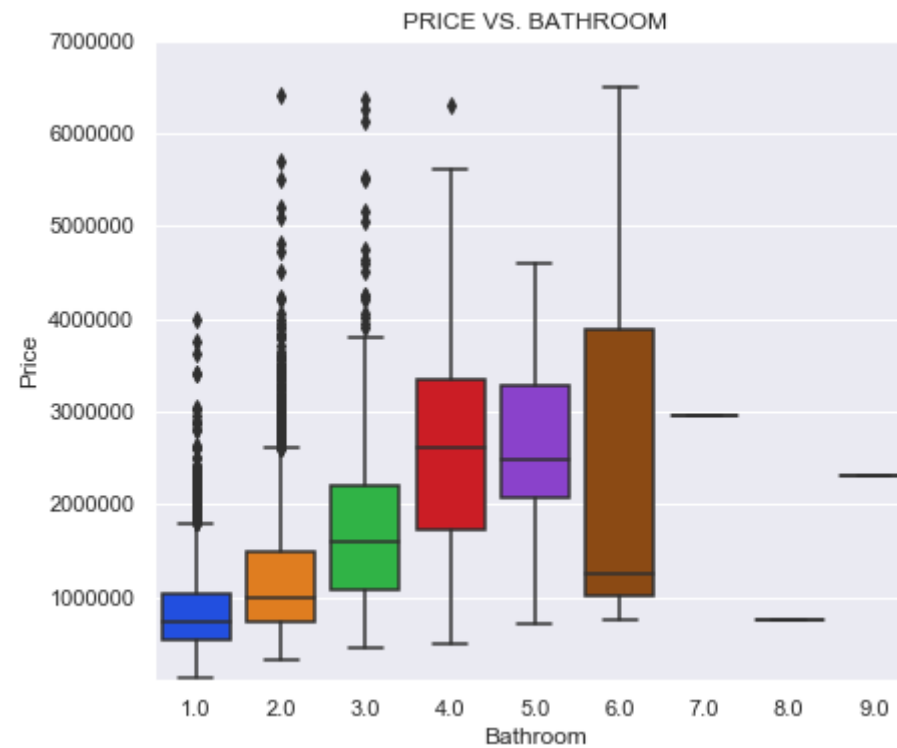
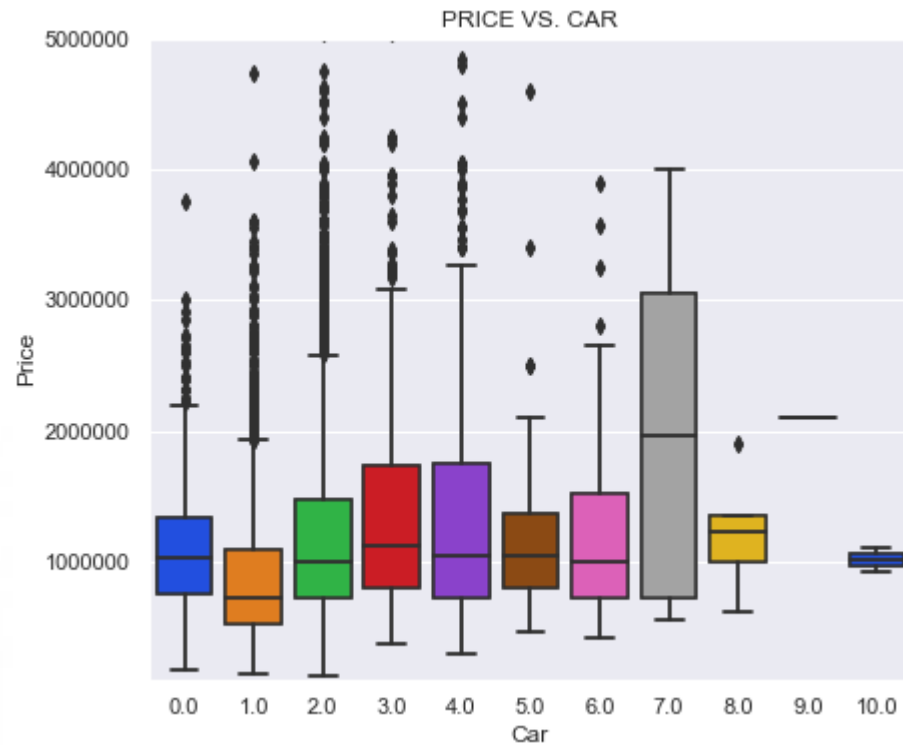
- ❑ Regions closer to CBD like North, South and West Metro are expensive with less land and building size and have fewer car spaces.
- ❑ Whereas farther regions like North and West Victoria considering higher land and building size have more car spaces
- ❑ No of Bathrooms is correlated with rooms **0.62**. From previous analysis it can be inferred that for a property with three rooms there are 1.5 to two Bathrooms available.

Price Analysis using Box Plots – I



- ❑ The prices of homes appear to be higher followed by town homes and then units.
- ❑ South Metro region closest to CBD appears to have the highest price followed by E and SE Metro regions.

Price Analysis using Box Plots – II



- ❑ Properties with higher no of car spaces and Bathrooms appear to be costly.
- ❑ Whereas properties with 3 car spaces and 3 Bathrooms have median values of approximately 1 to 1.5 Million Dollars

Problem Statement

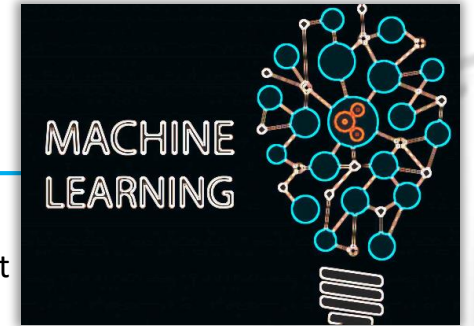
As analyzed and discussed in exploratory analysis the home prices in Melbourne depend upon several different factors.

The purpose of this project is to identify the following.

- ☐ Discover important factors which drive prices in the Melbourne housing market.
- ☐ Explore relationship between different factors and their relationship with price.
- ☐ Analyze the most significant factors that impact housing prices
- ☐ Prepare a Model based upon identified factors and accurately predict housing prices.
- ☐ Compare Models for accuracy
 - Our Target or Dependent variable for this analysis will be Price.
 - Where as important features will be selected using a machine learning algorithm.



Machine Learning Models



- ❑ For this project we will be using two different algorithms, Random Forest and Multiple Linear Regression in order to predict price of housing in Melbourne Australia.
- ❑ **Random forest** is an ensemble learning method for classification and regression which builds multiple decision trees and merges them together to get a more accurate and stable prediction.
- ❑ **Multiple Linear Regression** is used to explain the relationship between one continuous dependent variable and two or more independent variables.
 - *In present scenario our dependent variable is Price whereas the multiple independent variables include distance, land size, building area, room etc.*
- ❑ Reason for choosing the models:
 - ✓ Random Forest – The dataset appears to have large no of variables. The algorithm will help us identify most important features.
 - No of original variables 21
 - No of extended variables 229
 - ✓ Multiple Linear Regression – The algorithm will tell how much does our dependent variable *Price* changes with an amount of change in the independent variables.

Model Parameters and Results

Model Parameters

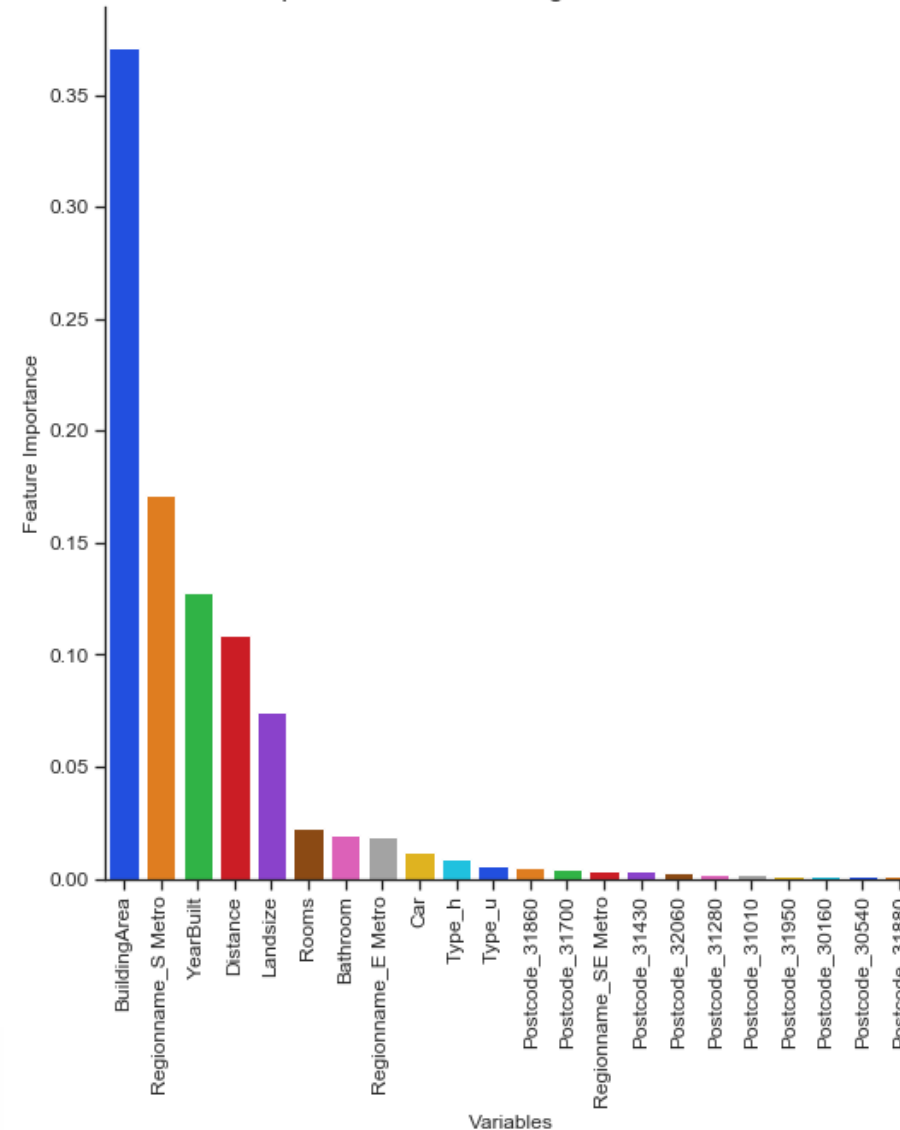
- Variables: 212
- n_estimators: 250
- max_depth: 15

Model results

Random Forest: 212 Variables

- R² Train: 0.9662708569930585**
- R² Test: 0.8078177350596782**
- Mean Absolute Error: 171398.5471691893
- Mean Squared Error: 94793273740.64745
- Root Mean Squared Error: 307885.16323565744

Top 20 Features Percentage - Random Forest



Random Forest Results with Max & Min Features

Random Forest: 212 Variables

- **Train Accuracy - R^2** : 0.9662708569930585
- **Test Accuracy - R^2** : 0.8078177350596782
- Mean Absolute Error: 171398.5471691893
- Mean Squared Error: 94793273740.64745
- Root Mean Squared Error: 307885.16323565744

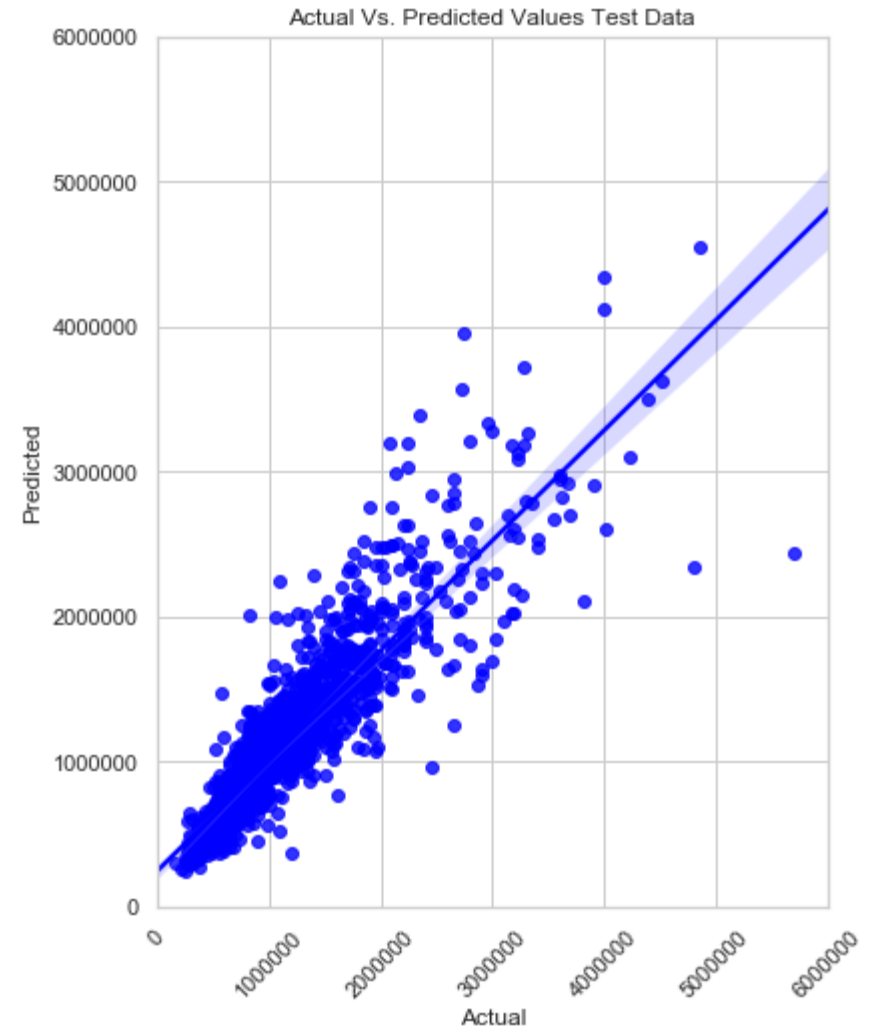
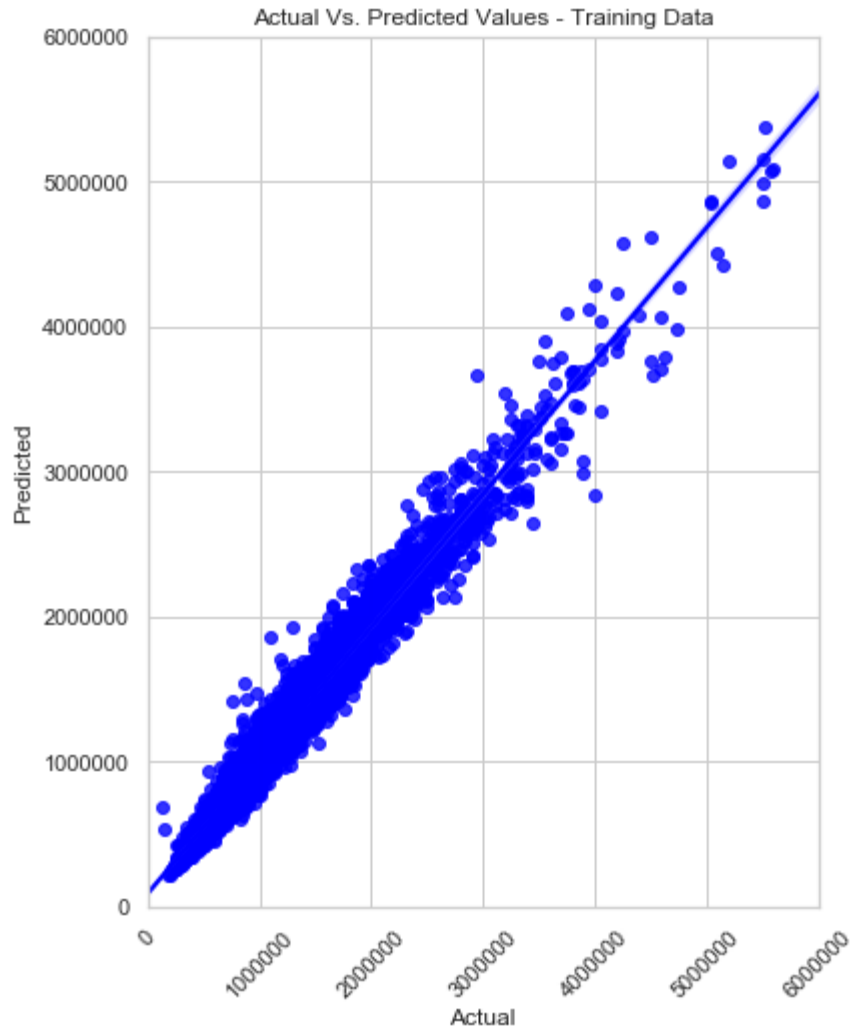


Random Forest: 22 Variables

- **Train Accuracy - R^2** : 0.9677837765735764
- **Test Accuracy - R^2** : 0.8047399497892608
- Mean Absolute Error: 174401.20392445527
- Mean Squared Error: 96311381260.84035
- Root Mean Squared Error: 310340.7502421175

1. Rooms
2. Distance
3. Bathroom
4. Car
5. Landsize
6. BuildingArea
7. YearBuilt
8. Type_h
9. Type_u
10. Postcode_30160
11. Postcode_30540
12. Postcode_31010
13. Postcode_31280
14. Postcode_31430
15. Postcode_31700
16. Postcode_31860
17. Postcode_31880
18. Postcode_31950
19. Postcode_32060
20. Regionname_E Metro
21. Regionname_S Metro
22. Regionname_SE Metro

Random Forest Results Actual Vs. Predicted Values



Linear Regression using Important Features

1. Rooms
2. Distance
3. Bathroom
4. Car
5. Landsize
6. YearBuilt
7. Type_h
8. Type_t
9. Type_u
10. Regionname_N Metro
11. Regionname_S Metro
12. Regionname_SE Metro
13. Regionname_W Metro
14. Regionname_N Vic
15. Regionname_W Vic
16. Regionname_E Vic

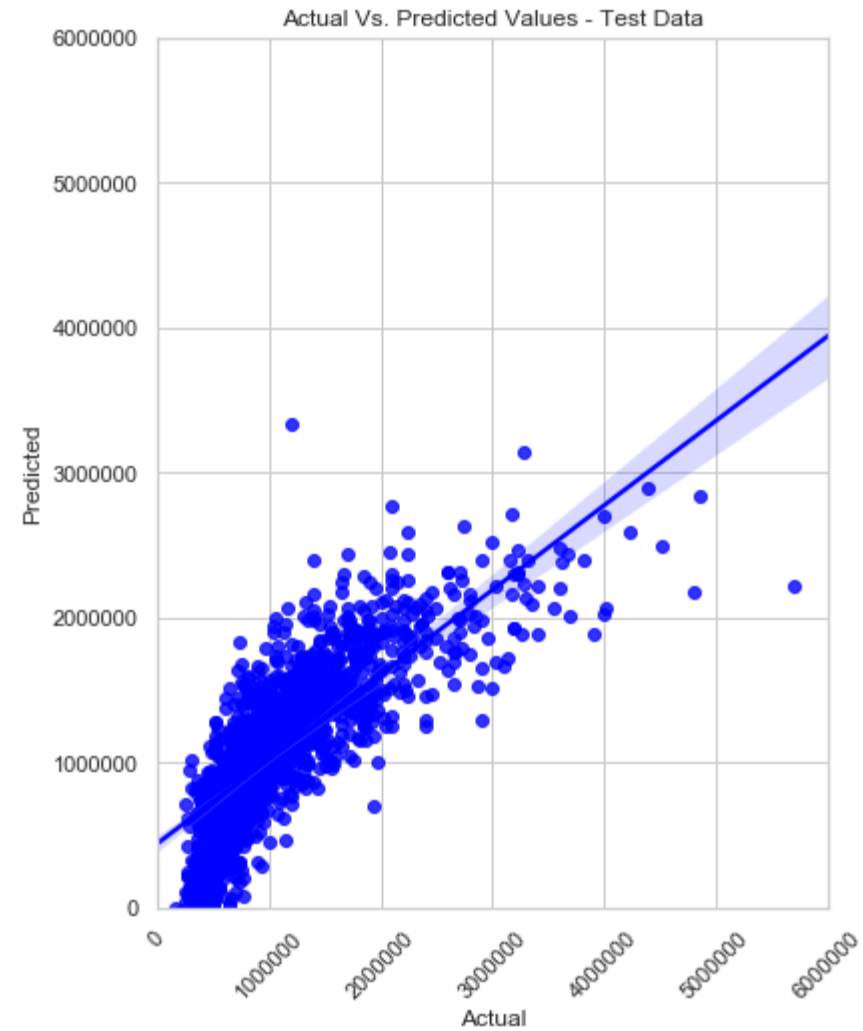
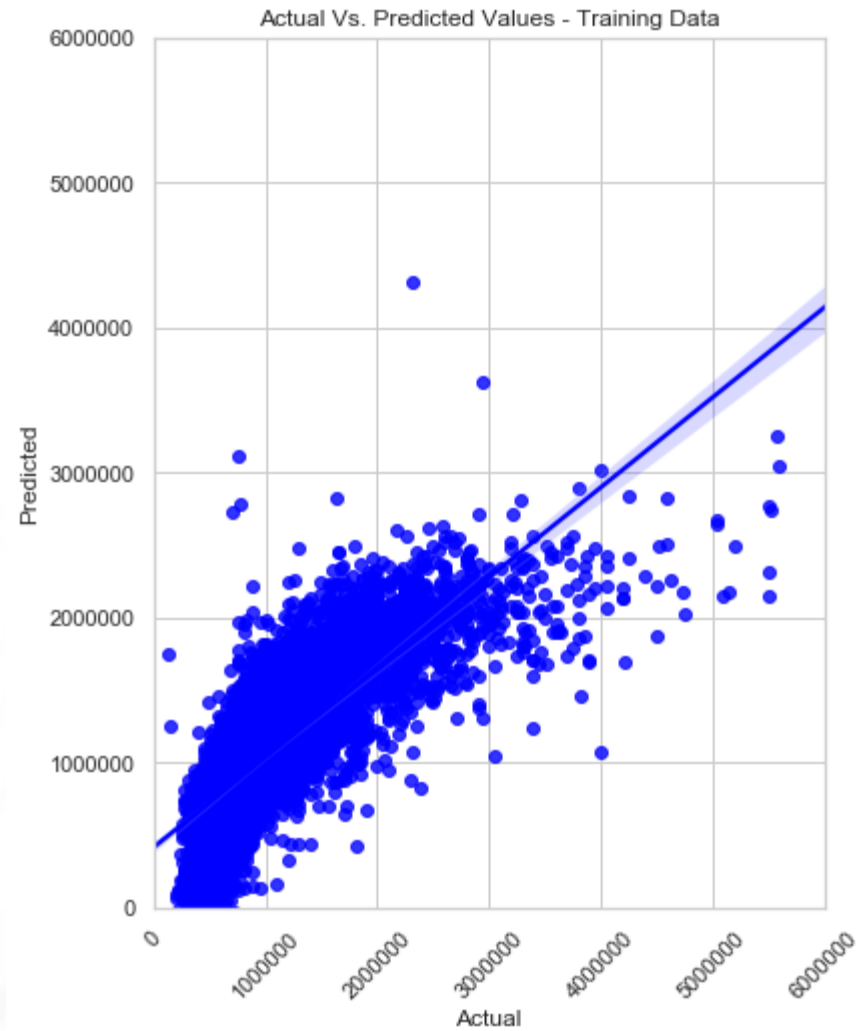
❑ Linear Regression

Linear Regression: 16 Variables

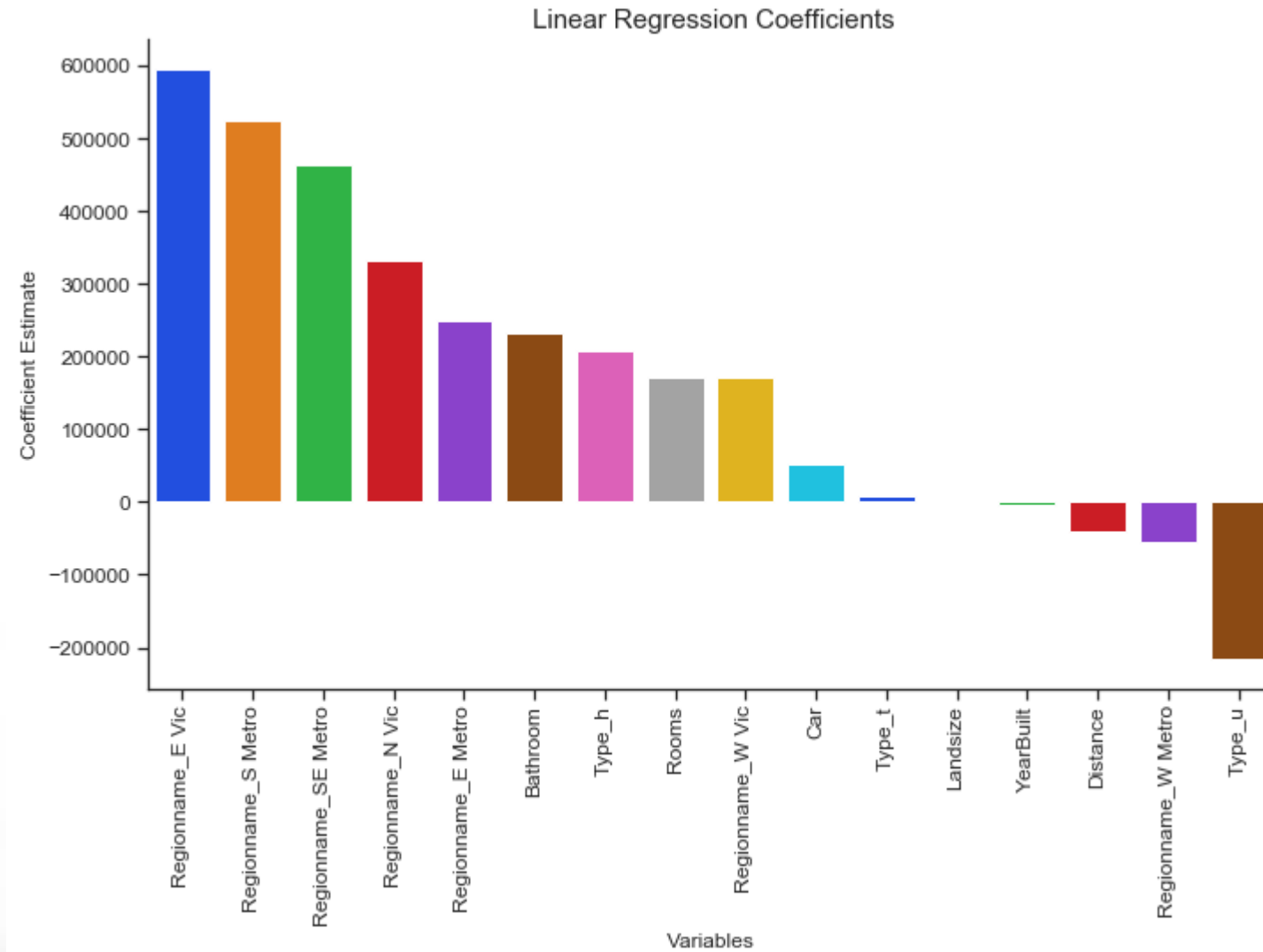
- **Train Accuracy - R^2 :** 0.6202410353433015
- **Test Accuracy - R^2 :** 0.6068245415946455
- **Intercept:** 5985283.630972797
- **Mean Absolute Error:** 279719.61707274447
- **Mean Squared Error:** 193932509164.13568
- **Root Mean Squared Error:** 440377.6892215768

	Variables	Coefficient Estimate
0	Regionname_E Vic	596619.124946
1	Regionname_S Metro	526268.527683
2	Regionname_SE Metro	463700.727390
3	Regionname_N Vic	331481.918438
4	Regionname_E Metro	248955.731742
5	Bathroom	233409.021030
6	Type_h	207884.319893
7	Rooms	172711.992977
8	Regionname_W Vic	171434.528418
9	Car	51259.354077
10	Type_t	7623.116522
11	Landsize	24.594512
12	YearBuilt	-2945.909631
13	Distance	-39451.159089
14	Regionname_W Metro	-54597.998791
15	Type_u	-215507.436416

Linear Regression Actual Vs. Predicted Values



Multiple Linear Regression Actual Vs. Predicted Values



FINDINGS

❑ Most important factors impacting price of housing in Melbourne are ranked as follows.

1. Rooms
2. Distance
3. Bathroom
4. Car
5. Landsize
6. BuildingArea
7. YearBuilt
8. Type_h
9. Type_u
10. Regionname_S Metro



CONCLUSION

- ❑ E Vic, N Vic, W Vic appear to have cheaper price, larger distance, larger land size, more car space with 80's and 90's construction.
- ❑ E Metro, N Metro, S Metro appears to have higher price, short distance, less land size, slightly less car space and 50's, 60's construction
- ❑ SE Metro appears to be expensive located at a distance with 2 car spaces and mid 70's construction.
- ❑ W Vic appears cheapest in price, 3rd largest in distance, small land size, larger building area, more number of rooms, car spaces and less bathrooms.
- ❑ The supervised learning model appears to have high variance and low bias.