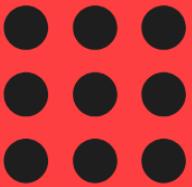


FEBRUARY 12, 2020



ANALYSIS OF MELBOURNE HOUSING



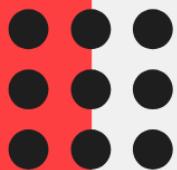
Presentation Outline

- Introduction to Dataset
- Problem Statement
- Exploratory Data Analysis
- Data Modelling
- Model Optimization
- Model Comparison
- Conclusion



DATA OVERVIEW

**Melbourne Housing
Dataset**



SOURCE

Kaggle



DIMENSIONS

34857 rows, 21 columns



DEPENDENT VARIABLE

Price



PERIOD

2016-2018

PROBLEM STATEMENT

Predicting the house price and
analyzing the market trends

DATA CLEANING

REMOVING MISSING DATA COLUMNS

Building Area

REMOVING OUTLIER'S Room, Bedroom, Landsize

REPLACING NA'S & MISSING VALUES

Land size, Car

CORRECTION OF DATATYPES

Changing as per need

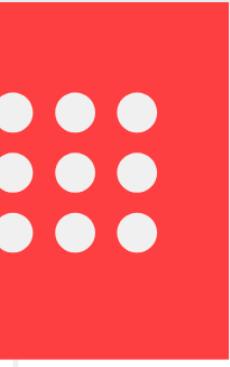


KEY CONSIDERATION

- **LAND SIZE < 120**

No land size can be less than 120 square feet
- **BATHROOM MISSING VALUES FIXING**

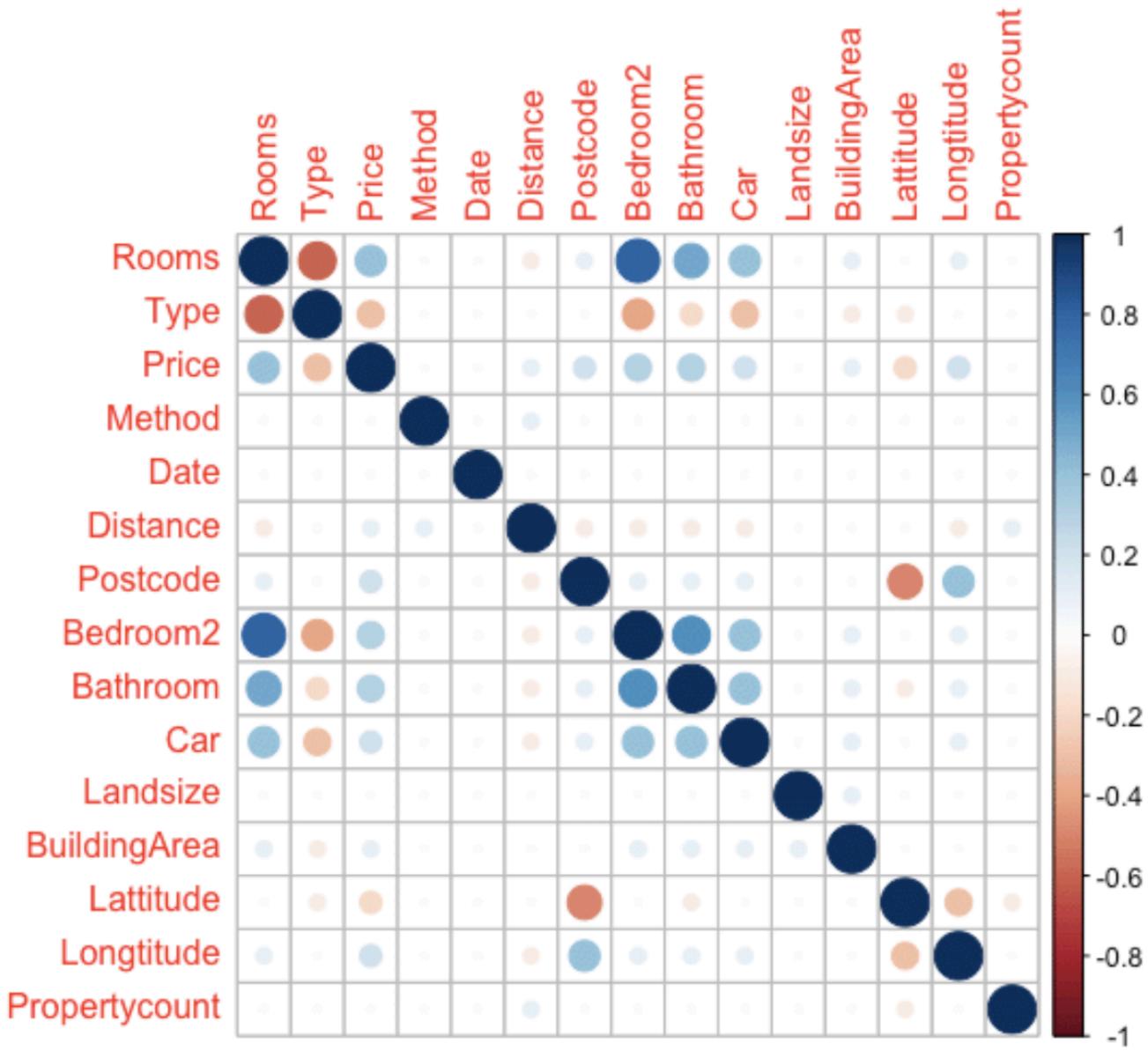
Maintained 3:2 ratio to maximize desirability
- **CORRELATION BETWEEN ROOM & BEDROOM**
- **Replace missing and NA value by median**



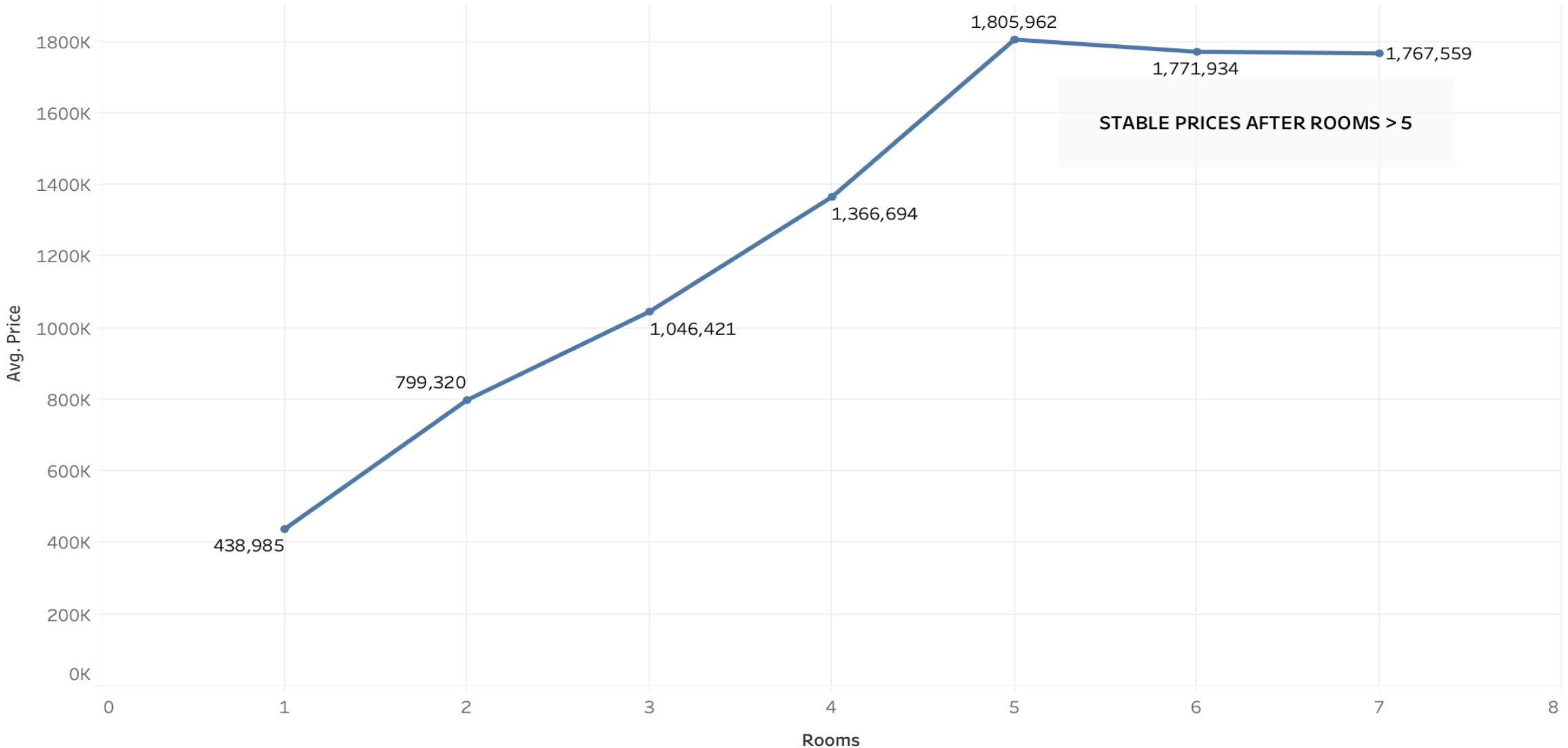
EXPLORATORY DATA ANALYSIS



CORRELATION MATRIX

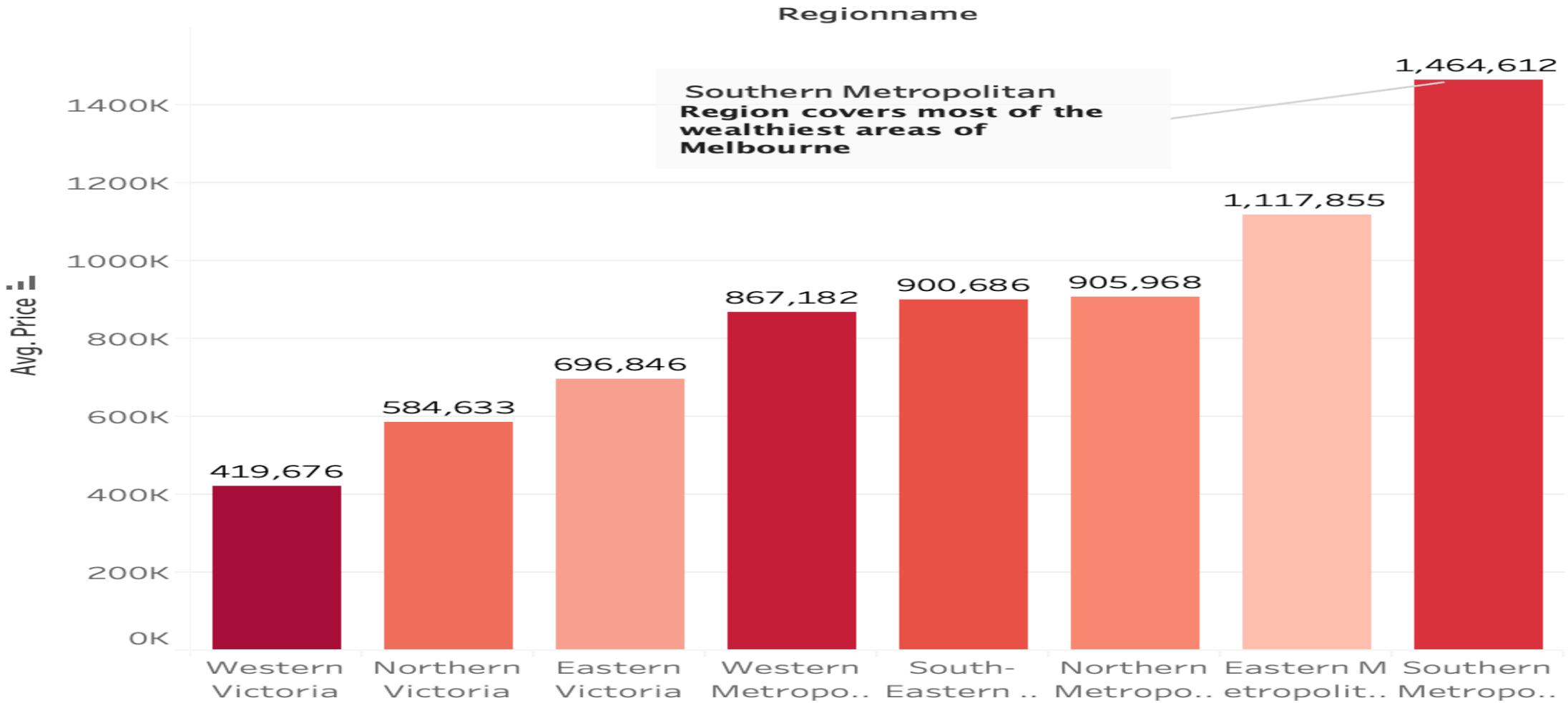


Average Price Per Room in Melbourne City.(AUD)



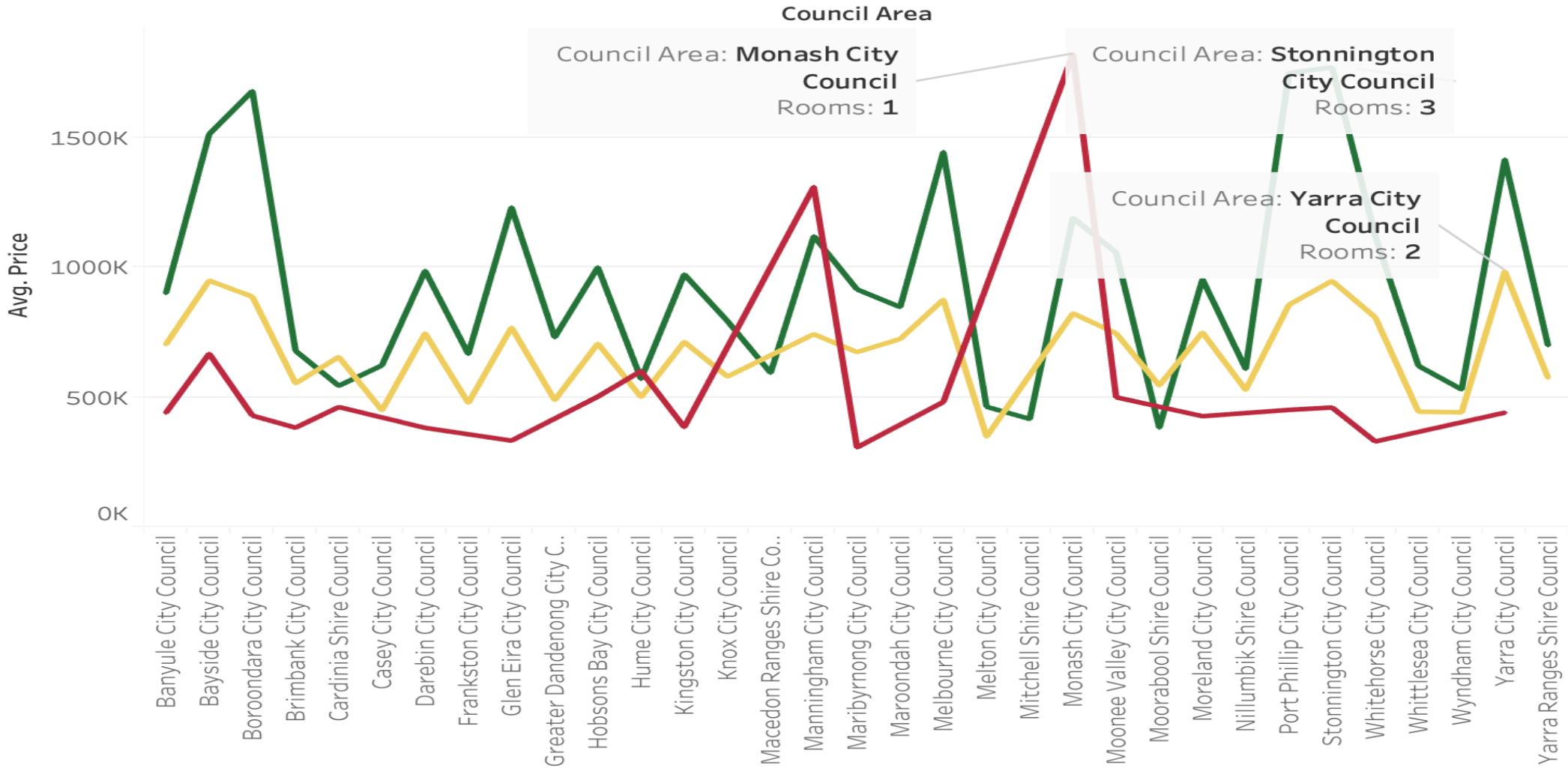
Expensive Side of Town.

Region Name compared with the median Prices.



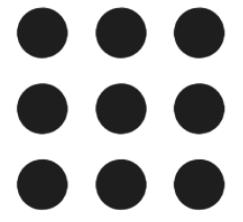
Where should I not be looking for 1-2-3 room Houses.

Average Price per Council compared with Number of Rooms.





DATA MODELING



LINEAR REGRESSOR

MODEL ANALYSIS METRIC

R-square

46.84%

ACUURACY PERCENTAGE

SIGNIFICANT VARIABLE

Council Area, Type,Latitude, longitude, Season, Rooms, Method,
Distance, Bathroom, Car, landsize, Postcode

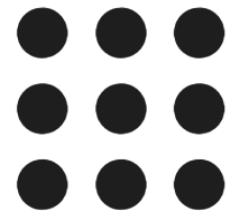
Did you know?



MODEL OPTIMIZATION

- 
- ONE-HOT ENCODING
 - STEP WISE SELECTION
 - NORMALIZATION
 - FEATURE ENGINEERING





LINEAR REGRESSOR

Did you know?

MODEL ANALYSIS METRIC

R-square

62.13%

ACCURACY PERCENTAGE

SIGNIFICANT VARIABLE

Positively Significant: Season, Rooms, Method, Distance, Bathroom, Car, landsize

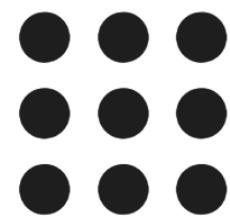
Negatively Significant: Council Area, Type,Latitude & longitude,





MODEL 2: RANDOM FOREST REGRESSOR

- Least Overfitting
- Ensemble Learning
- Uncorrelated Models
- Robust



RANDOM FOREST REGRESSOR

MODEL ANALYSIS METRIC

R-square

80.06%

ACCURACY PERCENTAGE

IMPORTANT VARIABLE

Latitude, longitude, Rooms, Type, Bathroom, landsize, Council Area

Did you know?



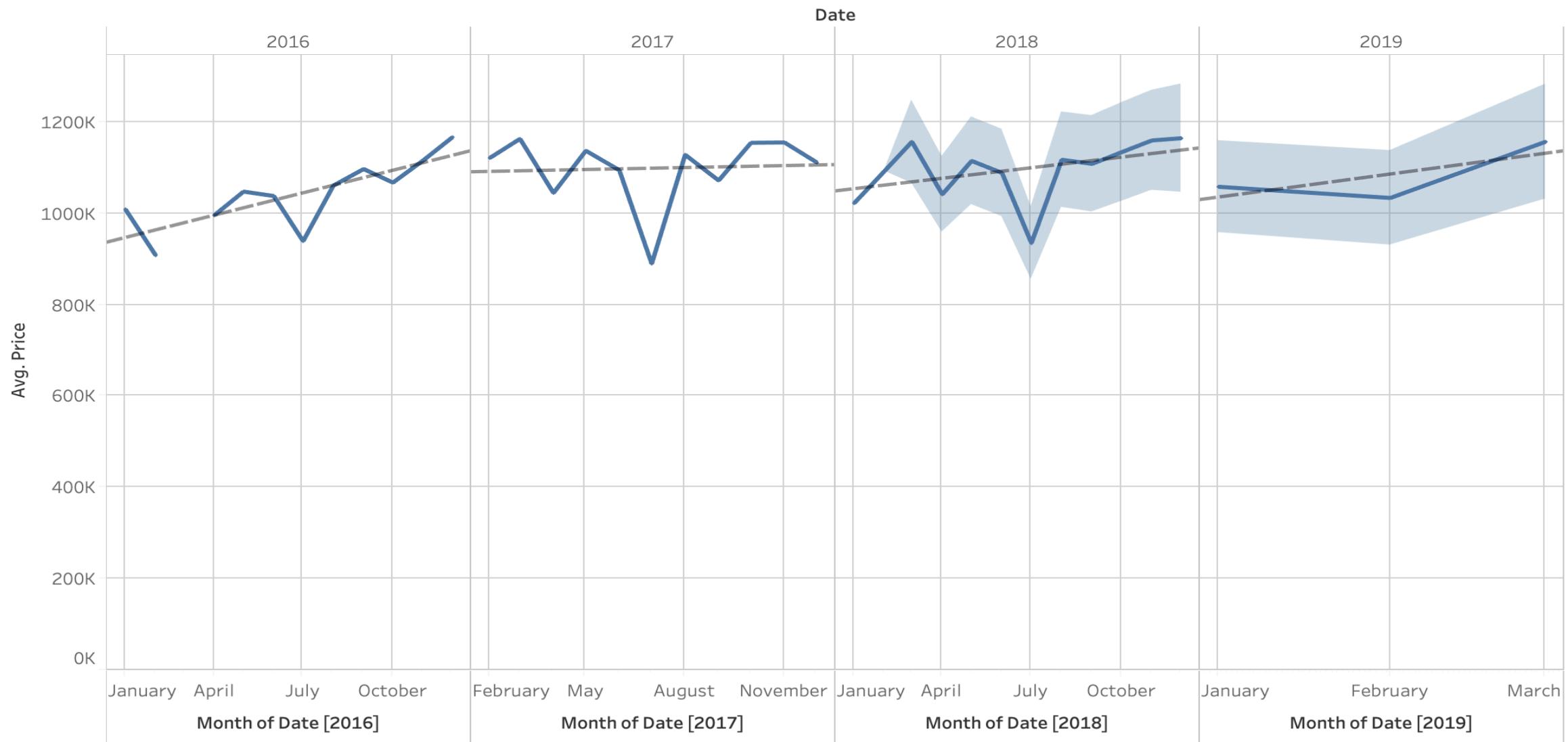


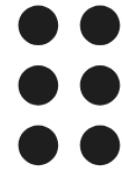
CONCLUSION

- **Southern Metropolitan is the most expensive side of town**
- **Considerable Variables: Rooms, Distance from CBD, Bathroom, Land size, Council Area**
- **Based on the updated data model can predict the future prices for there investment**
- **Best Model: Random Forest Regressor**



Monthwise Trends in Melbourne Housing Market.

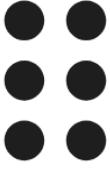




REFERENCES

- <https://hoa.org.uk/2017/05/room-ratio/>
- [/www.godownsize.com/minimum-house-square-footage/](http://www.godownsize.com/minimum-house-square-footage/)
- <https://www.kaggle.com/anthonypino/melbourne-housing-market>





THANK YOU

