Melbourne Housing Prices

Agenda

01 Background & beneficiaries

02 Data collection & cleansing

03 Exploratory data analysis

04 Model building

05 Conclusion

Background

- Melbourne is in the midst of a population boom, leading houses' prices to spike year after year

- Prospect property owner or real estate agent would like to estimate the value of a given estate to avoid overpricing

- Construction companies would like to examine the impact of different characteristics on a property's final price

Data collection & cleansing

- Properties' characteristics collected from [Kaggle](#), including distance from centre, yearbuilt, and general information about house sizes

- Visualisation was made using data from Foursquare API

- After cleansing duplicates and heavily correlated variables, the final dataset contains 22 columns and 8,842 rows
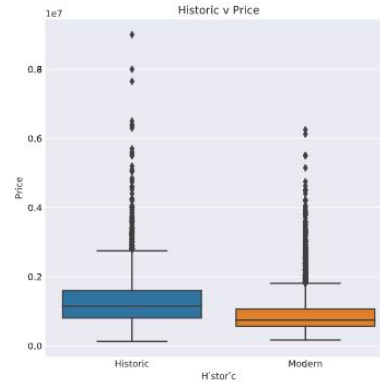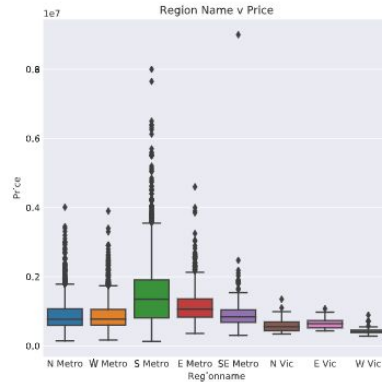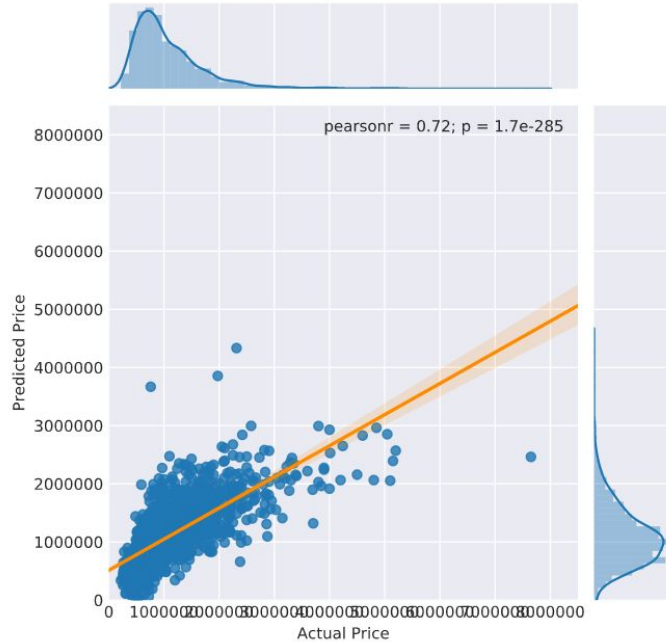
Clusters of homes

Distance & age impact

Region Name v Price

Historic v Price
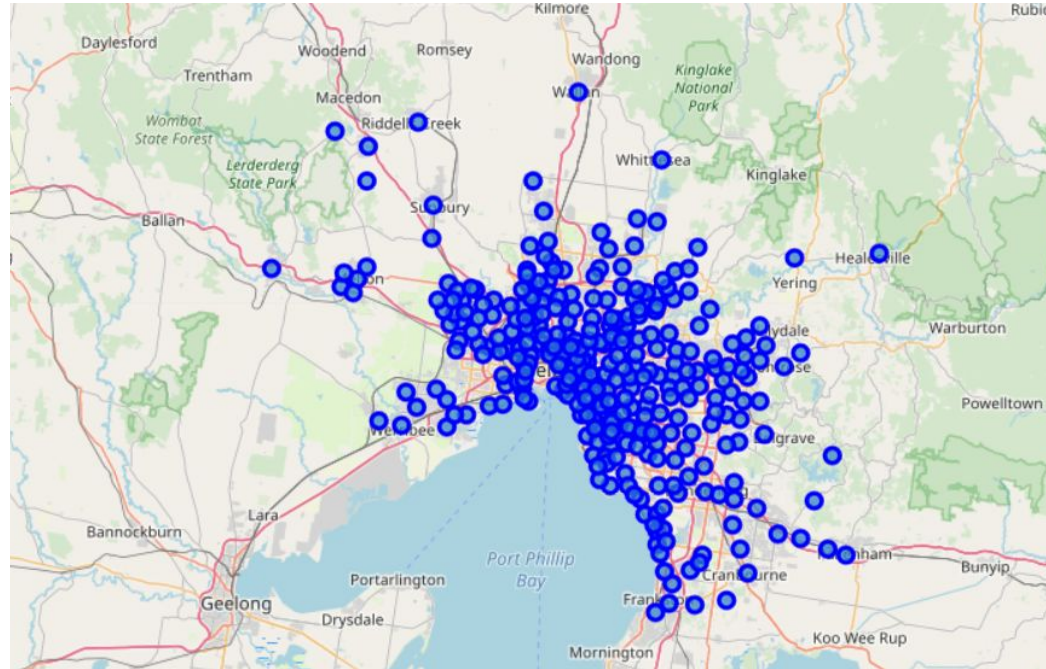
- Houses close to the Metro Area tend to be more expensive.

- Older houses generally are sold for higher but have larger variance

Coefficients of variables

	Coefficients
Rooms	130781.994615
Distance	-28481.473952
Bathroom	255950.485267
Car	49936.398377
Landsize	24.128532
BuildingArea	2160.118289
Propertycount	-1.229619
Age	5491.481614

Evaluation scores

Mean absolute error	Root mean squared error	R squared	Cross validation score
312564.47	470194.87	0.52	0.5

Actual vs Fitted Values for Price

Conclusion

- The number of rooms has largest impact on prices, while land size has the least (distance to the centre is more important)

- R squared and cross validation score hovering around the 0.5 mark leaves room for improvement (perhaps using SVM or classification models)

- Further studies can consider more features such as furniture, public transportation, amenities, etc.

Thank you