The Analysis of NSW House Price and Growth

With relation to demographic characteristics and socioeconomic factors

2021

Group 99

Members

Felix Rosenberger

46165991

Christopher Rudolph

45241805

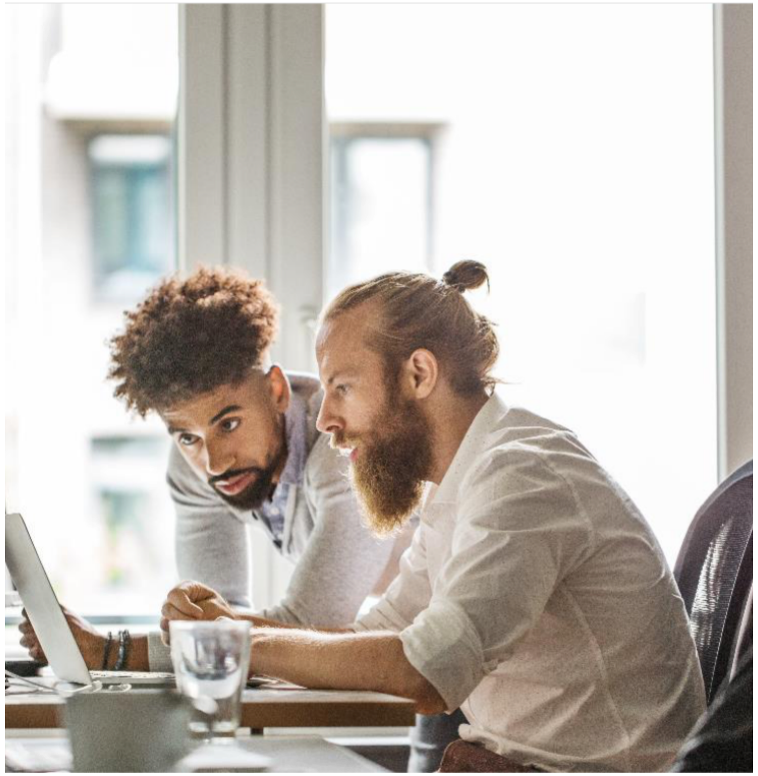

Ken Walther Sy

46162291

Jingran Zhao

46317848

Key Topics



1. **Trends** in the NSW property sales and rental market in the context of COVID pandemic.

2. **Correlation** between housing market activities and selected demographic & socioeconomic features

3. **Predicting** 2021 Q1 house price at postcode level

4. **Clustering** similar postcodes and predicting house prices based on individual clusters

5. **Predicting** areas with high growth potential

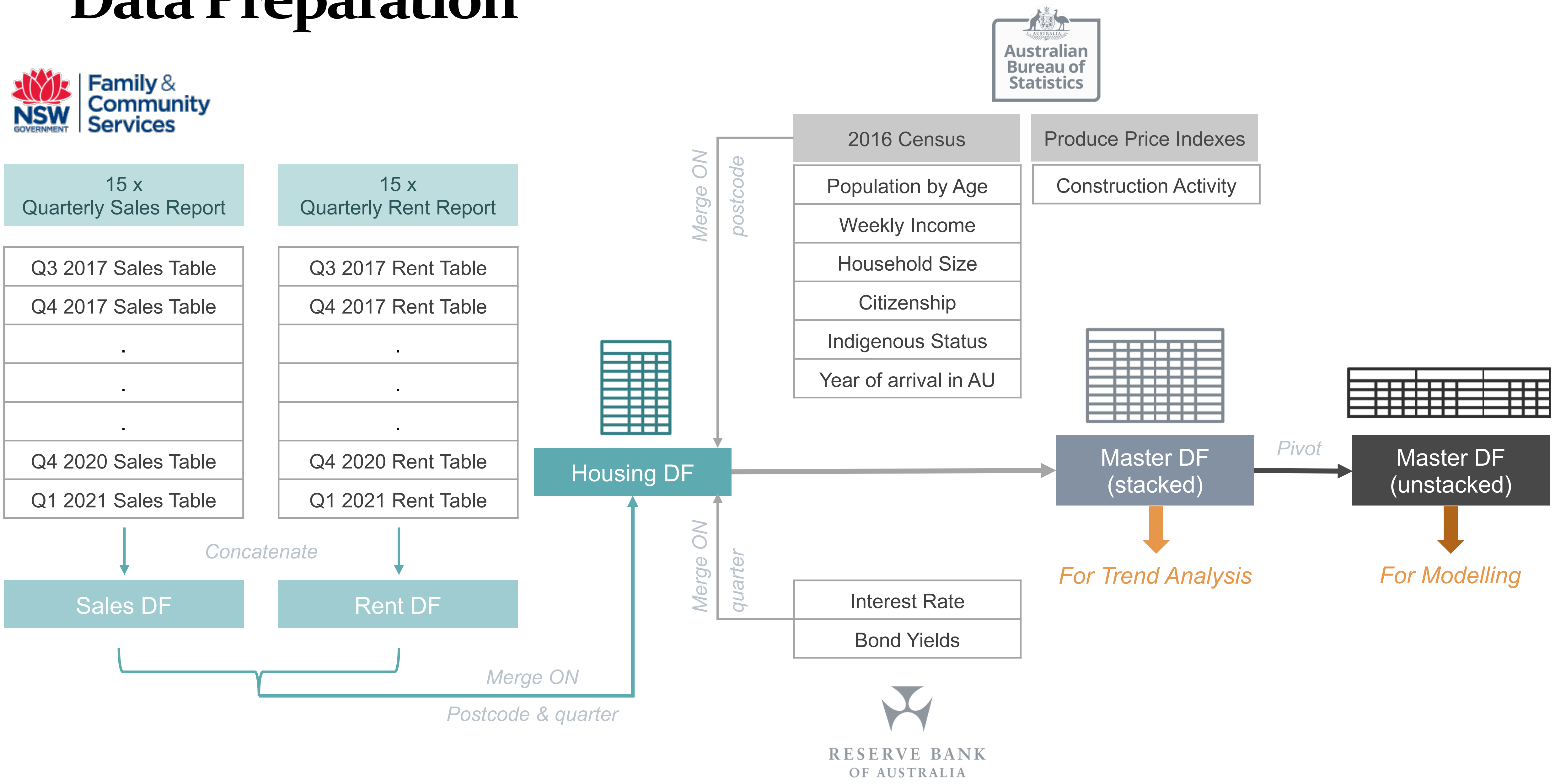
Data Sources

Australian Bureau of Statistics TableBuilder

NSW Family & Community Services

Reserve Bank of Australia

Data Preparation



Exploratory Data Analysis

Master DF
(stacked)

For Trend Analysis

Univariate Analysis

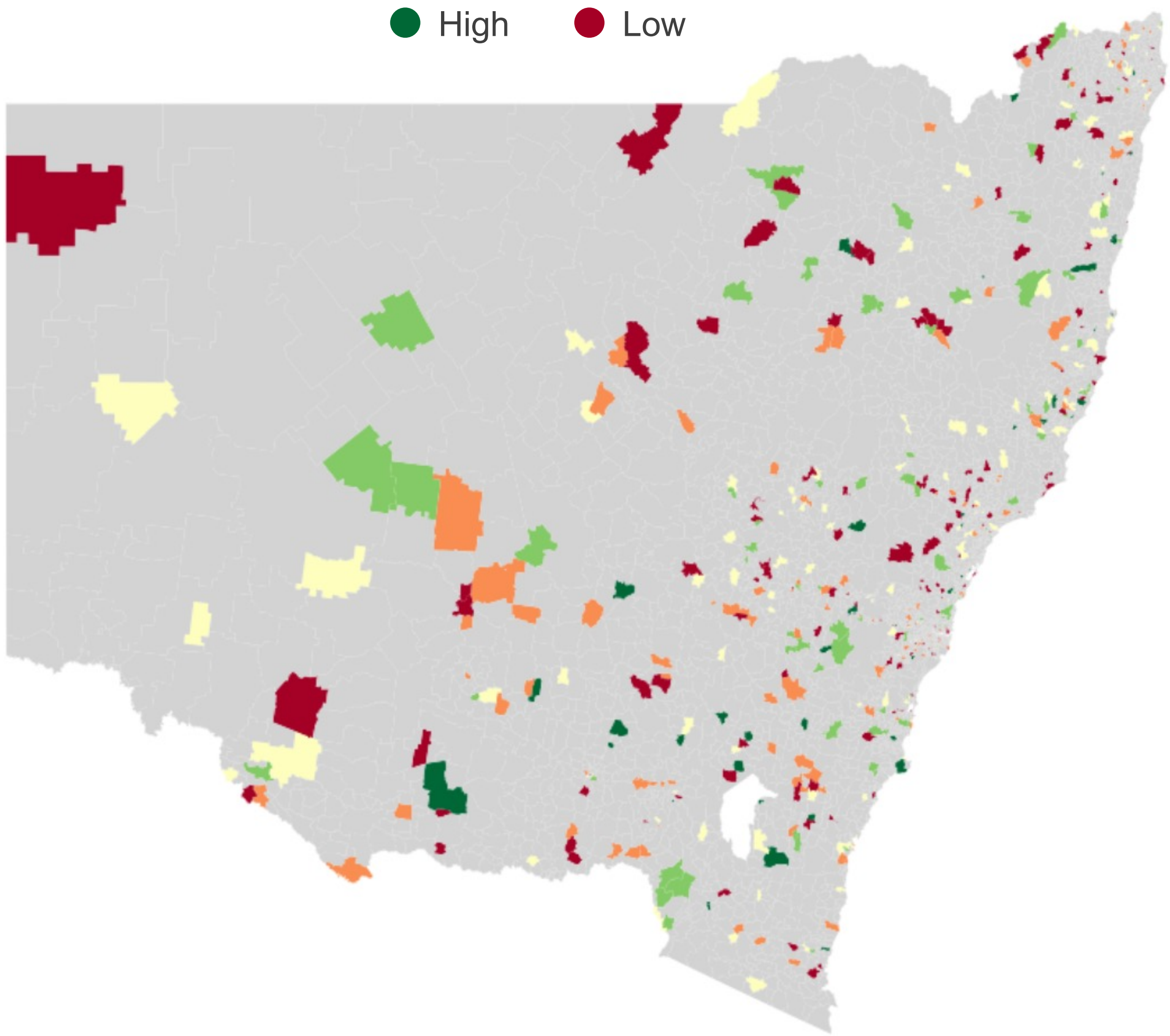
Analysis of each feature contained in the dataset regarding its distribution and outliers.

- 2.2 Univariate Analysis
  - 2.2.1 Median Price
  - 2.2.2 Sales Numbers
  - 2.2.3 Median Rent
  - 2.2.4 New Bonds
  - 2.2.5 Total Bonds
  - 2.2.6 Income
  - 2.2.7 Number of People in Household
  - 2.2.8 Age Intervals
  - 2.2.9 Population
  - 2.2.10 Citizenship
  - 2.2.11 Year of Arrival in Australia

Bivariate Analysis

Analysis of the relationship of different features with price.

- 2.3 Bivariate Analysis with Target Variable
  - 1
  - 2.3.1 Over Time
    - 2.3.1.1 Sales Number
    - 2.3.1.2 Median Rent
    - 2.3.1.3 New Bonds
    - 2.3.1.4 Construction Costs
    - 2.3.1.5 Bond Yields
    - 2.3.1.6 Interest Rates
  - 2
  - 2.3.2 By Postcode
    - 2.3.2.1 Median Rent
    - 2.3.2.2 New Bonds
    - 2.3.2.3 Age Brackets
    - 2.3.2.4 ATSI
    - 2.3.2.5 High Income
    - 2.3.2.6 Year of Arrival
    - 2.3.2.7 Non-AUS Citizen
  - 3
  - 2.3.3 For transformed Features
    - 2.3.3.1 Sales Proportion
    - 2.3.3.2 New Bonds Proportion
    - 2.3.3.3 Total Bonds Proportion
    - 2.3.3.4 Income Proportion
    - 2.3.3.5 Household Size Proportion
    - 2.3.3.6 Age Proportion
    - 2.3.3.7 Citizenship Proportion
    - 2.3.3.8 Year of Arrival Proportion



Which postcodes have the highest prices?

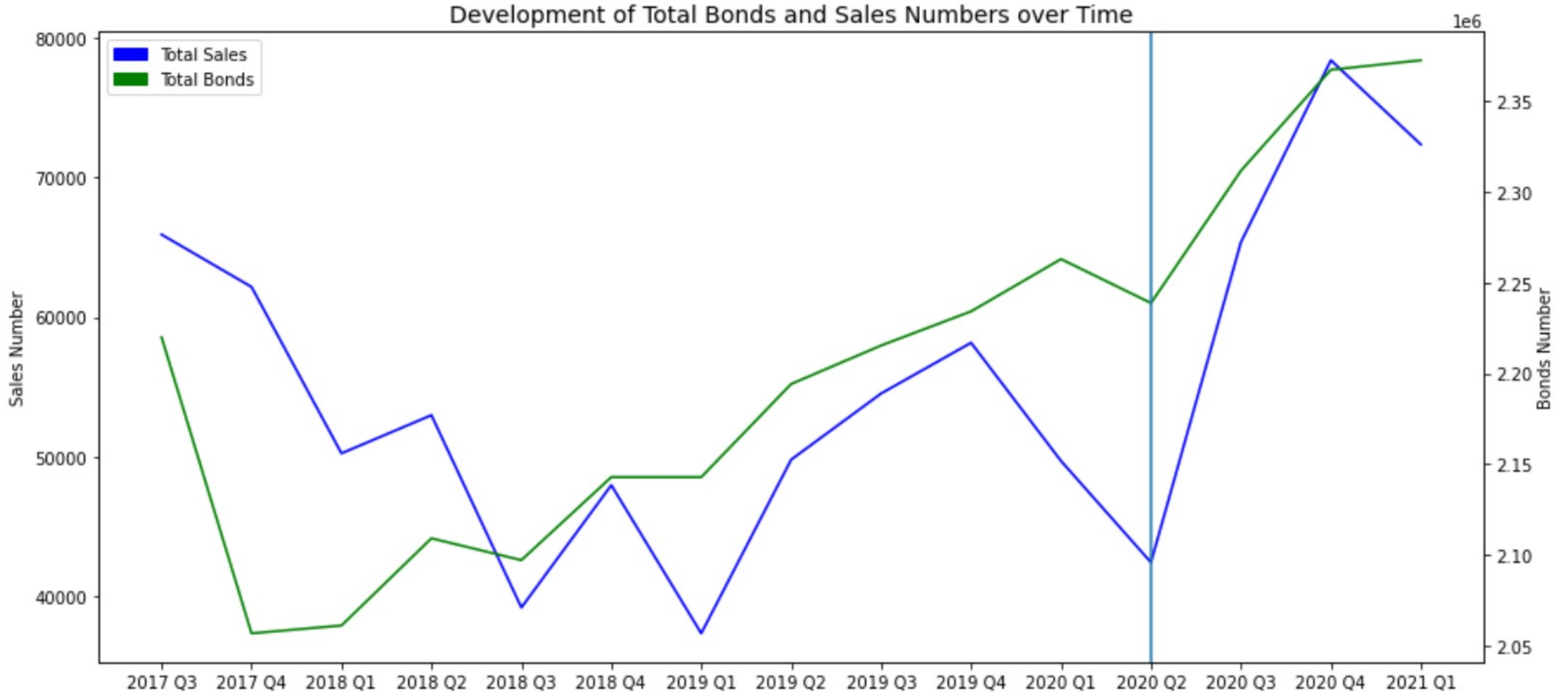
As measured in median house price

Postcode	Suburb	LGA	Median Price in AUD$
2030	Dover Heights	Waverly	2.921 mil.
2108	Coasters Retreat	Northern Beaches	2.659 mil.
2023	Bellevue Hill	Woollahra	2.651 mil.
2024	Bronte	Waverly	2.209 mil.
2088	Mosman	Mosman	2.165 mil.

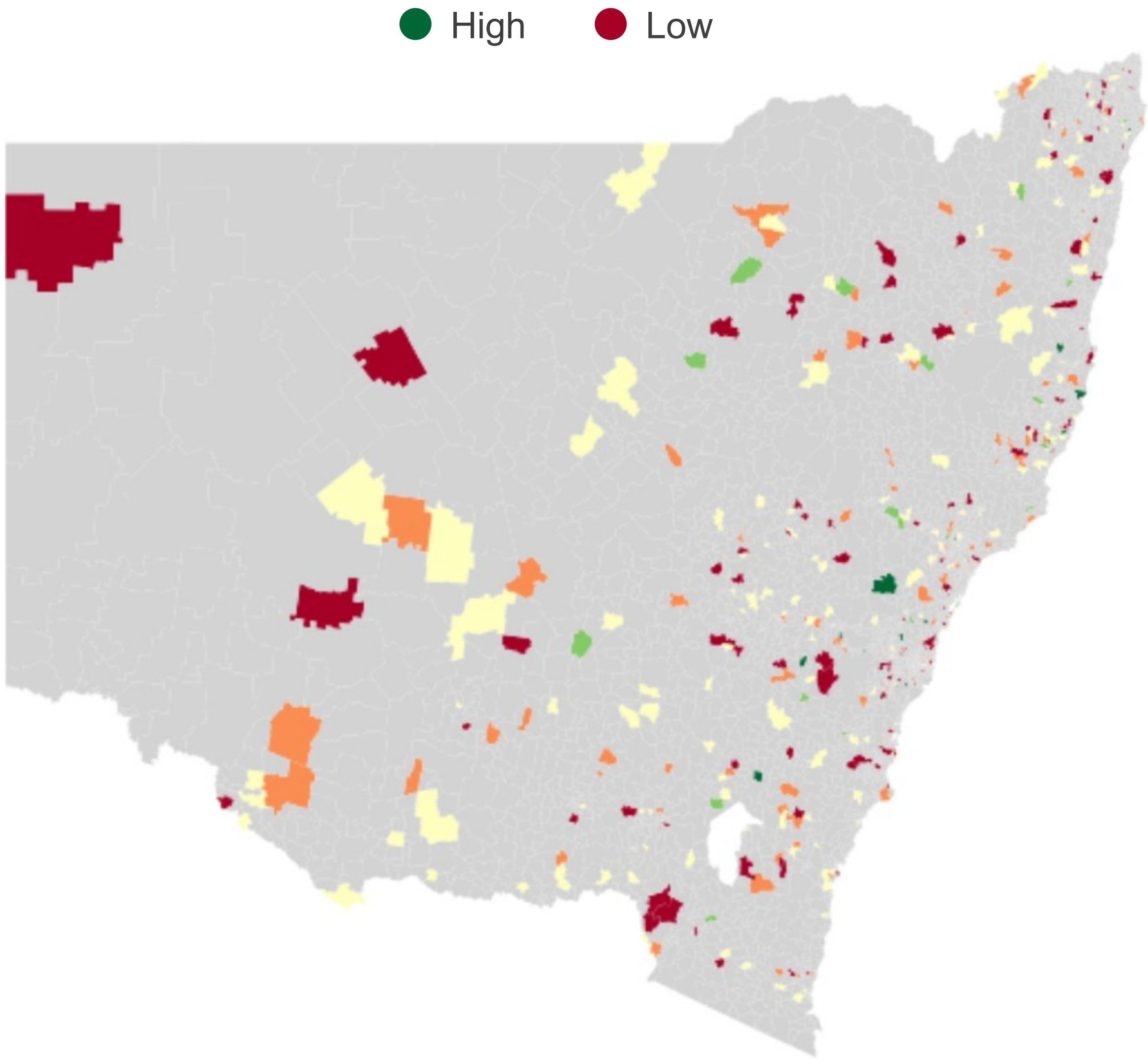
The most expensive areas are directly in or close to Sydney.

Have sales activities changed during the pandemic?

As measured in absolute sales numbers



- Sales activity rose by around 70% since the Covid low
- Amount of bonds rose by around 6% only in the same time

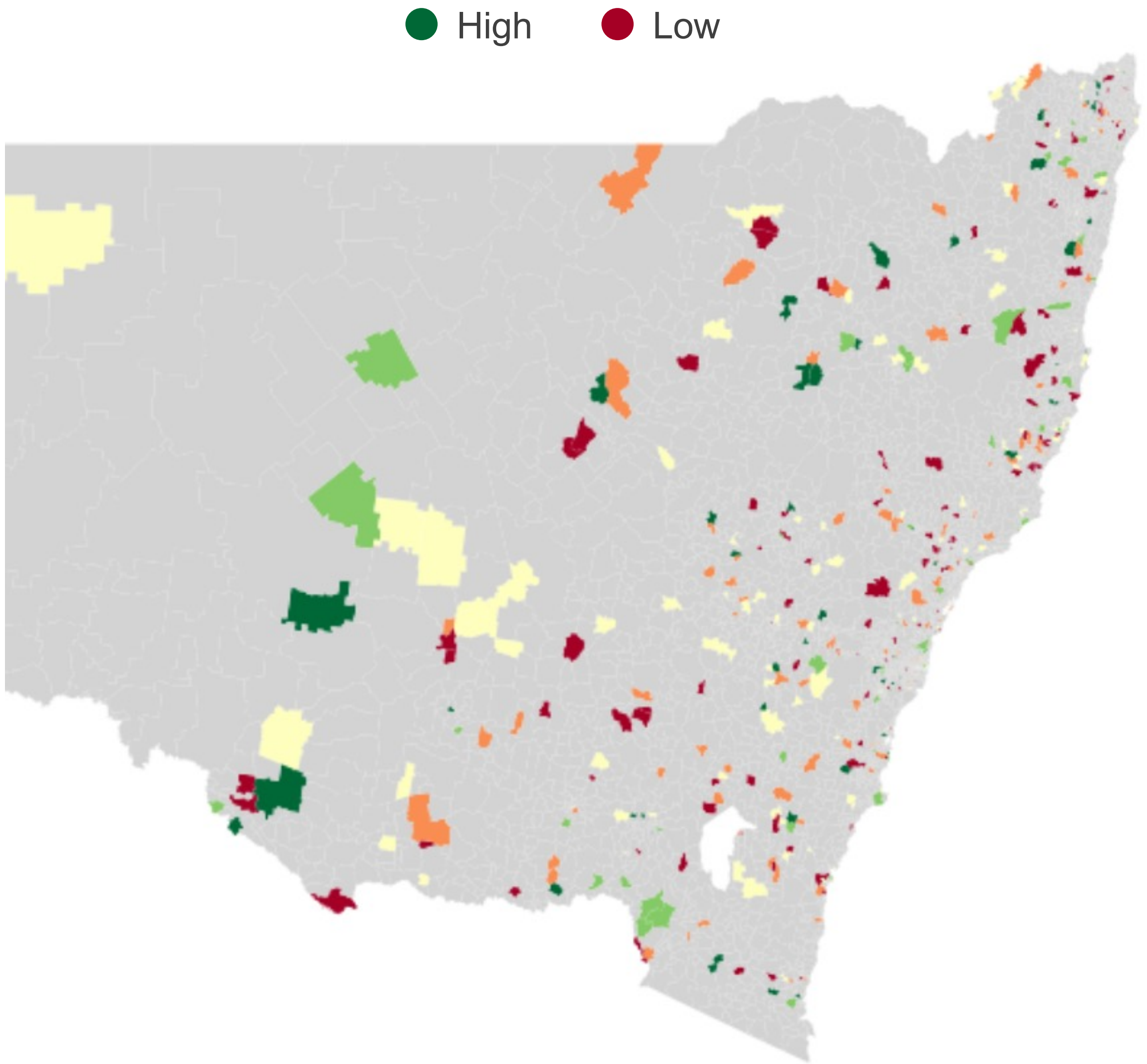


Which postcodes had strong sales increase during the Covid outbreak?

As measured in percentage change

Postcode	Suburb	%-Change
2292	Broadmeadow	300
2329	Cassilis	300
2352	Kootingal	300
2358	Arding	300
2372	Back Creek	300

The highest sales activites are found in rural areas.



Which postcodes had strong sales increase since the Covid outbreak?

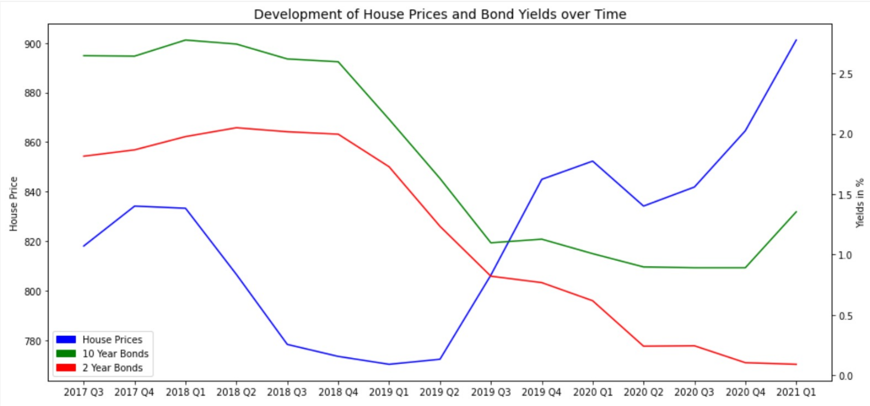
As measured in percentage change

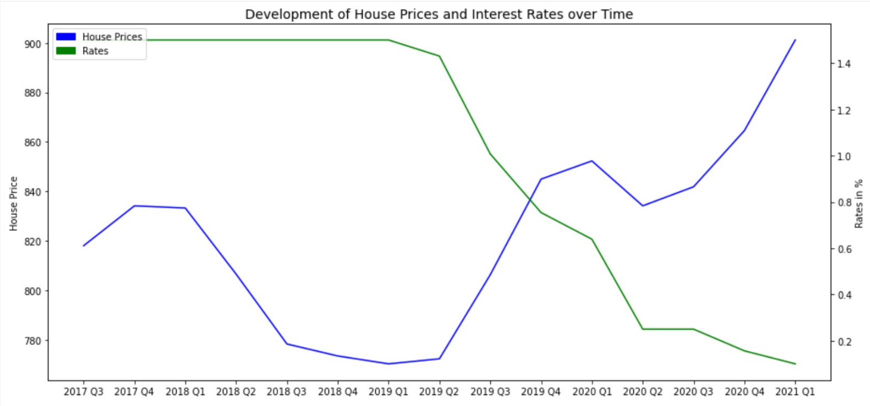
Postcode	Suburb	%-Change
2174	Edmondson Park	1120
2427	Crowdy Head	710
2762	Schofields	374
2116	Rydalmere	374
2082	Berowra Heights	350

The highest sales activites are mostly found in Sydney areas.

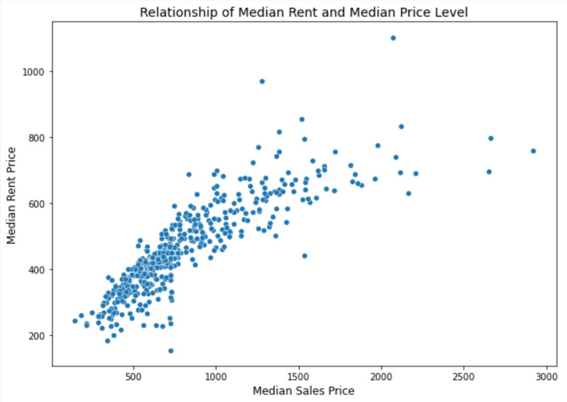
Are there any interesting relationships between features and price?

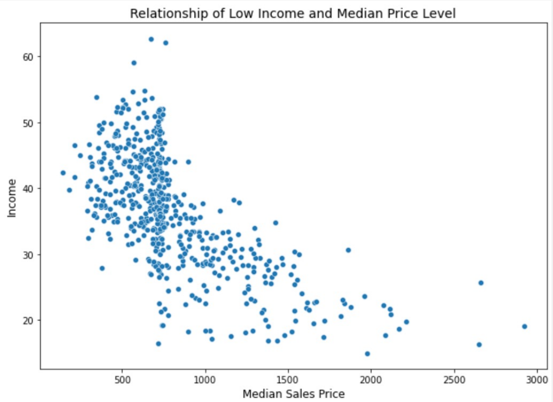
Over Time





By Postcode





Predicting House Prices with Linear Regression

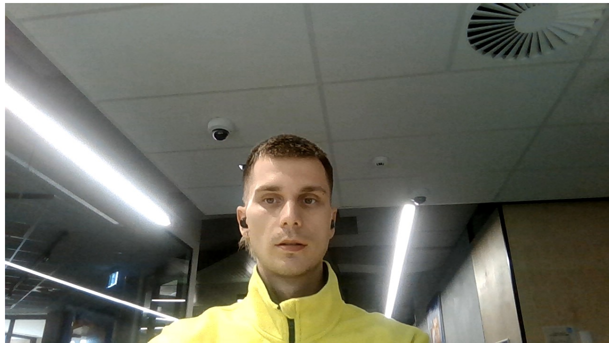
	Parameters Tuning	Feature Selection	R^2 (cross-validated)	Test MSE
Model 0 (Base)	-	All (67)	86.78%	28,344
Model 1	-	RFE (41)	88.04%	27,082
Model 2	GridSearch (LR)	RFE (41)	88.30%	27,768
Model 3 (Final)	-	GridSearch (RFE) (35)	87.42%	30,104

Final Model Performance

- **R^2 = 0.87**
  87% of the variation in the Q1'21 median prices is explained by the model
- **MSE = 30,104**

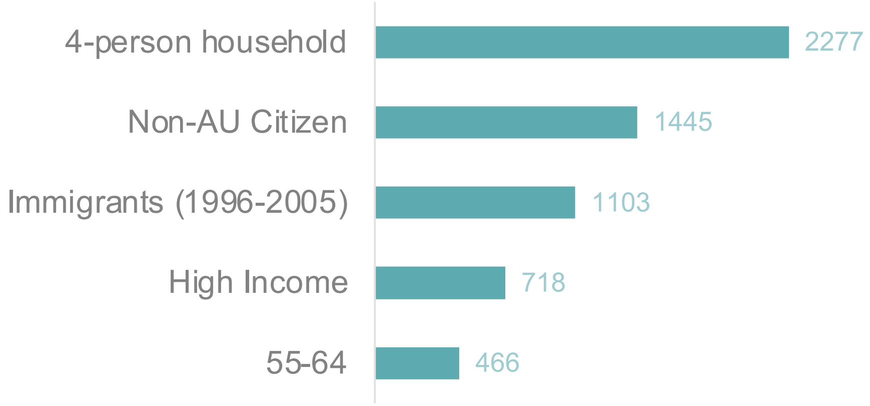
Interpretation of Results

- A R^2 score of 1.0 indicates a model with perfect fit
- Similarly, a MSE is the squared average error in the prediction, lower is better
- Feature importance is depicted by the coefficients in the model.

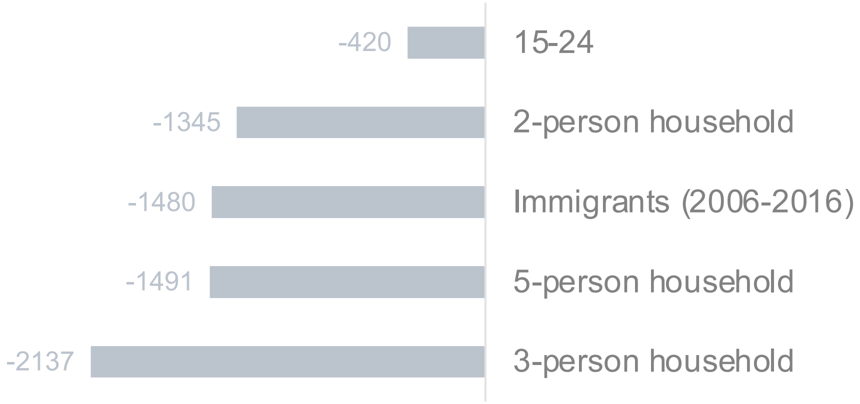


Features of the model

Top 5 Positive Predictors of House Price



Top 5 Negative Predictors of House Price



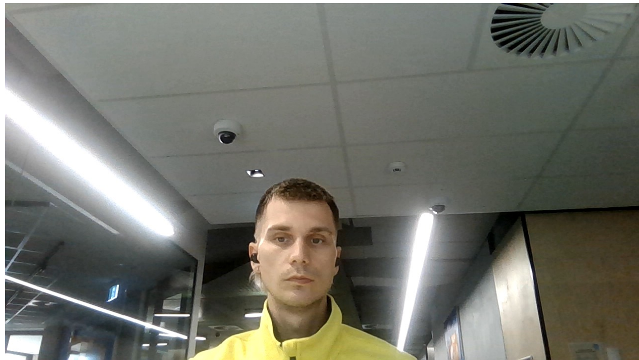
I. Features shown are their proportions to the population of each postal area

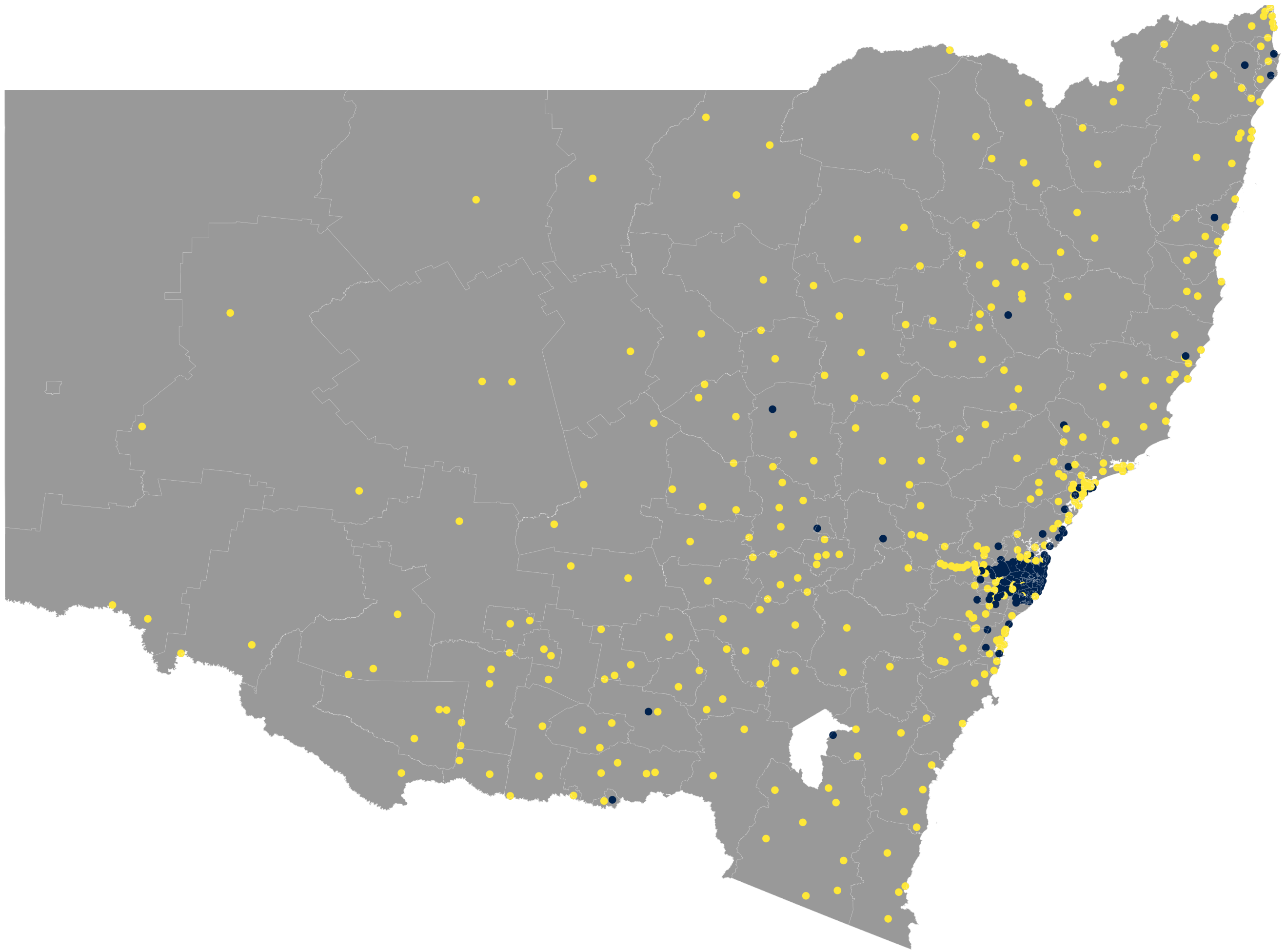
II. Each feature’s coefficient represents the impact 1 unit change in their value has on the predicted price

III. It is clear the number of people in a single household has the strongest predicting power

- 4-person households (+)
- 2, 3 and 5-person households (-)

IV. Immigration, Income and age also play their parts





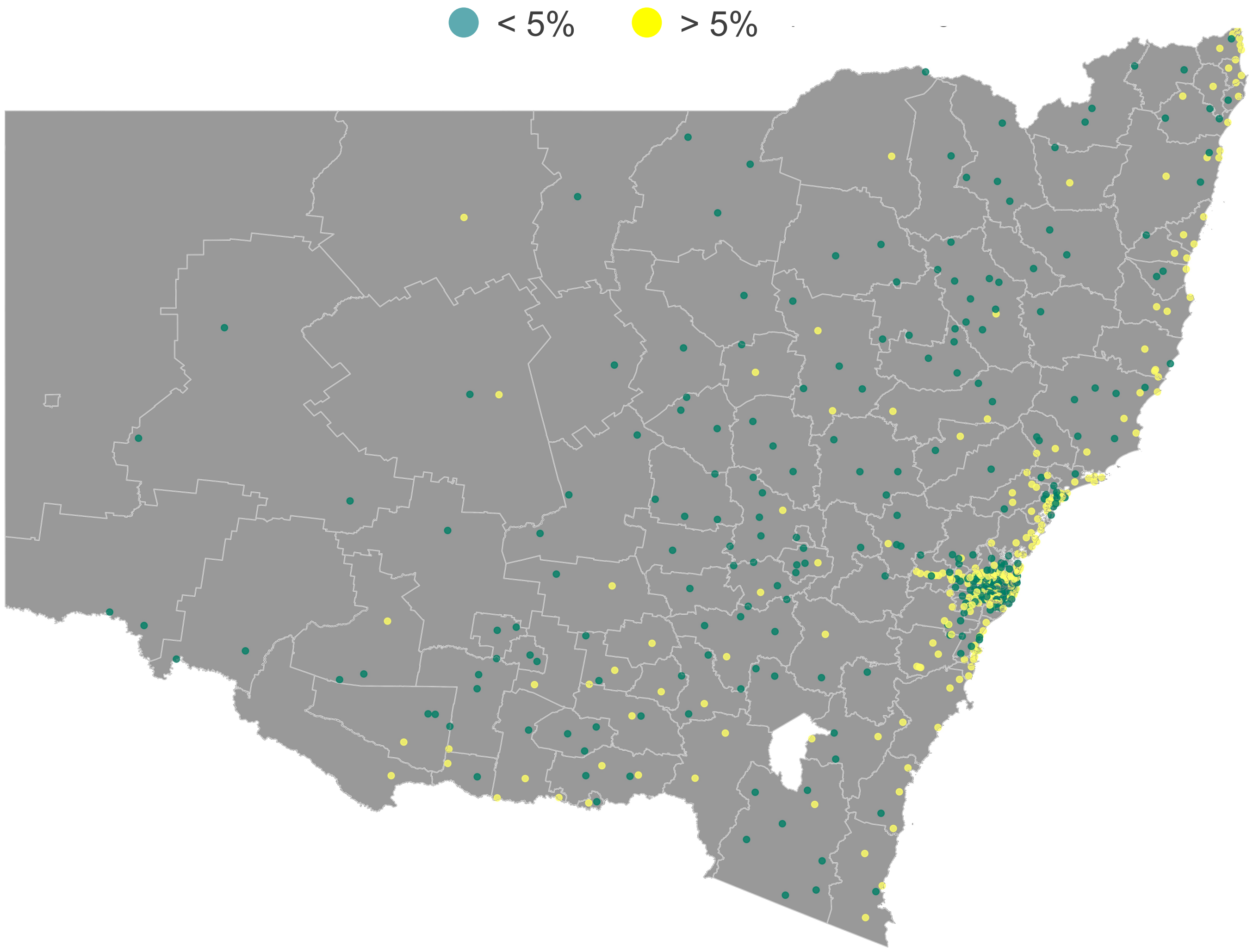
Which postcodes are similar to one another?

Using KMeans algorithm

There are some postcodes outside Sydney that were classified as the same group. However, it may be more difficult to predict given the smaller dataset.

Linear Regression on clusters

Cluster	Train R^2	Test R^2
Blue	93%	82%
Yellow	82%	63%



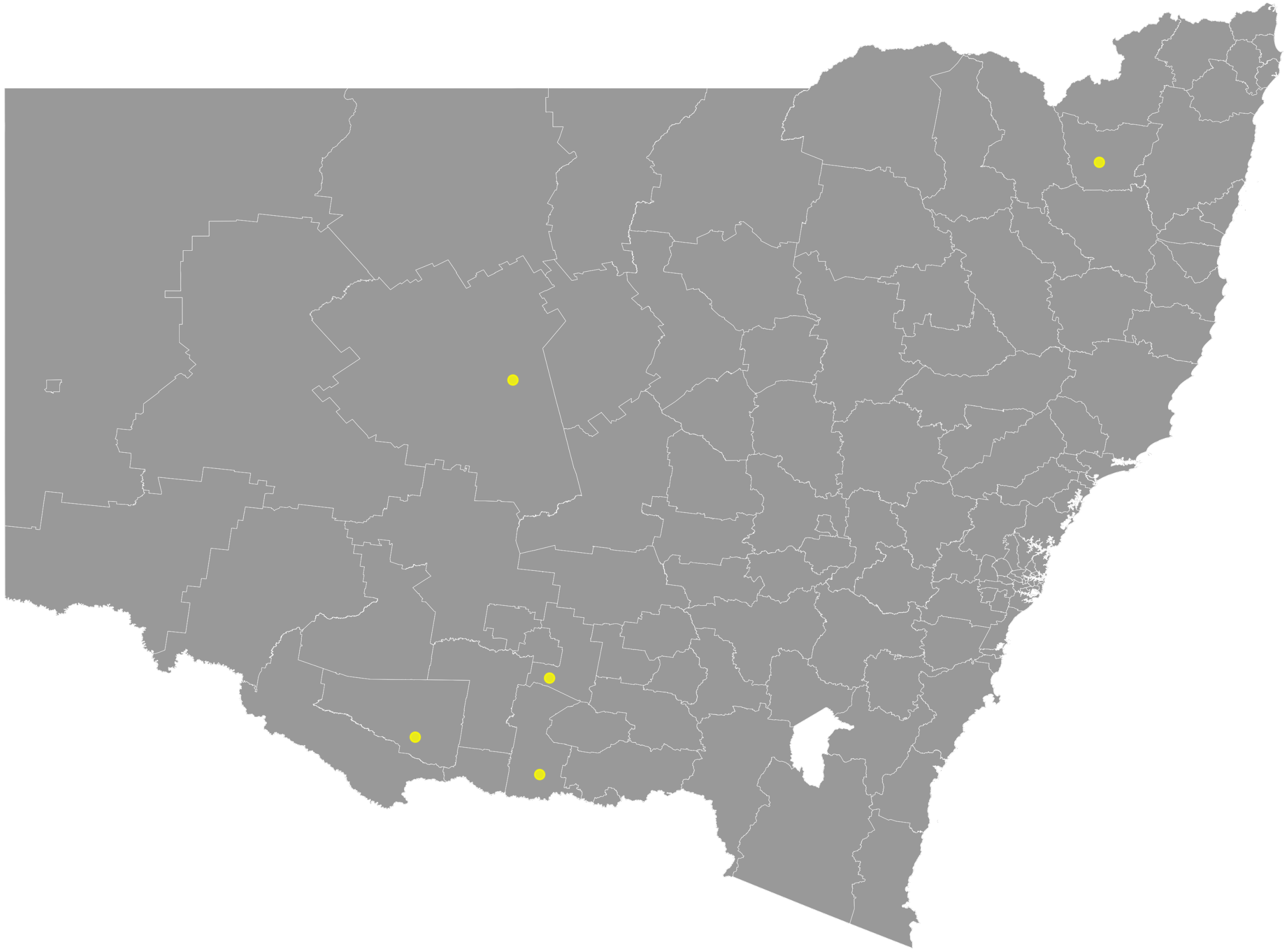
Which postcodes will grow more than 5%?

As measured in quarterly median prices

	Random Forest Classifier	MLP Classifier
Ave accuracy	**71%**	68%
Std deviation	0.02	0.07
Recall	67%	70%
Precision	64%	58%

Top 3 predictors

1. Previous quarter housing data
2. Rental market
3. Where recent immigrants settle



Which postcodes are the cheapest?

Amongst that are predicted to grow >5%

Postcode	Suburb	Price ($)
2835	Canbelego	195 K
2700	Bundure	230 K
2710	Booroorban	273 K
2370	Kingsland	275 K
2646	Nyora	298 K

Disclaimer

All contents discussed do not constitute investment advice