Property Analysis User Guide

Lei Huang

November 3, 2020

1 Introduction

This shiny app is designed for analysing sale properties. Property details are scarped from real estate website: `https://www.domain.com.au/`. In order to get prompt data size, we only limited sale apartments from four major cities, not included surrounding suburbs and other house types (Figure 1.1). In order to show what a data with API looks like, I pre-processed adelaide-demo. Several methods are applied to collect raw data. General property details are scraped by `rvest`. If there is a avaiable google API, Longitude and latitude of properties are collected from `ggmap`. There are video demos available, one is with API and the other is without.

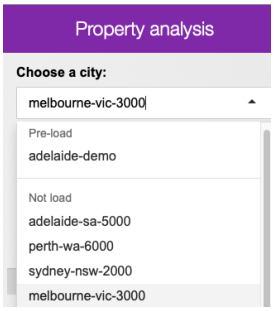

Figure 1.1: Select city

There are two aims in this analysis: 1) Make a good prediction on property price; 2) Find out is there any property descriptions are similar. For the first aim, the analysis could be found under "price prediction" (Figure 1.2). "Similarity analysis" part is designed for second aim (Figure 1.2).

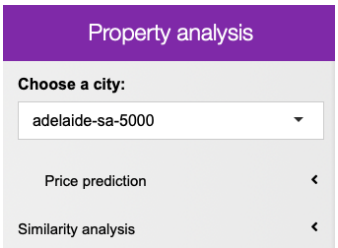

Figure 1.2: Main select pannel

2 Price prediction

The figure below is an overview of price prediction section (Figure 2.1). The select panel is on the left hand side (Figure 2.2).

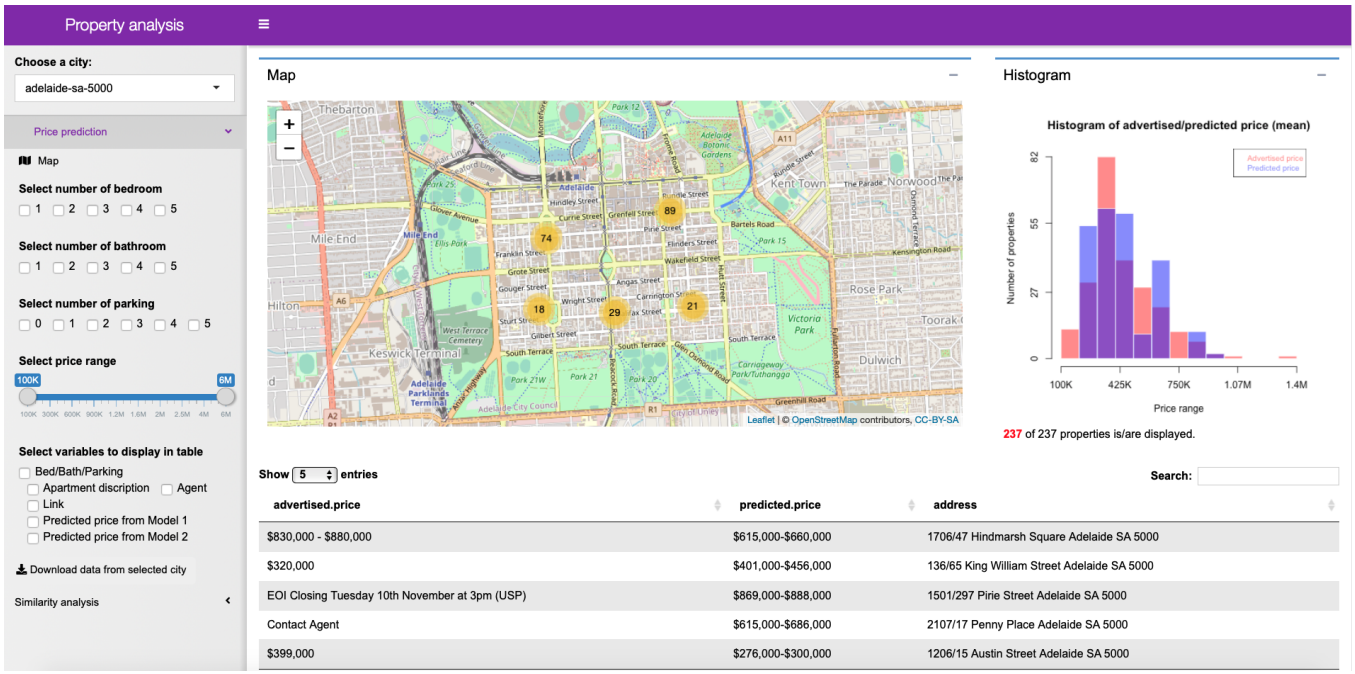

Figure 2.1: Overview of price prediction

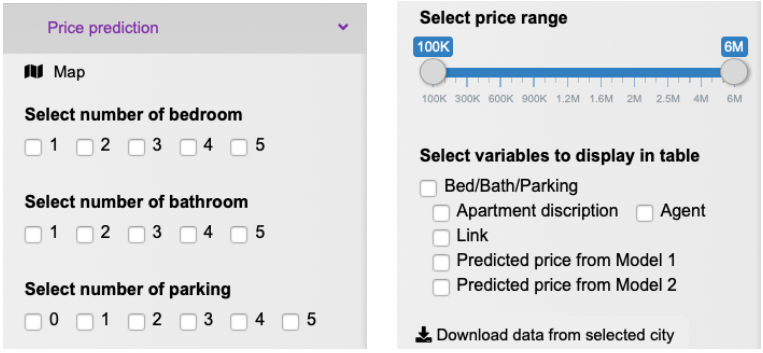

Figure 2.2: Select panel under price prediction

User can choose any number of bedrooms (from 1 bedroom to 5 bedrooms), bathrooms (from 1 bathroom to 5 bathrooms)and parking (from 0 parking to 5 parkings) to display. If no numbers are selected, all properties from selected city will be displayed. User could also check any properties in particular price range by dragging the price range sidebar. The checkbox at bottom of the select panel can control displaying information in the table. The default three columns are: advertised price (collected from webpage), predicted price (predicted price from model) and address. If user would like to have a look more information about property, user could tick any boxes in select panel. Predicted price is calculated from two models: Model one is linear model with number of Beds/baths/parking as key parameter. Model two is generalised additive model where Number of Beds/baths/parking and apartment level are considered as key

parameters in this price prediction model. Both models is well fitted with R^2 around 0.8. Users can choose to display predict price from both two models in table by tick boxes in select panel.

There are three subsections in output: Map, Histogram and Information table. All properties under select conditions will be displayed in the map if a google API has been set up. Click any pop-up, user can see address, advertised price and predicted price range which contains predict price from both models (Figure 2.3). If user would like to see extra property information such as photos, the hyperlink will bring user to the sale page on domain. if user does not set up a google API, this popup function is not available. This shiny app will automatically check whether user has successfully set up a google API. For set up google API, please read the following website: `https://cran.r-project.org/web/packages/ggmap/readme/README.html`.

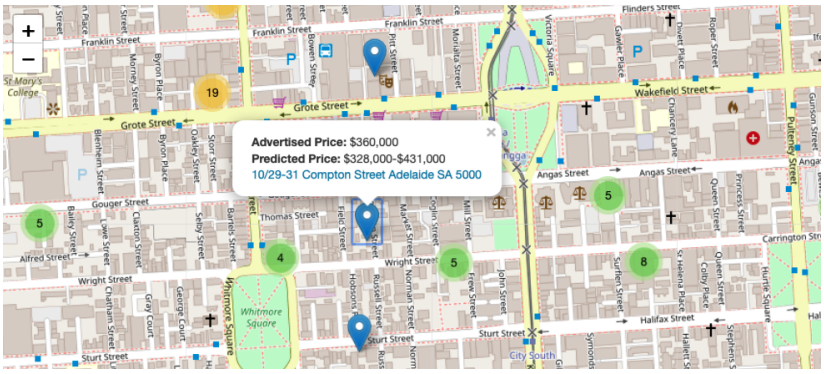

Figure 2.3: Map detail

Histogram (Figure 2.4) is used to show mean of advertised and mean of predicted price information from selected properties where advertised price is in pink and predicted price is in light purple. The overlapped part is in dark purple. Y-axis is the number of properties and X-axis is the price range where "K" stands for thousand and "M" means million. Total number of selected properties is listed at bottom of histogram.

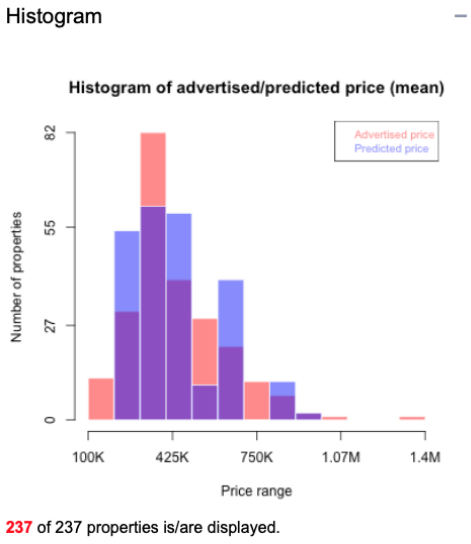

Figure 2.4: Histogram detail

User can tick more box to show more information in the table. The default number of rows to display is 5. User could choose to show more rows from top left of table and search any properties with keyword. Besides, if user only interested in the table, user could fold map and histogram by click "-" button on top right of each box, the table will take whole page to display (Figure 2.5).

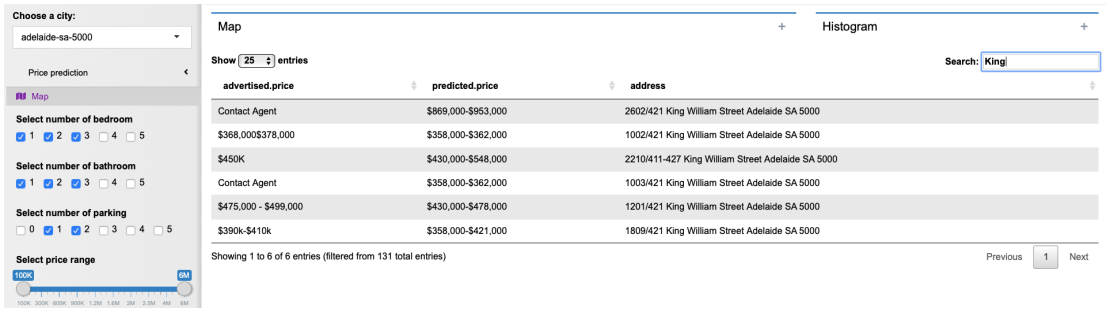

Figure 2.5: Fold map and histogram

3 Similarity analysis

This part is analysing the similarity of property descriptions. Words are grouped into one word, two words and three words. Under each word panel, user could select number of bedrooms from 1 to 5 bedrooms (Figure 3.1). If none of bedroom is selected, all properties will be used. Choosing number of bathroom and parking is not allowed, because some of bed/bath/parking combinations contain too few properties for similarity analysis. If there is no available data for given selected bedrooms, nothing will be displayed.

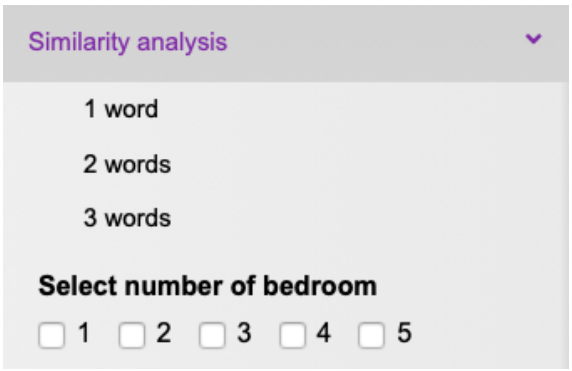

Figure 3.1: Main panel on similarity analysis

Under each word panel, there are five main parts: a wordcloud with top 100 words, a barplot of top 10 capitalised word and a barplot of top 10 common word (Figure 3.2). They can be folded by click "-" button on top right of each box.

The next part is a cluster dendrogram which shows the similarities among descriptions from selected properties. Properties are grouped by description similarity. Only properties with high

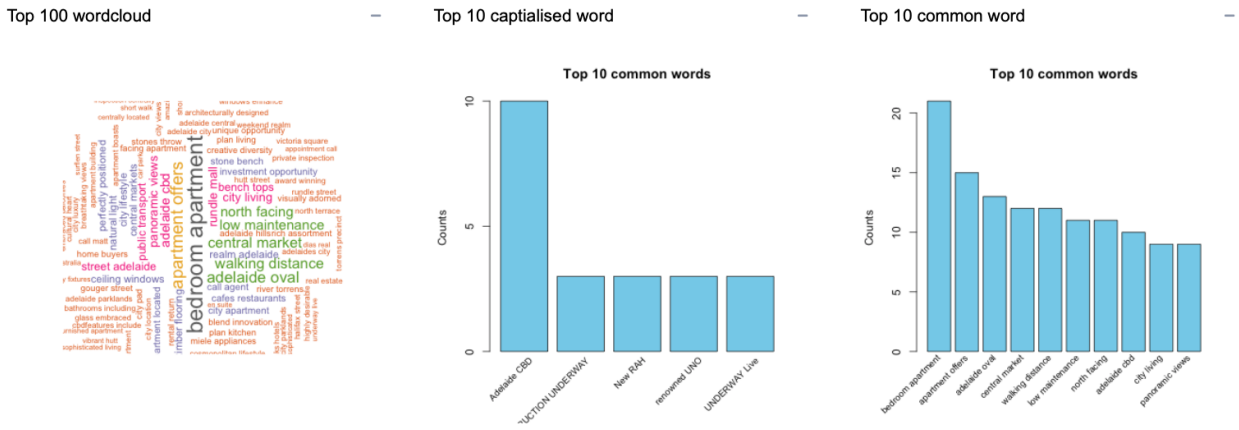

Figure 3.2: Wordcloud and top 10

level of description similarity are named in the cluster plot (Figure 3.3). A red dash line indicates the similarity cutoff where height(h)) is at the lowest 5%. This part can be folded as well.

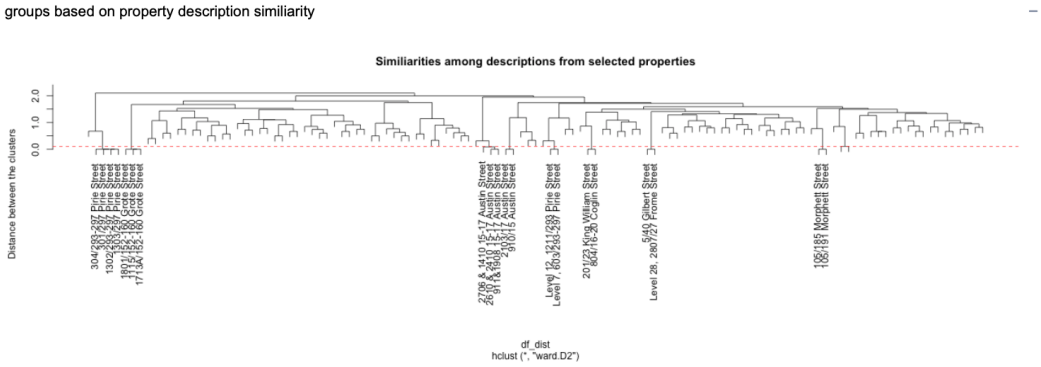

Figure 3.3: Similarity

There are four columns in table: address, description, agent and group. Group indicates which group of the given property is in. If these named properties in same group which come from same building and same agent, it will not be considered to have potential copying problem. Therefore, the details of these properties will not be listed. Only properties in same group from different agent will be listed (Figure 3.4).

98	304/293-297 Pirie Street Adelaide SA 5000	A reflection of the areas creative diversity, the East End apartments are visually adorned with a rich assortment of materials and forms that blend innovation with the heritage and culture of the east end. Glass-embraced balconies and floor-to-ceiling windows enhance the panoramic views of Adelaides city, parklands and hills. High end finishes such as timber flooring and stone bench tops work to	Korn Real Estate	group-3
125	201/23 King William Street Adelaide SA 5000	Set in the heart of the CBD and over looking the hustle and bustle of King William Street apartments like this at this price don't come along everyday.Wheather your looking for a good investment with great returns or somewhere to call home this one will certainly tick a lot of the boxes.Featuring two bedrooms, one bathroom and open plan kitchen and lounge with neutral decor.The location	Harcourts Tagni	group-4

Figure 3.4: Similarity