STAT 344 Project

(November 10, 2022)

Name: Jun Zheng **Student Number: 10827335**
Role:
Name: Hanxi Chen **Student Number: 96132642**
Role:
Name: Chenyue Qian **Student Number: 74299629**
Role:
Name: Tianyang Wang **Student Number: 50797802**
Role:

1 Introduction:

1.1 Objectives

Melbourne experienced a large real estate bubble from 2017 to 2018. During this period, Melbourne's house prices were more volatile compared to the stable period. Therefore, our first objective is that we want to obtain quantitative data on the price of properties during the real estate bubble. We will therefore estimate the average price of properties sold in Melbourne at the time. In addition, in the present (2022), many households have more than one car and the demand for parking spaces is growing. So our second goal is to get more specific data about the parking spaces of the houses sold during the fluctuation period. For example, we want to estimate how many homes sold during the volatility period have two or more parking spaces. To achieve this, we estimate the number of sold homes with different numbers of parking spaces by using parking spaces as a variable.

1.2 Background

Choosing a livable city as an investment in the future is a goal that everyone strives for. Melbourne, Australia is a suitable choice for many people in comparison to its climate and cultural background. As a famous metropolis in Australia, Melbourne has a good development and history, as well as good educational resources, and is the center of Australia's cultural, dining, economic and other diversity. [1]

Melbourne was one of the first cities in Australia to develop, so while it has a strong foundation, it is also attracting people from outside to join the city. The constant influx of new residents has led to an oversupply of houses for a certain period of time, which in turn has led to an increase in property prices, making many people concerned about the value of their houses while purchasing them for residential purposes. In the current social environment, cars have become the means of travel for almost every household, so the configuration of the house has naturally become a point of concern. Choosing a good location, having one or more parking spaces, and matching the price with the right price has become a key concern. Therefore, we have selected the price data and the number of parking spaces in each area of Melbourne to filter and analyze, so that we can meet the needs of our customers and let them see more intuitively the houses that match their needs and the price of the houses so that they can choose the most suitable house to buy. The housing bubble - refers to the rapid decline in property prices after excessive growth. After the growth of prices in 2016 and reach to the peak in 2017, Melbourne's house prices are facing rapid price reductions (We refer to this as "The Housing Bubble" in later context.), 'Compared with their 2017 peak, Sydney and Melbourne's house prices have now dropped 11.1% and 7.2% respectively,'[2]. Commonly, the rapid price reduction allows many people to purchase properties at low prices for residential or investment needs, therefore, analyzing the trend of the data can be a good way to avoid the risk of excessive devaluation while satisfying the residential needs of investors.

2 Data Collection and Summaries:

2.1 Data Description

Our dataset was found from Kaggle (Melbourne Housing Market), which includes information regarding all the traded houses in Melbourne during the housing price bubble from 2016 to beginning of 2018. These data was scraped by Tony Pino from the Domain, which is an Australian commercial real estate portal. In this case, we assume this dataset covers all the traded houses over the housing price bubble. All the variables related to our study along with their corresponding descriptions are included in the table below.

Variable Name	Description
Price	House Price in Australian dollars
Type	House Types: br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential
Rooms	Number of rooms in the house
Distance	Distance from CBD in kilometres
Bedroom2	Number of bathrooms in the house
Bathroom	Number of bathrooms in the house
Car	Number of car spots
Region name	Name of the General Region (West, North West, North, North east ... etc)
Satisfactory Rate	Binary variable created based on the variable Car, whereas 2 car spots denotes as satisfactory

2.2 Target Population

The target population is all the house in Melbourne that have been trade during the housing bubble. Since we have access to the whole dataset, we have information about the total number of population $N = 20424$.

2.3 Parameter of Interest

- Proportion of houses with two or more parking spaces during the housing bubble. In this case, we consider 2 or more car spots as satisfactory, so we give it the name satisfactory rate.
- The mean price of the houses traded during The Housing Bubble in AUD.

3 Sampling Method and Procedure:

For this study, we would use two sampling methods: Simple Random Sample and Stratified Random Sample. In this section, we would discuss the sampling procedures of these two samples, respectively, mainly focusing on how we set up the sample size, how we determine the strata and how we get the sample.

3.1 Simple Random Sample

3.1.1 Sample Size

In order to use R to draw the sample from data, we would need to decide the sample size n in advance. In this study, we are interested in two parameters from the population, one being binary and one being continous. Since we did not have reference to any prior study, we cannot make a good guess for the sample variance (s^2_{guess}). Therefore, we choose to get the sample size based on the binary one.

To make our estimate have a high accuracy, we demand a relative small width for the 95% confidence interval. Hence, we decide the width to be 10%. In this case, the half-width (δ) would be 5% and $z_{\alpha/2}$ is 1.96, and we use the conservative guess about the population proportion, i.e. setting $p_{guess} = 0.5$. Besides, the accessibility to the population provides us knowledge about the population size ($N = 20423$), which allows us to implement FPC here. With these information, formula and consequent calculation can be applied to get sample size n:

$$n_0 = \frac{(z_{\alpha/2})^2 \times p_{guess} \times (1 - p_{guess})}{\delta^2} = \frac{1.96^2 \times 0.5 \times (1 - 0.5)}{0.05^2} = 384.16$$

$$n = \frac{n_0}{1 + n_0/N} = \frac{384.16}{1 + 384.16/20423} = 377.067 \rightarrow n = 338$$

In this case, a total number of 338 houses is necessary for our sample in order to achieve a 95% confidence interval with width less or equal to 10%.

3.1.2 Sampling Process

Sample could now be drawn from population using R (detail codes provided in Appendix). Here, we give an overview of our sample by providing a histogram for the house price and a bar plot for the satisfactory rate.