NM2207 Data Story: Final Project

Marzuki Nooranas

2023-11-03

Week-9: Webpage + Diary Entry

Question 1:

(1) What is the topic that you have finalized? (Answer in 1 or 2 sentences)

Answer: I have finalised and chosen to explore the changes in the prices of resale flats in Singapore. The goal is to analyse the nominal price growth from 1990 to 2023, an compare it with the adjusted inflation rate to understand the real increase in flat prices over the years

Question 2:

(2) What are the data sources that you have curated so far? (Answer 1 or 2 sentences).

Answer: I have curated two datasets from https://beta.data.gov.sg/collections/189/view – Singapore's open data collection. The one data set contains transacted prices of resale flats from 1990 to 1999 and the other contains the resale prices from January 2017 onwards with an additional column "remaining_lease" for the lease duration.

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.2     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

##      month       town flat_type block      street_name storey_range
## 1  1990-01 ANG MO KIO    1 ROOM   309 ANG MO KIO AVE 1     10 TO 12
## 2  1990-01 ANG MO KIO    1 ROOM   309 ANG MO KIO AVE 1     04 TO 06
## 3  1990-01 ANG MO KIO    1 ROOM   309 ANG MO KIO AVE 1     10 TO 12
## 4  1990-01 ANG MO KIO    1 ROOM   309 ANG MO KIO AVE 1     07 TO 09
## 5  1990-01 ANG MO KIO    3 ROOM   216 ANG MO KIO AVE 1     04 TO 06
## 6  1990-01 ANG MO KIO    3 ROOM   211 ANG MO KIO AVE 3     01 TO 03
## 7  1990-01 ANG MO KIO    3 ROOM   202 ANG MO KIO AVE 3     07 TO 09
```

```
## 8  1990-01 ANG MO KIO    3 ROOM   235 ANG MO KIO AVE 3     10 TO 12
## 9  1990-01 ANG MO KIO    3 ROOM   235 ANG MO KIO AVE 3     04 TO 06
## 10 1990-01 ANG MO KIO    3 ROOM   232 ANG MO KIO AVE 3     01 TO 03
##    floor_area_sqm     flat_model lease_commence_date resale_price
## 1              31       IMPROVED                1977         9000
## 2              31       IMPROVED                1977         6000
## 3              31       IMPROVED                1977         8000
## 4              31       IMPROVED                1977         6000
## 5              73 NEW GENERATION                1976        47200
## 6              67 NEW GENERATION                1977        46000
## 7              67 NEW GENERATION                1977        42000
## 8              67 NEW GENERATION                1977        38000
## 9              67 NEW GENERATION                1977        40000
## 10             67 NEW GENERATION                1977        47000

##      month       town flat_type block       street_name storey_range
## 1  2017-01 ANG MO KIO    2 ROOM   406 ANG MO KIO AVE 10     10 TO 12
## 2  2017-01 ANG MO KIO    3 ROOM   108  ANG MO KIO AVE 4     01 TO 03
## 3  2017-01 ANG MO KIO    3 ROOM   602  ANG MO KIO AVE 5     01 TO 03
## 4  2017-01 ANG MO KIO    3 ROOM   465 ANG MO KIO AVE 10     04 TO 06
## 5  2017-01 ANG MO KIO    3 ROOM   601  ANG MO KIO AVE 5     01 TO 03
## 6  2017-01 ANG MO KIO    3 ROOM   150  ANG MO KIO AVE 5     01 TO 03
## 7  2017-01 ANG MO KIO    3 ROOM   447 ANG MO KIO AVE 10     04 TO 06
## 8  2017-01 ANG MO KIO    3 ROOM   218  ANG MO KIO AVE 1     04 TO 06
## 9  2017-01 ANG MO KIO    3 ROOM   447 ANG MO KIO AVE 10     04 TO 06
## 10 2017-01 ANG MO KIO    3 ROOM   571  ANG MO KIO AVE 3     01 TO 03
##    floor_area_sqm     flat_model lease_commence_date    remaining_lease
## 1              44       Improved                1979 61 years 04 months
## 2              67 New Generation                1978 60 years 07 months
## 3              67 New Generation                1980 62 years 05 months
## 4              68 New Generation                1980  62 years 01 month
## 5              67 New Generation                1980 62 years 05 months
## 6              68 New Generation                1981           63 years
## 7              68 New Generation                1979 61 years 06 months
## 8              67 New Generation                1976 58 years 04 months
## 9              68 New Generation                1979 61 years 06 months
## 10             67 New Generation                1979 61 years 04 months
##    resale_price
## 1        232000
## 2        250000
## 3        262000
## 4        265000
## 5        265000
## 6        275000
## 7        280000
## 8        285000
## 9        285000
## 10       285000
```

Week 10: Diary entry

An update from Week 9:

Due to the massive volume of data from the original data set, I have narrowed down to two years specifically - 1990 and 2023. The 33 year gap provides a unique opportunity to compare long-term trends and the evolution of the housing market. The selection can potentially provide valuable insights in to the factors that have influence the property market over the years. For example, the estate or the size (in square meters). After which, I will provide an analysis of a price for the next 5-10-15 years using a projected approach based on historical data gathered and the expected inflation rates. This could offer a realistic forecast, taking to account the likely impact on future property prices, helping to paint a picture of future market condition.

Question 1:

(1) What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story)

Theme: Investigating Singapore's Resale Flat Prices, from 1990 to 2023

Background:

In Singapore, housing affordability has been an increasing worry for future Singaporeans. However, in 2012, the then Deputy Prime Minister (DPM) Tharman Shanmugaratnam - stating household income of SGD $1000 was sufficient to own a home, sparked discussions among netizen online (Tan, 2012). While Mr Tharman argued the housing grants could provide support for low-income household, many Singaporeans expressed concern about the long-term viability of such claims, especially in the face of the constant rising cost of living and property prices.

The Question: Is public housing in Singapore truly affordable for the average Singaporean in the future? How will the evolving housing market impact the affordability of homes for future generations of Singaporeans?

The above question will uncover the complexities of the housing market, analyse potential long-term trend, also seek to understand that while variables such as estate, size or remaining lease play significant roles, this analysis aims to identity the most influential factors.

References:

(1) Tan, J. (2012, March 2). Online uproar over DPM Tharman's '$1k flat' statement. Yahoo Singapore. https://sg.news.yahoo.com/blogs/singaporescene/online-uproar-over-dpm-tharman-1k-flat-statement-132522414.html

Question 2:

(2) Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., "According to the United Nations…" to justify why the question you have chosen is important)

First, according to an article by Liew (2023), the resale prices from Housing Development Board saw a constant increase over 14 consecutive quarters since mid-2020. While the increase is has slowed down - 1.5% from 2nd Quarter in 2023 to 1.2% in 3rd quarter - the same year highlights the growing concerns of housing affordability in the future. The report also mentioned the alarming increase in million-dollar transactions in mature estates - 105 transactions from 2nd Quarter in 2023 to 128 transactions in 3rd quarter (Tan, as cited by Liew, 2023).

Next, in another article by Liew (2023), the Housing Development Board (HDB), more new home owners are opting for a resale flat instead of the 'Build-To-Order' flats. The sales delayed from August 2023 to October 2023 and the wait time of 4-years contributed to the shift in preference towards resale flats. Therefore, the shift highlights the urgent housing needs, the prices Singaporeans are willing to pay for immediate availability shows the importance of this analysis.

Lastly, to evaluate the affordability for future Singaporeans, it is important to delve into the analysis on income used for housing, and the median housing prices over time. For instance, according to the Central Provident Fund Board (CPF), Singaporeans should allocate a maximum of 30% of income for housing - Mortgage Servicing Ratio (MSR). Shown below, according to Kumar (2022), in order to afford a 3-room average flat in Singapore, buyers should have a minimum combined salary of $5,200. This highlights the critical housing afforability for Singaporeans' financial sustainability for the long run.

The Salary We Need To Buy Our Dream Home

Housing Type	Median Resale Housing Price	Minimum Downpayment	Monthly Repayments	Average Salary Per Spouse
HDB 3-Room	$360,000	$54,000	$1,388	$2,615
HDB 4-Room	$554,000	$83,100	$2,136	$3,560
HDB 5-Room	$652,000	$97,800	$2,514	$4,190
HDB Executive	$695,000	$104,250	$2,680	$4,467

Estimation of Average Salary Per Spouse (Kumar, 2022)

References:

(1) Liew, I. (2023, October 2). HDB resale prices up 1.2% in Q3, lower than quarter average in 2022; more flats sold. The Straits Times. https://www.straitstimes.com/singapore/housing/hdb-resale-prices-up-12-in-q3-lower-than-quarter-average-in-2022-more-flats-sold#:~:text=SINGAPORE%20%E2%80%93%20Prices%20of%20Housing%20Board,from%20HDB%20showed%20on%20Monday.

(2) Liew, I. (2023, October 1). About 1,900 BTO flats in 2 Choa Chu Kang projects to be offered in early Oct launch. The Straits Times. https://www.straitstimes.com/singapore/housing/about-1900-bto-flats-in-2-choa-chu-kang-projects-to-be-offered-in-early-oct-launch

(3) Kumar, S. (2022, April 11). Here's The Salary You Need To Earn To Afford These Homes In Singapore [2022 Edition]. DollarsAndSense.sg. https://dollarsandsense.sg/salary-need-earn-afford-homes-singapore/

Question 3:

(3) Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use). Include the challenges and errors that you faced and how you overcame them.

The overall plan:

(1) To get a price estimation of the flats based on Room Type (3-Room, 4-Room, 5-Room, Executive) for both years.

(2) Use variables like the remaining lease, estates, size to see which plays a bigger role in setting the prices of Resale HDB Flats.

(3) Use median income guidelines from SingStats to determine affordability of housing in current market, followed by adjusing to inflation.

To answer part (1) and (2) of why the question is important, we will be looking at two files:

File 1: ResaleFlatPrices1999.csv

```
resaleprice1999 <- read.csv("ResaleFlatPrices1999.csv")
head(resaleprice1999, 3)
```

```
##     month       town flat_type block      street_name storey_range
## 1 1990-01 ANG MO KIO    3 ROOM   216 ANG MO KIO AVE 1     04 TO 06
## 2 1990-01 ANG MO KIO    3 ROOM   211 ANG MO KIO AVE 3     01 TO 03
## 3 1990-01 ANG MO KIO    3 ROOM   202 ANG MO KIO AVE 3     07 TO 09
##   floor_area_sqm     flat_model lease_commence_date resale_price
## 1             73 NEW GENERATION                1976        47200
## 2             67 NEW GENERATION                1977        46000
## 3             67 NEW GENERATION                1977        42000
```

```
names(resaleprice1999)
```

```
##  [1] "month"              "town"               "flat_type"
##  [4] "block"              "street_name"        "storey_range"
##  [7] "floor_area_sqm"     "flat_model"         "lease_commence_date"
## [10] "resale_price"
```

File 2: ResaleFlatPrices2022.csv

```
resaleprice2022 <- read.csv("ResaleFlatPrices2022.csv")
head(resaleprice2022,3)
```

```
##     month       town flat_type block      street_name storey_range
## 1 2022-01 ANG MO KIO    2 ROOM   323 ANG MO KIO AVE 3     07 TO 09
## 2 2022-01 ANG MO KIO    3 ROOM   320 ANG MO KIO AVE 1     07 TO 09
## 3 2022-01 ANG MO KIO    3 ROOM   225 ANG MO KIO AVE 1     07 TO 09
##   floor_area_sqm     flat_model lease_commence_date    remaining_lease
## 1             44       Improved                1977 54 years 05 months
## 2             73 New Generation                1977 54 years 05 months
## 3             67 New Generation                1978  55 years 01 month
##   resale_price
## 1       245000
## 2       358000
## 3       355000
```

```
names(resaleprice2022)
```

```
##  [1] "month"              "town"               "flat_type"
##  [4] "block"              "street_name"        "storey_range"
##  [7] "floor_area_sqm"     "flat_model"         "lease_commence_date"
## [10] "remaining_lease"    "resale_price"
```

Columns to be used:

(1) **'month' :** To determine the time frame, and be used to track the resale prices overtime. For example, in 1990 September and compared against 2022 September to give a more accurate representation.

(2) **'town':** Contains the different estates in Singapore, it is used to explore the geographical price variation and impact on prices. For example, if a similar 'flat_type', 'storey_range', but different 'town' and determine the more affordable versus expensive estate.

(3) **'flat_type':** Helps to categorise the estamate prices of the flats depending on room types. For example, "3-Room, 4-Room, 5-Room and Executive"

(4) **'storey_range':** Although not significant, it might still give insights into the price variations based on the floor level

(5) **'floor_area_sqm':** Might be able to determine how size of the flat impact resale value - despite being the same flat-type

(6) **'lease_commence_data':** The variable indicates the starting year of the lease, can potentially be used to calculate remaining lease for further comparision against resale price.

(7) **'resale price':** This will be analysed against other independent variables to understand the trends or pricing factors that may effect affordability.

Additional Column in resaleprice2022

(8) **'remaining_lease':** By calculating the remaining lease (the amount of time before the flat gets released back to the government), it determines how it could significantly impact resale prices

To answer part (3), another data set of household income with CPF contribution. This data set will be used in gauging the affordability of housing in the current market. By using the median income from SingStats, and comparing it against the resale prices of HDB in 2022, it could give a clearer perspective if housing affordability is achieved.

Challenges:

(1) **Volume of Original File:** As previously mentioned, due to the massive volume of data from the original data set, I have narrowed down to two years specifically - 1990 and 2023. Initially, there were over 300,000 rows of data on both files and I would be extremely tedious for comparision without pre-filtering. To solve this, I proceeded with further pre-filtering to concentrate on more common flat types (3-Room to Executive) and more typical flat levels (From Level 1 to 15). This data filtering can help to facilitate a more manageable and focused analysis so that the core question can be addressed more efficiently.

(2) **SingStat Untidy Data:** Currently, still trying to tidy the data.This data set contained titles, descriptions, and other information that could interfere with the accurate reading and interpretation of the dataset. What i have done is removing titles, and non essential information. By reformating the data to a more structured format, the raw data can help with analysis and provide a more accessible framework to draw insights and form conclusions.

(3) **Missing 'remaining_lease' on "ResaleFlatPrices1999.csv":** The data set lacked a crutial information which is viatal for a comparative analysis with the 2022 dataset. Therefore, to solve this, a new column is created by calculating the remaining lease through subtracting the 'lease_commence_data' from the year 1999, giving the remaining lease at the point in time.

(4) **Trouble finding the current resale flat ownership data set:** While there is potential data set from SingStats, I am still finalising the last data set that I require. This dataset can roughly gauge the ownership rates and trends overtime. This can provide a clearer picture of how the evolving housing market is impacting the ownership raters, and potentially the broader socio-economic landscape in Singapore.

References:

(1) Tan, J. (2012, March 2). Online uproar over DPM Tharman's '$1k flat' statement. Yahoo Singapore. https://sg.news.yahoo.com/blogs/singaporescene/online-uproar-over-dpm-tharman-1k-flat-statement-132522414.html

(2) Household income - latest data. (n.d.). Singstat. https://www.singstat.gov.sg/find-data/search-by-theme/households/household-income/latest-data

Week-11: Diary Entry

Update from Week 10: For week 11, I explored potential visualisation that can be used to provide a more clearer understanding. I also researched on other possible ways to make in interactive.

Question 1:

(1) List the visualizations that you are going to use in your project

For now, I am planning to create 4 analysis and visualisation. These visualisation will provide a clearer understanding of the prices of resale flats in Singapore. The 4 visualisations are:

(1) Trend of HDB Resale Prices in Singapore - Line Graph and Bar Graph: This analyses the trend of resale flat prices fron 1999 to 2022. A line plot would be created to show the average resale prices of that year, based on room types as well. The bar graph on the other hand, will be used for direct comparison between two years - they will be grouped base on the room types. For example, 1999 compared with 2022 for 3-room flats, and another one for 4-room flats and so on.

(2) The Distribution of Resale Flat Prices for different flat types - A Box Plot: A box plot is ideal to show the distribution of numerical variable - in this case the transacted amount of the resale flat - across the different flat types. This can also show if there are any outliers - to further understand the number of high prices as compared to the average ones.

(3) The Impact of Remaining Lease on Resale Prices - Scatter Plot: This helps to understand how the length of the remaining affects the resale prices. Therefore, the use of scatter plot, together with a regression line will be ideal. The scatter plot will depict the resale transactions, in this case, the x-axis will be the remaining lease, and the y-axis will be the resale price. The regression line will also visualise the overall trend, while quantifying the relationship between the resale prices, and remaining lease. The line will also highlight if there is a correlation of a shorter lease and a lower selling price, a major consideration for flat buyers.

(4) The influence of Estate on Resale Prices - Violin Plot: The violin plot will be able to visualise the distribution of resale flat prices accoss the estates and flat types. The plot will show mediam price and full distribution of prices - this allows comparison on central tendency but also the density of prices and variability. Also this can show how certain estates have wider variances in prices or if there are any similar price points where many flats are concentrated. The violin plot - although still considering - can be segmented further into flat types within the estates or be colour coded within one plot - this will be explored in Week 12/13.

Question 2:

(2) How do you plan to make it interactive?

In order to create a more interactive webpage, I am planning to use the capabilities of ‘ggplot2’ to create and render the initial plots and diagrams. After which, I will utilise ‘Shiny

UI' for the user interface and interactivity. Lastly, I will use R Markdown and R file to integrate the story and present it in one cohesive webpage with sections. For each visualisation, here is how i am planning to do so:

'ggplot2' will be used to create the initial foundation plots for line and bar graphs, box plot, scatter plot and violin plot.

'Shiny UI' components like 'selectInput', 'sliderInput, 'radioButtons', 'hover' and 'coloour inout will be used to create interactivity on the graphs. Users will be able to interact with the dynamic capabilities of Shiny UI, giving them a more engaging, user-driven exploration of the data.

'R Markdown': The integration of 'ggplot2' and 'Shiny' will be shown through the Quarto Document. Therefore, serving as a narrative structure guiding users through the data story of resale flats, providing context and explaining the potential insights derived from the visualisation.

Question 3:

(3) What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Create a table)

Weeks	Topics
Week 1	Creating and uploading photos on RMarkdown and Quarto
Week 2	Understanding ggplot2 and Shiny
Week 3	Variables and Data Structures to explore the data-sets
Week 4	Making new variables, choosing columns and rows for clearer visualisation
Week 5	-
Week 6	-
Week 7	Making integration of Shiny Dashboard with Quarto, Visualising data using ggplot and Shiny
Week 8	Visualising data using ggplot and Shiny, Creating an R app for visualisation
Week 9	-
Week 10	Storytelling
Week 11	-
Self-Thought	Still exploring possible topics

Question 4:

(4) Challenges faced and how I go about it?

1. Currently facing difficulties when trying to deploy interactive Shiny Application into the main quarto page. May need to look into past challenges and code along to refresh.

2. As of now, the placeholders on the webpage is still in static mode - looks very plain and dull. Will explore the other decorative options in the coming days. For now, focusing on creating the plots for data sets as well.

3. Data cleaning for household income is surprisingly tedious as well. Looking to streamline the datasets further to ensure easy readability when creating plots to measure affordability of the prices of HDB Resale flats.

Week-12: Diary Entry

Week-13: Diary Entry