Final Writeup

Niki

2023-11-24

Final Writeup In order to offer thorough insights into market trends, distribution, and variables impacting rental pricing across various planning areas, my data story analyses Singapore's HDB rental prices. For individuals looking for reasonably priced home options in Singapore, it is vital to comprehend HDB rental prices. Understanding the market dynamics is essential for making well-informed decisions, since the demand for rental properties rises along with the rental market.

A source on Singapore's rental market for 2023–2024 claims that demand for rental homes is rising due to an expanding renter market and rising interest rates (Jacob, 2023). This emphasises how crucial it is to investigate rental costs in order to assist prospective tenants in effectively navigating the market.

A strong foundation for answering the question is laid by the carefully chosen data sources, which include commuting time data, rental data, and planning area polygons. Geographic coordinates are provided by the planning area polygons, which makes it possible to create an interactive map that shows rental costs in various locations. A thorough examination of average costs based on flat types and planning areas is made possible by the rental data that covers a period of 3 years. Aside from that, the data on commute times also sheds light on the connection between the duration of travel to work and rental costs.

Many important insights are revealed by the data. First, as the size of HDB apartments increases, rental prices plateau. According to the data, renting a five-room apartment can be more advantageous than a four-room apartment when space is taken into account. This is because the hike in mean prices between four-room and five-room flats is not steep. However, it seems that the spread of prices for five-room flats is the largest amongst all room types, hence such a deduction may not hold accurately for a large range of five-room flats. Furthermore, the interactive map shows the average rental prices in planning regions; it highlights the most expensive locations, such as Bukit Timah and Bukit Merah, and the more reasonably priced districts, such as Yishun.

The most common HDB type throughout planning regions is the 4-room apartment. This is seen from the distribution of HDB kinds. For instance, despite the fact that 4-room flats are the most common kind in Bukit Timah, the area's reputation for wealth keeps the average rental price high. The trend analysis over three years indicates a consistent overall increase in rental prices, with notable fluctuations in specific areas like Bukit Timah and Central. Interestingly, there is not much of a correlation between rental prices and commute times as the plots are scattered throughout the graph, and there is no obvious trend. On the other hand, concentrating on Bukit Timah reveals intriguing trends that may point to a connection between commute times and real estate costs whereby the proportion of Bukit Timah commuters take a shorter time to travel to work.

There were difficulties and educational opportunities during implementation. It was first difficult to load the map into R, but I was able to find solutions by searching online, especially in YouTube and GitHub posts. Reading up on ways to effectively show data through an interactive chart led me to use Plotly and HighCharter - I thought these were great in allowing users to interact with the graphs. Another challenge I faced was coding the Shiny App. Initially, the map was intended to be a Shiny App to calculate the mean prices based on the HDB flat types. Unfortunately, I came to realise that I was not able to do it because one StackOverflow post mentions that mapview is no longer supported on Shiny. Hence, I had to find other ways to make the data story still flow, and thereby linking it with the count of flat types.

Lastly, understanding logic and resolving issues like case sensitivity in column names helped overcome challenges in data cleaning, such as merging rental data with map data. The project was intentionally elevated beyond static graphs by opting for visually appealing displays and interactive features through the use of shiny apps.

Reference

"How to Read in KML File Properly in R, or Separate out Lumped Variables into Columns." Stack Overflow, stackoverflow.com/questions/50775357/how-to-read-in-kml-file-properly-in-r-or-separate-out-lumped-variables-into-col. Accessed 24 Nov. 2023.

Jacob, Charmaine. "Singapore Rent Prices Are Still Sky-High. Long-Time Expats Are Having to Consider Drastic Changes." CNBC, 12 Apr. 2023, www.cnbc.com/2023/04/13/singapore-rent-prices-are-soaring-what-this-means-for-expats.html.

"Master Plan 2019 Planning Area Boundary (No Sea) — Data.gov.sg." Beta.data.gov.sg, beta.data.gov.sg/collections/2104/view. Accessed 24 Nov. 2023.

"Renting out of Flats 2023 — Data.gov.sg." Beta.data.gov.sg, beta.data.gov.sg/collections/166/view. Accessed 24 Nov. 2023.

"Resident Working Persons Aged 15 Years and Over." Https://Tablebuilder.singstat.gov.sg/Table/CT/8347, Department of Statistics.

Previous Diary entries

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

The topic I've selected to work on is to analyse HDB rental prices in Singapore.

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

For data visualisation, I intend to do an interactive hover map. I'll be using the API data set from: https://beta.data.gov.sg/collections/2104/view (Planning Area Polygon), to obtain the coordinates of the different towns in Singapore.

As for rental data, I will be using: https://beta.data.gov.sg/collections/166/view

And to further expand the analysis, I found a data set about the time taken for people in each planning area to go to work.

(3) 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),

Looking to rent a HDB flat in Singapore? Let's see what's the best for you!

(4) 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),

It'll be important to anyone looking for a more affordable place to stay in Singapore in comparison to condominiums or landed properties. Given that rental prices are increasing, it indirectly correlates to an increasing demand for rental properties.

Source: https://buycondo.sg/rental-market-in-singapore-2023-2024/#:~:text=The%20renter%20market%20is%20booming,wit

(5) 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. I will be looking at slicing the data into towns first - and then analyse the number of flat types (eg no of 1-room, 2-room etc), and also the average pricing per room type per estate. Thereafter, if there is indeed a trend that certain estates are more expensive than others, I may include another data set showing whether there is a correlation between transport nodes and rental prices based on estate.

The challenge so far is trying to load the map onto R. . . I have been Struggling, but thankful to the many resources on YouTube and other GitHub posts on how to do it! I have yet to clean this set of data, I'm looking to merge the rental data with the map data. . .

(6) List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?) a) A map plotting town coordinates and monthly rent

    b) Bar chart plotting room type and monthly rent

    c) Line graph plotting year and monthly rent (to show trend, pending data set)

    d) A map plotting transport nodes on top of the map with town coordinates

(7) How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive) a) A map where you can hover to get the bite sized information on average rental price (draft already on the website!)

    b) An animated gg-plot showing how rental price trends (pending data set)

    c) Interactive bar charts

(8) What concepts incorporated in your project were taught in the course and which ones were self-learnt? Concepts learnt from module: GG-Plot, Shiny, Cleaning data through manipulation and creating new data frames Self-learnt: Map making! And the interactive visualisation tools (highcharter and plotly)

(9) Include the challenges and errors that you faced and how you overcame them (if any) a) The first major challenge was the map. Initially I was keen on using the API data from onemapsg, but for some reason I could not manage to load the GeoJSON file onto R. After a while I realised that I needed to manipulate the data but I didn't know how :". . .

Previously I loaded a KML file and it worked! But discarded it because the coordinates were on SMCs and GRC districts, so the geography was a bit off. . . Thankfully after scouring through the net, I found the actual KML file for the planning areas.

    b) The second challenge was figuring out different ways of displaying the data. Bar graphs always come to mind because they're so easy. . . But after much thought, I will use the remaining week to create shiny apps to make my project more interactive and visually appealing. Further, aside from ensuring that the graphs were plotted correctly, it was also a good learning process to code the graphs with the end user in mind - to make sure that it's actually intuitive. One chart that I think needs further improvement would be my last scatterplot graph - since the percentage y-axis may be misleading.

    c) Cleaning data had a few roadblocks. . . But after figuring out the logic, it went smoothly! For eg when leftjoining columns into an existing data set, at first it didn't load properly. . . Just because the one data set was in all caps, and the other was in small letters xD . . .