Project submission 4

Niki

2023-11-10

Topic changed!! So new 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 ...