Final Submission

Annette

2023-11-24

Final Submission

Q1: What is the topic you have finalised? My data story is about exploring how the rising costs of living in Singapore are affecting different income groups, and whether education is truly the one-stop solution for meritocracy. This topic is one that's close to my heart, driven by observations of my peers, friends, and family, as well as conversations I've witnessed on social media platforms. Witht he rising costs, many young working adults in Singapore, like myself, grapple with concerns about affording our lifestyles while pursuing personal goals like work-life balance and leisure. I want to use data to uncover insights beyond surface level and use data to show that new policies or new approaches may have to be explored to help people climb the social ladder.

Q2: What are the data sources you have curated so far? I have found data that is provided publicly by the government on https://www.singstat.gov.sg/ and https://beta.data.gov.sg/, about Singapore's CPI (Consumer Price Index) for different income groups with 2019 as the base year. I also got data from singstat websites (from studies by MOE) for the proportion of youths enrolled in tertiary education each year, where the total number of youths is the theoretical age group for that level of education.

Q3: Why is this an important question?

This question addresses the fundamental issue of financial stability for individuals and families in Singapore. It shows the possible need for more opportunities and "paths" to be presented to people to climb the social ladder, or a change in direction. It also has implications for policymakers, as it highlights potential inequalities in income distribution and the need for policies to address these disparities.

Q4: Which rows and columns of the dataset will be used to answer these questions?

1. Consumer Price Index, 2019 as base year
   Row: "all items"
   Columns: all columns 2022-1993
2. Consumer Price Index by Income Group, Lowest 20%, 2019 as base year
   Row: "all items"
   Columns: all columns 2022-1993
3. Consumer Price Index by Income Group, Middle 60%, 2019 as base year
   Row: "all items"
   Columns: all columns 2022-1993
4. Consumer Price Index by Income Group, Highest 20%, 2019 as base year
   Row: "all items"
   Columns: all columns 2022-1993

5. Combined And Gross Enrolment Ratio For Primary, Secondary & Tertiary Education

   Row: "Gross Enrolment Ratio - Tertiary Education (Total)"

   Column: all columns 2020-1990 https://tablebuilder.singstat.gov.sg/table/TS/M850681

Q5: 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?)

1. Line graph- plotting the trends of different income groups' CPI

   This can help me see the trends of increase in different income groups' CPI, and identify income groups that are most vulnerable to rising costs. After ploting the graph, I uncovered unexpected results taht are explained in the webpage.

2. Stacked bar chart- proportion of youth enrolled in tertiary education over the years

   This visualization will show the proportion of youth enrolling in tertiary education, which I think is a good indication of how involved and committed youths are to getting a 'better' education. I want to show that education along cannot help most people to climb the social ladder significantly, as the increase in proportion of enrollment should mean a closing of gap in income disparities but that is not what is observed in Singapore (rising gini coefficient etc).

Q6: How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

For my first plot, I plan to use Shiny's input widgets like: 1. selectInput() - this lets the user choose the time period that they want to view at once, and they can see the rate of increase in CPI over specific portions of the graph 2. checkboxGroupInput() - this lets users select the line graphs they want to see at once. This is useful im letting them explore how the different income groups' CPI fare against each other, and can make comparison easier. 3. sidebarPanel() and sidebarLayout - these allowed me to have the side bar to make my visualisation interactive and let teh user choose their graphs and time periods to look at. It was inspired from one of the data story examples made by shiny.

For both visualisations, I also used "Plotly" to help me make my visualisations more interactive and aesthetic. For example in both graph I wanted to let the user hover over a certain parts of the graph and seeing a snapshot of data at that specific timestamp, so I used Hoverext() and tooltip.

Q7: What concepts incorporated in your project were taught in the course and which ones were self-learnt?

Topics	Week
Data Manipulation	3
Tidying Data	4
Functions	5
Data visualisation	8
Shiny apps	9
Plotly	
sidebarLayout	
sidebarInput	
observeEvent	
if()	
rbind()	

I faced a lot of challenges in conceptualising my data and what I wanted to use it to answer my questions. In the previous weeks, I had a rough idea of my data and I thought that the data sets I had were sufficient, however only in week 10 when I was starting to work on my visualisation I realised that I needed some

slightly different data, or that I needed to use different forms of visualisation. Thus although a bit late, I had to find new datasets and created new shiny apps that were more aligned with what I wanted to do.

I also faced a lot of difficulty in creating my shiny apps, as I had a vision in mind of how I wanted it to be but I had a lot of difficulty executing them. For my first plot, I had trouble in creating the logic for when the user selected different lines to compare using the if() function, but after experimenting and searching the internet it solved it.

For both plots, I found plotly interesting as it had so many functions but since they were not taught in class I had a bit of difficulty integrating them into my codes. I overcame this challenge by searching the internet and luckily enough, my visualisations were not too complicated so I was able to rectify them.

References:

1. SINGAPORE DEPARTMENT OF STATISTICS. (2023b, May 23). Consumer Price Index (CPI), 2019 As Base Year. Tablebuilder.singstat.gov.sg. https://tablebuilder.singstat.gov.sg/table/TS/M212931
2. SINGAPORE DEPARTMENT OF STATISTICS. (2023b, July 24). Consumer Price Index (CPI) By Household Income Group, Highest 20%, 2019 As Base Year. Tablebuilder.singstat.gov.sg. https://tablebuilder.singstat.gov.sg/table/TS/M213081
3. SINGAPORE DEPARTMENT OF STATISTICS. (2023, February 23). Consumer Price Index (CPI) By Household Income Group, Middle 60%, 2019 As Base Year. Tablebuilder.singstat.gov.sg. https://tablebuilder.singstat.gov.sg/table/TS/M213041
4. SINGAPORE DEPARTMENT OF STATISTICS. (2023b, July 24). Consumer Price Index (CPI) By Household Income Group, Lowest 20%, 2019 As Base Year. Tablebuilder.singstat.gov.sg. https://tablebuilder.singstat.gov.sg/table/TS/M213061
5. MINISTRY OF EDUCATION. (2020, October 4). Combined And Gross Enrolment Ratio For Primary, Secondary & Tertiary Education. Tablebuilder.singstat.gov.sg. https://tablebuilder.singstat.gov.sg/table/TS/M850681