Data Translation Challenge

6/9/2023

92/100 Points

Attempt 1 ⌄	◯	**Review Feedback** **6/9/2023**	Attempt 1 Score: **92/100**	💬 View Feedback

Unlimited Attempts Allowed

⌄ **Details**

Auto Module Navigation Red

Business Problem

It's currently April, 2022. You work for a company in the retail sector. Your company knows how well they're weathering the pandemic, but they are having difficulty figuring out what the effect has been on the rest of the retail sector and the rest of the economy. Everyone is being tight with their numbers!

So, they've brought you on and pointed you towards the **IPUMS release** ⤷ **(https://cps.ipums.org/cps/)** of the **Current Population Survey** ⤷ **(https://www.census.gov/programs-surveys/cps.html)** , which the government often uses to understand changes in employment.

Why employment data? **They figure that revenues can be misrepresented, or might be affected by government aid. But employment in an industry can tell you a lot about how well that industry is doing!**

They don't know exactly what analyses they want you to run or variables they want you to use - figuring that out is your job. But they know the general questions they'd like to be answered, each of which will probably require more than one regression analysis.

1. How has COVID affected the health of the retail industry, as measured by employment?
2. How has retail fared relative to other industries?
3. Retail needs to worry about who has money to spend - what has changed about *who* is working and earning money?

Analysis

1. Explore the data on the IPUMS CPS website and see what is in there. (Note: perform this analysis as though you were asked to do it in April 2022. Don't use data more recent than April 2022 - it just gets big and unwiedly).
2. Figure out which analyses to run that will help answer these questions
3. Run those analyses
4. Write up your results in an RMarkdown document with your analysis code inside, and prepare a presentation of your results (RMarkdown works here too, or Google Slides or Powerpoint or whatever you like)
5. Your entire project should be kept track of as an R Project (.Rproj), in a folder with files kept in a standard file structure (code/, data/, results/, etc), and uploaded to a GitHub repository

You may bring in outside information aside from your data (for example, "lockdowns strengthened in month X, so our analysis will…") but this is not expected/required.

Communication

In the Writeup

Your writeup should be in an RMarkdown document. Don't worry about length - if you're hitting all the points below for each of your analyses, that is what you want. Format the writeup as a report with sections for the main questions, not a list of bullet points. Information to include:

1. Why you are running the analyses you are running
2. How the analyses answer the question being asked, and what the result is
3. Carefully interpreting the results
4. Presenting the results in an appealing way. Graphs are great, `sumtable()` is great, `etable()` is great - put a little effort into formatting tables and figures to make them look nice! At the very least, variable names should be in English rather than statistics-package ('Education' not 'EDUC'). If you

aren't comfortable enough with `ggplot` to make its visualizations look nice, feel free to make graphics in Excel or Tableau or anything you like, and include them in your RMarkdown doc as images. Econometric analyses should be in R.

5. Acknowledging the assumptions you are making in each analysis, how plausible those assumptions are in the context of your data, and any evidence you can provide backing up those assumptions

6. After doing all analyses related to a given question, provide a generalized answer to the main questions.

I do not expect undisputable flawless results - the data can only do so much, and we always have to rely on assumptions. However, an analysis with big flaws goes down a lot easier if you can very accurately interpret the results, point out the flaws or implausible assumptions, discuss how those flaws affect the results, and perhaps suggest an improved analysis you would run if it were feasible. **Don't claim more than your results can actually show. I'd rather you accurately discuss where your results are flawed, than try to sell them to me as perfect.**

You can work collaboriatively on a Markdown file using GitHub, which is recommended. If you prefer, you can instead use Google Docs with the **Markdown Preview** ⤴ **(https://gsuite.google.com/marketplace/app/markdown_preview/410253451848)** add-in, or you can use **Draft** ⤴ **(https://draftin.com/)** . These will not run the R code in the document, but they will let you work together on text and code.

In the Presentation

Your group presentation should be between 15 and 20 minutes long, plus a few minutes for questions. As many or as few of the people in your group can present as you like. There are no bonus points for being a presenter. 20 minutes is a *maximum* length.

For your presentation:

1. Choose two of the main questions you feel you can answer best
2. Present your analyses related to those questions. Depending on how many you have you may want to select only a subset of them.
3. Make clear to the audience how your analysis works, how to interpret the results, and what the general answer to the question is
4. Be prepared to answer questions about your analyses and the assumptions behind them

---

RESOURCES:

For full instructions and links to resources, see **Data-Translation-Assignment.html (https://seattleu.instructure.com/courses/1608897/files/69203994/download?wrap=1)** ↓ **(https://seattleu.instructure.com/courses/1608897/files/69203994/download?download_frd=1)**

Industry code file: **indnames.csv (https://seattleu.instructure.com/courses/1608897/files/69453093/download?wrap=1)** ↓ **(https://seattleu.instructure.com/courses/1608897/files/69453093/download?download_frd=1)**

https://github.com/ankhanhquach/ECON-FINAL-PROJECT

You are unable to submit to this assignment as your enrollment in this course has been concluded.