Job Posting for Undergraduate Employment

Econ 145 Summer 2023

Application Due Date: 2023-05-08 at 5p.m. PDT.

The Economics department is hiring undergraduate students for Summer 2023 to help set up ***Econ 145 Data Wrangling for Economics in R***. Students hired for the summer will most likely be offered a position Fall quarter 2023. Econ 145 focuses on teaching students the fundamentals of cleaning and analyzing data. The ideal candidate will have:

1. Knowledge of ‘R‘. This means you can do the following:

- Create graphs using ‘ggplot2‘

  - Perform basic manipulations in R
  - Write loops and functions

2. Familiarity with `tidyverse` (specifically `dplyr`, `magrittr`, `stringr`),
3. Strong verbal skills in English.
4. Taken Econ 145 (**preferred**)

The economics department looking for multiple undergraduates. **The positions will be for 8 weeks working 20 hours per week for $17 per hour.**

All applicants are required to complete the prompts. In addition, there should be an additional attachment **MasterData.csv**. You will use this data to solve the Econ 145 prompt.

Please email your answers to econ-econ145@ucsb.edu with the subject line **Econ 145 TA**. In the email, please include:

1) Name

2) Class standings (e.g. freshman, sophomore, junior, senior)
3) Major/minor
4) Unofficial transcript
5) Coding experience (including classwork, internships, kaggle competitions, etc.)
6) Summer availability

Feel free to contact us with any questions on comments.

Prompt - Econ 145

You are working for a local retail firm. They have many, many clients and are interested in some basic information about them. The firm has supplied you with the excel sheet **Masterdata.csv**. They have also asked you to do some basic analysis on the dataset. You are asked to do the following:

Basic

1) How many missing values are there in each column? How many missing values are there total? Put your answers in a tibble and name it **missing**.

2) How many unique states does the company operate in? Are there any typos in this column? If so, correct the typos. Save the corrected dataframe as **data_cleaned**. Unless otherwise told, work with **data_cleaned**.

3) Find the average invoice and the percent of invoices paid. Save them as **avg_inv** and **avg_inv_paid**.

Advanced

4) Create a new dataframe where the "Names" column is split into 2 columns: *First Name* and *Last Name*. Nothing else needs to be changed. Save it as **data_cleaned_split**. For part 5, use **data_cleaned_split**.

5) Write a function named **client_status** that inputs *last name*, *first name* and outputs: i) state they live in, ii) number of invoices that need to be paid, iii) outstanding debt. Here, outstanding debt means the sum of the value of the invoices not yet paid. Have your code return an error if someone enters a name not in **data_cleaned_split**. Provide a line of code that runs your function for *Jordyn Kang*.

6) Create a new dataframe named *p6* where you:

i) Drop all rows that have at least one NA value
ii) Group by state
iii) Return the number of invoices per state as well as the average number of invoices paid

7) I wrote a function that compares the number of invoices between two states using *p6*. It is supposed to return to me which state has more invoices. It's not working and I don't know why. In 3 sentences or less, explain to me why my code won't work. Then fix the code so that it returns which state has more invoices[1].

```
comp <- function(state1,state2){
  if(p6$state1 > p6$state2){
    return(state1)
  }else{
    return(state2)
  }
}
#Running the code
comp("Arizona","Alaska")
```

8) Create a barplot that compares the average value of paid invoices to unpaid invoices to missing invoice information. Label the graph accordingly.

[1] If you have extra time, is there a way to run the function without putting quotes around "Arizona" and "Alaska"? What about extending to more than 2 states?

Technical Suggestions

- Please set your working directory to source file location and call the original dataset *data*.
- Please comment your code.
- Your code will be run using the original dataset you were provided. This means every change that is done to the data must be done within your script.
- If anything is confusing, please ask! You may email startz@ucsb.edu.