Research Assistant Programming Assessment 2021-22

As a research assistant at the Bank of Canada, you will be required to assist economists and researchers with various data-related work. This includes obtaining datasets, checking their accuracy and manipulating them to make it easier to perform statistical analysis. The following questions are examples of the kind of work that you will be expected to perform at the Bank.

Question 1:

For Question 1, you should show all your work and any external code used for part 4 in the same, clearly laid out, Excel document.

Part 1: Gathering Data

a) Using StatsCan Table 14-10-0287-01, download total population, total employment, and total unemployment for both sexes (combined) over the age of 15 in Newfoundland and Labrador, Alberta, Ontario, and Canada from January 1977 until December 2019 (inclusive). All variables should be seasonally adjusted.

Part 2: Transforming Data

b) Convert total employment in Ontario to a quarterly frequency by taking the average monthly values. Which quarter resulted in the largest quarter-over-quarter percentage increase (not annualized) in employment?
c) For each of the 3 provinces, calculate the monthly unemployment rate over the entire period.

Part 3: Interpreting Data

d) For the 3 unemployment rates series you calculated in part 2c, create 1 chart showing this data that could be used in a report or presentation.
e) Using the chart you just created, answer the following questions:
    i. What pattern do you see in Alberta starting in late 2014? Name 1 reason why this may have happened.
    ii. Does one of the provinces seem different from the other 2? Name 2 reasons why this might be.

Part 4: Computing Descriptive Statistics and Regressions

For the below question, you can use the statistical software of your choice (i.e. Excel, R, Stata, MATLAB, Python, etc.). **Please paste any external code that you use, as well as your answers, in your Excel workbook for marking.**

f) Download series found at the below links from 1965Q1-2019Q4
    i. Seasonally Adjusted Quarterly Unemployment Rate, Canada: https://fred.stlouisfed.org/series/LRUNTTTTCAQ156S
    ii. Seasonally Adjusted Quarterly Real (Chained 2012) Gross domestic product at market prices, Canada (v62305752): https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610010401

g) Compute the quarter-over-quarter growth[1] of GDP and the quarterly percentage point change in the unemployment rate.
h) What is the correlation between the two series calculated in part g)?
i) Perform an ordinary least squares regression (of the form $y_t = a*x_t + c$) of quarter-over-quarter real GDP growth (dependant variable or y_t) and the quarterly percentage point change in the unemployment rate (independent variable or x_t).
    i. What is this relationship usually called in economics?
    ii. Give a brief interpretation of the regression results.
    iii. Describe 2 ways the regression could be improved.

Question 2: For the following parts, use your preferred coding language (or pseudo-code) to answer the questions. Only use basic built in functions/packages. The goal of this is to assess each candidate's understanding of coding practices and ability to use logic to solve a problem. Present your code in a Word, *.pdf, or *.txt document.

a) The .csv file attached to this assignment has temperature data for Ottawa going back until 1890. Read the .csv into your code and create a structure (e.g. dataframe, list, matrix, etc) with the following variables:
    Year (LOCAL_YEAR)
    Month (LOCAL_MONTH)
    Day (LOCAL_DAY)
    Mean Temperature (MEAN_TEMPERATURE)
    Maximum Temperature (MAX_TEMPERATURE)
    Minimum Temperature (MIN_TEMPERATURE)
    Total Rain (TOTAL_RAIN)
    Total Snow (TOTAL_SNOW)

Year	Month	Day	Mean Temperature	Maximum Temperature	Minimum Temperature	Total Rain	Total Snow
1890	1	1					
…	…	…					
2020	12	31					

b) Create a code that sums the total amount of rain for all days in the dataset.
c) Create a code that counts the number of days that it snowed in the dataset.
d) Create a code that counts the number of days that it snowed in May, June, July, and August in the dataset.
e) Create a code that computes the annual values for the following variables and saves them in a new structure (e.g. dataframe, list, matrix, etc): Mean Temperature, Min Temperature, Max Temperature.

Year	Mean Temperature	Minimum Temperature	Maximum Temperature

[1] The formula is $100*(x_t/x_{t-1})-1$

1890			
…			
2020			

f) Create a code that will add the temperature range as a column to the structure in Part e)

Year	Mean Temperature	Minimum Temperature	Maximum Temperature	Temperature Range
1890 (example*)	0	-100	100	200
…				
2020				

*These numbers are for example only, your answer should be different from them

Bonus: Display the numerical values from Question 2 parts b), c), d) in a PDF or Word document.