# US, Mexico, and Canada Economic Trends — Case Study Rubric

**DS 4002 – Spring 2024 – Miranda Khoury**
**Due: TBD**
**Submission format: Upload link to github repo on Canvas**

**Individual Assignment**

**Why am I doing this?** This is an opportunity to develop your data science skills by analyzing time series data using two common analysis techniques, Granger causality and autoregressive integrated moving average (ARIMA). By working on this project, you will gain an understanding of how data science can give people insights into important real-life trends.

**What am I going to do?** The GitHub repository for this case study can be found at https://github.com/MirandaKhoury/DS4002_CS3. You will obtain six time series data, found in the DATA file in the GitHub repository. Then, you will use R to conduct exploratory data analysis. You may do whatever exploratory data analyses you wish, but it is recommended that you at least plot all time series so you can observe the trends of the times series. Next, you will implement Granger causality analysis in R to answer three research questions: 1) Can PPP be used to predict inflation or vice versa? 2) Can one country's inflation rate changes predict inflation rate changes in another country? 3) Can one country's PPP changes predict changes in another country's PPP? Finally, you will conduct ARIMA modeling in R to produce forecasts for at least three of the time series.

**Deliverables include:**

- A data dictionary of the data tables used for analysis
- At least three exploratory data analysis visualizations (e.g. an initial plot of one of the time series data)
- Answers to the three research questions, supported by data (e.g. "Yes, inflation rates from one country can predict inflation rates in another," supported by Granger causality p values)
- Forecasts for at least three time series, generated by ARIMA
- Well documented code
- Github repository – to provide resources like code and data and to present the results of your analysis. Should include all of the above deliverables

All of this will be submitted electronically via a link to a github repository.

**Tips for success:**

- Don't overthink it. You can spend hours trying to optimize ARIMA model parameters with little improvement to the forecasting model. Start with running an automatic ARIMA; if you're getting reasonable results and the parameters it selects for you make sense, stick with that. If it ain't broke, don't fix it!
- Keep your code and variable names organized. You will likely run many combinations of Granger causality tests – for instance, you might test whether Canada inflation predicts Mexico inflation

and also whether Mexico inflation predicts Canada inflation. Come up with a simple and *consistent* naming convention for variables so you know what is what.

- Talk to the instructors. ARIMA can be non-intuitive to understand, so ask for help if you're having trouble selecting parameters or if your forecast is not performing well. Make use of their expertise!

**How will I know I have Succeeded?** You will meet expectations for this case study when you follow the criteria in the rubric below.

| | |
|---|---|
| Formatting | ● Repository – a GitHub repository (submitted via link on Canvas) containing all materials<br>○ Submit a link to the repo<br>○ Everything is contained in the repo or linked to it if appropriate<br>○ Contents<br>■ README.md<br>● Link your references at the bottom of this document<br>■ DATA folder<br>● Put raw data tables here OR links to the webpages from which the data were acquired<br>■ RESULTS folder<br>● Put results visualizations here, e.g. forecast graphs, initial EDA visualizations, results of any hypothesis tests<br>■ SCRIPTS folder<br>● Put all code necessary to generate the results here, *cleaned and commented* |
| README.md | ● <u>Goal</u>: This file serves as an orientation to everyone who comes to your repository. It should enable them to get their bearings.<br>● Use markdown headers to divide content.<br>● Section 1: Brief overview<br>○ A 1 paragraph summary of what the case study entails and the key results. This does not have to be very detailed but should summarize the key points.<br>● Section 2: Software section<br>○ The coding languages you used for the project and the IDEs used.<br>○ The names of any libraries used.<br>○ The platform (e.g., Windows, Mac, or Linux) you used.<br>● Section 3: Documentation map<br>○ Outline showing the hierarchy of folders and files contained in the GitHub repo |

| | |
|---|---|
| | ● Section 4: Instructions for reproducing your results<br>○ Explicit step-by-step instructions to reproduce your results. Must be clear enough to make it possible for an interested user to reproduce your results without much difficulty.<br>● Section 5: References |
| RESULTS folder | ● This is where you will include most of the project deliverables, including:<br>o At least three exploratory data analysis visualizations (e.g. an initial plot of one of the time series data)<br>o Answers to the three research questions, supported by data (e.g. "Yes, inflation rates from one country can predict inflation rates in another," supported by Granger causality p values)<br>o Forecasts for at least three time series, generated by ARIMA |
| DATA folder | ● Include all the raw data here so others can reproduce your results<br>● Also include a data dictionary of the data tables used for analysis |
| SCRIPTS folder | ● Include all code necessary to produce the results<br>○ The code must be cleaned and commented so as to be easily readable by others |

Acknowledgements: Special thanks to professors Rasero and Anand for providing the rubric structure.