

*A structured abstract is not a single paragraph but rather several labeled sections. The template below reflects and expands on the [data science process](#). Please answer all questions below. Aim for 250-300 total words, not including section labels and the questions written in this template.*

**Name:** Candice Xie

**Anticipated graduation month/year:** May 2024

**Capstone Advisor:** Anny-Claude Joseph

**Draft Title:** The effects of Joining EU on Economic Growth

**GitHub link:** [https://github.com/xzhcandice/EU\\_effects.git](https://github.com/xzhcandice/EU_effects.git)

**State an interesting question that can be answered with data.**

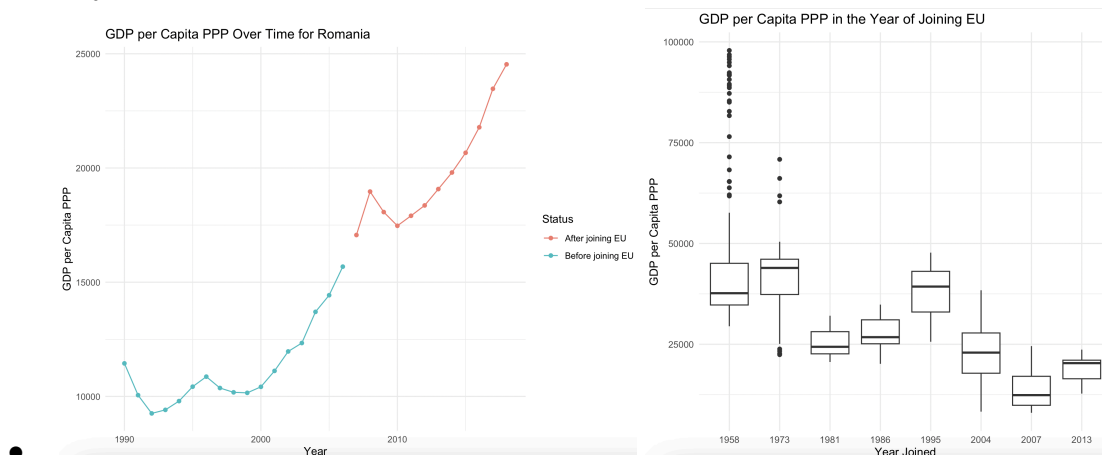
*How strongly does joining EU impact or correlate with a country's economic growth?*

**Get the data.** At the time when you submit this abstract, you should have a data set.

- The data source is the World Development Index from 1960 to 2018.
- It is the primary World Bank collection of development indicators, compiled from officially recognized international sources.
- Related variables: Year, country, GDP, FDIInflow, TourismReceipts, GDP\_PC\_PPP, ODAAndAid\_Received, Gini index, and unemployment rates

**Explore and Visualize the data.**

- I need to clean the countries to make a yearjoined and inEU variable to estimate the time-related effects and differentiate between those inEU countries and nonEU countries.
- There is a lot of missing data for some countries. I will remove rows with >50% of missing predictors for nonEU countries.
- There is no data for the year 2019, so I will remove it. Also some countries joined EU before 1960, and I do not have data for years before 1960, so I will need to find another way to deal with those countries.



**Model the data.**

- Model selection: I will use the Synthetic Control Method to construct a counterfactual for each EU country that joined the EU at different times. I will select predictors (GDP, GDP per capita PPP, FDI inflow, etc.) and control units (countries not in the EU or those that joined at different times).
- Fitting: For each treated unit (country joining the EU), I will use pre-treatment periods to fit the synthetic control, optimizing weights assigned to control units to best match the treated unit's pre-treatment trajectory. Then I will compare the post-treatment outcomes of treated units with their synthetic controls and assessing the significance of the observed treatment effects.
- Multiple Models: I will also compare the SCM results with traditional difference-in-differences models and fixed effects panel regression models to check for consistency.
- Visualization: I will plot the GDP growth trajectories of treated countries against their synthetic controls to visually assess the impact of EU membership.
- Create visual comparisons between actual post-treatment outcomes and those predicted by synthetic controls and other models.

**Summarize the results so far.**

- The data may suggest a positive effect of joining the EU on economic growth, FDI inflow, and possibly on tourism, depending on the country and period analyzed. However, the effects on aid received and the Gini index might be less straightforward, requiring deeper analysis.
- Limitations: The lack of data for certain countries or time periods may limit my analysis.
- Further questions: Are there differential effects of EU membership based on the economic structure or region of the member countries? How do specific policies tied to EU membership contribute to observed economic outcomes?

**Data ethics.**

- Bias: Since I might exclude certain countries or time periods, this may cause bias in the representativeness of the data.
- Data Collection: Since the data is collected by the World Bank, it adheres to GDPR and relevant data protection regulations.

**Is this capstone a continuation of a class project?** In the original class project, we conducted a preliminary analysis on the economic impact of joining the European Union (EU), utilizing the basic econometric models: simple/multiple Regression model and the fixed effect model. Data was sourced primarily from the World Development Indicators for the period from 2000 to 2015. However, this capstone project extends the analysis from 1960 to the present and employs more sophisticated econometric techniques, including the Synthetic Control Method (SCM). Beyond GDP and FDI, the capstone project will incorporate additional economic and social indicators, such as the Gini index and unemployment rates to provide a more holistic view of the impact of EU membership.