

1. Challenges Faced

a) What difficulties did you encounter while searching for and selecting a dataset?

- Navigating the world bank databases and filtering through multiple datasets to find 3 suitable ones was time-consuming.
- Some datasets had too broad or too narrow scopes, making them difficult to use.
- Picking the series that might work best with the indicators chosen in other datasets was difficult.
- Understanding series within dataset for research question formation.
- Many datasets had missing values or incomplete records. Some datasets included only a few countries, so we had to find datasets that had consistent data.

b) Were there any obstacles in understanding the structure and variables of the dataset?

- We spent time researching the meaning of certain indicators as some datasets lacked clear documentation or metadata, making it difficult to interpret variables.
- Some variable names were abbreviated or unclear, requiring additional research to understand their meaning.

2. Difficulties in Formulating Research Questions

a) What challenges arose in defining relevant and researchable questions based on your dataset?

- Needed to frame questions carefully to avoid misleading interpretations.
- Framing research questions that were relevant to the 3 datasets combined were challenging.
- Indicators were measured differently in different countries, so we had to spend time writing questions that wouldn't give us anomalies.
- There was missing data for certain years for certain countries so framing questions that wouldn't introduce bias was challenging.

b) Did you struggle with aligning the dataset to meaningful analytical methods?

- Determining if regression, classification, clustering, or time series analysis was appropriate for the datasets chosen was challenging as missing values and inconsistent time periods limited the application of predictive models.

- Refining research questions to ensure it might not lead to multicollinearity or misleading conclusions was a little tough because we haven't worked on the data yet.

3. Strategies Used to Overcome Obstacles

a) What steps did you take to address these challenges?

- We worked on understanding the data by doing research on the series involved.
- We applied filters like country, year to ensure uniform data and consistency.
- We selected datasets with minimal missing values.

b) Did you seek external resources, peer discussions, or additional research?

- We had a lot of meetings with our teammates to discuss the challenges.
- We also discussed with other teams for peer discussions.
- Utilized metadata and documentation provided by Worldbank.org to better understand the available indicators.
- We did additional research from watching YouTube tutorials, Google and ChatGPT to answer our questions and doubts.

4. Reflections and Lessons Learned

a) What would you do differently if you were to start this process again?

- We would start with clearer research questions before selecting datasets to ensure better alignment with study objectives.
- Allocate more time for data exploration and to figure out missing values and inconsistencies efficiently.
- Develop a structured approach for dataset comparison, ranking them based on relevance, completeness, and usability using Python.

c) How has this experience shaped your approach to research and data analysis?

- This experience has helped us understand the importance of defining clear objectives before picking 3 databases for the project.
- Emphasized the importance of selecting databases that have consistent data for countries and timeframes.
- Highlighted the necessity of framing and refining research questions that we can focus on.