

Submit a proposal with the following topics covered as a PDF:

- Topic - Describe and name your project in 1-2 sentences max
- Business Problem - Describe the business problem your project is trying to solve and/or the research questions you will explore
- Datasets - where are you getting your data? Describe the data that you will use to solve the problem
- Methods - What analysis methods will you use to complete this project? Note: this is just a proposal, your project can adapt as you work on it
- Ethical Considerations - What are some potential ethical concerns of this topic or analyzing the data?
- Challenges/Issues - What are some issues and challenges do you think you might face?
- References - What sources will you use to validate your results and support your project topic?

P

Topic:

Create a list of micronations that still exist today into a database, and provide information on each of these nations.

Business Problem:

There are several micronations that you can google upon and find more information about. This is usually through Wikipedia articles on pages on each of these nations. It is not easy to find one central source of information on all the micronations that aren't a part of larger countries or territories. By creating a single table with all the information required that we are looking for, this could be put into a database and pulled from. I will be looking to get data from various sources, various websites, and API calls, and doing data munging to generate a clean dataset that will be placed in a SQL database that can be pulled from.

Methods:

The Wikipedia page is micronations is the first starting point for this project. I will look to import all the various tables for each micronation on this page and combine them in one master table. I will then do some EDA and find some statistics like the population of each territory, the average age, and the GDP of each. Lastly, I will upload this all to an SQL database that can be pulled from and generate a webpage that references the data that I can present to.

Ethical Considerations:

The main consideration that is to make sure the data I am getting from various sources is not contradicting and finding sources of truth. I have a feeling that because a lot of the data I am looking for is on a niche topic, some of the data I find might now be trustworthy. So, I will need to be very conscious of the places I pull data from in order to make a clean dataset. I want to make sure I represent each nation in the best light and not present any false data on them.

Shaquiel Pashtunyar
DSC680 Project 1
Milestone 1
Challenges/Issues:

The issues and challenges are mainly that I will be scraping a lot of data from various online sources. There isn't really a Kaggle file that is already cleaned or something of that sort but rather Wikipedia tables and articles that will have me entering and altering data into a data frame. This will need to be done in a clean manner to make a final table that I can enter some of my visualization tools when I get to the final step. Managing all the incoming data formats will be the main challenge I see, and once all the data wrangling is complete, the visualizations and presentation I think will be easier to handle.

References:

My Wikipedia starting pages with many charts, and links to more datapoints one each nation

https://en.wikipedia.org/wiki/List_of_micronations