Project 1 Write-Up

Project: World Development Indicators

~~Potential Projects: World Development Indicators, Credit Card Fraud Detection, Bitcoin Historical Data, Google Trends Data, Onion or Not, Trending YouTube Video Statistics, Fruit Recognition, Super Bowl Ads, Student Performance in Exams, TMDB Movies, The Economic Freedom Index, Wine Reviews~~

Core Message/Hypothesis: How do the countries compare across economic development indices?

How well do the fruits of economic development (Outputs) relate to GDP & GNI per Capita?

Questions:

Sources of terms- Kaggle Datasets and <https://bizfluent.com/info-7746218-indicators-economic-development.html>

* What are the baseline indicators of economic development?
  + GDP per Capita
  + ~~Gross National Income (GNI) per Capita~~
* Does life expectancy increase with economic development across countries?
  + It is assumed that there is a strong positive correlation between life expectancy and economic development.
* Do a~~dult &~~ infant mortality go down with economic development across countries?
  + It is assumed that there is a strong negative correlation between maternal & infant mortality rates and economic development.
* What are the import & export rates across countries and how do they compare to economic development?
  + The question is asked because we theorize that there is a strong correlation between import/export rates and economic development.
  + Developing nations need to rely on production & exports to grow, while more advanced economies tend to move from manufacturing to services, so the expectation is:
    - Export rates and economic development: strong negative correlation.
    - Import rates and economic development: strong positive correlation.
* ~~Does the poverty rate decrease with economic development?~~
  + ~~It is assumed that there is a negative correlation between poverty rate and economic development.~~
* ~~Does the GINI coefficient decrease with economic development?~~
  + ~~It is assumed that the inequality in a nation increases as it is developing, but that it decreases after reaching a certain level of development.~~

Sources of Data

* Outputs
  + Kaggle Datasets
    - Life Expectancy,
    - Mortality Rates,
    - Imports & Exports
* Baseline Indicators (Inputs)
  + <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>
    - GDP per Capita
    - GNI per Capita

Types of Joins/Merges:

Country.csv & CountryNotes.csv: CountryCode

Country.csv & Footnotes.csv: CountryCode

Country.csv & Indicators.csv: CountryCode

Indicators.csv & Series.csv: IndicatorName

Series.csv & SeriesNotes.csv: SeriesCode

API\_NY.GDP.PCAP.CD\_DS2\_en\_csv\_v2\_821084 & Country.csv: CountryCode

Steps to Answer Questions:

* Merge files to create Master DataFrame with Jupyter Notebook.
* Clean the data with Jupyter Notebook.
  + Use DropNa function with how = “any” argument to remove rows with null values.
  + Sort values of whole master dataframe by the GDP per capita values.
* Create DataFrame for “Top 5” countries and one for “Bottom 5” countries based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980).
* Create Bar Charts showing “Top 5” & “Bottom 5” countries together based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980)
* Create Life Expectancy Scatter Plot comparing life expectancy vs. “Top 5” & “Bottom 5” countries together based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980)
  + Set the plot title, xlabel, & ylabel
  + Create regression parameters
  + Create the line equation.
  + Plot the regress values on the Scatter Plots.
  + Annotate the line equations on the Scatter Plots.
* Create Infant Mortality Rate Scatter Plot comparing life expectancy vs. “Top 5” & “Bottom 5” countries together based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980)
  + Set the plot title, xlabel, & ylabel
  + Create regression parameters
  + Create the line equation.
  + Plot the regress values on the Scatter Plots.
  + Annotate the line equations on the Scatter Plots.
* Create Export Rate Scatter Plot comparing life expectancy vs. “Top 5” & “Bottom 5” countries together based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980)
  + Set the plot title, xlabel, & ylabel
  + Create regression parameters
  + Create the line equation.
  + Plot the regress values on the Scatter Plots.
  + Annotate the line equations on the Scatter Plots.
* Create Import Rate Scatter Plot comparing life expectancy vs. “Top 5” & “Bottom 5” countries together based on GDP per capita.
  + One for earliest common year (i.e. 1960)
  + One for latest common year (i.e. 1980)
  + Set the plot title, xlabel, & ylabel
  + Create regression parameters
  + Create the line equation.
  + Plot the regress values on the Scatter Plots.
  + Annotate the line equations on the Scatter Plots.

Tasks:

Jemi – Merge the databases and clean the master DataFrame.

Maria – Create bar charts showing GDP per capita. (Reference the Steps to Answer Questions section)

Tom – Create scatter plots for life expectancy vs GDP per capita and infant mortality vs GDP per capita. (Reference the Steps to Answer Questions section)

Ben – Create scatter plots for imports and exports.