**ETL Mini Project**

**Soumya, Jenn, and Regina**

**10/16/19**

**Summary:**

For this project, we used data Economic Freedom Data of the latest year and how internal factors (income group, population, etc.) and an external factor (tourist count) affect the economic freedom score. Here we have taken three csv files to do this analysis. The sources of data that we have extracted from are given below. The cleaned data is joined and the final data is loading to the Relational Database Management System - Postgresql with the PgAdmin tool.

**Data Sources:**

1. <https://www.kaggle.com/lewisduncan93/the-economic-freedom-index>
2. <https://data.worldbank.org/indicator/ST.INT.ARVL>

**ETL Process/Steps Taken:**

**Step1:** Extracted data  
**Step 2:** Cleaned csv file "tourism\_data\_1.csv" and creating dataframe "tourism\_df1\_clean"  
**Step 3:** Cleaned csv file "tourism\_data\_2.csv" and creating dataframe "tourism\_df2\_clean"  
**Step 4:** Cleaned csv file "mini\_project\_economic\_freedom.csv" and creating dataframe "economic\_df\_clean"  
**Step 5:** Joined dataframes "tourism\_df1\_clean" and "tourism\_df2\_clean" and making a single dataframe called "merge\_with\_replace"  
**Step 6:** Joined dataframes "economic\_df\_clean" and "merge\_with\_replace" to get the final dataframe "final\_data\_to\_load" .  
**Step 7:** Created html file from final data(final\_data\_to\_load).  
**Step 8:** Connection to database(RDBMS)-POSTGRESQL  
**Step 9:** Created table in the database, creating the sql file.¶  
**Step 10:** Checked for the "etl\_economic\_freedom" table in POSTGRESQL  
**Step 11:** Confirmed the data had been added to POSTGRESQL by querying the table.  
**Step 12:** Created visualizations by using the data in Jupyter Notebook.

**Step 13**: Took df\_to\_html to create html table.

**Step 14:** Created website to showcase html data table and analysis visualizations.

**Analysis**

The analysis shows that the economic freedom score of countries depends upon income. Per the latest report, Hong-Kong scores the highest on economic freedom with a score of 90.2, and the Asia-Pacific region has the highest economic score. The majority of top countries are from Europe, and they have comparatively high income. Sub-Saharan Africa has a comparatively low economic freedom score, and they have mostly low or lower middle income. The high-income group is comparatively more than other income groups, and their average economic freedom score is above 70. The average economic freedom score of the low-income group is below 55. The tourist count is higher in regions such as Europe and Asia-Pacific where economic freedom scores are higher. The sub-Saharan Africa region has low economic freedom, and their tourist count is also low.