International Business Voyager

Data base creation.

By: Carlos J., Denis S., and Ian W.

Purpose:

The aim of our report is to indicate countries that may be potentially beneficial for the business expansions of our company. Due to growth, and pace of growth, we have concluded that it is appropriate to create some type of data base that can fit key variables required when analyzing a potential country to expand our business too.

We will conduct our database creation by using the method of ETL. ETL is short for Extraction, Transformation, and Loading, which is an essential practice for data scientist creating databases and analyzing.

Extraction

First, we took into consideration some basic variables/information of a many countries. In particular, a dataset with a unique number of countries which we assigned the primary key and use this as our reference to other datasets when using SQL. We also gathered other datasets on Economic Freedom, Ease of Doing Business and Terrorism.

For our Economic Freedom dataset, this file includes many variables that relate to legal issues, tariffs, government, credit, and etc. We see this data fit to help develop an analysis to whether a country is fit to host one of our business locations. Fortunately, we were able to fit this dataset into our database.

For our Ease of Doing Business dataset, this file includes many variables such as how easy it is to open a business, registering a business, starting a business and ease of doing business. we see this data fit to help develop an analysis to whether a country is fit to host one of our business locations. Fortunately, we were able to fit this dataset into our database.

For our Terrorism dataset, our idea was to gather data on past threats on a country. This would give us some type of indication of how prone one place is to violent events which may impede our business conduct. Therefore, we see this data fit to help develop an analysis to whether a country is fit to host one of our business locations. Unfortunately, our Terrorism dataset had a few complicated which we did not have enough time to clean and add to the database.

Transform

For our transformation stage. Our biggest concern was applying identity keys to each dataset, particularly on with respect to the county names.

We ran into a few problems with respect to proper spelling, and modifying each names, so all datasets have the same spelling and format of names in each country column for each dataset.

We overcame these problems for all datasets except for our terrorism datasets. With Terrorism, there was just a problem with the country name that we didn’t have enough time to fix, therefore we left the terrorism dataset aside so we can focus on the other datasets for now. Eventually, this problem will be solved but for now it’s on stand-by.

Loading

We used SQLAlchemy for the loading part.

Conclusion

For our conclusion, we have gathered relevant datasets together, transformed and compiled them into a database and have them readily available for future insights and references.

We are glad to have this database and have future plans to include more data as the days go on.

We are confident in our finding and complete our ETL initiative.