ETL Assignment

Andrew Johnson

Laurie Grant

Sean Evans

The sources of data that you will extract from.

\* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

\* The type of final production database to load the data into (relational or non-relational).

\* The final tables or collections that will be used in the production database.

You will be required to submit a final technical report with the above information and steps

\* \*\*E\*\*xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

\* \*\*T\*\*ransform: what data cleaning or transformation was required.

\* \*\*L\*\*oad: the final database, tables/collections, and why this was chosen.

We as a group delved into three datasets to help examine the agricultural imports and exports of goods throughout the world. This report highlights the importance of trade and how much people rely on cooperation for their daily goods.

Extract

1 – The first was a cumulative amount of meat and livestock imported and exported to the United States.

2 – Our next dataset was consisted of national ISO country codes. This information provided an opportunity to join our separate datasets on a mutual variable – the list of country calling codes. Extracted through web scrapping, we were able to secure the data with Python

3 - The final dataset was pulled as a CSV file from the website Kaggle. Titled “Who Eats the Food That We Grow” the data is comprised of the food supply for nearly 245 countries since 1961. Provided by the Food and Agricultural Organization of the United Nations, the dataset is comprised of over 21,000 rows of data that provides Country, Country Area Code, the Foods universally utilized by each listed country since the year 1961.

Transform

The group utilized three separate methods of Transforming the data.

- Our first dataset comprised of the cumulative imports/exports was cleaned to successfully upload the data. We sectioned the information from the original table from Excel into separate sheets, which were then uploaded as individual CSV files onto MySQL using the Table Import Wizard. Some of the data would not import onto MySql. While uploading the original 201 300

-The third dataset was initially uploaded into SQL as a table. Unfortunately, the final table only contained just 4,000 fields of data – a drastic loss compared to the original 21,000. To solve this issue, we uploaded the

Load

The final load of the database was into a SQL file. Each transformed dataset was