# ETL Assignment, Week 13

**Team Members:** Andrew Johnson, Laurie Grant, Sean Evans

18 May 2019

**Summary:** 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. The datasets allow reporting by ISO Country Code of the consumption of meat and the import and export of meat for a given country.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Extract | **Transform** | **Load** |
| 1 - Meat and Livestock Import and Exports |  | Following the transformation, each data set was properly uploaded onto a SQL Database labeled ETL. We were able to join the three separate datasets utilizing the similar variable ISO Codes. The purpose of creating this SQL - ETL Database. |
| The first was a cumulative amount of meat and livestock imported and exported to the United States. This data was pulled from the United States Department of Agriculture as an Excel document.  <https://www.ers.usda.gov/data-products/livestock-and-meat-international-trade-data/livestock-and-meat-international-trade-data/#Zipped%20CSV%20files>  CSV: <https://www.ers.usda.gov/webdocs/DataFiles/81475/LivestockMeat_Yearly.xlsx?v=8782.9> | Our first dataset, comprised of the cumulative imports/exports, was initially cleaned on Excel. We sectioned the information from the original table onto separate Excel Sheets. Some of this data would not import onto MySql and required returning to Excel to remove the blank cells. |
| 2 – Country ISO Codes |  |
| The ISO country codes provided an opportunity to join our separate datasets on a mutual variable. This information was extracted in *three* distinct ways.   * Simple Copy/Paste from Website to CSV which was very fast. * Import the Pandas Table which was also fast, but had many data cleanup issues. * Webscraping – not fast due to the unstructured HTML code, but produced cleaner data.   <https://www.nationsonline.org/oneworld/country_code_list.htm> | There were three separate ways of data cleaning the ISO data set.   1. The initial extraction onto an excel file required no clean up. 2. The second method of extraction required the Pandas Data Frame and manipulating 3. The third cleaning method of Webscrapping    * The CSV looked clean, but once imported into SQL there was much cleanup for the encoding.    * The pandas dataframe was not pursued due to the amount of cleanup needed.    * Web scraping required no cleanup. |
| 3 – Consumers of Food |  |
| The final dataset was pulled as a clean 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.  <https://www.kaggle.com/dorbicycle/world-foodfeed-production> | The third dataset containing the data from Kaggle was initially uploaded directly into SQL as a CSV file. Unfortunately, the final table only contained just 4,000 fields of data – a drastic loss compared to the original 21,000 rows. To solve this issue, we utilized Pandas to create a data frame which then manipulated the ISO to make it more upload friendly. |