Russian Alcohol ETL Project

Report

Team Members:

Ben Lamkin, David Witton, Nick Riffel, & Nick Sain

Objectives:

**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.

1. Found data sources at:
   1. <https://www.kaggle.com/dwdkills/russian-demography>
   2. <https://www.kaggle.com/dwdkills/alcohol-consumption-in-russia>
2. Download Data to CSV & store in folder
3. Load CSV’s into Jupiter Notebook
4. Create Data frames from CSV’s
5. Remove auto index.
6. Set index as Year for both Data frames.
7. Use Pandas to remove date’s before 2000 & after 2016 for both data frames
8. Export CSV’s for both Data frames.
9. Created two tables in PosgresSQL
10. Imported data from CSV’s
11. Filter out to only include(Year, Region, Birth Rate, Death Rate, Wine, Beer, Vodka, Champagne, & Brandy.
12. Perform Interjoin between the two Tables.