

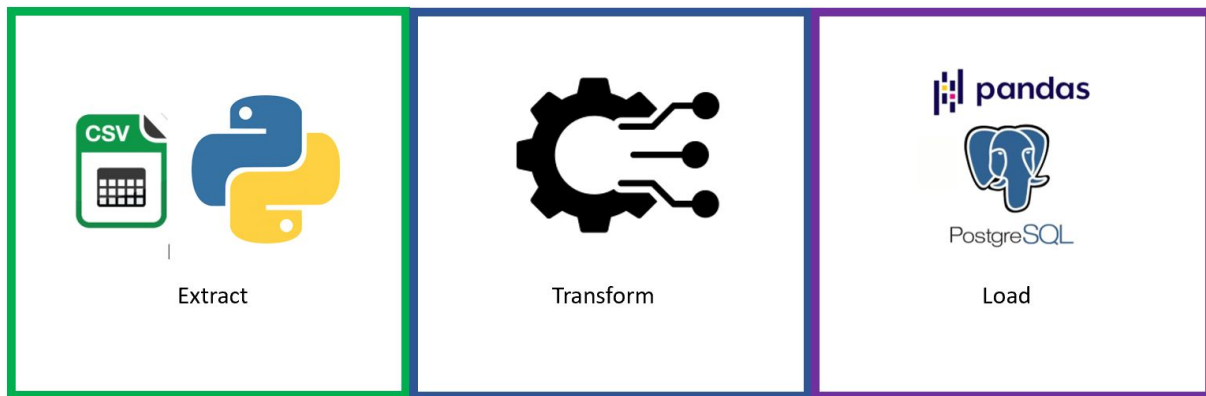
ETL Project Report

Team Members: Brian Roberts, Vanessa Howell, Amanda Vital
December 14, 2020

Happiness Report vs. Alcohol Consumption:

Project Proposal: *Do countries with lower Happiness Index scores tend to consume more alcohol?*

ETL



Extract:

Sources of Data:

- Kaggle - Drinks by Country
(<https://www.kaggle.com/faressayah/data-visualization-seaborn-matplotlib-tutorial/comments>)
- Kaggle - World Happiness Report
(<https://www.kaggle.com/mathurinache/world-happiness-report>)

The fields of Interest include the following:

- Country
- Happiness Index
- Beer Servings
- Spirit Servings
- Wine Servings
- Total Litres of Alcohol

Our original two data sources will both be found on Kaggle. The first is a 2020 combined survey from the World Happiness Report. It is a csv that includes summary statistical data for each countries' population's perceived happiness, based on a 1-10 scale. This dataset also includes social data such as; Logged GDP per capita, social support, health life expectancy, and perceived corruption. The second dataset shows what we believe to be the number of alcoholic beverages consumed in a year on average by residents of each country who choose to consume alcohol.

Transform:

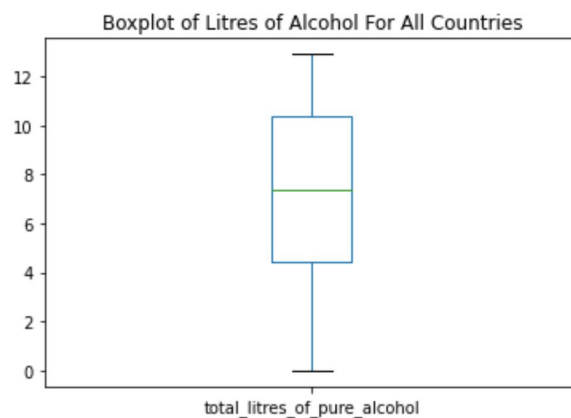
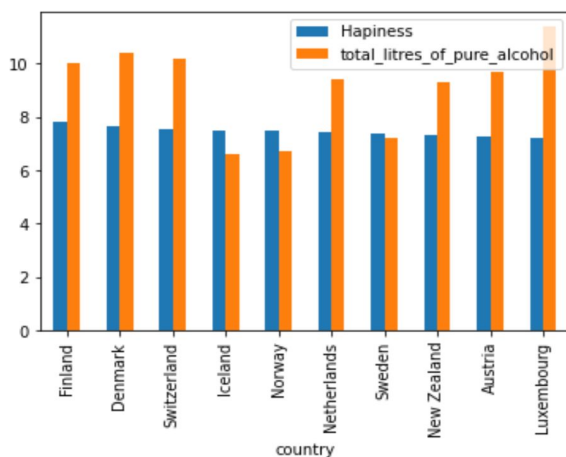
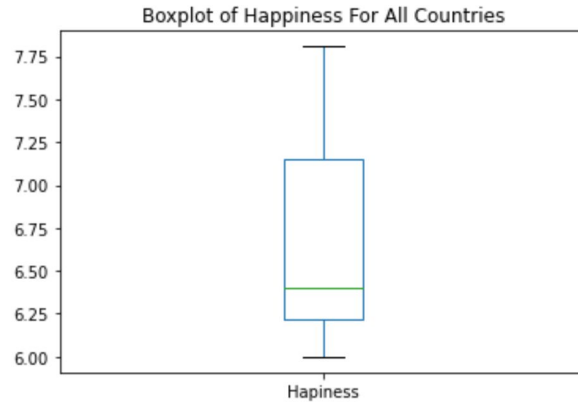
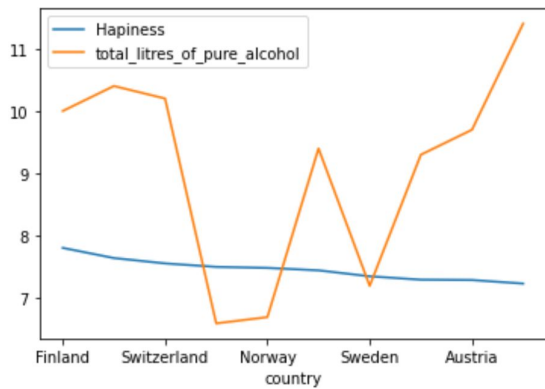
Both datasets will need to be cleaned before they can be combined into a final database. While both datasets can be joined by country name, they have different regional/continental indicators that will need to be matched accordingly. The world happiness dataset also has a large number of superfluous columns that must be dropped. Some of the columns in this dataset should also be renamed to provide a clearer understanding of what they look to portray. Thankfully, the Drinks by Country dataset is more concise and requires no manipulation at this time. These functions will be performed via Pandas/Python on a Jupyter Notebook.

Load:

Once the datasets are cleaned to our liking the data from the clean csvs will be pushed into pre-generated SQL tables within the same SQL database. We feel that Object Relational Mapping is the database format that best suits the type of data we are transforming. This is because the ORM tool automatically generates the data access code we need to write. Therefore, if later on we were to create an application that interacts with this data model, the amount of code we would need to write would be drastically reduced. Finally, the two tables will be joined on their primary key, which in this case will be 'country_name', producing our final product. A two sample t-test may be run on the data to determine whether or not our hypothesis was correct. A box-and-whisker plot based on continent for both Happiness Index and Alcohol Consumption may also be interesting to visualize.

Visualizations/Tests:

We used a relational database (PostgreSQL) to link the data by our common identifier, Country.



Sample Queries:

As stated in Science Daily, it is easy to say that alcohol makes you temporarily happy but we wanted to look at a wider scope to see if drinking patterns among people in different countries relate to how they view their overall happiness.

This database can be used to look at relations between specific types of alcohol and their relation to happiness (i.e. beer, wine, spirits) as well as total consumption in litres. However, it's important to keep in mind that alcohol may have different a

Challenges:

- Finding data that can be used for the project and could be joined together
- Finding data based on countries
- Dropping unnecessary columns and unneeded data from the csv's
- Creating tables using SQL and Jupyter Notebook

Summary/Conclusion:

We cannot conclude that happiness has a correlation with alcohol consumption. The data assumptions that were made were that the data comes from a random sample as well as an adequate sample size and that the measurement system being used is accurate and precise.

Citations:

University of Kent. "Alcohol makes you momentarily happier but not more satisfied." ScienceDaily. ScienceDaily, 4 May 2016. <www.sciencedaily.com/releases/2016/05/160504121645.htm>.