ELT Project - 2019

Team Members: Brickey LeQuire and Alex Koynoff

Team Name: Drunken Zip Codes

Summary: Following ETL procedures, we will analyze the count of bars by zip code and merge the data with various demographic data points such as population and churches by zip code. Then we will create some computed information, such as bars per capita, count of bars by zip code (concentration), church to bar ratio, etc, and load that into a database.

Extract: Gather the data from the sources below in csv/excel and sav files.

Transform:

- Review the datasets and select the columns to load.
- Clean the datasets by dropping blank values as appropriate.
- Perform merging of data as appropriate.
- Clean the zip codes for all data sets to make them consistent (XXXXX vs XXXXX-XXXX)
- Derive calculated values (e.g. count of bars by zip code).
- Create a database and tables before loading:
    - Database: Bars_db
        - Tables:
            - List of bars and their zip codes
            - Demographics and zip codes
            - Churches and zip codes
            - Calculated columns(done via Python) and zip codes

Load: Load data to a relational database (MySQL).

Sources:

Bars info: https://data.world/datafiniti/breweries-brew-pubs-in-the-usa

Zip Code info:

- https://www.kaggle.com/census/us-population-by-zip-code
- https://www.pewforum.org/dataset/pew-research-center-2014-u-s-religious-landscape-study/