ETL Project

Submitted January 22, 2020

Amanda Belloma

Dallas Diaz

Jamie Lewis

Tim Raynor

# Data Sources

## United States Income by State

Income by state data was sourced from the United States Census Bureau. Data includes median household income by state from 1984 to 2018. Standard deviations are included as well.

URL: [https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households](https://www2.census.gov/programs-surveys/cps/tables/time-series/historical-income-households/h08.xls)

## Wineries

Data related to the number of wineries in the United States was obtained from wine-road.com (source data attributed to The National Association of American Wineries). Data includes the number of wineries grouped by state.

URL: <http://www.wine-road.com/education/articles/state-winery-rankings.php>

## Breweries

Data related to the number of breweries in the United States was obtained kaggle.com. Data includes brewery name, type, address, website, state, total breweries in state.

URL: <https://www.kaggle.com/brkurzawa/us-breweries>

## Populate by State

Population by state for several years, using 2018 as the most current to match the income by state from the United States Census Bureau.

URL: <https://www.census.gov/newsroom/press-kits/2018/pop-estimates-national-state.html>

# Data Wrangling

1. Dropped columns not used
2. Renamed columns
3. Performed counts
4. Uppercase state columns to allow for better matching
5. Searched data frames to validate data could be pulled

# Data Schema

