**Wine Global Point Values and Reviews**

**Order of operations for the data processing and upload:**

1. First, load and run the Jupyter Notebook: wine\_additional.ipynb
   1. This Notebook processes a web scape of raw data that is transformed into a Jupyter notebook table and then further process just for the worldwide wine consumption. The Notebook does contain images as well that could be used later for an app or website depending on the use case.

-1-image US Alcohol Consumption'https://fivethirtyeight.com/wp-content/uploads/2014/08/chalabi-datalab-alcohol-1.png?w=610 1x, https://fivethirtyeight.com/wp-content/uploads/2014/08/chalabi-datalab-alcohol-1.png?w=1220 2x'

-2-Image List of all of the information in the table: 'https://fivethirtyeight.com/wp-content/uploads/2014/08/chalabi-datalab-alcohol-table-2.png?w=610 1x, https://fivethirtyeight.com/wp-content/uploads/2014/08/chalabi-datalab-alcohol-table-2.png?w=1220 2x'

1. Second, run the Jupyter Notebook named: wine review.ipynb
   1. This notebook processes two separate data files and merges them into one. It contains the review information for the worldwide wine reviews along with the titles, points, and lots of additional information.
   2. The wines are separated into columns for 100 points and another table that is 90 to 99 points. This allows for sorting to find the top wines by their cost.
   3. The unique titles for all the wines at 100 points was found.
   4. The main “wine\_merge” dataframe contains all the information including wine review were available within the dataset. The reviewer’s Twitter account is also listed to gain more information if needed.
   5. Notebook also contains charts that can be used.
2. After running the notebooks, they are all exported into new tables to csv files that can be imported into a main SQL database or use for any other data needs.
3. For this database, the files were all loaded into a Postgres SQL to be used for future updates. The tables can all be used to either search for top times, search for reviews, etc. The data can also be loaded into a flask Mongo or used for processing an app to make searching easy for the user to find top notch wine in one easy app. These are all ideas for data use down stream after this initial extraction, transforming, and loading to keep. A great use case would be to keep a database and update with new wines and reviews and create an app so customers can easily search for wines by country, wine title, points of the wine, reviews for the wine, and in which province they are from if that is wanted or needed by the customer. The use cases are endless. Currently, the database is a holding point to save all the information in one place. A use case is in progress.

**Data Sources scraped or borrowed for this project from the following:**

1. **wine\_additional.ipynb:** scrapped data off the website: <https://fivethirtyeight.com/features/dear-mona-followup-where-do-people-drink-the-most-beer-wine-and-spirits/> borrowed from article by: Mona Chalabi is data editor at the Guardian US, and a columnist at New York Magazine. She was previously a lead news writer for FiveThirtyEight. @monachalabi

**wine review.ipynb:**

1. Data obtained from Kaggle which was provided in Json format by ChristopheIV, <https://www.kaggle.com/christopheiv/winemagdata130k>
2. : <https://data.world/markpowell/global-wine-points/workspace/file?filename=Wines.xlsx> (<https://query.data.world/s/mplvtacphx3svygfg3uu4gycii2dbt>) Global Wine Points, Dataset BY Mark Powell, data provided by “Data.World”, Dataset updated Mar 15, 2021