**Data of different counties’ Population and household Income (2012 and 2017):**

**Source of Dataset:**

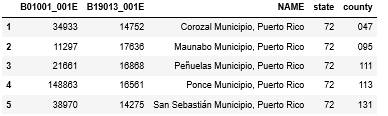
Data was downloaded from https://api.census.gov

From the dataset County was considered instead of Metro because Government often change the boundaries of the Metro Areas in every census. So there was a possibility of a wrong parameter to judge the data set.

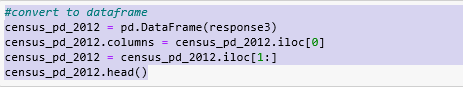
**Convert dataset into DataFrame:**

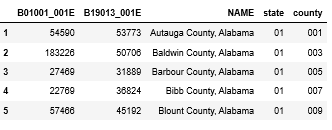
This dataset included the population and household income of counties’ in different states in the year 2017. API was the source of access to get the data. B01001\_001E was the code for total population and B19013\_001E was the code for total household income. json was the reader. Data for population and income was filtered by .iloc[] function. Then dataset was converted into DataFrame.





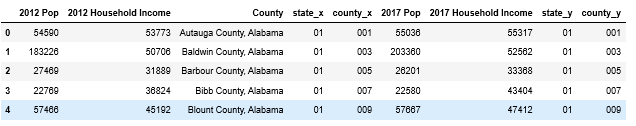
Same process followed for 2012 dataset. B01001\_001E was the code for total population and B19013\_001E was the code for total household income.





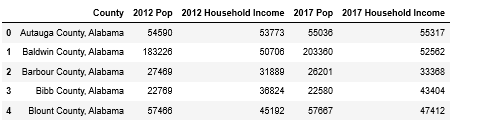
Next step was merged the two data by inner join





Drop the State and county code columns’ which are not required after tally. Final dataset is ready. Convert the dataset into csv file for further analysis.







**Data of Yelp rating of Five Guys, Halal Guys, McDonalds, Panera Bread, Shake Shack, Taco Bell, Texas Roadhouse in different Metro Areas:**

**Source of Dataset:**

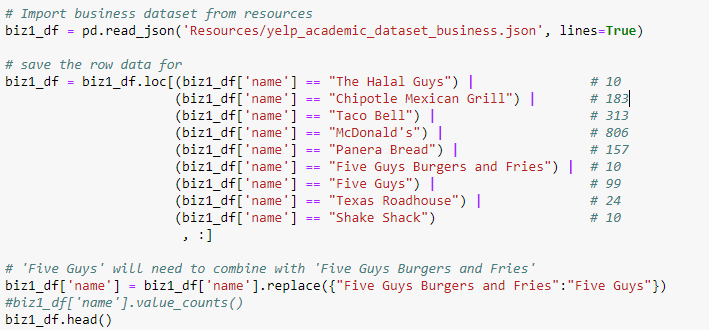
This dataset was downloaded from Kaggle: <https://www.kaggle.com/yelp-dataset/yelp-dataset>

This dataset showcased restaurants and businesses found on Yelp and information concerning the location and user reciprocity. Overall, it seemed outdated and incomplete as some major cities, such as Los Angeles, didn’t seem to have entries. However, the dataset did include entries on all restaurant franchises that interested this project. Since it was downloaded, there was the benefit of not having to rely on APIs to access the data.

For each entry, the columns of information were:

1. Address (location)
2. Attributes: keywords Categories: Another list of keywords to describe the type of food and service e.g. “Fast Food”
3. and phrases used to describe the business e.g. “Accepts Credit Cards”
4. Business ID: Unique code for this location. Unique among separate locations in a franchise.
5. City (location)
6. Hours Open
7. Latitude (location)
8. Longitude (location)
9. Name of Business/Franchise
10. Postal Code
11. Review Count: Number of reviews this location has received
12. Stars: Average star rating of above review count.
13. State (location)
14. **Cleaning the Raw Data to Only Include Franchises Related to this Project**

First, the data was retrieved from the Resources directory it was downloaded into. Then, all entries not related to the eight franchises selected for the project were filtered out using the .loc() function. Finally, all entries with names ‘Five Guys Burgers and Fries’ were renamed to ‘Five Guys’ since they represent the same franchise. This was done with the. replace () function.

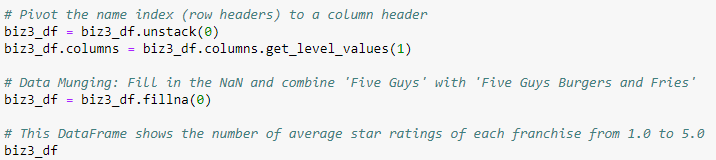


1. **Manipulating the Data for the ‘Review Count Spread’ DataFrame**

First, create a new DataFrame with only the columns needed (name, stars, review\_count), then create the double groupby via name then stars.



Second, unstack and level the DataFrame so a pivot level is created comparing star numeric ranking and franchise.



1. **Manipulating the Data for ‘Top Ten Most Popular Cities for Each Franchise’**

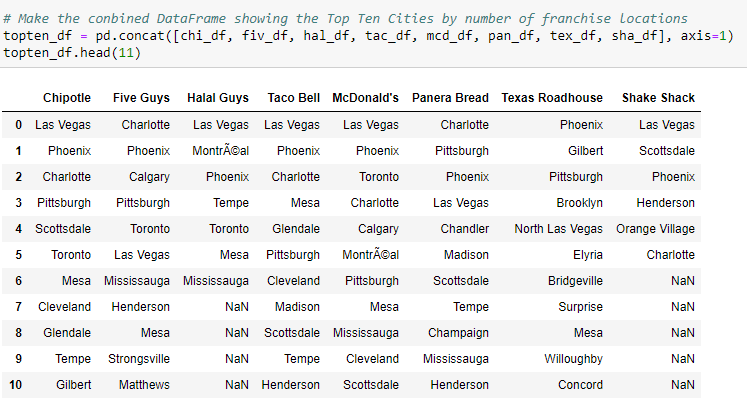
First, a new DataFrame was created with relevant columns (name, city, stars).



Second, create new, single column DataFrames for each franchise ranking the cities by number of franchise locations.



Lastly, combine the single column DataFrames of each franchise into one with the .concat() function to see the top ten cities.



**Data of social network rating of Chipotle, Five Guys, Halal Guys, McDonalds, Panera Bread, Shake Shack, Taco Bell, Texas Roadhouse in different states in US of last 5 years:**

**Source of Dataset:**

**Steps of Data Cleaning:**

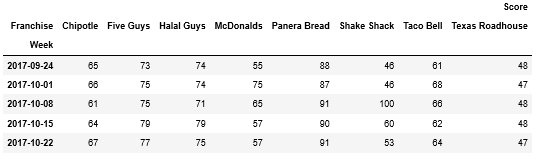
After download dataset to join the dataset use inner join function. After that combine into a single DataFrame with the .concat() function to see the allover rating of those franchises in different states in each week.

****





Unstack and level the DataFrame so a pivot level is created comparing franchise and score in each week.



Finally convert the dataset into csv for further analysis.

