The Battle of Neighborhoods

9/25/2020

# Capstone Project — The Battle of Neighborhoods

## 1.1 Introduction

Phoenix is the capital city of Arizona and is in the central region of the state. People may know it for its year-round sun, desert beauty, and world-class resorts and golf, but as the fifth-largest city in the U.S., it also offers sophisticated urbans capes, southwest culture, and lots of outdoor adventure.

As the heart of Arizona, Phoenix having a huge amount of people working and living here. It is also the fifth-most populous city in the United States and the largest state capital by population, and the only state capital with a population of more than one million residents.

Also, lots of foreign visitors and immigrates come to live here. It has a mix of culture and people came come to multiple countries to live and open a restaurant right here. With its diverse cultures, comes diverse food items. There are many restaurants in Phoenix, each belonging to different categories like Japanese, Italian, French, etc.

So as part of this project, we will list and visualize all major parts of Phoenix.

Questions that can be asked using the above-mentioned datasets

What is the best location in Phoenix for Brunch?

What places have the best restaurants in Phoenix?

Which areas have a large/small number of Brunch Market?

Which is the best place to stay if I prefer Brunch?

## 1.2 Data

For this project we need the following data:

Phoenix Restaurants data that contains list Locality, Restaurant name, Rating along with their latitude and longitude.

Data source: yelp Kaggle dataset: "https://www.kaggle.com/yelp-dataset/yelp-dataset"

Description: This data set contains the required information. And we will use this data set to explore various locality of Phoenix.

Nearby places in each locality of Phoenix.

Data source: Foursquare API : "https://developer.foursquare.com/"

Description: By using this API we will get all the venues in each neighborhood.

## 1.3 Approach

Collect the Phoenix city data from yelp Kaggle dataset

Using Foursquare API we will find all venues for each neighborhood in Phoenix.

Filter out all brunch in Phoenix.

Using aggregative rating for each restaurant to find the best places.

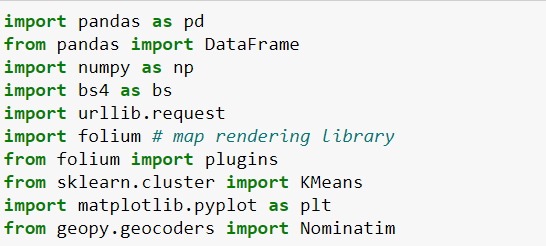
## 1.4 Foursquare API

Foursquare API is used to find all venues for each neighborhood and filter that are nearby by population density.

Also, the process will use aggregative rating for each restaurant to find the best places.

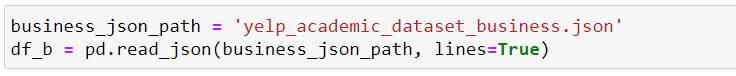
Visualize the Ranking of neighborhoods using folium library(python) to find the best place to open a new brunch

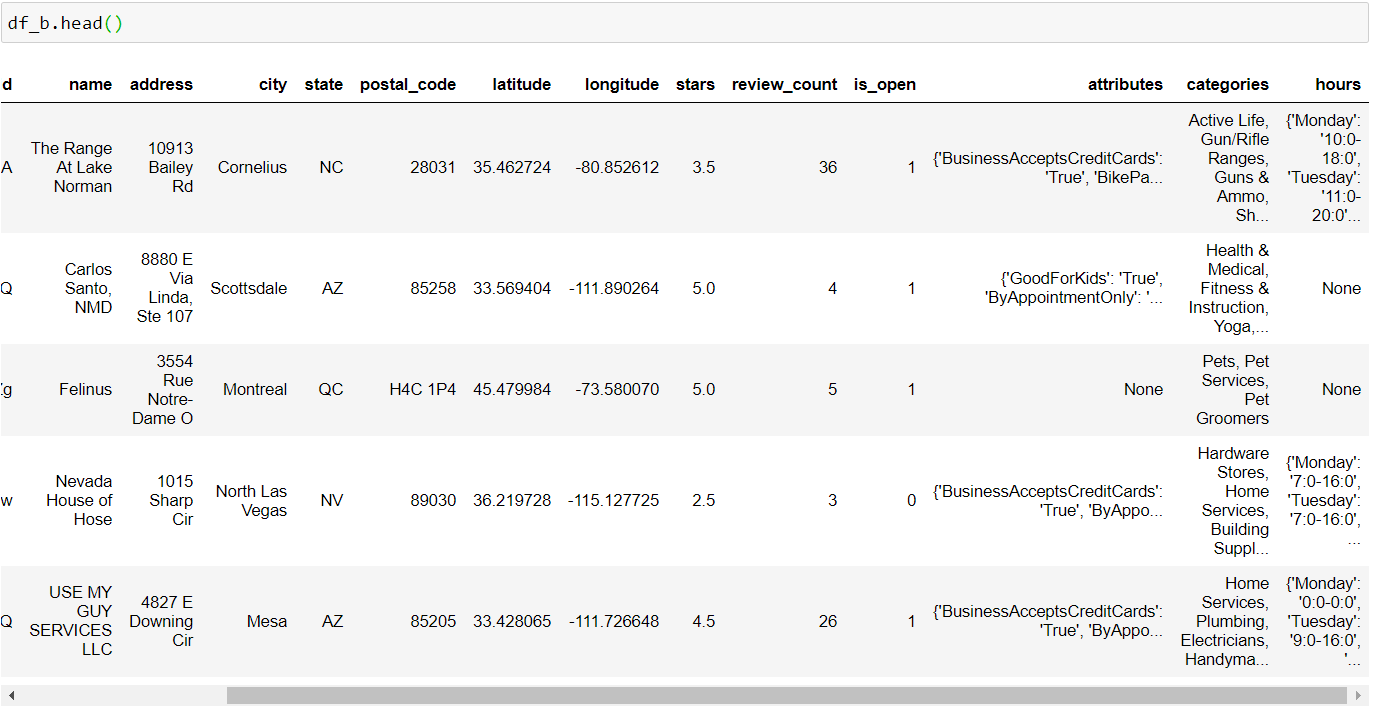
## 1.5 Packages include in the project



# Data Mining

## 2.1 Read the yelp\_academic\_dataset\_business and convert to df



## 2.2 Data Cleaning

### 2.2.1 Filer the useless data out of the data frame

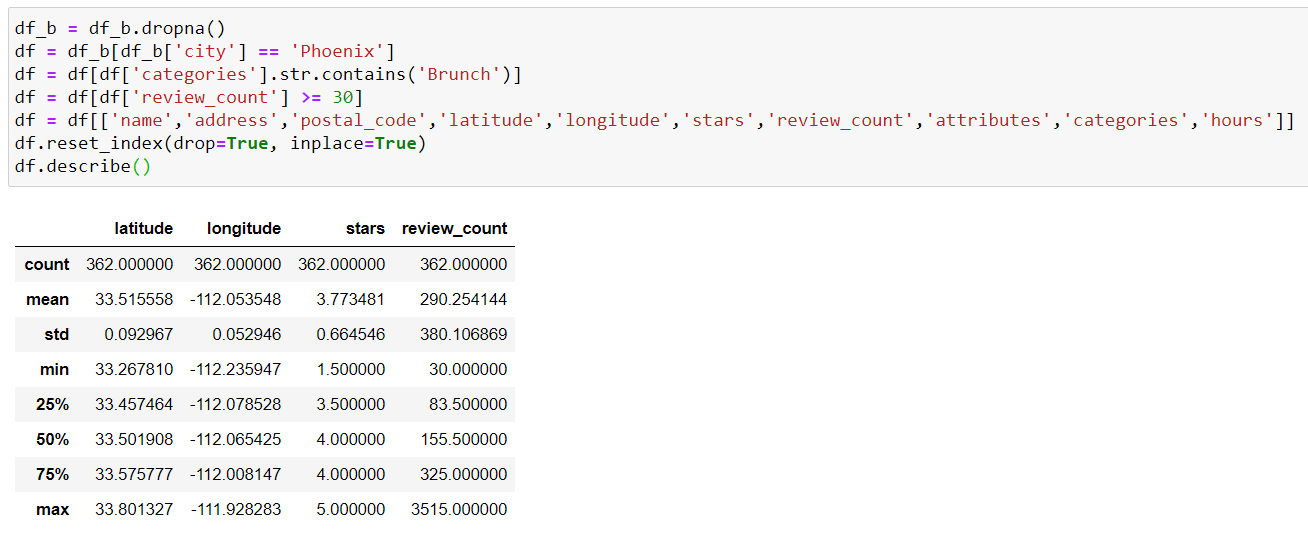
We only need data in Phoenix, data frame will only leave where state == AZ and City = Phoenix

Attributes which left are: 'name’, ‘address', 'postal\_code', 'latitude', 'longitude', 'stars', 'review\_count', 'attributes','categories', and 'hours'

Clear data with review\_count less than 30, since too little reviews means the data is not reliable.

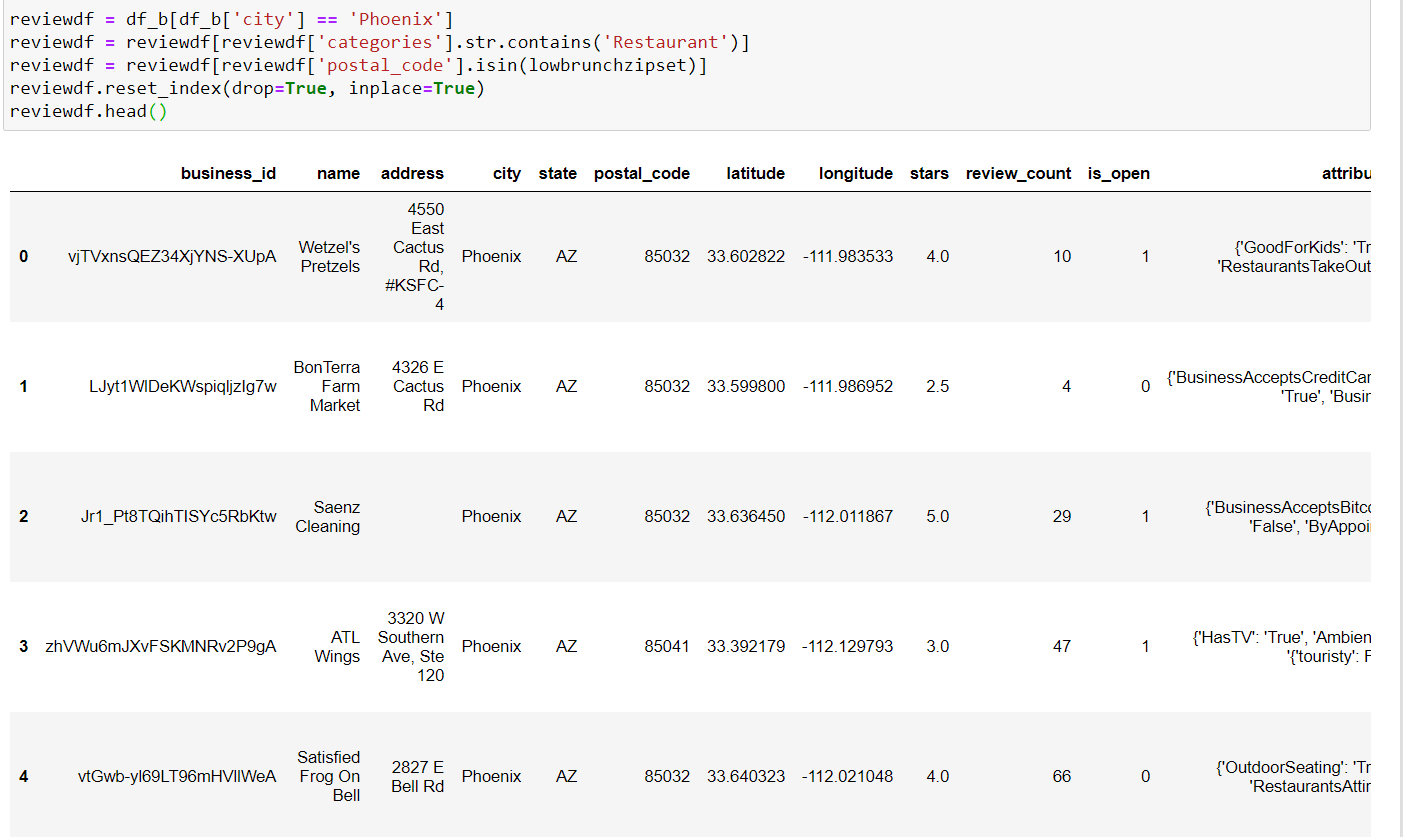
Data frame 1 : all brunches

* figure by categories contains brunch

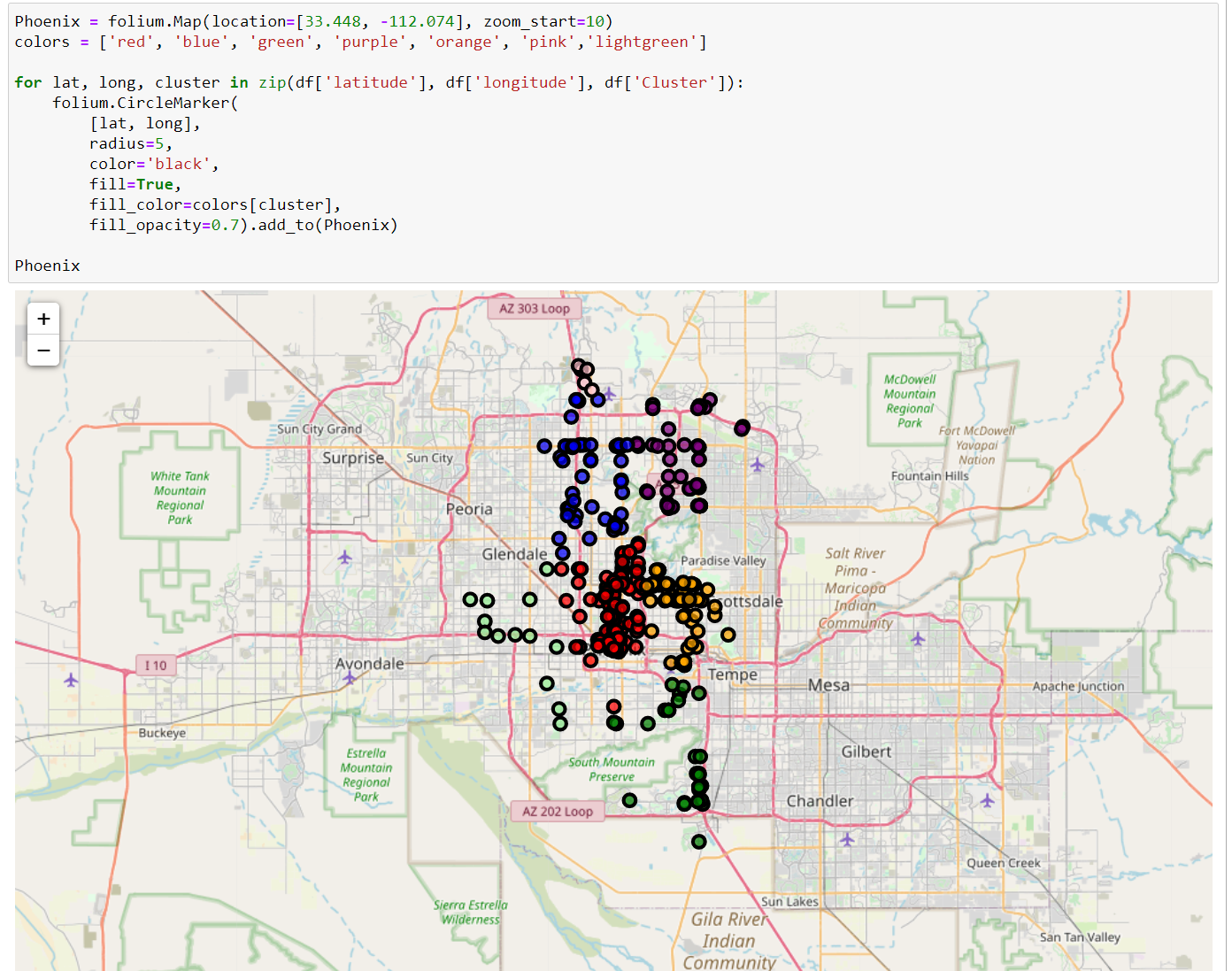


Data frame 1 : all restaurants

* figure by categories contains restaurants



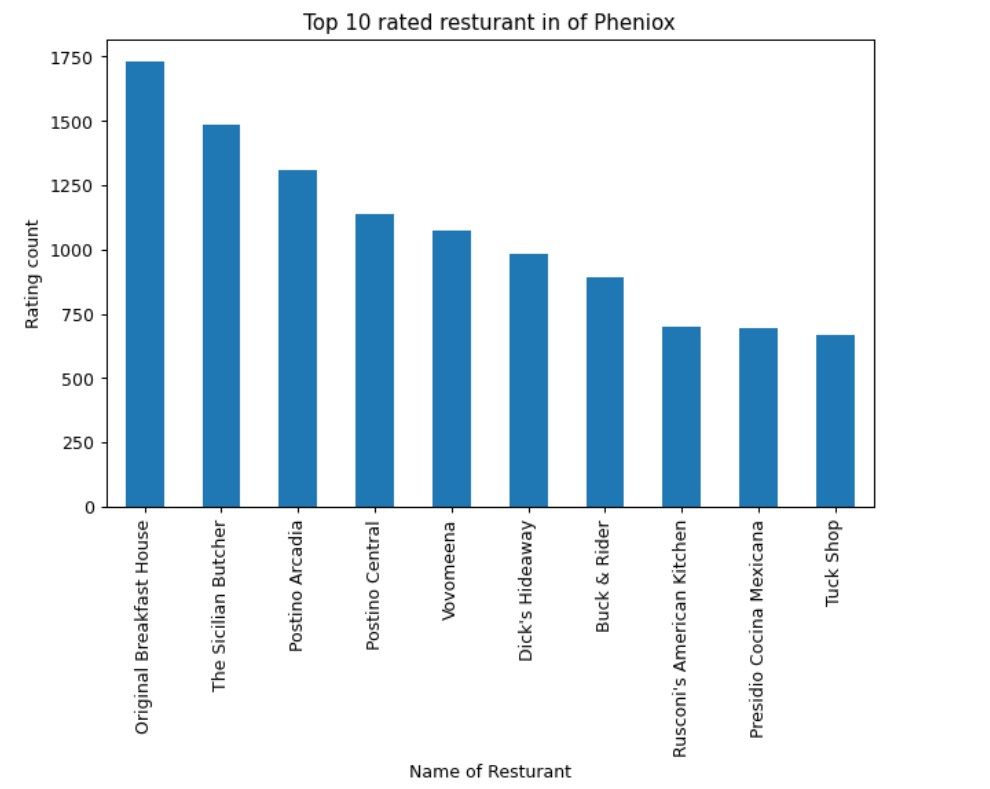
## 2.3 Clustering the city by location



## 2.4 Information for best and worst brunches in Phoenix and where they locate

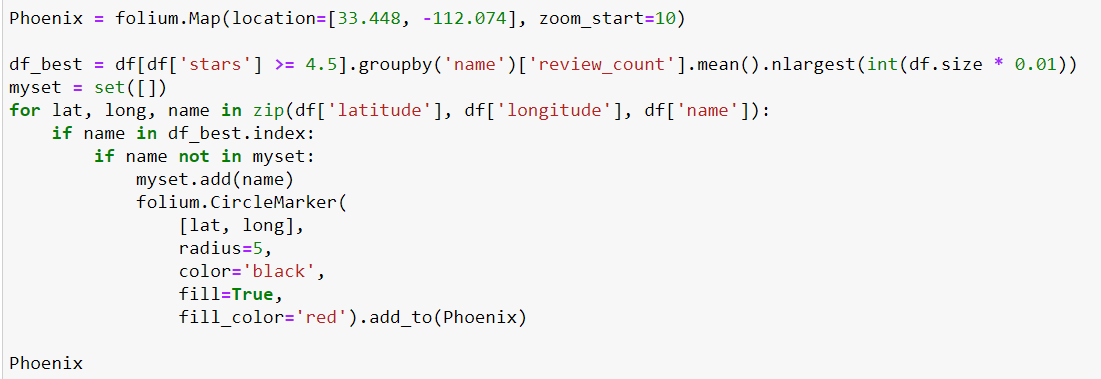
### 2.4.1 Best brunches

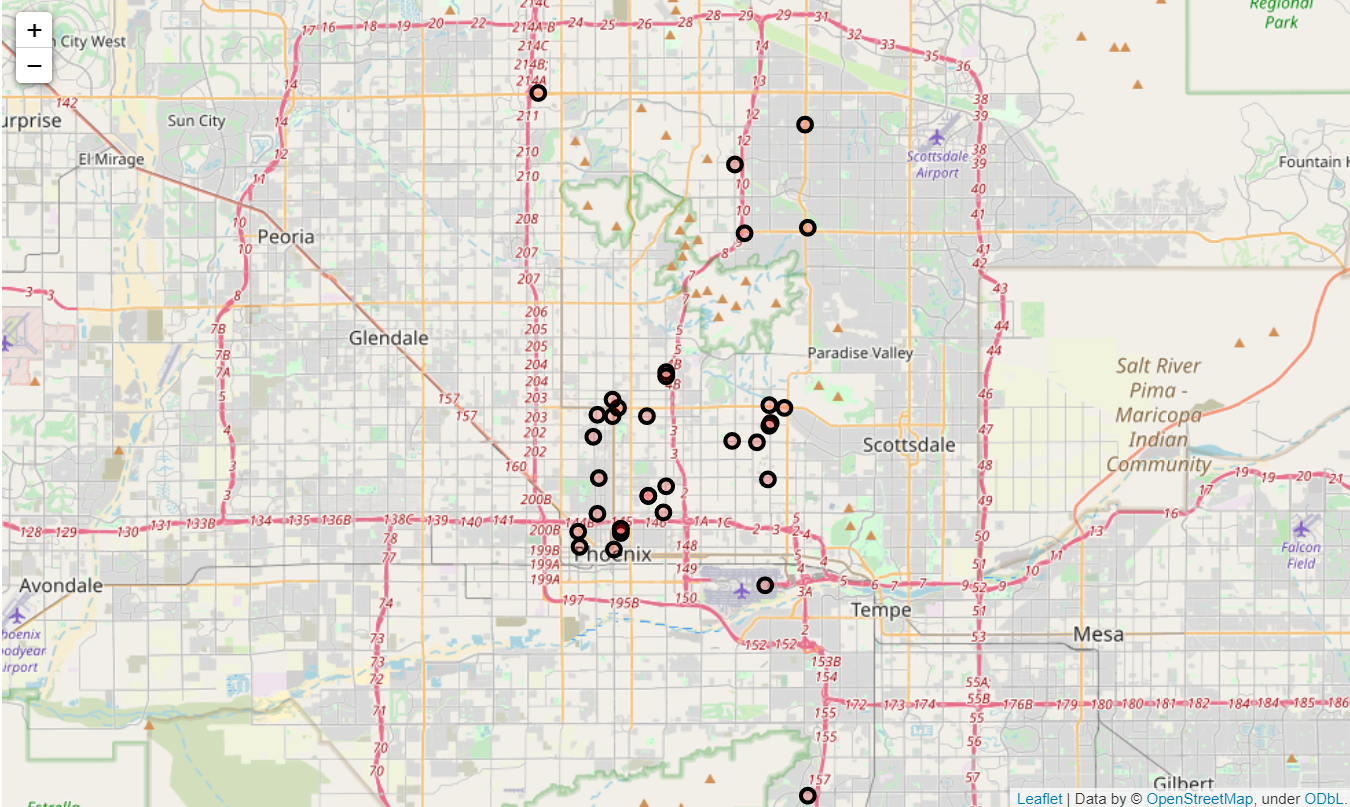
- Take the brunches with 4.5 stars and rank by review numbers

location of these brunches

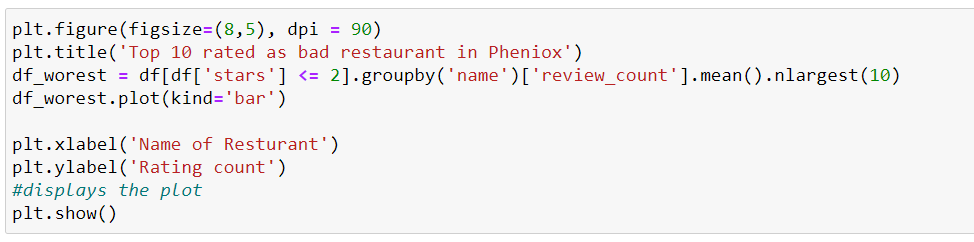
* mostly locates in Cluster red and yellow (center of the city)

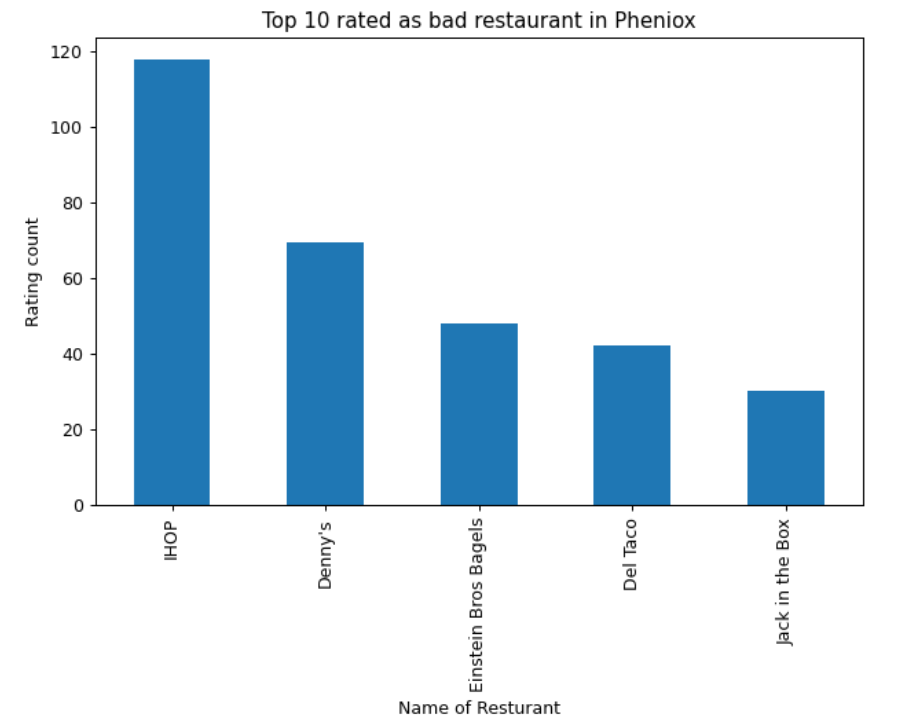


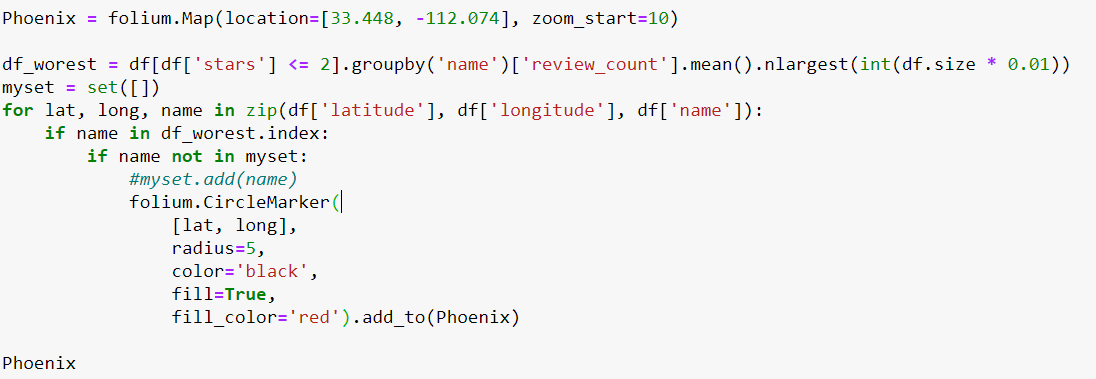


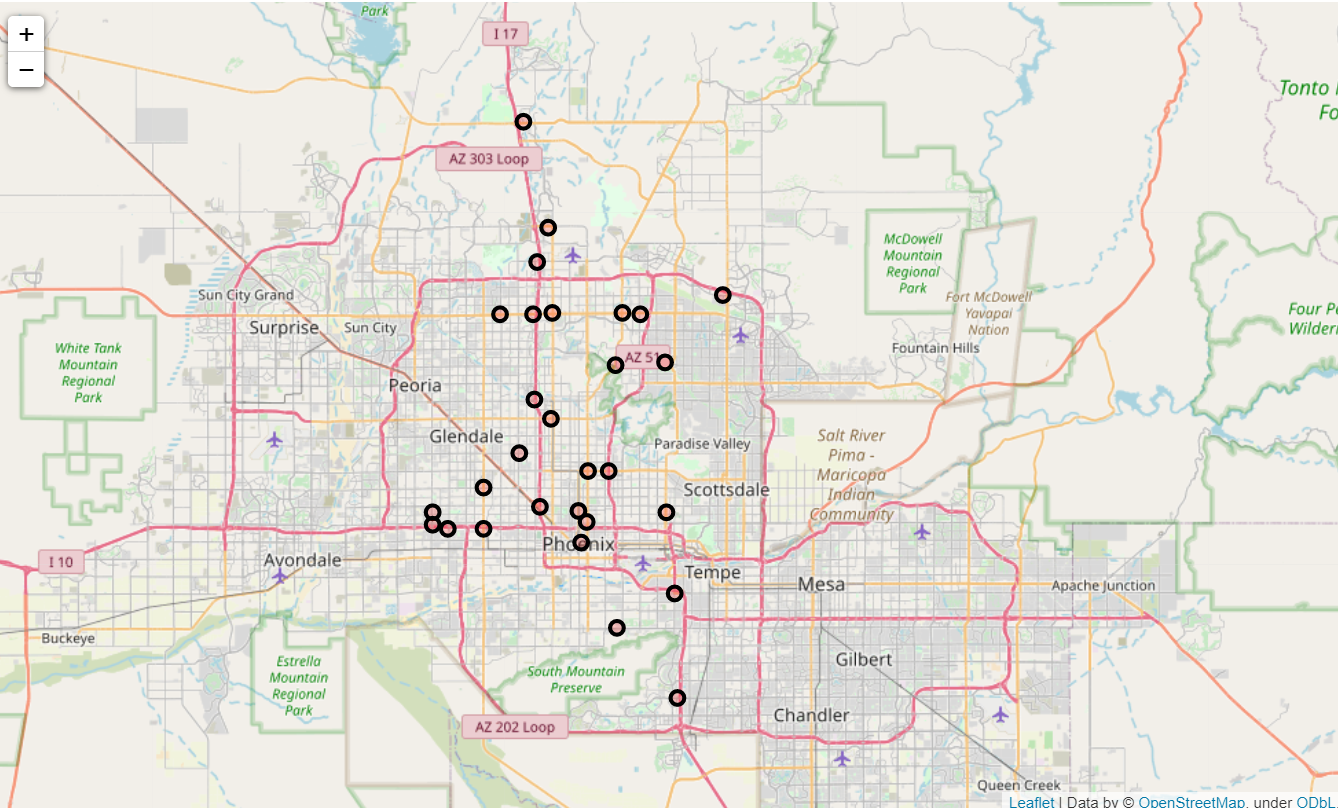
### 2.4.1 Worst brunches

- Take the brunches under 2 stars and rank by review numbers









* According to the picture, the worst two are IHOP and Denny’s which is Fast food chain
* This makes the reason why the locations are spread on the map

## 2.5 Find the population density for the city

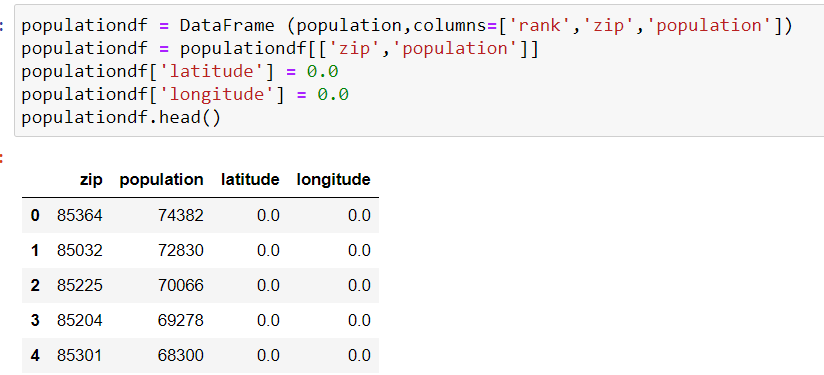
### 2.5.1 get population by zip code

source: <https://www.arizona-demographics.com/zip_codes_by_population>

1. use beautiful soup to download the table on the website

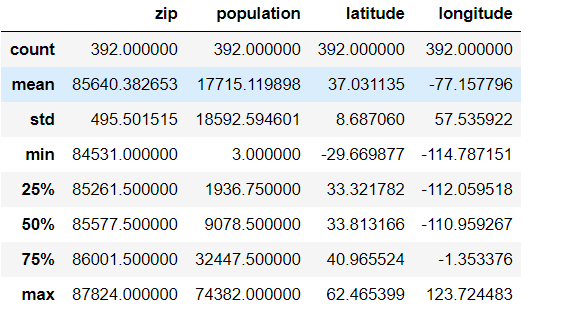


1. use the zip code to get the lat&lon



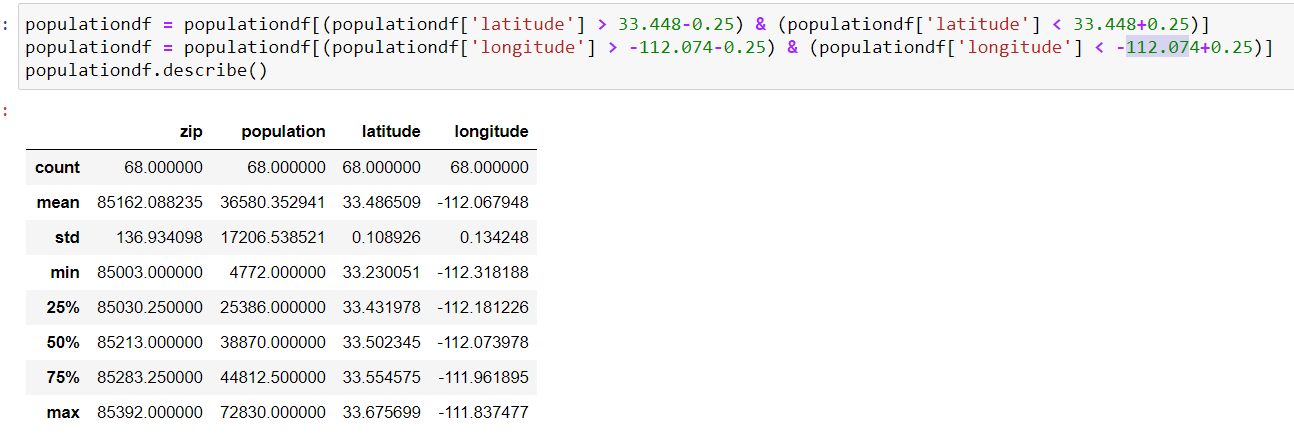


* the process takes huge time since each API request is slow.
  + do it once and download as csv file for future use

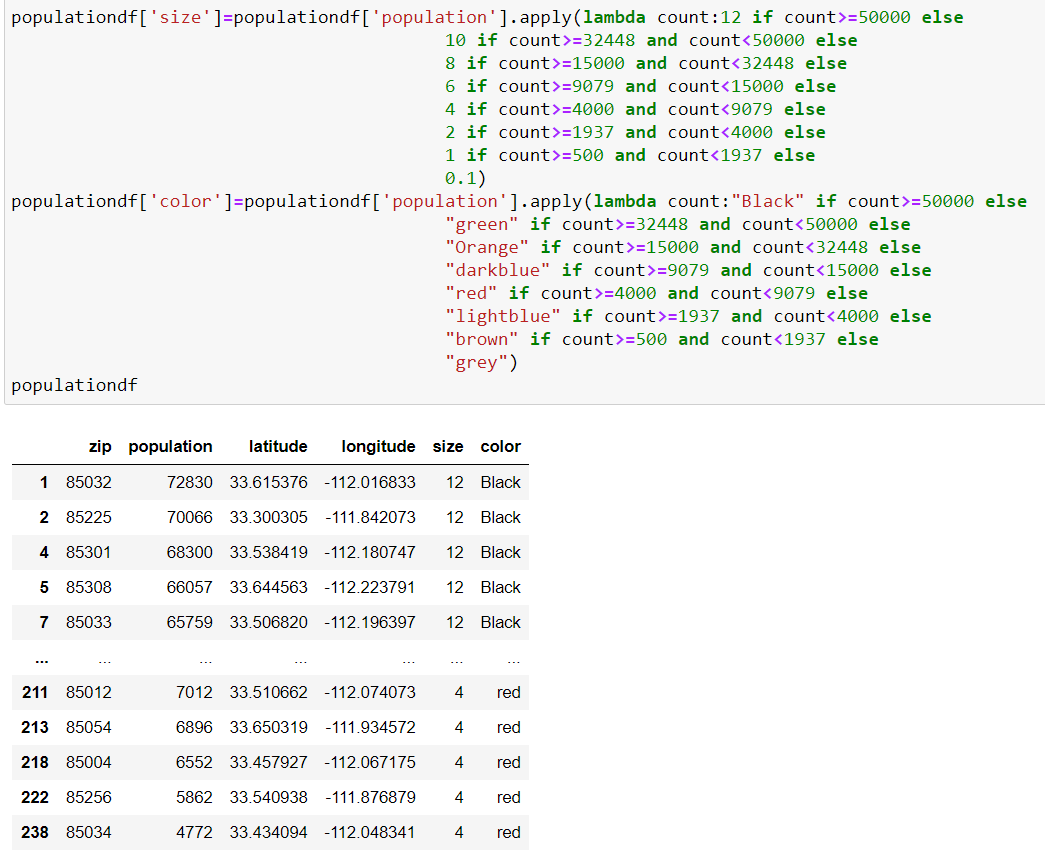


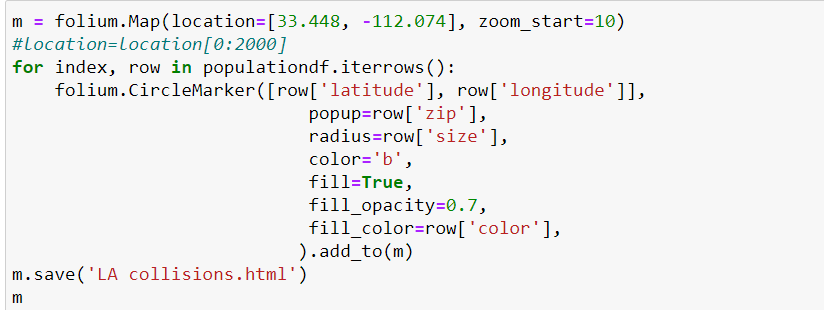
### clean the population data

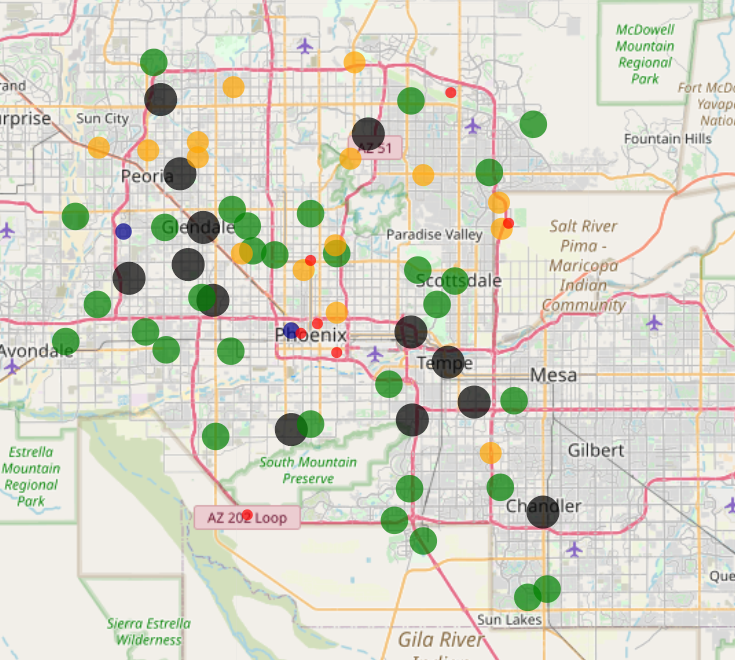
* use lat = 33.448 and lon = -112.074 as center
* radius +- 0.25

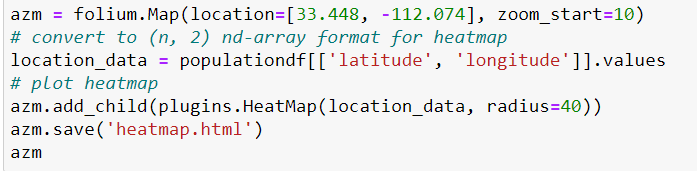


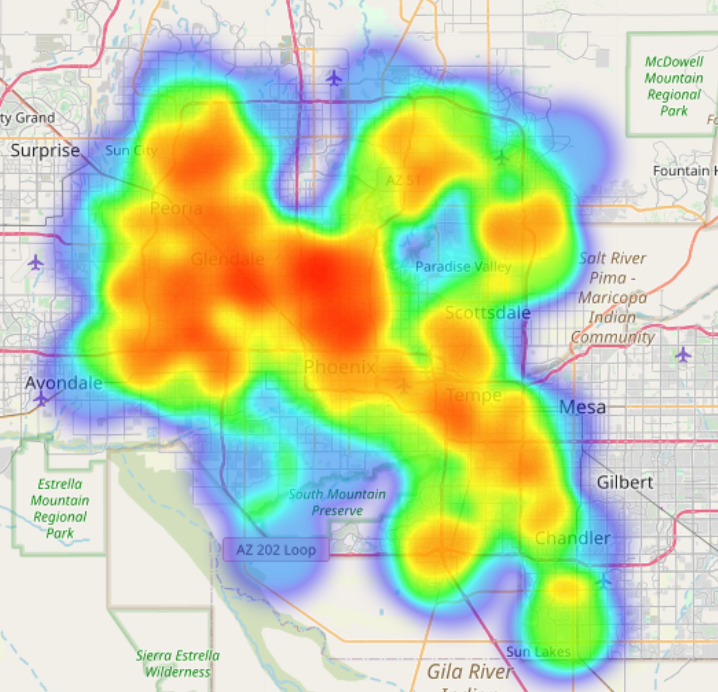
### 2.5.3 give color and value for each zip code and create a heat map







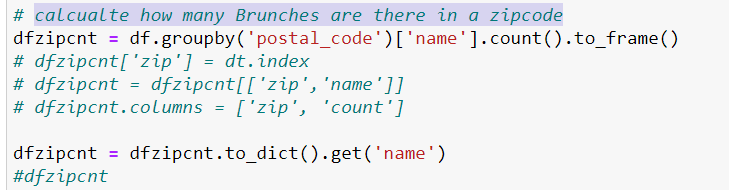




if compare with the cluster map, we can figure that most people live near cluster blue and res

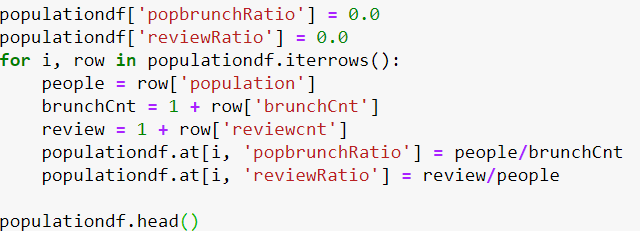
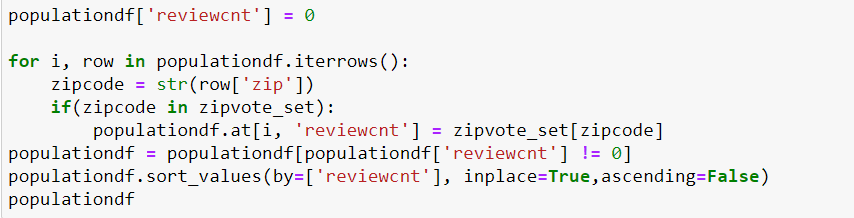
## 3 Calculation

### 3.1 calcualte how many Brunches are there in a zipcode



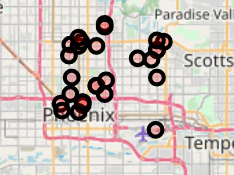
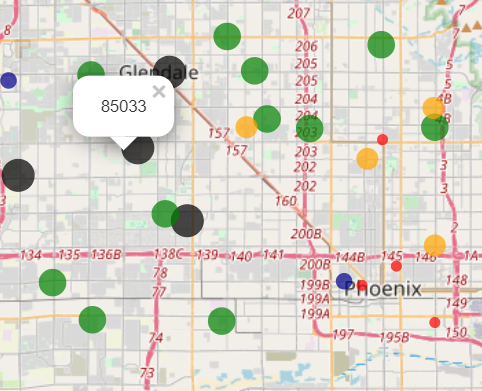
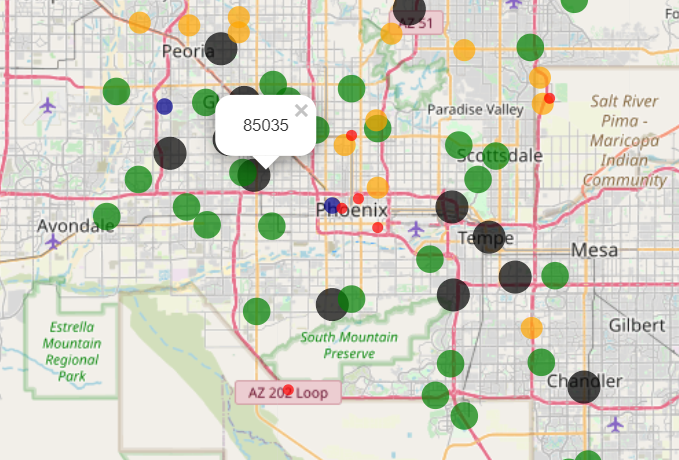


3.2 Calculate the review ratio for each location



## 4 Conclusion

* The top 5 place to a new brunch is ‘85032’, ’85008’, ‘85035’, 85033’, 85037’
* zip code 85032 has the best review ratio and population, but it has a lot brunch already
* 85035 and 85033 are not good place since they locate near the best brunches



* Compare with 85008 and 80537, 85008 has the best review ratio means a lot people nearby or work around love to eat in this place.

### 4.1 Final Conclusion

I will suggest myself to open a brunch in zip code 85008.

1. little competitors nearby
2. people nearby love to eat outside
3. although there is some brunch already, the big number of populations is still a good support for a new business. The ratio is almost 10,000: 1.