

# The Battle of Neighborhoods

Finding top US cities with highest number and density of Peruvian restaurants



# Problem



---

A foods distributor is interested in importing and delivering autochthonous Peruvian groceries and produce to Peruvian restaurants in United States.

---

He is not sure where to start his business and would like to know:

---

What are the top 5 cities by population

---

What city has the lowest restaurant mean distances to their mean coordinates, so he can maximize the food delivery process.

---

With that information, what is the best two city for his business according to his requirements above.

# Data

- Two different data sources will be used on this project to solve the problem:
- List of United States cities by population
- Source: [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population)
- Foursquare API
- Source: <https://foursquare.com/developers/>

# Metodology

- Top US cities data was retrieved from Wikipedia

2018rank		City	State[c]	2018estimate	2010Census	Change	2016 land area	2016 land area.1	2016 population density	2016 population density.1	Location
0	1	New York[d]	New York	8398748	8175133	+2.74%	301.5 sq mi	780.9 km2	28,317/sq mi	10,933/km2	40°39'49"N 73°56'19"W / 40.6635°N 73.9387°W
1	2	Los Angeles	California	3990456	3792621	+5.22%	468.7 sq mi	1,213.9 km2	8,484/sq mi	3,276/km2	34°01'10"N 118°24'39"W / 34.0194°N 118.4108°W
2	3	Chicago	Illinois	2705994	2695598	+0.39%	227.3 sq mi	588.7 km2	11,900/sq mi	4,600/km2	41°50'15"N 87°40'54"W / 41.8376°N 87.6818°W
3	4	Houston[3]	Texas	2325502	2100263	+10.72%	637.5 sq mi	1,651.1 km2	3,613/sq mi	1,395/km2	29°47'12"N 95°23'27"W / 29.7866°N 95.3909°W
4	5	Phoenix	Arizona	1660272	1445632	+14.85%	517.6 sq mi	1,340.6 km2	3,120/sq mi	1,200/km2	33°34'20"N 112°05'24"W / 33.5722°N 112.0901°W

- Then, data was wrangled to update headers and keep only the required columns

Rank	City	State	Population
1	New York[d]	New York	8398748
2	Los Angeles	California	3990456
3	Chicago	Illinois	2705994
4	Houston[3]	Texas	2325502
5	Phoenix	Arizona	1660272

# Metodology

- Foursquare API was called to get the restaurant list for each city

```
{ "New York, NY":  
  0      Pío Pío      604 10th Ave  
  1      Flor de Mayo 484 Amsterdam Ave  
  2      Pío Pío Salon 702 Amsterdam Ave  
  3      Pío Pío      210 E 34th St  
  4      Pío Pío      1746 1st Ave  
  5      Llama-San    359 Avenue of the Americas  
  6      Chirp        369 W 34th St  
  7      Llamita      80 Carmine St  
  8      Panca        92 7th Ave  
  9      Riko Peruvian Cuisine 409 8th Ave  
 10      Morocho Peruvian Fusion West 52nd St.
```

- Also the total amount of restaurants by city was retrieved

```
Total number of Peruvian restuarants in "New York, NY" = 44  
Showing Top 100  
Total number of Peruvian restuarants in Los Angeles, CA = 34  
Showing Top 100  
Total number of Peruvian restuarants in Chicago, IL = 17  
Showing Top 100  
Total number of Peruvian restuarants in Houston, TX = 17  
Showing Top 100  
Total number of Peruvian restuarants in Phoenix, AZ = 7  
Showing Top 100
```



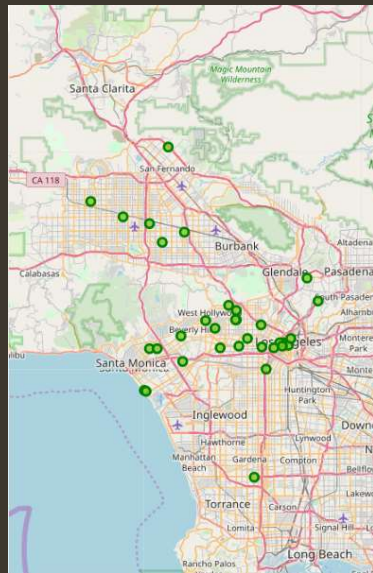
# Methodology

- Folium was used in this project to obtain a visual representation of the quantity and location of the Peruvian restaurants by city and to get geographical insights.

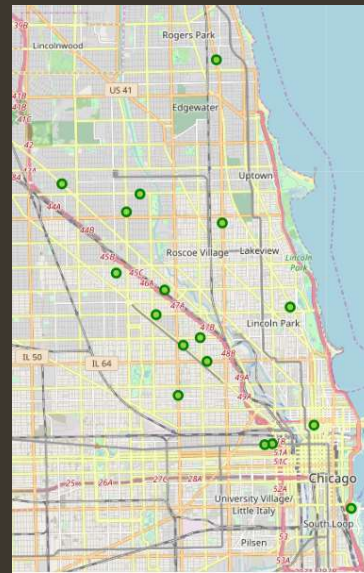
New York: 44



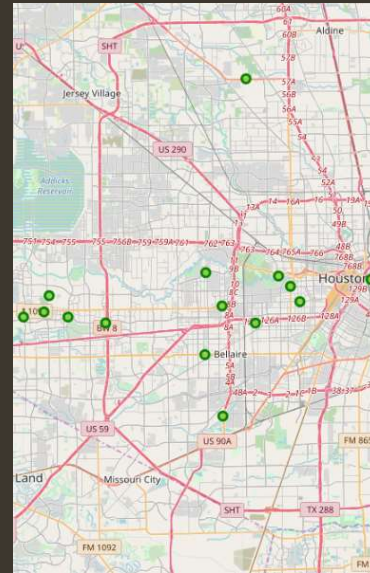
Los Angeles: 34



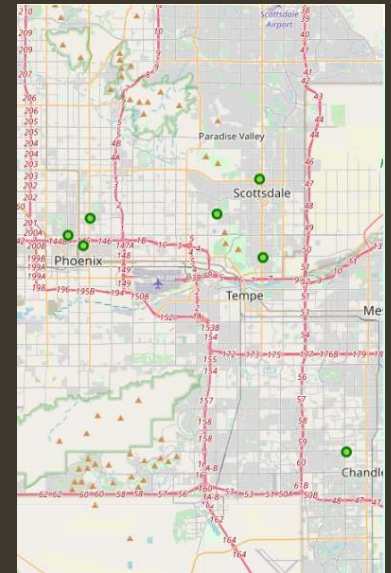
Chicago: 17



Houston: 7



Phoenix: 7



# Metodology

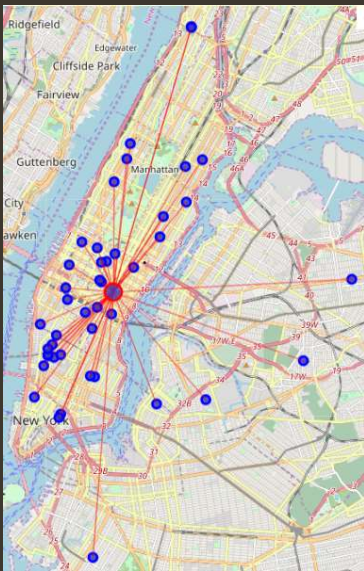
- Getting the mean location of the restaurants in each city and calculating the average distance of the restaurants to the mean coordinates.
- Results:

```
"New York, NY"  
Mean Distance from Mean coordinates  
0.03403361834108446  
Los Angeles, CA  
Mean Distance from Mean coordinates  
0.1333208138726697  
Chicago, IL  
Mean Distance from Mean coordinates  
0.03921846798477392  
Houston, TX  
Mean Distance from Mean coordinates  
0.10194325397112565  
Phoenix, AZ  
Mean Distance from Mean coordinates  
0.08928591991231158
```

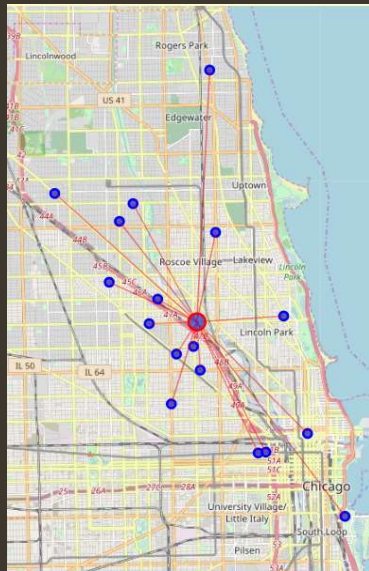
# Methodology

- Showing the Mean coordinate and the mean distances to mean coordinate(MDMC).
- Cities ordered from lower to higher mean distances:

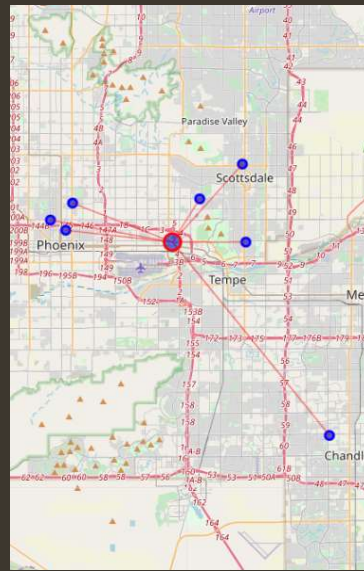
New York: 0.034



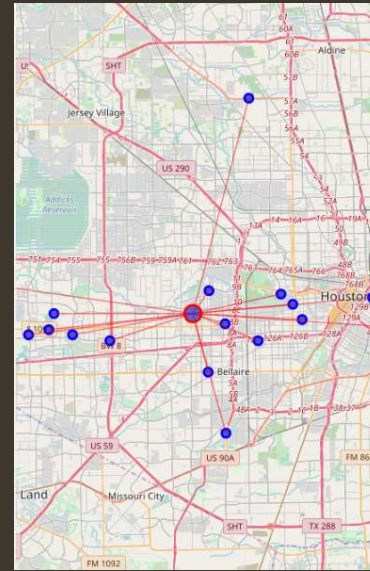
Chicago: 0.039



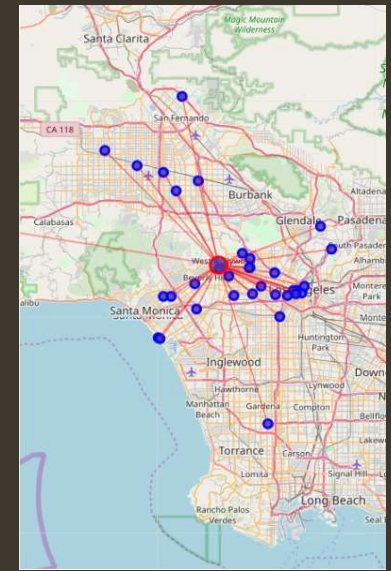
Phoenix: 0.089



Houston: 0.101



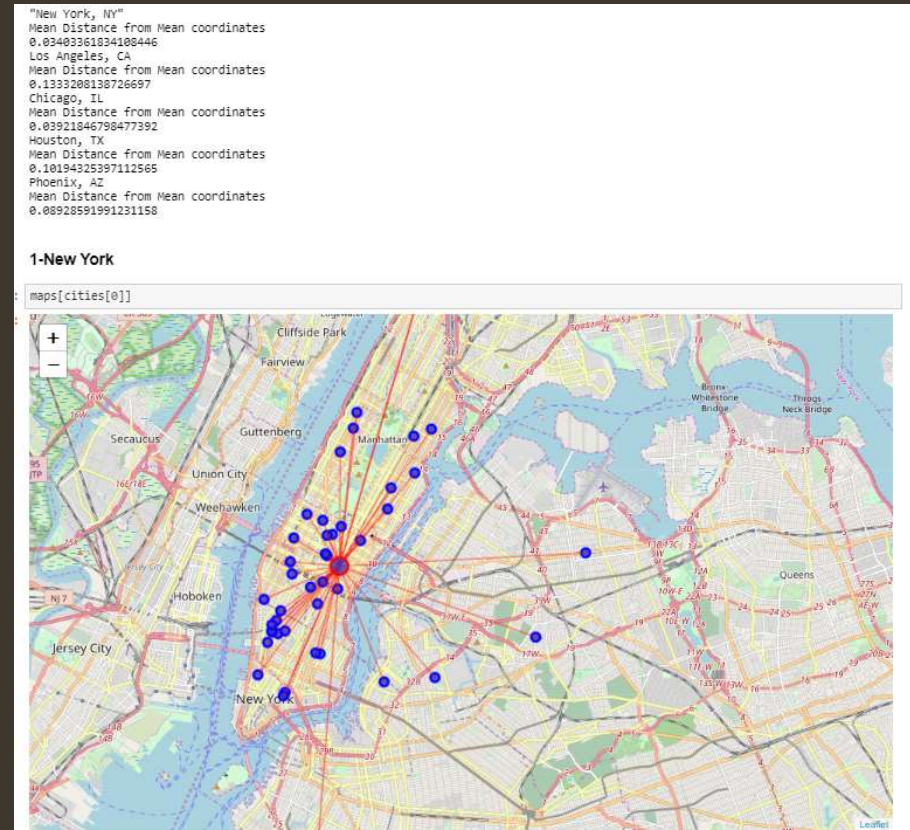
Los Angeles: 0.133





# Results

- It was determined that New York city has the lowest mean distance (0.034) compared to the other cities, and therefore the highest restaurant density among all cities.
- Another advantage New York has is the total amount of Peruvian restaurants (44), which is the highest and combined the lowest mean distance, makes it the best option for goods delivery, as restaurants are closer together, maximizing delivery routes and delivery times; and increasing potential sales, therefore this is the option recommended to the customer.



# Discussion

- Even though customer request did not mention the number of restaurants per city as a factor to decide the best option, it is a very important factor.
- By analyzing the best (lowest) mean distances, it was shown that Chicago was the second-best option; however the total amount of restaurants (170) was just half of Los Angeles amount (34), and you might think that the most restaurants you have available to deliver your goods, the more chance of getting more profits; therefore this outcome has to be mentioned to the customer, so he can take into consideration and make a more educated decision.
- Data also showed one city (Phoenix) could be ruled out right away, as it showed considerable bigger mean distances and its restaurant numbers were very low (Just 7).
- Finally, considering the highest populated cities does not mean they have the biggest amount of (Peruvian restaurants) in this case, and that is something we -as future data scientists- should take into consideration, to understand the business needs and anticipate any variables to deliver a better result than just what the customer requested.

# Conclusion

- There are many real-life situations where data along with technology can be used to find solutions to most problems; either by just analysis or by training existing data to predict future outcomes.
- Having the right tools and the knowledge can make a big difference for the future of a business; for instance, the fictional problem developed in this project got great insights from using Foursquare API and data science tools to determine the best options, therefore decision could be made based in facts, and even though, they were just a small sample of all the variables to take in consideration, they showed how important data is.
- This project was just a small taste of what can be accomplished by implementing data science to the decision-making process.

Thank you!

