Eat, Pray and Love Across Continents

Applied Data Science Capstone
Week 4 & 5

Authored by: Heba Mushtaha

1. Introduction

This project is about Traveling around the continents to the largest city in each continent. We choose the largest city in each continent as a start, and then we can evolve our dataset with more cities and data. The project shall compare neighborhoods across these cities and shall further recommend which city to visit first according to further criteria that the user sets.

2. Problem/Business Understanding

If you decide to take a tour around the world starting with the largest city in each continent, there is a challenge regarding which Continent/City to start with.
In addition, you have certain preferences (i.e. food, touristic destinations…etc.) and you can set these preferences and accordingly decide which city in each continent to visit first.

Problem as a Question

Part 1: If you decide to take a tour across continents' largest cities (largest city in each continent), which continent/city shall you start with according to the city's neighborhoods and venues?

Part 2: If you decide to take a tour across continents' largest cities (largest city in each continent), which continent/city shall you start with according to your specific preferences (i.e. food, touristic destinations…etc.)?

3. Analytic Approach

This project analysis is divided into two parts:

- **Part 1: Battle of Neighborhood** - Analyzing city's neighborhoods and their venues to get more insight about the city's neighborhoods clusters. We shall cluster the neighborhoods (equal clusters) in each city and compare and you choose the city of your preference according to the clustering results.
- **Part 2: Recommendation System** - Analyzing each city's neighborhoods venues and classifying these venues under three starting classifications (Eat, Pray, Love). You shall set your preference on a scale of 1 to 5 for these categories, and accordingly the recommender shall recommend the cities to visit in sequence.

4. Data Requirements

Working with six (6) cities at each continent, the cities within the scope of this project are:
1. North America: Mexico City, Mexico.

2. South America: Sao Paulo, Brazil.
3. Asia: Tokyo, Japan.
4. Australia: Sydney, Australia.
5. Europe: Paris, France.
6. Africa: Greater Cairo, Egypt.

We need the following Data per city:
1. Neighborhoods in each city (we used the central neighborhoods only due to size, processing and time limitations).
2. Venues and venues categories in each neighborhood.
3. User's preferences.

5. Data Collection

1. Neighborhoods in each city collected in csv files (we used Wikipedia as a source for each city's neighborhood/boroughs/wards/districts/councils).
2. Venues and venues categories in each neighborhood (we used FourSquare APIs).
3. User's preferences (this is set on a scale of 1 to 5)

6. Data Understanding & Preparation

- For All: Cleaned the dataframes for redundancy and inaccurate data.
- For the Recommender System: we further cleaned the data and dropped venues categories whose count is less than 10

7. Methodology/Modeling

Battle of Neighborhood

Clustering Technique
- Divided the neighborhoods in each city into 5 clusters (with the same *limit* and *radius* passed to all), with venues and venues categories in each cluster.
- Add all the neighborhoods' clusters to one dataframe so that we have one dataframe consisting of all neighborhoods (with venues and venues categories in each city) in the 6 cities.

Recommendation System

Classification Method
- Classified each venue category as either Eat/Pray/Love

- Add all the results of venues lying under Eat/Pray/Love in one dataset for comparison and using in the recommender system.

  Note: _This is a "biased" classification where we set the Eat/Pray/Love according to our criteria. Further enhancement will be to use un-biased classification_

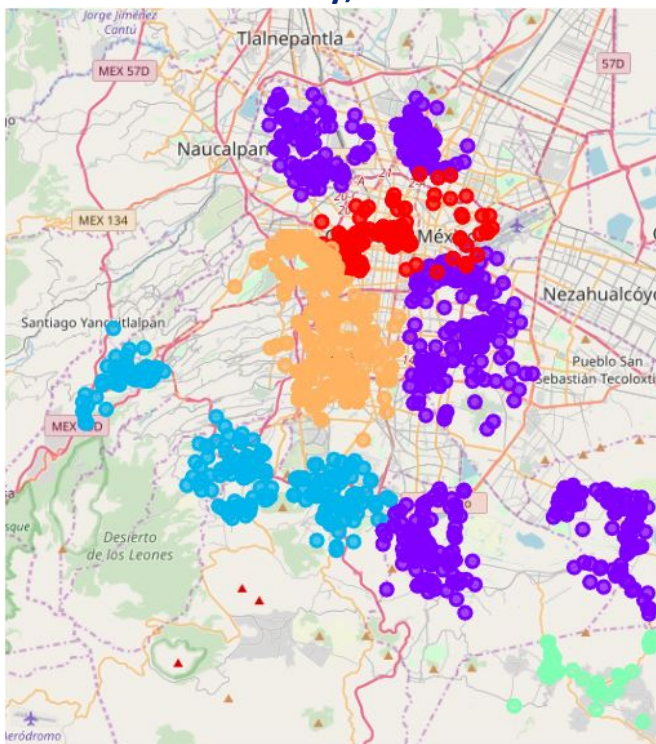### Content-based Recommendation System

- Using the output of the classification method above, we now set the user's preference for Eat/Pray/Love. for example, user gives the following scores for Eat/Pray/Love:
  - Eat equals 3
  - Pray equals 5
  - Love equals 4
- Add the user's preferences to the user's profile; using the output of the classification method and content-based recommendation system the user gets recommendations for the cities to visit by order.
- We also repeated this exercise for many users.
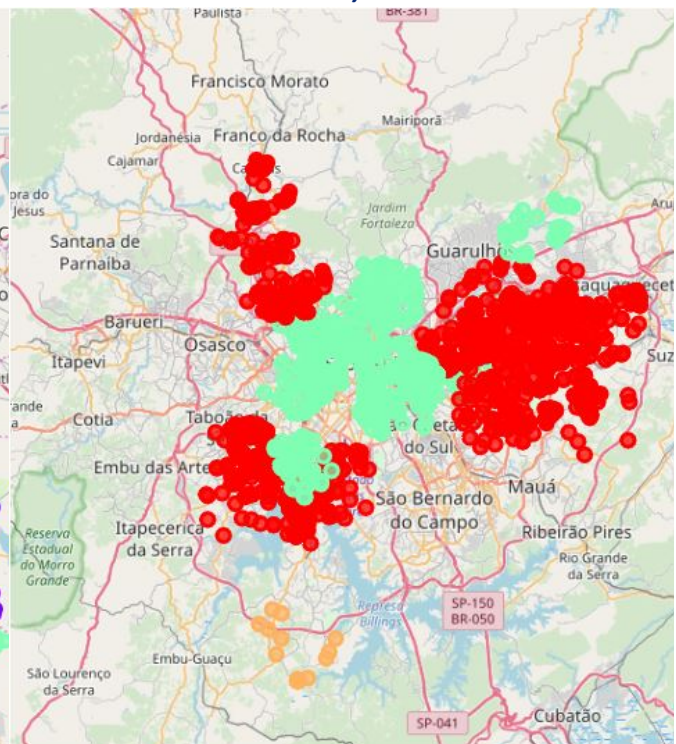
8. Results

Battle of Neighborhood

- The following are the generated clusters per city

| Mexico City, Mexico | Sao Paulo, Brazil |

Tokyo, Japan

Sydney, Australia

Paris, France

Greater Cairo, Egypt

- Clusters in all neighborhoods in all cities are added to one dataframe
- Arranged the dataframe descending according to number of venues (my preference)
- Results for cities with largest number of venues were Sao Paulo then Greater Cairo.

- Resulting dataframe is displayed on the notebook (not here as there are 30 clusters – 6 cities, 5 clusters each)

Recommendation System

From the results of classifying venues by Eat, Pray and Love, the following results are generated:

	Continent Name	Country	City	Eat	Pray	Love
0	North America	Mexico	Mexico City	728	120	23
1	South America	Brazil	Sao Paulo	1153	148	130
2	Asia	Japan	Tokyo	1292	153	140
3	Australia	Australia	Sydney	1061	169	197
4	Europe	France	Paris	816	341	178
5	Africa	Egypt	Greater Cairo	629	348	126

Corresponding chart for these results is as follows:
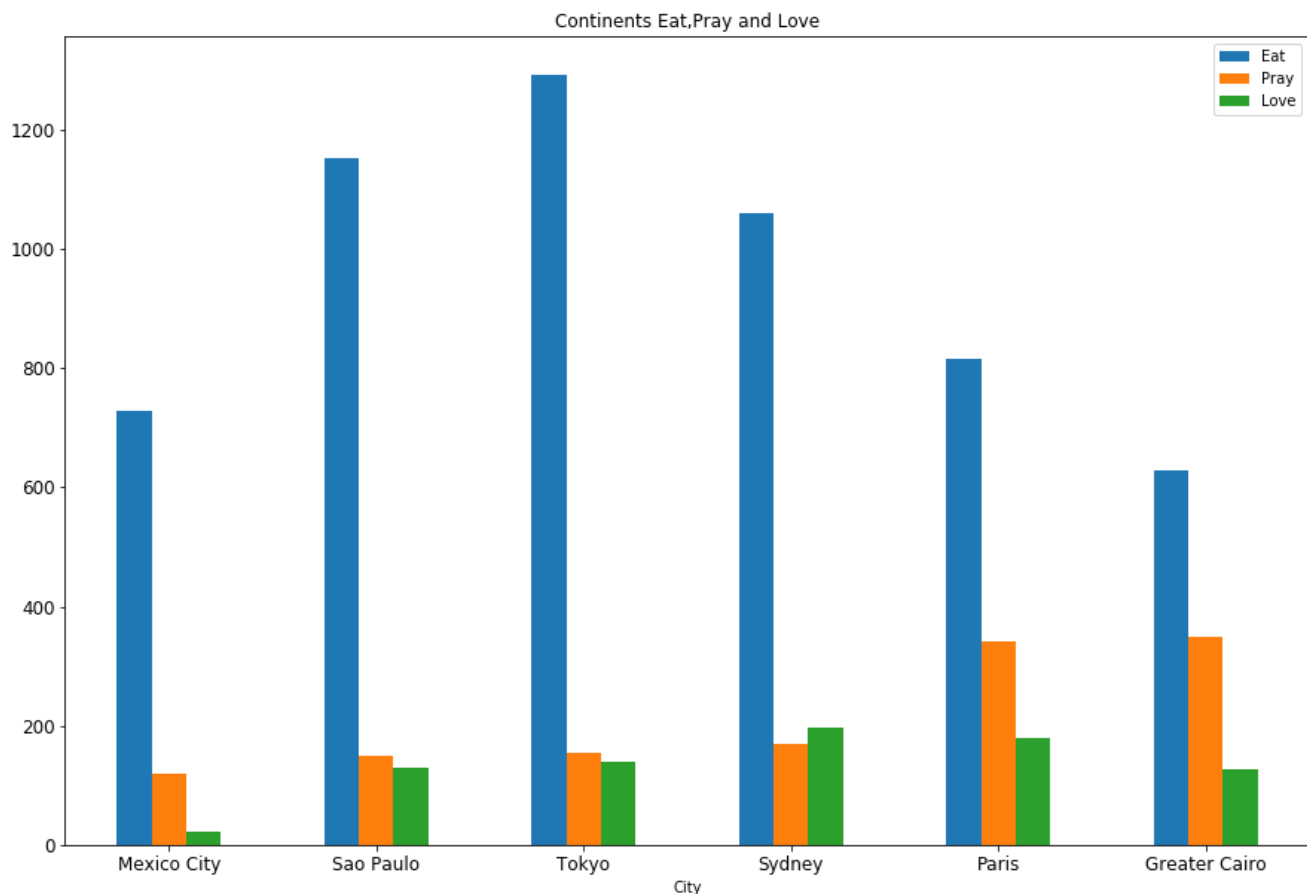
This chart reflects that if you are fan of visiting countries to try new food, then Tokyo comes in the first place.

If you are fan of visiting sites, parks, museums and similar, then you should consider Paris and Cairo.

6

If you are fan of shopping and anything that is fun i.e. shopping, spa and similars, then you should consider Sydney and Paris.

For one-user preference, with the user entering a score for Eat, Pray and Love (on a scale of 1 to 5), the following results are generated accordingly:

	Continent Name	Country	City	Eat	Pray	Love	MyScore
2	Asia	Japan	Tokyo	1292	153	140	433.416667
4	Europe	France	Paris	816	341	178	405.416667
3	Australia	Australia	Sydney	1061	169	197	401.333333
1	South America	Brazil	Sao Paulo	1153	148	130	393.250000
5	Africa	Egypt	Greater Cairo	629	348	126	344.250000
0	North America	Mexico	Mexico City	728	120	23	239.666667

One-User Preferences
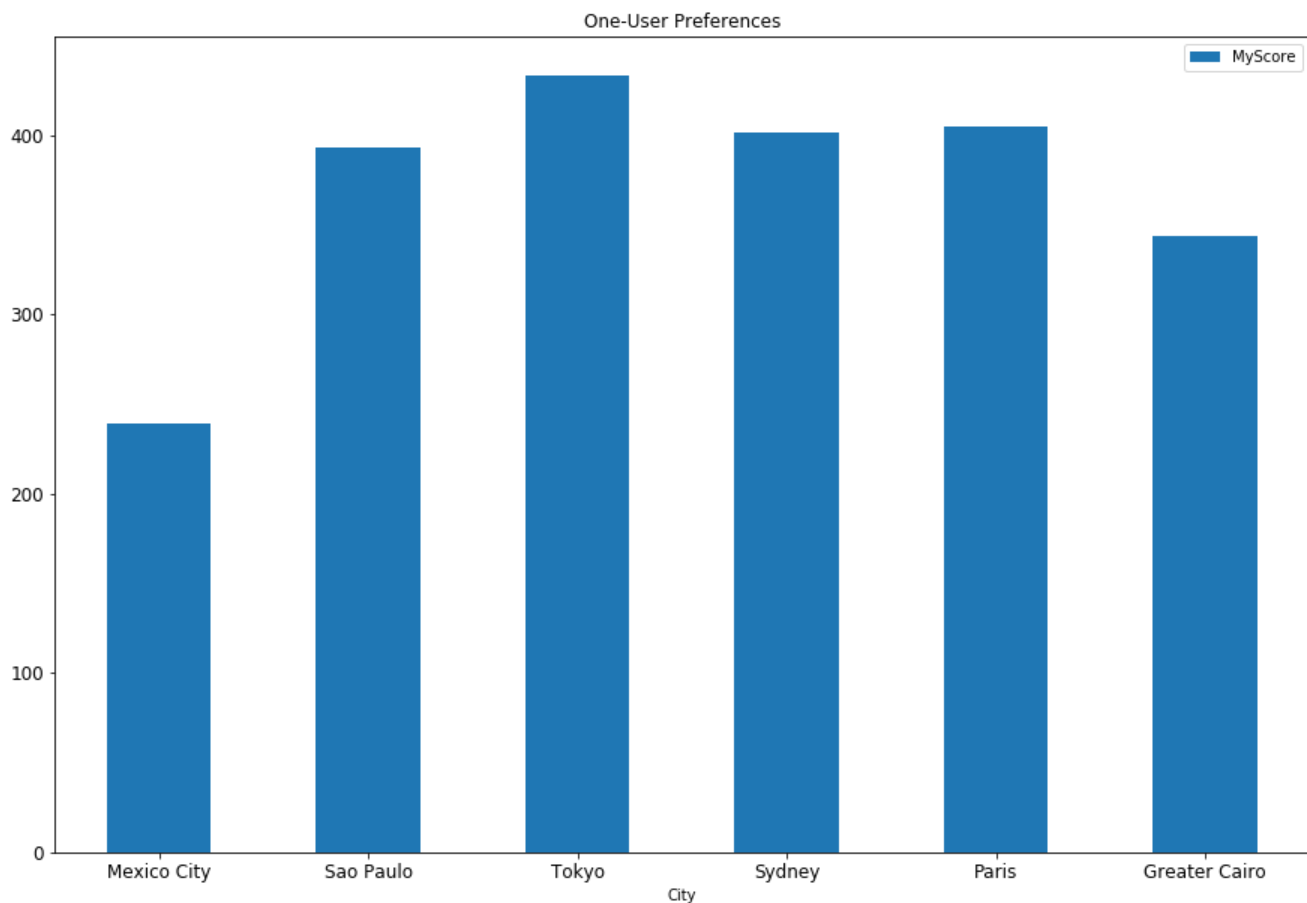
Therefore, the results will be as follows:

Results here shows that the city that is recommended to visit first (According to my criteria) is Tokyo followed by Paris (results are available on the notebook too).

```
According to your user profile and your rating for Eat, Pray and Love, You can visit the following cities by the foll
owing order:
-------------------------------------------------------------------------------------------------------------------
---------------
2                   Tokyo in Japan at Asia
4                   Paris in France at Europe
3           Sydney in Australia at Australia
1       Sao Paulo in Brazil at South America
5           Greater Cairo in Egypt at Africa
0       Mexico City in Mexico at North America
```

For Multi –User Recommendation System, the following diagram also shall be used by recommendation system to group people of similar preferences and consequently recommend other cities.

Users' Preferences for Cities based on their profiles
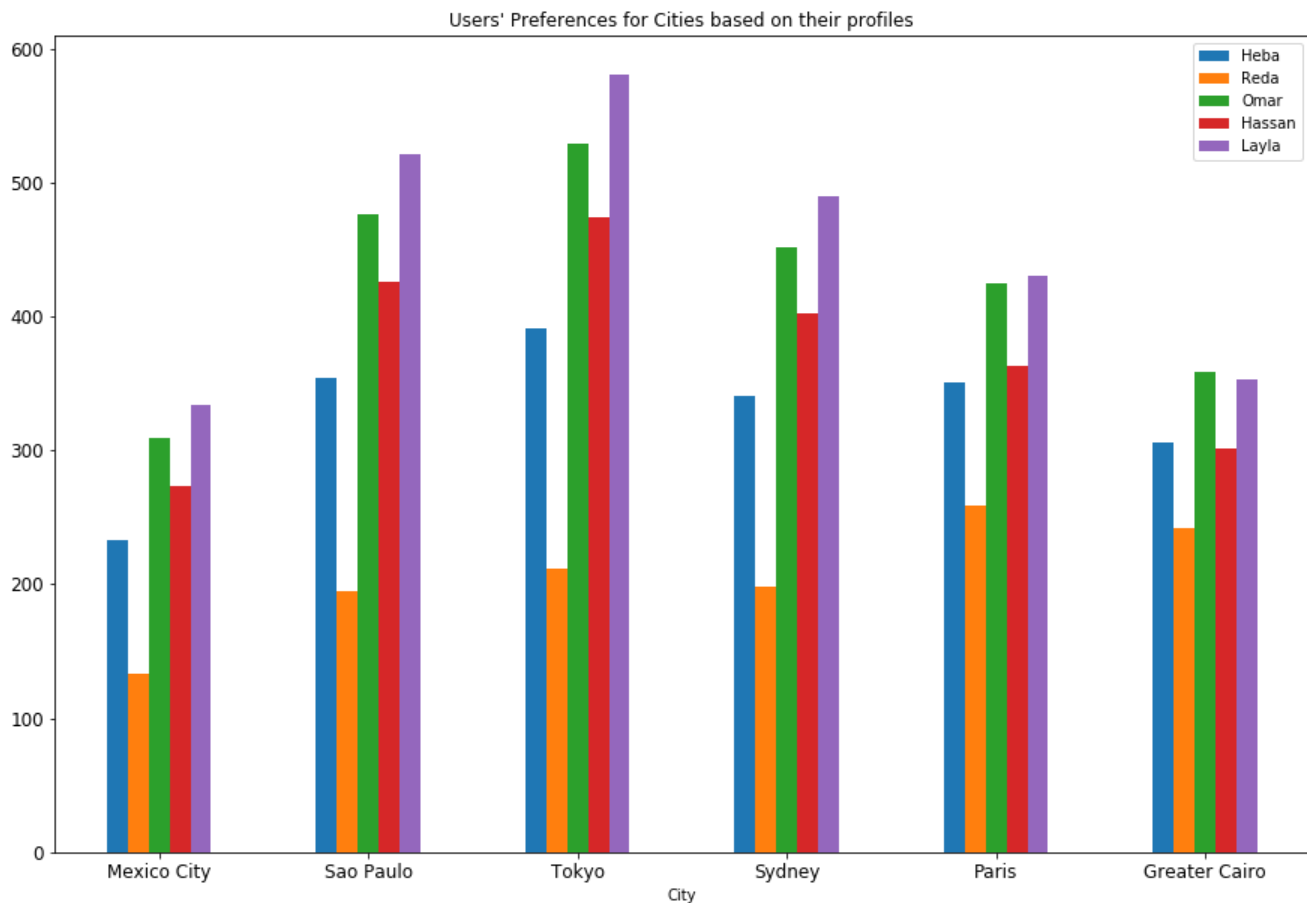
9. Discussion

From the results, there are some observations:
- Neighborhoods' clustering gives more insight about the neighborhoods and venues in the city.
- Adding user preferences of Eat, Pray and Love – will give recommender system input so that a route of cities suiting your preferences shall be recommended.
- You can compare the first city you chose from the clustering exercise against the first city recommended from the recommender system, and you might be surprised!

10. Conclusion

As a wrap-up for this report that the problem that we are trying to solve is to help you in choosing a travelling destination based on your own preferences. Preferences can be as follows:

- Battle of Neighborhoods – choosing among neighborhood clusters in each city, according to your criteria i.e. number of venues, number of venue categories, and number of neighborhoods included in the cluster (So that you can visit more neighborhood than other clusters).
- Recommendation System – After analyzing each city's venues, you can set your preferences and accordingly a route of the cities to visit will be recommended accordingly.