Must-visit places for foodies in European capitals

BATTLE OF THE NEIGHBOURHOODS

Capstone Project

Authored by: Ana Costa

1. Introduction / Business problem

A foodie is a person who eats food not only out of hunger but due to their interest or hobby and passionate about food. Different cultures and cuisines all over the World provide exquisite experiences and typical foodie interests and activities include travelling around the world and discover a whole new set of cultural gastronomical experiences. Therefore, this project aims to look for food-related venues with particular restaurants in the capitals of European countries, cluster them and ultimately reveal the different cities in terms of food venues, even if they have closer geography.

So, if you're a food lover planning your next big adventure, then we've got you covered. Whether you fancy indulging in the finest fresh seafood or simply sampling new twists on your favourite cuisines, check out our suggestions to the best foodie destinations in Europe to get on your bucket list.

2. Data

To meet the abovementioned objective, two data sources were used.

a) **The first one**, contain the coordinates of the world's capitals, available on: http://techslides.com/list-of-countries-and-capitals. The raw dataset has 245 rows x 6 columns, comprising the index, Country Name, Capital Name, Capital Latitude, Capital Longitude, Country Code and Continent Name. In Figure 1, there is a snapshot of the first five rows of the raw dataset. This dataset was then filtered to create a dataset containing only the capitals located at the European Continent (further described in the Methodology), containing 49 rows x 6 columns. Note that, there were only considered the countries that are listed as "Sovereign states" in this webpage: https://en.wikipedia.org/wiki/Category:Capitals_in_Europe).

	Country Name	Capital Name	Capital Latitude	Capital Longitude	Country Code	Continent Name
0	Afghanistan	Kabul	34.516667	69.183333	AF	Asia
1	Aland Islands	Mariehamn	60.116667	19.900000	AX	Europe
2	Albania	Tirana	41.316667	19.816667	AL	Europe
3	Algeria	Algiers	36.750000	3.050000	DZ	Africa
4	American Samoa	Pago Pago	-14.266667	-170.700000	AS	Australia

Figure 1 – Snapshot of the first five rows from the raw data gathered from http://techslides.com/list-of-countries-and-capitals.

b) The second one comprised the use of Foursquare API, which is a location data provider used to make RESTful API calls to get data about venues. In this project, the venues were retrieved for the top 15 food venues in a 500m radius from each European Capital city center for the category "Food". The results output included "Venue", "Venue Latitude", "Venue Longitude" and "Venue Category" and retrieved a total of 621 unique venues, 641 total venues and 102 unique subcategories within the broader category "Food".

3. Methodology

3.1. Download and Explore the World's Coordinates Dataset to get only the coordinates of the City Center of Capitals within the European Continent.

To download the raw dataset, we used read_html from pandas to read in tables from a html, determined the index of raw data and created a data frame. Then, we filtered the data to get a new dataframe containing only the data from the European Continent (i.e. filter the dataframe to have 'Continent Name' that contains 'Europe'), specifically recognized as European Sovereign States. The curated dataframe returned a total of 49 rows x 6 columns, comprising the index, Country Name, Capital Name, Capital Latitude, Capital Longitude, Country Code and Continent Name (Figure 2).

	Country Name	Capital Name	Capital Latitude	Capital Longitude	Country Code	Continent Name
0	Albania	Tirana	41.316667	19.816667	AL	Europe
1	Andorra	Andorra la Vella	42.500000	1.516667	AD	Europe
2	Armenia	Yerevan	40.166667	44.500000	AM	Europe
3	Austria	Vienna	48.200000	16.366667	AT	Europe
4	Azerbaijan	Baku	40.383333	49.866667	AZ	Europe

Figure 2 – Snapshot of the first five rows of the curated dataframe.

We then used the curated dataframe to visualize a map of European Continent with superimposed tag of the respective capital city center, by using the library “folium” (Figure 3).

Figure 3 – Map of European Continent with superimposed tag of Capital’s City Center.

3.2. RESTful API Calls to Foursquare

We used the Foursquare API to explore the European Capitals and segment them, which required the proper setting of “CLIENT_ID”, “CLIENT_SECRET” and “VERSION”.
Although a wide set of venue categories are available, we only explored the ones from “Food” category, by addressing the categoryId in the URI.
As during the labs sessions we had the opportunity to further understand and explore GET requests, we did not start by exploring the returns from only one capital. So, considering the aim of the project, we readily moved to the segmentation of all the 49 European capitals considered,

by “Food”. In order to create an API request URL for each capital, a function “getNearbyFood” was created, with radius of 500m and LIMIT of 15 (maximum venues returned for each capital).
In this function, after the GET request, only the relevant information for each nearby venue was retained (venue name, venue latitude, venue longitude and venue category), by appending it to a list. This list was further unfolded to append it to the dataframe returned by the function. Also, the python library “Pickle” was needed to serialise the information retrieved from GET request, which was then deserialised to get an exact python object structure. Figure 4 illustrates a snapshot of the returned dataframe.

	Capital Name	Capital Latitude	Capital Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tirana	41.316667	19.816667	Sophie Caffe & Snacks	41.318426	19.814562	Coffee Shop
1	Tirana	41.316667	19.816667	Era Restaurant & Pizzeria	41.320253	19.814534	Pizza Place
2	Tirana	41.316667	19.816667	Sophie Caffe & Snacks	41.317183	19.818986	Café
3	Tirana	41.316667	19.816667	D'angelo coffee shop	41.319406	19.818964	Coffee Shop
4	Tirana	41.316667	19.816667	Hotel Dinasty	41.314882	19.812469	Bed & Breakfast

Figure 4 – Dataframe resulting from the GET request to the Foursquare API.

3.3. Data cleaning

In an exploratory data analysis, we determined how many unique categories and venues there are in the dataframe resulting from aforementioned methodology. So, we retrieved 621 unique venues, 641 total venues and 102 unique subcategories within the broader category “Food”. This means that we have repeated venues in the dataset. To clean the dataset, we dropped out the duplicated venues, obtained a total unique venues and total venues of 621 and 102 unique subcategories.

3.4. Analysis of each European Capital

To analyse the most common type of food served within the 500m vicinity of the European capitals’ city centers, we used a “one hot encoding” function, which converts category variables into a “friendly” format for algorithmic further analysis. We obtained a new dataframe (named “europe_onecode”), add the capitals’ column back to the dataframe and move it to the first column. The size of this new dataframe is 621 x 103.

The top 10 food categories in European capitals were found by counting their occurrences (Figure 5).

```
# Let's find out the top 10 food categories in European capitals
venue_counts_described = venue_counts.describe().transpose()
venue_top10 = venue_counts_described.sort_values('max', ascending=False)[0:10]
venue_top10
```

	count	mean	std	min	25%	50%	75%	max
Café	48.0	1.833333	1.860374	0.0	0.0	1.5	3.00	8.0
Italian Restaurant	48.0	0.541667	1.071057	0.0	0.0	0.0	1.00	5.0
Coffee Shop	48.0	1.145833	1.352532	0.0	0.0	1.0	2.00	5.0
Restaurant	48.0	1.208333	1.147770	0.0	0.0	1.0	2.00	5.0
Japanese Restaurant	48.0	0.145833	0.618495	0.0	0.0	0.0	0.00	4.0
Fast Food Restaurant	48.0	0.250000	0.635811	0.0	0.0	0.0	0.00	3.0
Spanish Restaurant	48.0	0.062500	0.433013	0.0	0.0	0.0	0.00	3.0
Tea Room	48.0	0.145833	0.504852	0.0	0.0	0.0	0.00	3.0
Pizza Place	48.0	0.500000	0.771845	0.0	0.0	0.0	1.00	3.0
Ice Cream Shop	48.0	0.333333	0.663111	0.0	0.0	0.0	0.25	3.0

Figure 5 – Top 10 food categories in European capitals

We then grouped the rows by capital name and by taking the mean of the frequency of occurrence of each category. We sorted the venues in descending order and then create the new dataframe and displayed the top 10 venues for each European capital (Figure 6).

```
europe_venues_sorted
```

	Capital Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amsterdam	Restaurant	Café	Coffee Shop	Ice Cream Shop	Pizza Place	Italian Restaurant	Breakfast Spot	Australian Restaurant	Hotel	French Restaurant
1	Andorra la Vella	Restaurant	Portuguese Restaurant	Coffee Shop	Diner	Café	Breakfast Spot	BBQ Joint	Bakery	Snack Place	Clothing Store
2	Ankara	Café	Coffee Shop	Tea Room	Chocolate Shop	Comfort Food Restaurant	Kebab Restaurant	Steakhouse	Fast Food Restaurant	Falafel Restaurant	General Entertainment
3	Athens	Coffee Shop	Café	Dessert Shop	Bar	Souvlaki Shop	Sushi Restaurant	Restaurant	Other Nightlife	Wine Bar	Vegetarian / Vegan Restaurant
4	Baku	Tea Room	Café	Restaurant	Seafood Restaurant	Modern European Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Eastern European Restaurant	Afghan Restaurant	Diner

Figure 6 – Dataframe displaying the 10 top venues for each European Capital.

3.5. Clustering

On this project, we used “K-means”, an unsupervised machine learning algorithm, which aggregates data points according to the similarity, to aggregate the different European capitals according to the food venues. We started by determining the optimal number of clusters, using the elbow method. We varied the k from 1 to 29 (Figure 7). We considered a k = 6 to pursue our analysis.

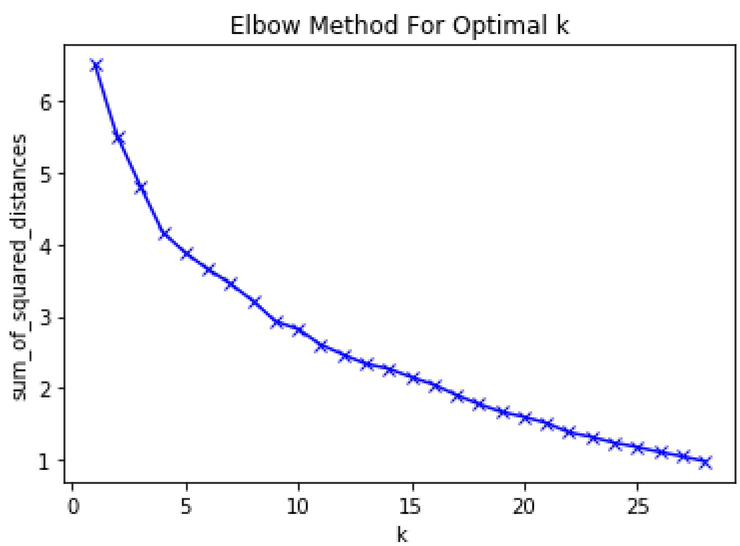

Figure 7 – Elbow method to choose the optimal k, for our case study.

Then, we added the cluster labels to the dataframe to get the segmentation of the European capitals based upon the most common venues in its vicinity. Following this, ‘europe_venues_sorted’ was merged with ‘df_final’ to add the Capital Name, Capital Latitude and Capital Longitude for each capital (Figure 8).

	Cluster Labels	Capital Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Country Name
0	1	Amsterdam	Restaurant	Café	Coffee Shop	Ice Cream Shop	Pizza Place	Italian Restaurant	Breakfast Spot	Australian Restaurant	Hotel	French Restaurant	Netherlands
1	3	Andorra la Vella	Restaurant	Portuguese Restaurant	Coffee Shop	Diner	Café	Breakfast Spot	BBQ Joint	Bakery	Snack Place	Clothing Store	Andorra
2	1	Ankara	Café	Coffee Shop	Tea Room	Chocolate Shop	Comfort Food Restaurant	Kebab Restaurant	Steakhouse	Fast Food Restaurant	Falafel Restaurant	General Entertainment	Turkey
3	3	Athens	Coffee Shop	Café	Dessert Shop	Bar	Souvlaki Shop	Sushi Restaurant	Restaurant	Other Nightlife	Wine Bar	Vegetarian / Vegan Restaurant	Greece
4	1	Baku	Tea Room	Café	Restaurant	Seafood Restaurant	Modern European Restaurant	Middle Eastern Restaurant	Chinese Restaurant	Eastern European Restaurant	Afghan Restaurant	Diner	Azerbaijan

Figure 8 - merge europe_venues_sorted with df_final to add latitude/longitude for each European capital.

Finally, we visualized the results using python library "folium", showing the segmentation of European capitals according to food category (Figure 9).

Figure 9 - segmentation of European capitals according to food category.

4. Results

4.1. Analysis of Cluster 0

This cluster comprises the capitals: Brussels, Kyiv, Bucharest, Athens, Dublin, Tallinn, Lisbon, Tbilisi, Madrid, Skopje, Riga, Nicosia, Andorra la Vella, London and Valletta. The 1st most common venues are coffee shops and restaurants, with 7 and 4 occurrences, respectively. Also, the 2nd most common venue are Cafés with 3 occurrences (Figure 10).

```
--------------------------------------------
Coffee Shop             7
Restaurant              4
Sandwich Place          1
Cocktail Bar            1
Portuguese Restaurant   1
Spanish Restaurant      1
Name: 1st Most Common Venue, dtype: int64
--------------------------------------------
Café                    3
Romanian Restaurant     1
Greek Restaurant        1
Portuguese Restaurant   1
Bakery                  1
Bar                     1
English Restaurant      1
Thai Restaurant         1
Coffee Shop             1
Chinese Restaurant      1
Tapas Restaurant        1
Dessert Shop            1
Indian Restaurant       1
Name: 2nd Most Common Venue, dtype: int64
--------------------------------------------
```

Figure 10 – Analysis of cluster 0.

4.2. Analysis of Cluster 1

This cluster comprises only one European capital: Bern. The 1st most common venues is Japanese Restaurant. Furthermore, the 2nd most common venue are Wings Joint (Figure 11).

```
--------------------------------------------
Japanese Restaurant    1
Name: 1st Most Common Venue, dtype: int64
--------------------------------------------
Wings Joint    1
Name: 2nd Most Common Venue, dtype: int64
--------------------------------------------
```

Figure 11 – Analysis of cluster 1.

4.3. Analysis of Cluster 2

This cluster comprises the capitals: Vienna, Podgorica, Helsinki, Ankara, Bratislava and Paris. The 1st most common venues are Cafés, with 5 occurrences. The 2nd most common venue are Coffee Shops with 2 occurrences (Figure 12).

```
------------------------------------------
Café                     5
Japanese Restaurant      1
Name: 1st Most Common Venue, dtype: int64
------------------------------------------
Coffee Shop             2
Italian Restaurant      1
Café                    1
Bar                     1
Food Court              1
Name: 2nd Most Common Venue, dtype: int64
------------------------------------------
```

Figure 12 – Analysis of cluster 2.

4.4. Analysis of Cluster 3

This cluster comprises the capitals: Vatican City, Luxembourg, Chisinau, Tirana and Copenhagen. The 1st most common venues are Italian Restaurants, with 2 occurrences, respectively. Also, the 2nd most common venue are Pizza Places with 2 occurrences (Figure 13).

```
------------------------------------------
Italian Restaurant      2
Bakery                  1
Coffee Shop             1
Pizza Place             1
Name: 1st Most Common Venue, dtype: int64
------------------------------------------
Pizza Place              2
Romanian Restaurant      1
Café                     1
Sandwich Place           1
Name: 2nd Most Common Venue, dtype: int64
------------------------------------------
```

Figure 13 – Analysis of cluster 3.

4.5. Analysis of Cluster 4

This cluster comprises the capitals: Zagreb, Oslo, Amsterdam, Ljubljana, Reykjavik, Minsk, Busapest, Moscow, Rome, Baku, Stockholm, Berlin, Prague, Sarajevo, Sofia, Monaco, Yerevan and Warsaw. The 1st most common venues are coffee shops, Fast food restaurant and bakery, respectively with 9, 2 and 2 occurrences. The 2nd most common venue are Cafés, Ice Cream shop and Pizza places with, respectively, 4, 3 and 2 occurrences (Figure 14).

```
Café                       9
Fast Food Restaurant       2
Bakery                     2
Restaurant                 1
French Restaurant          1
Cocktail Bar               1
Food Court                 1
Tea Room                   1
Name: 1st Most Common Venue, dtype: int64
---------------------------------------------
Café                            4
Ice Cream Shop                  3
Pizza Place                     2
Steakhouse                      1
Coffee Shop                     1
Italian Restaurant              1
Fast Food Restaurant            1
Bakery                          1
Jewish Restaurant               1
Restaurant                      1
German Restaurant               1
South American Restaurant       1
Name: 2nd Most Common Venue, dtype: int64
---------------------------------------------
```

Figure 14 – Analysis of cluster 4.

4.6. Analysis of Cluster 5

This cluster includes Belgrade, Vilnius and Vaduz. The 1st most common venue is restaurant. The 2nd most common food venues are seafood restaurant, pizza place and bed & breakfast (Figure 15).

```
Restaurant    3
Name: 1st Most Common Venue, dtype: int64
---------------------------------------------
Seafood Restaurant    1
Pizza Place           1
Bed & Breakfast       1
Name: 2nd Most Common Venue, dtype: int64
---------------------------------------------
```
Figure 15 – Analysis of cluster 5.

5. Discussion / Conclusions

The analysis of food venues around the capital city centers of European Continent during this project showed that central European countries are clearly different from the western and southern Europe. However, coffee shops, cafés and restaurants are on the top list of the most common categories of food venues.
A complementary analysis to further explore this subject could include the consideration of the different cultural backgrounds and peoples' migration patterns.

6. References

Notebook created by Eklavya Saxena for the 'Applied Data Science Capstone' course on Coursera