Introduction:

“Amaretto” is a global chain of Italian Restaurants, headquartered at New York, with multiple branches across USA. In the last 5 years, they have also forayed into the international market, with new branches opening in Latin America as well as Europe. Currently, they run operation across 7 countries. With the appointment of the new CEO, they have developed an aggressive expansion strategy and are looking to further extend their business across Europe, with Berlin being a key target. Due to its economic status as well as the central location, Berlin is planned to become the European Headquarters in the coming years.

Business Problem:

Due to the strategic importance of Berlin, the location to be chosen for the restaurant is of vital importance. There is already a stiff competition in the food & beverages industry and Berlin being the economic hotspot of entire Europe, boasts of a huge number of restaurants and pubs. There are quite a few Italian Restaurants as well, spread across the city. In this scenario, it is urgent to adopt machine learning tools in order to assist ‘Amaretto’ to make wise and effective decisions regarding the location of the Berlin Restaurant. To solve this business problem, we are going to cluster Berlin neighbourhoods to recommend areas where people are more likely to go out to eat at restaurants and give good business. We will also look at the proximity of other Italian Restaurants in the area to be proposed for the new restaurant.

Data:

Berlin is divided into 12 different Boroughs or neighbourhoods. The data regarding the different Boroughs with their population was extracted from Wikipedia. The link of the extracted data is (https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin). The following fields were extracted: Name of Borough, Population, Area in Sq. Km and Population Density per Sq. Km. The Latitude-Longitude details for each borough was obtained by using Nominatim.

To explore and recommended locations across different boroughs according to the presence of other restaurants, data was obtained through FourSquare API interface and it was arranged as a dataframe for visualization. The following fields were extracted from FourSquare API: Name of Venue, Venue Latitude, Venue Longitude and Venue Category. By merging data on Berlin Boroughs and data of existing restaurants in a Borough from FourSquare API interface and further by using the k-Means Clustering method, recommended location options are provided for the new restaurant.

Methodology:

The Methodology section will describe the main components of our analysis and the statistical testing performed on the dataset. The Methodology section has four sub-sections as described below:

1. Data Collection
2. Data Inspection and exploration
3. Data preparation and pre-processing
4. Statistical Modelling

1. Data Collection

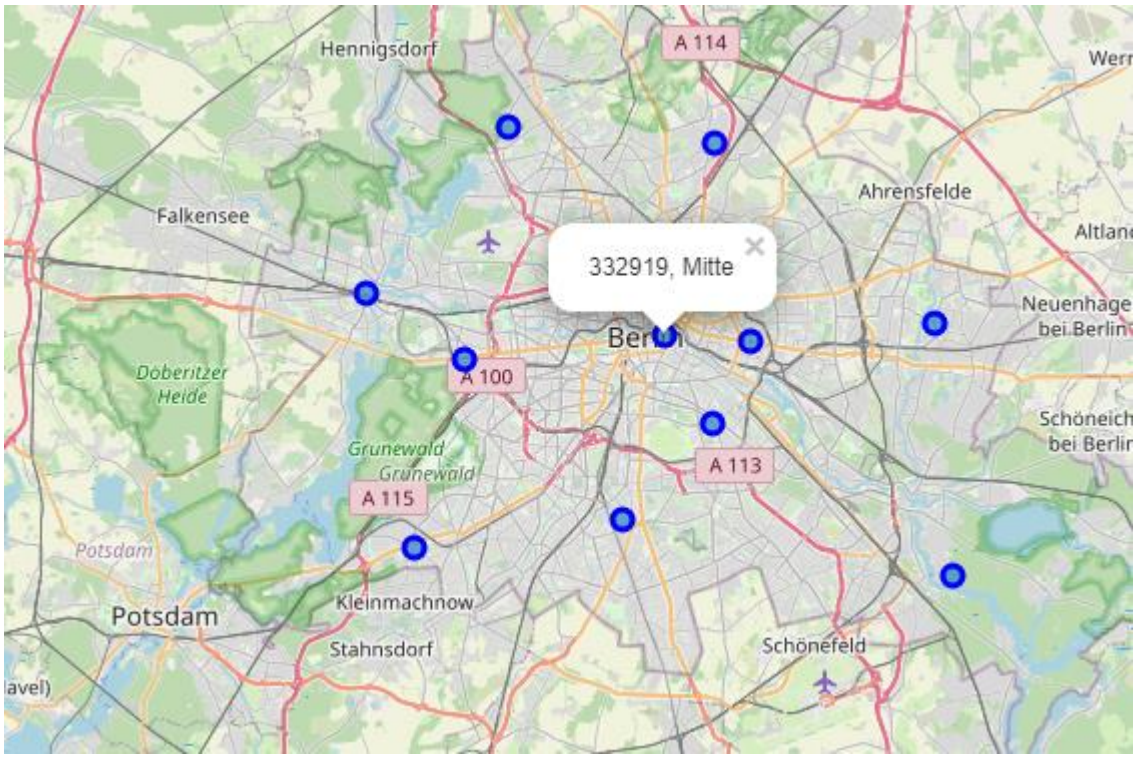
After importing the necessary libraries, the data regarding the different Boroughs of Berlin with their population was extracted from Wikipedia. The following fields were extracted: Name of Borough, Population, Area in Sq. Km and Population Density per Sq. Km. Before using data for analysis, it is imperative that it needs to be explored for better understanding

2. Data Inspection and exploration

We read the dataset that we collected from Wikipedia regarding Berlin Boroughs into a pandas' data frame and display the data. Our dataset consists of 12 rows and 5 columns. We will now prepare and pre-process data accordingly.

3. Data preparation and pre-processing

First, we drop the “Map column” from the extracted dataset, since it is not relevant. Using the Nominatim function, we extract the central location coordinates for each of the 12 Boroughs and it is added to the Pandas Dataframe. We further split the location coordinates to 2 separate columns named “Latitude” and “Longitude” and add the same in the dataframe. We create a map of Berlin with the central coordinates of each Borough marked in same.



In the next steps, using FourSqaure API, the information regarding the nearby venues of each Borough is collected. The following fields were extracted from FourSquare API: Name of Venue, Venue Latitude, Venue Longitude and Venue Category. The data on Berlin Boroughs and data of existing restaurants in a Borough from FourSquare API interface was merged into a new dataframe.

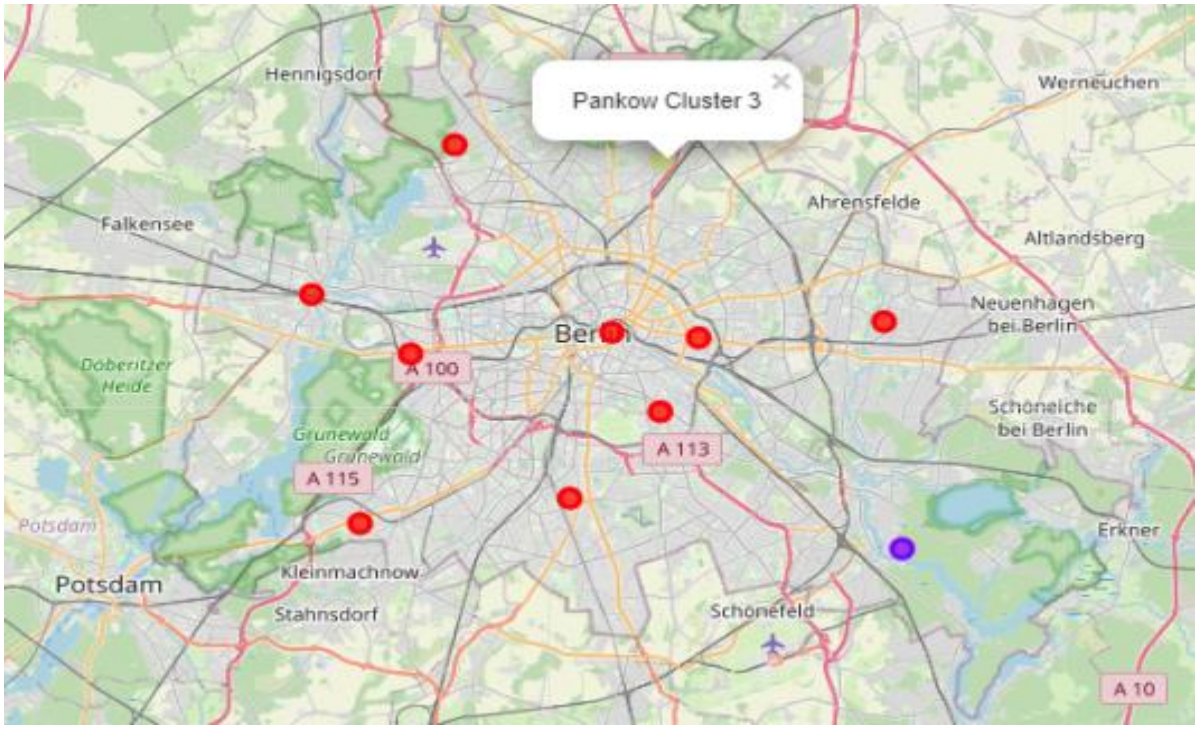
The number of existing Italian Restaurants per Borough was analysed along with the population density of the particular area.

Borough	Italian Restaurant	Population 31 March 2010	Area in km²	Density per km²	loc_coord	Latitude	Longitude
Charlottenburg-Wilmersdorf	2	319628	64.72	4878	(52.5078558, 13.2639518)	52.507856	13.263952
Friedrichshain-Kreuzberg	3	268225	20.16	13187	(52.5153063, 13.4616117)	52.515306	13.461612
Lichtenberg	0	259881	52.29	4952	(48.9212959, 7.4812268)	48.921296	7.481227
Marzahn-Hellersdorf	0	248264	61.74	4046	(52.5225225, 13.5876634)	52.522523	13.587663
Mitte	1	332919	39.47	8272	(52.5176896, 13.4023757)	52.517690	13.402376
Neukölln	3	310283	44.93	6804	(52.4811497, 13.4353501)	52.481150	13.435350
Pankow	0	366441	103.01	3476	(52.597636699999995, 13.436373975411648)	52.597637	13.436374
Reinickendorf	0	240454	89.46	2712	(52.6047631, 13.2952872)	52.604763	13.295287
Spandau	1	223962	91.91	2441	(52.535788, 13.1977924)	52.535788	13.197792
Steglitz-Zehlendorf	3	293989	102.50	2818	(52.4292052, 13.2299741)	52.429205	13.229974
Tempelhof-Schöneberg	0	335060	53.09	6256	(52.4406033, 13.3737035)	52.440603	13.373703
Treptow-Köpenick	0	241335	168.42	1406	(52.417893, 13.6001848)	52.417893	13.600185

We also figured out the top 5 most visited venues in each Borough.

4. Statistical Modelling

At this stage, we prepare our dataset for the modelling process and opted for the K-means clustering method to denote the Boroughs into Clusters. The K value was chosen as 4. Based on the 4 clusters, each cluster was analysed as per the type of most frequently visited venues:



Cluster 0:

Italian Restaurant_x	loc_coord	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
2	(52.5078558, 13.2639518)	52.507856	13.263952	Italian Restaurant	Bus Stop	Indian Restaurant	Pizza Place	Hotel	0
3	(52.5153063, 13.4616117)	52.515306	13.461612	Bar	Café	Pizza Place	Thai Restaurant	Falafel Restaurant	0
0	(52.5225225, 13.5876634)	52.522523	13.587663	Supermarket	ATM	Trail	Doner Restaurant	Drugstore	0
1	(52.5176896, 13.4023757)	52.517690	13.402376	German Restaurant	Museum	History Museum	Art Gallery	Hotel	0
3	(52.4811497, 13.4353501)	52.481150	13.435350	Bar	Café	Coffee Shop	Dive Bar	Bistro	0

Cluster 1:

Italian Restaurant_x	loc_coord	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	(52.417893, 13.6001848)	52.417893	13.600185	Lake	Beach	Zoo Exhibit	Gastropub	Doner Restaurant	1

Cluster 2:

Italian Restaurant_x	loc_coord	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	(48.9212959, 7.4812268)	48.921296	7.481227	Hostel	Historic Site	Zoo Exhibit	Gastropub	Doner Restaurant	2

Cluster 3:

Italian Restaurant_x	loc_coord	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	(52.597636699999995, 13.436373975411648)	52.597637	13.436374	Tram Station	Asian Restaurant	Turkish Restaurant	Farmers Market	Fried Chicken Joint	3

Results:

In the first section of analysis, it was found that there are existing Italian restaurants in 6 of the 12 Boroughs: 3 each in Friedrichshain, Neukolin & Steglitz, 2 in Charlottenburg and 1 each in Mitte & Spandau

Borough	Italian Restaurant	Population 31 March 2010	Area in km²	Density per km²
Charlottenburg-Wilmersdorf	2	319628	64.72	4878
Friedrichshain-Kreuzberg	3	268225	20.16	13187
Lichtenberg	0	259881	52.29	4952
Marzahn-Hellersdorf	0	248264	61.74	4046
Mitte	1	332919	39.47	8272
Neukölln	3	310283	44.93	6804
Pankow	0	366441	103.01	3476
Reinickendorf	0	240454	89.46	2712
Spandau	1	223962	91.91	2441
Steglitz-Zehlendorf	3	293989	102.50	2818
Tempelhof-Schöneberg	0	335060	53.09	6256
Treptow-Köpenick	0	241335	168.42	1406

Secondly, the most frequently visited places in each Borough are given below:

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Charlottenburg-Wilmersdorf	Italian Restaurant	Light Rail Station	Comedy Club	Bus Stop	Café
1	Friedrichshain-Kreuzberg	Bar	Café	Pizza Place	Thai Restaurant	Falafel Restaurant
2	Lichtenberg	Hostel	Historic Site	Zoo Exhibit	Garden	Doner Restaurant
3	Marzahn-Hellersdorf	Supermarket	Trail	Doner Restaurant	Drugstore	Shopping Mall
4	Mitte	German Restaurant	Museum	History Museum	Café	Hotel
5	Neukölln	Bar	Café	Coffee Shop	Cocktail Bar	Bistro
6	Pankow	Tram Station	Asian Restaurant	Turkish Restaurant	Farmers Market	Fried Chicken Joint
7	Reinickendorf	Zoo Exhibit	Hostel	Nature Preserve	Liquor Store	Plaza
8	Spandau	Clothing Store	Drugstore	Bakery	Café	Restaurant
9	Steglitz-Zehlendorf	Italian Restaurant	Bus Stop	Liquor Store	Garden	Donut Shop
10	Tempelhof-Schöneberg	Industrial Estate	Beer Bar	Pet Store	Climbing Gym	Brewery
11	Treptow-Köpenick	Lake	Beach	Zoo Exhibit	Gastropub	Donut Shop

Finally, after the K-means clustering analysis, it was found that Treptow , Lichtenburg, Pantow lie in Cluster 1, 2 & 3 respectively while rest of the Boroughs are in Cluster 0.

Borough	Italian Restaurant_x	Population 31 March 2010	Area in km²	Density per km²	loc_coord	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
Charlottenburg-Wilmersdorf	2	319628	64.72	4878	(52.5078558, 13.2639518)	52.507856	13.263952	Italian Restaurant	Bus Stop	Indian Restaurant	Pizza Place	Hotel	0
Friedrichshain-Kreuzberg	3	268225	20.16	13187	(52.5153063, 13.4616117)	52.515306	13.461612	Bar	Café	Pizza Place	Thai Restaurant	Falafel Restaurant	0
Lichtenberg	0	259881	52.29	4952	(48.9212959, 7.4812268)	48.921296	7.481227	Hostel	Historic Site	Zoo Exhibit	Gastropub	Doner Restaurant	2
Marzahn-Hellersdorf	0	248264	61.74	4046	(52.5225225, 13.5876634)	52.522523	13.587663	Supermarket	ATM	Trail	Doner Restaurant	Drugstore	0
Mitte	1	332919	39.47	8272	(52.5176896, 13.4023757)	52.517690	13.402376	German Restaurant	Museum	History Museum	Art Gallery	Hotel	0
Neukölln	3	310283	44.93	6804	(52.4811497, 13.4353501)	52.481150	13.435350	Bar	Café	Coffee Shop	Dive Bar	Bistro	0
Pankow	0	366441	103.01	3476	(52.597636699999995, 13.436373975411648)	52.597637	13.436374	Tram Station	Asian Restaurant	Turkish Restaurant	Farmers Market	Fried Chicken Joint	3
Reinickendorf	0	240454	89.46	2712	(52.6047631, 13.2952872)	52.604763	13.295287	Zoo Exhibit	Hostel	Liquor Store	Nature Preserve	Plaza	0
Spandau	1	223962	91.91	2441	(52.535788, 13.1977924)	52.535788	13.197792	Bakery	Drugstore	Clothing Store	Pub	Café	0
Steglitz-Zehlendorf	3	293989	102.50	2818	(52.4292052, 13.2299741)	52.429205	13.229974	Italian Restaurant	Bus Stop	Liquor Store	Gastropub	Doner Restaurant	0
Tempelhof-Schöneberg	0	335060	53.09	6256	(52.4406033, 13.3737035)	52.440603	13.373703	Beer Bar	Bus Stop	Supermarket	Motorcycle Shop	Climbing Gym	0
Treptow-Köpenick	0	241335	168.42	1406	(52.417893, 13.6001848)	52.417893	13.600185	Lake	Beach	Zoo Exhibit	Gastropub	Doner Restaurant	1

Discussion:

First, as per the results of K-means cluster, it is noted the most visited places in Boroughs located in Cluster 1, 2 & 3 are historical or natural places, which signifies that the people living in these areas enjoy visiting to different places rather than eating out. However, in Cluster 0, the top 5 visited places usually contains restaurants, pubs, bars and other Food & Beverages Places, which means Boroughs belonging to Cluster 0 are more suitable for a new Italian Restaurant.

Next, we should look at the existing Italian restaurants in the Boroughs in Cluster 0.

Marzahn-Hellerdorf, Reinickendorf, Templehof-Schoneberg belong to Cluster 0, however they do not have any Italian restaurants in these Boroughs. Out of these 3 Boroughs, Templehof-Schoneberg has the highest population density and their most visited place is a beer bar. So it seems to be a good choice for the new venue.

However, on further analysis, it is noted that Borough Mitte has the second highest population density among the Boroughs, however has only one Italian Restaurant in the area. Further, it has a Central location and has historical monuments and hotels in the top 5 visited places, which ensures a large footfall of external people to the Borough along with their residential members. Further, the most visited place in the Borough is a restaurant.

Therefore, given the fact, that the Berlin branch is poised to become the European headquarters in near future, it is essential that it garners high visibility along with profitability. Therefore, Borough Mitte seems to be the perfect choice for opening this restaurant and is recommended as per the analysis performed.

Conclusion:

To sum up, “Amaretto”, a global chain of Italian Restaurants, headquartered at New York, is looking to further extend their business across Europe, with Berlin being a key target. Due to its economic status as well as the central location, Berlin is planned to become the European Headquarters in the coming years

There is already a stiff competition in the food & beverages industry and Berlin being the economic hotspot of entire Europe, boasts of a huge number of restaurants and pubs. There are quite a few Italian Restaurants as well, spread across the city. In this scenario, it is urgent to adopt machine learning tools in order to assist ‘Amaretto’ to make wise and effective decisions regarding the location of the Berlin Restaurant

To explore and recommended locations across different boroughs according to the presence of other restaurants, data was obtained through Wikipedia, FourSquare API and location coordinates was obtained by using Nominatim. The data was further analysed using k-means clustering method

Finally, we drew the conclusion that Marzahn-Hellerdorf, Reinickendorf, Templehof-Schoneberg are areas where people like to go out for dining and they do not have any Italian restaurants in these Boroughs. Out of these 3 Boroughs, Templehof-Schoneberg has the highest population density and their most visited place is a beer bar. So, it was a good choice for the new venue.

However, on further analysis, it is noted that Borough Mitte has the second highest population density among the Boroughs, however has only one Italian Restaurant in the area. Further, it has a Central location and has historical monuments and hotels in the top 5 visited places, which ensures a large footfall of external people to the Borough along with their residential members. Further, the most visited place in the Borough is a restaurant.

Therefore, given the fact, that the Berlin branch is poised to become the European headquarters in near future, it is essential that it garners high visibility along with profitability. Therefore, Borough Mitte seems to be the perfect choice for opening this restaurant and is recommended as per the analysis performed.