

**OPENING RESTAURANT IN
BERLIN**

BUSINESS PROBLEM AND STAKEHOLDERS

An exclusivist Canadian restaurateur that has a strong presence in North American HoReCa is looking to expand reach into European markets. After considering several Western European capitals for the opening of its first unit in the continent, the chain management has selected Berlin as the first stop. As part of a preliminary analysis of the HoReCa market, the management is looking at the main competition in the city, and where the competition is located.

In order to provide a complete overview of the competitive environment in Berlin, the analytic strategy is to select all restaurants in the city and cluster them based on popularity information, traffic statistics and pricing policies. A complete map of these clusters is to be presented to the management to gain a preliminary understanding of the market and potential customers.

DATA DESCRIPTION

A comprehensive listing of the restaurants in Berlin will be extracted from Foursquare based on Berlin latitude and longitude data. For each venue id extracted from the Foursquare database, a full overview of

unit popularity,

traffic statistics

... will be crawled from Foursquare database.

Namely, each venue ID will be associated with: a Category, CheckinsCount, usersCount, tipCount, ratings. Once the above variables are preprocessed (dummy coding, scaling where necessary/available), the restaurants in Berlin area will be clustered using Kmeans algorithm. The resulting clusters will be plotted on a map using Folium to depict areas with high competition and market potential.

METHODOLOGY

For this problem, several machine learning algorithms were used.

In a first step, I looked at the state of the HoReCa market in Berlin.

- I extracted the longitude and latitude data from 4square database
- I looked at all the restaurants, cafes, theaters, bars that exist in 4square platform.
- From the initial list extracted, I removed those units that were not relevant for the problem at hand, namely *'Boutique', 'Supermarket', 'Outdoor Sculpture', 'Exhibit', 'Drugstore', 'Furniture / Home Store', 'Church', 'Museum', 'Art Museum', 'Souvenir Shop', 'Optical Shop', 'Memorial Site', 'Department Store', 'Clothing Store', 'Monument / Landmark', 'Historic Site', 'Concert Hall', 'Gym / Fitness Center', 'History Museum', 'Cosmetics Shop'*
- After cleaning the list with only relevant units that would accurately depict the competitive landscape for the Restaurateur, I extracted the relevant metrics for popularity and traffic proxied by:
 - No of photos in the database ('photos.count')
 - No of tips by the 4Square users ('tips.count')
 - Total rating ('rating'),
 - No of raters ('ratingSignals')
- I clustered these venues into 4 segments that represent the top most and least competitive units in Berlin
- I mapped these restaurants and clusters in order to identify the location of the competition

METHODOLOGY

For this problem, several machine learning algorithms were used.

In a first step, I looked at the state of the HoReCa market in Berlin.

- I extracted the longitude and latitude data from 4square database
- I looked at all the restaurants, cafes, theaters, bars that exist in 4square platform.
- From the initial list extracted, I removed those units that were not relevant for the problem at hand, namely *'Boutique', 'Supermarket', 'Outdoor Sculpture', 'Exhibit', 'Drugstore', 'Furniture / Home Store', 'Church', 'Museum', 'Art Museum', 'Souvenir Shop', 'Optical Shop', 'Memorial Site', 'Department Store', 'Clothing Store', 'Monument / Landmark', 'Historic Site', 'Concert Hall', 'Gym / Fitness Center', 'History Museum', 'Cosmetics Shop'*
- After cleaning the list with only relevant units that would accurately depict the competitive landscape for the Restaurateur, I extracted the relevant metrics for popularity and traffic proxied by:
 - No of photos in the database ('photos.count')
 - No of tips by the 4Square users ('tips.count')
 - Total rating ('rating'),
 - No of raters ('ratingSignals')
- I clustered these venues into 4 segments that represent the top most and least competitive units in Berlin
- I mapped these restaurants and clusters in order to identify the location of the competition

RESULTS – COMPETITIVE OVERVIEW

Category of venue	☑ Average of No of photos	Average of Tips	Average of Rating	Average of NoRaters
Asian Restaurant	40.3	18.3	8.1	82.7
Austrian Restaurant	99.0	47.0	8.4	184.0
Bar	71.0	42.0	8.4	207.5
Burrito Place	3.0	28.0	8.0	99.0
Café	227.0	101.7	8.1	496.3
Chocolate Shop	488.0	523.5	9.0	840.5
Cocktail Bar	87.0	80.5	8.7	260.0
Coffee Shop	135.5	45.5	7.9	221.0
Frozen Yogurt Shop	61.0	162.0	8.4	243.0
German Restaurant	357.0	110.0	8.0	503.8
Gourmet Shop	45.3	4.7	8.6	34.0
Hookah Bar	5.0	5.0	7.5	12.0
Hotel Bar	54.0	3.0	7.6	18.0
Indian Restaurant	12.0	11.0	8.6	31.0
Italian Restaurant	98.0	74.0	8.2	266.0
Piano Bar	16.0	2.0	8.0	22.0
Restaurant	165.5	87.5	8.4	413.0
Roof Deck	6.0	27.0	8.6	25.0
Salad Place	25.0	15.0	7.8	74.0
Sandwich Place	158.0	100.0	9.1	378.0
Steakhouse	405.0	128.0	9.0	543.0
Sushi Restaurant	307.0	128.0	8.4	604.0
Vegetarian / Vegan Restaurant	179.0	105.0	9.3	449.0
Wine Bar	59.0	49.0	8.3	133.0
Grand Total	146.0	80.9	8.3	271.7

➤ Popularity for restaurants specialized in local cuisine / steakhouses

➤ Ratings reflect high consumption of coffes and restaurants that serve coffee variants

➤ Popularity for restaurants that serve sweets and deserts

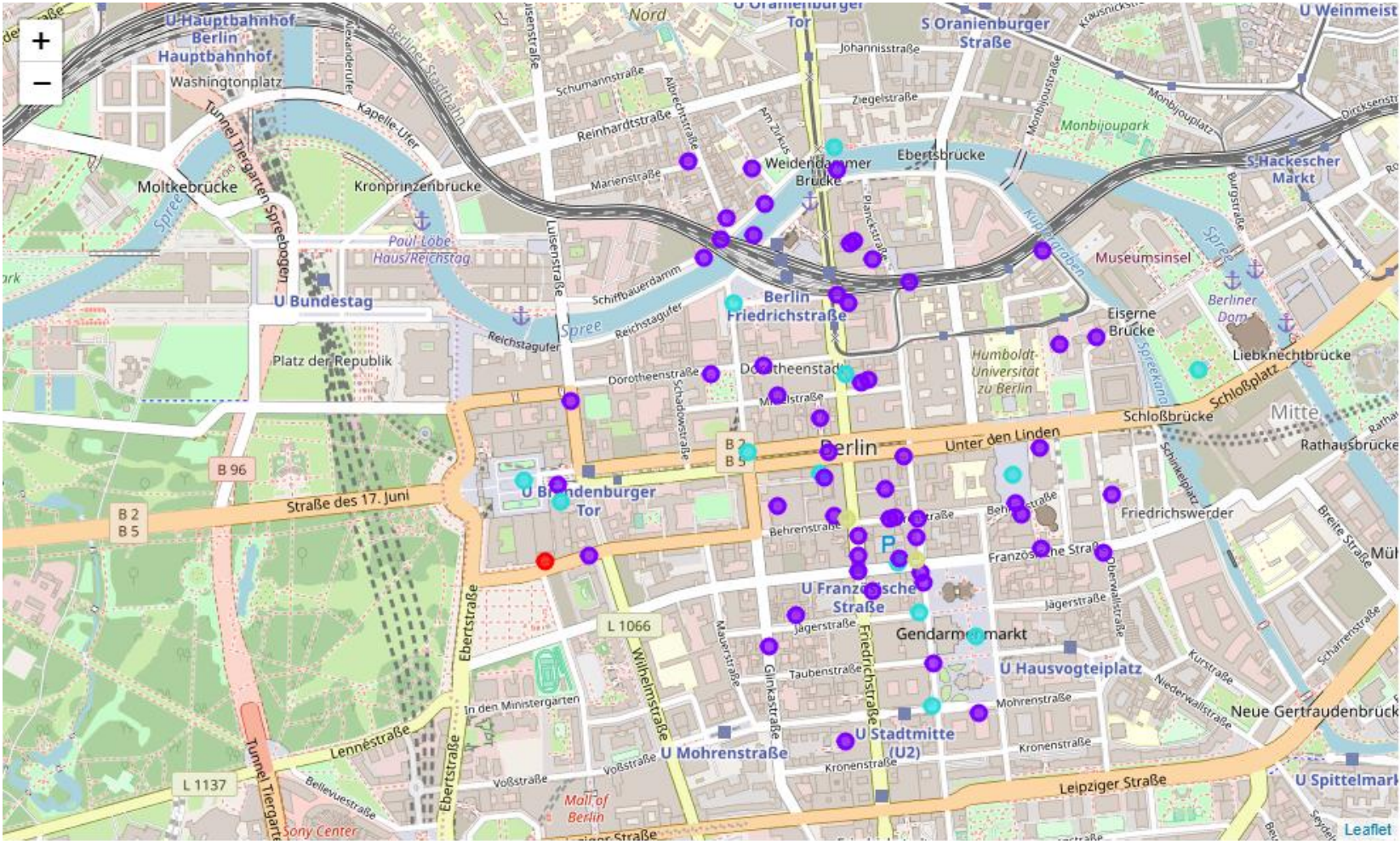
➤ Vegan / vegetarian dishes are highly popular, probably among youth.

RESULTS - CLUSTERS

Cluster	# Units	Average of No of photos	Average of Tips	Average of Rating	Average of NoRaters
0	1	12	11	8.6	31
1	54	112.2	39.6	8.3	173.6
2	13	599.8	187.0	9.0	732.2
3	2	54	3	7.55	18
Grand Total	70	199.7	65.52857143	8.398571429	270.8285714

The most popular venues are in cluster 2, which contains 13 units. The average rating is 9 in this cluster and have been rated by an average of 732 users.

RESULTS - WHERE ARE THEY LOCATED



CONCLUSIONS

- Restaurants are highly popular in Berlin and are highly popular among youth, given their rate of offering recommendations and tips on 4square platform.
- In terms of consumption behavior, the most preferred venues are those that serve local cuisine dishes, meat dishes, a high variety of deserts, vegetarian options and coffee varieties in terms of drinks.
- Berliners do not seem to be very attracted by foreign cuisines, given the low scores for Italian, Indian or Asian restaurants.
- The relevant competition for the Canadian restaurateur is Cluster 2 venues, which boast an average rating of 9, 187 tips from users and an average number of raters of 732. Additionally, it seems that venues with a strong online presence gather more positive feedback and more raters than those venues that don't display many photos on their pages.
- In terms of location, the best rated venues are located at the heart of the city, along The Friedrichstraße
- Street, which is a major cultural and shopping street in central Berlin. It is also one of the most frequented shopping areas in Berlin.