

UTKARSH SHARMA  
JUNE 2019

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

---

# THE BATTLE OF NEIGHBORHOODS (ITALIAN RESTAURANT IN BERLIN, GERMANY)

IF SOMEONE IS LOOKING TO OPEN AN ITALIAN RESTAURANT IN THE CITY OF BERLIN, WHERE WOULD YOU RECOMMEND THAT THEY OPEN IT?

The background image shows a panoramic view of the Berlin skyline under a clear blue sky. The most prominent feature is the Fernsehturm (TV Tower) in the center-right. To its left is the modern glass facade of the Sony Center. Further left is the Reichstag dome with its green copper roof. In the foreground, there's a mix of modern office buildings and older residential structures. A large area of green space with autumn-colored trees is visible in the lower half of the image.

BUSINESS PROBLEM

# TARGET AUDIENCE AND DATA REQUIREMENTS

- ▶ **Target Audience:** This project is particularly useful for restaurant chain owners and/or investors looking to open and/or invest in an Italian restaurant in the capital of Germany.
- ▶ **Data Requirements:** To solve this problem, we will need the following data:
  - List of Neighborhoods in Berlin
  - Latitude and Longitude coordinates of all the Neighborhoods
  - Venue data (powered by Foursquare), particularly related to Italian Restaurants

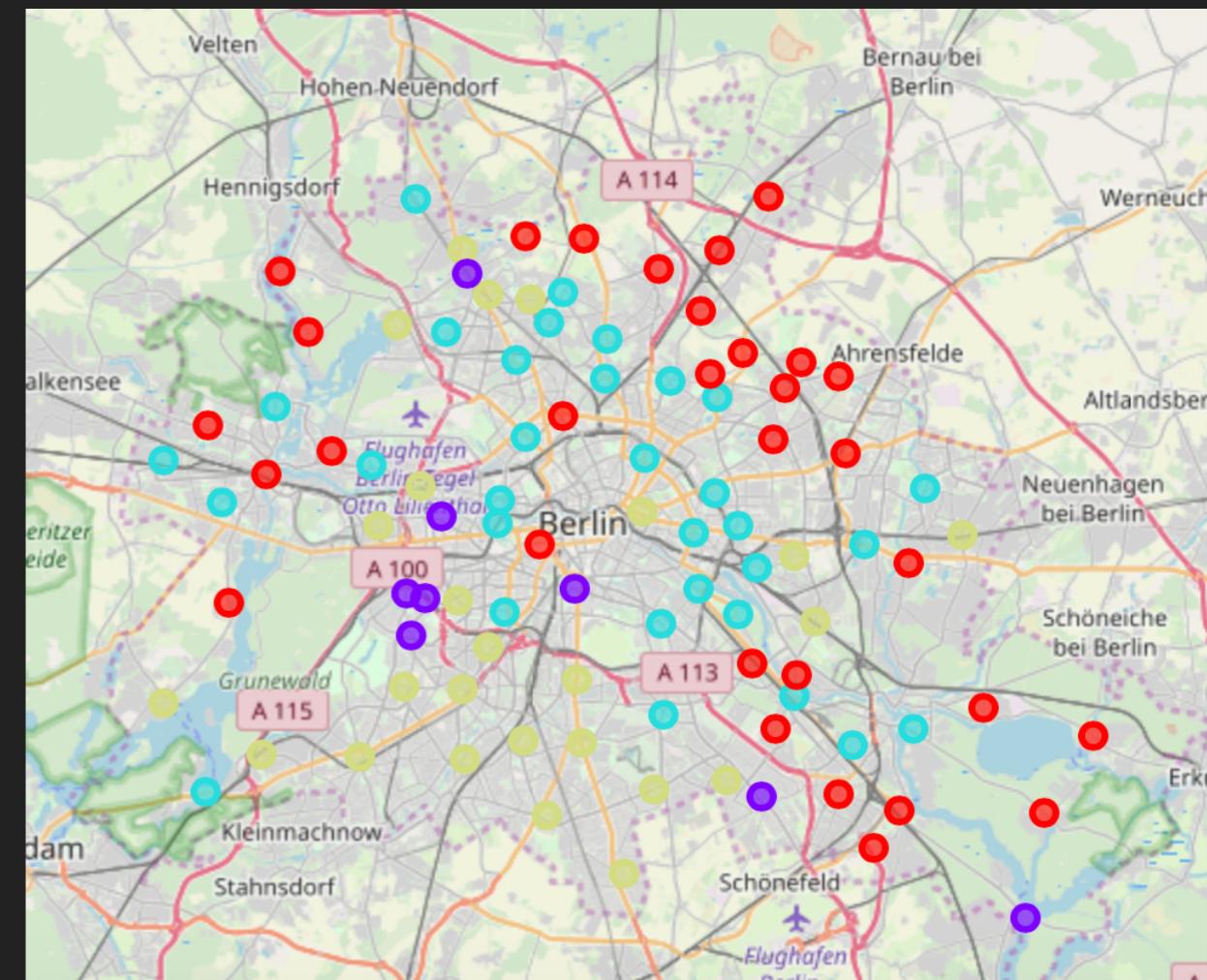
To get a list of neighborhoods in the city of Berlin. This data is extracted from the Wikipedia page titled, 'Localities of Berlin' ([https://en.wikipedia.org/wiki/Category:Localities\\_of\\_Berlin](https://en.wikipedia.org/wiki/Category:Localities_of_Berlin)). Using the BeautifulSoup and Requests packages of Python, the required data is scraped from the webpage. In order to use the Foursquare API, we fetch the location data of all these neighbourhoods from the Python Geocoder package. Next, the Foursquare API is used to get the top 100 venues in a radius of 2000 meters of a particular neighborhoods. To prepare the data for K-means clustering, we group the data frame by neighborhoods. Lastly, K-means clustering is performed on this data set to return 4 clusters, or categories of neighborhoods in terms of number of Italian Restaurants.

## METHODOLOGY

# BATTLE OF NEIGHBORHOODS

## RESULTS

- ▶ The output from the clustering algorithm shows the division of all the neighborhoods into 4 groups in terms of number of Italian restaurants present in that area.
- Cluster 0: Neighborhoods with very low number of Italian restaurants
- Cluster 1: Neighborhoods with low number of Italian restaurants
- Cluster 2: Neighborhoods with a significant number of Italian restaurants
- Cluster 3: Neighborhoods with heavy competition of Italian restaurants



# DISCUSSIONS AND SUGGESTIONS

- ▶ As observed from the clusters, opening a restaurant in Berlin could be tricky business. Going for Restaurants in Category 0 can lead to less footfall, but also non-existence of competition, hence, this is not recommended.
  
- ▶ On the other hand, Category 3 suffers from an intense competition and this is also not recommended. Italian restaurants with a USP like innovative menu, dirt cheap prices etc. can go for Category 1 but Category 2 is recommended more than Category 1 due to a significantly higher expected target audience.