

Data Science Capstone Project

FINDING THE BEST SPOT IN BERLIN FOR OPENING A CAFÉ CUM RESTAURANT

Alpesh Laxman Vora



OVERVIEW

- ❖ Introduction
- ❖ Data – Acquisition and Cleaning
- ❖ Examining Relevant Parameters
- ❖ Foursquare Data Analysis
- ❖ K-Means Clustering
- ❖ Conclusion

INTRODUCTION

❖ Berlin – Background Discussion

- Capital and largest city of Germany by population and area
- Top European destination - Hosts to 48 million tourists in year 2019 - ranked third after London and Paris
- Europe's leading economic force – rapidly growing economy for several years – GDP € 153.3 billion in 2019 – total number of businesses 192199

❖ Business Opportunity in Berlin

- Tourists, visitors and most business employees rely on Cafés and Restaurants for their dietary needs
- Significant number of businesses and tourists in Berlin provides huge opportunities for café or restaurant business

DATA – ACQUISITION AND CLEANING

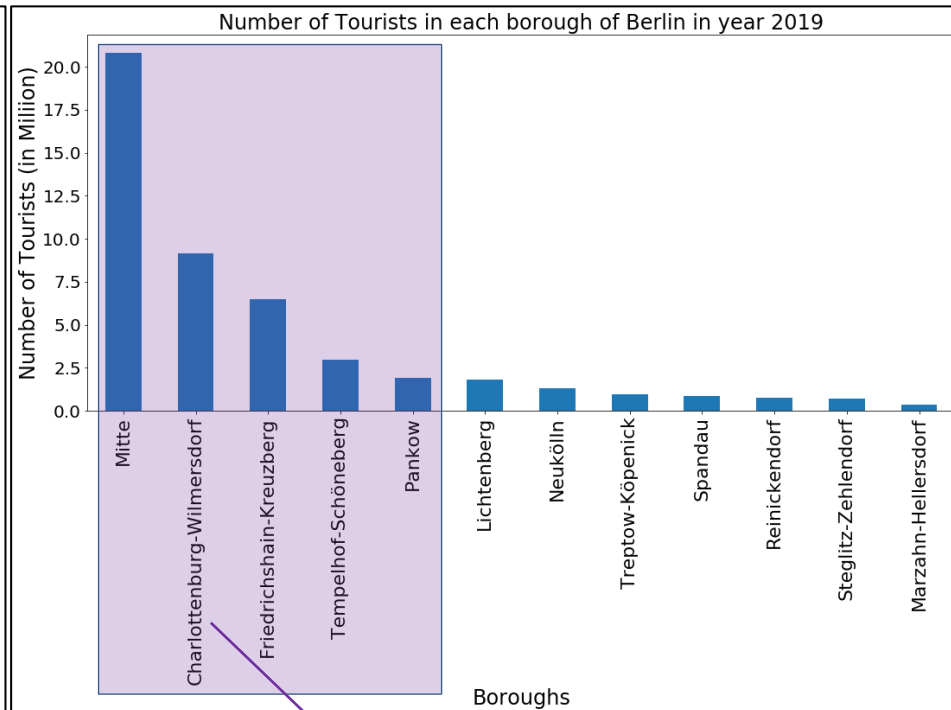
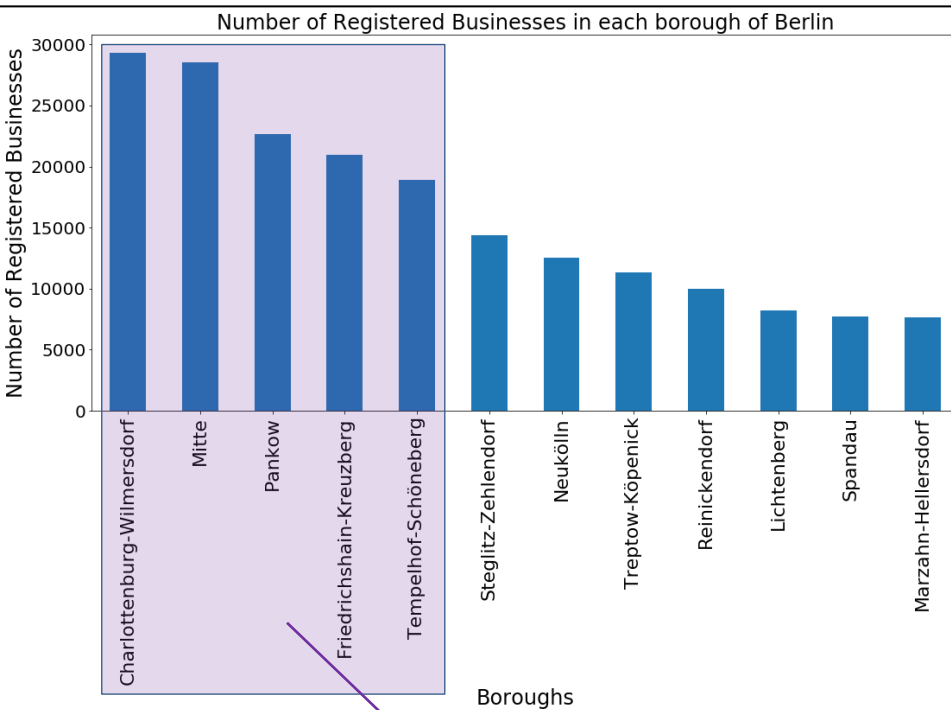
❖ Acquisition of data

- ❖ Berlin Boroughs, Neighborhoods, Area and Population data is scraped from Wikipedia page – “Verwaltungsgliederung Berlins”
- ❖ Berlin Registered Business data and Berlin Registered Tourist data are downloaded from “Amt für Statistik Berlin-Brandenburg” website
- ❖ Rent price of property data is scraped from “Homeday – mein Immobilienmakler” website

❖ Total 4 datasets: 2 neighborhood based, and 2 boroughs based

- ❖ Converted data from German conventions to standard conventions, cleaned and merged all into one Panda data frame
- ❖ Import Geospatial data using GeoPy geocoders
- ❖ Cleaned data contains 8 features

EXAMINE RELEVANT PARAMETERS



These 5 boroughs have 62.7% of total Businesses

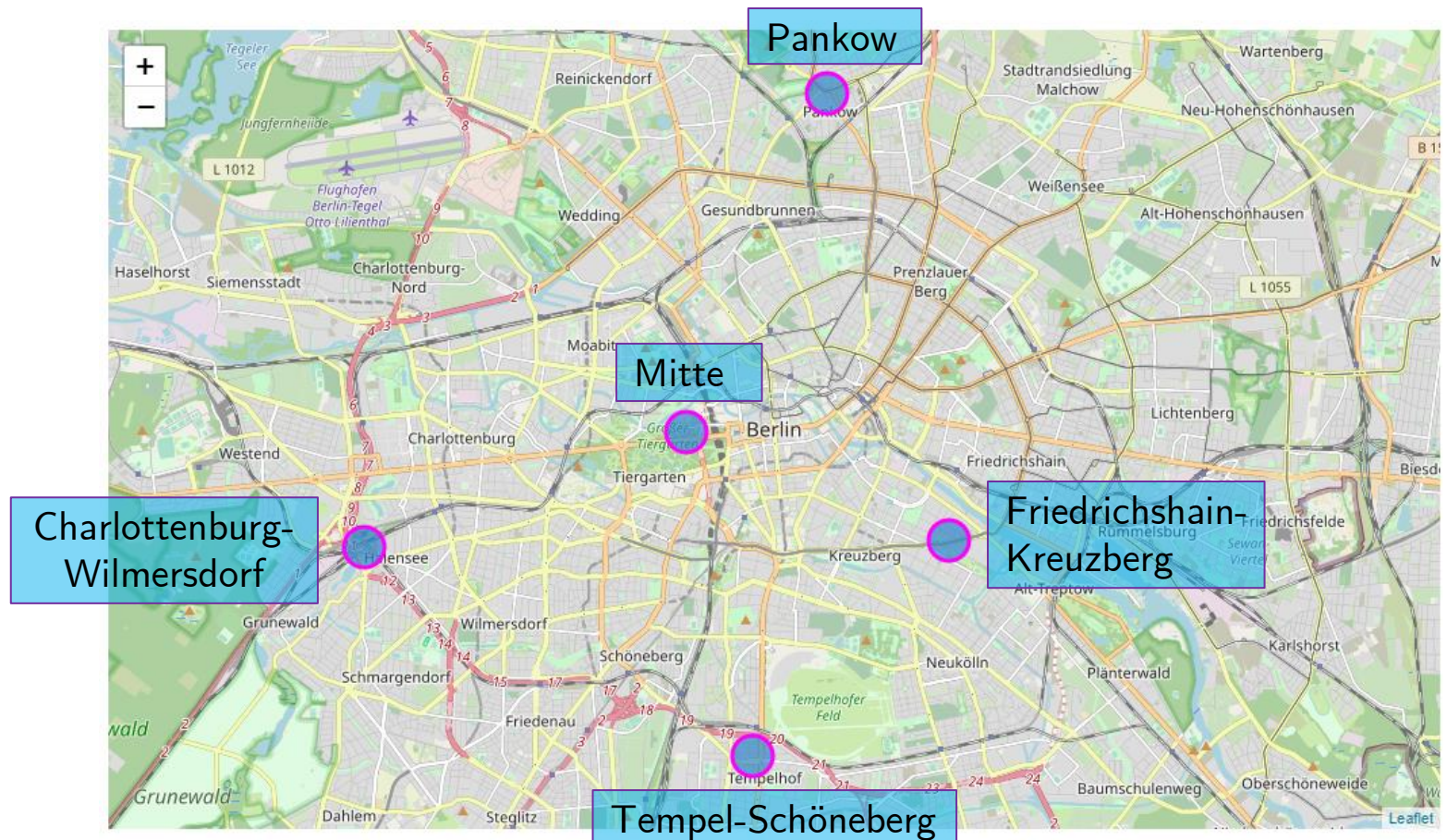
These 5 boroughs hosts to 86% of total Tourists

Narrowed down 5 boroughs
for further analysis →

- Charlottenburg-Wilmersdorf
- Mitte
- Pankow
- Friedrichshain-Kreuzberg
- Tempel-Schöneberg

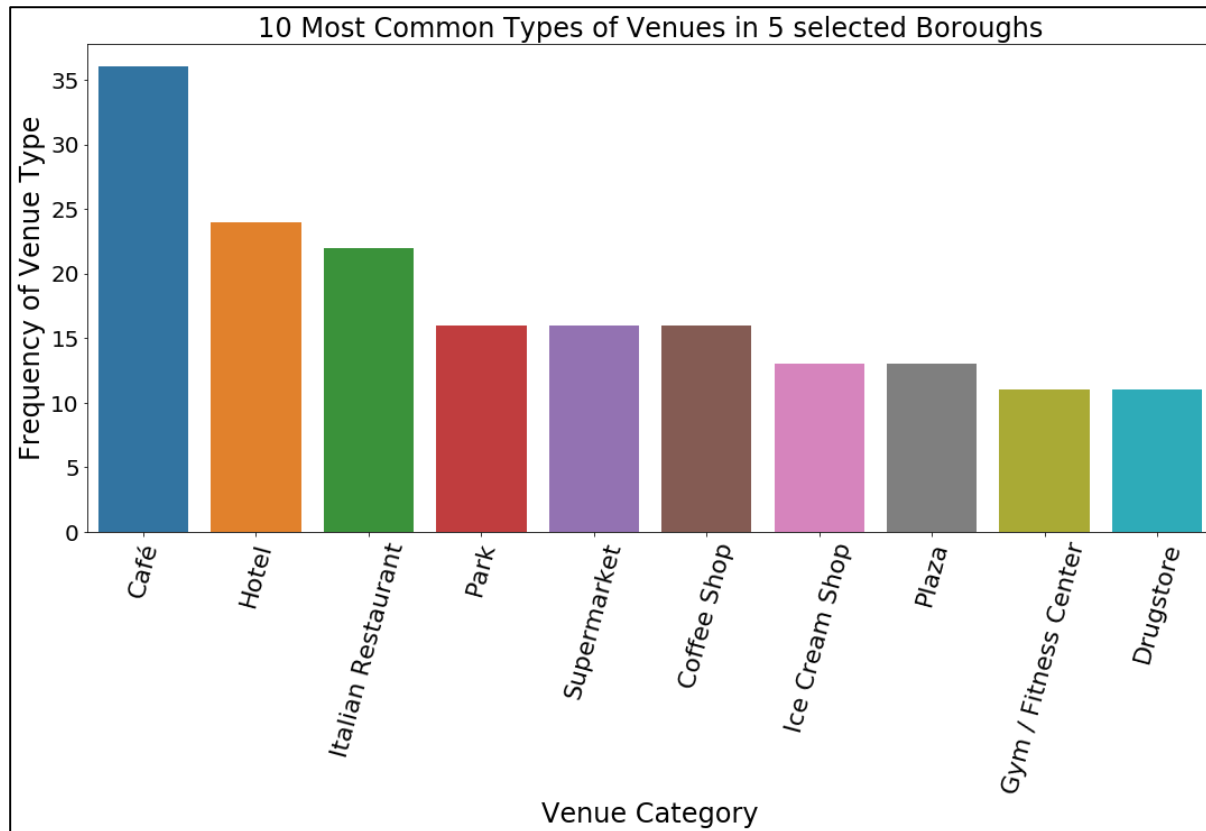
NARROWED DOWN 5 BOROUGHES

Selected top 5 boroughs for further analysis on leaflet map



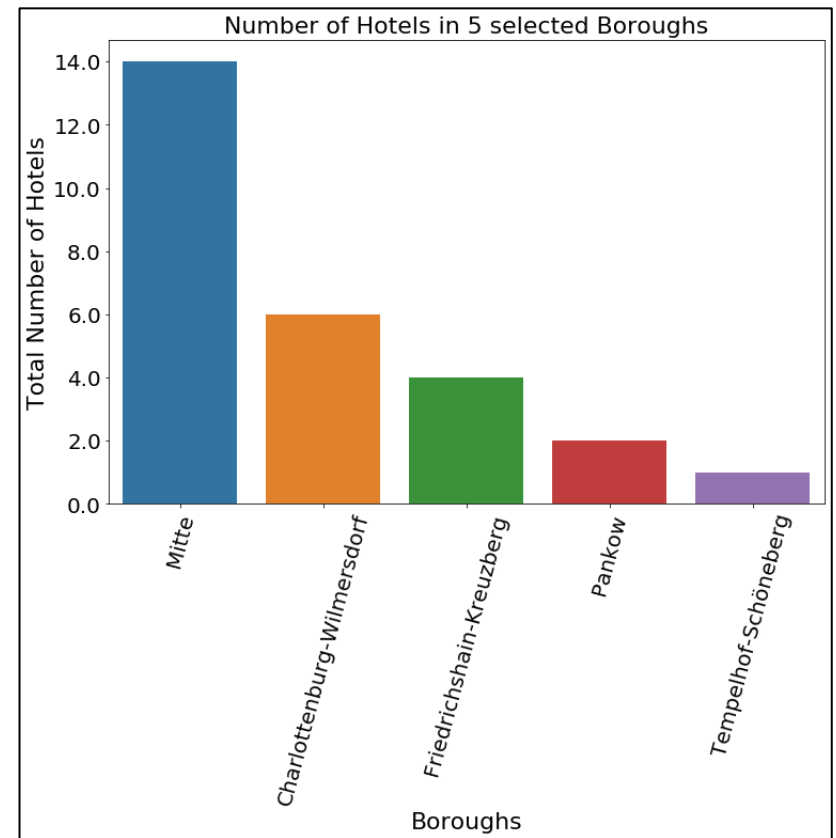
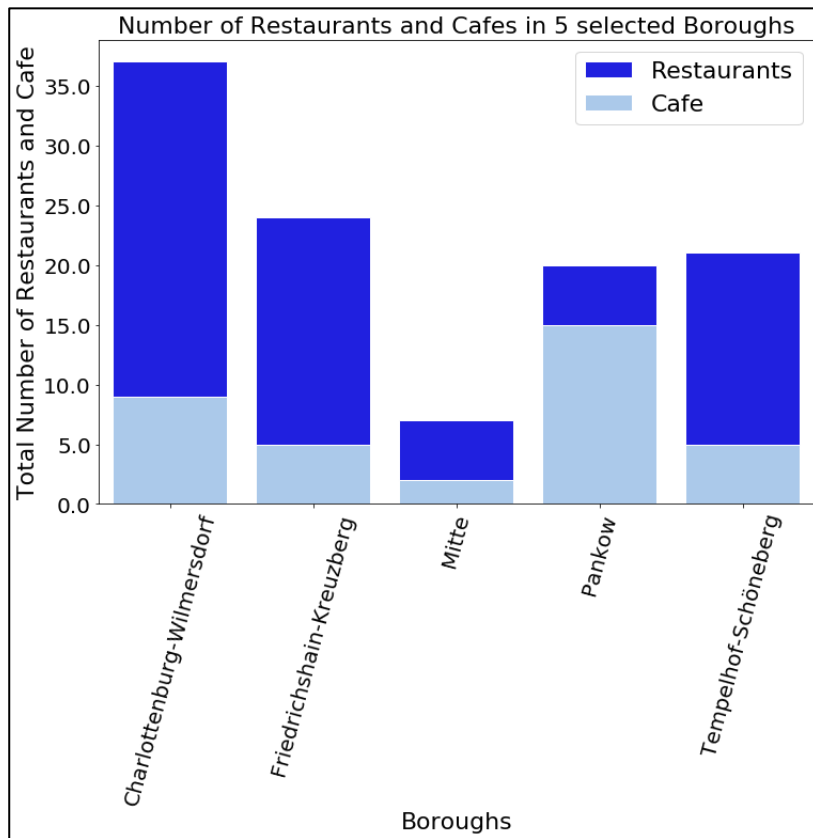
FOURSQUARE DATA ANALYSIS

- ❖ Total returned nearby venues by Foursquare API are 500
- ❖ Total unique categories are 157
- ❖ Café is most common popular type of venue in 5 boroughs overall



FOURSQUARE DATA ANALYSIS

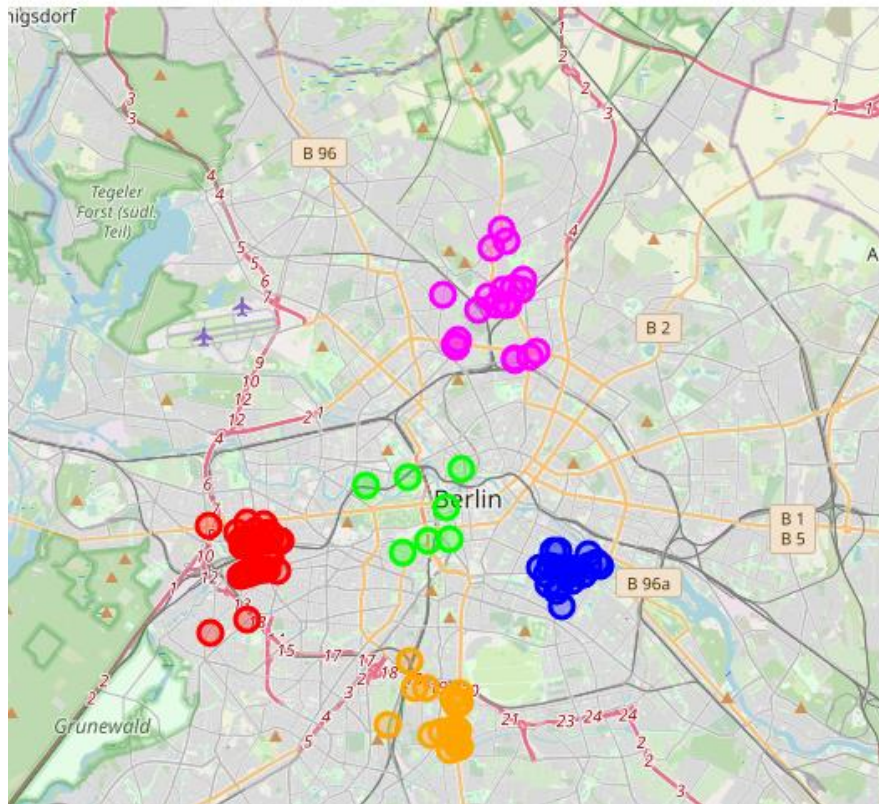
- ❖ Charlottenburg-Wilmersdorf & Friedrichshain-Kreuzberg boroughs, which covers 30% total businesses, have Restaurants and Cafés as most popular venues
- ❖ Mitte borough, which hosts to 43% of total tourists, is dominated by Hotel



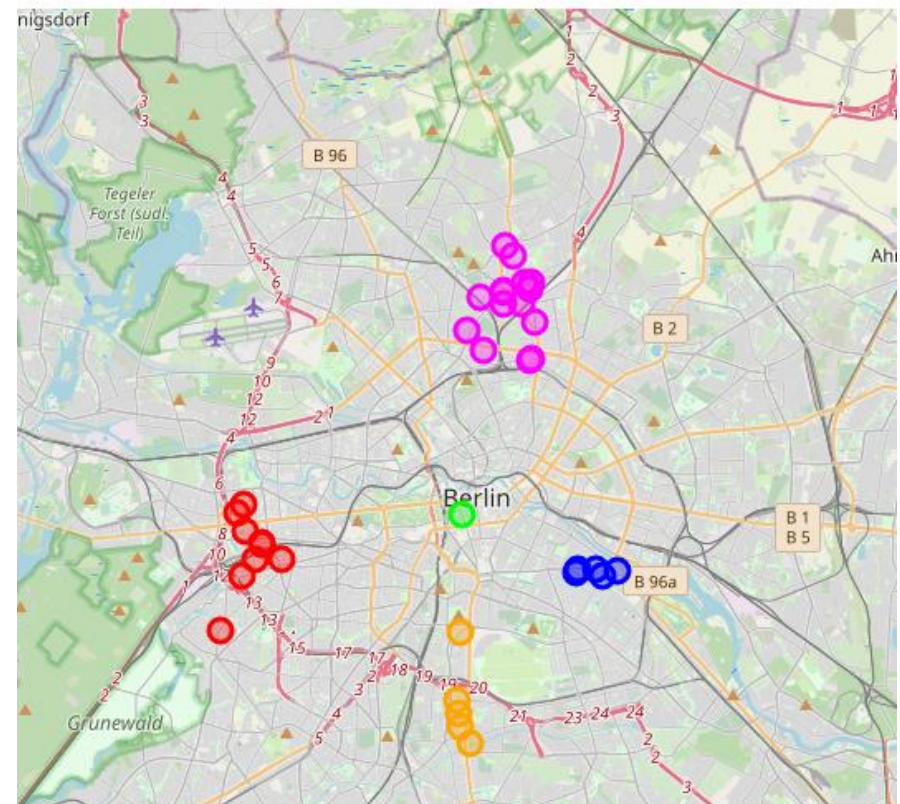
FOURSQUARE DATA ANALYSIS

- ❖ Most visited restaurants and cafés in 5 selected boroughs on leaflet map

Most frequently visited restaurants in 5 selected Boroughs



Most frequently visited cafe in 5 selected Boroughs



Marked circles for below boroughs:

Charlottenburg-Wilmersdorf, Mitte, Pankow, Tempelhof-Schöneberg, Friedrichshain-Kreuzberg

BOROUGH CLUSTERING

- ❖ K-Means clustering is used to group the 5 selected boroughs
- ❖ Clustered into 3 groups based on their popular venue categories in each borough

Clustered group 1:

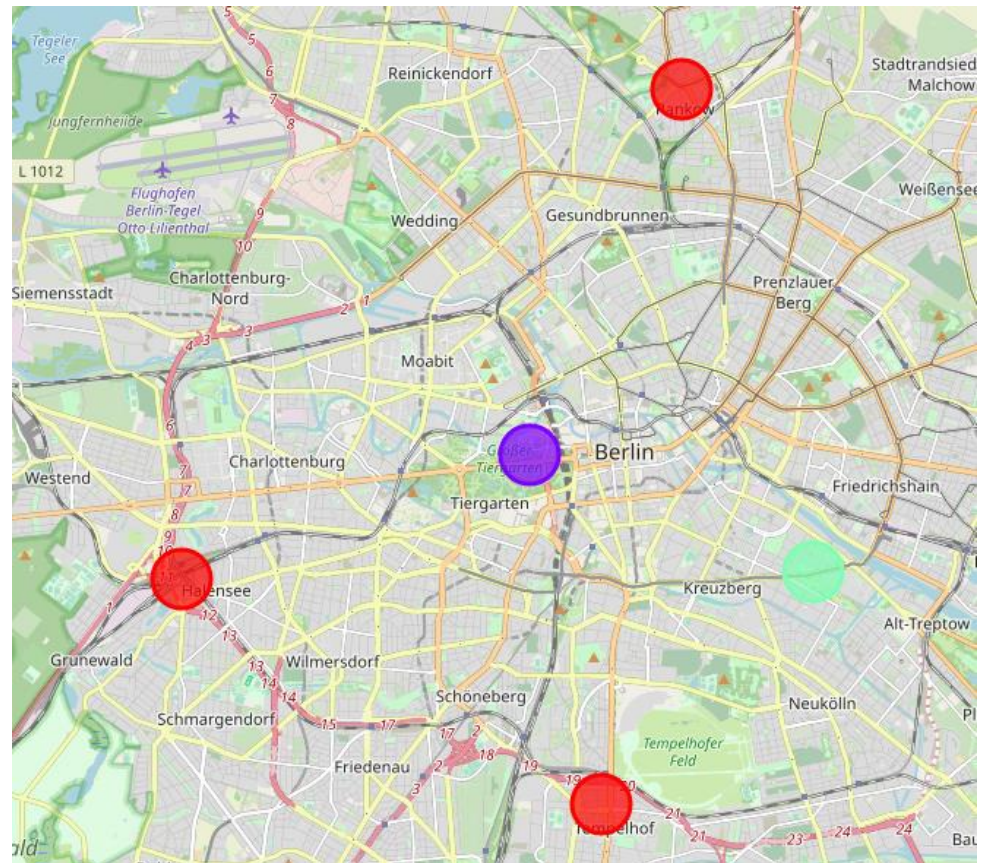
- Charlottenburg-Wilmersdorf
- Pankow
- Tempelhof-Schöneberg

Clustered group 2:

- Mitte

Clustered group 3:

- Friedrichshain-Kreuzberg



BOROUGHS CLUSTERING

- ❖ 3 Clustered groups with size of each circles represents the total number of Restaurants and Cafés together listed as popular venues

Clustered group 1:

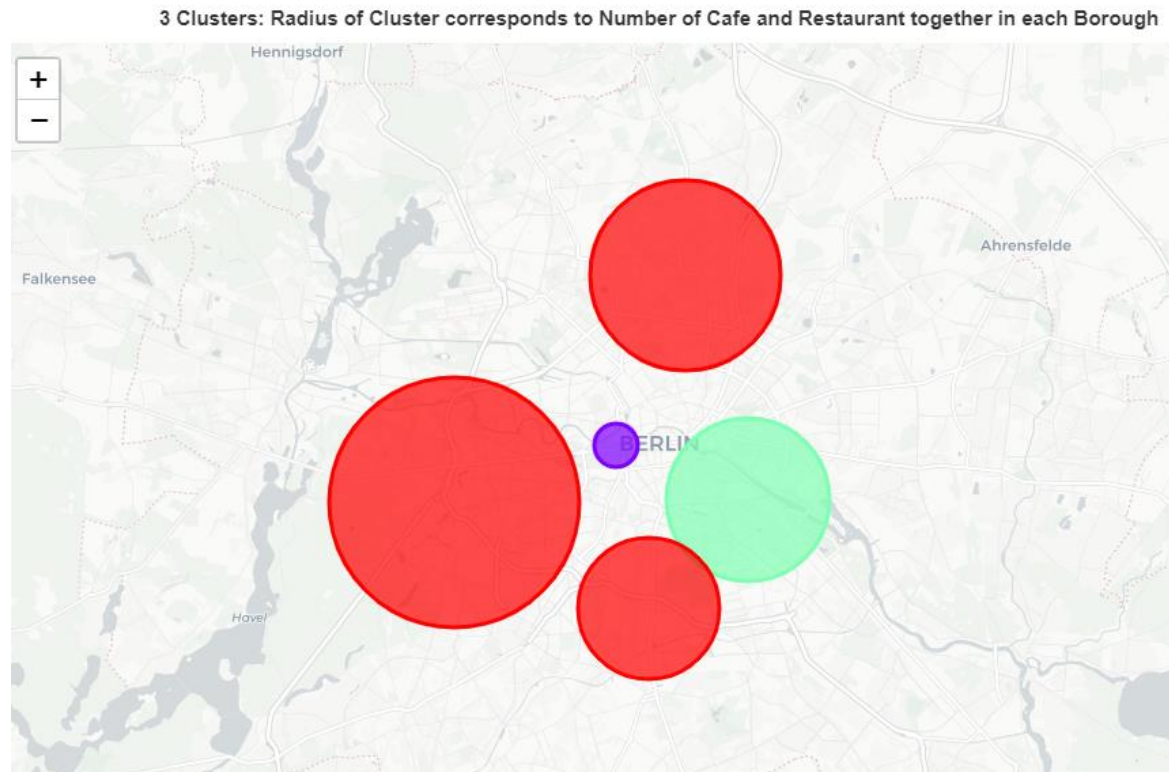
- Charlottenburg-Wilmersdorf
- Pankow
- Tempelhof-Schöneberg

Clustered group 2:

- Mitte

Clustered group 3:

- Friedrichshain-Kreuzberg



CONCLUSION

Analysis:

- ❖ Café is the most common venues in the preferred 5 boroughs
- ❖ Mitte and Charlottenburg-Wilmersdorf boroughs have 30% of total businesses
- ❖ 43% of total tourists in Berlin are hosted by only Mitte borough
- ❖ Mitte and Friedrichshain-Kreuzberg are as own independent clusters
- ❖ Mitte has very low frequency of café-restaurant as common venue

Conclusion:

Mitte borough can be best choice to open a café cum restaurant due to

- ❑ High foot traffic by business employees and tourists
- ❑ Least competition due to low frequency of café cum restaurant as common venues

THANK YOU FOR YOUR KIND ATTENTION

