Berliner Food Recommender System

By: Özge Celenk

October 2019

Methodology

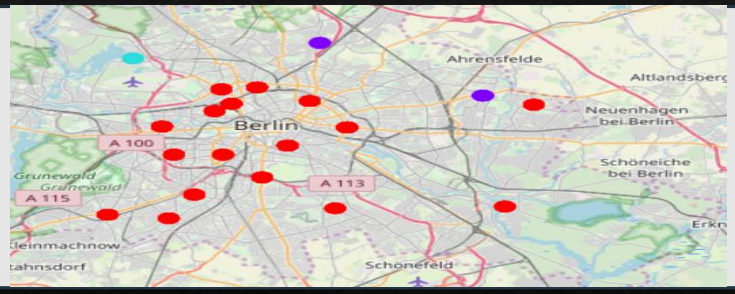
Data Collection

	Neighborhood	Borough	Population	Area(km2)	Latitude	Longitude	Avg_Income_Euro	Normalized_income	Normalized_population	Normalized_area
0	Neukölln	Neukölln	167248	12.0	52.4408	13.4445	1550	0.738095	1.000000	0.342857
1	Prenzlauer Berg	Pankow	160127	11.0	52.5392	13.4242	1850	0.880952	0.957423	0.314286
2	Kreuzberg	Friedrichshain-Kreuzberg	153887	10.0	52.4983	13.4066	1675	0.797619	0.920113	0.285714
3	Friedrichshain	Friedrichshain-Kreuzberg	127189	9.9	52.5158	13.4540	1675	0.797619	0.760481	0.282857
4	Charlottenburg	Charlottenburg-Wilmersdorf	126800	11.0	52.5166	13.3041	1800	0.857143	0.758156	0.314286

Data preprocessing

	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Neukölln	Neukölln	52.4408	13.4445	P.A.M. - Pizza And More	52.437406	13.446066	Fast Food Restaurant
1	Neukölln	Neukölln	52.4408	13.4445	Wochenmarkt Britz-Süd	52.437445	13.446328	Food & Drink Shop
2	Neukölln	Neukölln	52.4408	13.4445	H U Britz-Süd	52.437049	13.447439	Bus Stop
3	Neukölln	Neukölln	52.4408	13.4445	U Britz-Süd	52.437038	13.448485	Metro Station
4	Prenzlauer Berg	Pankow	52.5392	13.4242	Grand Tang Xi Yu	52.537738	13.423279	Chinese Restaurant

K-Means Clustering



Model Testing

```
--------- Charlottenburg ---------
                              Category  Frequency
0  Venue Category_Italian Restaurant       0.10
1              Venue Category_Bakery       0.10
2         Venue Category_Supermarket       0.10
3               Venue Category_Café       0.10
4         Venue Category_Pizza Place       0.07


--------- Friedrichshain ---------
                           Category  Frequency
0            Venue Category_Café       0.13
1  Venue Category_Ice Cream Shop       0.07
2     Venue Category_Pizza Place       0.07
3              Venue Category_Pub       0.07
4     Venue Category_Coffee Shop       0.07
```

```
print("There are {} neighborhoods which has similar venue characteristics to Marzahn.".format(possible_neighborhoods.shape[0]))
```

```
There are 11 neighborhoods which has similar venue characteristics to Marzahn.
```

Datasets

Step 1: Finding the neighborhoods

	Neighborhood	Borough	Population	Area(km2)
0	Neukölln	Neukölln	167248	12.0
1	Prenzlauer Berg	Pankow	160127	11.0
2	Kreuzberg	Friedrichshain-Kreuzberg	153887	10.0
3	Friedrichshain	Friedrichshain-Kreuzberg	127189	9.9
4	Charlottenburg	Charlottenburg-Wilmersdorf	126800	11.0

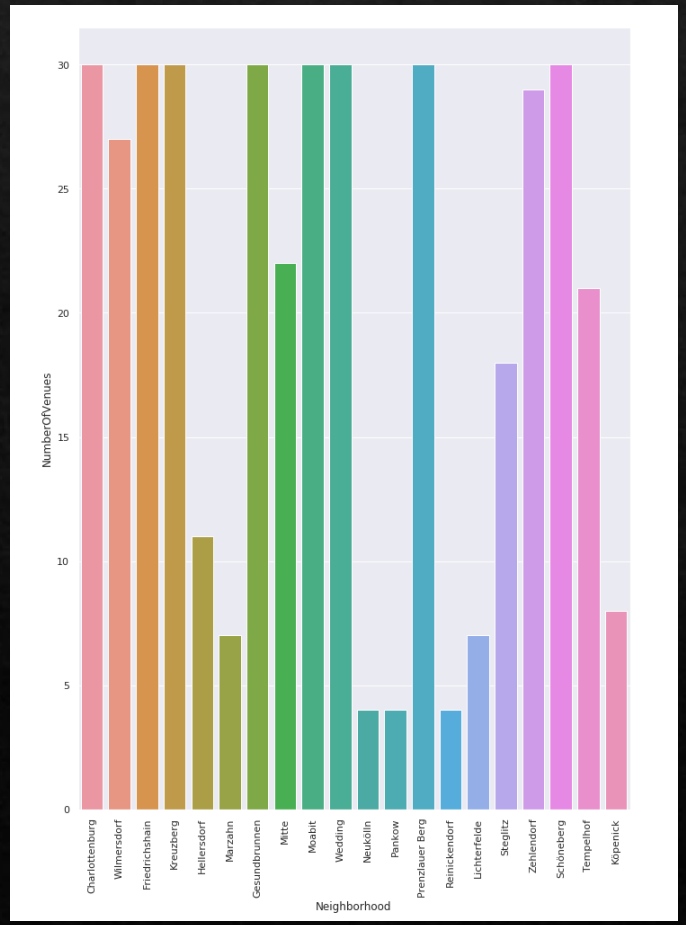
Step 2: Adding geographical coordinates and average income per neighborhood

	Neighborhood	Borough	Population	Area(km2)	Latitude	Longitude	Avg_Income_Euro	Normalized_income	Normalized_population	Normalized_area
0	Neukölln	Neukölln	167248	12.0	52.4408	13.4445	1550	0.738095	1.000000	0.342857
1	Prenzlauer Berg	Pankow	160127	11.0	52.5392	13.4242	1850	0.880952	0.957423	0.314286
2	Kreuzberg	Friedrichshain-Kreuzberg	153887	10.0	52.4983	13.4066	1675	0.797619	0.920113	0.285714
3	Friedrichshain	Friedrichshain-Kreuzberg	127189	9.9	52.5158	13.4540	1675	0.797619	0.760481	0.282857
4	Charlottenburg	Charlottenburg-Wilmersdorf	126800	11.0	52.5166	13.3041	1800	0.857143	0.758156	0.314286

Step 3: Finding the venues per neighborhood using the Foursquare API

	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Neukölln	Neukölln	52.4408	13.4445	P.A.M. - Pizza And More	52.437406	13.446066	Fast Food Restaurant
1	Neukölln	Neukölln	52.4408	13.4445	Wochenmarkt Britz-Süd	52.437445	13.446328	Food & Drink Shop
2	Neukölln	Neukölln	52.4408	13.4445	H U Britz-Süd	52.437049	13.447439	Bus Stop
3	Neukölln	Neukölln	52.4408	13.4445	U Britz-Süd	52.437038	13.448485	Metro Station
4	Prenzlauer Berg	Pankow	52.5392	13.4242	Grand Tang Xi Yu	52.537738	13.423279	Chinese Restaurant

Data Preprocessing



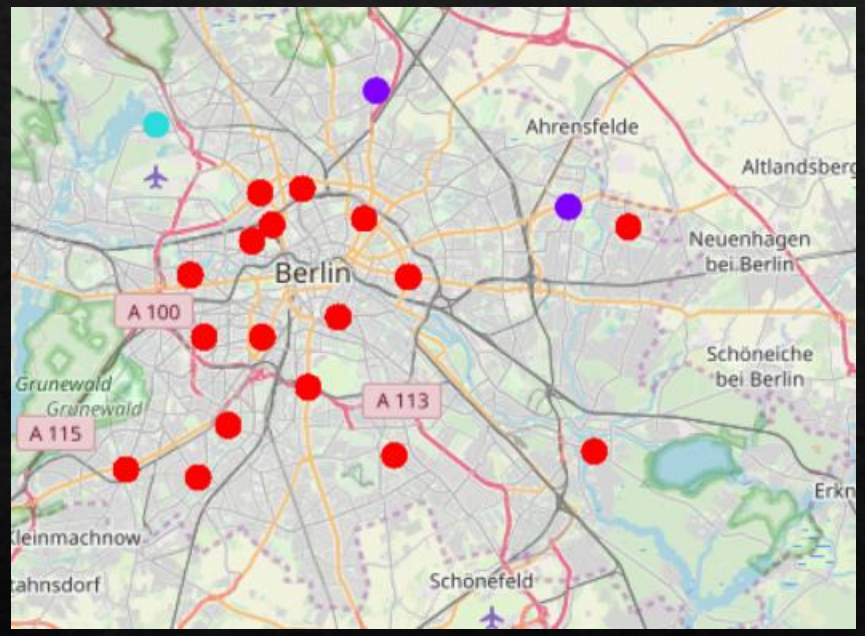
Visualizing the number of venues per neighbourhood

```
--------- Charlottenburg ---------
                                 Category  Frequency
0  Venue Category_Italian Restaurant          0.10
1              Venue Category_Bakery          0.10
2         Venue Category_Supermarket          0.10
3               Venue Category_Café          0.10
4         Venue Category_Pizza Place          0.07


--------- Friedrichshain ---------
                          Category  Frequency
0             Venue Category_Café          0.13
1  Venue Category_Ice Cream Shop          0.07
2     Venue Category_Pizza Place          0.07
3             Venue Category_Pub          0.07
4     Venue Category_Coffee Shop          0.07
```

Analyzing each neighborhood for venue categories

Results



K-Means Clustering

```
possible_neighborhoods = neighborhoods_venues_sorted[neighborhoods_venues_sorted['Cluster']==target_cluster]
possible_neighborhoods.tail()
```

]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Venue Category_African Restaurant	Venue Category_Art Gallery	Categ Re
152	Marzahn	Venue Category_Tram Station	Venue Category_Asian Restaurant	Venue Category_Supermarket	Venue Category_German Restaurant	Venue Category_Ice Cream Shop	0.0	0.0	
209	Pankow	Venue Category_Tram Station	Venue Category_Asian Restaurant	Venue Category_Supermarket	Venue Category_Hotel	Venue Category_Electronics Store	0.0	0.0	
210	Pankow	Venue Category_Tram Station	Venue Category_Asian Restaurant	Venue Category_Supermarket	Venue Category_Hotel	Venue Category_Electronics Store	0.0	0.0	
211	Pankow	Venue Category_Tram Station	Venue Category_Asian Restaurant	Venue Category_Supermarket	Venue Category_Hotel	Venue Category_Electronics Store	0.0	0.0	
212	Pankow	Venue Category_Tram Station	Venue Category_Asian Restaurant	Venue Category_Supermarket	Venue Category_Hotel	Venue Category_Electronics Store	0.0	0.0	

5 rows × 152 columns

```
print("There are {} neighborhoods which has similar venue characteristics to Marzahn.".format(possible_neighborhoods.shape[0]))
```

```
There are 11 neighborhoods which has similar venue characteristics to Marzahn.
```

Recommender System Output

Discussion

- Model works accurately, but still needs improvement
- Future Work: Adding more neighborhoods to the recommender system