Datascience Capstone project - Usage of Foursquare Data to compare popular venues of US and German Cities

by Andreas Ediger

Introduction

Comparing US and German Cities

- Habits
- Popular Places
- Business opportunities

Description of the data and how it will be used to solve the problem.

- Usage of Foursquare API
  - Extraction of popular places in US and German City
  - Extraction of venue category
  - Allows to conclude on habits and preferences

- Jupyter Notebooks

- Pandas for data handling

Analysis

- Diving Berlin and New York into 400 segments and extracting venues

```python
lat_nw=52.589423
long_nw=13.280060

lat_ne=52.589423
long_ne=13.504031


lat_sw=52.427776
long_sw=13.280060

lat_se=52.427776
long_se=13.504031

lat_step=(lat_nw-lat_sw)/20
lat_iter=np.arange(lat_sw,lat_nw,lat_step)

long_step=(long_se-long_sw)/20
long_iter=np.arange(long_sw,long_se,long_step)

name='Berlin'
#radius=500
#LIMIT=100

Berlin_venues = pd.DataFrame(columns=['City',
                    'Venue',
                    'Venue Latitude',
                    'Venue Longitude',
                    'Venue Category','Venue id'])

for lat_ in lat_iter:
    for long_ in long_iter:
        print('Retrieving venues for GPS position: %f, %f' %(lat_,long_))
        try:
            #temp_venues=getVenues(name,lat_,long_,radius=500,LIMIT=100)
            temp_venues=getVenues(name,lat_,long_,radius=500,LIMIT=100)
            #print('        ' , temp_venue.iloc[[0]]['Venue'])
            #print(temp_venue)
            #Berlin_venues=pd.concat([Berlin_venues,temp_venue],axis=0)
            Berlin_venues=Berlin_venues.append(temp_venues,ignore_index=True)
            print("Got %i venues. Total venues: %i" % (temp_venues.shape[0],Berlin_venues.shape[0]))
        except:
            print('Could not retreive venues for GPS position: %f, %f' %(lat_,long_))


Berlin_venues.head()
```

```
Retrieving venues for GPS position: 52.427776, 13.280060
Got 5 venues. Total venues: 5
```

Analysis

- Each venue list has between 7300 and 8500 unique entries

```python
print("Berlin Dataframe shape: %ix%i" % Berlin_venues.shape)
Berlin_unique_venues=Berlin_venues.drop_duplicates(['Venue id'], keep='last')
print("Berlin Dataframe shape after removing duplicates: %ix%i" % Berlin_unique_venues.shape)
```
```
Berlin Dataframe shape: 8494x6
Berlin Dataframe shape after removing duplicates: 7320x6
```

```
Got 4 venues. Total venues. 8554
New York Dataframe shape: 8554x6
New York Dataframe shape after removing duplicates: 8554x6
```

Analysis

- Analysing most popular venue category for each city

```
The most frequent venue categories in Berlin are:
Café                        427
Supermarket                 312
Hotel                       278
Italian Restaurant          255
Bakery                      229
Bar                         217
Coffee Shop                 179
German Restaurant           163
Bus Stop                    151
Park                        142
Vietnamese Restaurant       135
Ice Cream Shop              128
Plaza                       114
Restaurant                  103
Drugstore                    94
Pizza Place                  92
Gym / Fitness Center         80
Cocktail Bar                 78
Asian Restaurant             72
Doner Restaurant             68
Name: Venue Category, dtype: int64
```

```
The most frequent venue categories in New York are:
Pizza Place                 390
Deli / Bodega               258
Coffee Shop                 207
Chinese Restaurant          202
Bakery                      202
Italian Restaurant          191
Bar                         179
Park                        162
Grocery Store               154
Donut Shop                  152
Pharmacy                    147
Sandwich Place              141
Mexican Restaurant          131
Café                        125
Ice Cream Shop              115
American Restaurant         111
Supermarket                 107
Bank                        104
Gym                         101
Fast Food Restaurant         99
Name: Venue Category, dtype: int64
```

Analysis

- Analysing most popular venue category for each city… and check how popular they are in the other city

	City	Category	NY Percentage	Berlin Percentage	NY/Berlin Ratio
0	NY	Pizza Place	4.559271	1.256831	3.627593
1	NY	Deli / Bodega	3.016133	0.314208	9.599170
2	NY	Coffee Shop	2.419921	2.445355	0.989599
3	NY	Chinese Restaurant	2.361468	0.710383	3.324221
4	NY	Bakery	2.361468	3.128415	0.754845
5	NY	Italian Restaurant	2.232874	3.483607	0.640966
6	NY	Bar	2.092588	2.964481	0.705887
7	NY	Park	1.893851	1.939891	0.976267
8	NY	Grocery Store	1.800327	0.491803	3.660666
9	NY	Donut Shop	1.776946	0.122951	14.452498
10	NY	Pharmacy	1.718494	0.286885	5.990180
11	NY	Sandwich Place	1.648352	0.368852	4.468864
12	NY	Mexican Restaurant	1.531447	0.314208	4.873997
13	NY	Café	1.461305	5.833333	0.250509
14	NY	Ice Cream Shop	1.344400	1.748634	0.768829
15	NY	American Restaurant	1.297639	0.081967	15.831190
16	NY	Supermarket	1.250877	4.262295	0.293475
17	NY	Bank	1.215805	0.368852	3.296184
18	NY	Gym	1.180734	0.245902	4.801652
19	NY	Fast Food Restaurant	1.157353	0.519126	2.229428

	City	Category	Berlin Percentage	NY Percentage	Berlin/NY Ratio
0	Berlin	Café	5.833333	1.461305	3.991867
1	Berlin	Supermarket	4.262295	1.250877	3.407446
2	Berlin	Hotel	3.797814	0.701426	5.414417
3	Berlin	Italian Restaurant	3.483607	2.232874	1.560145
4	Berlin	Bakery	3.128415	2.361468	1.324775
5	Berlin	Bar	2.964481	2.092588	1.416658
6	Berlin	Coffee Shop	2.445355	2.419921	1.010511
7	Berlin	German Restaurant	2.226776	0.093523	23.809802
8	Berlin	Bus Stop	2.062842	0.455927	4.524499
9	Berlin	Park	1.939891	1.893851	1.024310
10	Berlin	Vietnamese Restaurant	1.844262	0.303951	6.067623
11	Berlin	Ice Cream Shop	1.748634	1.344400	1.300679
12	Berlin	Plaza	1.557377	0.362404	4.297356
13	Berlin	Restaurant	1.407104	0.888473	1.583732
14	Berlin	Drugstore	1.284153	0.011690	109.846448
15	Berlin	Pizza Place	1.256831	4.559271	0.275665
16	Berlin	Gym / Fitness Center	1.092896	1.040449	1.050408
17	Berlin	Cocktail Bar	1.065574	0.748188	1.424206
18	Berlin	Asian Restaurant	0.983607	0.549451	1.790164
19	Berlin	Doner Restaurant	0.928962	0.000000	inf

Supermarkets occur 3.4 times more often in Berlin than in NY

Analysis

- Analysing which categories are the most „specific" to a city

	City	Category	Berlin Percentage	NY Percentage	Berlin/NY Ratio
14	Berlin	Drugstore	1.284153	0.011690	109.846448
27	Berlin	Tram Station	0.792350	0.023381	33.888798
20	Berlin	Organic Grocery	0.915301	0.035071	26.098270
7	Berlin	German Restaurant	2.226776	0.093523	23.809802
94	Berlin	Modern European Restaurant	0.232240	0.011690	19.865847
44	Berlin	Hostel	0.519126	0.035071	14.802004
118	Berlin	Austrian Restaurant	0.163934	0.011690	14.022951
74	Berlin	Pastry Shop	0.300546	0.023381	12.854372
31	Berlin	Falafel Restaurant	0.696721	0.058452	11.919508
132	Berlin	Science Museum	0.136612	0.011690	11.685792
135	Berlin	Climbing Gym	0.136612	0.011690	11.685792
35	Berlin	Bistro	0.642077	0.058452	10.984645
151	Berlin	Kebab Restaurant	0.122951	0.011690	10.517213
155	Berlin	Camera Store	0.109290	0.011690	9.348634
159	Berlin	Platform	0.109290	0.011690	9.348634
163	Berlin	Persian Restaurant	0.109290	0.011690	9.348634
168	Berlin	Rest Area	0.095628	0.011690	8.180055
55	Berlin	Indie Movie Theater	0.423497	0.058452	7.245191
185	Berlin	Garden Center	0.081967	0.011690	7.011475
187	Berlin	Auto Dealership	0.081967	0.011690	7.011475

	City	Category	NY Percentage	Berlin Percentage	NY/Berlin Ratio
37	NY	Food Truck	0.771569	0.013661	56.478840
25	NY	Caribbean Restaurant	1.005378	0.027322	36.796820
72	NY	New American Restaurant	0.350713	0.013661	25.67222
23	NY	Bus Station	1.028758	0.040984	25.101707
85	NY	Women's Store	0.280570	0.013661	20.537760
90	NY	Kids Store	0.268880	0.013661	19.682020
92	NY	Taco Place	0.257190	0.013661	18.826280
39	NY	Latin American Restaurant	0.736498	0.040984	17.970540
102	NY	Southern / Soul Food Restaurant	0.233809	0.013661	17.114800
31	NY	Discount Store	0.865092	0.054645	15.831190
15	NY	American Restaurant	1.297639	0.081967	15.831190
47	NY	Baseball Field	0.607903	0.040984	14.832827
114	NY	Cuban Restaurant	0.198737	0.013661	14.547580
9	NY	Donut Shop	1.776946	0.122951	14.452498
70	NY	Bubble Tea Shop	0.362404	0.027322	13.263970
22	NY	Bagel Shop	1.028758	0.081967	12.550853
132	NY	Moving Target	0.151976	0.013661	11.124620
1	NY	Deli / Bodega	3.016133	0.314208	9.599170
63	NY	Exhibit	0.385784	0.040984	9.413140
142	NY	Nail Salon	0.128595	0.013661	9.413140

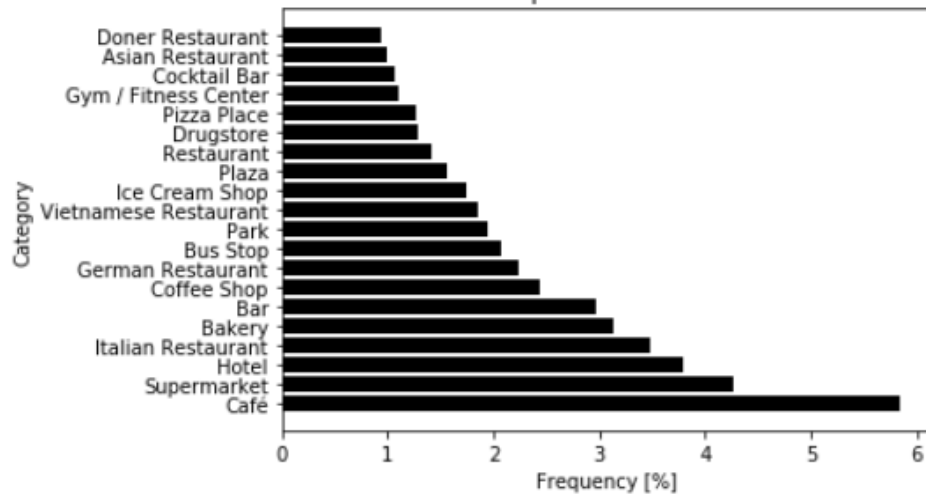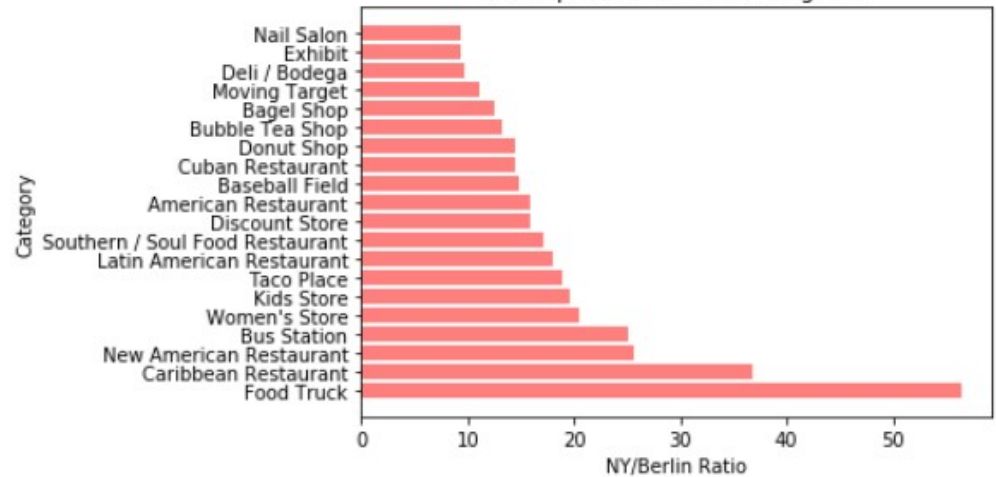Carribean Restaurants occur 36.8 times more often in NY than in Berlin
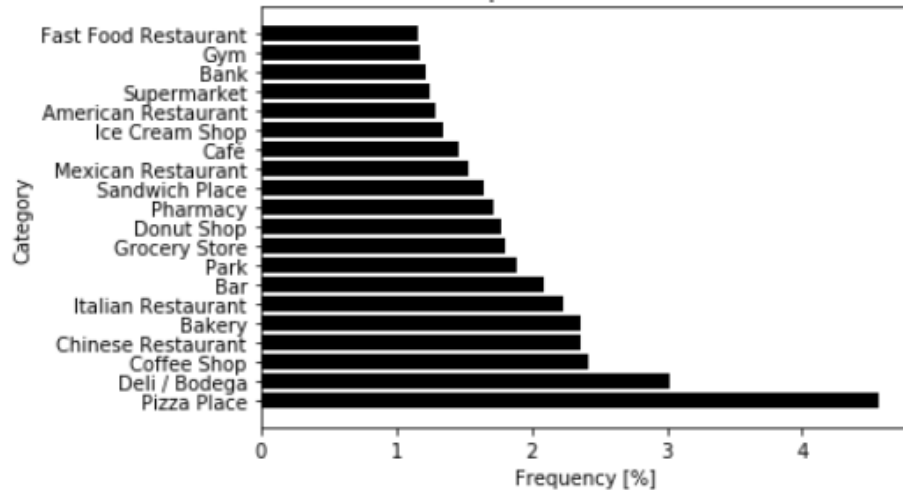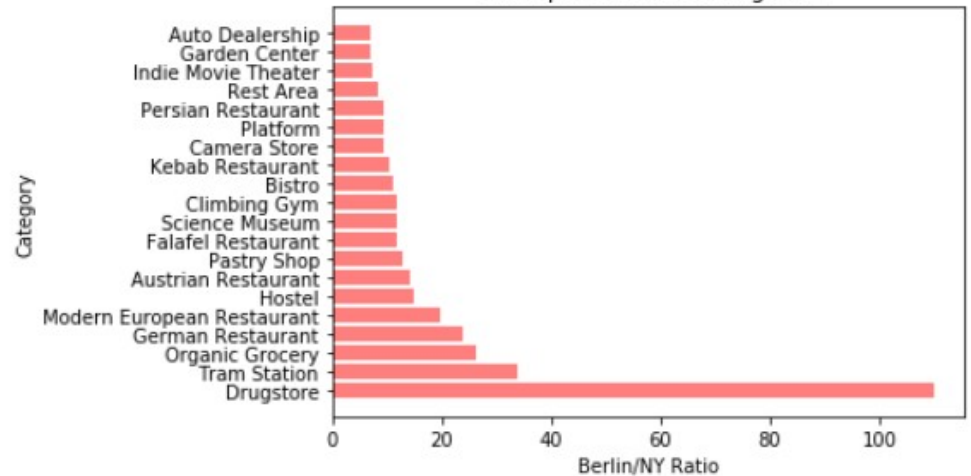
Results

Discussion

- Italian Food occurs in both cities frequently

- Both cities have more restaurants from countries nearby

- German cities have more supermarkets

- Parks and Gyms are equally popular in US and Germany

- Climbing is more popular in Germany than in the US

- Baseball is more popular in the US than in Germany