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

by Andreas Ediger

1. Introduction: Description of the problem and a discussion of the background.

This analysis shall compare venues of US and German Cities. It shall determine general live-style differences between US and German citizens. For example it can reveal differences or similarites between US and German citizens by extracting the density of certain venue groups. Imagine a certain venue category like coffee shops occurs much more often in Germany, we can conclude that Germany is the bigger Coffe-Drinker Nation. Or in the US, parks are much more frequent, this can show that US people like to spend their leasure time in Parks more, compared to Germans. These are just two imaginary examples but the analysis will find this kind of outcome. Another example can be popularity of venue categories. E.g. we could find out that bakeries are extremely popular for US citizens. This could encourage German bakers to start their business in the US (thinking about strong German Bakery-Tradition)
This analysis will be of particular interest of US or German people who plan to visit the other country, or even plan a longer stay up to migration. Also companies who want to expand their business in the other country will benefit from this analysis.

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

The data used for the analysis will mainly be retreived from the Foursquare API. This way we can geospatial data from specific German and US cities. Via the API we can get the occurence of particular venue categories such as Parks or Coffee-Shops. This data can be normalized, e.g. by diving a category occurence by the total number of venues in a specific area. This way we can avoid biasing by US or German citizens using Foursquare more. The Foursquare Data will be enriched by publically available data such as demographics, city size or income structure.

3. Analysis

In this section I will provide details of the actual analysis

3.1 Data collection

In this section we will retreive the data that will be examined in the further analysis. This data will come from Foursquare. As German and US cities shall be compared, venues from two major US cities and two major German cities will be retreived. First, the Foursquare API will be initialized with user credentials.

First, I will try to retreive the venues around a specific location. The location we choose is Berlin Alexanderplatz, a popular place in Berlin. We will search in a 500 m radius around Alexanderplatz. We will use the "explore" endpoint of the foursquare API.
Here is a loop over the first venues that are retrived by the „explore“ endpoint:

```
Venue 1 is a Cosmetics Shop called LUSH and has the venue id
4c28ad7997d00f47212d40ea
Venue 2 is a Hotel called Motel One Berlin-Alexanderplatz and has the venue
id 54aa71d0498e077a285b5eed
Venue 3 is a Historic Site called Ruine der Franziskaner-Klosterkirche and
has the venue id 4adcda7ff964a520eb4721e3
Venue 4 is a Sporting Goods Shop called Decathlon and has the venue id
5535f234498efc81eb55d6e6
Venue 5 is a Bike Rental / Bike Share called Fat Tire Bike Tours and has the
venue id 4bb8541f1261d13a118fe898
Venue 6 is a Gift Shop called Flying Tiger and has the venue id
588875294745863ba6709dcf
Venue 7 is a Burger Joint called momotaro tavern and has the venue id
598f4c672aff3159d8a373b2
Venue 8 is a Scenic Lookout called Berliner Fernsehturm and has the venue id
4adcda7cf964a520504721e3
Venue 9 is a Tour Provider called Alternative Berlin Tour and has the venue
id 4d049a6c37036dcb560112fb
Venue 10 is a Clothing Store called UNIQLO and has the venue id
5ba4b8d31ffed7002cd2e17e
Venue 11 is a Gourmet Shop called Galeria Gourmet and has the venue id
4f2c0d10e4b09e391d56db8d
Venue 12 is a Spanish Restaurant called EL COLMADO and has the venue id
59eb78736bdee6069ed97d77
Venue 13 is a Grocery Store called Asia Markt Lee and has the venue id
4b23b127f964a5201d5824e3
Venue 14 is a Tanning Salon called Sonnenstudio BlueBox and has the venue id
4da96aa1a86e771ea70a1693
Venue 15 is a Frozen Yogurt Shop called Wonderpots Frozen Yogurt and has the
venue id 4efe1962be7bbcd4f1ee9de5
Venue 16 is a Plaza called Alexanderplatz and has the venue id
4be5826a78e895219ea964ce
Venue 17 is a Bookstore called Thalia and has the venue id
4b3b7247f964a520f47325e3
Venue 18 is a Gym / Fitness Center called Superfit and has the venue id
4d98ab3e0caaa1439943b2b3
Venue 19 is a Coffee Shop called Coffee Fellows and has the venue id
5c94c9b72db4a9002c27ee75
Venue 20 is a Bed & Breakfast called B & B Hotels Alexanderplatz and has the
venue id 58bfc797af5c14780419
```

So, we know what is around Alexanderplatz. However, more venues around Berlin are required. For this purpose four „corners“ were selected (North-East, North-West, South-East, South-West) that describe the territory of Berlin. Within this square, the area was

divided into 400 parts. For each part an „explore“ call was made. Here is a sample output of the for-loop.

```
Retrieving venues for GPS position: 52.427776, 13.280060
Got 5 venues. Total venues: 5
Retrieving venues for GPS position: 52.427776, 13.291259
Got 1 venues. Total venues: 6
```

The results were stored into a Pandas-Dataframe.
Now, the most frequent venue-categories were extracted:

```
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
```

The same process was applied for the City of New York to cover a major US city as well:

```
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
```

Now we know which are the most common venue categories in New York and Berlin. Now we will loop through the individual top-categories of the two cities and see how popular they are in the other city. The following lists will list how frequent (in percent) the accoring

	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

category occur in the respective city. It will also show the ratio of the category between the two cities. If e.g. Bars occur in NY twice as often as in Berlin, the ration will be equal to 2.

If we observe the NY dataframe, the ratio will be quite interesting. A ratio >1 indicates that the according category occurs more often in New York. A ration <1 indicates that the according category occurs more often in Berlin.

	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 are 3.4 times more frequent in Berlin than in NY. You will find German Restaurants 23.8 times more often in Berlin than in NY. The famous Doener (actually

Turkish cuisine) is quite popular in Berlin (1% of all venues are Doener Restaurants) whereas it is not known in New York at all.

Now, let's sort the NY dataframe by the NY/Berlin Ratio to see, which are categories that are more popular in New York than in Berlin. Before we do that, we want to drop the "inf" values.

	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	25.672200
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

So the category that occur much more often in New York than in Berlin are Food Trucks, Caribbean Restaurant, New American Restaurant, etc.

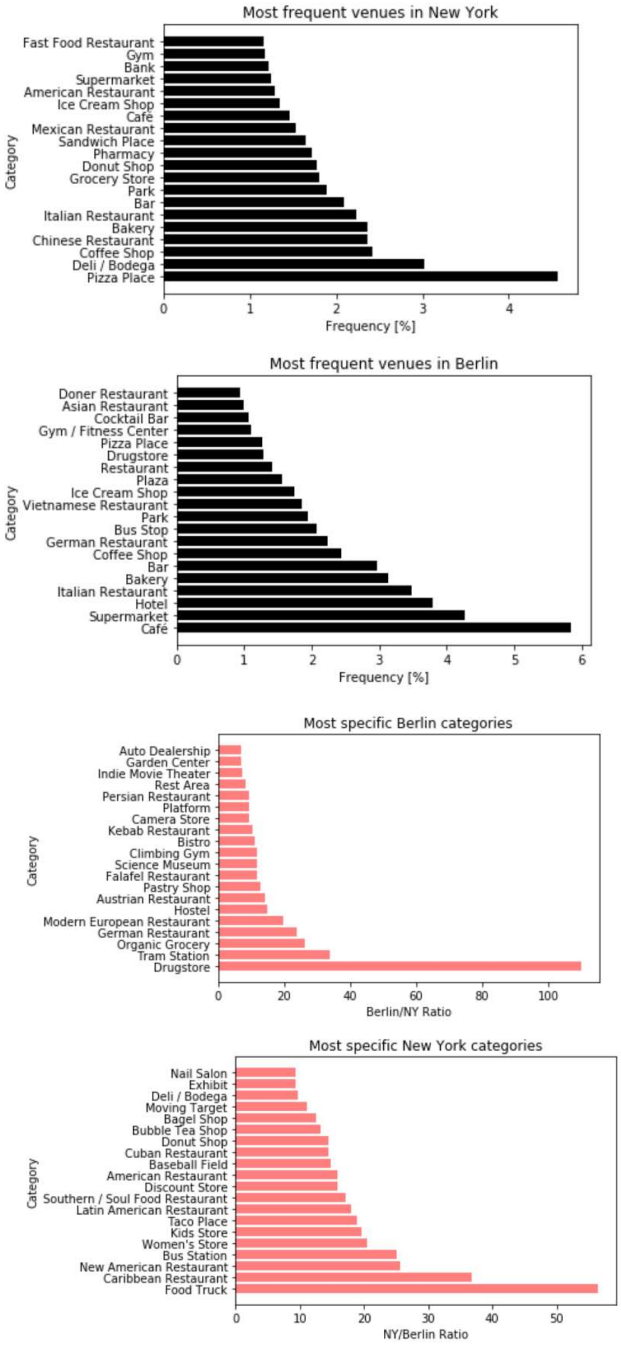
A few categories must be neglected as the naming is just different. Note the "Bus Station" Category is called "Bus Stop" in Berlin. I assume there are not much more bus stops in NY than in Berlin (speaking of normalized measures).
Some of the categories could be expected to occur more frequently in US cities such as Caribbean Restaurant, Latin American Restaurant or Baseball Field.
Now let's examine from the Berlin Perspective.

	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

Also here, many results could be expected. E.g. there are much more German Restaurants, Austrian Restaurant or Kebab (aka Doener) Restaurants in Berlin than in NY. But also there are some surprises such as "Science Museum", "Climbing Gym" or "Garden Center".

3.3 Results / Data Visualization

In this section we will visualize the results that we found earlier. Let's visualize Berlin's and New York's Top Categories



4. Discussion

In this analysis, the differences between German and US cities has been examined. The analysis shows some interesting results. Some of the results were expected, some were not. In the following section three areas shall be analyzed closer.

4.1 Eating

Both cities show a high concentration of Italian Food ("Italian Restaurant", "Pizza Place"). So italian food seems to be pretty international. When it comes to asian food, New York has a high concentration of Chines Restaurants, whereas Berlin shows a higher concentration of Vietnamese Restaurants. Both cities show a high concentration of local restaurants ("German Restaurant", "American Restaurant"). Berlin shows the particularity to have many Turkish Restaurants ("Doener","Kebeb") whereas New York shows many "Mexican Restaurants". In general both cities show a higher conentration of restaurants from "close-by" countries compared to the oter cities. So in Berlin "Austrian" Restaurants are common, compared to New York, but show low occurence in General. The same applies for "Caribbean Restaurants" in New York

4.2 Supermarkets/Groceries

Supermarkets occur 3.4 more frequently in Berlin compared to New York. This can be explained by the size of supermarkets. In Germany one would rather find smaller supermarket but better distributed over the city, whereas the US shows bigger supermarkets bundled at certain locations. This obersvation even holds true if you consider the category "Grocery store" along with the category "supermarket" in New York

4.3 Leisure Time

People in Berlin and New York seem to have different habbits when it comes to how they spend their leisure time. Parks and Gyms show an equal concentration in both cities. However Germans seem to like Cafés more than Americans. This might be because of more frequent pedestrian zones in German cities. In New York, Climbing Gyms seem quite rare. In Berlin the concentration is about ten times higher. The opposite applies for Baseball fields, which are much more common in New York

5. Conclusion

The report shows the main differences and commonality between US and German Cities. Also it gives a look into the history of the city. For example vietnamese restaurants in Berlin are frequent because of good relationships of former eastern Germany and former socialist vietnam. A further analysis of more US and German cities would be interesting to further prove the discovered results.