Battle of the Neighborhoods -What to Eat in a Financial Metropolis?

March 10, 2021

by Lutz Wimmer

0.1 Table of Contents:

0.2 Introduction

Global financial centers are some of the most wealthy regions. They are spacially very concentrated and a hot spot for luxury products. In the aftermath of the coronavirus pandemic, thousands of restaurants shut down for good, even in the busy financial centers. But ones downfall is the others chance and people always need places to go and eat. New restaurants are needed! **The idea is to open a restaurant in Frankfurt am Main** (short: FFM), Germany. As a restaurant, especially nowadays, you want to minimize risk. So this is part of a whole **market study** an investor could rely on for his decisions. So what kind of restaurant? What does the international financial elite love to eat? Is there a prevalent type of restaurant between those cities? Let's find out!

0.3 The Data

0.3.1 Geolocation Data

Different cultures, different food, right? But do mondane people favor their native kitchen? Or do they all agree on one trendy type of food all over the world? To get a better overview, six cities with big financial centers will be analyzed: **New York, Toronto, Tokyo, London, Paris and Frankfurt**. The data used will be provided by foursquare.com, automatically downloaded via their API. The focus will lie on restaurants only, excluding cafes, bars, bistros and so on.

Adding "&query=food" to the Foursquare API url will get us only food-related venues. This way the search radius can be expanded without hitting the limit of 100 venues per city. The request and subsequent data frame building is looped over all 5 cities. A radius of 10 km for example would be too high and include restaurants far outside FFM and thus too far away to travel to for a business lunch. The radius of 7 km seems to be a good fit between the maximum number of venues and travel distance.

```
<ipython-input-2-112cf702a079>:42: FutureWarning: pandas.io.json.json_normalize
is deprecated, use pandas.json_normalize instead
  nearby_venues = json_normalize(venues) # flatten JSON
```

[2]:

```
                      name                categories        lat        lng  \
0                 Manhatta  New American Restaurant  40.707654 -74.009138
1                Crown Shy               Restaurant  40.706187 -74.007490
2               sweetgreen              Salad Place  40.705626 -74.008282
3            Luke's Lobster       Seafood Restaurant  40.704488 -74.010915
4  Pisillo Italian Panini            Sandwich Place  40.710530 -74.007526

       city
0  New York
1  New York
2  New York
3  New York
4  New York
```

It would also been possible to store all the cities in one data frame, distinguishable by the "city" column. But I prefer to have them separated within a superstructure i can loop over. I guess it is just a matter of taste.

0.3.2 Data Refinement

Cleaning Restaurant Categories Although bistros and cafes also serve food, for simplicity, we use restaurant-type categories only.

[3]:

```
                                  name                categories  \
0                                Canoe                Restaurant
1                     Richmond Station       American Restaurant
2                                  Pai           Thai Restaurant
3  The Keg Steakhouse + Bar - York Street                Restaurant
4                       Byblos Toronto  Mediterranean Restaurant

        lat        lng     city
0  43.647452 -79.381320  Toronto
1  43.651569 -79.379266  Toronto
2  43.647923 -79.388579  Toronto
3  43.649987 -79.384103  Toronto
4  43.647615 -79.388381  Toronto
```

Some restaurants don't have a specification for their kitchen, simply called "Restaurant". These will be assigned a "miscellaneous" type.

```
Number of restaurants in New York :   60
Misc. Restaurants: 1
Number of restaurants in Toronto :   52
Misc. Restaurants: 7
Number of restaurants in Tokyo :   75
Misc. Restaurants: 1
Number of restaurants in London :   49
Misc. Restaurants: 7
Number of restaurants in Paris :   61
Misc. Restaurants: 4
Number of restaurants in Frankfurt :   48
Misc. Restaurants: 3
```

[5]:

```
                                     name                categories  \
0                                   Canoe  Miscellaneous Restaurant
1                        Richmond Station       American Restaurant
2                                     Pai           Thai Restaurant
3  The Keg Steakhouse + Bar - York Street  Miscellaneous Restaurant
4                          Byblos Toronto  Mediterranean Restaurant

         lat        lng     city
0  43.647452 -79.381320  Toronto
1  43.651569 -79.379266  Toronto
2  43.647923 -79.388579  Toronto
3  43.649987 -79.384103  Toronto
4  43.647615 -79.388381  Toronto
```

The word “Restaurant” isn’t needed anymore and will interfer with future visualization. Therefore we simply drop it.

[6]:

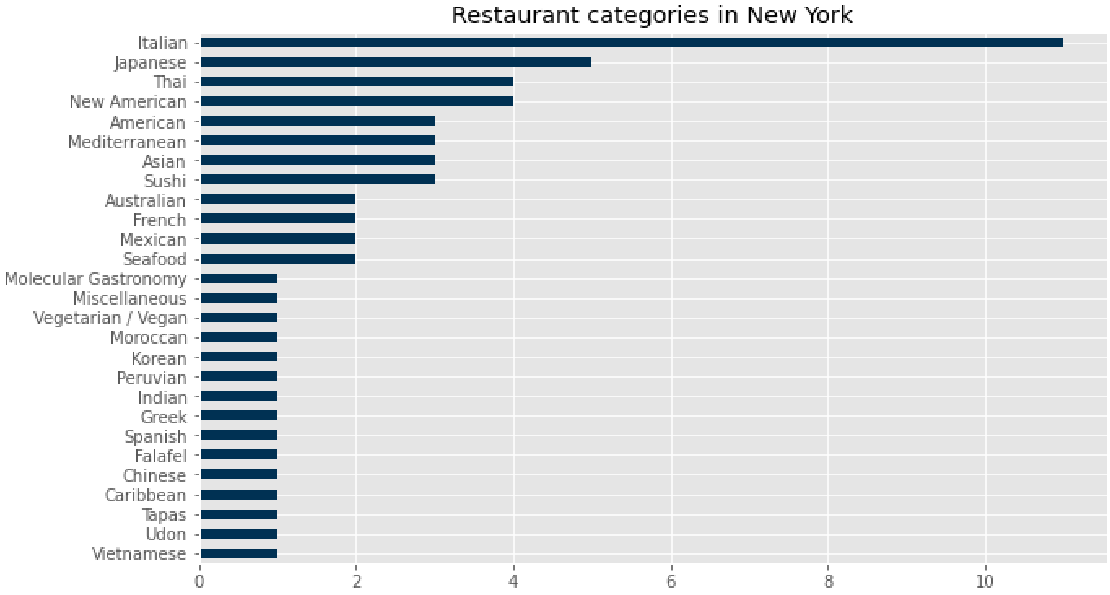
```
                                     name     categories        lat  \
0                                   Canoe  Miscellaneous  43.647452
1                        Richmond Station       American  43.651569
2                                     Pai           Thai  43.647923
3  The Keg Steakhouse + Bar - York Street  Miscellaneous  43.649987
4                          Byblos Toronto  Mediterranean  43.647615

         lng     city
0 -79.381320  Toronto
1 -79.379266  Toronto
2 -79.388579  Toronto
3 -79.384103  Toronto
4 -79.388381  Toronto
```

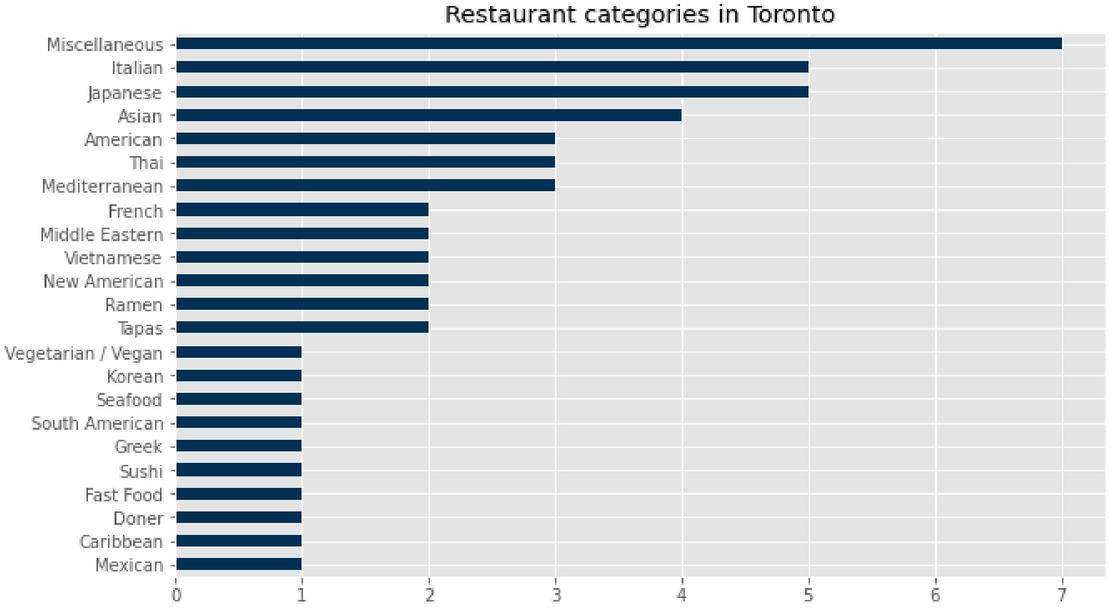
Finally, a map of the restaurants to see if we really got all the data we wanted or maybe even too much.

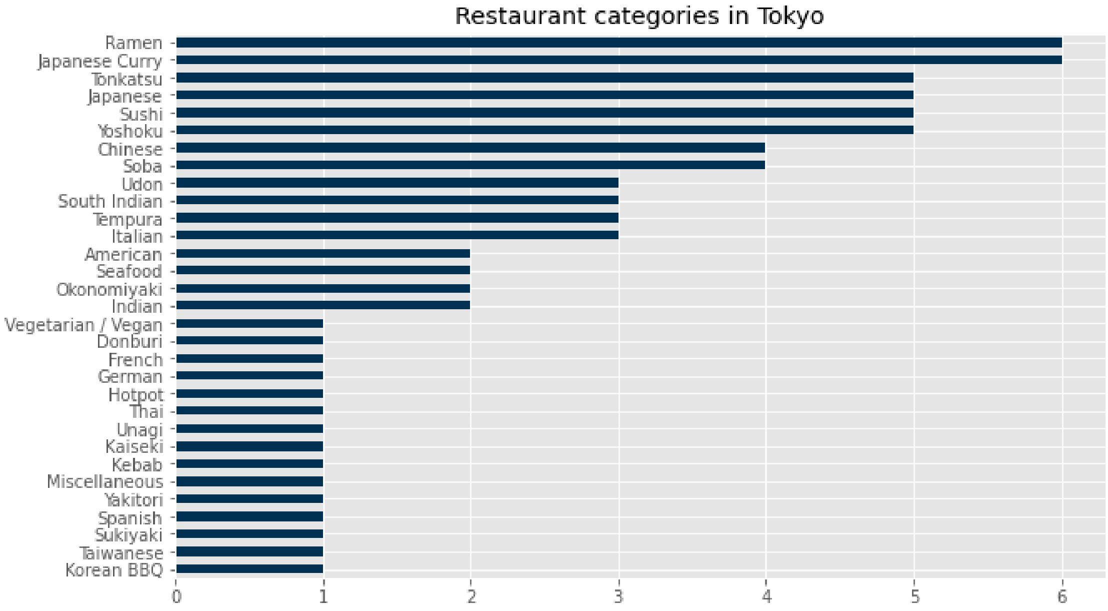
[7]:

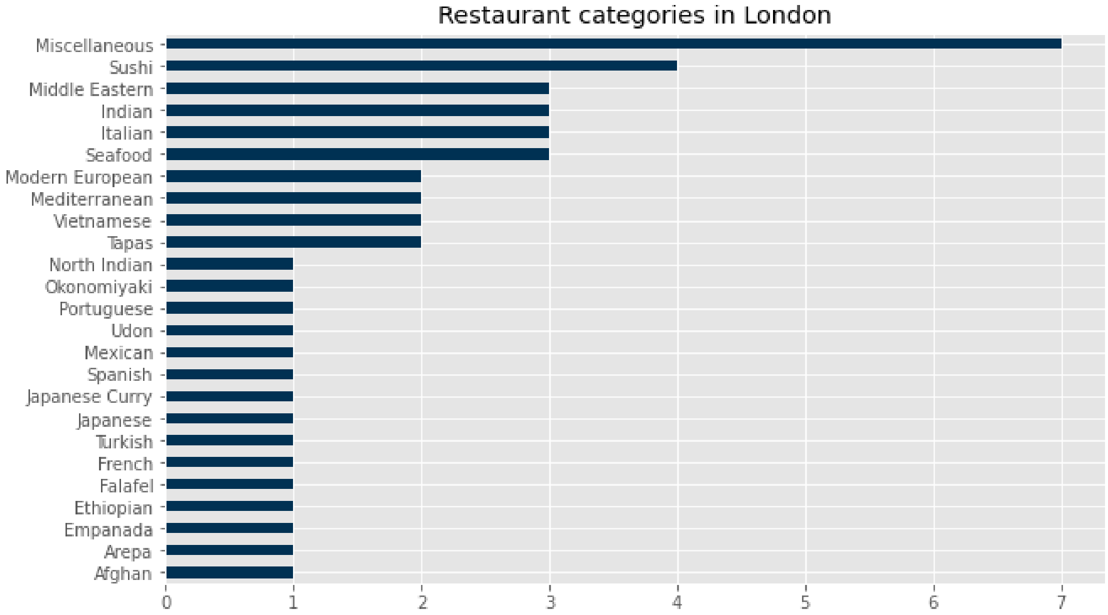
```
<folium.folium.Map at 0x28bf6c6ff70>
```

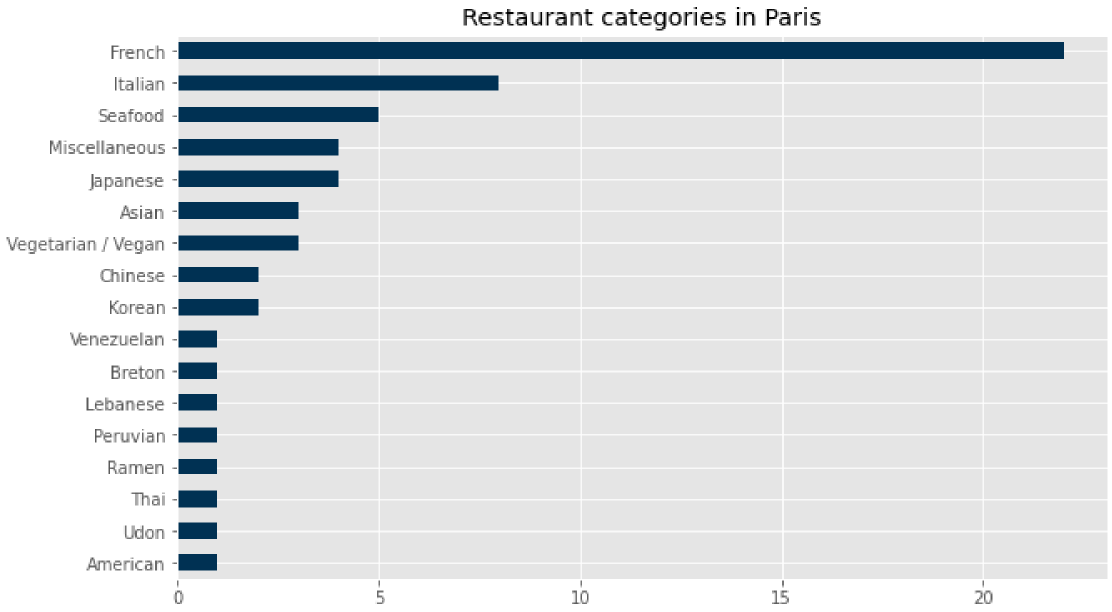
Consolidation of Redundant Categories First, let's visualize the data as horizontal bar plot.

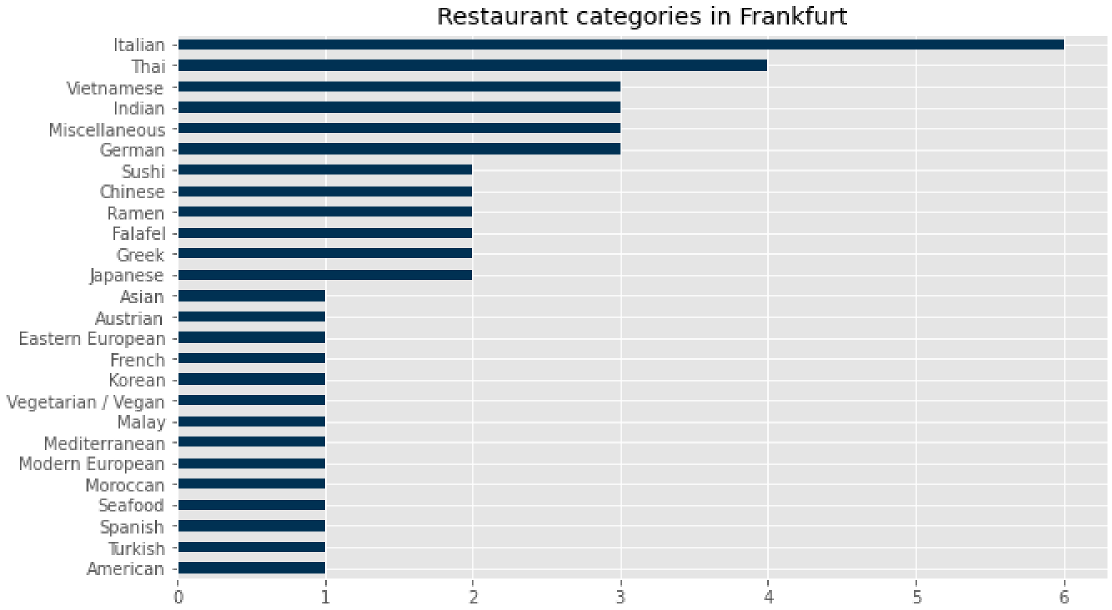





Restaurant categories in Tokyo
Ramen
Japanese Curry
Tonkatsu
Japanese
Sushi
Yoshoku
Chinese
Soba
Udon
South Indian
Tempura
Italian
American
Seafood
Okonomiyaki
Indian
Vegetarian / Vegan
Donburi
French
German
Hotpot
Thai
Unagi
Kaiseki
Kebab
Miscellaneous
Yakitori
Spanish
Sukiyaki
Taiwanese
Korean BBQ
0
1
2
3
4
5
6


Restaurant categories in London
Miscellaneous
Sushi
Middle Eastern
Indian
Italian
Seafood
Modern European
Mediterranean
Vietnamese
Tapas
North Indian
Okonomiyaki
Portuguese
Udon
Mexican
Spanish
Japanese Curry
Japanese
Turkish
French
Falafel
Ethiopian
Empanada
Arepa
Afghan
0
1
2
3
4
5
6
7





We can see, that in **Paris and Tokyo, the respective national kitchen** is by far the most favourite one, whereas in **New York and Franfurt the Italian cuisine** dominates. Another similarity is found between **London and Toronto with "miscellaneous" Restaurants being the top one**, followed by Italian cuisine. Now the "miscellaneous" category is a real problem here, since we don't know what kind of food they serve and we don't know if the restaurant type just wasn't specified by the Foursquare users despite belonging to a certain type. Another "problem" is Tokyo, or to be precise: the Japanese cuisine. Japanese restaurants are often more like small street food venues, serving just a specific Japanese dish. For comparability reasons, all the single Japanese dishes will be merged into the “Japanese” type, with exception of “Sushi”. Sushi places are found in every city and quite popular, except in Paris, where it might fall into the category “Seafood”. There are several more examples of duplicity, but we have to be careful not to simplify the data too much, since a specific subtype, like “Tapas”, might be much more favoured than the regional type it belongs to. Let’s start with Tokyo.

With the categories now nice and clean, we can finally start to analyze the similarities between the cities.

0.4 Methodology

First we try to establish the similarity between the cities. To do so, the Pearson correlation coefficient *r* is calculated. Since Frankfurt is our point of reference, we calculate the *r* value between Frankfurt and the other 5 cities. This gives us the advantage to ignore restaurant categories, that are not present in Frankfurt, but the other city it is paired to.

```
[16]:                Frankfurt  Paris  London  New York  Toronto  Tokyo
      categories
      Chinese                2      2       0         1        0      4
      Sushi                  2      0       4         3        1      5
      German                 3      0       0         0        0      1
      Miscellaneous          3      4       7         1        7      1
      Indian                 3      0       3         1        0      5
      Vietnamese             3      0       2         1        2      0
      Thai                   4      1       0         4        3      1
      Italian                6      8       3        11        5      3
```

Now we can run the Perason correlation provided by scipy. It is important to note, that the data for the correlation does not include *nan* values and both variables (cites) have the same data length. The correlation results are shown in the **Results** section. Another possibility is to include all categories over all cities, so we can calculate the correlation between all cities. For that, we only have to alter the data frame a bit.

```
[17]:             Frankfurt  Paris  London  New York  Toronto  Tokyo
      Afghan              0      0       1         0        0      0
      American            1      1       0         7        5      2
      Asian               1      3       0         3        4      0
      Australian          0      0       0         2        0      0
      Austrian            1      0       0         0        0      0
      Caribbean           0      0       0         1        1      0
      Chinese             2      2       0         1        0      4
      Doner               0      0       0         0        1      0
```

The difference here is, that the “Frankfurt” column also has 0-values and it includes all categories.

0.5 Results

0.5.1 Similarity between Cities

The Pearson correlation between the cities paired to Frankfurt point of reference are as followed:

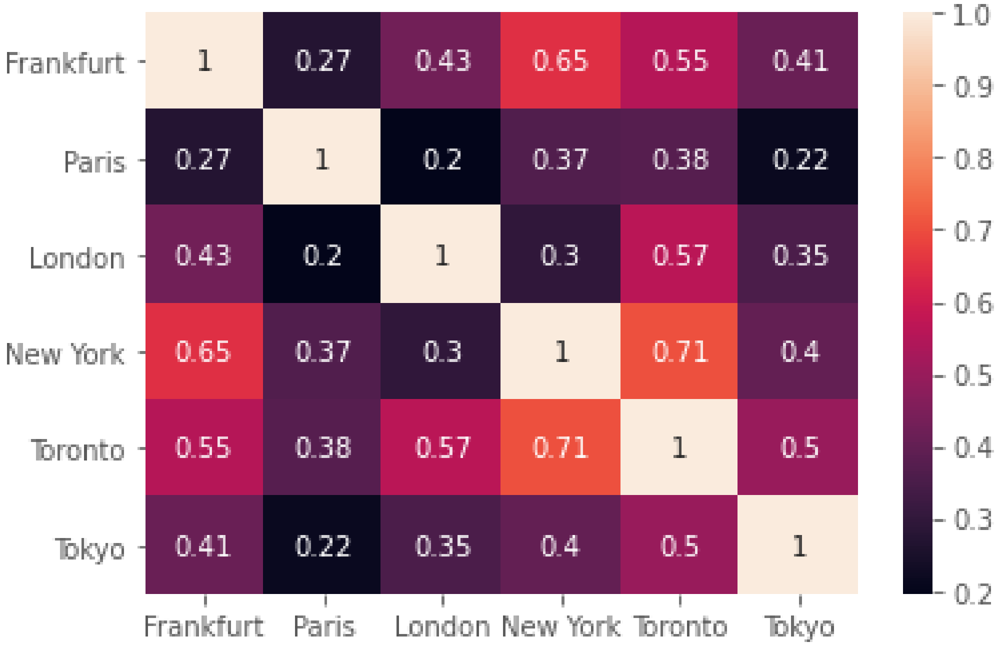
```
Frankfurt - Paris r: 0.115 | p-val.: 0.584
Frankfurt - London r: 0.433 | p-val.: 0.03
Frankfurt - New York r: 0.606 | p-val.: 0.001
Frankfurt - Toronto r: 0.498 | p-val.: 0.011
Frankfurt - Tokyo r: 0.386 | p-val.: 0.057
```

The highest correlation is found between Frankfurt and New York with **r = 0.6** and a **p-value < 0.05**, followed by Toronto (r ~ 0.5) and London (r = 0.43). Tokyo shows also a high correlation (r ~ 0.4) but it is not significant (p-value > 0.05). Most interesting is the low correlation between Frankfurt and Paris (r = 0.1) with a whopping p-value of 0.6, meaning not only that they are very dissimilar but also that the similarity they share is likely to be random. Now if we look at the second approach, correlating all cities with each other, we can easily calculate a correlation-matrix out of the data frame.

[19]:

```
            Frankfurt     Paris    London  New York   Toronto     Tokyo
Frankfurt    1.000000  0.273993  0.429313  0.650864  0.554981  0.413552
Paris        0.273993  1.000000  0.195868  0.367792  0.375848  0.217794
London       0.429313  0.195868  1.000000  0.296450  0.565341  0.354847
New York     0.650864  0.367792  0.296450  1.000000  0.705790  0.401946
Toronto      0.554981  0.375848  0.565341  0.705790  1.000000  0.503479
Tokyo        0.413552  0.217794  0.354847  0.401946  0.503479  1.000000
```

This can be visualized with a heat map using the *seaborn* library



This shows us that the highest correlation is found between New York and Toronto, not very surprising given their proximity. Correlation between Frankfurt and New York has slightly increased to **0.65**, due to the higher number of categories taken into account. The question now is, how to use this information about the similarity between Frankfurt and other cities to recommend a restaurant category. Another differentiation should be made between the "strategy" of the investor. This comes down to two different approaches: low risk and high risk.

0.6 Discussion

0.6.1 Low Risk Approach

"Low risk" in this context means that the recommendation for a restaurant category is based on what is already popular in Frankfurt and highly correlated cities. This investment poses the lowest risk of failing dependant on the restaurant category alone. This can also be subdivided into two different approaches:

a) we either recommend what is popular in both, Frankfurt and the other city. Or **b)** we recommend what is popular in highly correlated cites but not in Frankfurt.

This way the investor can open a restaurant that doesn't has much competition in Frankfurt, but is popular in cities similar to Frankfurt. (b) seems to be the best way to recommend a restaurant category, since recommending to open Italian restaurant number 7 probably won't fly with the investor.

We simply subtract the restaurant categories percentage of Frankfurt from the other cities. By that we get values between -1 and 1 with 1 being a category much more prevalent in Frankfurt and -1 much more prevalent in the other city. "0" means that either both cities have the same percentage of this restaurant category or neither has one. We will compare Frankfurt to New York and Toronto alone with both having a $r > 0.5$.

```
[21]: American                -0.097811
      Italian                 -0.061441
      Australian              -0.033898
      Mexican                 -0.033898
      Asian                   -0.030014
      Mediterranean           -0.030014
      Japanese                -0.018362
      Caribbean               -0.016949
      Molecular Gastronomy    -0.016949
      Peruvian                -0.016949
      Tapas                   -0.016949
      Seafood                 -0.013065
      French                  -0.013065
      Sushi                   -0.009181
      Afghan                   0.000000
      dtype: float64
```

So the **low risk (b)** recommendation from the comparison with New York would clearly be an **"American" restaurant** with only 1 American restaurant in Frankfurt. **"Italian"** restaurant would be definetely be the recommendation for the **low risk (a)** approach, being the most popular category in both cities.

```
[22]: American          -0.075321
      Miscellaneous     -0.072115
      Asian             -0.056090
      Japanese          -0.051282
      Middle Eastern    -0.038462
      Tapas             -0.038462
      Mediterranean     -0.036859
      Doner             -0.019231
      Mexican           -0.019231
      Fast Food         -0.019231
      Caribbean         -0.019231
      South American    -0.019231
      French            -0.017628
      Peruvian           0.000000
      Kebab              0.000000
      dtype: float64
```

The **low risk (b)** recommendation from Toronto is either an American or an "Miscellaneous" restaurant with 3 of latter being already in Frankfurt. But "Miscellaneous" could mean a lot and doesn't provide much information for the investor. The **low risk (a)** recommendation would be an "Asian" or "Japanese" restaurant. Since there are already a lot of Asian, Thai, Japanese and so on restaurants in Frankfurt, these restaurants seem to be a real evergreen in Frankfurt.

0.6.2 High Risk Approach

Under "high risk" we understand a recommendation of a category that isn't found in Frankfurt, but popular in Toronto or New York. Considering categories from cities with low similarity to Frankfurt and not existing there could be defined as "maximum risk" approach.

We simply subtract the category names of Frankfurt from New York/Toronto to get the categories only found in New York, then look at the amount of those categories in New York/Toronto.

```
[23]: Mexican                 2
      Australian              2
      Tapas                   1
      Molecular Gastronomy    1
      Peruvian                1
      Caribbean               1
      Name: New York, dtype: int64
```

```
[24]: Tapas             2
      Middle Eastern    2
      South American    1
      Fast Food         1
      Doner             1
      Mexican           1
```

```
       Caribbean         1
       Name: Toronto, dtype: int64

[25]: Tapas                      3.0
       Mexican                    3.0
       Caribbean                  2.0
       Australian                 NaN
       Doner                      NaN
       Fast Food                  NaN
       Middle Eastern             NaN
       Molecular Gastronomy       NaN
       Peruvian                   NaN
       South American             NaN
       dtype: float64
```

The combined **high risk** recommendation from New York and Toronto would be either a **Tapas** or a **Mexican** restaurant, with "Carribean" also being an alternative.

0.7 Conclusion

The investor now has his possible restaurant categories narrowed down to just a few (**American, Italian, Miscellaneous, Asian/Japanese, Tapas or Mexican**). Ultimately, he will decide between one of those based on their profit potenial. To estimate the potential profit of a restaurant category is very complex and many aspects go into that (e.g. the cost of food of that cuisine). Therefore it is vital to have just a narrow selection that needs to be analyzed. This approach is very similiar to "collaborative filtering" used by video streaming services like Netflix. Analoguous to the Netflix user, we evaluated the "preferences" of a whole city, compared them to the preferences of others and recommended based on different approaches.

Big thanks to IBM for this awesome course! It has been assured me that I really want to become a professional Data Science. Also shoutout to Coursera for being just the perfect platform for gaining skills!