Using Foursquare and eBay to analyze computer & hardware sellers in my area

Applied Data Science Capstone Project

Lukas Weidich

Contents

1 Introduction

1.1 Motivation

As a customer, I am always overwhelmed by all the buying options I am confronted with on a daily basis. Especially when looking for computers and hardware, there are seemingly endless stores and possibilities just around the corner. While these stores are professional businesses, I wonder whether private offerings are also that strongly represented within my local area. To find out, I will make use of business data provided by Foursquare and compare it to currently active eBay listings near me. By also visualizing the data, I hope to underline my findings even more.

1.2 Who will profit from the insights

Both, buyers and sellers, can profit from the results portrayed in this short report. Buyers can use the data to further understand the available options to choose from when buying computers or hardware. Sellers can increase their competitive advantages by getting to know their surroundings, such as other stores or active listings nearby. Also, anyone interested in the vast field of data science is able to get an impression of what to expect from the IBM Data Science Professional Certificate.

2 Data

I will use two different datasets to gain more insights into the questions I posed above. One will be Foursquare, which was introduced during the course. The other data will be provided by eBay. I will download the eBay data and load the file from my local directory, so I do not need to expose my credentials within the coding. The Forsquare data however will be fetched within the code cells.

2.1 Foursquare

Foursquare will provide general information about professional sellers and businesses offering computers in my area. I will use the location (longitude, latitude, city name) to cluster the data. Additionally, the entries are visualized by being shown on a map. Besides the given columns, I will also one hot encode the type of business (professional or private), depending on the entry itself.

2.2 eBay

By making use of the eBay API, I can access and filter all eBay listings near me that are currently active. Similar to the Foursquare data, I am mainly interested in the location of the seller, the sellers name and whether the seller is a private person or a professional business.

2.3 Additional information

Of course, I will use the same search radius for both API calls, so I have a completely comparable dataset to work with. However, the Foursquare call is limited to 200 results.

3 Methodology

3.1 Retrieving the data

I can access the active eBay listings near me by fetching data recursively from the eBay API. I wrote a small seeder script that stores the data in csv format and saves it locally. From there, I can easily read the csv in my notebook. The Foursquare data will be loaded within the code cells.

3.2 Handling missing information and combining the two datasets

After fetching the data, it is time to combine and unify the datasets, so I can work with coherent data in the next steps. For my purposes, it will be sufficient to have the following columns for both datasets:

- Name of business/seller/store/item

- City

- State

- Latitude

- Longitude

- Source (eBay/Foursquare)

This is what the two datasets look like after fetching:

```
df_foursquare.head()
```

	Name	Latitude	Longitude	Source
0	Computer Service Dröge	52.279037	8.902915	Foursquare
1	TM Computer e.K.	52.116941302856596	8.67265034442815	Foursquare
2	TM Computer e.K. - Thomas Müller (Büro)	52.15775752422116	8.63429950693619	Foursquare
3	Schormann-Computer	52.18319156583845	8.874798774686496	Foursquare
4	4you-computer.de	52.24314880371094	8.840249061584473	Foursquare

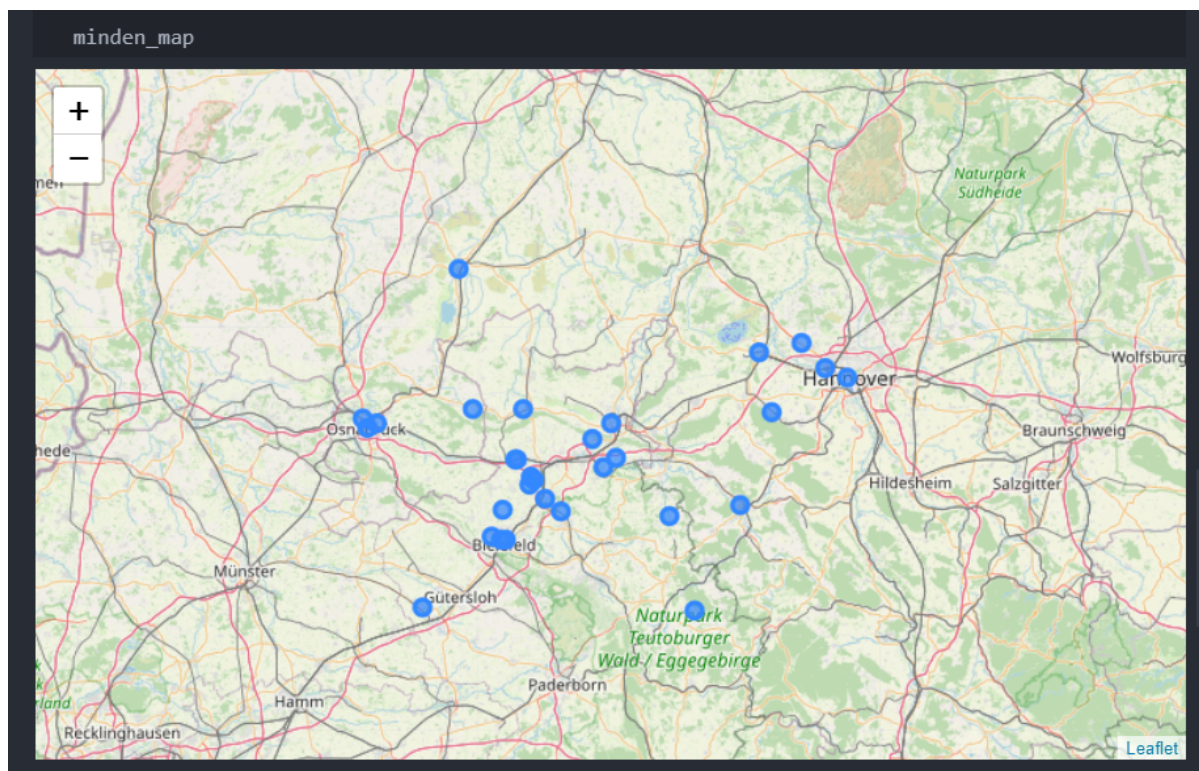Figure 3.1: Foursquare dataset right afer fetching with some columns already removed

Figure 3.2: Foursquare dataset visualized in map

Figure 3.3: Excerpt of eBay dataset right afer fetching

The Foursquare dataset is missing city and state, while the eBay dataset is missing coordinates unfortunately, as well as state. In the next step, I will preprocess the data and add the missing columns.

To preprocess the data, I will use OpenStreetMap[1]. It provides a reverse geocoding function, allowing to search for locations for example by name or coordinates and receiving vast information in return. See below for the code and result.

---

[1]https://geocoder.readthedocs.io/providers/OpenStreetMap.html

```
# add columns city, state to foursquare df
states = []
cities = []
for lat, lng in zip(df_foursquare["Latitude"], df_foursquare["Longitude"]):
    g = geocoder.osm(str(lat+","+lng)).json
    states.append(g["state"])
    if("town" in g):
        cities.append(g["town"])
    elif("city" in g):
        cities.append(g["city"])
    else:
        cities.append("-")
df_foursquare["State"] = states
df_foursquare["City"] = cities

df_foursquare.head()
```

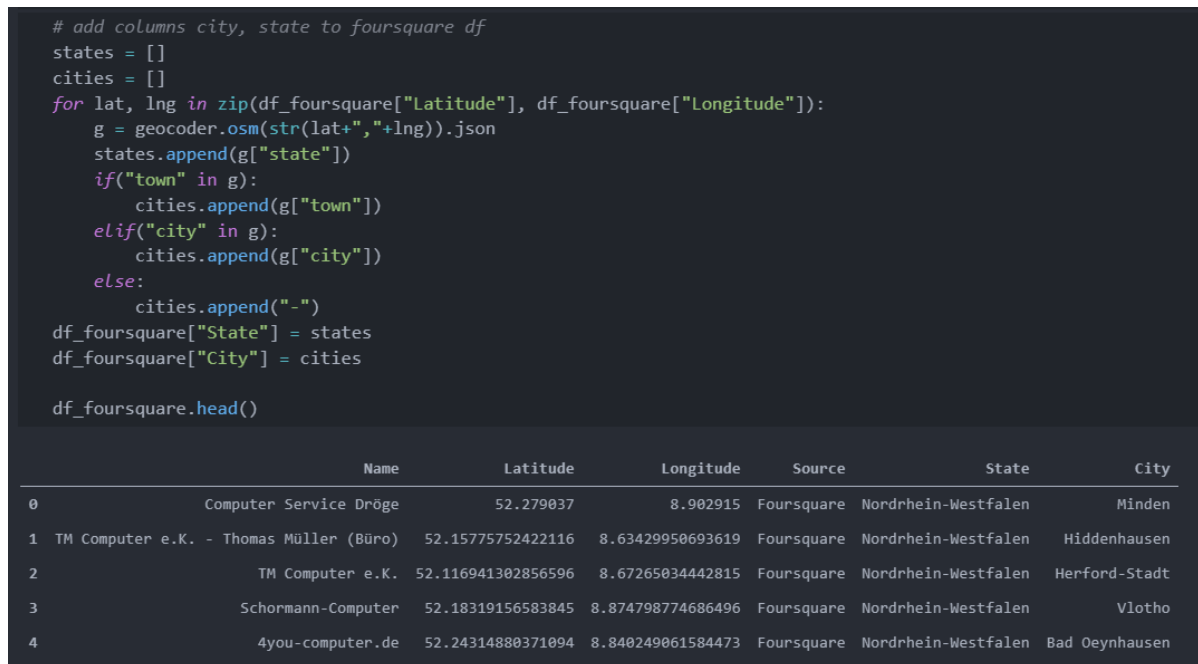|	Name	Latitude	Longitude	Source	State	City
0	Computer Service Dröge	52.279037	8.902915	Foursquare	Nordrhein-Westfalen	Minden
1	TM Computer e.K. - Thomas Müller (Büro)	52.15775752422116	8.63429950693619	Foursquare	Nordrhein-Westfalen	Hiddenhausen
2	TM Computer e.K.	52.116941302856596	8.67265034442815	Foursquare	Nordrhein-Westfalen	Herford-Stadt
3	Schormann-Computer	52.18319156583845	8.874798774686496	Foursquare	Nordrhein-Westfalen	Vlotho
4	4you-computer.de	52.24314880371094	8.840249061584473	Foursquare	Nordrhein-Westfalen	Bad Oeynhausen

Figure 3.4: Foursquare data with added state and city

The eBay data is missing state, latitude and longitude. These columns will also be added by using OpenStreetMap. See the following results.

```
df_ebay_corrected.head()
```

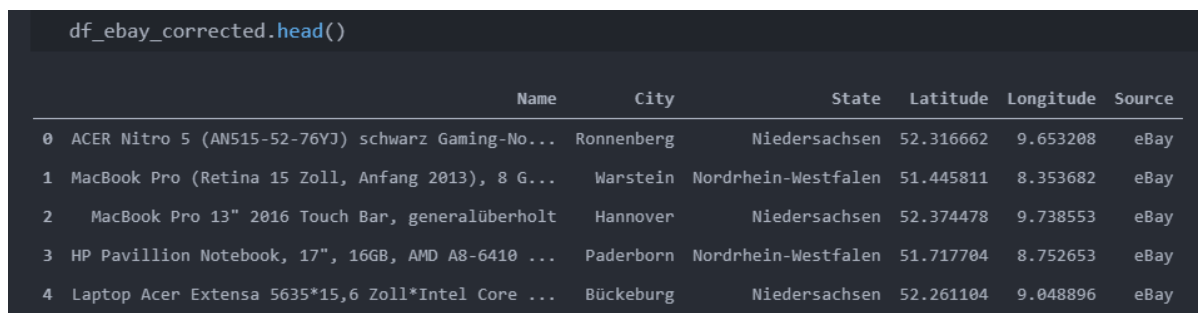|	Name	City	State	Latitude	Longitude	Source
0	ACER Nitro 5 (AN515-52-76YJ) schwarz Gaming-No...	Ronnenberg	Niedersachsen	52.316662	9.653208	eBay
1	MacBook Pro (Retina 15 Zoll, Anfang 2013), 8 G...	Warstein	Nordrhein-Westfalen	51.445811	8.353682	eBay
2	MacBook Pro 13" 2016 Touch Bar, generalüberholt	Hannover	Niedersachsen	52.374478	9.738553	eBay
3	HP Pavillion Notebook, 17", 16GB, AMD A8-6410 ...	Paderborn	Nordrhein-Westfalen	51.717704	8.752653	eBay
4	Laptop Acer Extensa 5635*15,6 Zoll*Intel Core ...	Bückeburg	Niedersachsen	52.261104	9.048896	eBay

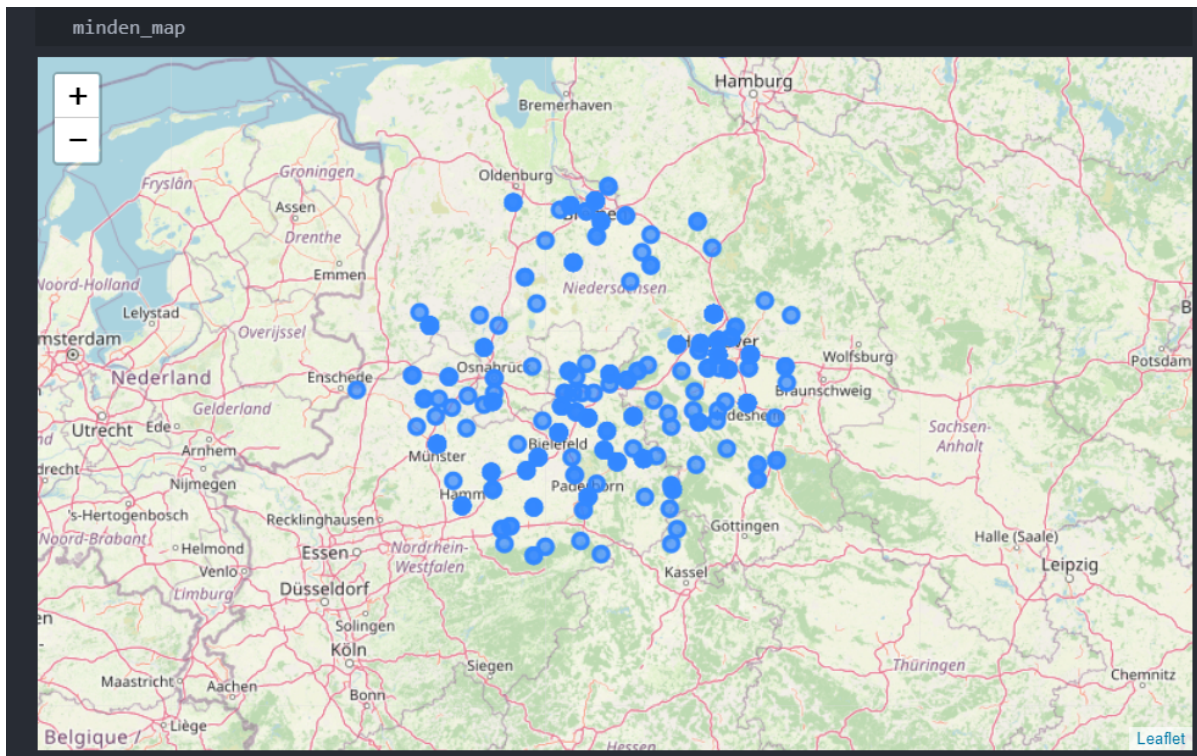Figure 3.5: eBay data with added state and coordinates

Figure 3.6: eBay dataset visualized in map

The two datasets are now combined and stored as a csv locally, to make sharing the data easier. This coherent structure now allows for some exploratory data analysis in the following subsections.

3.3 Exploratory data analysis

In this section, I will demonstrate how I managed to get a general overview over the data. This chapter only covers methodology, the results are discussed in 4.

3.3.1 Statistical key figures and indicators

To begin with some basic operations after creating a combined dataset, I described the given dataframe.

	Name	City	State	Latitude	Longitude	Source
count	630	630	630	630.000000	630.000000	630
unique	598	139	10	NaN	NaN	2
top	15.6 Zoll Laptop 4GB RAM 64GB SSD Intel Celero...	Wedemark	Niedersachsen	NaN	NaN	eBay
freq	4	106	345	NaN	NaN	600
mean	NaN	NaN	NaN	52.343838	8.919098	NaN
std	NaN	NaN	NaN	0.703020	1.630962	NaN
min	NaN	NaN	NaN	45.947330	-3.272988	NaN
25%	NaN	NaN	NaN	52.021274	8.531007	NaN
50%	NaN	NaN	NaN	52.374478	9.120040	NaN
75%	NaN	NaN	NaN	52.562287	9.703703	NaN
max	NaN	NaN	NaN	54.740692	11.954355	NaN

Figure 3.7: Description of dataset

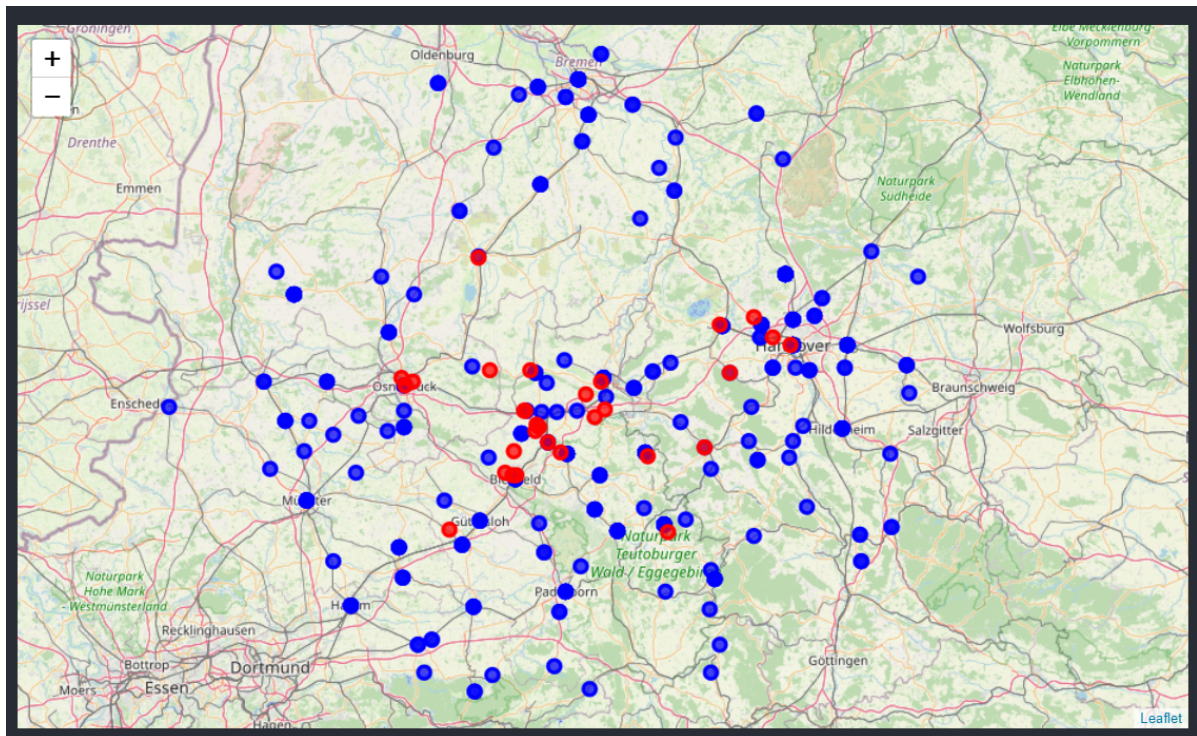## 3.3.2 Visualizing entries using a map

Figure 3.8: Visualization of dataset

8

This is what the map looks like when both sources are considered. Foursquare entries are colored red, eBay ones are blue.

3.4 Inferential statistical testing

3.4.1 Similarity between data sources

3.5 Machine learning

3.5.1 Clustering entries

4 Results

Results section where you discuss the results.

5 Discussion

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

6 Conclusion

Conclusion section where you conclude the report.