The Battle of Neighborhoods in Europe: Choosing A City For Relocation

By Sergei Perfilyev. June 2021

This notebook has been created for completing my capstone project for IBM/Coursera Applied Data Science specialization.

The code below is to implement the following tasks:

1. Collect, clean and prepare necessary data for our analysis. We are using several Python libraries including `Requests` for loading webpages, `BeautifulSoap` for parsing HTML, `Pandas` for tabular data processing, and etc. The objects to explore are: major European cities and top-ranked European universities.
2. Obtain geographical coordinates of various objects for further analysis. This is done with `GeoPy` library via HERE.COM API geocoder.
3. Visualize the results of our analysis using `Foluim` library for building geographical maps.
4. For the cities of interest, load info on recommended venues and their categorical distribution. This will be done with Foursquare REST API.
5. Prepare the data for processing and apply a clustering algorithm in order to group cities in partitions. We are using K-meand clustering method and `scikit-learn` Python package.

1.1. Install and import libraries for data preparation and processing

In [2]:

```
# Install BeautifulSoup library, we'll use it for HTML parsing
!conda install -c conda-forge bs4 --yes
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.
```

In [3]:

```
from bs4 import BeautifulSoup
import requests as rq
from requests.auth import HTTPDigestAuth
import json
import pandas as pd
import numpy as np
import re
from sklearn.cluster import KMeans
```

In [4]:

```
# Read my keys, passwords and secrets for various APIs and websites
with open('my_secrets.json', 'r') as infile:
    my_secrets = json.load(infile)
```

In [5]:

```
# Define a function to handle http responses
def get_html(response: rq.models.Response):
    if response.status_code // 100 == 2:
        print('HTTP request OK!')
        return response.text
    else:
        print('HTTP request failed!')
        return None
```

1.2. Prepare a dataset of European Union (EU) member countries

In [6]:

```
# Let's get data from Eurostat official website
url_countries = "https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:
Country_codes"
html_countries = get_html(rq.get(url_countries))

# Parse web page and create a data frame containing all countries in the European Union
(EU)
if html_countries:
    soup = BeautifulSoup(html_countries)
    cells = [cell.text.strip() for cell in soup.find('table').find_all('td')]
    df_countries = pd.DataFrame(data={'Code': cells[1::2], 'Country': cells[0::2]})
    df_countries.sort_values(by=['Country'], inplace=True, ignore_index=True)
    df_countries.Code = df_countries.Code.apply(lambda x: re.sub(r'[()]', '', x))
    df_countries.loc[df_countries.Country == 'Czechia', 'Country'] = 'Czech Republic'
    df_countries.set_index('Country', inplace=True)
```

```
HTTP request OK!
```

In [7]:

```
df_countries
```

Out[7]:

Country	Code
Austria	AT
Belgium	BE
Bulgaria	BG
Croatia	HR
Cyprus	CY
Czech Republic	CZ
Denmark	DK
Estonia	EE
Finland	FI
France	FR
Germany	DE
Greece	EL
Hungary	HU
Ireland	IE
Italy	IT
Latvia	LV
Lithuania	LT
Luxembourg	LU
Malta	MT
Netherlands	NL
Poland	PL
Portugal	PT
Romania	RO
Slovakia	SK
Slovenia	SI
Spain	ES
Sweden	SE

1.3. Prepare a dataset of EU top universities

In [8]:

```
# Define a function to request data from Times Higher Educations's website
# Use my credentials for authentication
request_TimesHE = lambda url: rq.get(
    url,
    auth=HTTPDigestAuth(my_secrets['TIMES_HE']['LOGIN'], my_secrets['TIMES_HE']['PASSWO
RD']),
    headers={'user-agent': 'Mozilla/5.0'}
)
```

In [9]:

```
# Let's get data from Times Higher Education's website
url_univ = "https://www.timeshighereducation.com/student/best-universities/best-univers
ities-europe"
html_univ = get_html(request_TimesHE(url_univ))

# Parse web page and create a data frame containing top european universities
if html_univ:
    univs = BeautifulSoup(html_univ).find('table').find_all('tr')
    df_univs = pd.DataFrame()
    for row in univs[1::]:
        row_cells = [cell.text.strip() for cell in row.find_all('td')]
        row_cells = row_cells[1:4] + [row.find('a').get('href')]
        df_univs = df_univs.append([row_cells], ignore_index=True)
    df_univs.columns = ['Rank', 'University', 'Country', 'URL']
    df_univs.set_index('Country', inplace=True)
    print(df_univs.shape)
```

```
HTTP request OK!
(531, 3)
```

In [10]:

```
df_univs.head(10)
```

Out[10]:

	Rank	University	URL
Country			
United Kingdom	1	University of Oxford	https://www.timeshighereducation.com/world-uni...
United Kingdom	2	University of Cambridge	https://www.timeshighereducation.com/world-uni...
United Kingdom	3	Imperial College London	https://www.timeshighereducation.com/world-uni...
Switzerland	4	ETH Zurich	https://www.timeshighereducation.com/world-uni...
United Kingdom	5	UCL	https://www.timeshighereducation.com/world-uni...
United Kingdom	6	London School of Economics and Political Science	https://www.timeshighereducation.com/world-uni...
United Kingdom	7	University of Edinburgh	https://www.timeshighereducation.com/world-uni...
Germany	8	LMU Munich	https://www.timeshighereducation.com/world-uni...
United Kingdom	9	King’s College London	https://www.timeshighereducation.com/world-uni...
Sweden	10	Karolinska Institute	https://www.timeshighereducation.com/world-uni...

In [11]:

```
df_univs.tail(10)
```

Out[11]:

	Rank	University	URL
Country			
Lithuania	=418	Vytautas Magnus University	https://www.timeshighereducation.com/world-uni...
Poland	=418	Warsaw University of Life Sciences – SGGW	https://www.timeshighereducation.com/world-uni...
Poland	=418	Warsaw University of Technology	https://www.timeshighereducation.com/world-uni...
Czech Republic	=418	University of West Bohemia	https://www.timeshighereducation.com/world-uni...
Poland	=418	Wrocław University of Environmental and Life S...	https://www.timeshighereducation.com/world-uni...
Poland	=418	Wrocław University of Science and Technology	https://www.timeshighereducation.com/world-uni...
Poland	=418	University of Wrocław	https://www.timeshighereducation.com/world-uni...
Ukraine	=418	Yuriy Fedkovych Chernivtsi National University	https://www.timeshighereducation.com/world-uni...
Croatia	=418	University of Zagreb	https://www.timeshighereducation.com/world-uni...
Slovakia	=418	University of Žilina	https://www.timeshighereducation.com/world-uni...

In [12]:

```
# Select top universities from the countries included in the EU
TOP_TO_SEARCH = 200
df_eu_univs = df_countries.join(df_univs, how='inner').reset_index()
df_eu_univs.Rank = pd.to_numeric(df_eu_univs.Rank.str.replace('=', ''))
df_eu_univs = df_eu_univs.sort_values(by='Rank', ignore_index=True).head(TOP_TO_SEARCH)
df_eu_univs.head(10)
```

Out[12]:

	Country	Code	Rank	University	URL
0	Germany	DE	8	LMU Munich	https://www.timeshighereducation.com/world-uni...
1	Sweden	SE	10	Karolinska Institute	https://www.timeshighereducation.com/world-uni...
2	Germany	DE	11	Technical University of Munich	https://www.timeshighereducation.com/world-uni...
3	Germany	DE	12	Heidelberg University	https://www.timeshighereducation.com/world-uni...
4	Belgium	BE	14	KU Leuven	https://www.timeshighereducation.com/world-uni...
5	France	FR	15	Paris Sciences et Lettres – PSL Research Unive...	https://www.timeshighereducation.com/world-uni...
6	Netherlands	NL	17	Wageningen University & Research	https://www.timeshighereducation.com/world-uni...
7	Netherlands	NL	18	University of Amsterdam	https://www.timeshighereducation.com/world-uni...
8	Netherlands	NL	19	Leiden University	https://www.timeshighereducation.com/world-uni...
9	Netherlands	NL	20	Erasmus University Rotterdam	https://www.timeshighereducation.com/world-uni...

In [13]:

```
df_eu_univs.tail(10)
```

Out[13]:

	Country	Code	Rank	University	URL
190	Czech Republic	CZ	279	Masaryk University	https://www.timeshighereducation.com/world-uni...
191	Greece	EL	279	University of Thessaly	https://www.timeshighereducation.com/world-uni...
192	Greece	EL	279	Athens University of Economics and Business	https://www.timeshighereducation.com/world-uni...
193	Austria	AT	279	Johannes Kepler University of Linz	https://www.timeshighereducation.com/world-uni...
194	Austria	AT	279	University of Graz	https://www.timeshighereducation.com/world-uni...
195	Germany	DE	279	Hamburg University of Technology	https://www.timeshighereducation.com/world-uni...
196	Germany	DE	279	University of Kaiserslautern	https://www.timeshighereducation.com/world-uni...
197	Portugal	PT	279	University of Aveiro	https://www.timeshighereducation.com/world-uni...
198	Portugal	PT	279	University of Beira Interior	https://www.timeshighereducation.com/world-uni...
199	Portugal	PT	279	University of Coimbra	https://www.timeshighereducation.com/world-uni...

In [14]:

```
df_eu_univs.drop(df_eu_univs.columns[0], axis=1, inplace=True)
df_eu_univs.to_csv('top_eu_univs.csv')
```

In [15]:

```
# Parse detailed webpages and get addresses of all universities into a list of strings
details = [''] * df_eu_univs.shape[0]
for i in df_eu_univs.index:
    html_univ_detail = get_html(request_TimesHE(df_eu_univs.loc[i, "URL"]))
    if html_univ_detail:
        print(i, ': Details downloaded OK!')
        soup = BeautifulSoup(html_univ_detail)
        details[i] = soup.find(class_="institution-info__contact-detail institution-inf
o__contact-detail--address").text.strip()
    else:
        print(i, ': Could not download details!')
```

```
HTTP request OK!
0 : Details downloaded OK!
HTTP request OK!
1 : Details downloaded OK!
HTTP request OK!
2 : Details downloaded OK!
HTTP request OK!
3 : Details downloaded OK!
HTTP request OK!
4 : Details downloaded OK!
HTTP request OK!
5 : Details downloaded OK!
HTTP request OK!
6 : Details downloaded OK!
HTTP request OK!
7 : Details downloaded OK!
HTTP request OK!
8 : Details downloaded OK!
HTTP request OK!
9 : Details downloaded OK!
HTTP request OK!
10 : Details downloaded OK!
HTTP request OK!
11 : Details downloaded OK!
HTTP request OK!
12 : Details downloaded OK!
HTTP request OK!
13 : Details downloaded OK!
HTTP request OK!
14 : Details downloaded OK!
HTTP request OK!
15 : Details downloaded OK!
HTTP request OK!
16 : Details downloaded OK!
HTTP request OK!
17 : Details downloaded OK!
HTTP request OK!
18 : Details downloaded OK!
HTTP request OK!
19 : Details downloaded OK!
HTTP request OK!
20 : Details downloaded OK!
HTTP request OK!
21 : Details downloaded OK!
HTTP request OK!
22 : Details downloaded OK!
HTTP request OK!
23 : Details downloaded OK!
HTTP request OK!
24 : Details downloaded OK!
HTTP request OK!
25 : Details downloaded OK!
HTTP request OK!
26 : Details downloaded OK!
HTTP request OK!
27 : Details downloaded OK!
HTTP request OK!
28 : Details downloaded OK!
HTTP request OK!
29 : Details downloaded OK!
HTTP request OK!
```

```
30 : Details downloaded OK!
HTTP request OK!
31 : Details downloaded OK!
HTTP request OK!
32 : Details downloaded OK!
HTTP request OK!
33 : Details downloaded OK!
HTTP request OK!
34 : Details downloaded OK!
HTTP request OK!
35 : Details downloaded OK!
HTTP request OK!
36 : Details downloaded OK!
HTTP request OK!
37 : Details downloaded OK!
HTTP request OK!
38 : Details downloaded OK!
HTTP request OK!
39 : Details downloaded OK!
HTTP request OK!
40 : Details downloaded OK!
HTTP request OK!
41 : Details downloaded OK!
HTTP request OK!
42 : Details downloaded OK!
HTTP request OK!
43 : Details downloaded OK!
HTTP request OK!
44 : Details downloaded OK!
HTTP request OK!
45 : Details downloaded OK!
HTTP request OK!
46 : Details downloaded OK!
HTTP request OK!
47 : Details downloaded OK!
HTTP request OK!
48 : Details downloaded OK!
HTTP request OK!
49 : Details downloaded OK!
HTTP request OK!
50 : Details downloaded OK!
HTTP request OK!
51 : Details downloaded OK!
HTTP request OK!
52 : Details downloaded OK!
HTTP request OK!
53 : Details downloaded OK!
HTTP request OK!
54 : Details downloaded OK!
HTTP request OK!
55 : Details downloaded OK!
HTTP request OK!
56 : Details downloaded OK!
HTTP request OK!
57 : Details downloaded OK!
HTTP request OK!
58 : Details downloaded OK!
HTTP request OK!
59 : Details downloaded OK!
HTTP request OK!
60 : Details downloaded OK!
```

```
HTTP request OK!
61 : Details downloaded OK!
HTTP request OK!
62 : Details downloaded OK!
HTTP request OK!
63 : Details downloaded OK!
HTTP request OK!
64 : Details downloaded OK!
HTTP request OK!
65 : Details downloaded OK!
HTTP request OK!
66 : Details downloaded OK!
HTTP request OK!
67 : Details downloaded OK!
HTTP request OK!
68 : Details downloaded OK!
HTTP request OK!
69 : Details downloaded OK!
HTTP request OK!
70 : Details downloaded OK!
HTTP request OK!
71 : Details downloaded OK!
HTTP request OK!
72 : Details downloaded OK!
HTTP request OK!
73 : Details downloaded OK!
HTTP request OK!
74 : Details downloaded OK!
HTTP request OK!
75 : Details downloaded OK!
HTTP request OK!
76 : Details downloaded OK!
HTTP request OK!
77 : Details downloaded OK!
HTTP request OK!
78 : Details downloaded OK!
HTTP request OK!
79 : Details downloaded OK!
HTTP request OK!
80 : Details downloaded OK!
HTTP request OK!
81 : Details downloaded OK!
HTTP request OK!
82 : Details downloaded OK!
HTTP request OK!
83 : Details downloaded OK!
HTTP request OK!
84 : Details downloaded OK!
HTTP request OK!
85 : Details downloaded OK!
HTTP request OK!
86 : Details downloaded OK!
HTTP request OK!
87 : Details downloaded OK!
HTTP request OK!
88 : Details downloaded OK!
HTTP request OK!
89 : Details downloaded OK!
HTTP request OK!
90 : Details downloaded OK!
HTTP request OK!
```

```
91 : Details downloaded OK!
HTTP request OK!
92 : Details downloaded OK!
HTTP request OK!
93 : Details downloaded OK!
HTTP request OK!
94 : Details downloaded OK!
HTTP request OK!
95 : Details downloaded OK!
HTTP request OK!
96 : Details downloaded OK!
HTTP request OK!
97 : Details downloaded OK!
HTTP request OK!
98 : Details downloaded OK!
HTTP request OK!
99 : Details downloaded OK!
HTTP request OK!
100 : Details downloaded OK!
HTTP request OK!
101 : Details downloaded OK!
HTTP request OK!
102 : Details downloaded OK!
HTTP request OK!
103 : Details downloaded OK!
HTTP request OK!
104 : Details downloaded OK!
HTTP request OK!
105 : Details downloaded OK!
HTTP request OK!
106 : Details downloaded OK!
HTTP request OK!
107 : Details downloaded OK!
HTTP request OK!
108 : Details downloaded OK!
HTTP request OK!
109 : Details downloaded OK!
HTTP request OK!
110 : Details downloaded OK!
HTTP request OK!
111 : Details downloaded OK!
HTTP request OK!
112 : Details downloaded OK!
HTTP request OK!
113 : Details downloaded OK!
HTTP request OK!
114 : Details downloaded OK!
HTTP request OK!
115 : Details downloaded OK!
HTTP request OK!
116 : Details downloaded OK!
HTTP request OK!
117 : Details downloaded OK!
HTTP request OK!
118 : Details downloaded OK!
HTTP request OK!
119 : Details downloaded OK!
HTTP request OK!
120 : Details downloaded OK!
HTTP request OK!
121 : Details downloaded OK!
```

```
HTTP request OK!
122 : Details downloaded OK!
HTTP request OK!
123 : Details downloaded OK!
HTTP request OK!
124 : Details downloaded OK!
HTTP request OK!
125 : Details downloaded OK!
HTTP request OK!
126 : Details downloaded OK!
HTTP request OK!
127 : Details downloaded OK!
HTTP request OK!
128 : Details downloaded OK!
HTTP request OK!
129 : Details downloaded OK!
HTTP request OK!
130 : Details downloaded OK!
HTTP request OK!
131 : Details downloaded OK!
HTTP request OK!
132 : Details downloaded OK!
HTTP request OK!
133 : Details downloaded OK!
HTTP request OK!
134 : Details downloaded OK!
HTTP request OK!
135 : Details downloaded OK!
HTTP request OK!
136 : Details downloaded OK!
HTTP request OK!
137 : Details downloaded OK!
HTTP request OK!
138 : Details downloaded OK!
HTTP request OK!
139 : Details downloaded OK!
HTTP request OK!
140 : Details downloaded OK!
HTTP request OK!
141 : Details downloaded OK!
HTTP request OK!
142 : Details downloaded OK!
HTTP request OK!
143 : Details downloaded OK!
HTTP request OK!
144 : Details downloaded OK!
HTTP request OK!
145 : Details downloaded OK!
HTTP request OK!
146 : Details downloaded OK!
HTTP request OK!
147 : Details downloaded OK!
HTTP request OK!
148 : Details downloaded OK!
HTTP request OK!
149 : Details downloaded OK!
HTTP request OK!
150 : Details downloaded OK!
HTTP request OK!
151 : Details downloaded OK!
HTTP request OK!
```

```
152 : Details downloaded OK!
HTTP request OK!
153 : Details downloaded OK!
HTTP request OK!
154 : Details downloaded OK!
HTTP request OK!
155 : Details downloaded OK!
HTTP request OK!
156 : Details downloaded OK!
HTTP request OK!
157 : Details downloaded OK!
HTTP request OK!
158 : Details downloaded OK!
HTTP request OK!
159 : Details downloaded OK!
HTTP request OK!
160 : Details downloaded OK!
HTTP request OK!
161 : Details downloaded OK!
HTTP request OK!
162 : Details downloaded OK!
HTTP request OK!
163 : Details downloaded OK!
HTTP request OK!
164 : Details downloaded OK!
HTTP request OK!
165 : Details downloaded OK!
HTTP request OK!
166 : Details downloaded OK!
HTTP request OK!
167 : Details downloaded OK!
HTTP request OK!
168 : Details downloaded OK!
HTTP request OK!
169 : Details downloaded OK!
HTTP request OK!
170 : Details downloaded OK!
HTTP request OK!
171 : Details downloaded OK!
HTTP request OK!
172 : Details downloaded OK!
HTTP request OK!
173 : Details downloaded OK!
HTTP request OK!
174 : Details downloaded OK!
HTTP request OK!
175 : Details downloaded OK!
HTTP request OK!
176 : Details downloaded OK!
HTTP request OK!
177 : Details downloaded OK!
HTTP request OK!
178 : Details downloaded OK!
HTTP request OK!
179 : Details downloaded OK!
HTTP request OK!
180 : Details downloaded OK!
HTTP request OK!
181 : Details downloaded OK!
HTTP request OK!
182 : Details downloaded OK!
```

```
HTTP request OK!
183 : Details downloaded OK!
HTTP request OK!
184 : Details downloaded OK!
HTTP request OK!
185 : Details downloaded OK!
HTTP request OK!
186 : Details downloaded OK!
HTTP request OK!
187 : Details downloaded OK!
HTTP request OK!
188 : Details downloaded OK!
HTTP request OK!
189 : Details downloaded OK!
HTTP request OK!
190 : Details downloaded OK!
HTTP request OK!
191 : Details downloaded OK!
HTTP request OK!
192 : Details downloaded OK!
HTTP request OK!
193 : Details downloaded OK!
HTTP request OK!
194 : Details downloaded OK!
HTTP request OK!
195 : Details downloaded OK!
HTTP request OK!
196 : Details downloaded OK!
HTTP request OK!
197 : Details downloaded OK!
HTTP request OK!
198 : Details downloaded OK!
HTTP request OK!
199 : Details downloaded OK!
```

In [16]:

```
print(details)
```

['Geschwister-Scholl-Platz 1, Munich, 80539, Germany', 'SE-171 77, Stockho
lm, Sweden', 'Arcisstraße 21, Munich, D-80333, Germany', '310 E. Market St
reet, Tiffin, Ohio, 44883, United States', 'Oude Markt 13, Leuven, 3000, B
elgium', '60 rue Mazarine, Paris, 75006, France', 'Droevendaalsesteeg 4, B
uilding nr. 104, Wageningen, 6702 PB, Netherlands', 'P.O. Box 19268, 1000
GG Amsterdam, Netherlands', 'PO Box 9500, Leiden, 2300, Netherlands', 'Bur
gemeester Oudlaan 50, Rotterdam, 3062 PA, Netherlands', 'Charitéplatz 1, 1
0117 Berlin, Germany', 'P.O Box 80125, TC Utrecht, 3508, Netherlands', 'Po
stbus 5, 2600 AA Delft, Netherlands', 'Geschwister-Scholl-Platz, Tubingen,
72074, Germany', 'Unter den Linden 6, D-10099 Berlin, Germany', 'PO Box 7
2, 9700 AB Groningen, Netherlands', 'Fahnenbergplatz, 79085 Freiburg, Germ
any', 'Nørregade 10 Postboks 2177, 1017 København K, Denmark', 'Route de S
aclay, Palaiseau, 91128, France', "21 Rue de l'Ecole-de-Medecine, Paris, 7
5006, France", 'P.O. Box 4, (Yliopistonkatu 3), 00014, Finland', 'Sint-Pie
tersnieuwstraat, B - 9000 Ghent, Belgium', 'Box 117, 221 00 Lund, Sweden',
'Nordre Ringgade 1 8000 Aarhus C, DK - Denmark, Denmark', 'Templergraben 5
5, 52056 Aachen, Germany', 'P.O. Box 256, SE-751 05, Uppsala, Sweden', 'Re
gina Pacis Weg 3, D-53012 Bonn, Germany', 'De Boelelaan 1105, Amsterdam, 1
081 HV, Netherlands', 'Kaiserswerther Str. 16-18, 14195 Berlin, Germany',
'Postbus 616, Maastricht, 6200 MD, Netherlands', 'Wilhelmsplatz, 37073 Göt
tingen, Germany', 'Mittelweg 177, 20148 Hamburg, Germany', '85 boulevard S
aint-Germain, cedex 06, Paris, Ile-de-France, 75006, France', 'Comeniuslaa
n 4, 6525 HP Nijmegen, Netherlands', 'Sanderring 2, Würzburg, 97070, Germa
ny', 'Straße des 17. Juni 135, D-10623 Berlin, Germany', 'Helmholtzstraße
16, Ulm, 89081, Germany', 'Schloss, 68131 Mannheim, Germany', 'Universität
sstraße 24, Cologne, 50931, Germany', 'D-01062 Dresden, Germany', 'Placa d
e la Merce 10-12, Barcelona, 08002, Spain', 'Postfach 10 01 31, D-33501 Bi
elefeld, Germany', "1 Place de l'Université, Louvain-la-Neuve, B-1348, Bel
gium", 'Universitätsring 1 1010, Vienna, Austria', 'Via Zamboni 33, Bologn
a, 40126, Italy', 'Prinsstraat 13, 2000 Antwerpen, Belgium', 'Piazza Marti
ri della Libertà,, Pisa, 56127, Italy', 'Espace Technologique, Bat. Discov
ery - RD 128 - 2e é, Saint-Aubin, 91190, France', 'Piazza dei Cavalieri,
7, Pisa, 56126, Italy', 'Placa Civica, Campus de la UAB, Barcelona, 08193,
Spain', 'SE-106 91, Stockholm, Sweden', 'Postbus 513, Eindhoven, 5600, Net
herlands', 'Anker Engelunds Vej 1, 2800 Kgs. Lyngby, Denmark', 'PO Box 10
0, SE-405 30, Gothenburg, Sweden', 'Schlossplatz 2, 48149 Munster, German
y', 'Schlossplatz 4, Erlangen, 91054, Germany', 'Gr. Via de les Corts Cata
lanes, 585, Barcelona, 08007, Spain', 'Piazzale Aldo Moro 5, Rome, 00185,
Italy', '19 place Marguerite Perey F-91120, Palaiseau, France', 'Universit
ätsstraße 2, Essen, 45141, Germany', 'Warandelaan 2, AB Tilburg, 5037, Net
herlands', 'PO Box 217, 7500 AE Enschede, Netherlands', 'Schloss Hohenheim
1, 70599 Stuttgart, Germany', 'Kaiserstraße 12, Karlsruhe, 76131, German
y', 'Avenue Franklin Roosevelt 50, 1050 Bruxelles, Belgium', "2, avenue de
l'Université, L-4365 Esch-sur-Alzette, Luxembourg", 'Universitätsring 1 10
10, Vienna, Austria', 'PO Box 18000, Aalto, 00076, Finland', 'Maskingränd
2, Göteborg, 412 58, Sweden', 'Universitätsplatz 3, A - 8010 Graz, Austri
a', 'Solbjerg Plads 3, DK-2000 Frederiksberg, Denmark', 'Fredrik Bajers Ve
j 5 P.O. Box 159, DK - 9100 Aalborg, Denmark', 'SE-100 44, Stockholm, Swed
en', 'Innrain 52, A-6020 Innsbruck, Austria', 'Campusvej 55 DK-5230, Odens
e M, Denmark', 'Universitätsstraße 10, Konstanz, 78464, Germany', 'Fürsten
graben 1, Jena, 07743, Germany', 'Grueneburgplatz 1/2nd floor, 60323 Frank
furt, Germany', 'Pleinlaan 2, 1050 Brussel, Belgium', 'Via 8 Febbraio 2, P
adova, 35122, Italy', 'Via Olgettina 58, Milano, 20123, Italy', 'Jacobs Un
iversity Bremen Campus Ring 1, 28759 Bremen, Germany', 'Ülikooli 18, Tart
u, 50090, Estonia', 'Am Neuen Palais 10, 14469 Potsdam, Germany', 'Christi
an-Albrechts-Platz 4, 24118 Kiel, Germany', "45 Rue d'Ulm, Cedex 05, Pari
s, 75230, France", '6-8 avenue Blaise-Pascal Cité Descartes, 77455 Champs-
sur-Marne, Marne-la-Vallée, France', 'Universitätsstraße 150, 44801 Bochu
m, Germany', 'Pamplona, Navarra, 31009, Spain', 'Domstraße 11, Greifswald,
17489, Germany', '163 rue Auguste Broussonnet, 34090 Montpellier, France',
'Karolinenplatz 5, Darmstadt, 64289, Germany', 'Innstraße 41, D-94032 Pass

au, Germany', 'Biegenstraße 10, D-35032 Marburg, Germany', 'Madrid, 28049,
Spain', 'Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austri
a', 'Pentti Kaiteran katu 1, 90570 Oulu, Finland', 'Saarstr. 21 D 55122, M
ainz, Germany', 'Via Calepina, 14, Trento, 38122, Italy', 'Kalevantie 4, T
ampere, 33100, Finland', '30 Archbishop Kyprianos Str, 3036 Lemesos, Cypru
s', 'Fakultetsgatan 1, Örebro SE 70182, Sweden', 'Universitätsstraße 30, 9
5440 Bayreuth, Germany', 'Place du 20-Août 7, Liege, 4000, Belgium', 'Kepl
erstraße 7, Stuttgart, 70174, Germany', 'SE-901 87, Umea, Sweden', 'Via Fe
sta del Perdono 7, Milano, 20122, Italy', 'SLU, P.O. Box 7070, SE-750 07,
Sweden', 'Via Festa del Perdono 7, Milano, 20122, Italy', 'Rethymnon, Cret
e, 74100, Greece', 'Jardin du Pharo, 58, bd Charles Livon, Marseille, 1328
4, France', "621 Central Avenue, Saint Martin d'Heres, 38400, France", 'La
Chantrerie 4 rue Alfred Kastler, BP 20722, Nantes, Cedex 3, 44307, Franc
e', '55 avenue de Paris, Versailles, 78035, France', 'Campus de Campolide,
1099-085 Lisboa, Portugal', 'Palma de Cima, 1649-023 Lisbon, Portugal', 'T
urku, 20014, Finland', 'Yliopistonkatu 34, Lappeenranta, 53850, Finland',
'Innrain 52, A-6020 Innsbruck, Austria', 'Martelarenlaan 42, 3500 Hasselt,
Belgium', 'Via Festa del Perdono 7, Milano, 20122, Italy', 'Piazza Umberto
I, Bari, 70121, Italy', 'C/ Isaac Peral, Madrid, 58 - 28040, Spain', 'Praç
a Gomes Teixeira, Porto, 4099-002, Portugal', 'Karlsplatz 13, 1040 Vienna,
Austria', '30 Panepistimiou Ave, 106 79 Athens, Greece', "Via dell'Artigli
ere 8, Verona, 37129, Italy", 'Piazza Guerrazzi, Benevento, 82100, Italy',
'Via Giovanni Paolo II, 132, Fisciano Salerno, 84084, Italy', 'Via Cracovi
a snc, Rome, 00133, Italy', 'Via Giovanni Amendola, 126/B, Bari, 70126, It
aly', 'Ülloi út 26, H - 1085 Budapest, Hungary', 'Linkoping, SE-581 83, Sw
eden', 'Lungarno Pacinotti 43, Pisa, 56126, Italy', 'Strada Nuova 65, Pavi
a, 27100, Italy', 'Corso Umberto I 40, Napoli, 80138, Italy', 'Via Univers
ità 4, Modena, 41121, Italy', 'Avda. Blasco Ibáñez 13, Valencia, 46010, Sp
ain', 'Piazza Universita, 1, Bozen-Bolzano, 39100, Italy', 'P.zza S.Marco
4, Firenze, Florence, 50121, Italy', 'Piazza del Mercato, 15, Brescia, 251
21, Italy', 'Via Giuseppe Verdi 8, Turin, 10124, Italy', 'Via Balbi, 5, Ge
nova, 16126, Italy', 'via Banchi di Sotto 55, Siena, 53100, Italy', '146 r
ue Léo Saignat, S 61292, Bordeaux, 33 076, France', 'PO Box 35, FI-40014,
Finland', 'Yliopistokatu 2 P.O. Box 111, FI-80101 Joensuu, Finland', '1 ru
e de la Noë, Nantes, Pays de la Loire, 44321, France', '41 Allées Jules Gu
esde, 31013 Toulouse, France', '42 Rue Scheffer, Paris, 75116, France', 'B
ibliothekstraße 1, Bremen, 28359, Germany', 'Ovocný trh 3-5, Prague 1, 116
36, Czech Republic', 'August-Schmidt-Straße 4, Dortmund, 44227, Germany',
'Ludwigstraße 23, 35390 Giessen, Germany', '9 rue Charles Fourier, Evry, 9
1011, France', 'Piazza Università, 2, Catania, 95131, Italy', 'via Ludovic
o Ariosto, 35, Ferrara, 44121, Italy', '27 Rue Saint Guillaume, 75337 Pari
s, France', 'Piazza Pugliatti, 1, Messina, 98122, Italy', 'Corso Duca degl
i Abruzzi 24, Torino, 10129, Italy', 'Piazza Università 21, Sassari, Ital
y', 'Piazzale Europa 1, Trieste, 34127, Italy', 'Via S.M. in Gradi n.4, Vi
terbo, 01100, Italy', 'Via Aurelio Saffi 2, Urbino, 61029, Italy', 'Alamed
a da Universidade Cidade Universitária, 1649 - 004 Lisboa, Portugal', 'Avd
a. de Séneca, 2, Ciudad Universitaria, Madrid, 28040, Spain', "Palazzo Cam
poneschi, 2 Piazza Santa Margherita, L'Aquila, Abruzzo, 67100, Italy", '43
Blvd du 11 Novembre 1918, 69622 Villeurbanne cedex, Lyon, France', 'ul. Go
lebia 24, Krakow, 31-007, Poland', 'Plateau de Moulon, 3 rue Joliot-Curie,
F-91192 Gif-sur-Yvette, France', 'Rechbauerstraße 12, 8010 Graz, Austria',
'Grand Château 28, avenue Valrose, BP 2135, Nice cedex 2, 06103, France',
'Welfengarten 1, Hannover, D-30167, Germany', 'Via Ravasi, 2, Varese, Lomb
ardia, 21100, Italy', '1 Panepistimiou Avenue, Aglantzia, Nicosia, Cypru
s', '4 rue Blaise Pascal CS 90032, F-67081 Strasbourg cedex, France', '8 K
rízkovskeho, Olomouc, 771 47, Czech Republic', '34 cours Léopold, CS 2523
3, Nancy, 54052, France', '1, quai de Tourville, BP 13522, Nantes, 44035,
France', '20 avenue Albert Einstein, 69621 Villeurbanne, France', 'Msida M
SD 2080, Malta', '12 place du Panthéon, 75231 Paris, France', 'P.O. Box 11
86, 45110 Ioannina, Greece', '42, rue Paul Duez, Lille, 59000, France', 'H
eroon Polytechniou 9, 15780 Zografou, Greece', 'Cra. de Valldemossa, Palm

```
a, Baleares, 07122, Spain', 'Raina bulvaris 19, Riga, LV 1586, Latvia', 'A
venida de las Universidades,, 24, Bizkaia, 48007, Spain', 'Via Palladio 8,
Udine, 33100, Italy', 'Piazza Tancredi, Lecce (LE), 73100, Italy', 'Žerotí
novo námestí 9 Rektorát, Brno-mesto, Brno, Czech Republic', 'Argonafton &
Filellinon, 38221 Volos, Greece', '76, Patission Str., GR10434 Athens, Gre
ece', 'Altenberger Straße 69, A-4040 Linz, Austria', 'Universitätsplatz 3,
A - 8010 Graz, Austria', 'Am Schwarzenberg, Hamburg, 21073, Germany', 'Got
tlieb-Daimler-Strasse, Kaiserslautern, 67663, Germany', 'Aveiro, 3810-193,
Portugal', 'Convento de Sto. Antonio, 6201-001 Covilha, Portugal', 'Paço d
as Escolas, Coimbra, 3004-531, Portugal']
```

In [17]:

```
# Append addresses as a new column to the dataset
df_eu_univs['Address'] = details
df_eu_univs.head()
```

Out[17]:

	Code	Rank	University	URL	Address
0	DE	8	LMU Munich	https://www.timeshighereducation.com/world-uni...	Geschwister-Scholl-Platz 1, Munich, 80539, Ger...
1	SE	10	Karolinska Institute	https://www.timeshighereducation.com/world-uni...	SE-171 77, Stockholm, Sweden
2	DE	11	Technical University of Munich	https://www.timeshighereducation.com/world-uni...	Arcisstraße 21, Munich, D-80333, Germany
3	DE	12	Heidelberg University	https://www.timeshighereducation.com/world-uni...	310 E. Market Street, Tiffin, Ohio, 44883, Uni...
4	BE	14	KU Leuven	https://www.timeshighereducation.com/world-uni...	Oude Markt 13, Leuven, 3000, Belgium

In [18]:

```
df_eu_univs.tail()
```

Out[18]:

	Code	Rank	University	URL	Address
195	DE	279	Hamburg University of Technology	https://www.timeshighereducation.com/world-uni...	Am Schwarzenberg, Hamburg, 21073, Germany
196	DE	279	University of Kaiserslautern	https://www.timeshighereducation.com/world-uni...	Gottlieb-Daimler-Strasse, Kaiserslautern, 6766...
197	PT	279	University of Aveiro	https://www.timeshighereducation.com/world-uni...	Aveiro, 3810-193, Portugal
198	PT	279	University of Beira Interior	https://www.timeshighereducation.com/world-uni...	Convento de Sto. Antonio, 6201-001 Covilha, Po...
199	PT	279	University of Coimbra	https://www.timeshighereducation.com/world-uni...	Paço das Escolas, Coimbra, 3004-531, Portugal

In [19]:

```
df_eu_univs.to_csv('top_eu_univs.csv')
#df_eu_univs = pd.read_csv('top_eu_univs.csv')
```

1.4. Prepare a dataset of EU major cities

In [21]:

```
# Create list of cities in EU (parse Wikipedia page)
url_cities = "https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_pop
ulation_within_city_limits"
html_cities = get_html(rq.get(url_cities))

if html_cities:
    cities = BeautifulSoup(html_cities).find('table')
    df_cities = pd.DataFrame()
    for row in cities.find_all('tr')[1::]:
        cells = [c.text.strip() for c in row.find_all('td')[1:4]]
        df_cities = df_cities.append([cells], ignore_index=True)
    df_cities.columns = ['City', 'Country', 'Population']
    df_cities.Population = pd.to_numeric(df_cities.Population.str.replace(',', ''))
    print(df_cities.dtypes)
    print(df_cities.shape)
```

```
HTTP request OK!
City            object
Country         object
Population       int64
dtype: object
(93, 3)
```

In [22]:

```
df_cities.head(10)
```

Out[22]:

	City	Country	Population
0	Berlin	Germany	3669495
1	Madrid	Spain	3348536
2	Rome	Italy	2856133
3	Bucharest	Romania	2155240
4	Paris	France	2140526
5	Vienna	Austria	1921153
6	Hamburg	Germany	1899160
7	Warsaw	Poland	1793579
8	Budapest	Hungary	1752286
9	Barcelona	Spain	1620343

In [23]:

```
df_cities.tail(10)
```

Out[23]:

	City	Country	Population
83	Cluj-Napoca	Romania	324960
84	Bari	Italy	320862
85	Constanța	Romania	317832
86	Münster	Germany	314319
87	Karlsruhe	Germany	313092
88	Catania	Italy	311584
89	Mannheim	Germany	309370
90	Nantes	France	306694
91	Craiova	Romania	305386
92	Galați	Romania	304050

2.1. Geocode our datasets: cities

In [24]:

```
# Install and import libraries for geocoding
!conda install -c conda-forge geopy --yes
from geopy.geocoders import Here
from geopy import distance
from geopy.location import Location
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - geopy


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    geographiclib-1.52         |     pyhd8ed1ab_0          35 KB  conda-fo
rge
    geopy-2.1.0                |     pyhd3deb0d_0          64 KB  conda-fo
rge
    ------------------------------------------------------------
                                           Total:          99 KB

The following NEW packages will be INSTALLED:

  geographiclib      conda-forge/noarch::geographiclib-1.52-pyhd8ed1ab_0
  geopy              conda-forge/noarch::geopy-2.1.0-pyhd3deb0d_0



Downloading and Extracting Packages
geographiclib-1.52   | 35 KB     | ##################################### |
100%
geopy-2.1.0          | 64 KB     | ##################################### |
100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
```

In [25]:

```
# Initialize a geocoder
geocoder = Here(apikey=my_secrets['HERE']['API_KEY'])
```

In [26]:

```
# Obtain pair of geo coords (Lat, Lon) for a chosen object defined by a name
# Return (None, None) if geocoder fails
def get_location(location_name: str):
    location = geocoder.geocode(location_name)
    if location is None:
        print('Cannot geocode specified object:', location_name)
        return (None, None)
    else:
        print(f"{location.address} = {location.latitude}, {location.longitude}")
        return (location.latitude, location.longitude)
```

In [27]:

```
# Obtain geo coords (Lat, Lon) for a chosen dataframe, append two columns 'Lat' and 'Lo
n'
def geocode_dataframe(df_to_modify, address_fields):
    geo_lats, geo_lons = zip(*[
        get_location(','.join([df_to_modify.loc[i, f] for f in address_fields])) for i
in df_to_modify.index
    ])
    df_to_modify['Lat'] = geo_lats[:]
    df_to_modify['Lon'] = geo_lons[:]
```

In [28]:

```
# Geocode all EU cities
geocode_dataframe(df_cities, address_fields = ['City', 'Country'])
```

Berlin, Deutschland, Berlin, Berlin 10117, DEU = 52.51605, 13.37691
Madrid, Comunidad de Madrid, España, Madrid, Comunidad de Madrid 28014, ES
P = 40.41956, -3.69196
Roma, Lazio, Italia, Roma, Lazio 00185, ITA = 41.90323, 12.49566
București, România, București 030171, ROU = 44.4343, 26.10298
Paris, Île-de-France, France, Paris, Île-de-France 75001, FRA = 48.85718,
2.34141
Wien, Österreich, Wien, Wien 1010, AUT = 48.20263, 16.36843
Hamburg, Deutschland, Hamburg, Hamburg 20354, DEU = 53.55562, 9.98746
Warszawa, Woj. Mazowieckie, Polska, Warszawa, Woj. Mazowieckie 00-941, POL
= 52.2356, 21.01038
Budapest, Magyarország, Budapest, Budapest 1061, HUN = 47.49973, 19.05508
Barcelona, Catalunya, Espanya, Barcelona, Catalunya 08007, ESP = 41.38804,
2.17001
München, Bayern, Deutschland, München, Bayern 80331, DEU = 48.13642, 11.57
755
Milano, Lombardia, Italia, Milano, Lombardia 20121, ITA = 45.46796, 9.1817
8
Praha, Hlavní město Praha, Česká Republika, Praha, Hlavní město Praha 120
00, CZE = 50.07913, 14.43303
София, България, София 1000, BGR = 42.69719, 23.32433
Köln, Nordrhein-Westfalen, Deutschland, Köln, Nordrhein-Westfalen 50667, D
EU = 50.94168, 6.95517
Stockholm, Stockholms län, Sverige, Stockholm, Stockholms län 111 53, SWE
= 59.33258, 18.06683
Napoli, Campania, Italia, Napoli, Campania 80133, ITA = 40.84016, 14.25222
Torino, Piemonte, Italia, Torino, Piemonte 10123, ITA = 45.06236, 7.67994
Amsterdam, Noord-Holland, Nederland, Amsterdam, Noord-Holland 1011 MG, NLD
= 52.36994, 4.90788
Marseille, Provence-Alpes-Côte d'Azur, France, Marseille, Provence-Alpes-C
ôte d'Azur 13001, FRA = 43.29338, 5.37132
Zagreb, Hrvatska, Zagreb 10000, HRV = 45.80724, 15.96757
København, Hovedstaden, Danmark, København, Hovedstaden 1620, DNK = 55.675
67, 12.56756
València, Comunitat Valenciana, Espanya, València, Comunitat Valenciana 46
002, ESP = 39.46895, -0.37686
Kraków, Woj. Małopolskie, Polska, Kraków, Woj. Małopolskie 31-109, POL = 5
0.06045, 19.93243
Frankfurt am Main, Hessen, Deutschland, Frankfurt am Main, Hessen 60311, D
EU = 50.11208, 8.68342
Sevilla, Andalucía, España, Sevilla, Andalucía 41001, ESP = 37.38788, -6.0
0197
Łódź, Woj. Łódzkie, Polska, Łódź, Woj. Łódzkie 90-136, POL = 51.77234, 19.
47502
Zaragoza, Aragón, España, Zaragoza, Aragón 50001, ESP = 41.65184, -0.88114
Αθήνα, Αττική, Ελληνικη Δημοκρατια, Αθήνα, Αττική 106 71, GRC = 37.97614,
23.7364
Palermo, Sicilia, Italia, Palermo, Sicilia 90133, ITA = 38.12207, 13.36112
Helsinki, Etelä-Suomi, Suomi, Helsinki, Etelä-Suomi 00100, FIN = 60.17116,
24.93266
Rotterdam, Zuid-Holland, Nederland, Rotterdam, Zuid-Holland 3011, NLD = 5
1.91439, 4.48717
Wrocław, Woj. Dolnośląskie, Polska, Wrocław, Woj. Dolnośląskie 50-075, POL
= 51.10825, 17.02692
Stuttgart, Baden-Württemberg, Deutschland, Stuttgart, Baden-Württemberg 70
178, DEU = 48.76779, 9.17203
Rīga, Latvija, Rīga 1050, LVA = 56.94599, 24.11487
Düsseldorf, Nordrhein-Westfalen, Deutschland, Düsseldorf, Nordrhein-Westfa
len 40217, DEU = 51.21564, 6.77662
Vilnius, Vilniaus Apskritis, Lietuva, Vilnius, Vilniaus Apskritis 01108, L
TU = 54.69063, 25.26981

Leipzig, Sachsen, Deutschland, Leipzig, Sachsen 04103, DEU = 51.3452, 12.38594
Dortmund, Nordrhein-Westfalen, Deutschland, Dortmund, Nordrhein-Westfalen 44137, DEU = 51.51661, 7.4583
Essen, Nordrhein-Westfalen, Deutschland, Essen, Nordrhein-Westfalen 45127, DEU = 51.45183, 7.01109
Göteborg, Västra Götalands län, Sverige, Göteborg, Västra Götalands län 411 38, SWE = 57.70068, 11.96823
Genova, Liguria, Italia, Genova, Liguria 16122, ITA = 44.41048, 8.93917
Málaga, Andalucía, España, Málaga, Andalucía 29015, ESP = 36.71847, -4.41965
Bremen, Deutschland, Bremen, Bremen 28195, DEU = 53.07537, 8.80454
Dresden, Sachsen, Deutschland, Dresden, Sachsen 01067, DEU = 51.05364, 13.74082
Dublin, Ireland, Dublin D01, IRL = 53.34807, -6.24827
Den Haag, Zuid-Holland, Nederland, Den Haag, Zuid-Holland 2514, NLD = 52.08409, 4.31732
Hannover, Niedersachsen, Deutschland, Hannover, Niedersachsen 30159, DEU = 52.37228, 9.73816
Poznań, Woj. Wielkopolskie, Polska, Poznań, Woj. Wielkopolskie 61-758, POL = 52.40947, 16.93828
Antwerpen, Vlaanderen, België, Antwerpen, Vlaanderen 2000, BEL = 51.22213, 4.39769
Nürnberg, Bayern, Deutschland, Nürnberg, Bayern 90403, DEU = 49.45435, 11.0735
Lyon, Auvergne-Rhône-Alpes, France, Lyon, Auvergne-Rhône-Alpes 69002, FRA = 45.75917, 4.82966
Lisboa, Portugal, Lisboa 1050-115, PRT = 38.72639, -9.14949
Duisburg, Nordrhein-Westfalen, Deutschland, Duisburg, Nordrhein-Westfalen 47051, DEU = 51.43148, 6.76356
Toulouse, Occitanie, France, Toulouse, Occitanie 31000, FRA = 43.60579, 1.44864
Gdańsk, Woj. Pomorskie, Polska, Gdańsk, Woj. Pomorskie 80-846, POL = 54.35311, 18.65106
Murcia, Región de Murcia, España, Murcia, Región de Murcia 30004, ESP = 37.98309, -1.13139
Tallinn, Eesti, Tallinn 10148, EST = 59.43642, 24.75258
Bratislava, Bratislavský kraj, Slovenská Republika, Bratislava, Bratislavský kraj 811 06, SVK = 48.14924, 17.10699
Palma, Illes Balears, Espanya, Palma, Illes Balears 07012, ESP = 39.57149, 2.64694
Szczecin, Woj. Zachodniopomorskie, Polska, Szczecin, Woj. Zachodniopomorskie 70-562, POL = 53.42521, 14.55549
Bologna, Emilia Romagna, Italia, Bologna, Emilia Romagna 40121, ITA = 44.50485, 11.34507
Brno, Jihomoravský kraj, Česká Republika, Brno, Jihomoravský kraj 602 00, CZE = 49.19728, 16.60368
Iași, România, Iași, ROU = 47.1594, 27.58733
Firenze, Toscana, Italia, Firenze, Toscana 50129, ITA = 43.78238, 11.25502
Las Palmas de Gran Canaria, Islas Canarias, España, Las Palmas de Gran Canaria, Islas Canarias 35010, ESP = 28.13026, -15.43973
Bochum, Nordrhein-Westfalen, Deutschland, Bochum, Nordrhein-Westfalen 44787, DEU = 51.488, 7.21399
Utrecht, Nederland, Utrecht, Utrecht 3511, NLD = 52.08979, 5.11415
Wuppertal, Nordrhein-Westfalen, Deutschland, Wuppertal, Nordrhein-Westfalen 42275, DEU = 51.27165, 7.19678
Aarhus, Midtjylland, Danmark, Aarhus, Midtjylland 8000, DNK = 56.15302, 10.20487
Bydgoszcz, Woj. Kujawsko-Pomorskie, Polska, Bydgoszcz, Woj. Kujawsko-Pomorskie 85-023, POL = 53.11931, 18.0081
Пловдив, България, Пловдив 4000, BGR = 42.13586, 24.74906

```
Bilbao, País Vasco, España, Bilbao, País Vasco 48014, ESP = 43.2689, -2.94
53
Malmö, Skåne län, Sverige, Malmö, Skåne län 211 43, SWE = 55.5967, 13.0011
Nice, Provence-Alpes-Côte d'Azur, France, Nice, Provence-Alpes-Côte d'Azur
06300, FRA = 43.70029, 7.27766
Lublin, Woj. Lubelskie, Polska, Lublin, Woj. Lubelskie 20-115, POL = 51.24
789, 22.56598
Варна, България, Варна 9002, BGR = 43.20631, 27.92524
Bielefeld, Nordrhein-Westfalen, Deutschland, Bielefeld, Nordrhein-Westfale
n 33604, DEU = 52.01548, 8.53232
Alicante, Comunidad Valenciana, España, Alicante, Comunidad Valenciana 030
01, ESP = 38.3441, -0.48043
Timișoara, România, Timișoara 300002, ROU = 45.75346, 21.22334
Bonn, Nordrhein-Westfalen, Deutschland, Bonn, Nordrhein-Westfalen 53113, D
EU = 50.73243, 7.10187
Córdoba, Andalucía, España, Córdoba, Andalucía 14009, ESP = 37.87064, -4.7
7862
Θεσσαλονίκη, Κεντρική Μακεδονία, Ελληνικη Δημοκρατια, Θεσσαλονίκη, Κεντρικ
ή Μακεδονία 546 30, GRC = 40.63957, 22.9371
Cluj-Napoca, România, Cluj-Napoca 400002, ROU = 46.7687, 23.58503
Bari, Puglia, Italia, Bari, Puglia 70122, ITA = 41.12588, 16.86666
Constanța, România, Constanța, ROU = 44.17827, 28.65116
Münster, Nordrhein-Westfalen, Deutschland, Münster, Nordrhein-Westfalen 48
143, DEU = 51.96302, 7.61782
Karlsruhe, Baden-Württemberg, Deutschland, Karlsruhe, Baden-Württemberg 76
131, DEU = 49.01094, 8.40846
Catania, Sicilia, Italia, Catania, Sicilia 95123, ITA = 37.51136, 15.06752
Mannheim, Baden-Württemberg, Deutschland, Mannheim, Baden-Württemberg 6816
1, DEU = 49.48651, 8.46679
Nantes, Pays de la Loire, France, Nantes, Pays de la Loire 44000, FRA = 4
7.21812, -1.55306
Craiova, România, Craiova, ROU = 44.3202, 23.79895
Galați, România, Galați, ROU = 45.43369, 28.05476
```

In [29]:

```
df_cities.head(10)
```

Out[29]:

	City	Country	Population	Lat	Lon
0	Berlin	Germany	3669495	52.51605	13.37691
1	Madrid	Spain	3348536	40.41956	-3.69196
2	Rome	Italy	2856133	41.90323	12.49566
3	Bucharest	Romania	2155240	44.43430	26.10298
4	Paris	France	2140526	48.85718	2.34141
5	Vienna	Austria	1921153	48.20263	16.36843
6	Hamburg	Germany	1899160	53.55562	9.98746
7	Warsaw	Poland	1793579	52.23560	21.01038
8	Budapest	Hungary	1752286	47.49973	19.05508
9	Barcelona	Spain	1620343	41.38804	2.17001

In [30]:

```
df_cities.tail(10)
```

Out[30]:

	City	Country	Population	Lat	Lon
83	Cluj-Napoca	Romania	324960	46.76870	23.58503
84	Bari	Italy	320862	41.12588	16.86666
85	Constanța	Romania	317832	44.17827	28.65116
86	Münster	Germany	314319	51.96302	7.61782
87	Karlsruhe	Germany	313092	49.01094	8.40846
88	Catania	Italy	311584	37.51136	15.06752
89	Mannheim	Germany	309370	49.48651	8.46679
90	Nantes	France	306694	47.21812	-1.55306
91	Craiova	Romania	305386	44.32020	23.79895
92	Galați	Romania	304050	45.43369	28.05476

In [31]:

```
df_cities.to_csv('top_eu_cities.csv')
```

2.2. Geocode our datasets: universities

In [32]:

```
# Geocode all EU universities
df_eu_univs = pd.read_csv('top_eu_univs.csv')
geocode_dataframe(df_eu_univs, address_fields = ['Address'])
```

Geschwister-Scholl-Platz 1, 80539 München, Deutschland, München, Bayern 80
539, DEU = 48.1505, 11.5803
121 77, Stockholm, Stockholms län, Sverige, Stockholm, Stockholms län 121
77, SWE = 59.29349, 18.08222
Arcisstraße 21, 80333 München, Deutschland, München, Bayern 80333, DEU = 4
8.14885, 11.568
310 E Market St, Tiffin, OH 44883, United States, Tiffin, OH 44883, USA =
41.11633, -83.16853
Oude Markt 13, 3000 Leuven, België, Leuven, Vlaanderen 3000, BEL = 50.8779
4, 4.70032
60 Rue Mazarine, 75006 Paris, France, Paris, Île-de-France 75006, FRA = 4
8.85468, 2.3376
Droevendaalsesteeg 4, 6708 PB Wageningen, Nederland, Wageningen, Gelderlan
d 6708 PB, NLD = 51.98633, 5.66794
Cannot geocode specified object: P.O. Box 19268, 1000 GG Amsterdam, Nether
lands
Cannot geocode specified object: PO Box 9500, Leiden, 2300, Netherlands
Burgemeester Oudlaan 50, 3062 PA Rotterdam, Nederland, Rotterdam, Zuid-Hol
land 3062 PA, NLD = 51.91906, 4.52516
Charitéplatz 1, 10117 Berlin, Deutschland, Berlin, Berlin 10117, DEU = 52.
52364, 13.37821
Cannot geocode specified object: P.O Box 80125, TC Utrecht, 3508, Netherla
nds
Delft, Zuid-Holland, Nederland, Delft, Zuid-Holland 2611 RV, NLD = 52.0087
8, 4.36535
Geschwister-Scholl-Platz, 72074 Tübingen, Deutschland, Tübingen, Baden-Wür
ttemberg 72074, DEU = 48.52435, 9.05997
Unter den Linden 6, 10117 Berlin, Deutschland, Berlin, Berlin 10117, DEU =
52.51762, 13.39376
Groningen, Nederland, Groningen, Groningen 9711, NLD = 53.21687, 6.57394
Fahnenbergplatz, 79098 Freiburg im Breisgau, Deutschland, Freiburg im Brei
sgau, Baden-Württemberg 79098, DEU = 47.9986, 7.84848
Nørregade 10, 1165 København K, Danmark, København K, Hovedstaden 1165, DN
K = 55.68012, 12.57151
Route de Saclay, 91120 Palaiseau, France, Palaiseau, Île-de-France 91120,
FRA = 48.71895, 2.21719
21 Rue de l'École de Médecine, 75006 Paris, France, Paris, Île-de-France 7
5006, FRA = 48.85123, 2.34038
Box, Sipoo, Etelä-Suomi, Suomi, Sipoo, Etelä-Suomi 01190, FIN = 60.30898,
25.39256
Sint-Pietersnieuwstraat, 9000 Gent, België, Gent, Vlaanderen 9000, BEL = 5
1.04475, 3.72653
Lund, Västerbottens län, Sverige, Lund, Västerbottens län 931 97, SWE = 6
4.74951, 20.89238
Nordre Ringgade 1, 8000 Aarhus C, Danmark, Aarhus C, Midtjylland 8000, DNK
= 56.17104, 10.19937
Templergraben 55, 52062 Aachen, Deutschland, Aachen, Nordrhein-Westfalen 5
2062, DEU = 50.77764, 6.07794
Uppsala, Västerås, Västmanlands län, Sverige, Västerås, Västmanlands län 7
25 95, SWE = 59.65727, 16.69322
Regina-Pacis-Weg 3, 53113 Bonn, Deutschland, Bonn, Nordrhein-Westfalen 531
13, DEU = 50.73404, 7.10384
De Boelelaan 1105, 1081 HV Amsterdam, Nederland, Amsterdam, Noord-Holland
1081 HV, NLD = 52.33459, 4.86648
Kaiserswerther Straße 16, 14195 Berlin, Deutschland, Berlin, Berlin 14195,
DEU = 52.44797, 13.286
Maastricht, Limburg, Nederland, Maastricht, Limburg 6211 LE, NLD = 50.8498
2, 5.68829
Wilhelmsplatz, 37073 Göttingen, Deutschland, Göttingen, Niedersachsen 3707
3, DEU = 51.53381, 9.93857
Mittelweg 177, 20148 Hamburg, Deutschland, Hamburg, Hamburg 20148, DEU = 5

3.56399, 9.99506
85 Boulevard Saint-Germain, 75006 Paris, France, Paris, Île-de-France 7500
6, FRA = 48.85191, 2.3401
Comeniuslaan 4, 6525 HP Nijmegen, Nederland, Nijmegen, Gelderland 6525 HP,
NLD = 51.81963, 5.85692
Sanderring 2, 97070 Würzburg, Deutschland, Würzburg, Bayern 97070, DEU = 4
9.78818, 9.93524
Straße des 17. Juni 135, 10623 Berlin, Deutschland, Berlin, Berlin 10623,
DEU = 52.51231, 13.32698
Helmholtzstraße 16, 89081 Ulm, Deutschland, Ulm, Baden-Württemberg 89081,
DEU = 48.42521, 9.96286
Schlossergasse, 68305 Mannheim, Deutschland, Mannheim, Baden-Württemberg 6
8305, DEU = 49.51322, 8.48993
Universitätsstraße 24, 50931 Köln, Deutschland, Köln, Nordrhein-Westfalen
50931, DEU = 50.93029, 6.92742
Dresden, Sachsen, Deutschland, Dresden, Sachsen 01067, DEU = 51.05364, 13.
74082
Plaça de la Mercè, 4, 08002 Barcelona (Barcelona), Espanya, Barcelona, Cat
alunya 08002, ESP = 41.37898, 2.17924
Cannot geocode specified object: Postfach 10 01 31, D-33501 Bielefeld, Ger
many
Place de l'Université 1, 1348 Ottignies-Louvain-la-Neuve, Belgique, Ottign
ies-Louvain-la-Neuve, Wallonie 1348, BEL = 50.6698128, 4.6155308
Universitätsring 1, 1010 Wien, Österreich, Wien, Wien 1010, AUT = 48.2130
1, 16.36086
Via Luigi Zamboni, 33, 40126 Bologna BO, Italia, Bologna, Emilia Romagna 4
0126, ITA = 44.49691, 11.35241
Prinsstraat 13, 2000 Antwerpen, België, Antwerpen, Vlaanderen 2000, BEL =
51.2221428, 4.4097577
Piazza Martiri della Libertà, 56127 Pisa PI, Italia, Pisa, Toscana 56127,
ITA = 43.7204, 10.40404
Saint-Aubin, Île-de-France, France, Saint-Aubin, Île-de-France 91190, FRA
= 48.71482, 2.14084
Piazza dei Cavalieri, 7, 56126 Pisa PI, Italia, Pisa, Toscana 56126, ITA =
43.71964, 10.40029
Plaça Cívica, 08193 Cerdanyola del Vallès (Barcelona), Espanya, Cerdanyola
del Vallès, Catalunya 08193, ESP = 41.50243, 2.10474
196 91, Kungsängen, Stockholms län, Sverige, Kungsängen, Stockholms län 19
6 91, SWE = 59.47845, 17.77873
Eindhoven, Noord-Brabant, Nederland, Eindhoven, Noord-Brabant 5611 CB, NLD
= 51.43598, 5.48533
Anker Engelunds Vej 1, 2800 Kongens Lyngby, Danmark, Kongens Lyngby, Hoved
staden 2800, DNK = 55.7857163, 12.5225252
425 30, Göteborg, Västra Götalands län, Sverige, Göteborg, Västra Götaland
s län 425 30, SWE = 57.79101, 11.99663
Schlossplatz 2, 48149 Münster, Deutschland, Münster, Nordrhein-Westfalen 4
8149, DEU = 51.96361, 7.61314
Schloßplatz 4, 91054 Erlangen, Deutschland, Erlangen, Bayern 91054, DEU =
49.59788, 11.00453
Gran Via de les Corts Catalanes, 585, 08007 Barcelona (Barcelona), Espany
a, Barcelona, Catalunya 08007, ESP = 41.38654, 2.16401
Piazzale Aldo Moro, 3, 00185 Roma RM, Italia, Roma, Lazio 00185, ITA = 41.
9015, 12.51243
1 Place Marguerite Perey, 91120 Palaiseau, France, Palaiseau, Île-de-Franc
e 91120, FRA = 48.71404, 2.20095
Universitätsstraße 2, 45141 Essen, Deutschland, Essen, Nordrhein-Westfalen
45141, DEU = 51.46313, 7.00332
Warandelaan 2, 5037 AB Tilburg, Nederland, Tilburg, Noord-Brabant 5037 AB,
NLD = 51.56327, 5.04171
Enschede, Overijssel, Nederland, Enschede, Overijssel 7514, NLD = 52.2236
1, 6.89551

Schloß Hohenheim 1, 70599 Stuttgart, Deutschland, Stuttgart, Baden-Württem
berg 70599, DEU = 48.71188, 9.21407
Kaiserstraße 12, 76131 Karlsruhe, Deutschland, Karlsruhe, Baden-Württember
g 76131, DEU = 49.0095, 8.4114
Avenue Franklin Roosevelt 50, 1050 Bruxelles, Belgique, Bruxelles, Bruxell
es 1050, BEL = 50.81172, 4.38108
2 Avenue de l'Université, L-4365 Esch-sur-Alzette, Luxembourg, Esch-sur-Al
zette, Luxembourg 4365, LUX = 49.50441, 5.94886
Universitätsring 1, 1010 Wien, Österreich, Wien, Wien 1010, AUT = 48.2130
1, 16.36086
Box, Sipoo, Etelä-Suomi, Suomi, Sipoo, Etelä-Suomi 01190, FIN = 60.30898,
25.39256
Maskingränd 2, SE-412 58 Göteborg, Sverige, Göteborg, Västra Götalands län
412 58, SWE = 57.6882, 11.97859
Universitätsplatz 3, 8010 Graz, Österreich, Graz, Steiermark 8010, AUT = 4
7.0776, 15.44954
Solbjerg Plads 3, 2000 Frederiksberg, Danmark, Frederiksberg, Hovedstaden
2000, DNK = 55.68155, 12.53045
Fredrik Bajers Vej 5, 9220 Aalborg Øst, Danmark, Aalborg Øst, Nordjylland
9220, DNK = 57.01502, 9.98686
120 44, Stockholm, Stockholms län, Sverige, Stockholm, Stockholms län 120
44, SWE = 59.29266, 18.02925
Innrain 52, 6020 Innsbruck, Österreich, Innsbruck, Tirol 6020, AUT = 47.26
292, 11.38442
Campusvej 55, 5230 Odense M, Danmark, Odense M, Syddanmark 5230, DNK = 55.
36832, 10.42772
Universitätsstraße 10, 78464 Konstanz, Deutschland, Konstanz, Baden-Württe
mberg 78464, DEU = 47.68953, 9.18823
Fürstengraben 1, 07743 Jena, Deutschland, Jena, Thüringen 07743, DEU = 50.
92977, 11.58959
Grüneburgweg 1, 60322 Frankfurt am Main, Deutschland, Frankfurt am Main, H
essen 60322, DEU = 50.12073, 8.67549
Pleinlaan 2, 1050 Elsene, België, Elsene, Brussel 1050, BEL = 50.82301, 4.
39262
Via 8 Febbraio 1848, 2, 35121 Padova PD, Italia, Padova, Veneto 35121, ITA
= 45.40667, 11.87691
Via Olgettina, 58, 20132 Milano MI, Italia, Milano, Lombardia 20132, ITA =
45.5069, 9.26745
Campus Ring 1, 28759 Bremen, Deutschland, Bremen, Bremen 28759, DEU = 53.1
6843, 8.64876
Ülikooli 18, Tartu, 50090 Tartu Maakond, Eesti, Tartu 50090, EST = 58.3810
7, 26.71993
Am Neuen Palais 10, 14469 Potsdam, Deutschland, Potsdam, Brandenburg 1446
9, DEU = 52.40064, 13.01365
Christian-Albrechts-Platz 4, 24118 Kiel, Deutschland, Kiel, Schleswig-Hols
tein 24118, DEU = 54.33881, 10.12262
45 Rue d'Ulm, 75005 Paris, France, Paris, Île-de-France 75005, FRA = 48.84
229, 2.34429
6 Avenue Blaise Pascal, 77420 Champs-sur-Marne, France, Champs-sur-Marne,
Île-de-France 77420, FRA = 48.84058, 2.58697
Universitätsstraße 150, 44801 Bochum, Deutschland, Bochum, Nordrhein-Westf
alen 44801, DEU = 51.44589, 7.26046
Avenida de Navarra, 31009 Pamplona (Navarra), España, Pamplona, Comunidad
Foral de Navarra 31009, ESP = 42.7992, -1.64776
Domstraße 11, 17489 Greifswald, Deutschland, Greifswald, Mecklenburg-Vorpo
mmern 17489, DEU = 54.09501, 13.37469
163 Rue Auguste Broussonnet, 34090 Montpellier, France, Montpellier, Occit
anie 34090, FRA = 43.61584, 3.87204
Karolinenplatz 5, 64289 Darmstadt, Deutschland, Darmstadt, Hessen 64289, D
EU = 49.8746998, 8.6554584
Innstraße 41, 94032 Passau, Deutschland, Passau, Bayern 94032, DEU = 48.56

653, 13.44961
Biegenstraße 10, 35037 Marburg, Deutschland, Marburg, Hessen 35037, DEU = 50.81023, 8.77399
28049, Comunidad de Madrid, España, Comunidad de Madrid 28049, ESP = 40.56456, -3.70194
Universitätsstraße 65, 9020 Klagenfurt am Wörthersee, Österreich, Klagenfurt am Wörthersee, Kärnten 9020, AUT = 46.61671, 14.26494
Pentti Kaiteran katu 1, FI-90570 Oulu, Suomi, Oulu, Pohjois-Suomi 90570, FIN = 65.05691, 25.46811
Saarstraße 21D, 55122 Mainz, Deutschland, Mainz, Rheinland-Pfalz 55122, DEU = 49.99585, 8.24637
Via Calepina, 14, 38122 Trento TN, Italia, Trento, Trentino-Alto Adige 38122, ITA = 46.06685, 11.12305
Kalevantie 4, FI-33100 Tampere, Suomi, Tampere, Länsi- ja Sisä-Suomi 33100, FIN = 61.49432, 23.78018
3036, Lemesos, Chypre, Lemesos 3036, CYP = 34.67465, 33.04547
Fakultetsgatan 1, SE-702 81 Örebro, Sverige, Örebro, Örebro län 702 81, SWE = 59.25471, 15.24844
Universitätsstraße 30, 95447 Bayreuth, Deutschland, Bayreuth, Bayern 95447, DEU = 49.92629, 11.587
Place du Vingt Août 7, 4000 Liège, Belgique, Liège, Wallonie 4000, BEL = 50.64078, 5.57634
Keplerstraße 7, 70174 Stuttgart, Deutschland, Stuttgart, Baden-Württemberg 70174, DEU = 48.78156, 9.17473
Cannot geocode specified object: SE-901 87, Umea, Sweden
Via Festa del Perdono, 7, 20122 Milano MI, Italia, Milano, Lombardia 20122, ITA = 45.46036, 9.19399
Cannot geocode specified object: SLU, P.O. Box 7070, SE-750 07, Sweden
Via Festa del Perdono, 7, 20122 Milano MI, Italia, Milano, Lombardia 20122, ITA = 45.46036, 9.19399
741 00, Κρήτη, Ελληνικη Δημοκρατια, Κρήτη 741 00, GRC = 35.31655, 24.50966
58 Boulevard Charles Livon, 13007 Marseille, France, Marseille, Provence-Alpes-Côte d'Azur 13007, FRA = 43.29215, 5.35912
Rue Anatole France, 38400 Saint-Martin-d'Hères, France, Saint-Martin-d'Hères, Auvergne-Rhône-Alpes 38400, FRA = 45.18366, 5.75224
4 Rue Alfred Kastler, 44300 Nantes, France, Nantes, Pays de la Loire 44300, FRA = 47.28195, -1.52002
55 Avenue de Paris, 78000 Versailles, France, Versailles, Île-de-France 78000, FRA = 48.79955, 2.14153
Campus de Campolide, 1099-085 Lisboa, Portugal, Lisboa 1099-085, PRT = 38.7339, -9.15997
Caminho de Palma de Cima 21, 1600-178 Lisboa, Portugal, Lisboa 1600-178, PRT = 38.74784, -9.16717
Turku, Lounais-Suomi, Suomi, Turku, Lounais-Suomi 20100, FIN = 60.4528, 22.25155
53650, Lappeenranta, Etelä-Suomi, Suomi, Lappeenranta, Etelä-Suomi 53650, FIN = 61.0137, 28.15952
Innrain 52, 6020 Innsbruck, Österreich, Innsbruck, Tirol 6020, AUT = 47.26292, 11.38442
Martelarenlaan 42, 3500 Hasselt, België, Hasselt, Vlaanderen 3500, BEL = 50.93364, 5.34233
Via Festa del Perdono, 7, 20122 Milano MI, Italia, Milano, Lombardia 20122, ITA = 45.46036, 9.19399
Piazza Umberto I, 70121 Bari BA, Italia, Bari, Puglia 70121, ITA = 41.12001, 16.8708
Calle de Isaac Peral, 58, 28040 Madrid (Madrid), España, Madrid, Comunidad de Madrid 28040, ESP = 40.44215, -3.71859
Praça de Gomes Teixeira 2, 4050-290 Porto, Portugal, Porto 4050-290, PRT = 41.1473723, -8.6151288
Karlsplatz 13, 1040 Wien, Österreich, Wien, Wien 1040, AUT = 48.19898, 16.3699

30 Πανεπιστημίου, 106 79 Αθήνα, Ελληνικη Δημοκρατια, Αθήνα, Αττική 106 79, GRC = 37.9805689, 23.7329373
Via dell'Artigliere, 8, 37129 Verona VR, Italia, Verona, Veneto 37129, ITA = 45.43855, 11.00403
Benevento, Campania, Italia, Benevento, Campania 82100, ITA = 41.12995, 14.78553
Via Giovanni Paolo II, 106, 84084 Fisciano SA, Italia, Fisciano, Campania 84084, ITA = 40.77727, 14.78501
Via Cracovia, 00133 Roma RM, Italia, Roma, Lazio 00133, ITA = 41.85156, 12.62991
Via Giovanni Amendola, 126, 70126 Bari BA, Italia, Bari, Puglia 70126, ITA = 41.11163, 16.88271
Üllői út 26, Budapest 1088, Magyarország, Budapest, Budapest 1088, HUN = 47.48715, 19.06738
Linköping, Östergötlands län, Sverige, Linköping, Östergötlands län 582 23, SWE = 58.41109, 15.62565
Lungarno Antonio Pacinotti, 43, 56126 Pisa PI, Italia, Pisa, Toscana 56126, ITA = 43.71665, 10.39882
Corso Strada Nuova, 65, 27100 Pavia PV, Italia, Pavia, Lombardia 27100, ITA = 45.18675, 9.15586
Corso Umberto I, 40, 80138 Napoli NA, Italia, Napoli, Campania 80138, ITA = 40.84525, 14.25771
Via dell'Università, 4, 41121 Modena MO, Italia, Modena, Emilia Romagna 41121, ITA = 44.64435, 10.92815
Avenida Blasco Ibáñez, 13, 46010 Valencia (Valencia), España, Valencia, Comunidad Valenciana 46010, ESP = 39.4793, -0.36413
Piazza Università, 1, 39100 Bolzano BZ, Italia, Bolzano, Trentino-Alto Adige 39100, ITA = 46.49849, 11.35073
Piazza di San Marco, 4, 50121 Firenze FI, Italia, Firenze, Toscana 50121, ITA = 43.77782, 11.25944
Piazza del Mercato, 15, 25122 Brescia BS, Italia, Brescia, Lombardia 25122, ITA = 45.53785, 10.21769
Via Giuseppe Verdi, 8, 10124 Torino TO, Italia, Torino, Piemonte 10124, ITA = 45.0691567, 7.6900353
Via Balbi, 5, 16124 Genova GE, Italia, Genova, Liguria 16124, ITA = 44.41485, 8.92662
Banchi di Sotto, 55, 53100 Siena SI, Italia, Siena, Toscana 53100, ITA = 43.31914, 11.33277
146 Rue Léo Saignat, 33000 Bordeaux, France, Bordeaux, Nouvelle-Aquitaine 33000, FRA = 44.82704, -0.60078
Box, Sipoo, Etelä-Suomi, Suomi, Sipoo, Etelä-Suomi 01190, FIN = 60.30898, 25.39256
Yliopistokatu 2, FI-80100 Joensuu, Suomi, Joensuu, Itä-Suomi 80100, FIN = 62.60361, 29.74763
1 Rue de la Noë, 44300 Nantes, France, Nantes, Pays de la Loire 44300, FRA = 47.2482139, -1.5508602
41 Allées Jules Guesde, 31000 Toulouse, France, Toulouse, Occitanie 31000, FRA = 43.59501, 1.45131
42 Rue Scheffer, 75116 Paris, France, Paris, Île-de-France 75116, FRA = 48.8625, 2.28159
Bibliothekstraße 1, 28359 Bremen, Deutschland, Bremen, Bremen 28359, DEU = 53.10609, 8.85241
Ovocný Trh, 110 00 Praha, Česká Republika, Praha, Hlavní město Praha 110 00, CZE = 50.08658, 14.42478
August-Schmidt-Straße 4, 44227 Dortmund, Deutschland, Dortmund, Nordrhein-Westfalen 44227, DEU = 51.48443, 7.41402
Ludwigstraße 23, 35390 Gießen, Deutschland, Gießen, Hessen 35390, DEU = 50.58053, 8.67705
9 Rue Charles Fourier, 91000 Évry-Courcouronnes, France, Évry-Courcouronnes, Île-de-France 91000, FRA = 48.62424, 2.44478
Piazza dell'Università, 2, 95131 Catania CT, Italia, Catania, Sicilia 9513

1, ITA = 37.5039, 15.08735
Via Ludovico Ariosto, 35, 44121 Ferrara FE, Italia, Ferrara, Emilia Romagn
a 44121, ITA = 44.8422, 11.61619
27 Rue Saint-Guillaume, 75007 Paris, France, Paris, Île-de-France 75007, F
RA = 48.85413, 2.3284
Piazza Salvatore Pugliatti, 98122 Messina ME, Italia, Messina, Sicilia 981
22, ITA = 38.18915, 15.55237
Corso Duca degli Abruzzi, 24, 10129 Torino TO, Italia, Torino, Piemonte 10
129, ITA = 45.06244, 7.66234
Piazza Università, 21, 07100 Sassari SS, Italia, Sassari, Sardegna 07100,
ITA = 40.72499, 8.55992
Piazzale Europa, 1, 34127 Trieste TS, Italia, Trieste, Friuli-Venezia Giul
ia 34127, ITA = 45.65871, 13.79335
Via Santa Maria in Gradi, 4, 01100 Viterbo VT, Italia, Viterbo, Lazio 0110
0, ITA = 42.41321, 12.11185
Via Aurelio Saffi, 2, 61029 Urbino PU, Italia, Urbino, Marche 61029, ITA =
43.72326, 12.63685
Alameda da Universidade, 1649-004 Lisboa, Portugal, Lisboa 1649-004, PRT =
38.75247, -9.15896
Avenida de Séneca, 2, 28040 Madrid (Madrid), España, Madrid, Comunidad de
Madrid 28040, ESP = 40.43694, -3.72474
Piazza Santa Margherita, 1, 67100 L'Aquila AQ, Italia, L'Aquila, Abruzzo 6
7100, ITA = 42.35145, 13.39772
43 Boulevard du 11 Novembre 1918, 69100 Villeurbanne, France, Villeurbann
e, Auvergne-Rhône-Alpes 69100, FRA = 45.7791, 4.86573
ulica Gołębia 24, 31-007 Kraków, Polska, Kraków, Woj. Małopolskie 31-007,
POL = 50.0609578, 19.9334482
3 Rue Joliot-Curie, 91190 Gif-sur-Yvette, France, Gif-sur-Yvette, Île-de-F
rance 91190, FRA = 48.70961, 2.16401
Rechbauerstraße 12, 8010 Graz, Österreich, Graz, Steiermark 8010, AUT = 4
7.06818, 15.4494
28 Avenue Valrose, 06100 Nice, France, Nice, Provence-Alpes-Côte d'Azur 06
100, FRA = 43.71823, 7.2668
Welfengarten 1, 30167 Hannover, Deutschland, Hannover, Niedersachsen 3016
7, DEU = 52.38224, 9.71776
Via Ravasi, 2, 21100 Varese VA, Italia, Varese, Lombardia 21100, ITA = 45.
81466, 8.82792
Οδός Πανεπιστημίου, 2116 Αγλαντζιά, Κύπρος, Αγλαντζιά 2116, CYP = 35.1428
5, 33.40549
4 Rue Blaise Pascal, 67000 Strasbourg, France, Strasbourg, Grand Est 6700
0, FRA = 48.58071, 7.76657
Křížkovského 511/8, 779 00 Olomouc, Česká Republika, Olomouc, Olomoucký kr
aj 779 00, CZE = 49.59513, 17.25922
34 Cours Léopold, 54000 Nancy, France, Nancy, Grand Est 54000, FRA = 48.69
615, 6.1766
1 Quai de Tourville, 44000 Nantes, France, Nantes, Pays de la Loire 44000,
FRA = 47.2094, -1.55589
20 Avenue Albert Einstein, 69100 Villeurbanne, France, Villeurbanne, Auver
gne-Rhône-Alpes 69100, FRA = 45.7834949, 4.8786717
Msida, Malta, Msida MSD, MLT = 35.89919, 14.48813
12 Place du Panthéon, 75005 Paris, France, Paris, Île-de-France 75005, FRA
= 48.8468, 2.34488
Ιωάννινα, Ήπειρος, Ελληνικη Δημοκρατια, Ιωάννινα, Ήπειρος 452 21, GRC = 3
9.66858, 20.85638
42 Rue Paul Duez, 59800 Lille, France, Lille, Hauts-de-France 59800, FRA =
50.63179, 3.07526
9 Ηρώων Πολυτεχνείου, 157 73 Ζωγράφος, Ελληνικη Δημοκρατια, Ζωγράφος, Αττι
κή 157 73, GRC = 37.9785993, 23.7734851
Carretera de Valldemossa, 07120 Palma (Illes Balears), Espanya, Palma, Ill
es Balears 07120, ESP = 39.64081, 2.64879
Raiņa bulvāris 19, Rīga, LV-1050, Latvija, Rīga 1050, LVA = 56.95063, 24.1

1578
Avenida Universidades, 24, 48007 Bilbao (Vizcaya), España, Bilbao, País Va
sco 48007, ESP = 43.2703, -2.93716
Via Andrea Palladio, 8, 33100 Udine UD, Italia, Udine, Friuli-Venezia Giul
ia 33100, ITA = 46.06629, 13.23273
Piazzetta Tancredi, 73100 Lecce LE, Italia, Lecce, Puglia 73100, ITA = 40.
34969, 18.16735
Žerotínovo náměstí 617/9, 602 00 Brno, Česká Republika, Brno, Jihomoravský
kraj 602 00, CZE = 49.19883, 16.60523
Αργοναυτών & Φιλελλήνων, 382 21 Βόλος, Ελληνικη Δημοκρατια, Βόλος, Θεσσαλί
α 382 21, GRC = 39.35707, 22.9509
Αθήνα, Αττική, Ελληνικη Δημοκρατια, Αθήνα, Αττική 106 71, GRC = 37.97614,
23.7364
Altenberger Straße 69, 4040 Linz, Österreich, Linz, Oberösterreich 4040, A
UT = 48.33733, 14.32256
Universitätsplatz 3, 8010 Graz, Österreich, Graz, Steiermark 8010, AUT = 4
7.0776, 15.44954
Am Schwarzenberg-Campus, 21073 Hamburg, Deutschland, Hamburg, Hamburg 2107
3, DEU = 53.46383, 9.96976
Gottlieb-Daimler-Straße, 67663 Kaiserslautern, Deutschland, Kaiserslauter
n, Rheinland-Pfalz 67663, DEU = 49.42259, 7.75349
3810-193, Aveiro, Portugal, Aveiro 3810-193, PRT = 40.63413, -8.65799
Rua de Santo António 1, 6200-811 Covilhã, Portugal, Covilhã 6200-811, PRT
= 40.19977, -7.54486
Rua das Escolas 525, 3060-711 Cantanhede, Portugal, Cantanhede 3060-711, P
RT = 40.31425, -8.75085

In [33]:

```
df_eu_univs.head(10)
```

Out[33]:

	Unnamed: 0	Code	Rank	University	URL	A
0	0	DE	8	LMU Munich	https://www.timeshighereducation.com/world-uni...	Geschwister Platz 1, 8053
1	1	SE	10	Karolinska Institute	https://www.timeshighereducation.com/world-uni...	SE- Stockholm,
2	2	DE	11	Technical University of Munich	https://www.timeshighereducation.com/world-uni...	Arcisstr Munich, D G
3	3	DE	12	Heidelberg University	https://www.timeshighereducation.com/world-uni...	310 E. Street, Tiffi 4488
4	4	BE	14	KU Leuven	https://www.timeshighereducation.com/world-uni...	Oude M Leuve I
5	5	FR	15	Paris Sciences et Lettres – PSL Research Unive...	https://www.timeshighereducation.com/world-uni...	60 rue M Paris,
6	6	NL	17	Wageningen University & Research	https://www.timeshighereducation.com/world-uni...	Droevendaal 4, Building W
7	7	NL	18	University of Amsterdam	https://www.timeshighereducation.com/world-uni...	P.O. Box 1 Ams Neth
8	8	NL	19	Leiden University	https://www.timeshighereducation.com/world-uni...	PO Bc Leide Neth
9	9	NL	20	Erasmus University Rotterdam	https://www.timeshighereducation.com/world-uni...	Burge Oud Rotterda

In [34]:

```
df_eu_univs.tail(10)
```

Out[34]:

	Unnamed: 0	Code	Rank	University	URL	A
190	190	CZ	279	Masaryk University	https://www.timeshighereducation.com/world-uni...	Že na Rektora mest
191	191	EL	279	University of Thessaly	https://www.timeshighereducation.com/world-uni...	Argo Filellinor Volos,
192	192	EL	279	Athens University of Economics and Business	https://www.timeshighereducation.com/world-uni...	76, F Str., G Athens,
193	193	AT	279	Johannes Kepler University of Linz	https://www.timeshighereducation.com/world-uni...	Alte Straß 40
194	194	AT	279	University of Graz	https://www.timeshighereducation.com/world-uni...	Universi 3, Graz
195	195	DE	279	Hamburg University of Technology	https://www.timeshighereducation.com/world-uni...	Schwarz H G
196	196	DE	279	University of Kaiserslautern	https://www.timeshighereducation.com/world-uni...	(I Kaisers
197	197	PT	279	University of Aveiro	https://www.timeshighereducation.com/world-uni...	Aveir 193, I
198	198	PT	279	University of Beira Interior	https://www.timeshighereducation.com/world-uni...	Conv Sto. 6 Covill
199	199	PT	279	University of Coimbra	https://www.timeshighereducation.com/world-uni...	P I Coimbr 531, I

In [35]:

```
# See the records with unsuccessful geocoding
uncoded_univs = list(df_eu_univs[df_eu_univs.Lat.isna()].index)
uncoded_univs
```

Out[35]:

```
[7, 8, 11, 41, 105, 107]
```

In [36]:

```
# Let's try to process the addresses where P.O.Boxes did not allow to geocode properly
for i in uncoded_univs:
    # if address contains P.O.Box then we remove this part of the address
    addr = df_eu_univs.Address[i].split(', ')
    for j in range(len(addr)):
        if re.search('Box [0-9]{3,}', addr[j]) or re.search('Postfach [0-9]*', addr[j
]):
            addr[j] = ''
    df_eu_univs.loc[i, 'Lat'], df_eu_univs.loc[i, 'Lon'] = get_location(', '.join(addr
))
```

```
Amsterdam, Noord-Holland, Nederland, Amsterdam, Noord-Holland 1011 MG, NLD
= 52.36994, 4.90788
Leiden, Zuid-Holland, Nederland, Leiden, Zuid-Holland 2311, NLD = 52.1536
4, 4.49381
Utrecht, Nederland, Utrecht, Utrecht 3511, NLD = 52.08979, 5.11415
Bielefeld, Nordrhein-Westfalen, Deutschland, Bielefeld, Nordrhein-Westfale
n 33604, DEU = 52.01548, 8.53232
Cannot geocode specified object: SE-901 87, Umea, Sweden
Cannot geocode specified object: SLU, , SE-750 07, Sweden
```

In [37]:

```
# See the results of this extra geocoding
df_eu_univs.iloc[uncoded_univs,]
```

Out[37]:

	Unnamed: 0	Code	Rank	University	URL	Addres
7	7	NL	18	University of Amsterdam	https://www.timeshighereducation.com/world-uni...	P.O. Bo: 19268 1000 GC Amsterdam Netherland:
8	8	NL	19	Leiden University	https://www.timeshighereducation.com/world-uni...	PO Bo: 9500 Leiden 2300 Netherland:
11	11	NL	23	Utrecht University	https://www.timeshighereducation.com/world-uni...	P.O Bo: 80125, TC Utrecht 3508 Netherland:
41	41	DE	68	Bielefeld University	https://www.timeshighereducation.com/world-uni...	Postfach 1(01 31, D 3350' Bielefeld German)
105	105	SE	164	Umeå University	https://www.timeshighereducation.com/world-uni...	SE-901 87 Umea Sweder
107	107	SE	164	Swedish University of Agricultural Sciences	https://www.timeshighereducation.com/world-uni...	SLU, P.O Box 7070 SE-750 07 Sweder

In [38]:

```
# Remove records that could not be geocoded eventually
df_eu_univs.dropna(axis=0, subset=['Lat', 'Lon'], inplace=True)
df_eu_univs.reset_index(drop=True, inplace=True)
print(df_eu_univs.shape)
```

```
(198, 8)
```

In [39]:

```
df_eu_univs.tail(10)
```

Out[39]:

	Unnamed: 0	Code	Rank	University	URL	A
188	190	CZ	279	Masaryk University	https://www.timeshighereducation.com/world-uni...	Že na Rektora mest
189	191	EL	279	University of Thessaly	https://www.timeshighereducation.com/world-uni...	Argo Fillellinor Volos,
190	192	EL	279	Athens University of Economics and Business	https://www.timeshighereducation.com/world-uni...	76, P Str., G Athens,
191	193	AT	279	Johannes Kepler University of Linz	https://www.timeshighereducation.com/world-uni...	Alte Straß 40
192	194	AT	279	University of Graz	https://www.timeshighereducation.com/world-uni...	Universit 3, / Graz
193	195	DE	279	Hamburg University of Technology	https://www.timeshighereducation.com/world-uni...	Schwarz H G
194	196	DE	279	University of Kaiserslautern	https://www.timeshighereducation.com/world-uni...	(I Kaisers
195	197	PT	279	University of Aveiro	https://www.timeshighereducation.com/world-uni...	Aveir 193, I
196	198	PT	279	University of Beira Interior	https://www.timeshighereducation.com/world-uni...	Conv Sto. . 6. Covill
197	199	PT	279	University of Coimbra	https://www.timeshighereducation.com/world-uni...	P I Coimbr 531, I

In [40]:

```
df_eu_univs.to_csv('top_eu_univs.csv')
```

2.3. Use location data to reduce number of cities to analyze

Now, we will narrow down our further analysis and exclude all the cities that do not have top-ranked universities located nearby.

That is, if the number of universities located in a specified range around the city (`CITY_UNIV_RADIUS`) is less than the specified minimum acceptable value (`UNIVS_IN_NEGHBORHOOD`), then the city will be excluded from our dataset.

In [41]:

```
# Calculate number of universities in radius of each city
CITY_UNIV_RADIUS = 50.0          # range in kilometers
UNIVS_IN_NEIGHBORHOOD = 1        # minimum acceptable numbers of universities around a ci
ty
city_has_univs = [0] * df_cities.shape[0]
for i in df_cities.index:
    univs_in_radius = 0
    for j in df_eu_univs.index:
        dist = distance.distance(
            (df_cities.Lat[i], df_cities.Lon[i]),
            (df_eu_univs.Lat[j], df_eu_univs.Lon[j])).km
        if dist <= CITY_UNIV_RADIUS:
            univs_in_radius += 1
    city_has_univs[i] = univs_in_radius
df_cities['HasUnivs'] = city_has_univs
df_cities.head(10)
```

Out[41]:

	City	Country	Population	Lat	Lon	HasUnivs
0	Berlin	Germany	3669495	52.51605	13.37691	5
1	Madrid	Spain	3348536	40.41956	-3.69196	3
2	Rome	Italy	2856133	41.90323	12.49566	2
3	Bucharest	Romania	2155240	44.43430	26.10298	0
4	Paris	France	2140526	48.85718	2.34141	14
5	Vienna	Austria	1921153	48.20263	16.36843	3
6	Hamburg	Germany	1899160	53.55562	9.98746	2
7	Warsaw	Poland	1793579	52.23560	21.01038	0
8	Budapest	Hungary	1752286	47.49973	19.05508	1
9	Barcelona	Spain	1620343	41.38804	2.17001	3

In [42]:

```
df_cities.tail(10)
```

Out[42]:

	City	Country	Population	Lat	Lon	HasUnivs
83	Cluj-Napoca	Romania	324960	46.76870	23.58503	0
84	Bari	Italy	320862	41.12588	16.86666	2
85	Constanța	Romania	317832	44.17827	28.65116	0
86	Münster	Germany	314319	51.96302	7.61782	1
87	Karlsruhe	Germany	313092	49.01094	8.40846	1
88	Catania	Italy	311584	37.51136	15.06752	1
89	Mannheim	Germany	309370	49.48651	8.46679	2
90	Nantes	France	306694	47.21812	-1.55306	3
91	Craiova	Romania	305386	44.32020	23.79895	0
92	Galați	Romania	304050	45.43369	28.05476	0

In [43]:

```
# Filter out cities with insufficient number of universities
df_cities = df_cities[df_cities.HasUnivs >= UNIVS_IN_NEIGHBORHOOD]
df_cities.reset_index(drop=True, inplace=True)
df_cities.shape
```

Out[43]:

```
(60, 6)
```

In [44]:

```
df_cities.sort_values(by=['Population'], ascending=False, ignore_index=True)
```

Out[44]:

	City	Country	Population	Lat	Lon	HasUnivs
0	Berlin	Germany	3669495	52.51605	13.37691	5
1	Madrid	Spain	3348536	40.41956	-3.69196	3
2	Rome	Italy	2856133	41.90323	12.49566	2
3	Paris	France	2140526	48.85718	2.34141	14
4	Vienna	Austria	1921153	48.20263	16.36843	3
5	Hamburg	Germany	1899160	53.55562	9.98746	2
6	Budapest	Hungary	1752286	47.49973	19.05508	1
7	Barcelona	Spain	1620343	41.38804	2.17001	3
8	Munich	Germany	1558395	48.13642	11.57755	2
9	Milan	Italy	1404239	45.46796	9.18178	6
10	Prague	Czech Republic	1324277	50.07913	14.43303	1
11	Cologne	Germany	1085664	50.94168	6.95517	2
12	Stockholm	Sweden	974073	59.33258	18.06683	3
13	Naples	Italy	959188	40.84016	14.25222	2
14	Turin	Italy	875698	45.06236	7.67994	2
15	Amsterdam	Netherlands	873289	52.36994	4.90788	4
16	Marseille	France	868277	43.29338	5.37132	1
17	Copenhagen	Denmark	794128	55.67567	12.56756	3
18	Valencia	Spain	791413	39.46895	-0.37686	1
19	Kraków	Poland	780981	50.06045	19.93243	1
20	Frankfurt	Germany	753056	50.11208	8.68342	3
21	Athens	Greece	664046	37.97614	23.73640	3
22	Helsinki	Finland	657674	60.17116	24.93266	3
23	Rotterdam	Netherlands	651870	51.91439	4.48717	4
24	Stuttgart	Germany	635911	48.76779	9.17203	3
25	Riga	Latvia	627487	56.94599	24.11487	1
26	Düsseldorf	Germany	619294	51.21564	6.77662	3
27	Dortmund	Germany	587010	51.51661	7.45830	3
28	Essen	Germany	583393	51.45183	7.01109	3
29	Gothenburg	Sweden	579281	57.70068	11.96823	2
30	Genoa	Italy	578000	44.41048	8.93917	1
31	Bremen	Germany	569352	53.07537	8.80454	2
32	Dresden	Germany	554649	51.05364	13.74082	1
33	The Hague	Netherlands	545273	52.08409	4.31732	4
34	Hanover	Germany	538068	52.37228	9.73816	1
35	Antwerp	Belgium	525935	51.22213	4.39769	4
36	Nuremberg	Germany	518365	49.45435	11.07350	1

	City	Country	Population	Lat	Lon	HasUnivs
37	Lyon	France	515695	45.75917	4.82966	2
38	Lisbon	Portugal	506654	38.72639	-9.14949	3
39	Duisburg	Germany	498590	51.43148	6.76356	3
40	Toulouse	France	479638	43.60579	1.44864	1
41	Palma de Mallorca	Spain	409661	39.57149	2.64694	1
42	Bologna	Italy	390636	44.50485	11.34507	3
43	Brno	Czech Republic	381346	49.19728	16.60368	1
44	Florence	Italy	378839	43.78238	11.25502	1
45	Bochum	Germany	364628	51.48800	7.21399	3
46	Utrecht	Netherlands	357676	52.08979	5.11415	6
47	Wuppertal	Germany	354382	51.27165	7.19678	4
48	Aarhus	Denmark	349977	56.15302	10.20487	1
49	Bilbao	Spain	345821	43.26890	-2.94530	1
50	Malmö	Sweden	344166	55.59670	13.00110	3
51	Nice	France	342637	43.70029	7.27766	1
52	Bielefeld	Germany	333786	52.01548	8.53232	1
53	Bonn	Germany	327258	50.73243	7.10187	2
54	Bari	Italy	320862	41.12588	16.86666	2
55	Münster	Germany	314319	51.96302	7.61782	1
56	Karlsruhe	Germany	313092	49.01094	8.40846	1
57	Catania	Italy	311584	37.51136	15.06752	1
58	Mannheim	Germany	309370	49.48651	8.46679	2
59	Nantes	France	306694	47.21812	-1.55306	3

In [45]:

```
df_cities.to_csv('top_eu_cities.csv')
#df_cities = pd.read_csv('top_eu_cities.csv')
```

2.4. The resulting dataset of EU cities

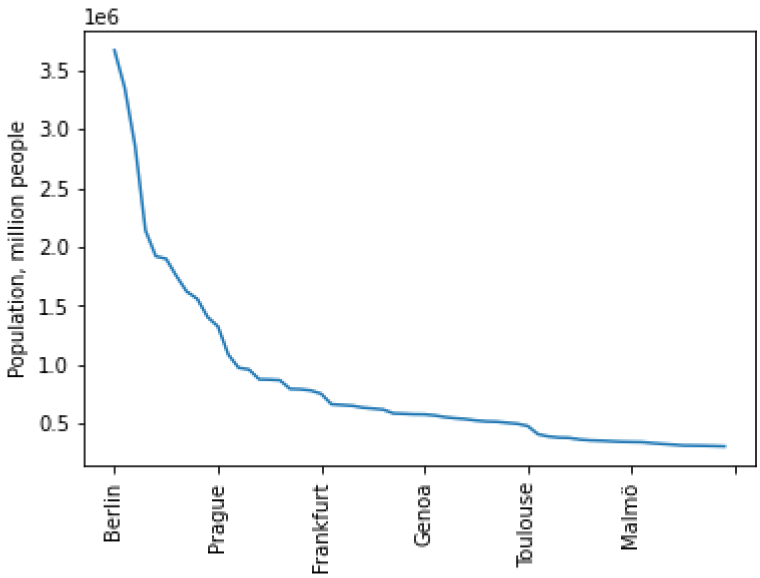
Now, let's take a look at how the cities of our interest are distributed in terms of size:

In [46]:

```
df_cities.plot(x='City', y='Population', xlabel='', ylabel='Population, million people'
, rot=90, legend=False)
```

Out[46]:

```
<AxesSubplot:ylabel='Population, million people'>
```

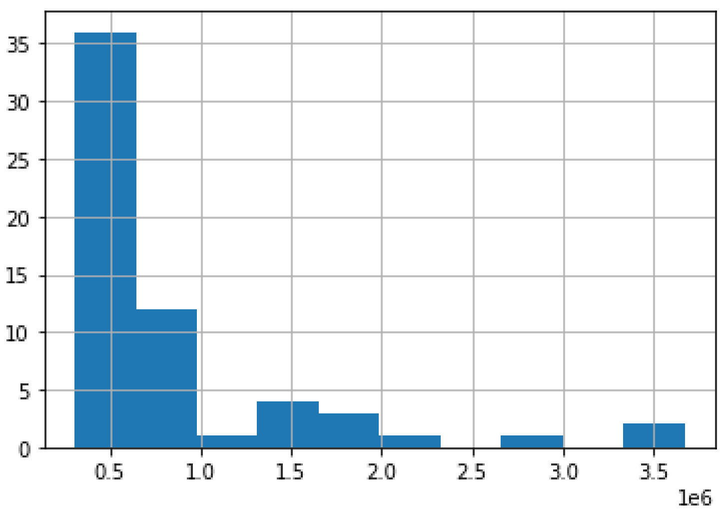


In [47]:

```
df_cities.Population.hist()
```

Out[47]:

```
<AxesSubplot:>
```



3. Visualizing the cities on a map

Now, let's see how the cities are located:

In [48]:

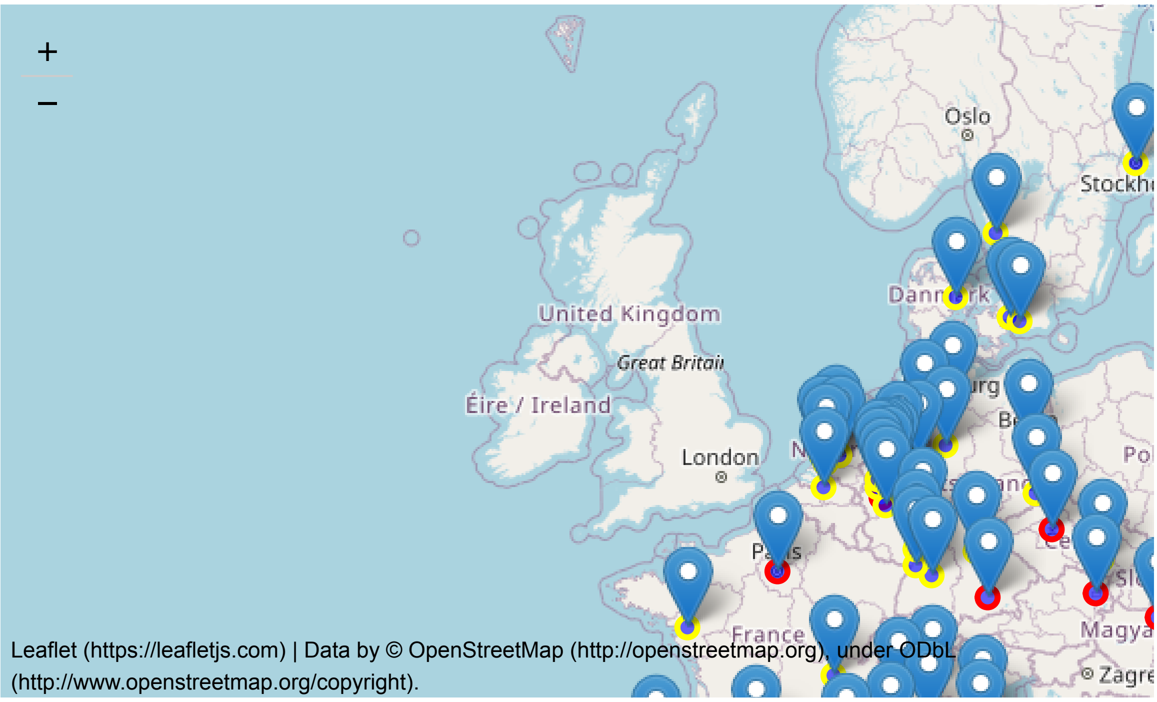
```
# Install Folium library to visualize cities on a map
!conda install -c conda-forge folium --yes
import folium
print('Folium installed and imported!')
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /home/jupyterlab/conda/envs/python

  added / updated specs:
    - folium


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    branca-0.4.2               |     pyhd8ed1ab_0          26 KB  conda-fo
rge
    folium-0.12.0              |     pyhd8ed1ab_1          64 KB  conda-fo
rge
    ------------------------------------------------------------
                                           Total:          90 KB

The following NEW packages will be INSTALLED:

  branca             conda-forge/noarch::branca-0.4.2-pyhd8ed1ab_0
  folium             conda-forge/noarch::folium-0.12.0-pyhd8ed1ab_1



Downloading and Extracting Packages
branca-0.4.2         | 26 KB     | ##################################### |
100%
folium-0.12.0        | 64 KB     | ##################################### |
100%
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
Folium installed and imported!
```

In [49]:

```
# Define the world map centered around Europe
location_center = df_cities.loc[df_cities['City'] == 'Munich', ['Lat', 'Lon']].values.t
olist()[0]
map_europe = folium.Map(location=location_center, zoom_start=4)

# Create and fill a feature group for the cities in the dataframe
feat_cities = folium.map.FeatureGroup()
for lat, lon, label, popul in zip(df_cities.Lat, df_cities.Lon, df_cities.City, df_citi
es.Population):
    folium.Marker([lat, lon], popup=label).add_to(map_europe)
    feat_color = 'red' if popul >= 1_000_000 else 'yellow'
    feat_cities.add_child(
        folium.features.CircleMarker(
            [lat, lon],
            radius=5,
            color=feat_color,
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )
map_europe.add_child(feat_cities)
map_europe
```

Out[49]:



4. Using Foursqare API to explore venues in the cities

In [50]:

```
# Define Foursquare API settings
API_VERSION = '20180605'     # API version
API_LIMIT = 100              # maximum records returned for one API request
API_RETRIES = 2              # how many times we retry an API request if an error occurs
```

In [51]:

```python
# Define a function that explores all the neighborhoods/cities
from time import sleep
def get_nearby_venues(names, latitudes, longitudes, radius=500):

    venues_list = []
    for name, lat, lon in zip(names, latitudes, longitudes):

        need_venues = 1
        api_offset = 0
        while api_offset < need_venues:

            # Create the API request URL
            url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_se
cret={}&v={}'\
                  '&ll={},{}&radius={}'\
                  '&limit={}&offset={}'.format(
                my_secrets['FOURSQUARE']['CLIENT_ID'],
                my_secrets['FOURSQUARE']['CLIENT_SECRET'],
                API_VERSION,
                lat, lon, radius,
                API_LIMIT, api_offset)

            # Make the GET request with retries
            print(name, lat, lon)
            current_attempt = 0
            while current_attempt <= API_RETRIES:
                print('Sending request to Foursquare API... ', end='')
                try:
                    response = rq.get(url).json()['response']
                    results = response['groups'][0]['items']
                    need_venues = int(response['totalResults'])
                    print('Success!')
                    print('N of venues total =', need_venues)
                    print('N of venues received =', len(results))
                    break
                except:
                    response = None
                    results = None
                    print('Error!')
                    current_attempt += 1
                    sleep(1)

            if results:
                # Return only relevant information for each nearby venue
                venues_list.append([(
                    name,
                    lat,
                    lon,
                    v['venue']['name'],
                    v['venue']['location']['lat'],
                    v['venue']['location']['lng'],
                    v['venue']['categories'][0]['name']) for v in results])
                api_offset += len(results)
            else:
                print('Could not retrieve data for', name)

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_
list])
```

```
    nearby_venues.columns = [
        'Neighborhood',
        'Neighborhood Latitude',
        'Neighborhood Longitude',
        'Venue',
        'Venue Latitude',
        'Venue Longitude',
        'Venue Category']
    return (nearby_venues)
```

In [52]:

```
# Let's select a subset of cities for further analysis.
# It may be useful to limit our dataset by some criteria,
# e.g. by lower boundary population and/or certain range of latitudes
cities_explored = df_cities[df_cities.Population >= 300_000]
cities_explored
```

Out[52]:

	City	Country	Population	Lat	Lon	HasUnivs
0	Berlin	Germany	3669495	52.51605	13.37691	5
1	Madrid	Spain	3348536	40.41956	-3.69196	3
2	Rome	Italy	2856133	41.90323	12.49566	2
3	Paris	France	2140526	48.85718	2.34141	14
4	Vienna	Austria	1921153	48.20263	16.36843	3
5	Hamburg	Germany	1899160	53.55562	9.98746	2
6	Budapest	Hungary	1752286	47.49973	19.05508	1
7	Barcelona	Spain	1620343	41.38804	2.17001	3
8	Munich	Germany	1558395	48.13642	11.57755	2
9	Milan	Italy	1404239	45.46796	9.18178	6
10	Prague	Czech Republic	1324277	50.07913	14.43303	1
11	Cologne	Germany	1085664	50.94168	6.95517	2
12	Stockholm	Sweden	974073	59.33258	18.06683	3
13	Naples	Italy	959188	40.84016	14.25222	2
14	Turin	Italy	875698	45.06236	7.67994	2
15	Amsterdam	Netherlands	873289	52.36994	4.90788	4
16	Marseille	France	868277	43.29338	5.37132	1
17	Copenhagen	Denmark	794128	55.67567	12.56756	3
18	Valencia	Spain	791413	39.46895	-0.37686	1
19	Kraków	Poland	780981	50.06045	19.93243	1
20	Frankfurt	Germany	753056	50.11208	8.68342	3
21	Athens	Greece	664046	37.97614	23.73640	3
22	Helsinki	Finland	657674	60.17116	24.93266	3
23	Rotterdam	Netherlands	651870	51.91439	4.48717	4
24	Stuttgart	Germany	635911	48.76779	9.17203	3
25	Riga	Latvia	627487	56.94599	24.11487	1
26	Düsseldorf	Germany	619294	51.21564	6.77662	3
27	Dortmund	Germany	587010	51.51661	7.45830	3
28	Essen	Germany	583393	51.45183	7.01109	3
29	Gothenburg	Sweden	579281	57.70068	11.96823	2
30	Genoa	Italy	578000	44.41048	8.93917	1
31	Bremen	Germany	569352	53.07537	8.80454	2
32	Dresden	Germany	554649	51.05364	13.74082	1
33	The Hague	Netherlands	545273	52.08409	4.31732	4
34	Hanover	Germany	538068	52.37228	9.73816	1
35	Antwerp	Belgium	525935	51.22213	4.39769	4
36	Nuremberg	Germany	518365	49.45435	11.07350	1

	City	Country	Population	Lat	Lon	HasUnivs
37	Lyon	France	515695	45.75917	4.82966	2
38	Lisbon	Portugal	506654	38.72639	-9.14949	3
39	Duisburg	Germany	498590	51.43148	6.76356	3
40	Toulouse	France	479638	43.60579	1.44864	1
41	Palma de Mallorca	Spain	409661	39.57149	2.64694	1
42	Bologna	Italy	390636	44.50485	11.34507	3
43	Brno	Czech Republic	381346	49.19728	16.60368	1
44	Florence	Italy	378839	43.78238	11.25502	1
45	Bochum	Germany	364628	51.48800	7.21399	3
46	Utrecht	Netherlands	357676	52.08979	5.11415	6
47	Wuppertal	Germany	354382	51.27165	7.19678	4
48	Aarhus	Denmark	349977	56.15302	10.20487	1
49	Bilbao	Spain	345821	43.26890	-2.94530	1
50	Malmö	Sweden	344166	55.59670	13.00110	3
51	Nice	France	342637	43.70029	7.27766	1
52	Bielefeld	Germany	333786	52.01548	8.53232	1
53	Bonn	Germany	327258	50.73243	7.10187	2
54	Bari	Italy	320862	41.12588	16.86666	2
55	Münster	Germany	314319	51.96302	7.61782	1
56	Karlsruhe	Germany	313092	49.01094	8.40846	1
57	Catania	Italy	311584	37.51136	15.06752	1
58	Mannheim	Germany	309370	49.48651	8.46679	2
59	Nantes	France	306694	47.21812	-1.55306	3

In [53]:

```
# Find all venues in the selected cities
df_eur_venues = get_nearby_venues(cities_explored.City, cities_explored.Lat, cities_exp
lored.Lon, radius=20_000)
```

```
Berlin 52.51605 13.37691
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Berlin 52.51605 13.37691
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Berlin 52.51605 13.37691
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 35
Madrid 40.41956 -3.69196
Sending request to Foursquare API... Success!
N of venues total = 229
N of venues received = 100
Madrid 40.41956 -3.69196
Sending request to Foursquare API... Success!
N of venues total = 229
N of venues received = 100
Madrid 40.41956 -3.69196
Sending request to Foursquare API... Success!
N of venues total = 229
N of venues received = 29
Rome 41.90323 12.49566
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 100
Rome 41.90323 12.49566
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 100
Rome 41.90323 12.49566
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 38
Paris 48.85718 2.34141
Sending request to Foursquare API... Success!
N of venues total = 205
N of venues received = 100
Paris 48.85718 2.34141
Sending request to Foursquare API... Success!
N of venues total = 205
N of venues received = 100
Paris 48.85718 2.34141
Sending request to Foursquare API... Success!
N of venues total = 205
N of venues received = 5
Vienna 48.20263 16.36843
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Vienna 48.20263 16.36843
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Vienna 48.20263 16.36843
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 41
Hamburg 53.55562 9.98746
```

```
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 100
Hamburg 53.55562 9.98746
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 100
Hamburg 53.55562 9.98746
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 34
Budapest 47.49973 19.05508
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Budapest 47.49973 19.05508
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Budapest 47.49973 19.05508
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Barcelona 41.38804 2.17001
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Barcelona 41.38804 2.17001
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Barcelona 41.38804 2.17001
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 41
Munich 48.13642 11.57755
Sending request to Foursquare API... Success!
N of venues total = 243
N of venues received = 100
Munich 48.13642 11.57755
Sending request to Foursquare API... Success!
N of venues total = 243
N of venues received = 100
Munich 48.13642 11.57755
Sending request to Foursquare API... Success!
N of venues total = 243
N of venues received = 43
Milan 45.46796 9.18178
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 100
Milan 45.46796 9.18178
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 100
Milan 45.46796 9.18178
Sending request to Foursquare API... Success!
N of venues total = 234
N of venues received = 34
Prague 50.07913 14.43303
Sending request to Foursquare API... Success!
```

```
N of venues total = 242
N of venues received = 100
Prague 50.07913 14.43303
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Prague 50.07913 14.43303
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Cologne 50.94168 6.95517
Sending request to Foursquare API... Success!
N of venues total = 200
N of venues received = 100
Cologne 50.94168 6.95517
Sending request to Foursquare API... Success!
N of venues total = 200
N of venues received = 100
Stockholm 59.33258 18.06683
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Stockholm 59.33258 18.06683
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Stockholm 59.33258 18.06683
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Naples 40.84016 14.25222
Sending request to Foursquare API... Success!
N of venues total = 214
N of venues received = 100
Naples 40.84016 14.25222
Sending request to Foursquare API... Success!
N of venues total = 214
N of venues received = 100
Naples 40.84016 14.25222
Sending request to Foursquare API... Success!
N of venues total = 214
N of venues received = 14
Turin 45.06236 7.67994
Sending request to Foursquare API... Success!
N of venues total = 193
N of venues received = 100
Turin 45.06236 7.67994
Sending request to Foursquare API... Success!
N of venues total = 193
N of venues received = 93
Amsterdam 52.36994 4.90788
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Amsterdam 52.36994 4.90788
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Amsterdam 52.36994 4.90788
Sending request to Foursquare API... Success!
N of venues total = 242
```

```
N of venues received = 42
Marseille 43.29338 5.37132
Sending request to Foursquare API... Success!
N of venues total = 172
N of venues received = 100
Marseille 43.29338 5.37132
Sending request to Foursquare API... Success!
N of venues total = 172
N of venues received = 72
Copenhagen 55.67567 12.56756
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Copenhagen 55.67567 12.56756
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Copenhagen 55.67567 12.56756
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Valencia 39.46895 -0.37686
Sending request to Foursquare API... Success!
N of venues total = 140
N of venues received = 100
Valencia 39.46895 -0.37686
Sending request to Foursquare API... Success!
N of venues total = 140
N of venues received = 40
Kraków 50.06045 19.93243
Sending request to Foursquare API... Success!
N of venues total = 183
N of venues received = 100
Kraków 50.06045 19.93243
Sending request to Foursquare API... Success!
N of venues total = 183
N of venues received = 83
Frankfurt 50.11208 8.68342
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Frankfurt 50.11208 8.68342
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Frankfurt 50.11208 8.68342
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 35
Athens 37.97614 23.7364
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Athens 37.97614 23.7364
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Athens 37.97614 23.7364
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 32
```

```
Helsinki 60.17116 24.93266
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Helsinki 60.17116 24.93266
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Helsinki 60.17116 24.93266
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 41
Rotterdam 51.91439 4.48717
Sending request to Foursquare API... Success!
N of venues total = 246
N of venues received = 100
Rotterdam 51.91439 4.48717
Sending request to Foursquare API... Success!
N of venues total = 246
N of venues received = 100
Rotterdam 51.91439 4.48717
Sending request to Foursquare API... Success!
N of venues total = 246
N of venues received = 46
Stuttgart 48.76779 9.17203
Sending request to Foursquare API... Success!
N of venues total = 171
N of venues received = 100
Stuttgart 48.76779 9.17203
Sending request to Foursquare API... Success!
N of venues total = 171
N of venues received = 71
Riga 56.94599 24.11487
Sending request to Foursquare API... Success!
N of venues total = 226
N of venues received = 100
Riga 56.94599 24.11487
Sending request to Foursquare API... Success!
N of venues total = 226
N of venues received = 100
Riga 56.94599 24.11487
Sending request to Foursquare API... Success!
N of venues total = 226
N of venues received = 26
Düsseldorf 51.21564 6.77662
Sending request to Foursquare API... Success!
N of venues total = 240
N of venues received = 100
Düsseldorf 51.21564 6.77662
Sending request to Foursquare API... Success!
N of venues total = 240
N of venues received = 100
Düsseldorf 51.21564 6.77662
Sending request to Foursquare API... Success!
N of venues total = 240
N of venues received = 40
Dortmund 51.51661 7.4583
Sending request to Foursquare API... Success!
N of venues total = 223
N of venues received = 100
Dortmund 51.51661 7.4583
```

```
Sending request to Foursquare API... Success!
N of venues total = 223
N of venues received = 100
Dortmund 51.51661 7.4583
Sending request to Foursquare API... Success!
N of venues total = 223
N of venues received = 23
Essen 51.45183 7.01109
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Essen 51.45183 7.01109
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Essen 51.45183 7.01109
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Gothenburg 57.70068 11.96823
Sending request to Foursquare API... Success!
N of venues total = 155
N of venues received = 100
Gothenburg 57.70068 11.96823
Sending request to Foursquare API... Success!
N of venues total = 155
N of venues received = 55
Genoa 44.41048 8.93917
Sending request to Foursquare API... Success!
N of venues total = 83
N of venues received = 83
Bremen 53.07537 8.80454
Sending request to Foursquare API... Success!
N of venues total = 189
N of venues received = 100
Bremen 53.07537 8.80454
Sending request to Foursquare API... Success!
N of venues total = 189
N of venues received = 89
Dresden 51.05364 13.74082
Sending request to Foursquare API... Success!
N of venues total = 99
N of venues received = 99
The Hague 52.08409 4.31732
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
The Hague 52.08409 4.31732
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
The Hague 52.08409 4.31732
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Hanover 52.37228 9.73816
Sending request to Foursquare API... Success!
N of venues total = 168
N of venues received = 100
Hanover 52.37228 9.73816
Sending request to Foursquare API... Success!
```

```
N of venues total = 168
N of venues received = 68
Antwerp 51.22213 4.39769
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Antwerp 51.22213 4.39769
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Antwerp 51.22213 4.39769
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 32
Nuremberg 49.45435 11.0735
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Nuremberg 49.45435 11.0735
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 100
Nuremberg 49.45435 11.0735
Sending request to Foursquare API... Success!
N of venues total = 232
N of venues received = 32
Lyon 45.75917 4.82966
Sending request to Foursquare API... Success!
N of venues total = 198
N of venues received = 100
Lyon 45.75917 4.82966
Sending request to Foursquare API... Success!
N of venues total = 198
N of venues received = 98
Lisbon 38.72639 -9.14949
Sending request to Foursquare API... Success!
N of venues total = 239
N of venues received = 100
Lisbon 38.72639 -9.14949
Sending request to Foursquare API... Success!
N of venues total = 239
N of venues received = 100
Lisbon 38.72639 -9.14949
Sending request to Foursquare API... Success!
N of venues total = 239
N of venues received = 39
Duisburg 51.43148 6.76356
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Duisburg 51.43148 6.76356
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 100
Duisburg 51.43148 6.76356
Sending request to Foursquare API... Success!
N of venues total = 235
N of venues received = 35
Toulouse 43.60579 1.44864
Sending request to Foursquare API... Success!
N of venues total = 67
```

```
N of venues received = 67
Palma de Mallorca 39.57149 2.64694
Sending request to Foursquare API... Success!
N of venues total = 230
N of venues received = 100
Palma de Mallorca 39.57149 2.64694
Sending request to Foursquare API... Success!
N of venues total = 230
N of venues received = 100
Palma de Mallorca 39.57149 2.64694
Sending request to Foursquare API... Success!
N of venues total = 230
N of venues received = 30
Bologna 44.50485 11.34507
Sending request to Foursquare API... Success!
N of venues total = 204
N of venues received = 100
Bologna 44.50485 11.34507
Sending request to Foursquare API... Success!
N of venues total = 204
N of venues received = 100
Bologna 44.50485 11.34507
Sending request to Foursquare API... Success!
N of venues total = 204
N of venues received = 4
Brno 49.19728 16.60368
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Brno 49.19728 16.60368
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 100
Brno 49.19728 16.60368
Sending request to Foursquare API... Success!
N of venues total = 242
N of venues received = 42
Florence 43.78238 11.25502
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 100
Florence 43.78238 11.25502
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 100
Florence 43.78238 11.25502
Sending request to Foursquare API... Success!
N of venues total = 238
N of venues received = 38
Bochum 51.488 7.21399
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Bochum 51.488 7.21399
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 100
Bochum 51.488 7.21399
Sending request to Foursquare API... Success!
N of venues total = 241
N of venues received = 41
```

```
Utrecht 52.08979 5.11415
Sending request to Foursquare API... Success!
N of venues total = 218
N of venues received = 100
Utrecht 52.08979 5.11415
Sending request to Foursquare API... Success!
N of venues total = 218
N of venues received = 100
Utrecht 52.08979 5.11415
Sending request to Foursquare API... Success!
N of venues total = 218
N of venues received = 18
Wuppertal 51.27165 7.19678
Sending request to Foursquare API... Success!
N of venues total = 180
N of venues received = 100
Wuppertal 51.27165 7.19678
Sending request to Foursquare API... Success!
N of venues total = 180
N of venues received = 80
Aarhus 56.15302 10.20487
Sending request to Foursquare API... Success!
N of venues total = 124
N of venues received = 100
Aarhus 56.15302 10.20487
Sending request to Foursquare API... Success!
N of venues total = 124
N of venues received = 24
Bilbao 43.2689 -2.9453
Sending request to Foursquare API... Success!
N of venues total = 84
N of venues received = 84
Malmö 55.5967 13.0011
Sending request to Foursquare API... Success!
N of venues total = 196
N of venues received = 100
Malmö 55.5967 13.0011
Sending request to Foursquare API... Success!
N of venues total = 196
N of venues received = 96
Nice 43.70029 7.27766
Sending request to Foursquare API... Success!
N of venues total = 154
N of venues received = 100
Nice 43.70029 7.27766
Sending request to Foursquare API... Success!
N of venues total = 154
N of venues received = 54
Bielefeld 52.01548 8.53232
Sending request to Foursquare API... Success!
N of venues total = 151
N of venues received = 100
Bielefeld 52.01548 8.53232
Sending request to Foursquare API... Success!
N of venues total = 151
N of venues received = 51
Bonn 50.73243 7.10187
Sending request to Foursquare API... Success!
N of venues total = 213
N of venues received = 100
Bonn 50.73243 7.10187
```

```
Sending request to Foursquare API... Success!
N of venues total = 213
N of venues received = 100
Bonn 50.73243 7.10187
Sending request to Foursquare API... Success!
N of venues total = 213
N of venues received = 13
Bari 41.12588 16.86666
Sending request to Foursquare API... Success!
N of venues total = 101
N of venues received = 100
Bari 41.12588 16.86666
Sending request to Foursquare API... Success!
N of venues total = 101
N of venues received = 1
Münster 51.96302 7.61782
Sending request to Foursquare API... Success!
N of venues total = 65
N of venues received = 65
Karlsruhe 49.01094 8.40846
Sending request to Foursquare API... Success!
N of venues total = 171
N of venues received = 100
Karlsruhe 49.01094 8.40846
Sending request to Foursquare API... Success!
N of venues total = 171
N of venues received = 71
Catania 37.51136 15.06752
Sending request to Foursquare API... Success!
N of venues total = 151
N of venues received = 100
Catania 37.51136 15.06752
Sending request to Foursquare API... Success!
N of venues total = 151
N of venues received = 51
Mannheim 49.48651 8.46679
Sending request to Foursquare API... Success!
N of venues total = 224
N of venues received = 100
Mannheim 49.48651 8.46679
Sending request to Foursquare API... Success!
N of venues total = 224
N of venues received = 100
Mannheim 49.48651 8.46679
Sending request to Foursquare API... Success!
N of venues total = 224
N of venues received = 24
Nantes 47.21812 -1.55306
Sending request to Foursquare API... Success!
N of venues total = 74
N of venues received = 74
```

In [54]:

```python
#df_eur_venues = pd.read_csv('top_eu_venues.csv')
print(df_eur_venues.shape)
df_eur_venues.head(20)
```

(11868, 7)

Out[54]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Ca
0	Berlin	52.51605	13.37691	Brandenburger Tor	52.516247	13.377786	Mor Lar
1	Berlin	52.51605	13.37691	Butter Lindner	52.517879	13.380450	G
2	Berlin	52.51605	13.37691	Dussmann English Bookshop	52.518223	13.389239	Boo
3	Berlin	52.51605	13.37691	Dussmann das KulturKaufhaus	52.518312	13.388708	Boo
4	Berlin	52.51605	13.37691	Philharmonie	52.509975	13.369776	C
5	Berlin	52.51605	13.37691	Freundschaft	52.518294	13.390344	Wi
6	Berlin	52.51605	13.37691	Pierre Boulez Saal	52.515333	13.396218	C
7	Berlin	52.51605	13.37691	Großer Tiergarten	52.514184	13.356886	
8	Berlin	52.51605	13.37691	ride.bln Studio Mitte	52.508761	13.391630	
9	Berlin	52.51605	13.37691	Gendarmenmarkt	52.513570	13.392720	
10	Berlin	52.51605	13.37691	Ischtar-Tor	52.520742	13.397205	
11	Berlin	52.51605	13.37691	Kin-Za	52.524928	13.395808	Cau Res
12	Berlin	52.51605	13.37691	Flamingo Fresh Food Bar	52.519541	13.385425	Sa
13	Berlin	52.51605	13.37691	Konzerthaus Berlin	52.513639	13.391795	C
14	Berlin	52.51605	13.37691	Die Espressonisten	52.507648	13.388159	
15	Berlin	52.51605	13.37691	Elemenza	52.503872	13.374201	
16	Berlin	52.51605	13.37691	BEN RAHIM	52.525168	13.401928	
17	Berlin	52.51605	13.37691	Das Stue	52.509876	13.346076	
18	Berlin	52.51605	13.37691	Rosengarten	52.513972	13.356888	(
19	Berlin	52.51605	13.37691	vabali spa	52.527603	13.360555	

In [55]:

```
# Make sure all cities have been explored successfully
all_cities = set(df_eur_venues.Neighborhood.values)
print(sorted(all_cities))
print(len(all_cities))
```

```
['Aarhus', 'Amsterdam', 'Antwerp', 'Athens', 'Barcelona', 'Bari', 'Berli
n', 'Bielefeld', 'Bilbao', 'Bochum', 'Bologna', 'Bonn', 'Bremen', 'Brno',
'Budapest', 'Catania', 'Cologne', 'Copenhagen', 'Dortmund', 'Dresden', 'Du
isburg', 'Düsseldorf', 'Essen', 'Florence', 'Frankfurt', 'Genoa', 'Gothenb
urg', 'Hamburg', 'Hanover', 'Helsinki', 'Karlsruhe', 'Kraków', 'Lisbon',
'Lyon', 'Madrid', 'Malmö', 'Mannheim', 'Marseille', 'Milan', 'Munich', 'Mü
nster', 'Nantes', 'Naples', 'Nice', 'Nuremberg', 'Palma de Mallorca', 'Par
is', 'Prague', 'Riga', 'Rome', 'Rotterdam', 'Stockholm', 'Stuttgart', 'The
Hague', 'Toulouse', 'Turin', 'Utrecht', 'Valencia', 'Vienna', 'Wuppertal']
60
```

In [56]:

```
# Let's see all categories of the venues
all_categories = set(df_eur_venues['Venue Category'].values)
print(sorted(all_categories))
print(len(all_categories))
```

['Abruzzo Restaurant', 'Afghan Restaurant', 'African Restaurant', 'Agritur
ismo', 'Airfield', 'Airport', 'Airport Lounge', 'Airport Service', 'Airpor
t Terminal', 'American Restaurant', 'Antique Shop', 'Apple Wine Pub', 'Aqu
arium', 'Arcade', 'Argentinian Restaurant', 'Art Gallery', 'Art Museum',
'Arts & Crafts Store', 'Asian Restaurant', 'Athletics & Sports', 'Auditori
um', 'Australian Restaurant', 'Austrian Restaurant', 'Auto Dealership', 'B
BQ Joint', 'Baby Store', 'Bagel Shop', 'Bakery', 'Ballroom', 'Bar', 'Baseb
all Stadium', 'Basketball Court', 'Basketball Stadium', 'Bathing Area', 'B
avarian Restaurant', 'Bay', 'Beach', 'Beach Bar', 'Bed & Breakfast', 'Beer
Bar', 'Beer Garden', 'Beer Store', 'Belgian Restaurant', 'Big Box Store',
'Bike Rental / Bike Share', 'Bike Shop', 'Bike Trail', 'Bistro', 'Board Sh
op', 'Boarding House', 'Boat or Ferry', 'Bookstore', 'Border Crossing', 'B
otanical Garden', 'Bougatsa Shop', 'Boutique', 'Bowling Alley', 'Brasseri
e', 'Bratwurst Joint', 'Brazilian Restaurant', 'Breakfast Spot', 'Brewer
y', 'Bridge', 'Bubble Tea Shop', 'Buffet', 'Building', 'Burger Joint', 'Bu
rrito Place', 'Bus Stop', 'Butcher', 'Café', 'Cajun / Creole Restaurant',
'Camera Store', 'Campground', 'Canal', 'Candy Store', 'Cantonese Restauran
t', 'Capitol Building', 'Caribbean Restaurant', 'Castle', 'Caucasian Resta
urant', 'Cemetery', 'Champagne Bar', 'Cheese Shop', 'Chinese Restaurant',
'Chocolate Shop', 'Church', 'Cigkofte Place', 'Circus', 'Circus School',
'City', 'City Hall', 'Climbing Gym', 'Clothing Store', 'Club House', 'Cock
tail Bar', 'Coffee Roaster', 'Coffee Shop', 'College Arts Building', 'Coll
ege Cafeteria', 'College Library', 'Comedy Club', 'Comfort Food Restauran
t', 'Comic Shop', 'Concert Hall', 'Construction & Landscaping', 'Convenien
ce Store', 'Convention Center', 'Cosmetics Shop', 'Courthouse', 'Coworking
Space', 'Creperie', 'Cretan Restaurant', 'Cruise Ship', 'Cuban Restauran
t', 'Cultural Center', 'Cupcake Shop', 'Currywurst Joint', 'Cycle Studio',
'Czech Restaurant', 'Dairy Store', 'Dam', 'Dance Studio', 'Daycare', 'Deli
/ Bodega', 'Department Store', 'Design Studio', 'Dessert Shop', 'Dim Sum R
estaurant', 'Diner', 'Disc Golf', 'Discount Store', 'Distillery', 'Dive Ba
r', 'Dive Spot', 'Dog Run', 'Doner Restaurant', 'Donut Shop', 'Drugstore',
'Dumpling Restaurant', 'Dutch Restaurant', 'Eastern European Restaurant',
'Electronics Store', 'Emilia Restaurant', 'English Restaurant', 'Escape Ro
om', 'Ethiopian Restaurant', 'Event Space', 'Exhibit', 'Factory', 'Falafel
Restaurant', 'Farm', 'Farmers Market', 'Fast Food Restaurant', 'Field', 'F
ilipino Restaurant', 'Fish & Chips Shop', 'Fish Market', 'Fish Taverna',
'Flea Market', 'Flower Shop', 'Food', 'Food & Drink Shop', 'Food Court',
'Food Service', 'Food Stand', 'Food Truck', 'Football Stadium', 'Forest',
'Fountain', 'Franconian Restaurant', 'Fraternity House', 'French Restauran
t', 'Fried Chicken Joint', 'Friterie', 'Frozen Yogurt Shop', 'Fruit & Vege
table Store', 'Furniture / Home Store', 'Gaming Cafe', 'Garden', 'Garden C
enter', 'Gas Station', 'Gastropub', 'Gay Bar', 'General Entertainment', 'G
erman Restaurant', 'Gift Shop', 'Gluten-free Restaurant', 'Go Kart Track',
'Golf Course', 'Gourmet Shop', 'Greek Restaurant', 'Grilled Meat Restauran
t', 'Grocery Store', 'Gym', 'Gym / Fitness Center', 'Gym Pool', 'Gymnastic
s Gym', 'Harbor / Marina', 'Hardware Store', 'Hawaiian Restaurant', 'Healt
h Food Store', 'Herbs & Spices Store', 'Hill', 'Himalayan Restaurant', 'Hi
storic Site', 'History Museum', 'Hobby Shop', 'Hockey Field', 'Hookah Ba
r', 'Hostel', 'Hot Dog Joint', 'Hotel', 'Hotel Bar', 'Hotel Pool', 'Hungar
ian Restaurant', 'IT Services', 'Ice Cream Shop', 'Imported Food Shop', 'I
ndian Restaurant', 'Indie Movie Theater', 'Indie Theater', 'Indonesian Res
taurant', 'Intersection', 'Irish Pub', 'Island', 'Israeli Restaurant', 'It
alian Restaurant', 'Japanese Restaurant', 'Jazz Club', 'Jewelry Store', 'J
ewish Restaurant', 'Juice Bar', 'Kafenio', 'Karaoke Bar', 'Kebab Restauran
t', 'Kitchen Supply Store', 'Korean Restaurant', 'Lake', 'Laser Tag', 'Lat
in American Restaurant', 'Leather Goods Store', 'Lebanese Restaurant', 'Li
brary', 'Light Rail Station', 'Lighthouse', 'Ligurian Restaurant', 'Linger
ie Store', 'Liquor Store', 'Lombard Restaurant', 'Lounge', 'Lyonese Boucho
n', 'Magirio', 'Malay Restaurant', 'Marijuana Dispensary', 'Market', 'Mart
ial Arts School', 'Massage Studio', 'Mattress Store', 'Mediterranean Resta
urant', "Men's Store", 'Mexican Restaurant', 'Meze Restaurant', 'Middle Ea

stern Restaurant', 'Military Base', 'Mini Golf', 'Miscellaneous Shop', 'Mobile Phone Shop', 'Modern European Restaurant', 'Modern Greek Restaurant', 'Molecular Gastronomy Restaurant', 'Monument / Landmark', 'Moroccan Restaurant', 'Motel', 'Motorcycle Shop', 'Mountain', 'Movie Theater', 'Multiplex', 'Museum', 'Music Store', 'Music Venue', 'National Park', 'Nature Preserve', 'Neighborhood', 'New American Restaurant', 'Night Market', 'Nightclub', 'Non-Profit', 'Noodle House', 'Nudist Beach', 'Office', 'Opera House', 'Optical Shop', 'Organic Grocery', 'Other Great Outdoors', 'Other Nightlife', 'Outdoor Event Space', 'Outdoor Gym', 'Outdoor Sculpture', 'Outdoor Supply Store', 'Outdoors & Recreation', 'Outlet Store', 'Ouzeri', 'Paella Restaurant', 'Paintball Field', 'Pakistani Restaurant', 'Palace', 'Paper / Office Supplies Store', 'Park', 'Pastry Shop', 'Pedestrian Plaza', 'Pelmeni House', 'Performing Arts Venue', 'Perfume Shop', 'Persian Restaurant', 'Peruvian Restaurant', 'Pet Café', 'Pet Store', 'Pharmacy', 'Photography Lab', 'Photography Studio', 'Piadineria', 'Pie Shop', 'Piedmontese Restaurant', 'Pier', 'Pizza Place', 'Planetarium', 'Playground', 'Plaza', 'Poke Place', 'Polish Restaurant', 'Pool', 'Pool Hall', 'Portuguese Restaurant', 'Provençal Restaurant', 'Pub', 'Public Art', 'Racecourse', 'Racetrack', 'Rafting', 'Ramen Restaurant', 'Record Shop', 'Recreation Center', 'Rental Car Location', 'Reservoir', 'Residential Building (Apartment / Condo)', 'Resort', 'Rest Area', 'Restaurant', 'Rhenisch Restaurant', 'River', 'Road', 'Rock Climbing Spot', 'Rock Club', 'Roman Restaurant', 'Roof Deck', 'Rooftop Bar', 'Rugby Stadium', 'Salad Place', 'Salon / Barbershop', 'Sandwich Place', 'Sauna / Steam Room', 'Scandinavian Restaurant', 'Scenic Lookout', 'Schnitzel Restaurant', 'Science Museum', 'Sculpture Garden', 'Seafood Restaurant', 'Shoe Store', 'Shopping Mall', 'Shopping Plaza', 'Sicilian Restaurant', 'Skate Park', 'Skating Rink', 'Ski Area', 'Ski Chairlift', 'Ski Trail', 'Slovak Restaurant', 'Smoke Shop', 'Snack Place', 'Soccer Field', 'Soccer Stadium', 'Soup Place', 'South American Restaurant', 'South Indian Restaurant', 'Southern / Soul Food Restaurant', 'Souvlaki Shop', 'Spa', 'Spanish Restaurant', 'Speakeasy', 'Sporting Goods Shop', 'Sports Bar', 'Sports Club', 'Stables', 'Stadium', 'State / Provincial Park', 'Steakhouse', 'Street Art', 'Street Food Gathering', 'Student Center', 'Supermarket', 'Surf Spot', 'Sushi Restaurant', 'Swabian Restaurant', 'Syrian Restaurant', 'Szechuan Restaurant', 'Taco Place', 'Tailor Shop', 'Tapas Restaurant', 'Taverna', 'Tea Room', 'Temple', 'Tennis Court', 'Tennis Stadium', 'Thai Restaurant', 'Theater', 'Theme Park', 'Theme Park Ride / Attraction', 'Theme Restaurant', 'Thrift / Vintage Store', 'Tibetan Restaurant', 'Town Hall', 'Toy / Game Store', 'Track', 'Track Stadium', 'Trail', 'Trailer Park', 'Train Station', 'Tram Station', 'Transportation Service', 'Trattoria/Osteria', 'Tunnel', 'Turkish Restaurant', 'Udon Restaurant', 'Vegetarian / Vegan Restaurant', 'Venezuelan Restaurant', 'Vietnamese Restaurant', 'Village', 'Vineyard', 'Volleyball Court', 'Warehouse Store', 'Water Park', 'Waterfall', 'Waterfront', 'Whisky Bar', 'Windmill', 'Wine Bar', 'Wine Shop', 'Winery', "Women's Store", 'Yoga Studio', 'Zoo', 'Zoo Exhibit']
453

In [57]:

```
# Let's see how many venues have been discovered in each city
df_eur_venues.groupby('Neighborhood').count()
```

Out[57]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Aarhus	124	124	124	124	124	124
Amsterdam	242	242	242	242	242	242
Antwerp	232	232	232	232	232	232
Athens	232	232	232	232	232	232
Barcelona	241	241	241	241	241	241
Bari	101	101	101	101	101	101
Berlin	235	235	235	235	235	235
Bielefeld	151	151	151	151	151	151
Bilbao	84	84	84	84	84	84
Bochum	241	241	241	241	241	241
Bologna	204	204	204	204	204	204
Bonn	213	213	213	213	213	213
Bremen	189	189	189	189	189	189
Brno	242	242	242	242	242	242
Budapest	242	242	242	242	242	242
Catania	151	151	151	151	151	151
Cologne	200	200	200	200	200	200
Copenhagen	242	242	242	242	242	242
Dortmund	223	223	223	223	223	223
Dresden	99	99	99	99	99	99
Duisburg	235	235	235	235	235	235
Düsseldorf	240	240	240	240	240	240
Essen	242	242	242	242	242	242
Florence	238	238	238	238	238	238
Frankfurt	235	235	235	235	235	235
Genoa	83	83	83	83	83	83
Gothenburg	155	155	155	155	155	155
Hamburg	234	234	234	234	234	234
Hanover	168	168	168	168	168	168
Helsinki	241	241	241	241	241	241
Karlsruhe	171	171	171	171	171	171
Kraków	183	183	183	183	183	183
Lisbon	239	239	239	239	239	239
Lyon	198	198	198	198	198	198
Madrid	229	229	229	229	229	229

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Malmö	196	196	196	196	196	196
Mannheim	224	224	224	224	224	224
Marseille	172	172	172	172	172	172
Milan	234	234	234	234	234	234
Munich	243	243	243	243	243	243
Münster	65	65	65	65	65	65
Nantes	74	74	74	74	74	74
Naples	214	214	214	214	214	214
Nice	154	154	154	154	154	154
Nuremberg	232	232	232	232	232	232
Palma de Mallorca	230	230	230	230	230	230
Paris	205	205	205	205	205	205
Prague	242	242	242	242	242	242
Riga	226	226	226	226	226	226
Rome	238	238	238	238	238	238
Rotterdam	246	246	246	246	246	246
Stockholm	242	242	242	242	242	242
Stuttgart	171	171	171	171	171	171
The Hague	242	242	242	242	242	242
Toulouse	67	67	67	67	67	67
Turin	193	193	193	193	193	193
Utrecht	218	218	218	218	218	218
Valencia	140	140	140	140	140	140
Vienna	241	241	241	241	241	241
Wuppertal	180	180	180	180	180	180

In [58]:

```
df_eur_venues.to_csv('top_eu_venues.csv')
df_eur_venues.groupby('Neighborhood').count().to_csv('top_eu_venues_grouped.csv')
```

In [59]:

```
print(f"We have found {len(df_eur_venues['Venue Category'].unique())} unique categories
of venues.")
print(f"We have {len(df_eur_venues['Neighborhood'].unique())} neighborhoods (cities) to
analyse.")
print(f"We have {df_eur_venues.shape[0]} records (venues) in our dataset.")
```

```
We have found 453 unique categories of venues.
We have 60 neighborhoods (cities) to analyse.
We have 11868 records (venues) in our dataset.
```

5.1. Prepare for clustering the cities: get data for our hometown(s)

In the following parts of the project, we will apply k-means clustering algorithm, which is a well-known unsupervised ML method, to analyze the selected EU cities and group them into several partitions by certain features:

- The objects to study are the cities.
- The set of features for each city will be derived from the distribution of venues among categories for that particular city. The weight of i-th category in the overall number of venues for the city will correspond to the i-th feature of the city.
- The cities will be grouped by resemblance of how their venues are distributed among categories

Also, we want to compare the selected EU cities to our clients' hometowns:

- our clients currently reside in Novosibirsk, Western Siberia, Russia
- previously they used to live in Irkutsk, Eastern Siberia, Russia

In [60]:

```
# Define hometowns
my_cities = ['Novosibirsk', 'Irkutsk']
my_locations = [get_location(my_cities[0]), get_location(my_cities[1])]
print(my_locations)
```

```
Новосибирск, Сибирский федеральный округ, Россия, Новосибирск, Сибирский ф
едеральный округ 630132, RUS = 55.03977, 82.91017
Иркутск, Сибирский федеральный округ, Россия, Иркутск, Сибирский федеральн
ый округ 664005, RUS = 52.30026, 104.24686
[(55.03977, 82.91017), (52.30026, 104.24686)]
```

In [61]:

```
# Find all venues in our clients' hometowns
df_my_venues = get_nearby_venues(
    my_cities,
    [my_locations[0][0], my_locations[1][0]],
    [my_locations[0][1], my_locations[1][1]],
    radius=20_000)
```

```
Novosibirsk 55.03977 82.91017
Sending request to Foursquare API... Success!
N of venues total = 95
N of venues received = 95
Irkutsk 52.30026 104.24686
Sending request to Foursquare API... Success!
N of venues total = 50
N of venues received = 50
```

In [62]:

```
print(df_my_venues.shape)
df_my_venues.head()
```

```
(145, 7)
```

Out[62]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venu Categor
0	Novosibirsk	55.03977	82.91017	Академия Кофе	55.031400	82.914292	Gamin Caf
1	Novosibirsk	55.03977	82.91017	HookahPlace	55.033206	82.924697	Hooka Ba
2	Novosibirsk	55.03977	82.91017	Papa Carlo	55.048588	82.914456	Pizz Plac
3	Novosibirsk	55.03977	82.91017	Blackwood Coffee Roastery	55.049253	82.915482	Coffe Sho
4	Novosibirsk	55.03977	82.91017	Tom Yum Bar Gray	55.029187	82.910583	Thi Restaurai

In [63]:

```
df_my_venues.to_csv('top_my_venues.csv')
```

In [64]:

```
print(f"We have found {len(df_my_venues['Venue Category'].unique())} unique categories
 of venues in hometowns.")
print(f"We have {len(df_my_venues['Neighborhood'].unique())} neighborhoods (cities) in
 hometowns.")
print(f"We have {df_my_venues.shape[0]} records (venues) in hometowns.")
```

```
We have found 78 unique categories of venues in hometowns.
We have 2 neighborhoods (cities) in hometowns.
We have 145 records (venues) in hometowns.
```

5.2. Prepare for clustering the cities: dimensionality reduction

Since we have selected a substantial number of cities to analyze and several hundreds of features (venue categories), it would be helpful to reduce dimensionality of our problem. We can eliminate redundant features (categories), so the computational complexity of the task will decrease

In [65]:

```
# Do not consider categories irrelevant for residents of a city, they're interesting fo
r visitors only
cats_not_relevant = {'Hotel', 'Hostel', 'Hotel Bar'}
```

In [66]:

```
# Categories for 1st hometown
my_cats0 = set(df_my_venues[df_my_venues['Neighborhood'] == my_cities[0]]['Venue Catego
ry'])
my_cats0 -= cats_not_relevant
print(sorted(my_cats0))
print(len(my_cats0))
```

```
['Airport Lounge', 'Airport Service', 'Arcade', 'Asian Restaurant', 'BBQ J
oint', 'Bank', 'Bar', 'Bath House', 'Beer Bar', 'Beer Store', 'Big Box Sto
re', 'Brewery', 'Burger Joint', 'Café', 'Cocktail Bar', 'Coffee Shop', 'Co
smetics Shop', 'Deli / Bodega', 'Department Store', 'Dessert Shop', 'Dumpl
ing Restaurant', 'Duty-free Shop', 'Eastern European Restaurant', 'Fast Fo
od Restaurant', 'Flower Shop', 'Food & Drink Shop', 'Gaming Cafe', 'Grocer
y Store', 'Gym / Fitness Center', 'Health Food Store', 'Hookah Bar', 'Ital
ian Restaurant', 'Middle Eastern Restaurant', 'Movie Theater', 'Music Stor
e', 'Opera House', 'Park', 'Perfume Shop', 'Pizza Place', 'Pool', 'Pub',
'Restaurant', 'Russian Restaurant', 'Sculpture Garden', 'Skating Rink', 'S
nack Place', 'Soccer Field', 'Spa', 'Steakhouse', 'Tapas Restaurant', 'Tha
i Restaurant', 'Theater', 'Wine Bar', 'Zoo']
54
```

In [67]:

```
# Categories for 2nd hometown
my_cats1 = set(df_my_venues[df_my_venues['Neighborhood'] == my_cities[1]]['Venue Catego
ry'])
my_cats1 -= cats_not_relevant
print(sorted(my_cats1))
print(len(my_cats1))
```

```
['Accessories Store', 'Art Gallery', 'Australian Restaurant', 'Beer Stor
e', 'Bookstore', 'Café', 'Cocktail Bar', 'Coffee Shop', 'Cosmetics Shop',
'Cupcake Shop', 'Dessert Shop', 'Farm', 'Food Truck', 'Garden Center', 'Ga
stropub', 'Gym / Fitness Center', 'History Museum', 'Hookah Bar', 'Islan
d', 'Italian Restaurant', 'Karaoke Bar', 'Lingerie Store', 'Mountain', 'Pa
rk', 'Pedestrian Plaza', 'Photography Studio', 'Pub', 'Rest Area', 'Scenic
Lookout', 'Ski Area', 'Soccer Stadium', 'Steakhouse', 'Theater', 'Train St
ation', 'Wine Bar']
35
```

In [68]:

```
# Let's see which categories are represented both in selected EU cities and in hometown
s
cats0 = sorted(my_cats0 & all_categories)
cats1 = sorted(my_cats1 & all_categories)
print(len(cats0), cats0)
print(len(cats1), cats1)
```

```
50 ['Airport Lounge', 'Airport Service', 'Arcade', 'Asian Restaurant', 'BB
Q Joint', 'Bar', 'Beer Bar', 'Beer Store', 'Big Box Store', 'Brewery', 'Bu
rger Joint', 'Café', 'Cocktail Bar', 'Coffee Shop', 'Cosmetics Shop', 'Del
i / Bodega', 'Department Store', 'Dessert Shop', 'Dumpling Restaurant', 'E
astern European Restaurant', 'Fast Food Restaurant', 'Flower Shop', 'Food
& Drink Shop', 'Gaming Cafe', 'Grocery Store', 'Gym / Fitness Center', 'He
alth Food Store', 'Hookah Bar', 'Italian Restaurant', 'Middle Eastern Rest
aurant', 'Movie Theater', 'Music Store', 'Opera House', 'Park', 'Perfume S
hop', 'Pizza Place', 'Pool', 'Pub', 'Restaurant', 'Sculpture Garden', 'Ska
ting Rink', 'Snack Place', 'Soccer Field', 'Spa', 'Steakhouse', 'Tapas Res
taurant', 'Thai Restaurant', 'Theater', 'Wine Bar', 'Zoo']
34 ['Art Gallery', 'Australian Restaurant', 'Beer Store', 'Bookstore', 'Ca
fé', 'Cocktail Bar', 'Coffee Shop', 'Cosmetics Shop', 'Cupcake Shop', 'Des
sert Shop', 'Farm', 'Food Truck', 'Garden Center', 'Gastropub', 'Gym / Fit
ness Center', 'History Museum', 'Hookah Bar', 'Island', 'Italian Restauran
t', 'Karaoke Bar', 'Lingerie Store', 'Mountain', 'Park', 'Pedestrian Plaz
a', 'Photography Studio', 'Pub', 'Rest Area', 'Scenic Lookout', 'Ski Are
a', 'Soccer Stadium', 'Steakhouse', 'Theater', 'Train Station', 'Wine Ba
r']
```

In [69]:

```
# Combine venues of EU cities and hometowns into one dataset for clustering
df_eur_venues = df_eur_venues.append(df_my_venues, ignore_index=True)
print(df_eur_venues.shape)
df_eur_venues.head()
```

```
(12013, 7)
```

Out[69]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Ver Catego
0	Berlin	52.51605	13.37691	Brandenburger Tor	52.516247	13.377786	Monum Landm
1	Berlin	52.51605	13.37691	Butter Lindner	52.517879	13.380450	Gourr Sh
2	Berlin	52.51605	13.37691	Dussmann English Bookshop	52.518223	13.389239	Bookst
3	Berlin	52.51605	13.37691	Dussmann das KulturKaufhaus	52.518312	13.388708	Bookst
4	Berlin	52.51605	13.37691	Philharmonie	52.509975	13.369776	Conc H

In [70]:

```
df_eur_venues.tail(100)
```

Out[70]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	C
11913	Novosibirsk	55.03977	82.91017	Чучвара	55.031730	82.914331	Re
11914	Novosibirsk	55.03977	82.91017	Сыроварня	55.030252	82.904885	Re
11915	Novosibirsk	55.03977	82.91017	JONATHAN Homemade Food & Beer	55.022428	82.923391	
11916	Novosibirsk	55.03977	82.91017	iBeauty	54.993425	82.893478	Co
11917	Novosibirsk	55.03977	82.91017	Бассейн СГУПС	55.067343	82.925641	
...	...	...	...	...	...	...	
12008	Irkutsk	52.30026	104.24686	Хурал	52.200695	104.070386	
12009	Irkutsk	52.30026	104.24686	Ст. Олха	52.157705	104.107754	
12010	Irkutsk	52.30026	104.24686	Горнолыжная база Олха	52.157167	104.100037	
12011	Irkutsk	52.30026	104.24686	Мельничный тракт	52.133957	104.326125	R
12012	Irkutsk	52.30026	104.24686	Аллея	52.427774	104.043546	

100 rows × 7 columns

In [71]:

```
# First, we'll filter dataset by categories
df_eur_venues.set_index('Venue Category', inplace=True)
df_eur_venues.head()
```

Out[71]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Venue Category						
Monument / Landmark	Berlin	52.51605	13.37691	Brandenburger Tor	52.516247	13.377786
Gourmet Shop	Berlin	52.51605	13.37691	Butter Lindner	52.517879	13.380450
Bookstore	Berlin	52.51605	13.37691	Dussmann English Bookshop	52.518223	13.389239
Bookstore	Berlin	52.51605	13.37691	Dussmann das KulturKaufhaus	52.518312	13.388708
Concert Hall	Berlin	52.51605	13.37691	Philharmonie	52.509975	13.369776

In [72]:

```
# Subset categories relevant for 1st hometown
df_venues0 = df_eur_venues.loc[cats0, ]
print(df_venues0.shape)
df_venues0.head(30)
```

(4533, 6)

Out[72]:

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Airport Lounge	Amsterdam	52.36994	4.90788	Privium ClubLounge	52.309003	4.765398
Airport Lounge	Frankfurt	50.11208	8.68342	Lufthansa First Class Terminal	50.049840	8.564382
Airport Lounge	Frankfurt	50.11208	8.68342	Lufthansa First Class Lounge B	50.047246	8.572217
Airport Lounge	Gothenburg	57.70068	11.96823	SAS Lounge	57.668075	12.293854
Airport Lounge	Gothenburg	57.70068	11.96823	Vinga Lounge by Menzies Aviation	57.668057	12.293839
Airport Lounge	Lyon	45.75917	4.82966	Montblanc Lounge	45.717086	5.078384
Airport Lounge	Lyon	45.75917	4.82966	Salon Air France	45.722105	5.081130
Airport Lounge	Lyon	45.75917	4.82966	Salon Confluence	45.716308	5.078130
Airport Lounge	Novosibirsk	55.03977	82.91017	S7 Business Lounge	55.009708	82.666600
Airport Service	Amsterdam	52.36994	4.90788	Sky Priority Check-In	52.309406	4.763865
Airport Service	Gothenburg	57.70068	11.96823	SAS Check In	57.667941	12.294907
Airport Service	Lyon	45.75917	4.82966	Security Check	45.717549	5.076735
Airport Service	Novosibirsk	55.03977	82.91017	Взлётно-посадочная полоса	55.010181	82.667421
Airport Service	Novosibirsk	55.03977	82.91017	Паспортный контроль / Passport Control	55.009589	82.671067
Airport Service	Novosibirsk	55.03977	82.91017	Зона досмотра пассажиров / Security Control (3...	55.009491	82.667545
Arcade	Munich	48.13642	11.57755	Chaos Computer Club	48.153618	11.560834
Arcade	Prague	50.07913	14.43303	ArcadeHry	50.073157	14.164236
Arcade	Kraków	50.06045	19.93243	Kraków Pinball Museum	50.052748	19.939833
Arcade	Bologna	44.50485	11.34507	Piscina Junior	44.416583	11.349241

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Arcade	Novosibirsk	55.03977	82.91017	Кёрлинг клуб Пингвин	54.999850	82.750312
Asian Restaurant	Rome	41.90323	12.49566	Thien Kim Ristorante Vietnamita	41.971203	12.433639
Asian Restaurant	Rome	41.90323	12.49566	Ristorante Yu Olgiata	42.020812	12.375398
Asian Restaurant	Rome	41.90323	12.49566	Xin Yi	41.778177	12.356236
Asian Restaurant	Paris	48.85718	2.34141	Bouddha Wok	48.826891	2.528196
Asian Restaurant	Vienna	48.20263	16.36843	BAO BAR	48.199302	16.351109
Asian Restaurant	Vienna	48.20263	16.36843	Sha Guo	48.194219	16.367629
Asian Restaurant	Vienna	48.20263	16.36843	Coconut Curry	48.216729	16.388234
Asian Restaurant	Vienna	48.20263	16.36843	IKI	48.186870	16.380092
Asian Restaurant	Vienna	48.20263	16.36843	L421	48.202425	16.255471
Asian Restaurant	Vienna	48.20263	16.36843	Sajado	48.342933	16.462590

In [73]:

```
# Subset categories relevant for 2nd hometown
df_venues1 = df_eur_venues.loc[cats1, ]
print(df_venues1.shape)
df_venues1.head(30)
```

(3300, 6)

Out[73]:

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Art Gallery	Berlin	52.51605	13.37691	Urban Nation	52.498676	13.356538
Art Gallery	Berlin	52.51605	13.37691	Liebermann-Villa am Wannsee	52.428896	13.164713
Art Gallery	Madrid	40.41956	-3.69196	Fundación Mapfre Recoletos	40.422450	-3.692151
Art Gallery	Madrid	40.41956	-3.69196	Tabacalera Promoción del Arte	40.406386	-3.703242
Art Gallery	Madrid	40.41956	-3.69196	Matadero Madrid	40.392270	-3.697500
Art Gallery	Madrid	40.41956	-3.69196	CA2M Centro de Arte 2 de Mayo	40.324614	-3.863249
Art Gallery	Milan	45.46796	9.18178	Palazzo Reale	45.462960	9.191348
Art Gallery	Milan	45.46796	9.18178	Pirelli Hangar Bicocca	45.520988	9.219257
Art Gallery	Prague	50.07913	14.43303	Pelléova Villa	50.099125	14.407441
Art Gallery	Stockholm	59.33258	18.06683	Artipelag	59.306431	18.346868
Art Gallery	Copenhagen	55.67567	12.56756	Cisternerne	55.669541	12.524161
Art Gallery	Valencia	39.46895	-0.37686	La Fábrica de Hielo	39.469638	-0.325085
Art Gallery	Kraków	50.06045	19.93243	MNK Kamienica Szołayskich	50.063440	19.935739
Art Gallery	Kraków	50.06045	19.93243	Cricoteka	50.047299	19.951285
Art Gallery	Athens	37.97614	23.73640	Lighthouse (SNFCC)	37.939493	23.691072
Art Gallery	Helsinki	60.17116	24.93266	Taidehalli	60.172127	24.931014
Art Gallery	Riga	56.94599	24.11487	Imanta tattoo	56.950765	24.125620
Art Gallery	Riga	56.94599	24.11487	Noass	56.945404	24.095235
Art Gallery	Essen	51.45183	7.01109	Ludwig Galerie	51.492101	6.860334
Art Gallery	Gothenburg	57.70068	11.96823	Röda sten	57.689290	11.901693
Art Gallery	Dresden	51.05364	13.74082	Kunsthofpassage	51.067898	13.754246
Art Gallery	Lyon	45.75917	4.82966	La Sucrière	45.736892	4.815079

Venue Category	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Art Gallery	Lyon	45.75917	4.82966	La Demeure du Chaos	45.837414	4.82668
Art Gallery	Lisbon	38.72639	-9.14949	ZDB - Galeria Zé dos Bois	38.711796	-9.14457
Art Gallery	Lisbon	38.72639	-9.14949	Centro Cultural Palácio do Egipto	38.691889	-9.31126
Art Gallery	Duisburg	51.43148	6.76356	Ludwig Galerie	51.492101	6.86033
Art Gallery	Duisburg	51.43148	6.76356	Schloss Oberhausen	51.492316	6.86113
Art Gallery	Palma de Mallorca	39.57149	2.64694	Fundació Pilar i Joan Miró	39.554874	2.60998
Art Gallery	Utrecht	52.08979	5.11415	Metaalkathedraal	52.081961	5.06713
Art Gallery	Malmö	55.59670	13.00110	Malmö Konsthall	55.595286	12.99878

5.3. Prepare for clustering the cities: split the dataset and the problem into independent parts

Since our two hometowns are quite different in terms of population, we decide to split the problem into separate parts: we will analyze and cluster large and medium cities independently.

In [74]:

```
# Combine EU cities and hometowns into one dataframe
df_cities = df_cities.append(
    pd.DataFrame(
        [[my_cities[0], 'Russia', 1_620_000, *my_locations[0], 5],
        [my_cities[1], 'Russia', 617_000, *my_locations[1], 4]],
        columns=list(df_cities.columns)),
    ignore_index=True)
df_cities
```

Out[74]:

	City	Country	Population	Lat	Lon	HasUnivs
0	Berlin	Germany	3669495	52.51605	13.37691	5
1	Madrid	Spain	3348536	40.41956	-3.69196	3
2	Rome	Italy	2856133	41.90323	12.49566	2
3	Paris	France	2140526	48.85718	2.34141	14
4	Vienna	Austria	1921153	48.20263	16.36843	3
...	...	...	...	...	...	...
57	Catania	Italy	311584	37.51136	15.06752	1
58	Mannheim	Germany	309370	49.48651	8.46679	2
59	Nantes	France	306694	47.21812	-1.55306	3
60	Novosibirsk	Russia	1620000	55.03977	82.91017	5
61	Irkutsk	Russia	617000	52.30026	104.24686	4

62 rows × 6 columns

Now, we will define the 1st group for clustering:

Large cities = 1st hometown + EU cities with population close to 1st hometown's

In [76]:

```
# Filter by city
df_venues0.reset_index(inplace=True)
df_venues0.set_index('Neighborhood', inplace=True)
print(df_venues0.shape)
df_venues0.head()
```

```
(4533, 6)
```

Out[76]:

	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Neighborhood						
Amsterdam	Airport Lounge	52.36994	4.90788	Privium ClubLounge	52.309003	4.765398
Frankfurt	Airport Lounge	50.11208	8.68342	Lufthansa First Class Terminal	50.049840	8.564382
Frankfurt	Airport Lounge	50.11208	8.68342	Lufthansa First Class Lounge B	50.047246	8.572217
Gothenburg	Airport Lounge	57.70068	11.96823	SAS Lounge	57.668075	12.293854
Gothenburg	Airport Lounge	57.70068	11.96823	Vinga Lounge by Menzies Aviation	57.668057	12.293839

In [77]:

```
# Filter by large cities
df_venues0 = df_venues0.loc[df_cities[df_cities.Population >= 800_000].City.values, ].r
eset_index()
print(df_venues0.shape)
df_venues0.head()
```

(1557, 7)

Out[77]:

	Neighborhood	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Berlin	BBQ Joint	52.51605	13.37691	Das märchenhafte Ribhouse	52.621963	13.48915
1	Berlin	Bar	52.51605	13.37691	Etc:Bar	52.534070	13.41966
2	Berlin	Bar	52.51605	13.37691	Lerchen und Eulen	52.502123	13.43074
3	Berlin	Bar	52.51605	13.37691	Rabu	52.454803	13.62803
4	Berlin	Beer Bar	52.51605	13.37691	BRÄUGIER BrewPub	52.542153	13.42385

In [78]:

```
df_venues0.to_csv('top_venues0.csv')
```

In [79]:

```
# Let's see which cities are in the 'Large' group
cities0 = sorted(df_venues0.Neighborhood.unique())
print(len(cities0), cities0)
print(df_venues0.groupby('Neighborhood').count().Venue.sum())
df_venues0.groupby('Neighborhood').count()
```

```
18 ['Amsterdam', 'Barcelona', 'Berlin', 'Budapest', 'Cologne', 'Hamburg',
'Madrid', 'Marseille', 'Milan', 'Munich', 'Naples', 'Novosibirsk', 'Pari
s', 'Prague', 'Rome', 'Stockholm', 'Turin', 'Vienna']
1557
```

Out[79]:

	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Neighborhood						
Amsterdam	105	105	105	105	105	105
Barcelona	102	102	102	102	102	102
Berlin	76	76	76	76	76	76
Budapest	104	104	104	104	104	104
Cologne	69	69	69	69	69	69
Hamburg	81	81	81	81	81	81
Madrid	94	94	94	94	94	94
Marseille	42	42	42	42	42	42
Milan	95	95	95	95	95	95
Munich	81	81	81	81	81	81
Naples	111	111	111	111	111	111
Novosibirsk	85	85	85	85	85	85
Paris	59	59	59	59	59	59
Prague	108	108	108	108	108	108
Rome	99	99	99	99	99	99
Stockholm	84	84	84	84	84	84
Turin	77	77	77	77	77	77
Vienna	85	85	85	85	85	85

Now, we will define the 2nd group for clustering:

Medium cities = 2nd hometown + EU cities with population close to 2nd hometown's

In [80]:

```
# Filter by city
df_venues1.reset_index(inplace=True)
df_venues1.set_index('Neighborhood', inplace=True)
print(df_venues1.shape)
df_venues1.head()
```

(3300, 6)

Out[80]:

	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Neighborhood						
Berlin	Art Gallery	52.51605	13.37691	Urban Nation	52.498676	13.356538
Berlin	Art Gallery	52.51605	13.37691	Liebermann-Villa am Wannsee	52.428896	13.164713
Madrid	Art Gallery	40.41956	-3.69196	Fundación Mapfre Recoletos	40.422450	-3.692151
Madrid	Art Gallery	40.41956	-3.69196	Tabacalera Promoción del Arte	40.406386	-3.703242
Madrid	Art Gallery	40.41956	-3.69196	Matadero Madrid	40.392270	-3.697500

In [81]:

```
# Filter by medium cities
df_venues1 = df_venues1.loc[df_cities[df_cities.Population < 1_000_000].City.values, ].
reset_index()
print(df_venues1.shape)
df_venues1.head()
```

(2440, 7)

Out[81]:

	Neighborhood	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Ve Longi
0	Stockholm	Art Gallery	59.33258	18.06683	Artipelag	59.306431	18.34
1	Stockholm	Bookstore	59.33258	18.06683	Science Fiction Bokhandeln	59.324047	18.07
2	Stockholm	Bookstore	59.33258	18.06683	Papercut	59.317183	18.05
3	Stockholm	Bookstore	59.33258	18.06683	Söderbokhandeln	59.316034	18.07
4	Stockholm	Café	59.33258	18.06683	Café Pascal	59.342019	18.05

In [82]:

```
df_venues1.to_csv('top_venues1.csv')
```

In [83]:

```
# Let's see which cities are in the 'Medium' group
cities1 = sorted(df_venues1.Neighborhood.unique())
print(len(cities1), cities1)
print(df_venues1.groupby('Neighborhood').count().Venue.sum())
df_venues1.groupby('Neighborhood').count()
```

```
49 ['Aarhus', 'Amsterdam', 'Antwerp', 'Athens', 'Bari', 'Bielefeld', 'Bilb
ao', 'Bochum', 'Bologna', 'Bonn', 'Bremen', 'Brno', 'Catania', 'Copenhage
n', 'Dortmund', 'Dresden', 'Duisburg', 'Düsseldorf', 'Essen', 'Florence',
'Frankfurt', 'Genoa', 'Gothenburg', 'Hanover', 'Helsinki', 'Irkutsk', 'Kar
lsruhe', 'Kraków', 'Lisbon', 'Lyon', 'Malmö', 'Mannheim', 'Marseille', 'Mü
nster', 'Nantes', 'Naples', 'Nice', 'Nuremberg', 'Palma de Mallorca', 'Rig
a', 'Rotterdam', 'Stockholm', 'Stuttgart', 'The Hague', 'Toulouse', 'Turi
n', 'Utrecht', 'Valencia', 'Wuppertal']
2440
```

Out[83]:

Neighborhood	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Aarhus	29	29	29	29	29	29
Amsterdam	72	72	72	72	72	72
Antwerp	64	64	64	64	64	64
Athens	86	86	86	86	86	86
Bari	42	42	42	42	42	42
Bielefeld	24	24	24	24	24	24
Bilbao	18	18	18	18	18	18
Bochum	85	85	85	85	85	85
Bologna	92	92	92	92	92	92
Bonn	33	33	33	33	33	33
Bremen	30	30	30	30	30	30
Brno	92	92	92	92	92	92
Catania	50	50	50	50	50	50
Copenhagen	71	71	71	71	71	71
Dortmund	51	51	51	51	51	51
Dresden	17	17	17	17	17	17
Duisburg	58	58	58	58	58	58
Düsseldorf	62	62	62	62	62	62
Essen	71	71	71	71	71	71
Florence	88	88	88	88	88	88
Frankfurt	77	77	77	77	77	77
Genoa	19	19	19	19	19	19
Gothenburg	28	28	28	28	28	28
Hanover	37	37	37	37	37	37
Helsinki	70	70	70	70	70	70
Irkutsk	46	46	46	46	46	46
Karlsruhe	27	27	27	27	27	27
Kraków	57	57	57	57	57	57
Lisbon	58	58	58	58	58	58
Lyon	61	61	61	61	61	61
Malmö	56	56	56	56	56	56
Mannheim	46	46	46	46	46	46
Marseille	39	39	39	39	39	39
Münster	21	21	21	21	21	21
Nantes	11	11	11	11	11	11

	Venue Category	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Neighborhood						
Naples	79	79	79	79	79	79
Nice	28	28	28	28	28	28
Nuremberg	54	54	54	54	54	54
Palma de Mallorca	42	42	42	42	42	42
Riga	75	75	75	75	75	75
Rotterdam	56	56	56	56	56	56
Stockholm	66	66	66	66	66	66
Stuttgart	36	36	36	36	36	36
The Hague	49	49	49	49	49	49
Toulouse	23	23	23	23	23	23
Turin	52	52	52	52	52	52
Utrecht	38	38	38	38	38	38
Valencia	20	20	20	20	20	20
Wuppertal	34	34	34	34	34	34

5.3. Prepare for clustering the cities: transform the dataframe

In [84]:

```
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

In [85]:

```
# Encode features for large cities
venues0_onehot = pd.get_dummies(df_venues0[['Venue Category']], prefix="", prefix_sep=
"")
print(venues0_onehot.shape)
venues0_onehot.head(10)
```

```
(1557, 50)
```

Out[85]:

	Airport Lounge	Airport Service	Arcade	Asian Restaurant	BBQ Joint	Bar	Beer Bar	Beer Store	Big Box Store	Brewery	Burger Joint	C
0	0	0	0	0	1	0	0	0	0	0	0	
1	0	0	0	0	0	1	0	0	0	0	0	
2	0	0	0	0	0	1	0	0	0	0	0	
3	0	0	0	0	0	1	0	0	0	0	0	
4	0	0	0	0	0	0	1	0	0	0	0	
5	0	0	0	0	0	0	1	0	0	0	0	
6	0	0	0	0	0	0	0	1	0	0	0	
7	0	0	0	0	0	0	0	0	1	0	0	
8	0	0	0	0	0	0	0	0	1	0	0	
9	0	0	0	0	0	0	0	0	1	0	0	

In [86]:

```
# Encode features for medium cities
venues1_onehot = pd.get_dummies(df_venues1[['Venue Category']], prefix="", prefix_sep=
"")
print(venues1_onehot.shape)
venues1_onehot.head(10)
```

(2440, 34)

Out[86]:

	Art Gallery	Australian Restaurant	Beer Store	Bookstore	Café	Cocktail Bar	Coffee Shop	Cosmetics Shop	Cupcake Shop	Dess Sh
0	1	0	0	0	0	0	0	0	0	
1	0	0	0	1	0	0	0	0	0	
2	0	0	0	1	0	0	0	0	0	
3	0	0	0	1	0	0	0	0	0	
4	0	0	0	0	1	0	0	0	0	
5	0	0	0	0	1	0	0	0	0	
6	0	0	0	0	1	0	0	0	0	
7	0	0	0	0	1	0	0	0	0	
8	0	0	0	0	1	0	0	0	0	
9	0	0	0	0	1	0	0	0	0	

In [87]:

```
# Insert column 'Neighborhood' into both encoded dataframes
venues0_onehot = pd.concat([df_venues0.Neighborhood, venues0_onehot], axis=1)
venues1_onehot = pd.concat([df_venues1.Neighborhood, venues1_onehot], axis=1)
# See the dimensions for both dataframes (N of venues * (1 + N of categories))
print(venues0_onehot.shape, venues1_onehot.shape)
```

(1557, 51) (2440, 35)

In [89]:

```
venues0_onehot.head()
```

Out[89]:

	Neighborhood	Airport Lounge	Airport Service	Arcade	Asian Restaurant	BBQ Joint	Bar	Beer Bar	Beer Store	Big Box Store	Brev
0	Berlin	0	0	0	0	1	0	0	0	0	
1	Berlin	0	0	0	0	0	1	0	0	0	
2	Berlin	0	0	0	0	0	1	0	0	0	
3	Berlin	0	0	0	0	0	1	0	0	0	
4	Berlin	0	0	0	0	0	0	1	0	0	

In [90]:

```
venues1_onehot.head()
```

Out[90]:

	Neighborhood	Art Gallery	Australian Restaurant	Beer Store	Bookstore	Café	Cocktail Bar	Coffee Shop	Cosmetics Shop
0	Stockholm	1	0	0	0	0	0	0	0
1	Stockholm	0	0	0	1	0	0	0	0
2	Stockholm	0	0	0	1	0	0	0	0
3	Stockholm	0	0	0	1	0	0	0	0
4	Stockholm	0	0	0	0	1	0	0	0

The final step of data preparation is to group venues by city and normallize features (i.e. transform absolute numbers to weights)

In [91]:

```
# Large cities
venues0_grouped = venues0_onehot.groupby('Neighborhood').mean().reset_index()
print(venues0_grouped.shape)
venues0_grouped
```

(18, 51)

Out[91]:

	Neighborhood	Airport Lounge	Airport Service	Arcade	Asian Restaurant	BBQ Joint	Bar	Beer Bar
0	Amsterdam	0.009524	0.009524	0.000000	0.009524	0.000000	0.057143	0.038095
1	Barcelona	0.000000	0.000000	0.000000	0.000000	0.019608	0.049020	0.029412
2	Berlin	0.000000	0.000000	0.000000	0.000000	0.013158	0.039474	0.026316
3	Budapest	0.000000	0.000000	0.000000	0.000000	0.009615	0.028846	0.028846
4	Cologne	0.000000	0.000000	0.000000	0.014493	0.000000	0.028986	0.000000
5	Hamburg	0.000000	0.000000	0.000000	0.037037	0.012346	0.012346	0.000000
6	Madrid	0.000000	0.000000	0.000000	0.000000	0.021277	0.031915	0.000000
7	Marseille	0.000000	0.000000	0.000000	0.000000	0.000000	0.119048	0.000000
8	Milan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.021053
9	Munich	0.000000	0.000000	0.012346	0.012346	0.012346	0.024691	0.012346
10	Naples	0.000000	0.000000	0.000000	0.000000	0.000000	0.009009	0.000000
11	Novosibirsk	0.011765	0.035294	0.011765	0.011765	0.035294	0.011765	0.011765
12	Paris	0.000000	0.000000	0.000000	0.016949	0.000000	0.033898	0.016949
13	Prague	0.000000	0.000000	0.009259	0.037037	0.000000	0.037037	0.018519
14	Rome	0.000000	0.000000	0.000000	0.030303	0.010101	0.000000	0.000000
15	Stockholm	0.000000	0.000000	0.000000	0.011905	0.000000	0.023810	0.047619
16	Turin	0.000000	0.000000	0.000000	0.000000	0.000000	0.012987	0.012987
17	Vienna	0.000000	0.000000	0.000000	0.070588	0.000000	0.023529	0.000000

In [92]:

```
# Medium cities
venues1_grouped = venues1_onehot.groupby('Neighborhood').mean().reset_index()
print(venues1_grouped.shape)
venues1_grouped
```

(49, 35)

Out[92]:

	Neighborhood	Art Gallery	Australian Restaurant	Beer Store	Bookstore	Café	Cocktail Bar	Coffee Shop
0	Aarhus	0.000000	0.000000	0.000000	0.000000	0.137931	0.034483	0.172414
1	Amsterdam	0.000000	0.013889	0.013889	0.041667	0.180556	0.013889	0.222222
2	Antwerp	0.000000	0.000000	0.000000	0.031250	0.046875	0.046875	0.265625
3	Athens	0.011628	0.000000	0.023256	0.023256	0.174419	0.046512	0.220930
4	Bari	0.000000	0.000000	0.000000	0.047619	0.238095	0.071429	0.047619
5	Bielefeld	0.000000	0.000000	0.000000	0.000000	0.250000	0.041667	0.041667
6	Bilbao	0.000000	0.000000	0.000000	0.055556	0.166667	0.000000	0.000000
7	Bochum	0.000000	0.000000	0.011765	0.023529	0.270588	0.011765	0.023529
8	Bologna	0.000000	0.000000	0.000000	0.021739	0.119565	0.021739	0.010870
9	Bonn	0.000000	0.000000	0.030303	0.000000	0.151515	0.060606	0.030303
10	Bremen	0.000000	0.000000	0.000000	0.000000	0.266667	0.033333	0.066667
11	Brno	0.000000	0.000000	0.010870	0.000000	0.326087	0.032609	0.076087
12	Catania	0.000000	0.000000	0.020000	0.020000	0.300000	0.080000	0.000000
13	Copenhagen	0.014085	0.000000	0.014085	0.014085	0.211268	0.070423	0.183099
14	Dortmund	0.000000	0.000000	0.019608	0.019608	0.274510	0.000000	0.039216
15	Dresden	0.058824	0.000000	0.058824	0.000000	0.117647	0.058824	0.117647
16	Duisburg	0.034483	0.000000	0.000000	0.017241	0.241379	0.017241	0.017241
17	Düsseldorf	0.000000	0.000000	0.016129	0.016129	0.258065	0.016129	0.064516
18	Essen	0.014085	0.000000	0.000000	0.014085	0.239437	0.014085	0.028169
19	Florence	0.000000	0.000000	0.000000	0.011364	0.181818	0.011364	0.011364
20	Frankfurt	0.000000	0.000000	0.000000	0.000000	0.285714	0.038961	0.038961
21	Genoa	0.000000	0.000000	0.000000	0.000000	0.052632	0.105263	0.052632
22	Gothenburg	0.035714	0.000000	0.000000	0.035714	0.285714	0.000000	0.178571
23	Hanover	0.000000	0.027027	0.027027	0.000000	0.135135	0.054054	0.135135
24	Helsinki	0.014286	0.000000	0.014286	0.000000	0.328571	0.028571	0.114286
25	Irkutsk	0.021739	0.021739	0.021739	0.021739	0.086957	0.065217	0.086957
26	Karlsruhe	0.000000	0.000000	0.000000	0.000000	0.185185	0.037037	0.111111
27	Kraków	0.035088	0.000000	0.000000	0.035088	0.245614	0.017544	0.087719
28	Lisbon	0.034483	0.000000	0.000000	0.017241	0.189655	0.034483	0.086207
29	Lyon	0.032787	0.000000	0.000000	0.016393	0.081967	0.016393	0.032787
30	Malmö	0.017857	0.000000	0.000000	0.017857	0.285714	0.035714	0.125000
31	Mannheim	0.000000	0.000000	0.000000	0.000000	0.260870	0.043478	0.108696
32	Marseille	0.000000	0.000000	0.000000	0.000000	0.025641	0.000000	0.076923
33	Münster	0.000000	0.000000	0.000000	0.047619	0.428571	0.000000	0.190476
34	Nantes	0.000000	0.000000	0.000000	0.000000	0.000000	0.090909	0.000000
35	Naples	0.000000	0.000000	0.000000	0.012658	0.278481	0.063291	0.050633

	Neighborhood	Art Gallery	Australian Restaurant	Beer Store	Bookstore	Café	Cocktail Bar	Coffee Shop
36	Nice	0.000000	0.000000	0.000000	0.000000	0.142857	0.071429	0.035714
37	Nuremberg	0.000000	0.000000	0.000000	0.037037	0.277778	0.018519	0.148148
38	Palma de Mallorca	0.023810	0.000000	0.000000	0.000000	0.238095	0.119048	0.166667
39	Riga	0.026667	0.000000	0.053333	0.026667	0.120000	0.026667	0.040000
40	Rotterdam	0.000000	0.000000	0.017857	0.000000	0.214286	0.000000	0.160714
41	Stockholm	0.015152	0.000000	0.000000	0.045455	0.333333	0.060606	0.151515
42	Stuttgart	0.000000	0.000000	0.000000	0.000000	0.222222	0.027778	0.027778
43	The Hague	0.000000	0.000000	0.000000	0.020408	0.183673	0.000000	0.122449
44	Toulouse	0.000000	0.000000	0.000000	0.043478	0.000000	0.043478	0.043478
45	Turin	0.000000	0.000000	0.000000	0.000000	0.134615	0.038462	0.019231
46	Utrecht	0.026316	0.000000	0.052632	0.000000	0.052632	0.026316	0.131579
47	Valencia	0.050000	0.000000	0.000000	0.000000	0.150000	0.000000	0.050000
48	Wuppertal	0.000000	0.000000	0.000000	0.000000	0.352941	0.029412	0.029412

Now let's take a look at the top (i.e. most frequent) categories of venues for each city:

In [93]:

```
def list_top_venues(venues_grouped, num_top_venues=5):
    for city in venues_grouped.Neighborhood:
        print("----" + city + "----")
        temp = venues_grouped[venues_grouped.Neighborhood == city].T.reset_index()
        temp.columns = ['venue','freq']
        temp = temp.iloc[1:]
        temp['freq'] = temp['freq'].astype(float)
        temp = temp.round({'freq': 2})
        print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num
_top_venues))
        print('\n')
```

In [94]:

```
# Large cities
list_top_venues(venues0_grouped)
```

```
----Amsterdam----
                venue  freq
0         Coffee Shop  0.15
1                Café  0.12
2                Park  0.11
3          Restaurant  0.08
4  Italian Restaurant  0.08


----Barcelona----
              venue  freq
0              Park  0.12
1  Tapas Restaurant  0.12
2        Restaurant  0.10
3       Pizza Place  0.09
4       Coffee Shop  0.08


----Berlin----
          venue  freq
0          Park  0.20
1   Coffee Shop  0.16
2          Café  0.14
3      Wine Bar  0.08
4  Cocktail Bar  0.05


----Budapest----
          venue  freq
0   Coffee Shop  0.16
1          Park  0.11
2  Dessert Shop  0.10
3   Pizza Place  0.08
4  Burger Joint  0.08


----Cologne----
                  venue  freq
0                  Café  0.19
1                  Park  0.16
2    Italian Restaurant  0.14
3  Gym / Fitness Center  0.04
4            Restaurant  0.04


----Hamburg----
         venue  freq
0         Café  0.22
1         Park  0.14
2  Coffee Shop  0.09
3  Pizza Place  0.07
4   Steakhouse  0.06


----Madrid----
                venue  freq
0          Restaurant  0.19
1                Park  0.11
2  Italian Restaurant  0.11
3    Tapas Restaurant  0.10
4         Coffee Shop  0.09
```

```
----Marseille----
                 venue  freq
0  Fast Food Restaurant  0.29
1                   Bar  0.12
2            Steakhouse  0.10
3           Coffee Shop  0.07
4                  Park  0.07


----Milan----
               venue  freq
0  Italian Restaurant  0.31
1                Park  0.16
2         Pizza Place  0.13
3                Café  0.07
4        Dessert Shop  0.05


----Munich----
               venue  freq
0                Café  0.23
1  Italian Restaurant  0.16
2                Park  0.14
3        Cocktail Bar  0.05
4          Steakhouse  0.04


----Naples----
               venue  freq
0                Café  0.20
1         Pizza Place  0.20
2  Italian Restaurant  0.15
3        Burger Joint  0.05
4          Restaurant  0.05


----Novosibirsk----
         venue  freq
0  Coffee Shop  0.11
1          Pub  0.06
2         Park  0.05
3      Theater  0.04
4    BBQ Joint  0.04


----Paris----
               venue  freq
0                Park  0.25
1  Italian Restaurant  0.14
2         Pizza Place  0.07
3            Wine Bar  0.05
4                 Pub  0.05


----Prague----
             venue  freq
0              Café  0.26
1              Park  0.17
2       Coffee Shop  0.09
```

```
3  Asian Restaurant  0.04
4           Theater  0.04


----Rome----
                  venue  freq
0    Italian Restaurant  0.25
1           Pizza Place  0.18
2                  Park  0.12
3  Gym / Fitness Center  0.06
4                  Café  0.06


----Stockholm----
         venue  freq
0         Café  0.26
1         Park  0.17
2  Coffee Shop  0.12
3  Pizza Place  0.06
4     Beer Bar  0.05


----Turin----
                  venue  freq
0           Pizza Place  0.27
1    Italian Restaurant  0.22
2                  Café  0.09
3                  Park  0.05
4  Fast Food Restaurant  0.05


----Vienna----
                venue  freq
0                Park  0.14
1                Café  0.13
2  Italian Restaurant  0.09
3         Coffee Shop  0.08
4          Restaurant  0.08
```

In [95]:

```
# Medium cities
list_top_venues(venues1_grouped)
```

```
----Aarhus----
                venue  freq
0         Coffee Shop  0.17
1            Wine Bar  0.14
2                Park  0.14
3                Café  0.14
4  Gym / Fitness Center  0.10


----Amsterdam----
                venue  freq
0         Coffee Shop  0.22
1                Café  0.18
2                Park  0.17
3  Italian Restaurant  0.11
4            Wine Bar  0.04


----Antwerp----
                  venue  freq
0           Coffee Shop  0.27
1    Italian Restaurant  0.20
2                  Park  0.11
3  Gym / Fitness Center  0.06
4                  Café  0.05


----Athens----
          venue  freq
0   Coffee Shop  0.22
1          Café  0.17
2  Dessert Shop  0.15
3          Park  0.12
4  Cocktail Bar  0.05


----Bari----
                venue  freq
0  Italian Restaurant  0.29
1                Café  0.24
2                 Pub  0.07
3        Cocktail Bar  0.07
4        Dessert Shop  0.07


----Bielefeld----
                venue  freq
0                Café  0.25
1                Park  0.25
2           Gastropub  0.08
3  Italian Restaurant  0.08
4             Theater  0.08


----Bilbao----
            venue  freq
0            Café  0.17
1            Park  0.17
2       Gastropub  0.11
3  Scenic Lookout  0.11
4             Pub  0.11
```

```
----Bochum----
               venue  freq
0               Café  0.27
1               Park  0.18
2  Italian Restaurant  0.13
3      History Museum  0.08
4                Pub  0.05


----Bologna----
               venue  freq
0  Italian Restaurant  0.58
1               Café  0.12
2               Park  0.07
3        Cupcake Shop  0.04
4                Pub  0.03


----Bonn----
               venue  freq
0  Italian Restaurant  0.18
1               Café  0.15
2               Park  0.12
3    Pedestrian Plaza  0.06
4        Cocktail Bar  0.06


----Bremen----
               venue  freq
0  Italian Restaurant  0.27
1               Café  0.27
2               Park  0.10
3           Gastropub  0.07
4       Train Station  0.07


----Brno----
         venue  freq
0         Café  0.33
1          Pub  0.15
2         Park  0.10
3  Coffee Shop  0.08
4      Theater  0.04


----Catania----
               venue  freq
0               Café  0.30
1  Italian Restaurant  0.26
2        Dessert Shop  0.08
3                Pub  0.08
4        Cocktail Bar  0.08


----Copenhagen----
          venue  freq
0          Café  0.21
1   Coffee Shop  0.18
2          Park  0.17
```

```
3      Wine Bar  0.07
4  Cocktail Bar  0.07


----Dortmund----
                venue  freq
0                Café  0.27
1                Park  0.24
2  Italian Restaurant  0.10
3      History Museum  0.06
4             Theater  0.04


----Dresden----
         venue  freq
0         Park  0.29
1  Coffee Shop  0.12
2    Gastropub  0.12
3         Café  0.12
4  Art Gallery  0.06


----Duisburg----
                venue  freq
0                Café  0.24
1                Park  0.14
2  Italian Restaurant  0.14
3      Scenic Lookout  0.07
4                 Pub  0.07


----Düsseldorf----
                  venue  freq
0                  Café  0.26
1                  Park  0.16
2    Italian Restaurant  0.15
3            Steakhouse  0.08
4  Gym / Fitness Center  0.06


----Essen----
                venue  freq
0                Café  0.24
1                Park  0.17
2  Italian Restaurant  0.11
3                 Pub  0.07
4      History Museum  0.06


----Florence----
                venue  freq
0  Italian Restaurant  0.47
1                Café  0.18
2                Park  0.09
3        Dessert Shop  0.07
4                 Pub  0.03


----Frankfurt----
                  venue  freq
0                  Café  0.29
```

```
1    Italian Restaurant  0.18
2                  Park  0.17
3  Gym / Fitness Center  0.08
4            Steakhouse  0.05


----Genoa----
                venue  freq
0      Scenic Lookout  0.26
1                 Pub  0.16
2  Italian Restaurant  0.16
3        Cocktail Bar  0.11
4             Theater  0.05


----Gothenburg----
         venue  freq
0         Café  0.29
1          Pub  0.25
2  Coffee Shop  0.18
3         Park  0.07
4  Art Gallery  0.04


----Hanover----
                venue  freq
0  Italian Restaurant  0.22
1                Café  0.14
2         Coffee Shop  0.14
3                Park  0.14
4        Cocktail Bar  0.05


----Helsinki----
                  venue  freq
0                  Café  0.33
1                  Park  0.20
2           Coffee Shop  0.11
3  Gym / Fitness Center  0.07
4                Island  0.04


----Irkutsk----
              venue  freq
0              Café  0.09
1       Coffee Shop  0.09
2      Cocktail Bar  0.07
3  Pedestrian Plaza  0.07
4               Pub  0.04


----Karlsruhe----
                venue  freq
0                Café  0.19
1  Italian Restaurant  0.15
2       Train Station  0.11
3         Coffee Shop  0.11
4           Gastropub  0.11


----Kraków----
```

```
                venue  freq
0                Café  0.25
1                Park  0.18
2  Italian Restaurant  0.12
3         Coffee Shop  0.09
4                 Pub  0.07


----Lisbon----
            venue  freq
0            Park  0.22
1            Café  0.19
2  Scenic Lookout  0.12
3        Wine Bar  0.09
4     Coffee Shop  0.09


----Lyon----
           venue  freq
0  Train Station  0.48
1           Café  0.08
2     Steakhouse  0.07
3           Park  0.07
4    Art Gallery  0.03


----Malmö----
                  venue  freq
0                  Café  0.29
1                  Park  0.16
2           Coffee Shop  0.12
3    Italian Restaurant  0.09
4  Gym / Fitness Center  0.07


----Mannheim----
                  venue  freq
0                  Café  0.26
1    Italian Restaurant  0.17
2                  Park  0.15
3           Coffee Shop  0.11
4  Gym / Fitness Center  0.07


----Marseille----
            venue  freq
0   Train Station  0.36
1      Steakhouse  0.10
2     Coffee Shop  0.08
3            Park  0.08
4  Scenic Lookout  0.05


----Münster----
                venue  freq
0                Café  0.43
1         Coffee Shop  0.19
2  Italian Restaurant  0.14
3                Park  0.10
4       Train Station  0.05
```

```
----Nantes----
           venue  freq
0  Train Station  0.45
1           Park  0.36
2   Dessert Shop  0.09
3   Cocktail Bar  0.09
4      Rest Area  0.00


----Naples----
                venue  freq
0                Café  0.28
1  Italian Restaurant  0.22
2                 Pub  0.08
3        Cocktail Bar  0.06
4         Coffee Shop  0.05


----Nice----
                venue  freq
0  Italian Restaurant  0.21
1                Café  0.14
2       Train Station  0.11
3                Park  0.11
4      Scenic Lookout  0.11


----Nuremberg----
                  venue  freq
0                  Café  0.28
1                  Park  0.17
2           Coffee Shop  0.15
3    Italian Restaurant  0.11
4  Gym / Fitness Center  0.07


----Palma de Mallorca----
                venue  freq
0                Café  0.24
1         Coffee Shop  0.17
2  Italian Restaurant  0.17
3        Cocktail Bar  0.12
4          Steakhouse  0.05


----Riga----
                  venue  freq
0                  Park  0.32
1                  Café  0.12
2  Gym / Fitness Center  0.08
3        Scenic Lookout  0.07
4            Beer Store  0.05


----Rotterdam----
                venue  freq
0                Café  0.21
1         Coffee Shop  0.16
2                Park  0.14
3  Italian Restaurant  0.09
```

```
4           Gastropub  0.07


----Stockholm----
          venue  freq
0          Café  0.33
1          Park  0.21
2   Coffee Shop  0.15
3  Cocktail Bar  0.06
4     Bookstore  0.05


----Stuttgart----
                venue  freq
0                Park  0.28
1                Café  0.22
2  Italian Restaurant  0.11
3      Scenic Lookout  0.06
4           Gastropub  0.06


----The Hague----
                venue  freq
0                Park  0.24
1                Café  0.18
2         Coffee Shop  0.12
3  Italian Restaurant  0.08
4                 Pub  0.06


----Toulouse----
           venue  freq
0  Train Station  0.35
1            Pub  0.17
2           Park  0.09
3       Wine Bar  0.04
4      Bookstore  0.04


----Turin----
                venue  freq
0  Italian Restaurant  0.33
1                Café  0.13
2          Steakhouse  0.08
3                Park  0.08
4            Wine Bar  0.06


----Utrecht----
                venue  freq
0                Park  0.21
1         Coffee Shop  0.13
2          Steakhouse  0.11
3  Italian Restaurant  0.08
4                 Pub  0.05


----Valencia----
                venue  freq
0  Italian Restaurant  0.25
1                Park  0.25
```

```
2                Café  0.15
3        Dessert Shop  0.10
4             Theater  0.05


----Wuppertal----
                venue  freq
0                Café  0.35
1           Gastropub  0.15
2      History Museum  0.12
3  Italian Restaurant  0.09
4                Park  0.09
```

In [97]:

```
# Define a function that willl sort the top categories and store them to dataframe
def df_top_venues(venues_grouped, num_top_venues=10):

    def get_top_venues(row, num_top_venues):
        row_categories = row.iloc[1:]
        row_categories_sorted = row_categories.sort_values(ascending=False)
        return row_categories_sorted.index.values[0:num_top_venues]

    # create columns according to number of top venues
    indicators = ['st', 'nd', 'rd']
    columns = ['Neighborhood']
    for ind in np.arange(num_top_venues):
        try:
            columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
        except:
            columns.append('{}th Most Common Venue'.format(ind+1))

    # create a new dataframe
    venues_sorted = pd.DataFrame(columns=columns)
    venues_sorted['Neighborhood'] = venues_grouped['Neighborhood']
    for ind in np.arange(venues_grouped.shape[0]):
        venues_sorted.iloc[ind, 1:] = get_top_venues(venues_grouped.iloc[ind, :], num_t
op_venues)

    return venues_sorted
```

In [98]:

```
# Large cities
venues0_sorted = df_top_venues(venues0_grouped)
venues0_sorted.to_csv('top_venues0_sorted.csv')
venues0_sorted
```

Out[98]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
0	Amsterdam	Coffee Shop	Café	Park	Restaurant	Italian Restaurant	Bar	
1	Barcelona	Tapas Restaurant	Park	Restaurant	Pizza Place	Coffee Shop	Cocktail Bar	
2	Berlin	Park	Coffee Shop	Café	Wine Bar	Cocktail Bar	Pizza Place	
3	Budapest	Coffee Shop	Park	Dessert Shop	Pizza Place	Burger Joint	Gym / Fitness Center	Re
4	Cologne	Café	Park	Italian Restaurant	Restaurant	Cocktail Bar	Coffee Shop	
5	Hamburg	Café	Park	Coffee Shop	Pizza Place	Steakhouse	Wine Bar	Re
6	Madrid	Restaurant	Park	Italian Restaurant	Tapas Restaurant	Coffee Shop	Burger Joint	
7	Marseille	Fast Food Restaurant	Bar	Steakhouse	Park	Coffee Shop	Pub	Re
8	Milan	Italian Restaurant	Park	Pizza Place	Café	Dessert Shop	Brewery	
9	Munich	Café	Italian Restaurant	Park	Cocktail Bar	Steakhouse	Brewery	Re
10	Naples	Café	Pizza Place	Italian Restaurant	Pub	Fast Food Restaurant	Cocktail Bar	Re
11	Novosibirsk	Coffee Shop	Pub	Park	Theater	Airport Service	BBQ Joint	
12	Paris	Park	Italian Restaurant	Pizza Place	Wine Bar	Pub	Café	Co
13	Prague	Café	Park	Coffee Shop	Restaurant	Asian Restaurant	Bar	
14	Rome	Italian Restaurant	Pizza Place	Park	Café	Gym / Fitness Center	Restaurant	
15	Stockholm	Café	Park	Coffee Shop	Pizza Place	Restaurant	Cocktail Bar	E
16	Turin	Pizza Place	Italian Restaurant	Café	Fast Food Restaurant	Park	Steakhouse	V
17	Vienna	Park	Café	Italian Restaurant	Coffee Shop	Restaurant	Asian Restaurant	

In [99]:

```
# Medium cities
venues1_sorted = df_top_venues(venues1_grouped)
venues1_sorted.to_csv('top_venues1_sorted.csv')
venues1_sorted
```

Out[99]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Aarhus	Coffee Shop	Café	Park	Wine Bar	Gym / Fitness Center	Steakhouse
1	Amsterdam	Coffee Shop	Café	Park	Italian Restaurant	Dessert Shop	Wine Bar
2	Antwerp	Coffee Shop	Italian Restaurant	Park	Gym / Fitness Center	Cocktail Bar	Farm
3	Athens	Coffee Shop	Café	Dessert Shop	Park	Cocktail Bar	History Museum
4	Bari	Italian Restaurant	Café	Dessert Shop	Pub	Cocktail Bar	Coffee Shop
5	Bielefeld	Park	Café	Theater	Steakhouse	Italian Restaurant	Gastropub
6	Bilbao	Café	Park	Gastropub	Scenic Lookout	Pub	Theater
7	Bochum	Café	Park	Italian Restaurant	History Museum	Pub	Theater
8	Bologna	Italian Restaurant	Café	Park	Cupcake Shop	Pub	Dessert Shop
9	Bonn	Italian Restaurant	Café	Park	Pedestrian Plaza	History Museum	Steakhouse
10	Bremen	Café	Italian Restaurant	Park	Steakhouse	Train Station	Coffee Shop
11	Brno	Café	Pub	Park	Coffee Shop	Theater	Gastropub
12	Catania	Café	Italian Restaurant	Cocktail Bar	Dessert Shop	Pub	Steakhouse
13	Copenhagen	Café	Coffee Shop	Park	Wine Bar	Italian Restaurant	Cocktail Bar
14	Dortmund	Café	Park	Italian Restaurant	History Museum	Garden Center	Coffee Shop
15	Dresden	Park	Coffee Shop	Café	Gastropub	Soccer Stadium	Scenic Lookout
16	Duisburg	Café	Italian Restaurant	Park	Scenic Lookout	Pub	Gym / Fitness Center
17	Düsseldorf	Café	Park	Italian Restaurant	Steakhouse	Gym / Fitness Center	Coffee Shop
18	Essen	Café	Park	Italian Restaurant	Pub	History Museum	Theater
19	Florence	Italian Restaurant	Café	Park	Dessert Shop	Pub	Steakhouse

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
20	Frankfurt	Café	Italian Restaurant	Park	Gym / Fitness Center	Steakhouse	Coffee Shop
21	Genoa	Scenic Lookout	Pub	Italian Restaurant	Cocktail Bar	Coffee Shop	Theater
22	Gothenburg	Café	Pub	Coffee Shop	Park	Wine Bar	Bookstore
23	Hanover	Italian Restaurant	Park	Coffee Shop	Café	Steakhouse	Cocktail Bar
24	Helsinki	Café	Park	Coffee Shop	Gym / Fitness Center	Theater	Island
25	Irkutsk	Café	Coffee Shop	Cocktail Bar	Pedestrian Plaza	Garden Center	Pub
26	Karlsruhe	Café	Italian Restaurant	Coffee Shop	Train Station	Gastropub	Park
27	Kraków	Café	Park	Italian Restaurant	Coffee Shop	Pub	Wine Bar
28	Lisbon	Park	Café	Scenic Lookout	Wine Bar	Coffee Shop	Gym / Fitness Center
29	Lyon	Train Station	Café	Steakhouse	Park	Wine Bar	Pedestrian Plaza
30	Malmö	Café	Park	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Wine Bar
31	Mannheim	Café	Italian Restaurant	Park	Coffee Shop	Gym / Fitness Center	Gastropub
32	Marseille	Train Station	Steakhouse	Park	Coffee Shop	Scenic Lookout	Pub
33	Münster	Café	Coffee Shop	Italian Restaurant	Park	Theater	Train Station
34	Nantes	Train Station	Park	Dessert Shop	Cocktail Bar	Gym / Fitness Center	Gastropub
35	Naples	Café	Italian Restaurant	Pub	Cocktail Bar	Coffee Shop	Dessert Shop
36	Nice	Italian Restaurant	Café	Scenic Lookout	Train Station	Park	Cocktail Bar
37	Nuremberg	Café	Park	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Steakhouse
38	Palma de Mallorca	Café	Coffee Shop	Italian Restaurant	Cocktail Bar	Park	Steakhouse
39	Riga	Park	Café	Gym / Fitness Center	Scenic Lookout	Beer Store	Theater
40	Rotterdam	Café	Coffee Shop	Park	Italian Restaurant	Gastropub	Wine Bar

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
41	Stockholm	Café	Park	Coffee Shop	Cocktail Bar	Bookstore	Wine Bar
42	Stuttgart	Park	Café	Italian Restaurant	Gastropub	Scenic Lookout	Cocktail Bar
43	The Hague	Park	Café	Coffee Shop	Italian Restaurant	Theater	Pub
44	Toulouse	Train Station	Pub	Park	Wine Bar	History Museum	Dessert Shop
45	Turin	Italian Restaurant	Café	Steakhouse	Park	Wine Bar	Pub
46	Utrecht	Park	Coffee Shop	Steakhouse	Italian Restaurant	Pub	Gym / Fitness Center
47	Valencia	Italian Restaurant	Park	Café	Dessert Shop	Art Gallery	Pub
48	Wuppertal	Café	Gastropub	History Museum	Park	Italian Restaurant	Gym / Fitness Center

5.4. Clustering the cities: apply the k-means method

Let's cluster the large cities:

In [100]:

```
# Apply the algorithm
K_CLUSTERS0 = 5
venues0_clustered = venues0_grouped.drop(columns=['Neighborhood'])
kmeans0 = KMeans(n_clusters=K_CLUSTERS0, random_state=0).fit(venues0_clustered)
kmeans0.labels_
```

Out[100]:

```
array([0, 3, 4, 3, 0, 4, 3, 2, 1, 0, 1, 3, 0, 4, 1, 4, 1, 0], dtype=int32)
```

In [101]:

```
# Take a look at results
cluster_sizes0 = [list(kmeans0.labels_).count(i) for i in range(K_CLUSTERS0)]
biggest_cluster_index0 = np.array(cluster_sizes0).argmax()
print(f'Cluster sizes are: {cluster_sizes0}.')
print(f'The biggest cluster is {biggest_cluster_index0}, it contains {cluster_sizes0[bi
ggest_cluster_index0]} neighborhoods.')
```

```
Cluster sizes are: [5, 4, 1, 4, 4].
The biggest cluster is 0, it contains 5 neighborhoods.
```

In [102]:

```
# Insert cluster labels into dataframe with top categories
venues0_sorted.insert(0, 'Cluster Labels', kmeans0.labels_)
venues0_sorted.sort_values(by=['Cluster Labels'], inplace=True, ignore_index=True)
venues0_sorted
```

Out[102]:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Com V
0	0	Amsterdam	Coffee Shop	Café	Park	Restaurant	Italian Restaurant	
1	0	Paris	Park	Italian Restaurant	Pizza Place	Wine Bar	Pub	
2	0	Munich	Café	Italian Restaurant	Park	Cocktail Bar	Steakhouse	Bre
3	0	Cologne	Café	Park	Italian Restaurant	Restaurant	Cocktail Bar	C
4	0	Vienna	Park	Café	Italian Restaurant	Coffee Shop	Restaurant	/ Resta
5	1	Turin	Pizza Place	Italian Restaurant	Café	Fast Food Restaurant	Park	Steakh
6	1	Naples	Café	Pizza Place	Italian Restaurant	Pub	Fast Food Restaurant	Co
7	1	Rome	Italian Restaurant	Pizza Place	Park	Café	Gym / Fitness Center	Resta
8	1	Milan	Italian Restaurant	Park	Pizza Place	Café	Dessert Shop	Bre
9	2	Marseille	Fast Food Restaurant	Bar	Steakhouse	Park	Coffee Shop	
10	3	Madrid	Restaurant	Park	Italian Restaurant	Tapas Restaurant	Coffee Shop	B
11	3	Novosibirsk	Coffee Shop	Pub	Park	Theater	Airport Service	BBQ
12	3	Barcelona	Tapas Restaurant	Park	Restaurant	Pizza Place	Coffee Shop	Co
13	3	Budapest	Coffee Shop	Park	Dessert Shop	Pizza Place	Burger Joint	(Fi C
14	4	Hamburg	Café	Park	Coffee Shop	Pizza Place	Steakhouse	Win
15	4	Berlin	Park	Coffee Shop	Café	Wine Bar	Cocktail Bar	Pizza P
16	4	Prague	Café	Park	Coffee Shop	Restaurant	Asian Restaurant	
17	4	Stockholm	Café	Park	Coffee Shop	Pizza Place	Restaurant	Co

Now, let's cluster the medium cities:

In [103]:

```
# Apply the algorithm
K_CLUSTERS1 = 10
venues1_clustered = venues1_grouped.drop(columns=['Neighborhood'])
kmeans1 = KMeans(n_clusters=K_CLUSTERS1, random_state=0).fit(venues1_clustered)
kmeans1.labels_
```

Out[103]:

```
array([0, 0, 9, 0, 2, 3, 6, 3, 5, 9, 2, 8, 2, 0, 3, 6, 3, 3, 3, 5, 3, 4,
       8, 9, 0, 9, 9, 3, 0, 1, 0, 3, 1, 0, 7, 2, 9, 0, 9, 6, 0, 0, 3, 0,
       1, 2, 6, 3, 3], dtype=int32)
```

In [104]:

```
# Take a look at results
cluster_sizes1 = [list(kmeans1.labels_).count(i) for i in range(K_CLUSTERS1)]
biggest_cluster_index1 = np.array(cluster_sizes1).argmax()
print(f'Cluster sizes are: {cluster_sizes1}.')
print(f'The biggest cluster is {biggest_cluster_index1}, it contains {cluster_sizes1[bi
ggest_cluster_index1]} neighborhoods.')
```

```
Cluster sizes are: [12, 3, 5, 12, 1, 2, 4, 1, 2, 7].
The biggest cluster is 0, it contains 12 neighborhoods.
```

In [105]:

```
# Insert cluster labels into dataframe with top categories
venues1_sorted.insert(0, 'Cluster Labels', kmeans1.labels_)
venues1_sorted.sort_values(by=['Cluster Labels'], inplace=True, ignore_index=True)
venues1_sorted
```

Out[105]:

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t Co
0	0	Aarhus	Coffee Shop	Café	Park	Wine Bar	Gym / Fitness Center	Steal
1	0	The Hague	Park	Café	Coffee Shop	Italian Restaurant	Theater	
2	0	Stockholm	Café	Park	Coffee Shop	Cocktail Bar	Bookstore	W
3	0	Rotterdam	Café	Coffee Shop	Park	Italian Restaurant	Gastropub	W
4	0	Nuremberg	Café	Park	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Steal
5	0	Münster	Café	Coffee Shop	Italian Restaurant	Park	Theater	
6	0	Malmö	Café	Park	Coffee Shop	Italian Restaurant	Gym / Fitness Center	W
7	0	Lisbon	Park	Café	Scenic Lookout	Wine Bar	Coffee Shop	I
8	0	Copenhagen	Café	Coffee Shop	Park	Wine Bar	Italian Restaurant	C
9	0	Helsinki	Café	Park	Coffee Shop	Gym / Fitness Center	Theater	
10	0	Amsterdam	Coffee Shop	Café	Park	Italian Restaurant	Dessert Shop	W
11	0	Athens	Coffee Shop	Café	Dessert Shop	Park	Cocktail Bar	M
12	1	Lyon	Train Station	Café	Steakhouse	Park	Wine Bar	Ped
13	1	Toulouse	Train Station	Pub	Park	Wine Bar	History Museum	D
14	1	Marseille	Train Station	Steakhouse	Park	Coffee Shop	Scenic Lookout	
15	2	Turin	Italian Restaurant	Café	Steakhouse	Park	Wine Bar	
16	2	Bari	Italian Restaurant	Café	Dessert Shop	Pub	Cocktail Bar	
17	2	Naples	Café	Italian Restaurant	Pub	Cocktail Bar	Coffee Shop	D
18	2	Bremen	Café	Italian Restaurant	Park	Steakhouse	Train Station	
19	2	Catania	Café	Italian Restaurant	Cocktail Bar	Dessert Shop	Pub	Steal
20	3	Bielefeld	Park	Café	Theater	Steakhouse	Italian Restaurant	Gas

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t Co
21	3	Stuttgart	Park	Café	Italian Restaurant	Gastropub	Scenic Lookout	C
22	3	Dortmund	Café	Park	Italian Restaurant	History Museum	Garden Center	
23	3	Duisburg	Café	Italian Restaurant	Park	Scenic Lookout	Pub	I
24	3	Essen	Café	Park	Italian Restaurant	Pub	History Museum	T
25	3	Düsseldorf	Café	Park	Italian Restaurant	Steakhouse	Gym / Fitness Center	
26	3	Mannheim	Café	Italian Restaurant	Park	Coffee Shop	Gym / Fitness Center	Gas
27	3	Bochum	Café	Park	Italian Restaurant	History Museum	Pub	T
28	3	Wuppertal	Café	Gastropub	History Museum	Park	Italian Restaurant	I
29	3	Valencia	Italian Restaurant	Park	Café	Dessert Shop	Art Gallery	
30	3	Frankfurt	Café	Italian Restaurant	Park	Gym / Fitness Center	Steakhouse	
31	3	Kraków	Café	Park	Italian Restaurant	Coffee Shop	Pub	W
32	4	Genoa	Scenic Lookout	Pub	Italian Restaurant	Cocktail Bar	Coffee Shop	T
33	5	Florence	Italian Restaurant	Café	Park	Dessert Shop	Pub	Steal
34	5	Bologna	Italian Restaurant	Café	Park	Cupcake Shop	Pub	D
35	6	Bilbao	Café	Park	Gastropub	Scenic Lookout	Pub	T
36	6	Utrecht	Park	Coffee Shop	Steakhouse	Italian Restaurant	Pub	I
37	6	Dresden	Park	Coffee Shop	Café	Gastropub	Soccer Stadium	L
38	6	Riga	Park	Café	Gym / Fitness Center	Scenic Lookout	Beer Store	T
39	7	Nantes	Train Station	Park	Dessert Shop	Cocktail Bar	Gym / Fitness Center	Gas
40	8	Brno	Café	Pub	Park	Coffee Shop	Theater	Gas

	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6t Co
41	8	Gothenburg	Café	Pub	Coffee Shop	Park	Wine Bar	Bo
42	9	Bonn	Italian Restaurant	Café	Park	Pedestrian Plaza	History Museum	Steal
43	9	Nice	Italian Restaurant	Café	Scenic Lookout	Train Station	Park	C
44	9	Hanover	Italian Restaurant	Park	Coffee Shop	Café	Steakhouse	C
45	9	Irkutsk	Café	Coffee Shop	Cocktail Bar	Pedestrian Plaza	Garden Center	
46	9	Karlsruhe	Café	Italian Restaurant	Coffee Shop	Train Station	Gastropub	
47	9	Antwerp	Coffee Shop	Italian Restaurant	Park	Gym / Fitness Center	Cocktail Bar	
48	9	Palma de Mallorca	Café	Coffee Shop	Italian Restaurant	Cocktail Bar	Park	Steal

In [106]:

```
venues0_sorted.to_csv('top_venues0_clustered.csv')
venues1_sorted.to_csv('top_venues1_clustered.csv')
```

5.5. Clustering the cites: look at results

In [138]:

```
# Remember labels for the clusters where our hometowns reside now.
# 1st hometown in some cluster among large cities.
# 2nd hometown in some cluster among medium cities.
cluster0_label = int(venues0_sorted[venues0_sorted.Neighborhood == my_cities[0]]['Clust
er Labels'])
cluster1_label = int(venues1_sorted[venues1_sorted.Neighborhood == my_cities[1]]['Clust
er Labels'])
print(f"Our 1st hometown {my_cities[0]} is in cluster # {cluster0_label} of large Europ
ean cities:")
print(venues0_grouped.loc[kmeans0.labels_ == cluster0_label, 'Neighborhood'].tolist())
print(f"Our 2nd hometown {my_cities[1]} is in cluster # {cluster1_label} of medium Euro
pean cities:")
print(venues1_grouped.loc[kmeans1.labels_ == cluster1_label, 'Neighborhood'].tolist())
```

```
Our 1st hometown Novosibirsk is in cluster # 3 of large European cities:
['Barcelona', 'Budapest', 'Madrid', 'Novosibirsk']
Our 2nd hometown Irkutsk is in cluster # 9 of medium European cities:
['Antwerp', 'Bonn', 'Hanover', 'Irkutsk', 'Karlsruhe', 'Nice', 'Palma de M
allorca']
```

In [139]:

```
# Large cities: select cluster of interest, join with other attributes of the cities
cluster0 = venues0_sorted[venues0_sorted['Cluster Labels'] == cluster0_label]
cluster0 = cluster0.rename(columns={'Neighborhood': 'City'})
cluster0 = cluster0.join(df_cities.set_index('City'), on='City')
cluster0.to_csv('top_cluster0.csv')
cluster0
```

Out[139]:

	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
10	3	Madrid	Restaurant	Park	Italian Restaurant	Tapas Restaurant	Coffee Shop	Burger Joint	
11	3	Novosibirsk	Coffee Shop	Pub	Park	Theater	Airport Service	BBQ Joint	
12	3	Barcelona	Tapas Restaurant	Park	Restaurant	Pizza Place	Coffee Shop	Cocktail Bar	
13	3	Budapest	Coffee Shop	Park	Dessert Shop	Pizza Place	Burger Joint	Gym / Fitness Center	Re

In [140]:

```python
# Medium cities: select cluster of interest, join with other attributes of the cities
cluster1 = venues1_sorted[venues1_sorted['Cluster Labels'] == cluster1_label]
cluster1 = cluster1.rename(columns={'Neighborhood': 'City'})
cluster1 = cluster1.join(df_cities.set_index('City'), on='City')
cluster1.to_csv('top_cluster1.csv')
cluster1
```

Out[140]:

	Cluster Labels	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
42	9	Bonn	Italian Restaurant	Café	Park	Pedestrian Plaza	History Museum	Steakhouse
43	9	Nice	Italian Restaurant	Café	Scenic Lookout	Train Station	Park	Cocktail Bar
44	9	Hanover	Italian Restaurant	Park	Coffee Shop	Café	Steakhouse	Cocktail Bar
45	9	Irkutsk	Café	Coffee Shop	Cocktail Bar	Pedestrian Plaza	Garden Center	Pub
46	9	Karlsruhe	Café	Italian Restaurant	Coffee Shop	Train Station	Gastropub	Park
47	9	Antwerp	Coffee Shop	Italian Restaurant	Park	Gym / Fitness Center	Cocktail Bar	Farm
48	9	Palma de Mallorca	Café	Coffee Shop	Italian Restaurant	Cocktail Bar	Park	Steakhouse

In [141]:

```
# Visualize the clusters of interest on a map of Europe
map_clusters = folium.Map(location=location_center, zoom_start=5)
feat_my_cities = folium.map.FeatureGroup()
for i, cluster in enumerate([cluster0, cluster1]):
    for lat, lon, label, popul in zip(cluster.Lat, cluster.Lon, cluster.City, cluster.P
opulation):
        popup_str = f'{label}.\nPopulation: {popul:,d}'
        folium.Marker([lat, lon], popup=popup_str).add_to(map_clusters)
        feat_my_cities.add_child(
            folium.features.CircleMarker(
                [lat, lon],
                radius=8 + popul // 400_000,
                color='red' if i == 0 else 'yellow',
                fill=True,
                fill_color='yellow' if i == 0 else 'blue',
                fill_opacity=0.6
            )
        )
map_clusters.add_child(feat_my_cities)
map_clusters
```

Out[141]:

In [ ]: