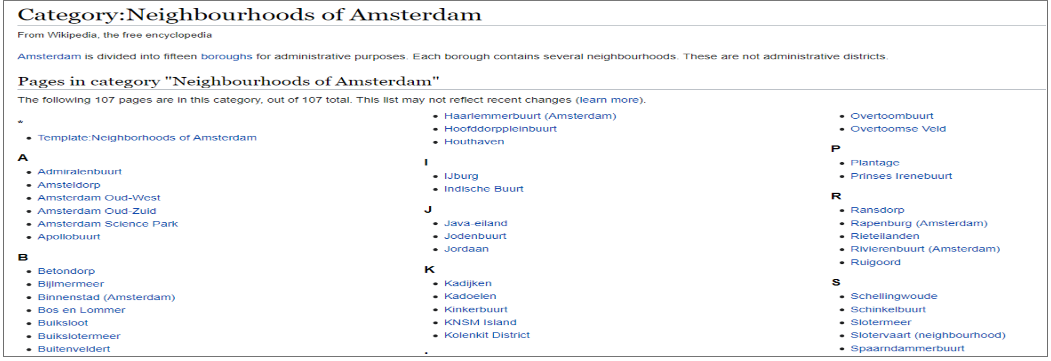
Description of the data

City neighborhood data

First we need to collect the neighborhoods of the cities. These can be found on different wikipedia websites. The python's webscraping package Beautifull Soup is going to be used to extract the data from these websites. The following sites are used:

1) Amsterdam: https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Amsterdam



Category:Neighbourhoods of Amsterdam

From Wikipedia, the free encyclopedia

Amsterdam is divided into fifteen boroughs for administrative purposes. Each borough contains several neighbourhoods. These are not administrative districts.

Pages in category "Neighbourhoods of Amsterdam"

The following 107 pages are in this category, out of 107 total. This list may not reflect recent changes (learn more).

*

- Template:Neighborhoods of Amsterdam

A

- Admiralenbuurt
- Amsteldorp
- Amsterdam Oud-West
- Amsterdam Oud-Zuid
- Amsterdam Science Park
- Apollobuurt

B

- Betondorp
- Bijlmermeer
- Binnenstad (Amsterdam)
- Bos en Lommer
- Buiksloot
- Buikslotermeer
- Buitenveldert

- Haarlemmerbuurt (Amsterdam)
- Hoofddorppleinbuurt
- Houthaven

I

- IJburg
- Indische Buurt

J

- Java-eiland
- Jodenbuurt
- Jordaan

K

- Kadijken
- Kadoelen
- Kinkerbuurt
- KNSM Island
- Kolenkit District

- Overtoombuurt
- Overtoomse Veld

P

- Plantage
- Prinses Irenebuurt

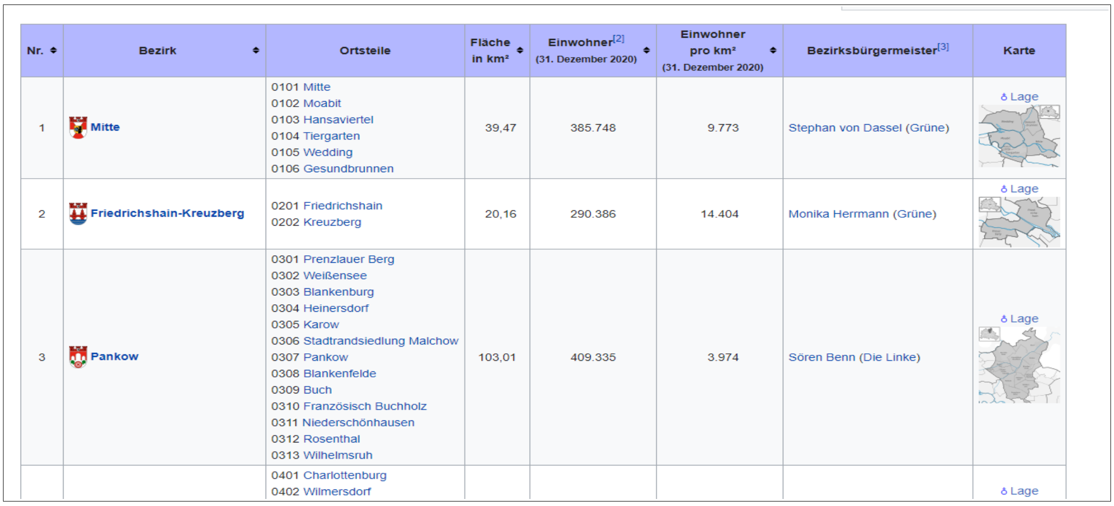
R

- Ransdorp
- Rapenburg (Amsterdam)
- Rieteilanden
- Rivierenbuurt (Amsterdam)
- Ruigoord

S

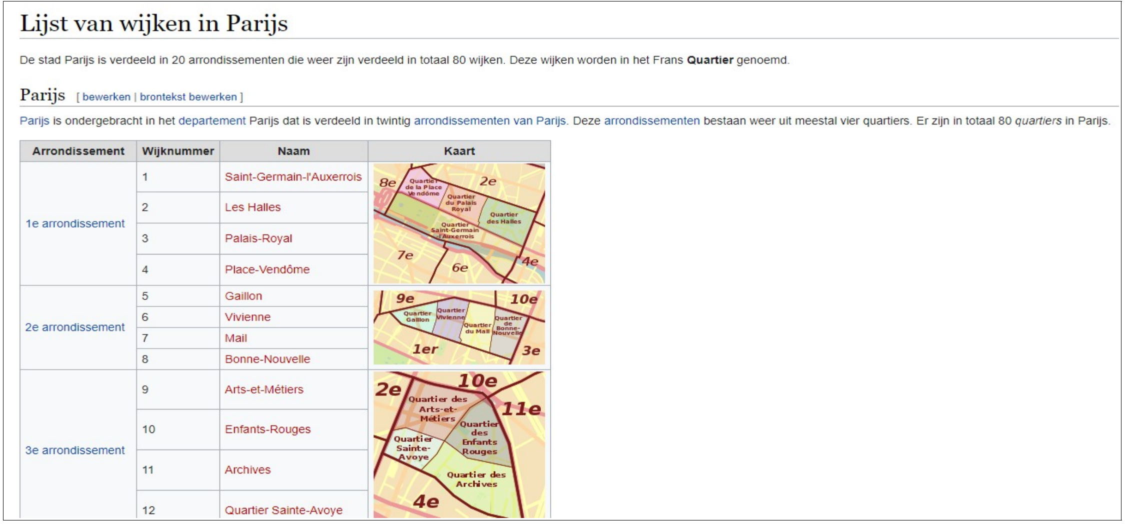
- Schellingwoude
- Schinkelbuurt
- Slotermeer
- Slotervaart (neighbourhood)
- Spaarndammerbuurt

2) Berlin: https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins



Nr.	Bezirk	Ortsteile	Fläche in km²	Einwohner[2] (31. Dezember 2020)	Einwohner pro km² (31. Dezember 2020)	Bezirksbürgermeister[3]	Karte
1	Mitte	0101 Mitte 0102 Moabit 0103 Hansaviertel 0104 Tiergarten 0105 Wedding 0106 Gesundbrunnen	39,47	385.748	9.773	Stephan von Dassel (Grüne)	Lage
2	Friedrichshain-Kreuzberg	0201 Friedrichshain 0202 Kreuzberg	20,16	290.386	14.404	Monika Herrmann (Grüne)	Lage
3	Pankow	0301 Prenzlauer Berg 0302 Weißensee 0303 Blankenburg 0304 Heinersdorf 0305 Karow 0306 Stadtrandsiedlung Malchow 0307 Pankow 0308 Blankenfelde 0309 Buch 0310 Französisch Buchholz 0311 Niederschönhausen 0312 Rosenthal 0313 Wilhelmsruh	103,01	409.335	3.974	Sören Benn (Die Linke)	Lage
		0401 Charlottenburg 0402 Wilmersdorf					Lage

3) Paris: https://nl.wikipedia.org/wiki/Lijst_van_wijken_in_Parijs



Lijst van wijken in Parijs

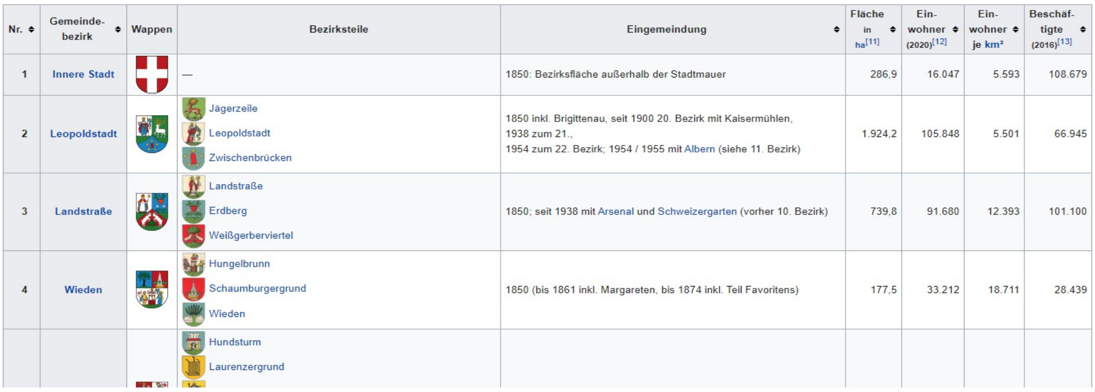
De stad Parijs is verdeeld in 20 arrondissementen die weer zijn verdeeld in totaal 80 wijken. Deze wijken worden in het Frans **Quartier** genoemd.

Parijs [bewerken | brontekst bewerken]

Parijs is ondergebracht in het departement Parijs dat is verdeeld in twintig arrondissementen van Parijs. Deze arrondissementen bestaan weer uit meestal vier quartiers. Er zijn in totaal 80 *quartiers* in Parijs.

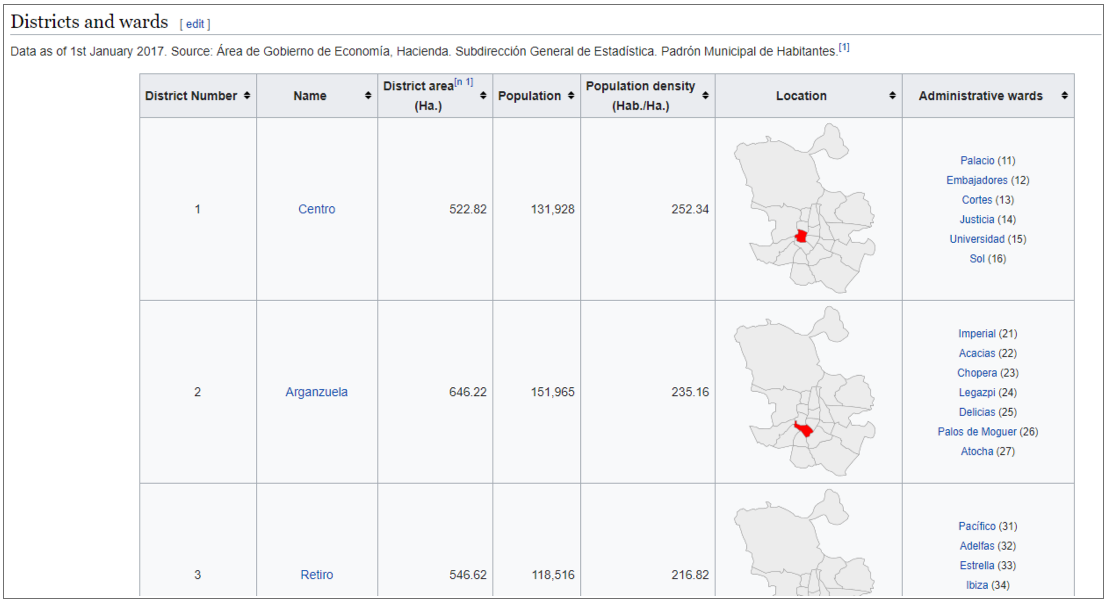
Arrondissement	Wijknummer	Naam	Kaart
1e arrondissement	1	Saint-Germain-l'Auxerrois	
	2	Les Halles	
	3	Palais-Royal	
	4	Place-Vendôme	
2e arrondissement	5	Gaillon	
	6	Vivienne	
	7	Mail	
	8	Bonne-Nouvelle	
3e arrondissement	9	Arts-et-Métiers	
	10	Enfants-Rouges	
	11	Archives	
	12	Quartier Sainte-Avoye	

4) Vienna: https://de.wikipedia.org/wiki/Wiener_Gemeindebezirke



Nr.	Gemeinde-bezirk	Wappen	Bezirksteile	Eingemeindung	Fläche in ha[11]	Ein-wohner (2020)[12]	Ein-wohner je km²	Beschäf-tigte (2016)[13]
1	Innere Stadt		—	1850: Bezirksfläche außerhalb der Stadtmauer	286,9	16.047	5.593	108.679
2	Leopoldstadt		Jägerzeile Leopoldstadt Zwischenbrücken	1850 inkl. Brigittenau, seit 1900 20. Bezirk mit Kaisermühlen, 1938 zum 21., 1954 zum 22. Bezirk; 1954 / 1955 mit Albern (siehe 11. Bezirk)	1.924,2	105.848	5.501	66.945
3	Landstraße		Landstraße Erdberg Weißgerberviertel	1850; seit 1938 mit Arsenal und Schweizergarten (vorher 10. Bezirk)	739,8	91.680	12.393	101.100
4	Wieden		Hungelbrunn Schaumburgergrund Wieden	1850 (bis 1861 inkl. Margareten, bis 1874 inkl. Teil Favoritens)	177,5	33.212	18.711	28.439
			Hundsturm Laurenzergrund					

5) Madrid: https://en.wikipedia.org/wiki/Districts_of_Madrid

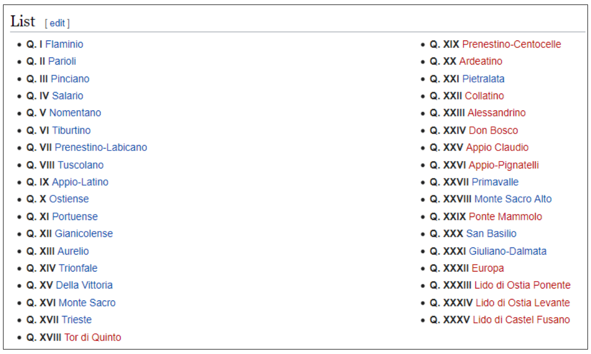


Districts and wards [edit]

Data as of 1st January 2017. Source: Área de Gobierno de Economía, Hacienda. Subdirección General de Estadística. Padrón Municipal de Habitantes.[1]

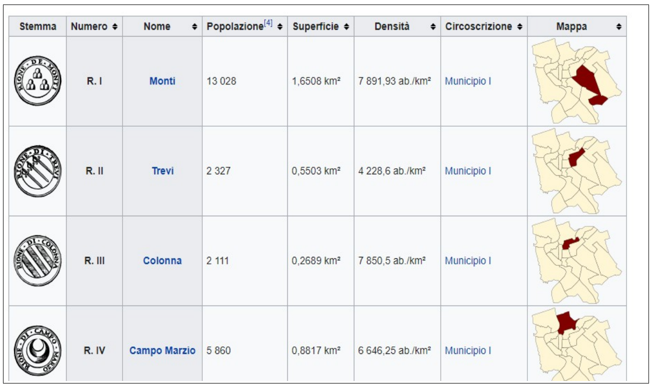
District Number	Name	District area[n 1] (Ha.)	Population	Population density (Hab./Ha.)	Location	Administrative wards
1	Centro	522.82	131,928	252.34		Palacio (11) Embajadores (12) Cortes (13) Justicia (14) Universidad (15) Sol (16)
2	Arganzuela	646.22	151,965	235.16		Imperial (21) Acacias (22) Chopera (23) Legazpi (24) Delicias (25) Palos de Moguer (26) Atocha (27)
3	Retiro	546.62	118,516	216.82		Pacífico (31) Adelfas (32) Estrella (33) Ibiza (34)

6) Rome: https://en.wikipedia.org/wiki/Quarters_of_Rome en https://it.wikipedia.org/wiki/Rioni_di_Roma



List [edit]

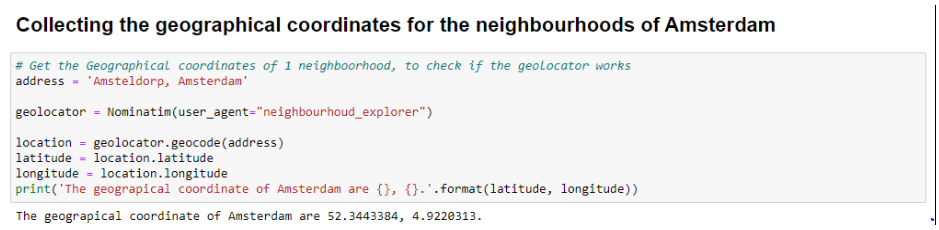
- **Q. I** Flaminio
- **Q. II** Parioli
- **Q. III** Pinciano
- **Q. IV** Salario
- **Q. V** Nomentano
- **Q. VI** Tiburtino
- **Q. VII** Prenestino-Labicano
- **Q. VIII** Tuscolano
- **Q. IX** Appio-Latino
- **Q. X** Ostiense
- **Q. XI** Portuense
- **Q. XII** Gianicolense
- **Q. XIII** Aurelio
- **Q. XIV** Trionfale
- **Q. XV** Della Vittoria
- **Q. XVI** Monte Sacro
- **Q. XVII** Trieste
- **Q. XVIII** Tor di Quinto
- **Q. XIX** Prenestino-Centocelle
- **Q. XX** Ardeatino
- **Q. XXI** Pietralata
- **Q. XXII** Collatino
- **Q. XXIII** Alessandrino
- **Q. XXIV** Don Bosco
- **Q. XXV** Appio Claudio
- **Q. XXVI** Appio-Pignatelli
- **Q. XXVII** Primavalle
- **Q. XXVIII** Monte Sacro Alto
- **Q. XXIX** Ponte Mammolo
- **Q. XXX** San Basilio
- **Q. XXXI** Giuliano-Dalmata
- **Q. XXXII** Europa
- **Q. XXXIII** Lido di Ostia Ponente
- **Q. XXXIV** Lido di Ostia Levante
- **Q. XXXV** Lido di Castel Fusano



Stemma	Numero	Nome	Popolazione[4]	Superficie	Densità	Circoscrizione	Mappa
	R. I	Monti	13 028	1,6508 km²	7 891,93 ab./km²	Municipio I	
	R. II	Trevi	2 327	0,5503 km²	4 228,6 ab./km²	Municipio I	
	R. III	Colonna	2 111	0,2689 km²	7 850,5 ab./km²	Municipio I	
	R. IV	Campo Marzio	5 860	0,8817 km²	6 646,25 ab./km²	Municipio I	

Geographical data

The geographical data for the neighborhoods will be collected with python's package Geopy, which can convert an address into latitude and longitude values (see example).



Collecting the geographical coordinates for the neighbourhoods of Amsterdam

```
# Get the Geographical coordinates of 1 neighboorhood, to check if the geolocator works
address = 'Amsteldorp, Amsterdam'

geolocator = Nominatim(user_agent="neighbourhoud_explorer")

location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Amsterdam are {}, {}.'.format(latitude, longitude))
```

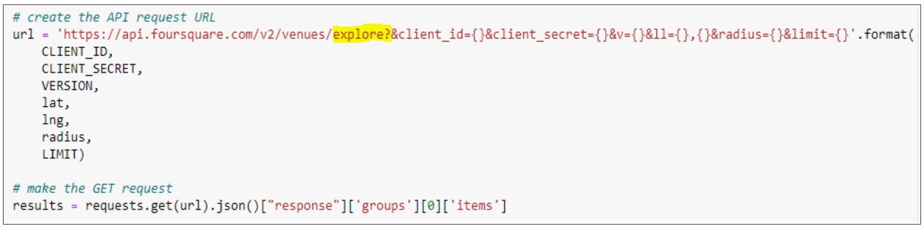
```
The geograpical coordinate of Amsterdam are 52.3443384, 4.9220313.
```

Neighborhoods without geographical coordinates will be dropped. The rest will be saved in a CSV document for later use.

Venue/Attraction data

Secondly we need to collect the attractions/venues of the neighborhoods. There are many location providers for instance Foursquare, Google places and Yelp. But in this project we will use Foursquare, because it's a part of the assignment. Foursquare is a US tech company from New York focusing on location data. Their technology and data is used by a lot of worldwide companies, for instance Uber, Twitter, etc.

One of the things they offer is an straightforward and easy to use API, which is recommended by a lot af people (see url on the left). The data obtained from this API in combination with the neighborhood information is going to give us the venue information we need. This information includes the name, ID, location , category, etc. of the venues. The category of the venues is going to be used to find the similarities/dissimilarities between the cities.

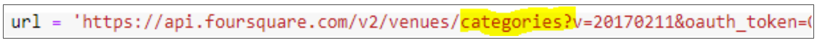


```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    radius,
    LIMIT)

# make the GET request
results = requests.get(url).json()["response"]['groups'][0]['items']
```

Venue category data

The venue categories in Foursquare are hierarchically ordered. For instance the main category “Food” consists of subcategory “Restaurant”, which again has subcategories like “French restaurant”, “Italian restaurant” and so on. This category tree can also be obtained with the Foursquare API (see url on the left). The category data is also saved in a CSV document and when necessary will be joined with the neighborhood data.



```
url = 'https://api.foursquare.com/v2/venues/categories?v=20170211&oauth_token=(
```