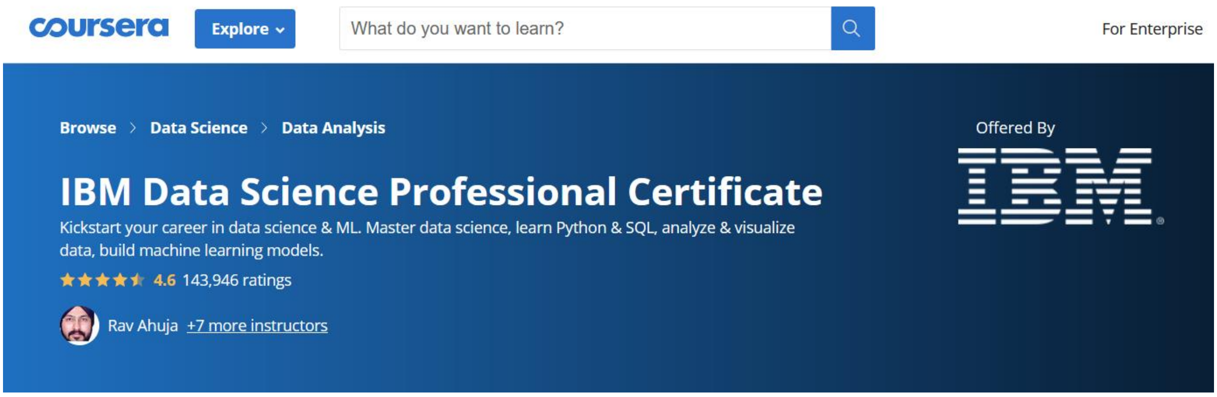


coursera
Explore
What do you want to learn?
For Enterprise
Browse > Data Science > Data Analysis
Offered By
IBM

IBM Data Science Professional Certificate

Kickstart your career in data science & ML. Master data science, learn Python & SQL, analyze & visualize data, build machine learning models.

★★★★☆ 4.6 143,946 ratings

Rav Ahuja +7 more instructors

IBM Data Science Professional Certificate

Capstone Project: The Battle of Neighbourhoods

Cologne: Where would I go when I am hungry

The Hunger Games in Cologne

Author: Vladyslav Borysenko

List of Figures

1. Introduction

1.1.Preamble

While reading this paper or article you might have noticed it is all about Cologne. Why? The reason is simple. At the moment of writing the paper I live in Schweinfurt, a small city in Germany in a province called Bavaria. Schweinfurt is a city that has a lot of firms, plants and manufacturing companies such as Schaeffler, ZF, Bosch, SKF, etc. Hence, there are not so many places for recreation and entertainment. Those that are here are good, but their low numbers make it difficult to practice clustering in Schweinfurt. You might say I could have tried to cluster plants, to see which one produces what. I could have, yet the project requirement is to use Foursquare's API which is not a perfect solution for clustering production plants at all. So, this is how I came up with an idea to focus the project on Cologne.

As my best friend lives and Cologne and I have visited it a few times, we have encountered an issue where we would wander around Cologne and, obviously, get hungry. However, Cologne is a huge city and even my friend does not know it that well. After, having our meals at a few random places, I have decided to use a more structured approach next time I go to Cologne.

Also, while doing the research I have found that this kind of project was already performed for Cologne (Johannes Wagner, 2020), yet I have attempted my own, to get results with the fresh data.

1.2.Introduction

As it was mentioned Cologne is a big city and has a lot of venues and places to go. This fact does not make it easy for tourists to make choices where to go and have their meal. And, believe me, tourists do it a lot. For many people it is the very reason the do tourism in the first place. Many people come to Cologne from different places and countries wanting to try various or, sometimes, a very particular types of cuisines.

But not only that, but often even locals want to try something new and travel to the other neighbourhood just to visit some places their friends were talking about or simply explore what is there.

Having these concerns in mind, I have basically narrowed it all down to a single question: "If one were to go to that part of Cologne, which kind of food places would they find there?"

Well, now it looks like a perfect case to apply some clustering. Hence, let the Cologne Hunger Games begin!

2. Data

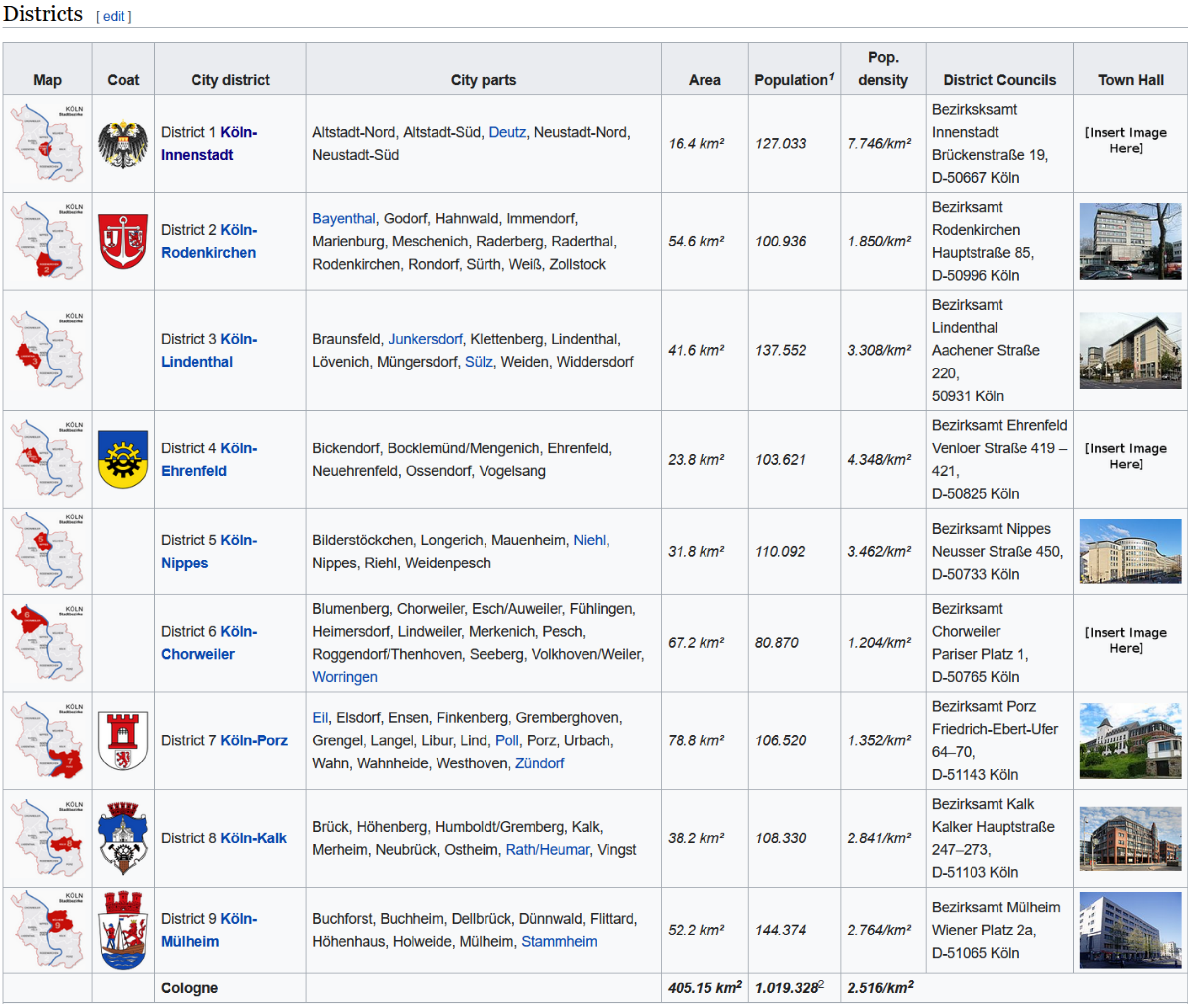
The data was collected from three main sources, it was then cleaned, organised and connected/joined to work in conjunction. In parallel, the resulting data frames were observed and analysed.

2.1.Data acquisition

The first step was to obtain the data on the neighbourhoods of Cologne. Luckily, there were many options available: the public Wikipedia, the official website of the City of Cologne, as well as some other websites that specialise on collecting data. Yet, the easiest way was to use the Wikipedia for some basic information about the districts.

2.1.1.Cologne Wikipedia

It was found out that Cologne has 9 (nine) main districts (Figure 2-1).



Districts [edit]

Map	Coat	City district	City parts	Area	Population[1]	Pop. density	District Councils	Town Hall
		District 1 **Köln-Innenstadt**	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-Nord, Neustadt-Süd	*16.4 km^2*	*127.033*	*7.746/km^2*	Bezirksamt Innenstadt Brückenstraße 19, D-50667 Köln	[Insert Image Here]
		District 2 **Köln-Rodenkirchen**	Bayenthal, Godorf, Hahnwald, Immendorf, Marienburg, Meschenich, Raderberg, Raderthal, Rodenkirchen, Rondorf, Sürth, Weiß, Zollstock	*54.6 km^2*	*100.936*	*1.850/km^2*	Bezirksamt Rodenkirchen Hauptstraße 85, D-50996 Köln	
		District 3 **Köln-Lindenthal**	Braunsfeld, Junkersdorf, Klettenberg, Lindenthal, Lövenich, Müngersdorf, Sülz, Weiden, Widdersdorf	*41.6 km^2*	*137.552*	*3.308/km^2*	Bezirksamt Lindenthal Aachener Straße 220, 50931 Köln	
		District 4 **Köln-Ehrenfeld**	Bickendorf, Bocklemünd/Mengenich, Ehrenfeld, Neuehrenfeld, Ossendorf, Vogelsang	*23.8 km^2*	*103.621*	*4.348/km^2*	Bezirksamt Ehrenfeld Venloer Straße 419 – 421, D-50825 Köln	[Insert Image Here]
		District 5 **Köln-Nippes**	Bilderstöckchen, Longerich, Mauenheim, Niehl, Nippes, Riehl, Weidenpesch	*31.8 km^2*	*110.092*	*3.462/km^2*	Bezirksamt Nippes Neusser Straße 450, D-50733 Köln	
		District 6 **Köln-Chorweiler**	Blumenberg, Chorweiler, Esch/Auweiler, Fühlingen, Heimersdorf, Lindweiler, Merkenich, Pesch, Roggendorf/Thenhoven, Seeberg, Volkhoven/Weiler, Worringen	*67.2 km^2*	*80.870*	*1.204/km^2*	Bezirksamt Chorweiler Pariser Platz 1, D-50765 Köln	[Insert Image Here]
		District 7 **Köln-Porz**	Eil, Elsdorf, Ensen, Finkenberg, Gremberghoven, Grengel, Langel, Libur, Lind, Poll, Porz, Urbach, Wahn, Wahnheide, Westhoven, Zündorf	*78.8 km^2*	*106.520*	*1.352/km^2*	Bezirksamt Porz Friedrich-Ebert-Ufer 64–70, D-51143 Köln	
		District 8 **Köln-Kalk**	Brück, Höhenberg, Humboldt/Gremberg, Kalk, Merheim, Neubrück, Ostheim, Rath/Heumar, Vingst	*38.2 km^2*	*108.330*	*2.841/km^2*	Bezirksamt Kalk Kalker Hauptstraße 247–273, D-51103 Köln	
		District 9 **Köln-Mülheim**	Buchforst, Buchheim, Dellbrück, Dünnwald, Flittard, Höhenhaus, Holweide, Mülheim, Stammheim	*52.2 km^2*	*144.374*	*2.764/km^2*	Bezirksamt Mülheim Wiener Platz 2a, D-51065 Köln	
		Cologne		***405.15 km^2***	***1.019.328***[2]	***2.516/km^2***		

Figure 2-1: Wikipedia article with Cologne's districts. Source: (Anon, 2020)

For better understanding Figure 2-2 shows the map that visualises the districts.

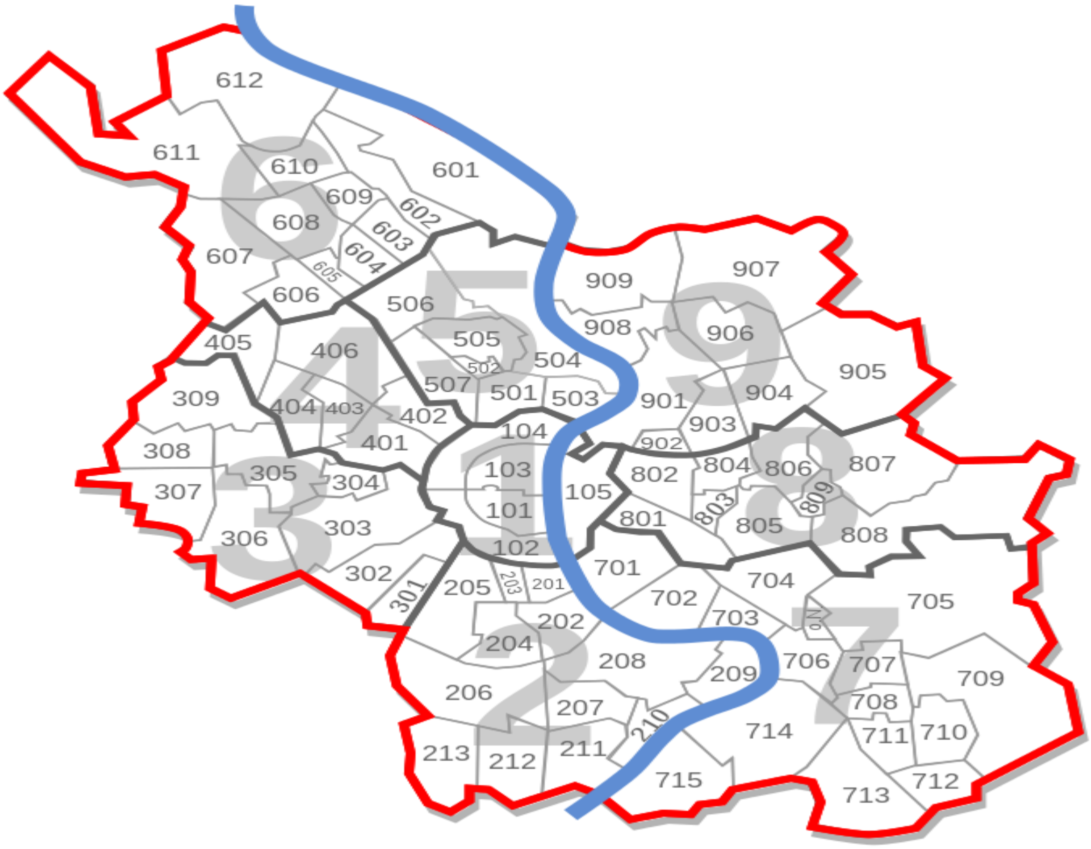
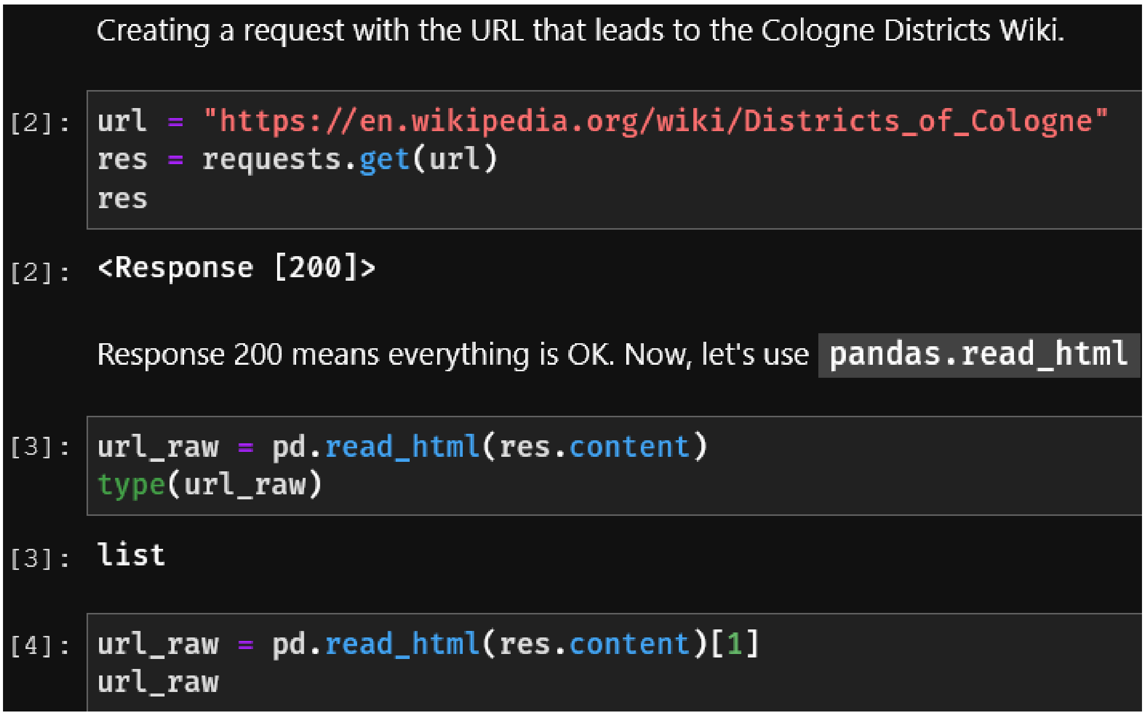


Figure 2-2: Cologne districts map. Source: (Anon, 2020)

With this all in hand it is easy to extract information from the Wikipedia using 'pandas.read_html' function (Figure 2-3).



Creating a request with the URL that leads to the Cologne Districts Wiki.

```
[2]: url = "https://en.wikipedia.org/wiki/Districts_of_Cologne"
     res = requests.get(url)
     res

[2]: <Response [200]>
```

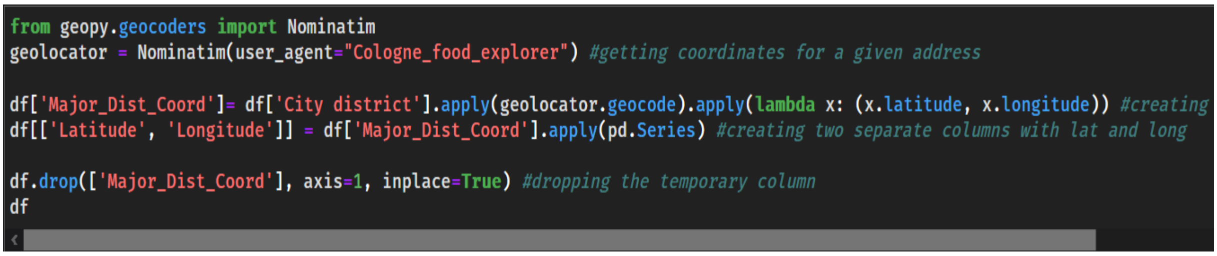
Response 200 means everything is OK. Now, let's use `pandas.read_html`

```
[3]: url_raw = pd.read_html(res.content)
     type(url_raw)

[3]: list

[4]: url_raw = pd.read_html(res.content)[1]
     url_raw
```

Figure 2-3: Extracting information from the Wikipedia.

2.1.2.Geocoding

The next data obtained was the geographical coordinates of Cologne and its districts. I gave another chance to geopy module and this time it worked perfectly, though producing the working code required some googling and efforts (Figure 2-4).

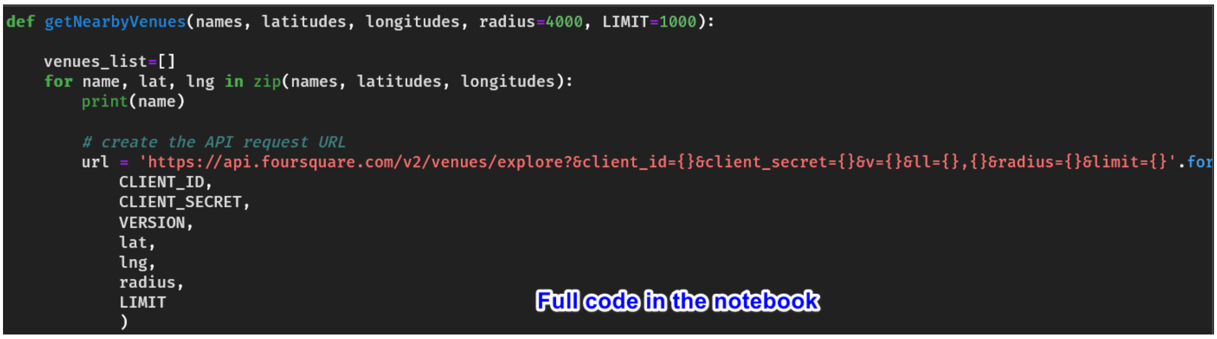


```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="Cologne_food_explorer") #getting coordinates for a given address

df['Major_Dist_Coord']= df['City district'].apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude)) #creating
df[['Latitude', 'Longitude']] = df['Major_Dist_Coord'].apply(pd.Series) #creating two separate columns with lat and long

df.drop(['Major_Dist_Coord'], axis=1, inplace=True) #dropping the temporary column
df
```

Figure 2-4: Using Geopy to obtain geographical coordinates.

2.1.3.Venues using Foursquare API

Finally, the last piece of data came with Foursquare API. Just like in previous labs, the request was created (actually multiple requests) and the list of venues was composed and put into a data frame (Figure 2-5).



```
def getNearbyVenues(names, latitudes, longitudes, radius=4000, LIMIT=1000):

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

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

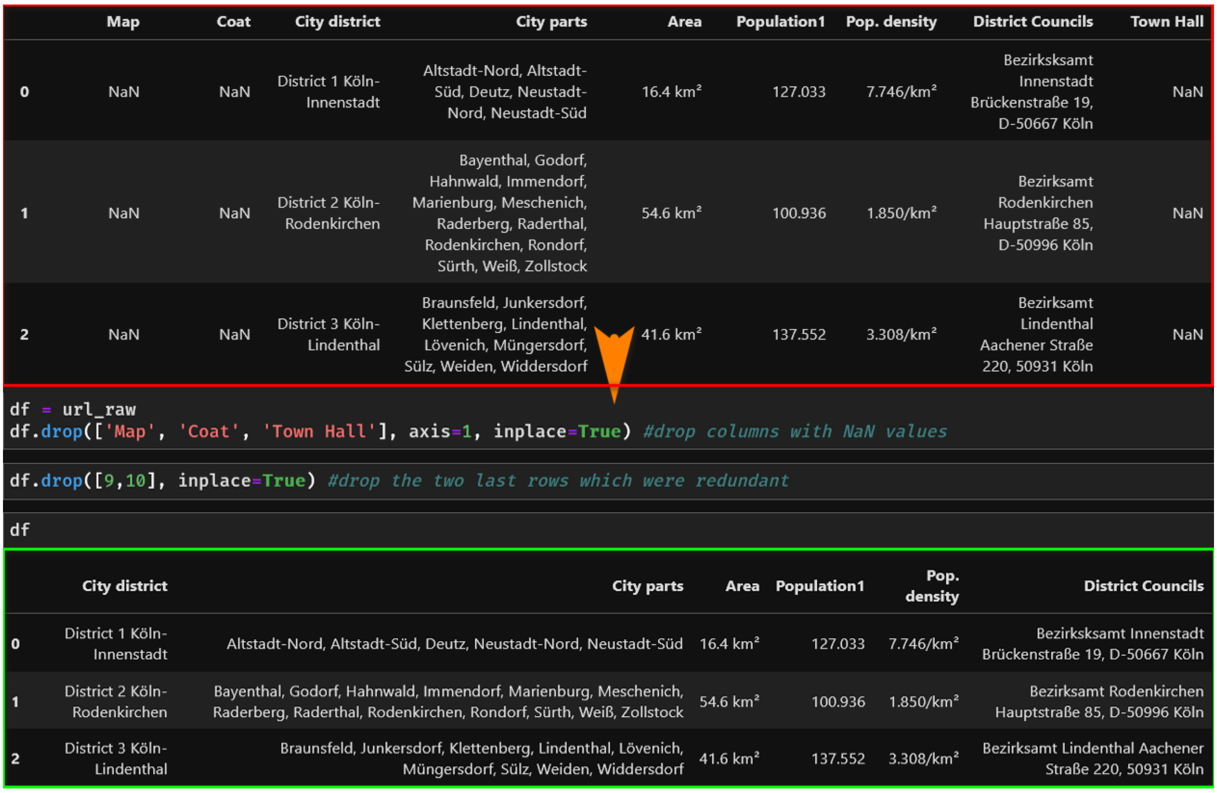
Figure 2-5: Venue explorer.

2.2.Data cleaning

Some of the data came with NaN values and additional unnecessary information. Therefore, the dataset was cleaned and redundant values removed (Figure 2-6).

This was achieved using pandas “drop” function with the “inplace” argument equal to “True”.

Later on, the final data frame was checked for null values (Figure 2-7) to determine whether further cleaning was necessary, but, luckily, no NaN values were found. In this case, data cleaning did not take much time as the data was structured pretty well and, also, there was not that much data so that missing values would be of a great concern. As our research was a product of curiosity, we also had some degree of freedom when working with the data. Yet, no missing data was found.



	Map	Coat	City district	City parts	Area	Population1	Pop. density	District Councils	Town Hall
0	NaN	NaN	District 1 Köln-Innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-Nord, Neustadt-Süd	16.4 km²	127.033	7.746/km²	Bezirksksamt Innenstadt Brückenstraße 19, D-50667 Köln	NaN
1	NaN	NaN	District 2 Köln-Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marienburg, Meschenich, Raderberg, Raderthal, Rodenkirchen, Rondorf, Sürth, Weiß, Zollstock	54.6 km²	100.936	1.850/km²	Bezirksamt Rodenkirchen Hauptstraße 85, D-50996 Köln	NaN
2	NaN	NaN	District 3 Köln-Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenthal, Lövenich, Müngersdorf, Sülz, Weiden, Widdersdorf	41.6 km²	137.552	3.308/km²	Bezirksamt Lindenthal Aachener Straße 220, 50931 Köln	NaN

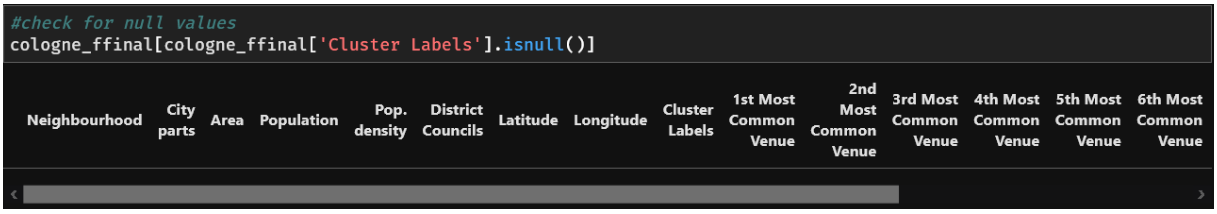
```
df = url_raw
df.drop(['Map', 'Coat', 'Town Hall'], axis=1, inplace=True) #drop columns with NaN values
```

```
df.drop([9,10], inplace=True) #drop the two last rows which were redundant
```

```
df
```

	City district	City parts	Area	Population1	Pop. density	District Councils
0	District 1 Köln-Innenstadt	Altstadt-Nord, Altstadt-Süd, Deutz, Neustadt-Nord, Neustadt-Süd	16.4 km²	127.033	7.746/km²	Bezirksksamt Innenstadt Brückenstraße 19, D-50667 Köln
1	District 2 Köln-Rodenkirchen	Bayenthal, Godorf, Hahnwald, Immendorf, Marienburg, Meschenich, Raderberg, Raderthal, Rodenkirchen, Rondorf, Sürth, Weiß, Zollstock	54.6 km²	100.936	1.850/km²	Bezirksamt Rodenkirchen Hauptstraße 85, D-50996 Köln
2	District 3 Köln-Lindenthal	Braunsfeld, Junkersdorf, Klettenberg, Lindenthal, Lövenich, Müngersdorf, Sülz, Weiden, Widdersdorf	41.6 km²	137.552	3.308/km²	Bezirksamt Lindenthal Aachener Straße 220, 50931 Köln

Figure 2-6: Removing redundant values.



```
#check for null values
cologne_ffinal[cologne_ffinal['Cluster Labels'].isnull()]
```

Neighbourhood	City parts	Area	Population	Pop. density	District Councils	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue

Figure 2-7: Missing data check.