

## Data tells where to build new hotels in Berlin

### II) Dataset description

The data required to solve the problem is the Foursquare data for interesting places in Berlin, Germany.

We first choose the city center as our starting point. From this point, we will explore venues within the radius as 10,000m (as *Assumption 3*).

We need to use Foursquare API to get the venues data within the radius as 10,000m from Berlin city center. There are several parameters which we can pass into the calling URL, so we need to modify to meet our need. As previously analyzed, tourists were those who visited places in Berlin and provide ratings, we will focus on these places. We will pass **query=tourist** into the URL to filter these values which will helpful for other tourists.

The dataset we receive contain of the list of interesting places in Berlin

	address	categories	cc	city	country	crossStreet	distance	formattedAddress	id	labeledLatLngs	lat	lng	name	neighborhood	postalCode	state
0	Behrenstr. 55-57	Opera House	DE	Berlin	Deutschland	NaN	188	[Behrenstr. 55-57, 10117 Berlin, Deutschland]	4adccda8af964a520bd4921e3	[{"label": "display", "lat": 52.51596828902978...	52.515968	13.386701	Komische Oper	Unter den Linden	10117	Berlin
1	Gendarmenmarkt	Concert Hall	DE	Berlin	Deutschland	Charlottenstr.	429	[Gendarmenmarkt (Charlottenstr.), 10117 Berlin...	4adccda7cf964a520584721e3	[{"label": "display", "lat": 52.513652, "lng":...	52.513652	13.391911	Konzerthaus Berlin	NaN	10117	Berlin
2	Markgrafenstr.	Plaza	DE	Berlin	Deutschland	Mohrenstr.	466	[Markgrafenstr. (Mohrenstr.), 10117 Berlin, De...	4adccda7df964a5206a4721e3	[{"label": "display", "lat": 52.51357005399756...	52.513570	13.392720	Gendarmenmarkt	NaN	10117	Berlin
3	Bebelplatz	Plaza	DE	Berlin	Deutschland	Unter den Linden	342	[Bebelplatz (Unter den Linden), 10117 Berlin, ...	4adccda7df964a520ab4721e3	[{"label": "display", "lat": 52.51653012045096...	52.516530	13.393847	Bebelplatz	NaN	10117	Berlin
4	Am Festungsgraben 1	Theater	DE	Berlin	Deutschland	NaN	481	[Am Festungsgraben 1, 10117 Berlin, Deutschland]	4adccda89f964a520954921e3	[{"label": "display", "lat": 52.51894077519612...	52.518941	13.395245	Maxim Gorki Theater	NaN	10117	Berlin

As we use the Foursquare API sandbox account, there is no rating of the place in the returned dataset yet. In order to obtain the ratings, we need to pass the ID of each location to URL, and call the Foursquare API. Since Foursquare provides also the number of likes for each venue, we collect this data in the call, too.

The consolidated dataset contains the aforementioned dataset, together with the rating and number of likes for each venue .

	address	categories	cc	city	country	crossStreet	distance	formattedAddress	id	labeledLatLngs	lat	lng	name	neighborhood	postalCode	state	no_of_likes	rating
0	Behrenstr. 55-57	Opera House	DE	Berlin	Deutschland		188	[Behrenstr. 55-57, 10117 Berlin, Deutschland]	4adccda8af964a520bd4921e3	[[{"label": "display", "lat": 52.51596828902978..., "lng": 13.386701}]]	52.515968	13.386701	Komische Oper	Unter den Linden	10117	Berlin	182.0	8.7
1	Gendarmenmarkt	Concert Hall	DE	Berlin	Deutschland	Charlottenstr.	429	[Gendarmenmarkt (Charlottenstr.), 10117 Berlin, Deutschland]	4adccda7cf964a520584721e3	[[{"label": "display", "lat": 52.513652, "lng": 13.391911}]]	52.513652	13.391911	Konzerthaus Berlin	NaN	10117	Berlin	327.0	9.2
2	Markgrafenstr.	Plaza	DE	Berlin	Deutschland	Mohrenstr.	466	[Markgrafenstr. (Mohrenstr.), 10117 Berlin, Deutschland]	4adccda7df964a5206a4721e3	[[{"label": "display", "lat": 52.51357005399756..., "lng": 13.392720}]]	52.513570	13.392720	Gendarmenmarkt	NaN	10117	Berlin	1293.0	9.4
3	Bebelplatz	Plaza	DE	Berlin	Deutschland	Unter den Linden	342	[Bebelplatz (Unter den Linden), 10117 Berlin, Deutschland]	4adccda7df964a520ab4721e3	[[{"label": "display", "lat": 52.51653012045096..., "lng": 13.393847}]]	52.516530	13.393847	Bebelplatz	NaN	10117	Berlin	176.0	8.8
4	Am Festungsgraben 1	Theater	DE	Berlin	Deutschland		481	[Am Festungsgraben 1, 10117 Berlin, Deutschland]	4adccda89f964a520954921e3	[[{"label": "display", "lat": 52.51894077519612..., "lng": 13.395245}]]	52.518941	13.395245	Maxim Gorki Theater	NaN	10117	Berlin	134.0	9.1

As we only focus on areas which have interesting places for tourists, we will drop all data rows that have ratings < 7.5 (out of 10). The remaining data is the dataset which we will use to solve the problem