Best Neighborhoods to Stay in Belo Horizonte, BRA

Author: David Guimarães Rocha, for the Coursera IBM Data Science Capstone Project.

Where to Stay? The question every traveler asks himself.

I was born in a great city in Brazil. It is called Belo Horizonte, the capital of Minas Gerais State. It is a city with great venues and touristic attractions and, it is known for the variety of amazing food, bars and great night life.

My friends from other places always ask me for a recommendation of a neighborhood to stay when going to Belo Horizonte, and it is always really hard to just chose one based on my experience. I was never a tourist in my own city. The best neighborhood to stay depends a lot on the visitors tastes and motivations for the trip.

So, how to recommend a Neighborhood to stay, depending on the visitors taste and objective? How to have the best information in order to provide that recommendation? Those are relevant information for people as well as for business. A tourism company can use such type information to provide better service for their costumers and increase profit. It is also important for the government to publicize these kinds of information that can be worth a lot to the citizens and to companies.

I started to search for quality data to back up my recommendations. I wanted some information that could relate to a visitor’s needs, such as infrastructure of a neighborhood, access to services, entertainment, etc.

Data

While surfing the City Council's website, I found data regarding an index on quality of life in urban centers. The Belo Horizonte Urban Quality of Life Index (IQVU-BH) is an index composed of several variables - such as, for example, urban infrastructure, security and education - that seek to quantify the availability of public and private goods and services in the city. The IQVU is calculated for the 80 Planning Units (UPs) in Belo Horizonte (neighborhoods).

Its calculation allows the delimitation of priority areas for public investments and a better understanding of the distribution of public and private goods and services between the regions of the city.

That IQVU provides a good starting point, for it supplies the knowledge of which neighborhoods have good infrastructure and access to services, things that are essential for a tourist.

The second part of the data is the one provided by Foursquare API. Once in possession of the best neighborhoods given by the IQVU index, the Foursquare data provides the information about the venues located in that neighborhood, the access to restaurants, bars, nightlife and other meaningful insights on the neighborhoods.

Neighborhoods and IQVU data gathering, transformation and analysis.

The tables about neighborhoods and IQVU for each neighborhood of the city of Belo Horizonte were retrieved at the City Council's website, specifically at their open data page (https://dados.pbh.gov.br/) and their geographic data page (http://bhmap.pbh.gov.br/).

Figure 1 – Neighborhoods base data frame with all the city's neighborhoods.

	FID	ID	CODIGO	NOME	AREA_KM2	PERIMETR_M	GEOMETRIA
0	BAIRRO.264	264	1257	Corumbiara	0.054	1253.896	MULTIPOLYGON (((603094.63678959 7787345.199450...
1	BAIRRO.507	507	1674	Vila Betânia	0.020	1547.980	MULTIPOLYGON (((604920.67306699 7792251.346766...
2	BAIRRO.510	510	1668	Ambrosina	0.055	1339.865	MULTIPOLYGON (((606432.210916944 7795405.36142...
3	BAIRRO.619	619	1976	Antônio Ribeiro de Abreu	0.059	2711.485	MULTIPOLYGON (((614811.833402637 7806598.93172...
4	BAIRRO.194	194	819	Vista Alegre	0.459	3986.841	MULTIPOLYGON (((604406.871931668 7793745.33072...

Through the geocoder library, the latitudes and longitudes for each neighborhood were added. For the neighborhoods that the geocoder was not able to find a location, their rows were dropped. Then the data frame was cleaned to result only in the columns of interest.

Figure 2 – Clean neighborhoods data frame with latitudes and longitudes.

	ID	CODIGO	NOME	latitude	longitude
0	264	1257	Corumbiara	-20.0093	-44.0135
1	507	1674	Vila Betânia	-19.977	-44.0185
2	510	1668	Ambrosina	-19.9313	-43.9828
3	619	1976	Antônio Ribeiro de Abreu	-19.8291	-43.8985
4	194	819	Vista Alegre	-19.955	-43.999

The IQVU base data frame was loaded and sorted by the IQVU value in descending order, to present the neighborhoods with the best IQVU's scores. For the purpose of this project, the top 15 neighborhoods were selected and sorted by score.

Figure 3 – Top 15 neighborhoods per IQVU score.

	NOME	IQVU
0	Barro Preto	0,856586868
1	Santa Lúcia	0,824468066
2	São Bento	0,824468066
3	Savassi	0,804586698
4	Centro	0,798798365
5	Lagoa da Pampulha	0,798016108
6	Santa Inês	0,794606199
7	Santa Tereza	0,785570033
8	Floresta	0,785570033
9	Mangabeiras	0,783575181
10	Serra	0,775935589
11	Santo Antônio	0,775673299
12	Pompéia	0,775655666
13	Belvedere	0,769292717
14	Barroca	0,765196443

The top 5 neighborhoods are: Barro Preto, Santa Lúcia, São Bento, Savassi and Centro.

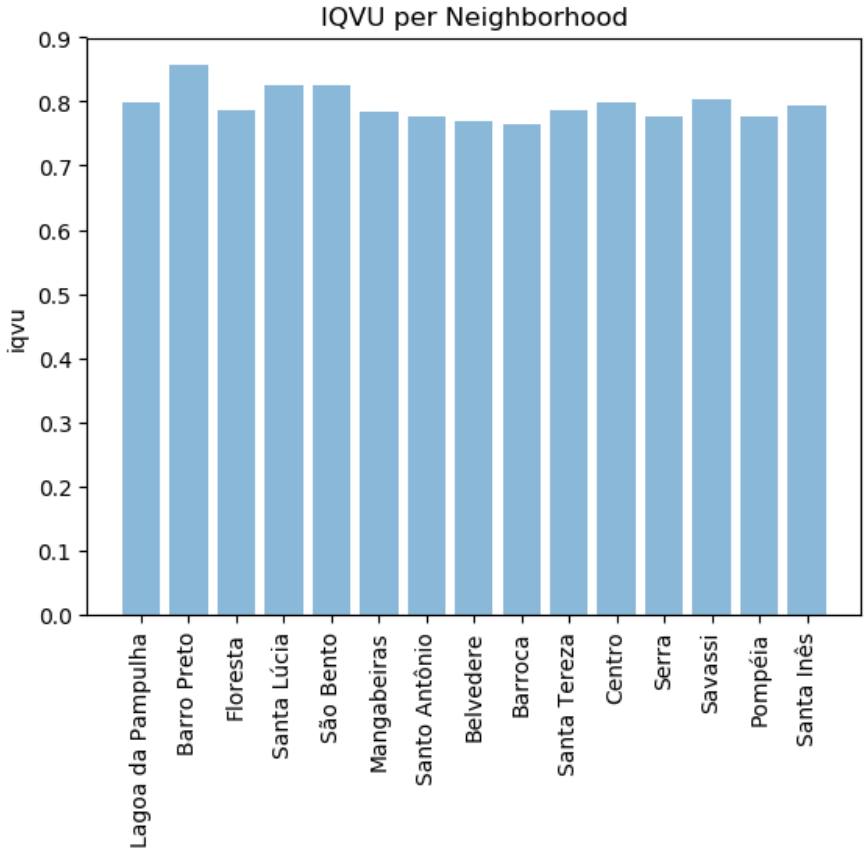
The IQVU table and the clean neighborhoods table, with the location, were filtered to a data frame containing all the information for the top 15 neighborhoods.

Figure 4 – Data frame with all the information of the top 15 neighborhoods.

	ID	CODIGO	NOME	latitude	longitude	iqvu
0	617	1972	Lagoa da Pampulha	-19.8532	-43.9777	0.798016
1	12	617	Barro Preto	-19.9234	-43.952	0.856587
2	68	679	Floresta	-19.9127	-43.9302	0.785570
3	144	768	Santa Lúcia	-19.9632	-43.9451	0.824468
4	155	779	São Bento	-19.9577	-43.9532	0.824468
5	101	718	Mangabeiras	-19.9554	-43.9116	0.783575
6	154	778	Santo Antônio	-19.9433	-43.9434	0.775673
7	13	618	Belvedere	-19.9803	-43.9466	0.769293
8	11	616	Barroca	-19.9337	-43.964	0.765196
9	150	774	Santa Tereza	-19.9149	-43.917	0.785570
10	37	642	Centro	-19.9193	-43.9403	0.798798
11	172	797	Serra	-19.9406	-43.9184	0.775936
12	217	1133	Savassi	-19.9352	-43.9344	0.804587
13	131	751	Pompéia	-19.9096	-43.9024	0.775656
14	143	767	Santa Inês	-19.8903	-43.9097	0.794606

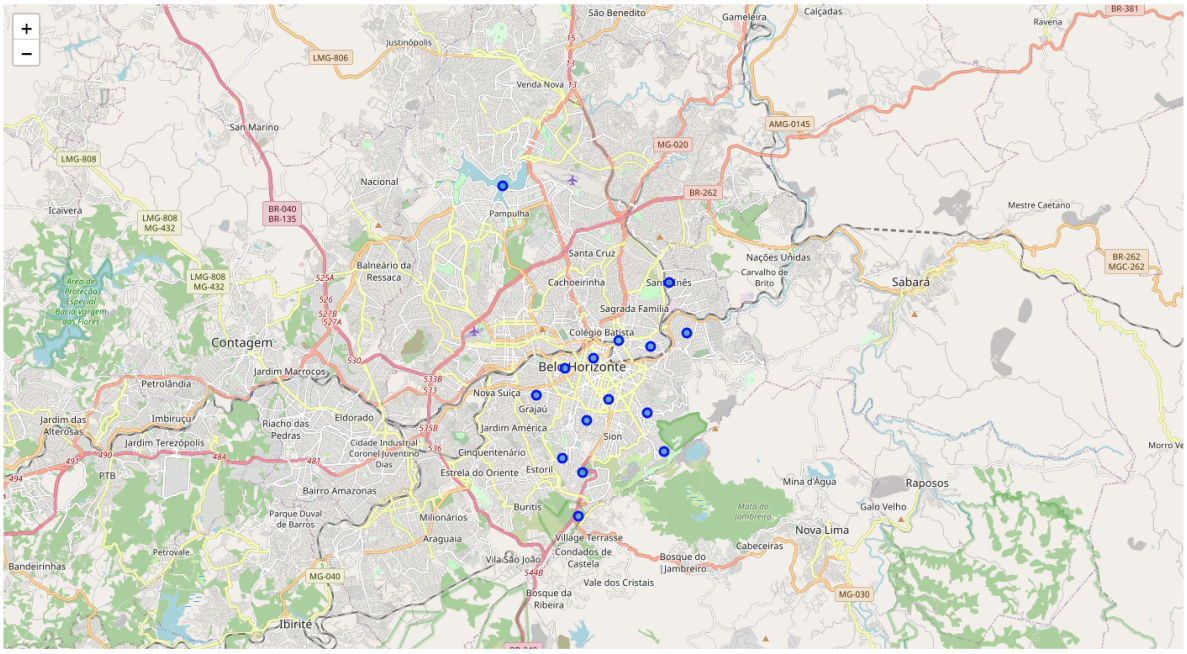
This bar chart shows the IQVU score for each neighborhood. As the IQVU index is between 0 and 1, the difference seems small, though it can represent a lot.

Figure 5 – Bar chart of IQVU scores per neighborhood.



The image bellow is the map of Belo Horizonte with the location of the 15 neighborhoods.

Figure 6 – Map of Belo Horizonte with the 15 neighborhoods with best IQVU score.



As we can see by the map, almost all of the neighborhoods are concentrated around the south-central region of the city, except the neighborhood of Lagoa da Pampulha.

Foursquare API dataset gathering, transformation and analysis

Once the dataset for the neighborhoods were processed, it was time to gather the data of the Foursquare API about venues. The API has restrictions on the types of call a developer can make with a “sandbox” account.

Considering that, it was decided to gather data about the different venues around the center location of the neighborhood in a ratio of 500 meters.

It is interesting to understand that, the most venues a neighborhood has around its center, the larger the concentration of venues are in that place. That is an important feature for a tourist that has no car and has to do everything by foot.

That said, the data for the venues around each neighborhood center was gathered.

Figure 7 – Venues for each neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category
0	Lagoa da Pampulha	-19.853209	-43.977749	Lagoa da Pampulha	4b69f01cf964a52025bd2be3	-19.854620	-43.979804	Lake
1	Lagoa da Pampulha	-19.853209	-43.977749	Pista de Corrida da Lagoa	4ec6edaf8b81dcfdc2cf5500	-19.855051	-43.976300	Racecourse
2	Lagoa da Pampulha	-19.853209	-43.977749	Casa Kubitschek	523b344a498e9f8caa6bf037	-19.854963	-43.980136	Museum
3	Lagoa da Pampulha	-19.853209	-43.977749	Guana Trainer - Lagoa da Pampulha	4e845b2a77c8a32bced089b0	-19.854170	-43.980451	Athletics & Sports
4	Lagoa da Pampulha	-19.853209	-43.977749	Iate Tênis Clube	4bcda17c937ca59309fdac92	-19.854157	-43.976040	Water Park

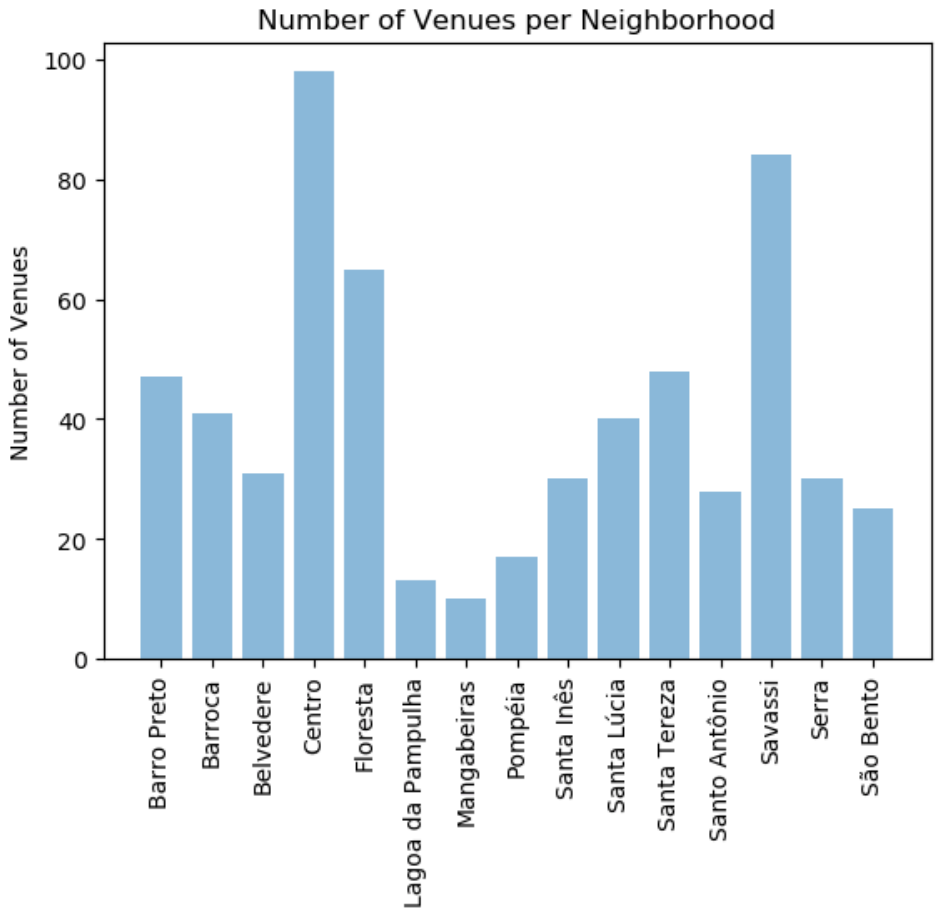
The first insight on the data is to look for the number of venues per neighborhood.

Figure 8 – Number of venues per neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category
0	Barro Preto	47	47	47	47	47	47	47
1	Barroca	41	41	41	41	41	41	41
2	Belvedere	31	31	31	31	31	31	31
3	Centro	98	98	98	98	98	98	98
4	Floresta	65	65	65	65	65	65	65
5	Lagoa da Pampulha	13	13	13	13	13	13	13
6	Mangabeiras	10	10	10	10	10	10	10
7	Pompéia	17	17	17	17	17	17	17
8	Santa Inês	30	30	30	30	30	30	30
9	Santa Lúcia	40	40	40	40	40	40	40
10	Santa Tereza	48	48	48	48	48	48	48
11	Santo Antônio	28	28	28	28	28	28	28
12	Savassi	84	84	84	84	84	84	84
13	Serra	30	30	30	30	30	30	30
14	São Bento	25	25	25	25	25	25	25

We can see the results in the bar chart below. The neighborhoods Centro, Savassi and Floresta have the most venues around their center.

Figure 9 – Quantity of venues per category in each neighborhood.



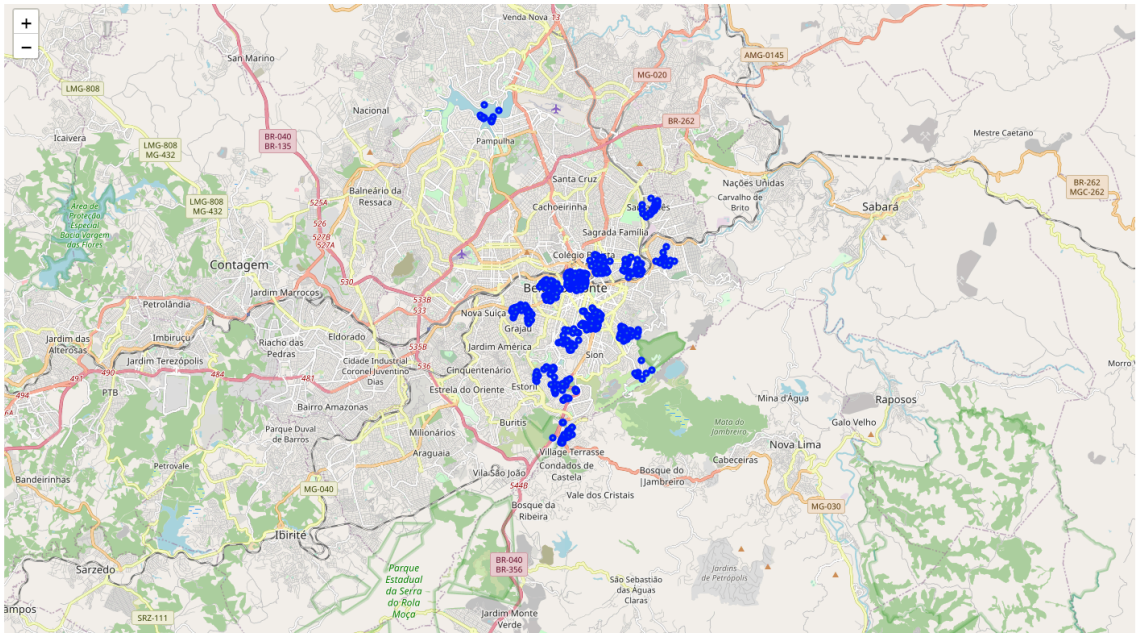
The data for the venues in each neighborhood was sorted by the venue's categories. This is a great way to see what types of venues there are in certain neighborhoods. If the visitor has a preference for a certain kind of food, or the need for a gym per example, it is possible to locate what neighborhoods provide such features.

Figure 10 – Bar chart of the number of venues per neighborhood.

		Neighborhood Latitude	Neighborhood Longitude	Venue	id	Venue Latitude	Venue Longitude
Neighborhood	Venue Category						
Barro Preto	ATM	1	1	1	1	1	1
	BBQ Joint	1	1	1	1	1	1
	Bakery	1	1	1	1	1	1
	Bar	1	1	1	1	1	1
	Brazilian Restaurant	6	6	6	6	6	6
	Buffet	1	1	1	1	1	1
	Burger Joint	1	1	1	1	1	1
	Café	2	2	2	2	2	2
	Chinese Restaurant	1	1	1	1	1	1
	Chocolate Shop	1	1	1	1	1	1

The map below shows all the venues for each neighborhood. The clusters are clearly defined, and it is possible to evaluate their density, which is a good indicator of how many venues and how close the venues are in each neighborhood.

Figure 11 – Map of the venues in each neighborhood.



Clustering the Neighborhoods by IQVU score and other insights on combining datasets.

The data showed above provide a good overview of a neighborhood feature set. In order to improve the analysis, clustering was performed utilizing their IQVU scores.
The algorithm utilized was k-means, an unsupervised machine learning clustering model that utilizes similarity measure to group objects. The results are shown below.

Figure 12 – Results of clustering by the IQVU score.

	ID	CODIGO	NOME	latitude	longitude	iqvu	Clus_km
0	617	1972	Lagoa da Pampulha	-19.853209	-43.977749	0.798016	2
1	12	617	Barro Preto	-19.923382	-43.952011	0.856587	3
2	68	679	Floresta	-19.912746	-43.930201	0.785570	0
3	144	768	Santa Lúcia	-19.963226	-43.945101	0.824468	1
4	155	779	São Bento	-19.957709	-43.953242	0.824468	1
5	101	718	Mangabeiras	-19.955434	-43.911609	0.783575	0
6	154	778	Santo Antônio	-19.943329	-43.943420	0.775673	0
7	13	618	Belvedere	-19.980275	-43.946570	0.769293	0
8	11	616	Barroca	-19.933743	-43.963996	0.765196	0
9	150	774	Santa Tereza	-19.914892	-43.917046	0.785570	0
10	37	642	Centro	-19.919331	-43.940349	0.798798	2
11	172	797	Serra	-19.940556	-43.918423	0.775936	0
12	217	1133	Savassi	-19.935220	-43.934446	0.804587	2
13	131	751	Pompéia	-19.909648	-43.902428	0.775656	0
14	143	767	Santa Inês	-19.890331	-43.909681	0.794606	2

- Only 1 neighborhood in cluster 3 (Barro Preto) -> highest IQVU score;
- 2 neighborhoods in cluster 1(Santa Lúcia and São Bento) -> very good IQVU scores;
- 4 neighborhoods in cluster 2 (Lagoa da Pampulha, Centro, Savassi and Santa Inês) -> good IQVU scores;
- 7 neighborhoods in cluster 0 (Floresta, Mangabeiras, Santo Antônio, Belvedere, Barroca, Santa Tereza, Serra and Pompéia) -> lowest IQVU scores.

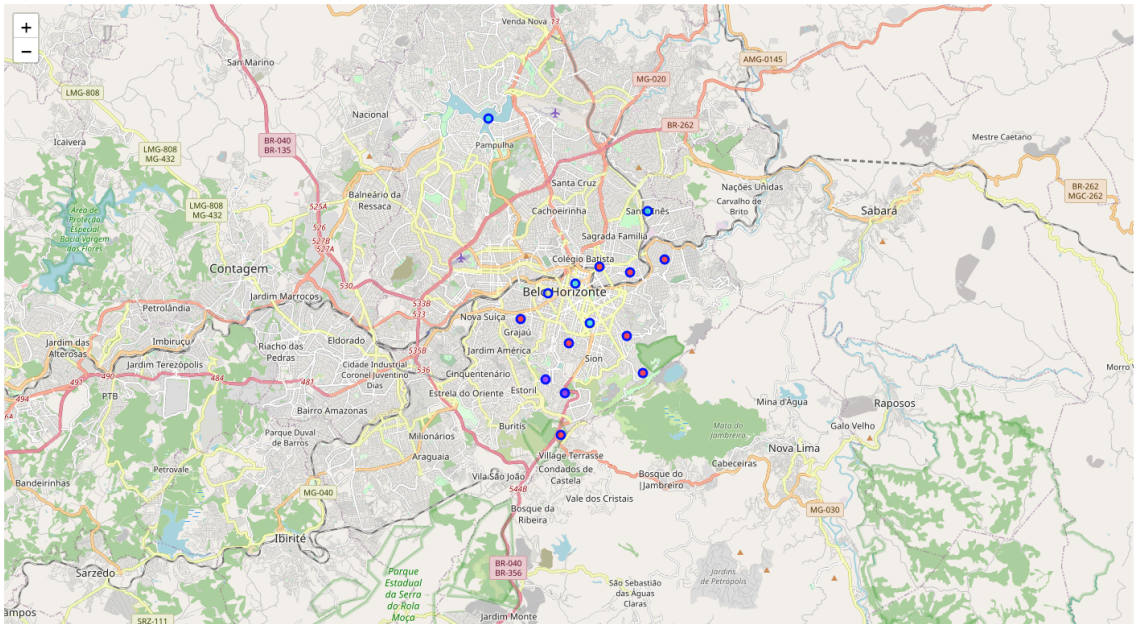
The table below shows the means of IQVU score for each cluster. Based on the values presented, it is possible to observe that the mean values for the clusters have similar distances between each other, although the distances decrease from the cluster with the highest value.

Figure 13 – IQVU means for each cluster.

Clus_km	ID	CODIGO	latitude	longitude	iqvu
0	100.0	716.375	-19.936328	-43.929212	0.777059
1	149.5	773.500	-19.960467	-43.949171	0.824468
2	253.5	1128.500	-19.899523	-43.940556	0.799002
3	12.0	617.000	-19.923382	-43.952011	0.856587

The next map shows the clusters of neighborhoods in the city space.

Figure 14 – Map clusters of neighborhoods.



Another analysis that was made on the data was the creation of a new index based on the standardized values of IQVU and number of venues in a neighborhood. This index aims to show the relation between good infrastructure and services and the availability of venues.

The index was calculated by the mean between the standardized values of IQVU and standardized number of venues for each neighborhood. The results are shown in the table below.

Figure 15 – Top index values for each neighborhood (column “mean”).

	ID	CODIGO	NOME	latitude	longitude	iqvu	Clus_km	num of places	mean
10	37	642	Centro	-19.919331	-43.940349	0.798798	2	98	2.458456
12	217	1133	Savassi	-19.935220	-43.934446	0.804587	2	84	2.002676
1	12	617	Barro Preto	-19.923382	-43.952011	0.856587	3	47	1.564052
2	68	679	Floresta	-19.912746	-43.930201	0.785570	0	65	0.823327
3	144	768	Santa Lúcia	-19.963226	-43.945101	0.824468	1	40	0.605484
9	150	774	Santa Tereza	-19.914892	-43.917046	0.785570	0	48	0.123200
4	155	779	São Bento	-19.957709	-43.953242	0.824468	1	25	-0.012275
14	143	767	Santa Inês	-19.890331	-43.909681	0.794606	2	30	-0.429537
8	11	616	Barroca	-19.933743	-43.963996	0.765196	0	41	-0.590260
11	172	797	Serra	-19.940556	-43.918423	0.775936	0	30	-0.819170
6	154	778	Santo Antônio	-19.943329	-43.943420	0.775673	0	28	-0.907012
7	13	618	Belvedere	-19.980275	-43.946570	0.769293	0	31	-0.916615
0	617	1972	Lagoa da Pampulha	-19.853209	-43.977749	0.798016	2	13	-1.058503
13	131	751	Pompéia	-19.909648	-43.902428	0.775656	0	17	-1.360403
5	101	718	Mangabeiras	-19.955434	-43.911609	0.783575	0	10	-1.483420

Discussion

Looking at the index, it is possible to see that the neighborhoods Centro e Savassi had the highest scores based on their IQVU (infrastructure and services) and the number of venues around their center. Therefore, it is concluded that they are good neighborhoods to stay in Belo Horizonte.

Of course, there are other factors to consider in making a choice, but they the information above provides a good indication of the best neighborhoods. For example, the analysis of the venues in these neighborhoods shows that Savassi has a highest number of hotels around its center, which demonstrates that this neighborhood is likely to be a touristic area. So that would be the ultimate choice for this case.

Conclusion

The data provided was able to provide meaningful information and insights about the problem in hand. The IQVU showed itself a relevant index that contributed in the understanding of different characteristics in a neighborhood that can be useful for a visitor in the city. The data originated by the Foursquare API is great, and it is possible to utilize it in a lot of different ways.

One difficulty was the limitation to make “premium” calls to the data base to access other types of data, such as venues ratings and combine everything to make different kind of analysis.

An interesting conclusion was the great value of indexes such as the IQVU. They can be calculated by different variables and measured in different ways so it is important to understand what king of information they can provide. But in general, is clearly observed their usefulness. Different kinds of industries, like tourism, can obtain helpful information to provide a better service and increase profit.

The possibilities of different types of analysis that can be made, and the numerous options of algorithms, models and visualization schemes provide a tremendous toolbox for the analyst. Longer one continues to reflect about a problem in hand, the more it opens for other paths to pursue.