Where should online store "Olist" open their first offline store?

Dataset Used

- Data is acquired from www.Kaggle.com which is scraped from www.olist.com
- Dataset has information of more than 100,000 orders from 2016 to 2018.
- Dataset has data about order status, price, payment, freight performance, customer location, product attributes, reviews, seller location and geolocation with zip code.

Things to consider and analyze

1. City that is generating revenue the most on Olist based on Consumers' address
2. In which city in Sao Paulo do customers live the most?
3. Boroughs that consist of various social venues since it means that is the area where many people goes for hangout so the store will be easily accessible and visible by customers.
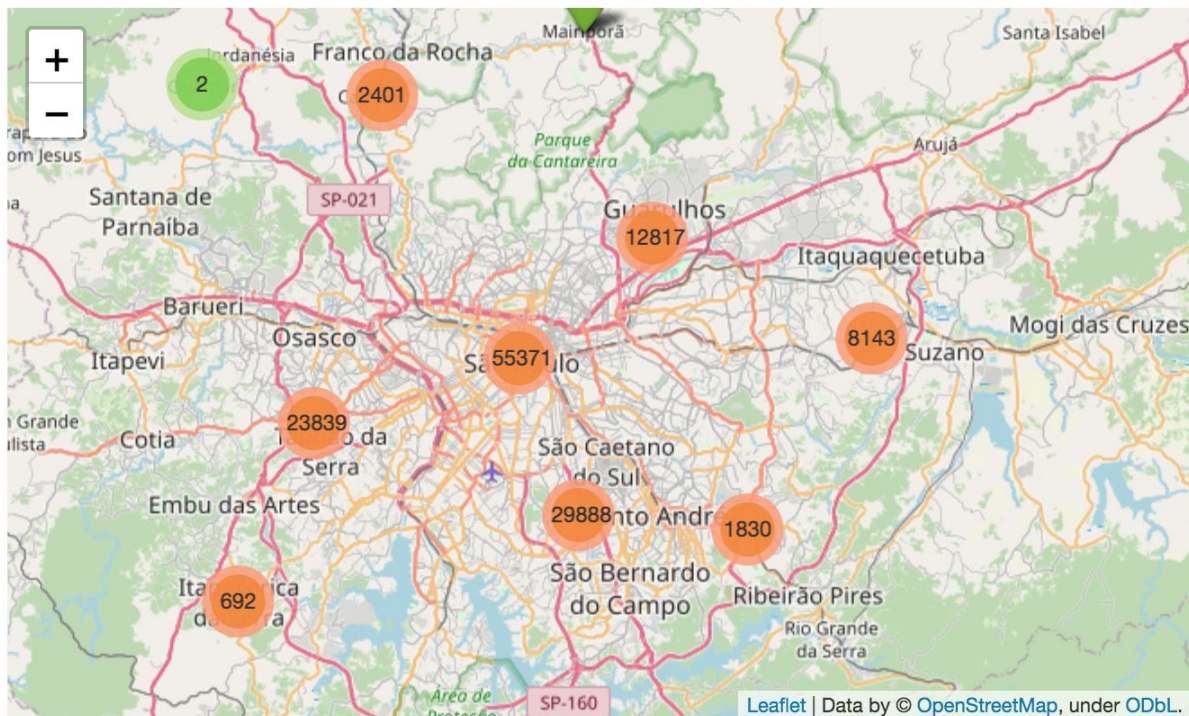
Top 5 states and cities with the revenues

	customer_state	payment_value
0	SP	5998226.96
1	RJ	2144379.69
2	MG	1872257.26
3	RS	890898.54
4	PR	811156.38

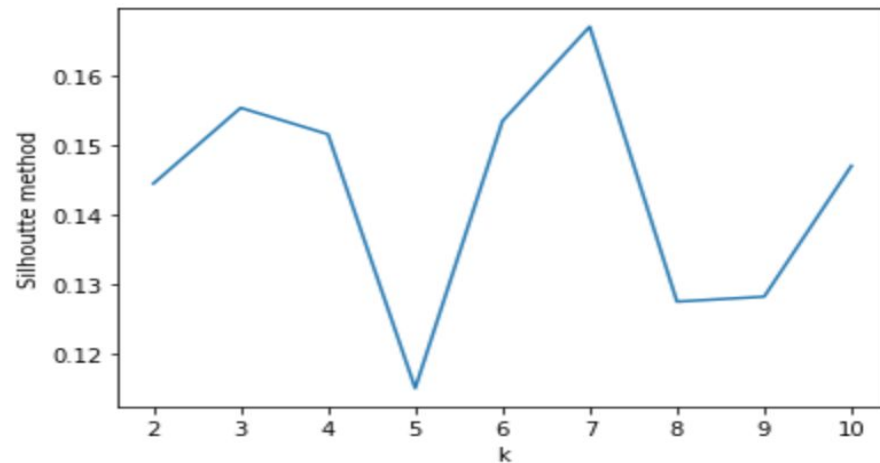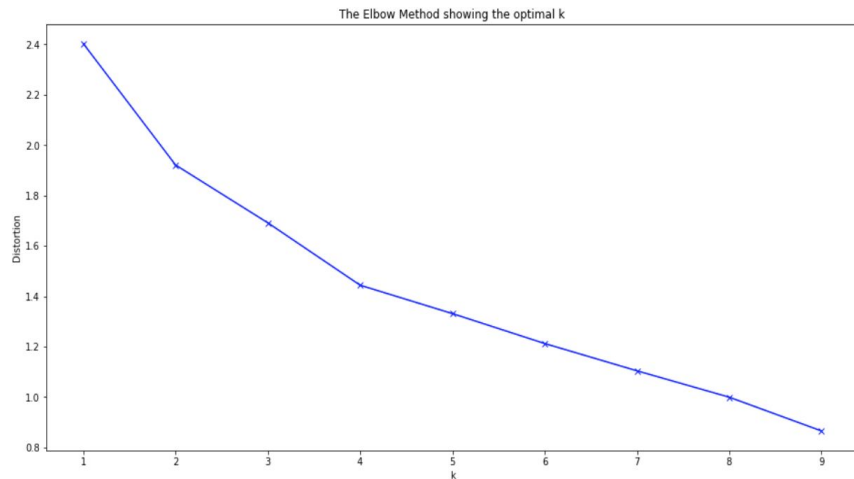|	customer_city	payment_value
0	sao paulo	2203373.09
1	rio de janeiro	1161927.36
2	belo horizonte	421765.12
3	brasilia	354216.78
4	curitiba	247392.48

City Sao Paulo in the state Sao Paulo is the city that generate revenue the most

City where the most customers of Olist live in the state Sao Paulo

City Sao Paulo has the most consumers of Olist in the state Sao Paulo, where there were 55371 customers.
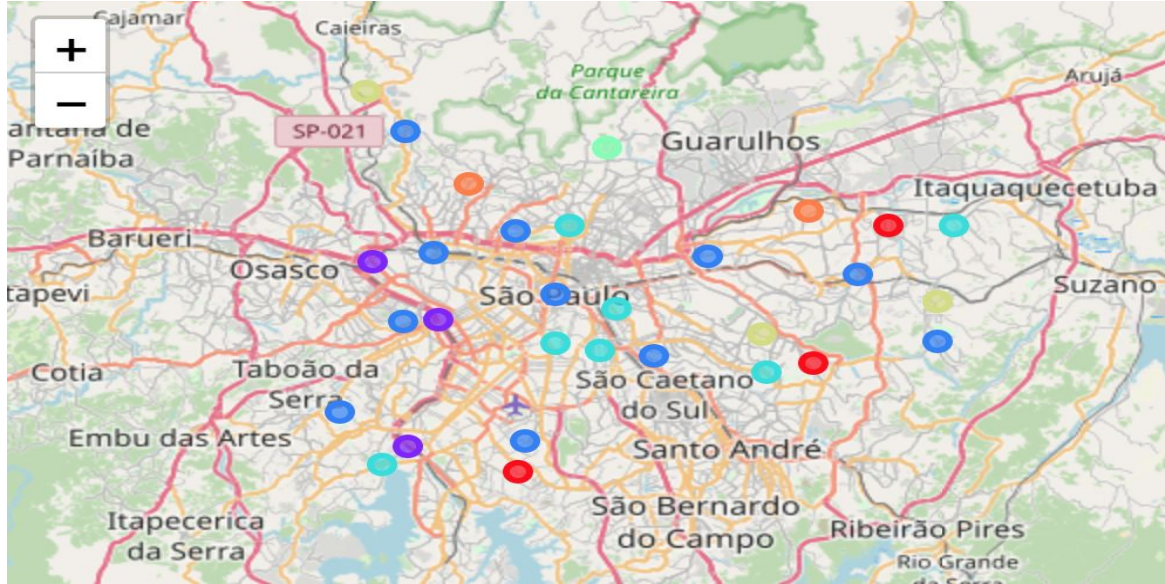
Find the optimal number of clusters "k"

The Elbow Method showing the optimal k

I used to Elbow Method to find the optimal "k" but it wasn't showing the clear "elbow". Therefore, I used the Silhouette Method.
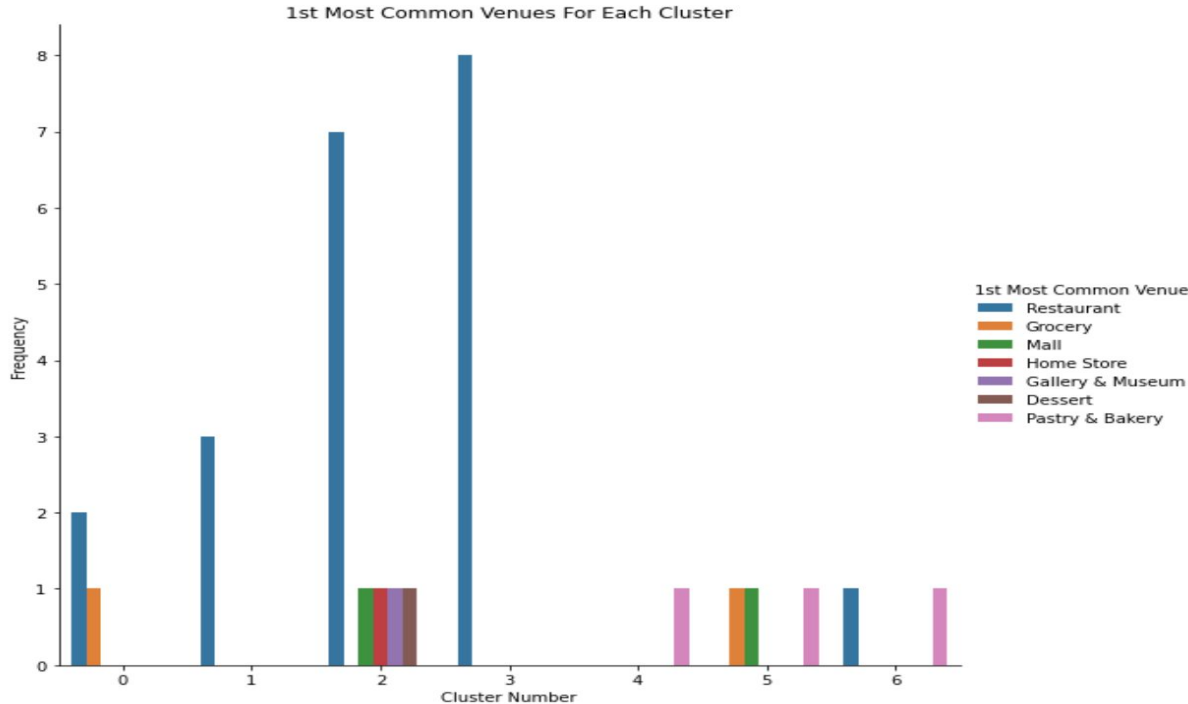
Based on the Silhouette Method, the optimal number of clusters is 7.

K-means Clustering

Using the latitude, longitude, zip code, name of the city and country with Fouresquare API, I ran K-means clutering.

Bar chart to find out what are the 1st most common venues of each borough

1st Most Common Venues For Each Cluster

Based on the bar graph, Cluster 2, where there are many restaurant and many other stores, venues at the same time is the place to consider opening the offline store.

Area where there are various types of venues is the area where many consumers visit.

Boroughs in Cluster 2 are Butanta, Campo Limpo, Casa Verde, Cidade Tiradentes, Itaquera, Jabaquara, Lapa, Penha, Pirituba-Jaragua, Se and Vila Prudente.

Conclusion

- City Sao Paulo in the state Sao Paulo in Brazil is the city where it generates the most revenue with the most customers on Olist.
- Butanta, Campo Limpo, Casa Verde, Cidade Tiradentes, Itaquera, Jabaquara, Lapa, Penha, Pirituba-Jaragua, Se and Vila Prudente are the best boroughs to open the offline store.
- There is a room for improvement for clustering method for grouping the data and performance method to find the optimal "k"
- Another possible questions to analyze to answer are:
1. What is the most popular product?
2. The month with the highest revenues (sales prediction)
3. Customer satisfaction using review data
4. Delivery performance (Consumption of time till the arrival of the parcels)
5. Who pays more for the delivery?