WHERE TO PUT A CHOCOLATE SHOP IN ARGENTINA

COURSERA IBM DATA SCIENCE PROFESIONAL CERTIFICATE FINAL CAPSTONE

INTRODUCTION TO THE PROJCET

- We want to perform an analysis to understand the distribution of chocolate shops within the Argentina territory. We want to know where are the dense areas, and where are the outliers. Thus, we will cluster the chocolate shops using DBSCAN.
- First, we will get cities from Argentina. Then we will geocode them. After this we will perform a search of all chocolate shops using Foursquare. Once we have this dataframe, we will perform the DBSCAN algorithm to cluster all these venues.
- We will use the following data: Wikipedia to get cities from Argentina, and Foursquare API to get the chocolate venues.

METHODOLOGY

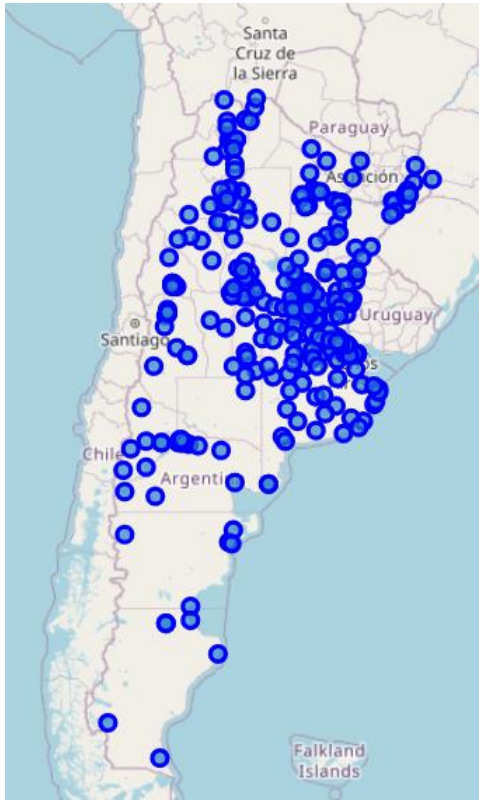
- Using wikiepdia to get the cities from Argentina

	City	Province
0	Buenos Aires	Autonomous city
1	Córdoba	Córdoba
2	Rosario	Santa Fe
3	La Plata	Buenos Aires
4	Mar del Plata	Buenos Aires

- Using arcgis to geocode each city

	City	Province	Latitude	Longitude
0	Buenos Aires	Autonomous city	-34.60849	-58.37344
1	Córdoba	Córdoba	-31.40718	-64.18571
2	Rosario	Santa Fe	-32.96780	-60.65924
3	La Plata	Buenos Aires	-34.91393	-57.94636
4	Mar del Plata	Buenos Aires	-37.99741	-57.54846

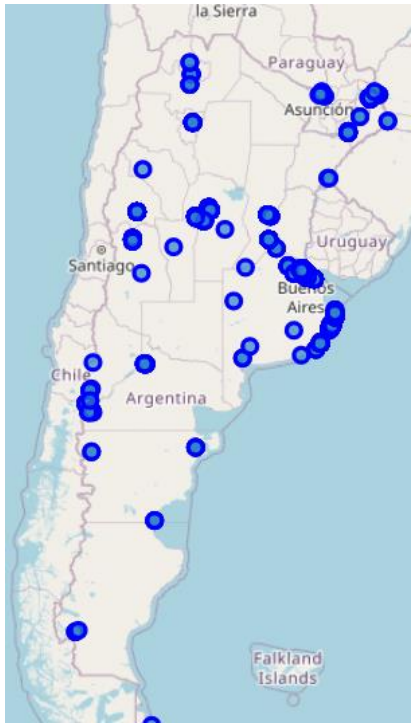
- Using Folium to create a map for the cities



METHODOLOGY

- Using Foursquare API we create a DataFrame with all the chocolate shop venues for each city

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude
0	Buenos Aires	-34.60849	-58.37344	Havanna	-34.606191	-58.375046
1	Buenos Aires	-34.60849	-58.37344	Rapa Nui	-34.599218	-58.513947
2	Buenos Aires	-34.60849	-58.37344	Havanna	-34.615760	-58.456093
3	Buenos Aires	-34.60849	-58.37344	Rapa Nui	-34.764566	-58.401615
4	Buenos Aires	-34.60849	-58.37344	Havanna	-34.611148	-58.363783



- We create a map with Folium to actually see all the chocolate shops.

METHODOLOGY

- We run DBSCAN to cluster the shops. We use a proximity parameter of 0,5, and mínimum samples of three.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Clus_Db
0	Buenos Aires	-34.60849	-58.37344	Havanna	-34.606191	-58.375046	0
1	Buenos Aires	-34.60849	-58.37344	Rapa Nui	-34.599218	-58.513947	0
2	Buenos Aires	-34.60849	-58.37344	Havanna	-34.615760	-58.456093	0
3	Buenos Aires	-34.60849	-58.37344	Rapa Nui	-34.764566	-58.401615	0
4	Buenos Aires	-34.60849	-58.37344	Havanna	-34.611148	-58.363783	0

- We get 19 clusters, and a 20th category with the ouliers.

```
clusters
```

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, -1, 12, 13, 14, 15,
       16, 17, 18], dtype=int64)
```

CONCLUSION

- Using a Folium Map show all clusters and outliers to understand the distribution of chocolate shops in Argentina based on density.