IBM Data Science specialization

Capstone Project: The Battle of the Neighbourhoods

Title: «Et voilà!» Restaurant in Buenos Aires

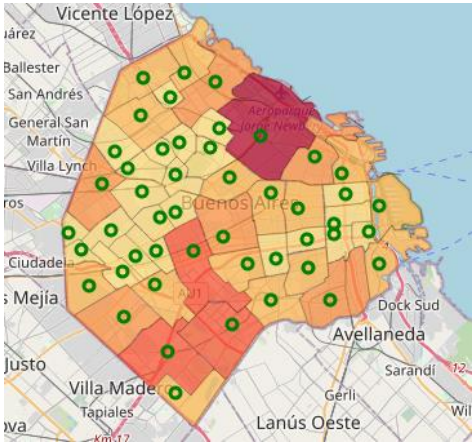
Opening a Restaurant in Buenos Aires

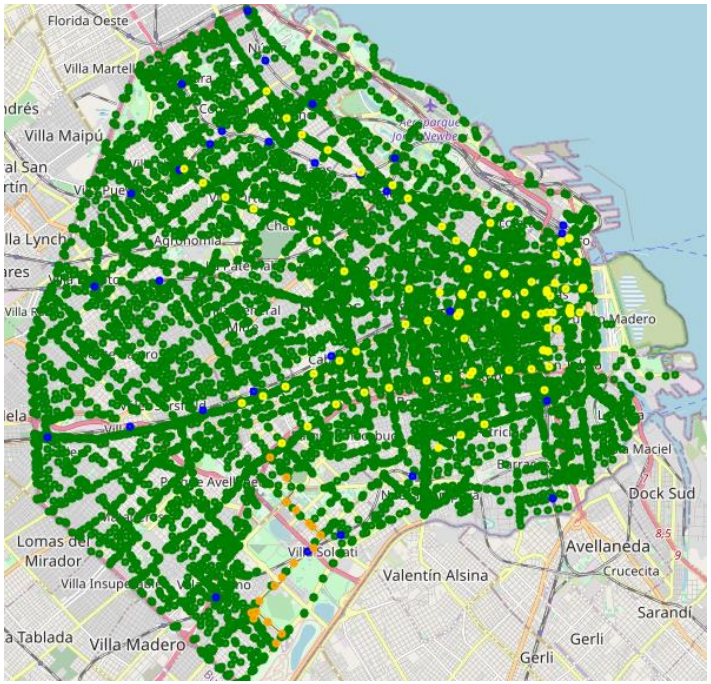
- Gastronomy is a field that could be very lucrative in a big city.
- When starting, competition and potential attendance are very influential factors to consider.
- These can be greatly determined by:
  - Number of restaurants already in the neighbourhood.
  - Population living in the neighbourhood.
  - Access to public transportation.

- **Therefore: strategically choosing where the restaurant will be is crucial for its success.**

Data used

- Foursquare API: extraction of local existing venues.
- ArcGIS API: extraction of geographical coordinates.
- Buenos Aires government datasets:
  - Neighbourhoods.
  - Population
  - Public transportation.





Methodology

- All data sources were cleaned and prepared, in order to obtain a comprehensive dataframe with information on the 48 neighbourhoods and the necessary features.
- The data was normalized to avoid predominance of a feature over the others.
- k-Means clustering was applied to find similar neighbourhoods.

Cluster Labels	Neighbourhood	Rest Density	Pop Density	Transp Density
0	Chacarita	0.184568	0.242984	0.486219
5	Paternal	0.018421	0.240815	0.199249
1	Villa Crespo	0.261268	0.684186	0.474720
7	Villa Del Parque	0.036248	0.478685	0.247408
4	Almagro	0.162244	1.000000	0.548360
1	Caballito	0.035973	0.781526	0.441336
7	Villa Santa Rita	0.000000	0.453499	0.240487
5	Monte Castro	0.031266	0.367644	0.213713
0	Villa Real	0.030788	0.279808	0.511280
7	Flores	0.023907	0.570661	0.322543

Results

- Labeled clusters were evaluated, preferring low competition (red) and, secondly, high population density (blue).
- Transport accessibility complemented in third place.

Results

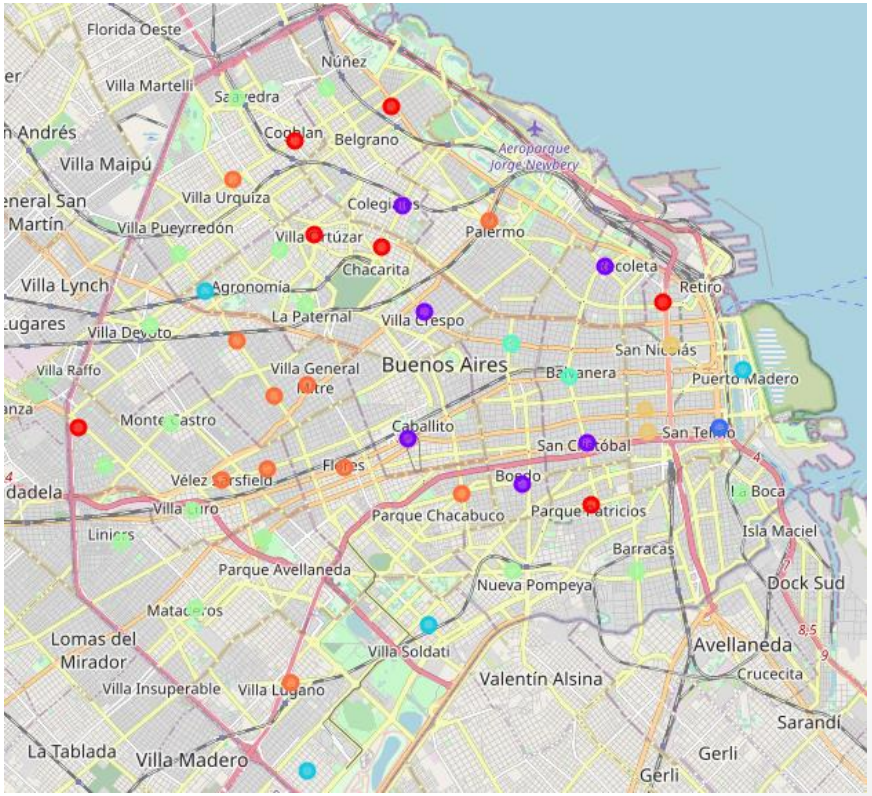
- Many of the clusters to the left side of the grouped bar chart present low competition.
- Only cluster 4 combines this with high population density.
- Inside this cluster, clearly Balvanera has the advantage, with the very low competition and very high population and transport densities.

Cluster Labels	Neighbourhood	Rest Density	Pop Density	Transp Density
4	Almagro	0.162244	1.000000	0.548360
4	Balvanera	0.047297	0.983371	0.787807

Discussion

- k-Means served as a filter, finding neighbourhoods with similar properties.
- After this filter, it is possible to easily notice how the different densities are distributed.

Geographically, neighbourhood clustering looks like this:



Conclusion

- The answer to the problem question is that Balvanera is the most apt choice of neighbourhood in which to open a restaurant.