Predicting the best place for opening a restaurant

By Samuel Sadek

Value in predicting best place for a new restaurant for investors

Thriving city and hub for the southern continent

Strong tourism, links and a stronghold settlement for international businesses

Food forms the foundational fabric of Spanish society and culture for good food and cuisine

Lots of shops, restaurants and other F&B establishment

Value in guiding investors where to place their investment and in which area

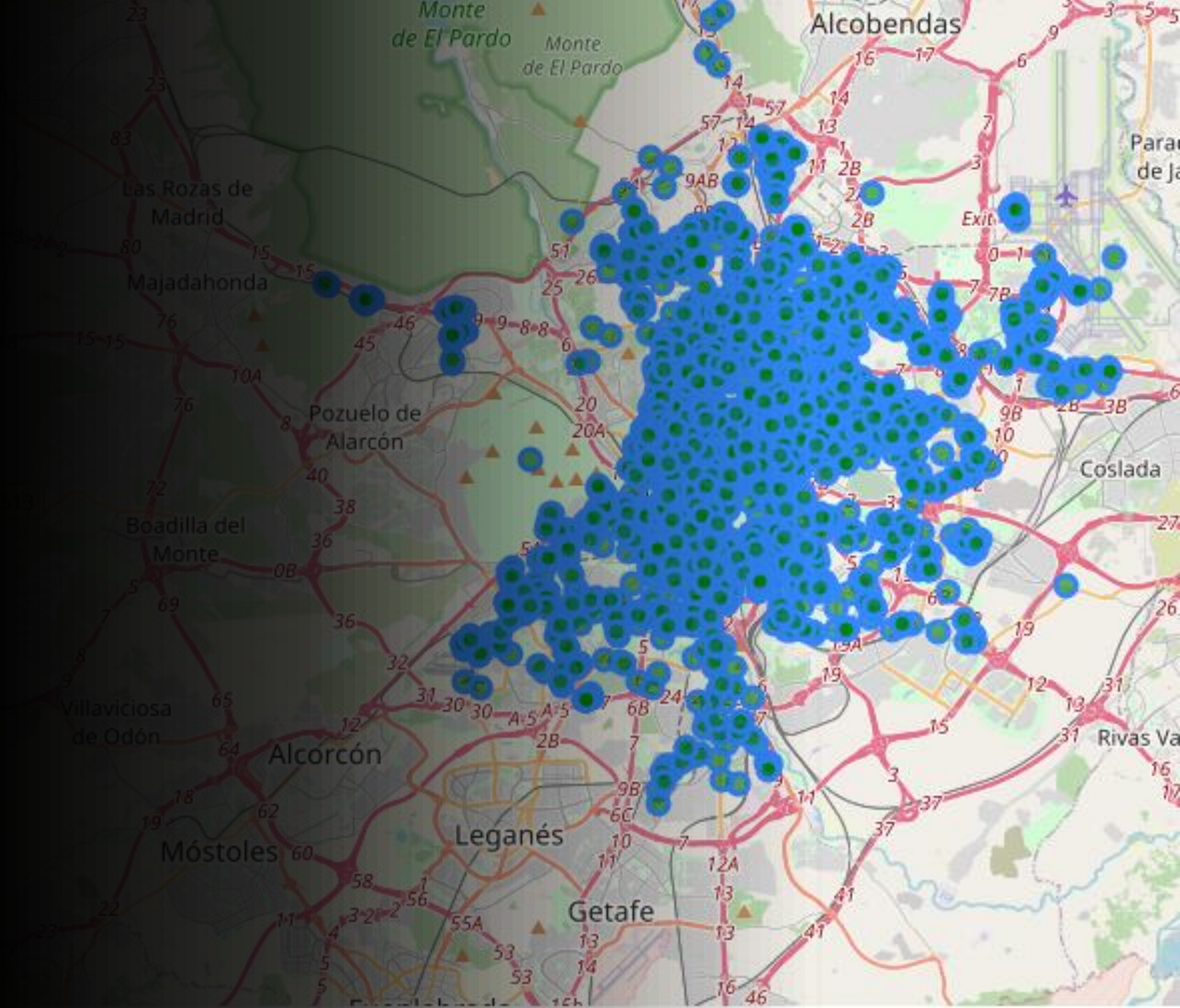
So, where would you recommend opening a new restaurant?

Data acquisition and cleaning

- ~ 6,300 food-related establishments found in Madrid with 1km distance (Foursquare API)
- ~ 130 neighbourhoods in Madrid via web scraping from Wikipedia
- In total, 6,280 rows and 7 features in the raw dataset
- Several duplicates, close proximity to each other and not pertaining to neighbourhoods they belonged to, were dropped
- Cleaned data contains 7 features.

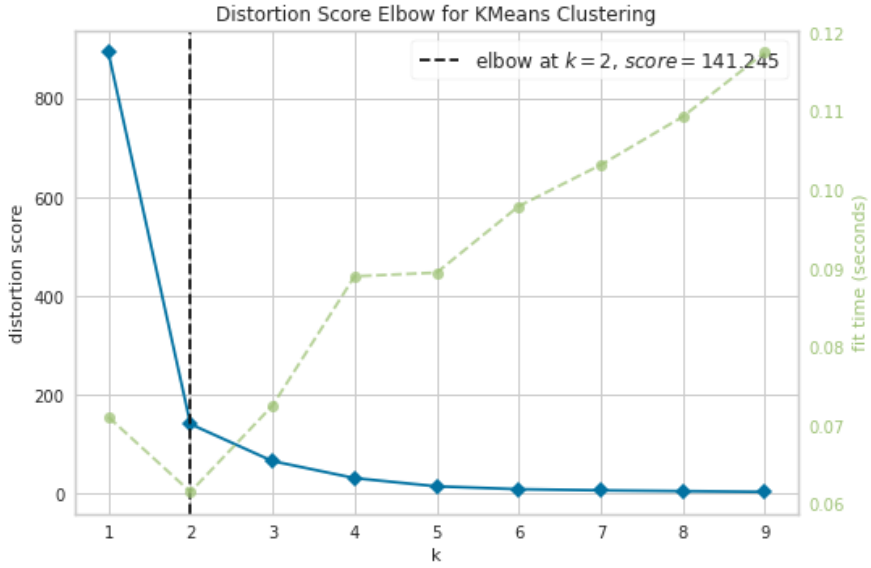
Data Pre-processing

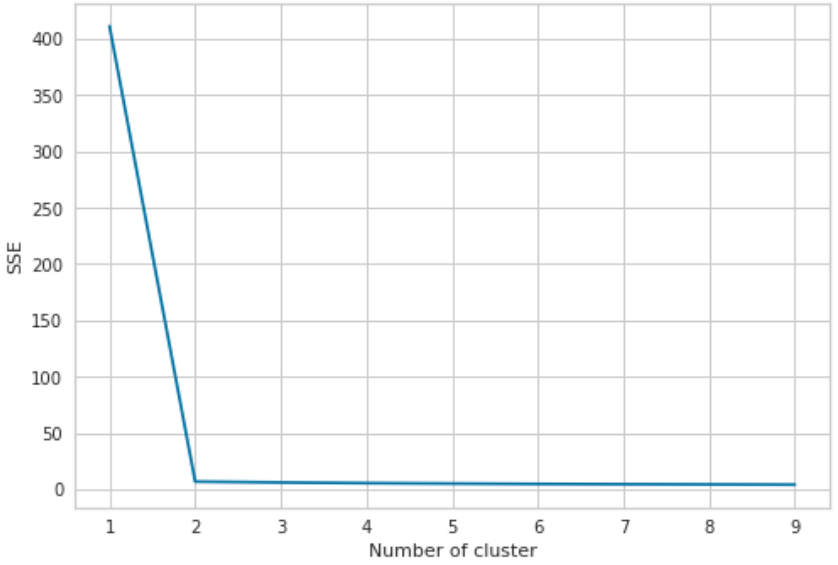
- Circa 3,200 venue data points plotted onto a Folium map
- Used one-hot encoding, a pre-processing feature, to convert a categorical predictor into numerical value (Venue Category)
- Calculated the mean of the frequency of occurrence of each category for each neighbourhood
- Calculated the top-10 most visited types of establishments for each neighbourhood



Clustering

- Adopted K-means clustering algorithm useful in unsupervised learning and working with unlabelled datasets
- Algorithm goal: to **find groups** in the data, with the number of groups or clusters represented by the **variable K**
- First step, to identify the best optimal value K through using a famous analytical approach: the **elbow method**
- From both the elbow and inertia test methods, we concluded the optimal value of clusters or K is **2**

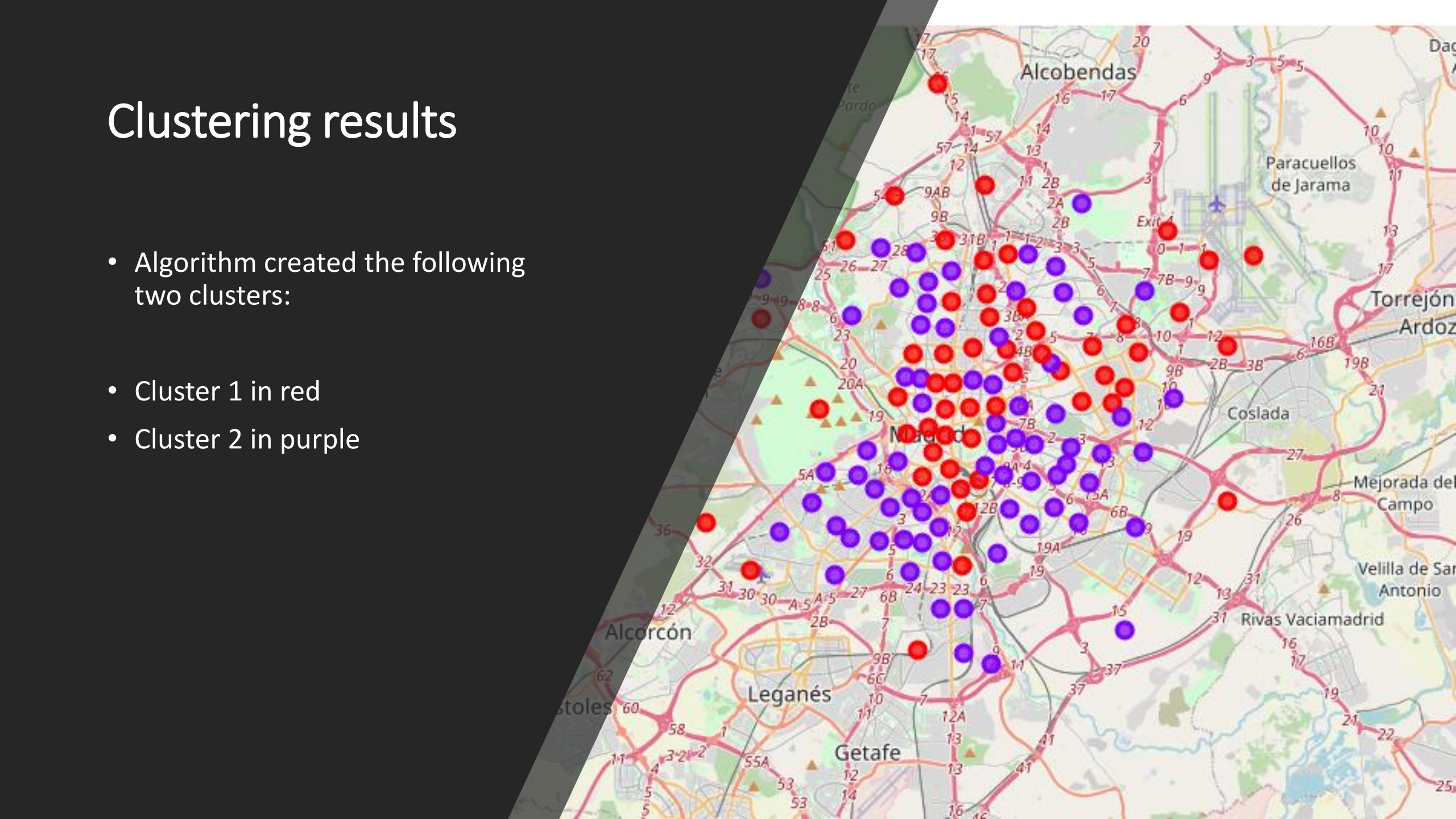


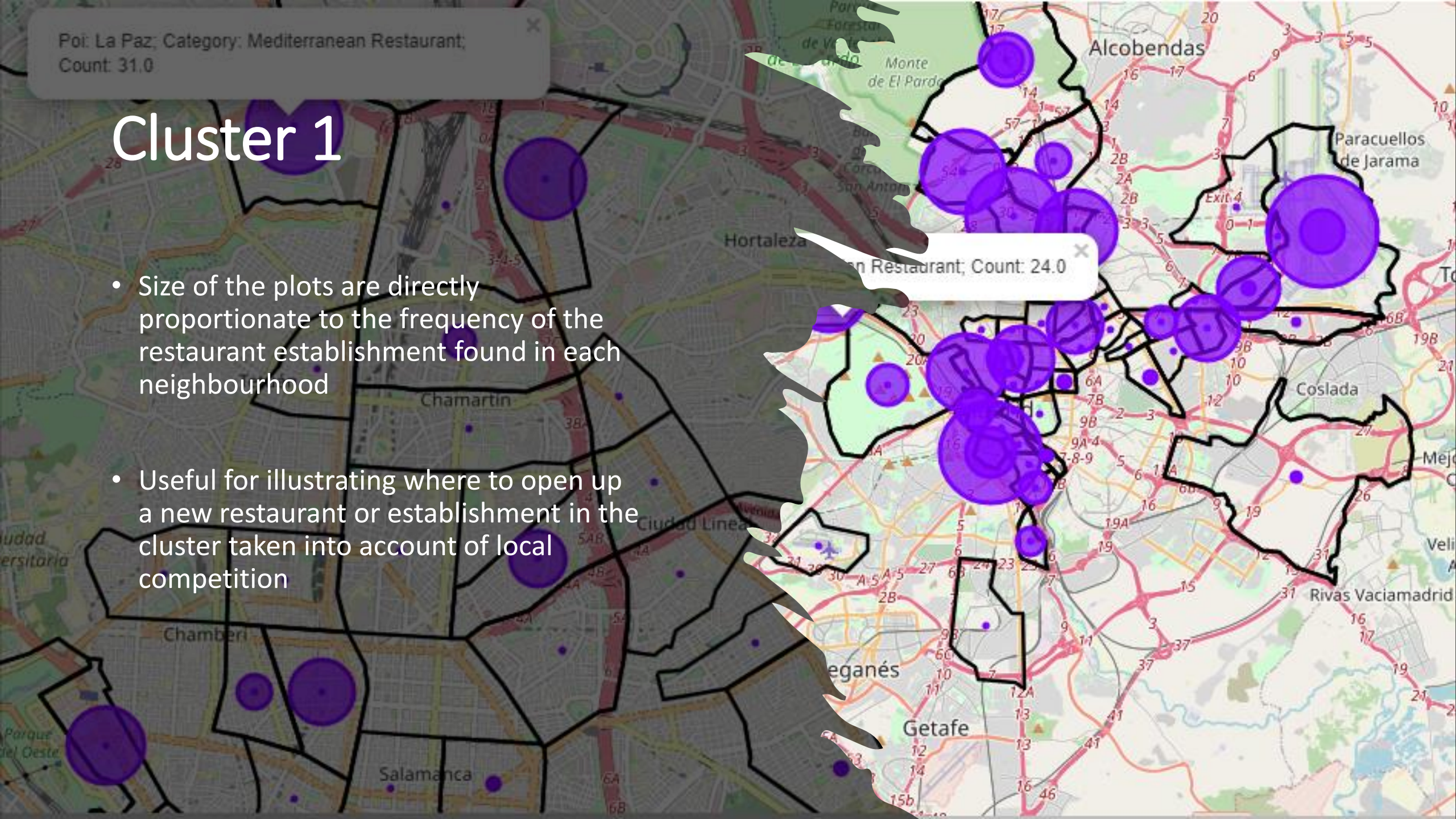


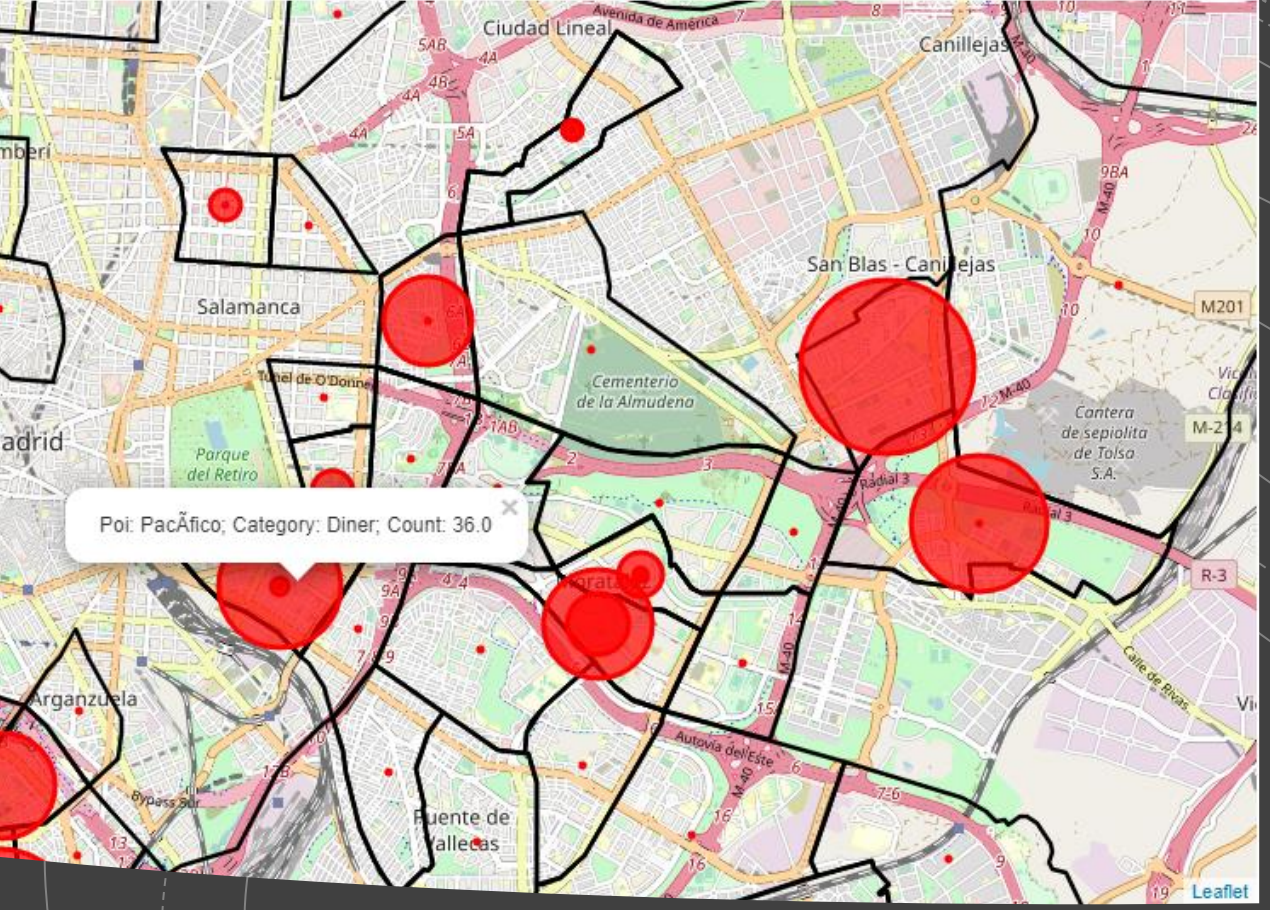
Clustering results

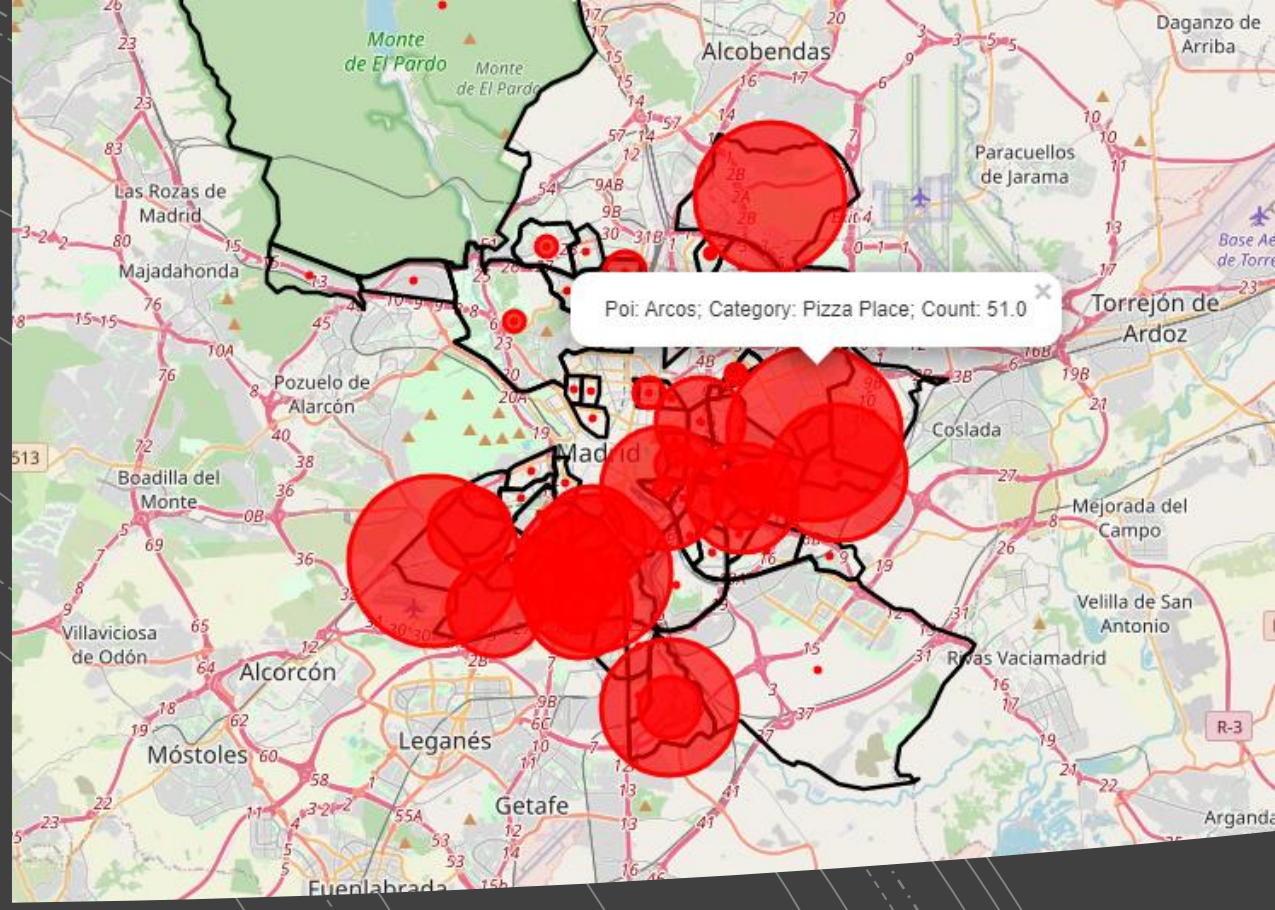
- Algorithm created the following two clusters:

- Cluster 1 in red
- Cluster 2 in purple




Poi: La Paz; Category: Mediterranean Restaurant; Count: 31.0
Cluster 1
Size of the plots are directly proportionate to the frequency of the restaurant establishment found in each neighbourhood
Useful for illustrating where to open up a new restaurant or establishment in the cluster taken into account of local competition
Restaurant; Count: 24.0
Alcobendas
Paracuellos de Jarama
Hortaleza
Chamartin
Coslada
Rivas Vaciamadrid
Getafe
Salamanca
Chamberi





Cluster 2

- In addition to the size and frequency, the plots have also been dissected into the relevant neighbourhood boundaries
- Useful for giving other insights about the locality of existing venues to aid in their analysis and final decision making for opening up a new establishment

Conclusion

- Built a useful clustering model to predict the optimal location for a new establishment
- Accuracy of the model has room for improvement
- Suggest using other clustering techniques such as DBSCAN to help discriminate between highly and densely populated areas
- In addition, incorporate some other useful information into the model providing invaluable insights to stakeholders:
  - For example, F&B revenues at city and neighbourhood-level
  - Knowledge about local city council plans for types of investments into the area, etc.