



Exploring Caracas Metropolitan District venues information available in FourSquare

Author: Rafael Blanco



Exploring Caracas Metropolitan District venues information available in FourSquare

- The Caracas Metropolitan District (CMD) is composed of five (5) municipalities and thirty two (32) parishes.
- There are popular web applications that require databases with information about popular places in the community, with their coordinates (geolocated data) and descriptions about the activities performed there.
- The aim of this study is to collect data available on FourSquare portal about venues in the CMD to identify the regions to implement a prototype of a new delivery food service application.



Data acquisition and cleaning (2)

- We collected and parsed information about venues around our points of interest to build a dataset, with special interest in the number of venues inside each searching radius and the categories these venues belong to. We obtained as result a dataset with the 5 most common category venues per parish.
- Parishes with less than 10 venues were excluded from this experiment.

Methodology

Unsupervised learning algorithm k-means used to make clusters of parishes based on the most popular categories of venue in that region.

Number of parishes	Number of venues	Color	Categories with greater presence
1	15	Purple	Spanish restaurant
8	368	Blue	Bakery / Shopping Mall / Restaurants
3	87	Orange	Pizza place
1	18	Yellow	Restaurants
1	12	Green	Pharmacy, Gym

Table 1: Clusters description

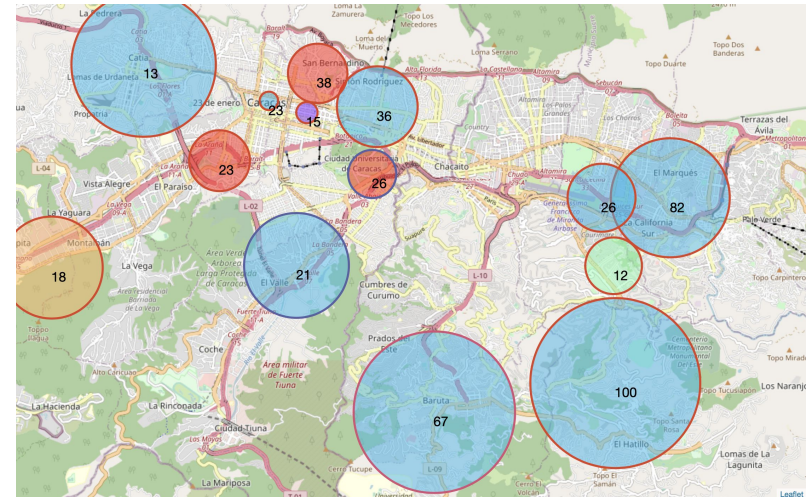


Figure 2: Clusters produced by k-means algorithm based on venues categories (k = 5)



Results

- After excluding the parishes with less than 10 venues registered, we kept 15 out of 32 parishes, representing 47 % of the total parishes in CMD. It means that in more than half of the parishes in the CMD the information about venues is null.
- There is a region on the south-west region of the map with the highest venues presence in the FourSquare database.
- Most of popular categories of venues in this region are bakeries, shopping malls and some restaurants.



Discussion and Conclusions

- We collected and explored geolocated data about venues in the CMD.
- The data recollected is not enough to provide conclusions and convince stakeholders to invest their money.
- This is an initial step of the process that should iterate starting from the data recollection phase, looking for more available sources.
- If we had to make a decision at this point, there is evidence that our prototype should be implemented on the south-west region of the CDM.
- Possible future work: analyze data from south-west parishes, trying to get insights about why there is more information available in this region and precise the interests of people who live there.