IBM Applied Data Science Capstone

COURSERA CAPSTONE

Business problem

- This project will analyze the Barcelona data. First, we will find the most visited hotels according to the number of check-ins, then we will try to find the districts that are lacking the selected type of shop which could be potential business opportunity. The target audience of this report is any one that is interested in opening a hotel but have no idea what kind of and in which district.

Data - Methodology

- This Wikipedia page
  (*https://es.wikipedia.org/wiki/Categor%C3%ADa:Barrios_de_Barcelona*) contains a list of districts in Barcelona. I will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then i will get the geographical coordinates of the districts using Python Geocoder package which will give me the latitude and longitude coordinates of the districts . After that, I will use Foursquare API to get the venue data for those districts . I am particularly interested in the hotel category in order to help us to solve the business problem

Results

- Categorized the districts into 3 clusters:

  - Cluster 0: Districts with low number of hotels: *Barrio de la Sagrada Familia, Baró de Viver, Font d'en Fargas, Horta-Guinardó, Nou Barris and Trinitat Vella*.

  - Cluster 1: Districts with high number of hotels: *Dreta de l'Eixample, Fort Pienc and La Vila Olimpica del Poblenou*.

  - Cluster 2: Districts with moderate number of hotels: *Barrio de San Antonio, Distrito de Ciutat Vella, Les Corts, Sants-Montjuic, Sarriá and Sarriá-San Gervasio.*

Discussion

- The majority of hotels are concentrated in downtown and near to the port / beach.

- A good place to open a new hotel will be in the suburb area like *Horta-Guinardó* or *Barrio de la Sagrada Familia* due to well connected with city center.

- *Cluster 0 has very low number of hotels, so could be interesting to attrack tourists.*