**Data**

**The necessary data for this project, based on the above stated requirements, are:**

**The metro stations in the karol bagh greater metropolitan area**

**Number of existing hotels near each station**

**In addition, the distance to the nearest hotels for every metro station will be used**

**In order to obtain the data, a combination of the geopy Python library and the Foursquare API will be used:**

**1. ‘Karol Bagh Market’ will be considered as the center of Karol Bagh,Delhi. It is indeed one of the most central location in the city. I will obtain its geospatial coordinates using the geopy library**

**2. Having the coordinates of the ‘center’ of Karol Bagh, the Foursquare API will be used to retrieve data for all the metro stations in Athens greater area in a radius of 15 km**

**3. To find the existing hotels near the metro stations, the Foursquare API will again be used for every station. I will obtain data for all the hotels located in a radius of 1000 meters of every metro station**

**Using the collected data, I will calculate the number of existing hotels near each station. I will also be able to determine the minimum distance to a hotel for every metro station from the 3rd step of the above process. This minimum distance to every metro station from a hotel, along with the number of already existing hotels near the station will be used as input to K-Means clustering algorithm to obtain the clusters of areas (metro stations).**