2. Description of the data:

Data sources and manipulation:

- Geographical coordinates of the negihborhoods in Buenos Aires: the data source of this coordinates is “Comunas.csv” available in https://data.buenosaires.gob.ar/dataset/comunas/archivo/Juqdkmgo-612221-resource.

  It consists on 7 columns:
  WKT: shows a list of coordinates that are the boundaries of the “comuna”.
  BARRIOS: a list of neighborhoods in the “comuna”
  PERIMETRO: perimeter lenght of the “comuna”
  AREA: área of the “comuna”
  COMUNAS: number of “Comuna”
  ID: id of the comuna
  OBJETO: useless straing saying “LIMIT OF THE COMUNA”

  In the analysis, the columns taken into account were WKT, BARRIOS and COMUNA. The column WKT was studied to get the coordinates of each of them instead of the coordinates of the boundaries. Another thing to say is that each comuna was divided in two parts to get more precisión. To do this, for each group of coordinates, the kmeans function with 2 cluster was applied to get the two centers. One of the centers was named with “A” an the other with “B”.

- Venues location and category from Foursquare.
  With the coordinates of the centers I sent a request to Foursquare and get a Json file with the venues and transformed i tinto a DataFrame with the data I wanted. After that i will filter the “Venues Categories” into the ones related to sports and sports shops.
  Then I clustered the neighborhoods taking into accounto only this two categories and decide which cluster is the most suitable for the store. To do this, the data was encoded so it could be analized with the KMeans algorithm.