

2. Data

2.1 Data Description

The primary data used for this capstone project is the geodata “Infrastructure/Collection of Innovation Hubs and Technology Centers in Kenya” from Kenya Open Data (referred to as “Innovation Data” in this report), including the types, location, target market information of the existing recorded innovation hubs and technology centers in Kenya.

Secondly, Foursquare API is used here to explore geographical information of the surrounding venues of the innovation hubs and technology centers as mentioned above.

The data origins are listed in the table below.

2.2 Data Source

Table 1 Data source

| Data Name | Form | Source |
|--|-----------|---|
| Infrastructure/Collection of Innovation Hubs and Technology Centers in Kenya (“Innovation Data”) | GeoJson | https://www.opendata.go.ke/datasets/infrastructure-collection-of-innovation-hubs-and-technology-centers-in-kenya |
| Foursquare | API Query | https://foursquare.com/ |

2.3 Data Usage

For the “Innovation Data”, firstly, it will be downloaded from the website and delivered to the panda dataframe which contains several categories of hubs’ detailed information:

- 1) ‘Type’: technology hubs and programs, venue funds or web portal.
- 2) ‘Name’: the names of the hubs such as “iHub”, “Naliab” etc.
- 3) ‘County_Location’: this category indicates the administrative location of the innovation hubs and technology centers in Kenya (i.e. county name).
- 4) ‘Target Market’: the targeted market sectors of the hubs.
- 5) ‘Latitude and Longitude’: It presents the geographical coordinates of a certain hub (web portal hubs have no location information).

The created dataset will be grouped with different categories as described above for initial analysis while the details will be discussed in the exploratory data analysis part. NAN values will be dropped for the convenience of further analysis.

The coordinates information of “innovation data” will be utilized jointly with the Foursquare API Query. Requests would be sent to return venue information in need. In this case, it refers to the venue information of tech-hub neighborhoods, which will be illustrated in the neighborhoods exploration part.