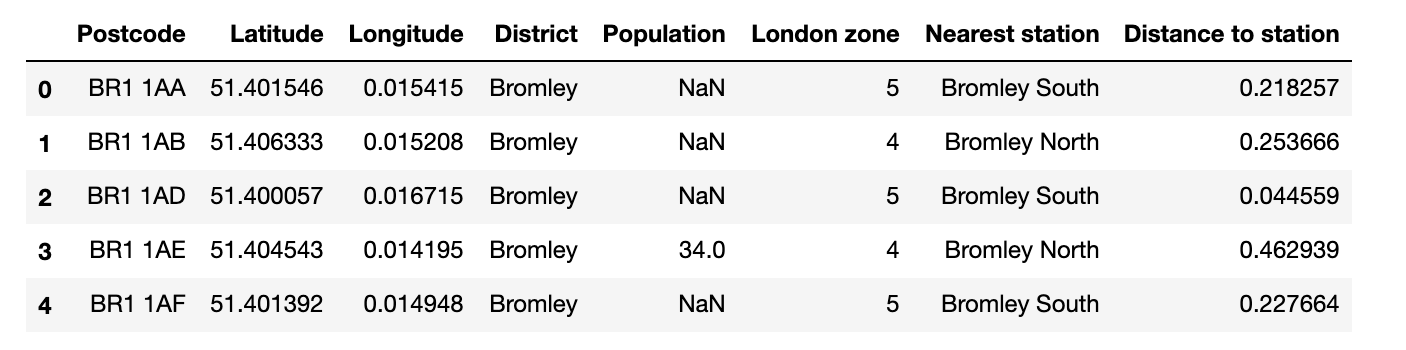
# Data section

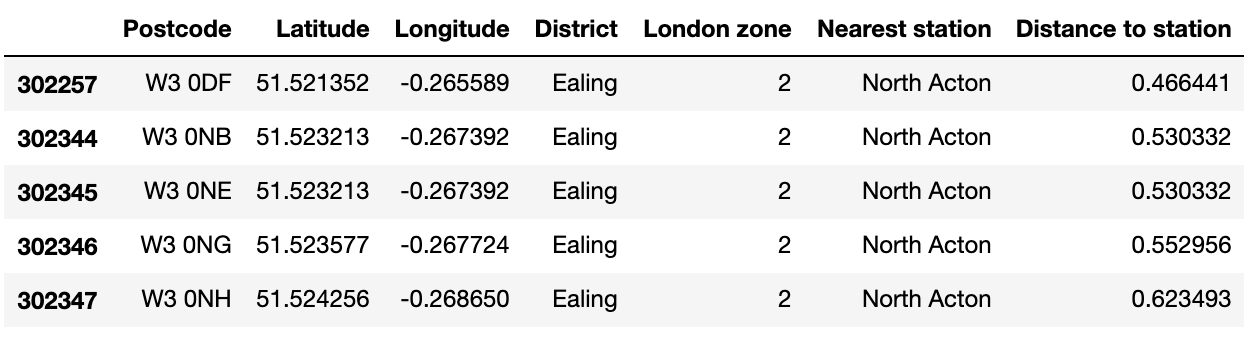
## Data obtain and cleaning

First, the coordinates and the postcode at Ealing are required. The data is obtained from: <https://www.doogal.co.uk/PostcodeDownloads.php>

This csv files contains all information needed including the postcode, coordinates, zone, station, and distance to the station.



The data can be processed to obtain the critical information for the analysis to answer the questions proposed in the introduction section, e.g. zone area for commuters, travelling distance from the station to the office. The data is cleaned to keep only the information needed: postcode, coordinates, distance to the station, station, and zone. The postcode is limited to W3 and the zone is limited to zone 2. The current location (W12 7SB) is also included in this dataframe for the clustering stage.

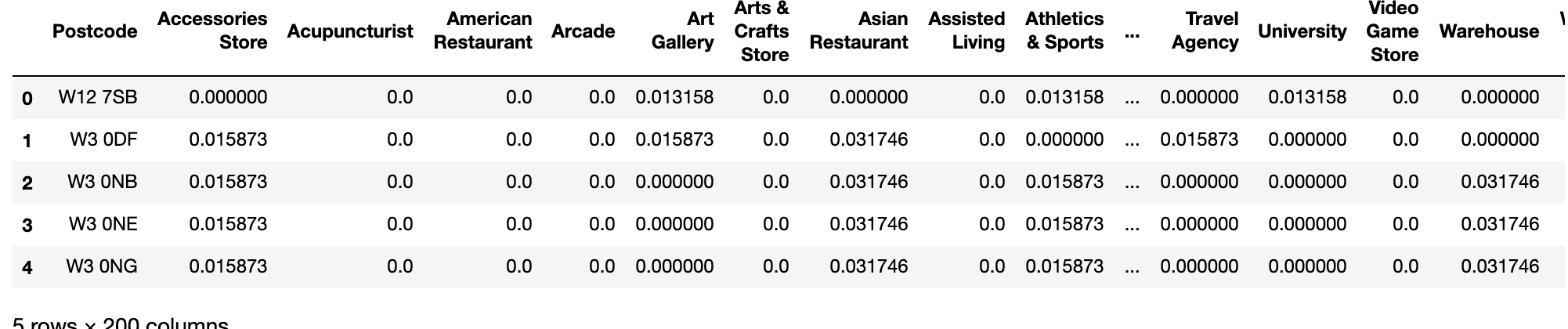


Second the coordinates from the cleaned data is used to loop through foursquare for the venues information at the area. The output venues are filtered to exclude the transportation, hotel, prison etc and only keep the meaningful information for this case study.

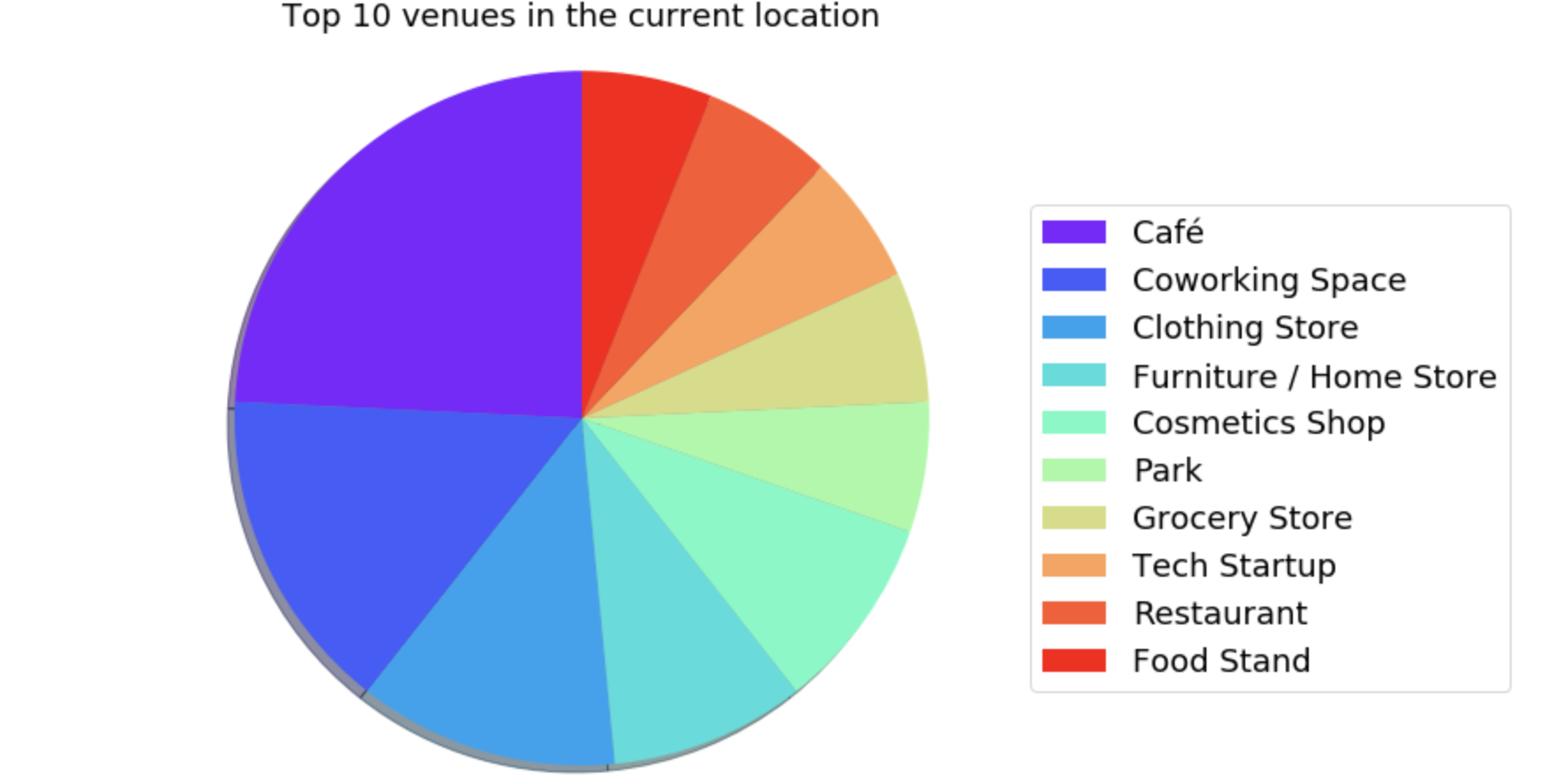


## Data processing and visualisation

The venue dataframe is transformed to dummies to find out the frequency for each venue appears in each postcode and obtain top 10 most common venues.



For comparison reason, the top-10 most common venues are put into a pie chart to visualise the venue composition around.



Folium is used to visualise the clustering and the potential new locations on the map.

