

Data Description:

The data used to solve this problem is geolocation data collected from FourSquare. Wikipedia is used to extract the postal codes for the Dublin postal districts. Using the names of the Dublin postal districts, latitude and longitude are generated from Python Geopy (Nominatim). These coordinates are then used to extract nearby venues using FourSquare API. Top 10 common venue categories are then identified for each Postal District and then grouped using unsupervised learning Kmeans clustering. The clustering will provide information to tourists as per their taste to choose best place for short stay. The clusters based on categories will provide business opportunities for opening new restaurants or cafes.