

Data

The Data we will use is the Foursquare location data. The Foursquare API allows us to explore areas by categories. In this case, we will use two categories that are frequently visited by tourists: Museums and Landmarks. First, we will explore from the center of London. The latitude and longitude of London will be obtained by a simple google search. After getting the locations of the museums and landmarks, we will cluster them using DBMeans. This method does not require us to state the initial number of clusters. Instead, it will do that for us. It will also find outliers that don't belong to a cluster.

After clustering, we will cleanup our dataframe by removing outlier rows. If there is any cluster that looks too large, we will cluster them further into two separate clusters using KMeans, since this time we know the number of clusters we need.

The resulting clusters will be mapped into a folium map of London where the travel company can use to determine the location of their information offices.