Introduction

There are lot of things which are already built across a place. Sometimes it is difficult to understand what kind of a thing can be built at a place. To understand a place and it surrounding we need to create clusters of that particular area and check which cluster contains minimum number places and what is the feasible thing that can be build around in every neighborhood in every cluster.

Business problem

A business can grow and develop when there are not much of competitions and when the product produced is of great quality. In any particular area there can be a lot of people with same business and products. Suppose in Brooklyn, if there are 5 gyms in the neighborhood there is no point in building one more gym. Rather a shop for gym equipments such as - weights, repairing of machines, protein products, gym wear can be opened which can help people as well as gym owners to buy products from the same. I am going to cluster the places around Booklyn and identify what business can be opened in every cluster.

Data

The data which will be used is newyork\_data.json which contains data about the Newyork city. From this json file latitude and longitude will be extracted and will be mapped on the map for visualization. Along with the json file I will use Foursquare api to collect data about the neighborhood within newyork regarding various places. After collecting data, it will be cleaned with only required fields for understanding the neighborhood. Then I will focus on Brooklyn and understand it surrounding. Various apis' and packages will be required for extraction, cleaning and working on data.

Methodology

K – means clustering is used to identify the clusters. First from newyork data, the data required for Brooklyn is identified. Then in Brooklyn all the neighborhood is identified and visualized. Later Brooklyn data is given to K mean for clustering with a cluster of 5. It creates 5 clusters and identifies the cluster with maximum number of businesses. We can visualize and see what can be opened in any cluster.

Result

As we can see in the notebook, cluster 2 and 3 have a highest number of businesses and cluster 1,4 and 5 do not have as much as businesses as of cluster 2 and 3, which identifies our location problem. Hence we can open a new business in cluster 1, 4 and 5.

Discussion

What all businesses can be opened: As we can see cluster 1 has a farm and farm market which gives an opportunity to build a Bakery or a Dairy. For crowded clusters where businesses are already in plenty, a shopping mall can be opened combining some business to open inside the mall.

Conclusion

Cluster 1,4 and 5 are ideal to open any new well established businesses such as Dairy, Bakery, Clothing store etc, whereas any new idea or innovation can be tested in cluster 2 and 3 as the amount of well established businesses are already present and the success rate of new business is high (if any new idea) as the population is high in that region.