Deciding the best place for restaurant

Introduction/Business Problem:

In a city of New York, Mr. John is looking to open a restaurant, we need to find the correct location for establishing it. So that he may get great amounts of profit and his business will also relish.

The target audience include the people of localities where they go to offices colleges, and may not have time to cook mostly the people of Indian domicile and hence they'll depend on the restaurant.we need to find the data of the localities where such people stay and report that data.

The target audience that we mentioned above will care because there is not possibility for them but the restaurant should meet the standards and should surpass the previously existing ones by the quality, service, and other aspects.

If we consider from the studies then we got to know that Indians living in other countries will prefer restaurants which offer Indian food so the target customers here in our case are Indian students, working professionals etc.,

Data section:

The data that will be used to solve is the location data of different colleges, hostels, and offices where the residence areas of different working peoples will be present.

For example the New York city has coordinates as 40.7128 ^{0} N and 74.0060^{0}W.If we take the geographical neighborhoods of that city by folium and explore them based on their similarity/nature. So it will be very easy to cluster them using KMean clustering technique in SkLearn. So our main concentrated data is on that cluster which has majority of schools , offices, business places like cinema halls, community centers, sports halls,etc.,

So to summarize we require the foursquare data of the New York city to explore its neighbor hoods and their coordinates .we also require the https://geo.nyu.edu/catalog/nyu_2451_34572 dataset inorder to access the neighborhoods of the city.

By following the above mentioned process we get the cluster with desired data, and choosing the location from that cluster would more easier and practical.

Methodology:

Here we perform analysis on the data that we obtained from the above link and thus first we need to find where the Indian people located and we need to cluster them. For that I use kmeans technique to create clusters of the data and classify them according to their living areas in New York.

Now we have the clusters of Indian people and then apply data frame operations on pandas and obtain of those who are of desired background like student, working professionals etc.,

By calculating the MSE from kmeans we will know how far from the households do the tentative restaurant is situated, and it will be the main factor in deciding the outcome.

Classification by logistic regression gives the probability of customer visiting the restaurants if it is situated there and averaging them from people present in that cluster gives the whole cluster probability .This will be the second most important factor.

We will find the Density of cluster by applying statistical operations on the kmeans model and that density gives the total number of people present in that are which will be the another deciding factor.

Results:

The results here obtained are the cluster with average of 3621.45 individuals out of which 62.59% are of desired background (students, working professionals, etc.,).By classification through logistic regression it was obtained that the probability is 0.63.

Discussion:

From the quoted above results It is clearly evident that this cluster is the most reasonable and the profitable cluster to establish a restaurant and it is recommended that the cluster of this area is recommended to start a restaurant which being Manhattan of 40.7831^0 N , 73.9712^0 W of radius 276.54m will be the best place and the recommended one.

Conclusion:

Thus the above mentioned place will be the more profitable and more suitable one for trying your business skills.