**Introduction**

**Background**

Gurgaon, also know as Gurugram, is located north of the Indian state of Haryana[1]. It is near the Delhi border. Gurgaon is now a leading financial and industrial hub in India. Gurgaon also has local offices for more than 250 Fortune 500 companies. It also has a population of 876,000. As such, Gurgaon is one of the fastest growing cities in the world. As such, there is a demand for more food cuisines. This has caused a surge in the types of cuisines available. It would be difficult for investors to choose the right location to open a new restaurant.

**Problem**

In this report, we would like to choose the best place to open a Mexican restaurant in Gurgaon, India. However, we do not intend to spend resources to scout for new locations where there are no restaurants open at the moment. We would like to

* Open a Mexican restaurant **among other existing cuisines** to target their market dominance
* Open a Mexican restaurant that is **not near any existing Mexican restaurants**

Considering these problems, we will map out areas where investors should avoid setting up shop and suggesting places they should target.

**Data**

**Dataset**

We will be using the foursquare API along with a Kaggle dataset[2] to assist us in mapping out existing restaurants to visualize potential areas.

**Data Cleaning**

A screenshot of a cell phone

Description generated with very high confidence

Table 1

Table 1 shows the raw data from our dataset. We will only be using City, Longitude, Latitude and Cuisines. As such, we will be dropping all our other tables.

A screenshot of a cell phone

Description generated with very high confidence

Table 2

Table 2 shows the cleaned data. Using this data, we would have to remove Mexican Cuisines as we are targeting other cuisines. We also remove categories such as bakery as we are targeting lunch and dinner restaurants.

After I have cleaned the data, I ended up with 106 points of interests. However, due to foursquare’s limitation, I had to reduce the points of interests to 50.

A screenshot of a cell phone

Description generated with very high confidence

Table 3

I also require another set of data that is specifically for Mexican cuisines. Table 3 is taken from the same data set but is filtered using City = Gurgaon and Cuisines = Mexican.

**Methodology**

Firstly, I would clean the data to make it easier to process. I will do this by dropping columns and filtering rows to get the most concentrated results.

Next, I would like to visualize both sets of data to get a feel of what is available and to determine if there are any existing segmentation that we can derive from there. I do this using Folium and pandas to visualize the points on a map.

With that, I will use the foursquare API to get existing venues near the locations to be able to cluster them together based on that. From the data obtained, I will use onehot encoding to assist in clustering. With that, I will proceed to cluster the restaurants based on the nearby venues. Using that, I will then plot them onto a map. Finally, I will superimpose existing Mexican restaurants onto the map to identify points of interest.

**Results**

**A picture containing text, map

Description generated with very high confidence**

Figure 1

Figure 1 shows the visualization of all non-Mexican restaurants in the area. From this plot, we can see that the density of restaurants near sector 26 and 27 is high. This is because that is the city center of Gurgaon, India.

**A close up of a map

Description generated with high confidence**

Figure 2

Figure 2 shows the visualization of all non-Mexican restaurants clustered. Using this, we are able to have a rough idea of where we would like to open our restaurant. From the figure, we can see that sector 27 may seem like a potential place to open a Mexican restaurant. Remember, my aim is to not open a restaurant in a totally new are but instead to be competition to existing cuisines. However, we now need to compare figure 2 with existing Mexican restaurants as we no not intend to compete with existing Mexican cuisines. To do this, we will superimpose existing Mexican restaurants physical location to get a better idea of which cluster would be the best.

A picture containing text, map

Description generated with very high confidence

Figure 3

Figure 3 shows all clusters in red and Mexican restaurants in blue. From this visualization, we will be able to decide where to open our Mexican restaurant.

**Discussion**

From the visualization, we can determine 2 main areas of interest, sector 48 and sector 27. Sector 48 is a really good location as there is no existing Mexican restaurants nearby. Having a high cluster density also means that there is a demand for restaurants in that area which supports our second main point. The only downside to this location is that it is further away from the city center where business may be higher.

The cluster density at sector 27 is high as well Although there is some Mexican restaurants near to sector 27, this will not heavily dampen our business if we opened a Mexican restaurant in sector 27. This is because the restaurant would be closer to the city center.

**Conclusion**

We can conclude that Sector 27 and sector 48 would be very good options to open a Mexican restaurant to achieve both our main points stated in the introduction. In this project we have using foursquare API and machine learning techniques such as k-means clustering to determine the best place(s) to open a Mexican restaurant.

**References**

1. https://en.wikipedia.org/wiki/Gurgaon
2. https://www.kaggle.com/shrutimehta/zomato-restaurants-data