Predicting suitable locations to open new cuisine restaurants.

Akshit Singhal

July 12, 2019

1. Introduction

1.1 Problem

In order to open a new restaurant, the owner need to assess the localities in a city to get insights about the different kind of possibilities, like a suitable location along with the type of cuisine that will be served in the restaurant.

So the problem is to find out a suitable locality to start new restaurants alongwith the proper choice of cuisine to be served at each restaurants.

1.2 Points to consider

Generally, restaurants are opened with keeping the locality in mind.

Restaurant should not be opened in an area where too many restaurants are already present. What kind of cuisine will be served also need to be picked wisely. Predicting suitable location helps the owner to target location to acquire. What restaurants in your area are successful and which are failures.

1.3 Interest

Obviously Foodies have interest as well. Nobody can say 'No' to delicious cuisine at new restaurants.

2. Data Acquisition

The dataset `zomato.csv` is downloaded from Kaggle dataset. Restaurant, City, Locality, Type of Cuisine served alongwith Price Range and Rating are the major columns for consideration. Latitude & Longitude for each restaurant are also available in the dataset.
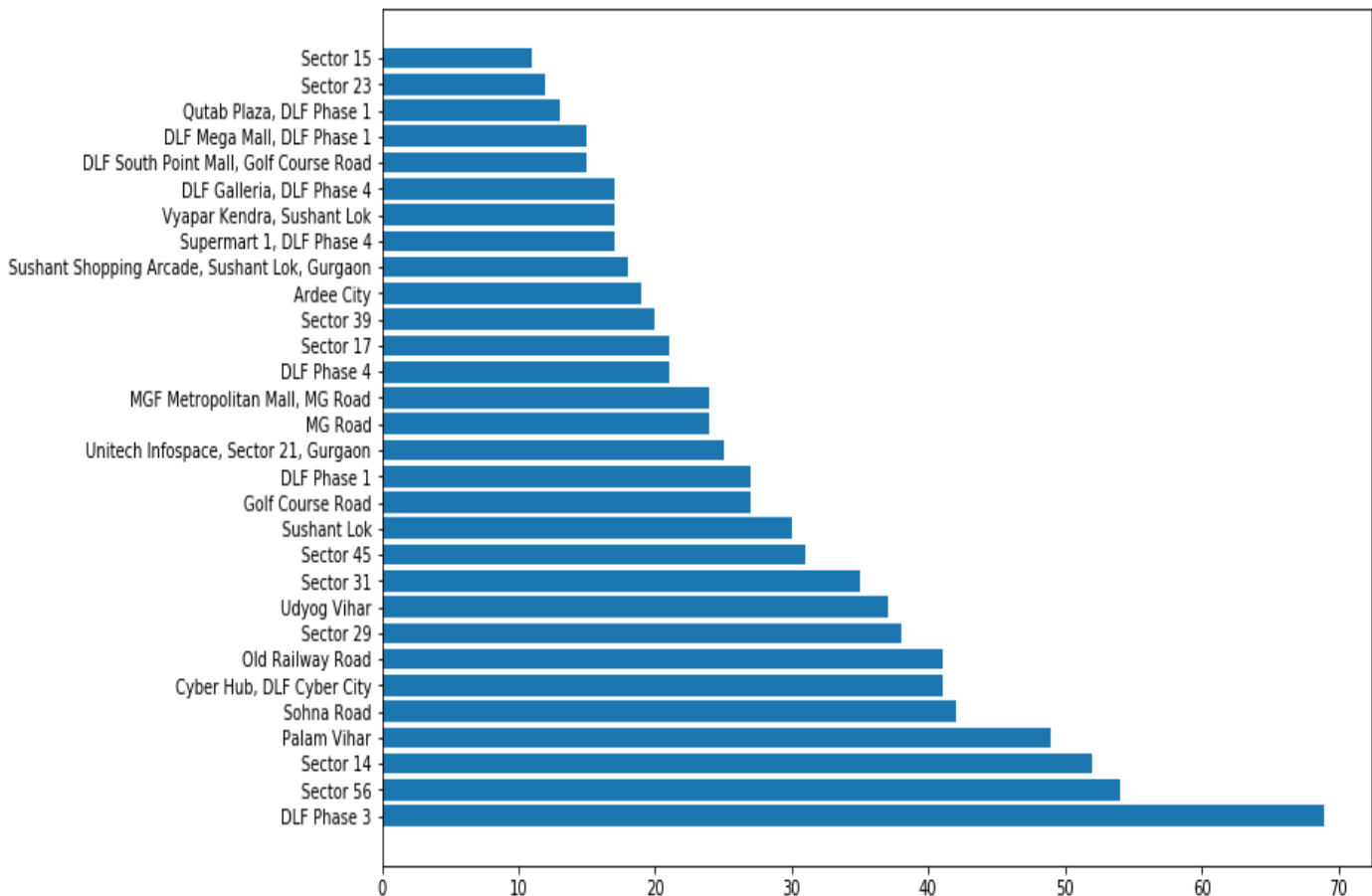
In total, 9551 rows and 21 columns.

3. Data Cleaning

Dataset consists of several countries data, but for sake of simplicity only data for India is considered and rest of the data is dropped. Now in India only data is divided according to cities.
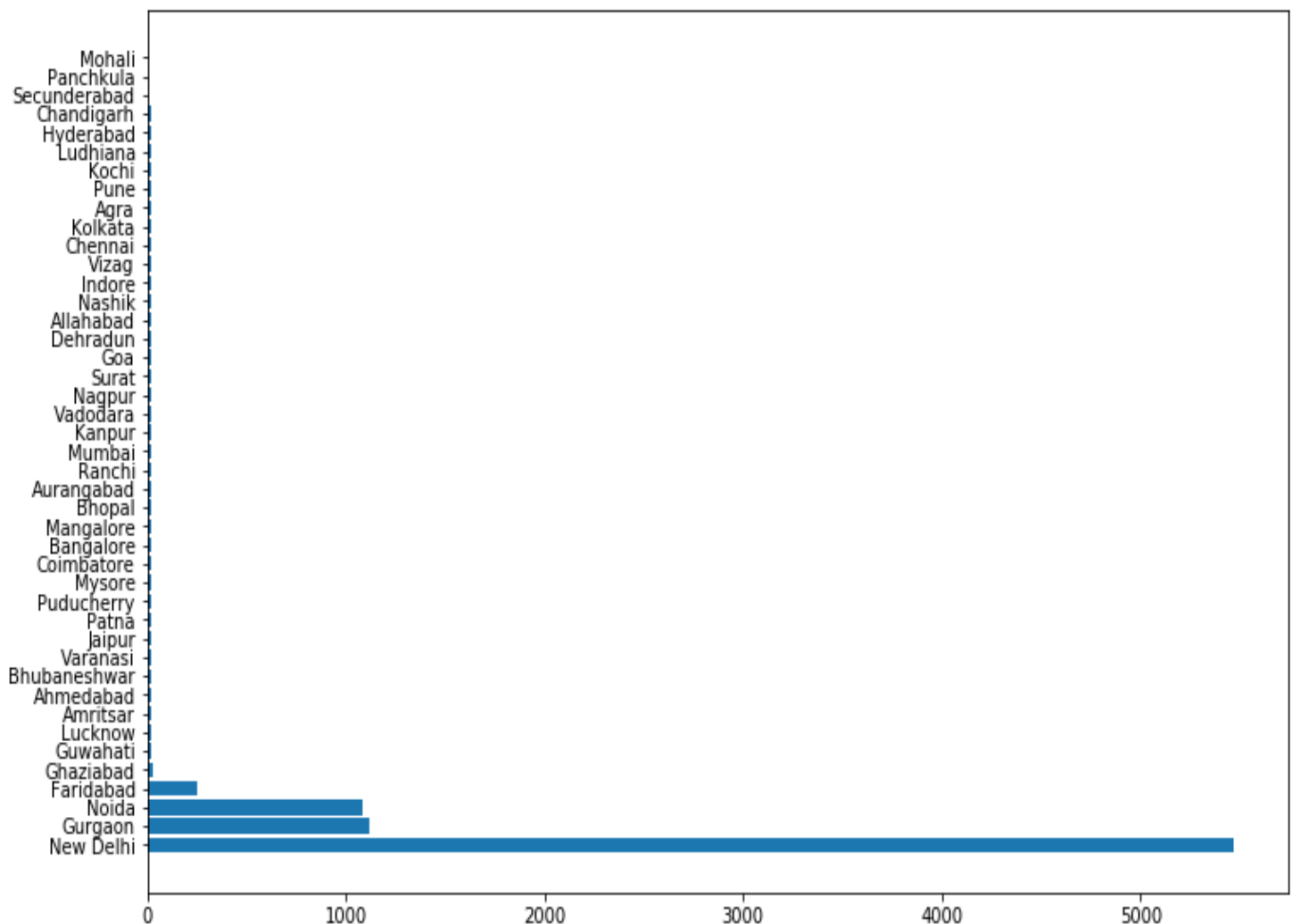
Since New Delhi already has more than 5500 restaurants, let's go for Gurgaon. Let see data for each locality in Gurgaon.

Clearly, DLF Phase 3 is already too much crowded with almost 70 restaurants. It is advisable not to open a restaurant in this area alongwith Sector 56 and 14 also.

Duplicate, highly similar or highly correlated features were dropped.

For example- Rating, Rating Color and Rating text all conveys the same information therefore any two of these columns can be dropped without any loss of information.

Analyzing columns `Aggregate rating` , `Rating color` and `Rating text` .

```
In [9]: rating={'Not rated':[],'Poor':[],'Average':[],'Good':[],'Very Good':[],'Excellent':[]}
        for c,t in zip(ggn['Aggregate rating'],ggn['Rating text']):
            rating[t].append(c)

        rating_c={'Not rated':[],'Poor':[],'Average':[],'Good':[],'Very Good':[],'Excellent':[]}
        for c,t in zip(ggn['Rating color'],ggn['Rating text']):
            rating_c[t].append(c)
        for t in rating:
            print(t,':',min(rating[t]),'-',max(rating[t]),':',*set(rating_c[t]))
            print('='*30)
```

```
Not rated : 0.0 - 0.0 : White
==============================
Poor : 2.0 - 2.4 : Red
==============================
Average : 2.5 - 3.4 : Orange
==============================
Good : 3.5 - 3.9 : Yellow
==============================
Very Good : 4.0 - 4.4 : Green
==============================
Excellent : 4.5 - 4.9 : Dark Green
==============================
```

It is evident that 'Not Rated' is associated with 'White' color, 'Poor' is associated with 'Red' color and so on. Therefore we can drop any of these columns without any loss of information.

4. Exploratory Data Analysis

After data cleaning, now time to choose famous cuisines in these areas.
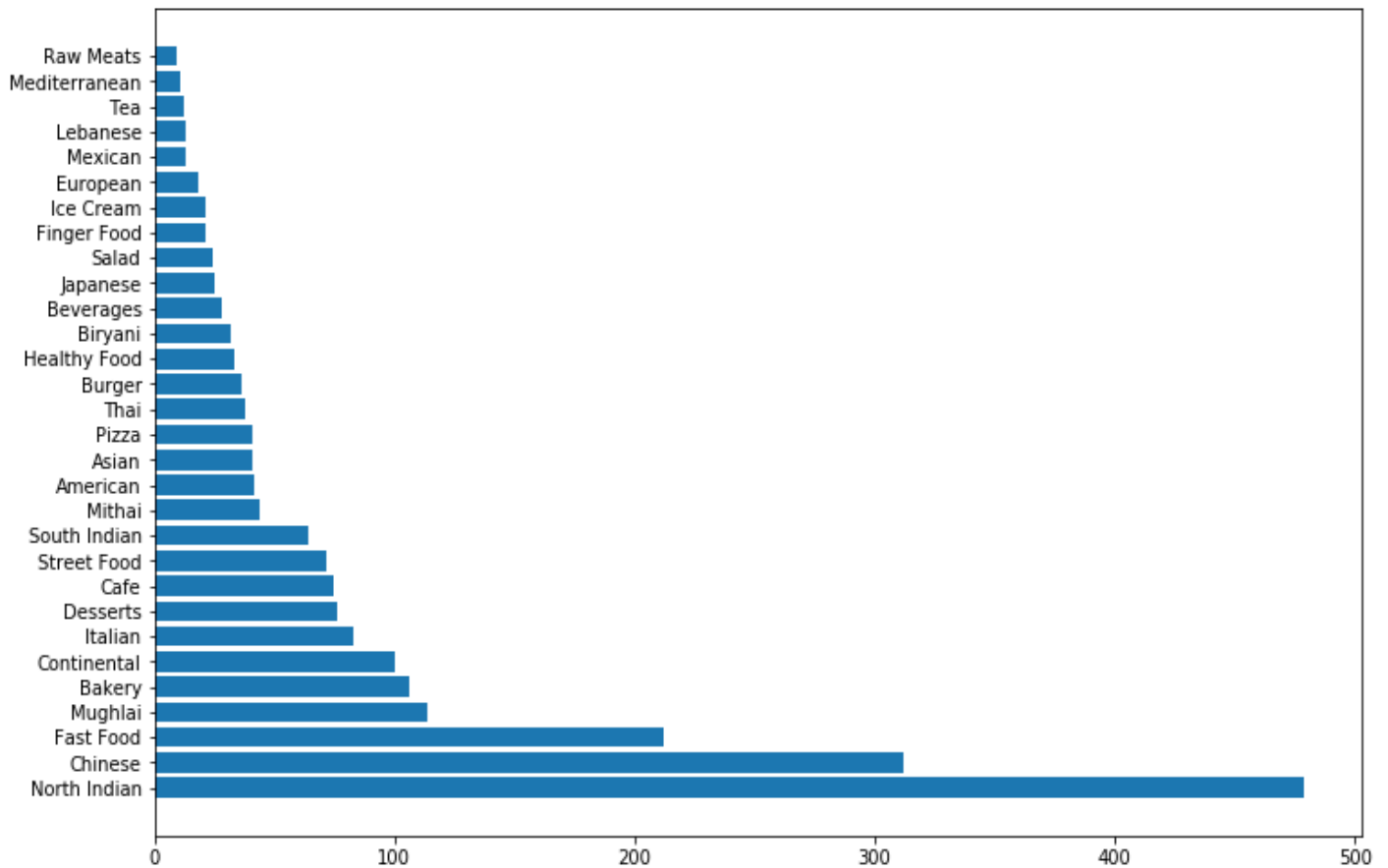
Around 470 restaurants already serves 'North Indian' food followed by Chinese and Fast Food. So it is not advisable to open a 'North Indian' restaurant, because mostly restaurant already serves this cuisine.

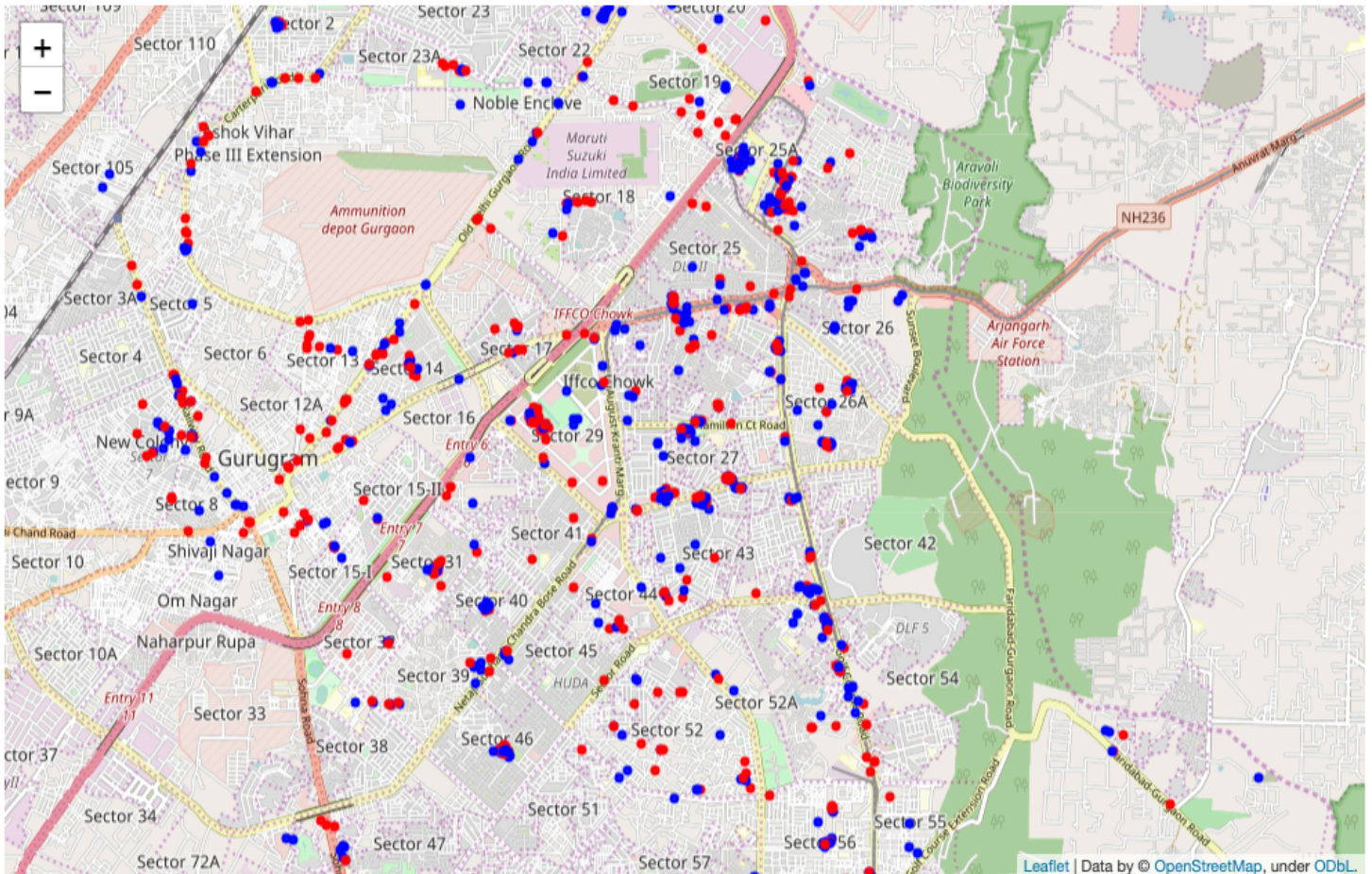Represent this cuisine data on map:

Figure – Map of 'North Indian' restaurants in Gurgaon.

Here Red represents 'North Indian' while Blue represents any other type of cuisine. So it is clear that there are already enough no. of North Indian restaurants all over the city.

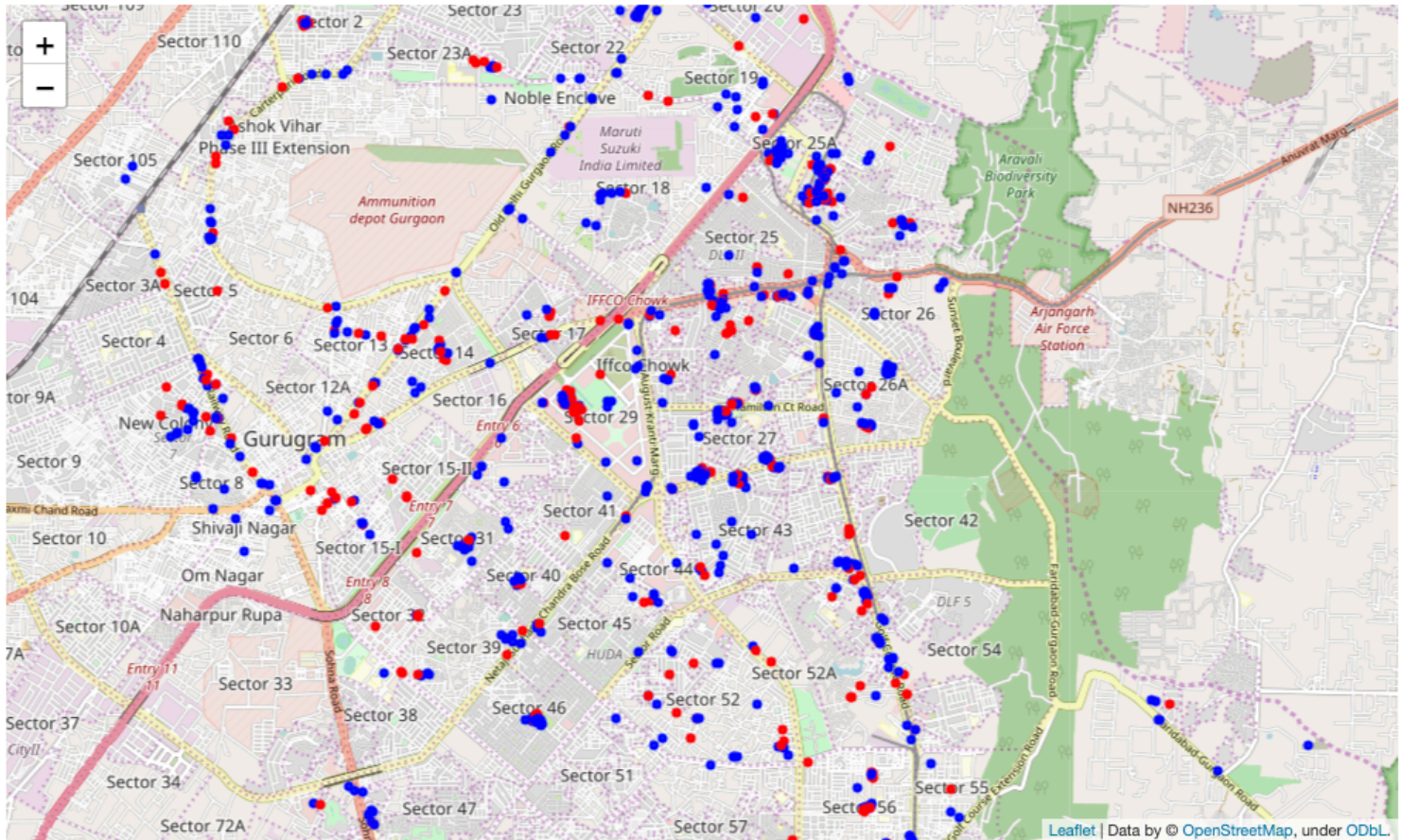Now plotting the same map for 'Chinese' cuisine.

Figure – Map of 'Chinese' restaurants in Gurgaon.

Here Red represents 'Chinese' while Blue represents any other type of cuisine. So it is clear that there are many Chinese restaurants but we can still open some more in few localities in the city.

5. Observations

✓ DLF Phase 3 is most crowded locality followed by Sector 56 & Sector 14. So it's not advisable to start a new restaurant here.
✓ North Indian is most sold cuisine followed by Chinese & Fast Food.
✓ There are still some locations where Chinese restaurant can be opened.

6. Suitable location with cuisine suggestion.

✓ *Café* can be opened in **'Palam Vihar'**
✓ *Chinese Restaurant* in **'Cyber Hub'**
✓ *Pizza store* on **'Sohna Road'**
✓ *Street Food* at **'Old Railway Road'**
✓ *Fast Food Centre* in **'Sector 29'**