PROJECT – ZOMATO-II

Question-1

Submitted by :-

Aryan Sharma

Q1.)

The dataset is highly skewed toward the cities included in Delhi-NCR. So, we will summarize all the other cities in Rest of India while those in New Delhi, Ghaziabad, Noida, Gurgaon, Faridabad to Delhi-NCR. Doing this would make our analysis turn toward Delhi-NCR v Rest of India.

1.1.)Plot the bar graph of number of restaurants present in Delhi NCR vs Rest of India

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('zomato.csv',encoding='latin-1')
data = data[data["Country Code"] == 1]
data_copy = data.copy()

NCR_cities = ["New Delhi" , "Ghaziabad" , "Noida" , "Gurgaon" , "Faridabad"]

count_dict = {"NCR":0 , "Rest_of_India" : 0}
```

```
def makeDict(cityname):
    if cityname in NCR_cities:
        count_dict["NCR"]+=1
    else:
        count_dict["Rest_of_India"]+=1

data_copy["City"].apply(makeDict)

keys = []
values = []
for key,value in count_dict.items() :
    keys.append(key)
    values.append(value)

plt.style.use("ggplot")
plt.bar(keys , values)
plt.xticks()
plt.xlabel("Location" , size = 20)
plt.ylabel("Number of restaurants",size = 20)
plt.show()
```

Explanation

- First the necessary libraries were included for the analysis.
- Then the given dataset/csv was loaded and the appropriate encoding which was latin-1 in this was specified.
- Since the analysis was supposed to be done for India so only the data from India was kept and for this we filtered using the Country Code which is 1 for India.
- A copy of the data was made so that the original data can be used again if needed.
- Since in the given question we have to plot a graph between number of restaurants in Delhi NCR vs Rest of India , a list of cities under Delhi NCR was made . The cities were New Delhi , Ghaziabad , Noida , Gurgaon and Faridabad .
- Now a dictionary was made which will keep a count of restaurants in Delhi NCR and Rest of India.

- Now we make a function named makeDict which takes in a cityname as an argument . This function is used to make the dictionary to keep count of the restaurant in Delhi-NCR or in Rest of India.
- The above function is then applied to the City column of the data and thus the dictionary is formed .
- Two lists namely keys and values are made in which we will store the keys and values from the dictionary keys being NCR or Rest of India.
- Finally Using the two lists we plot a bar graph of the number of restaurants in Delhi NCR and Rest of India.

Results

BAR GRAPH

From the bar graph we can see that there are a lot more restaurants in Delhi-NCR compared to Rest of India.

This result is on the basis of the given dataset which is not complete due to the Zomato API calls/day limit.

1.2) Find the cuisines which are not present in restaurant of Delhi NCR but present in rest of India.

Check using Zomato API whether this cuisines are actually not served in restaurants of Delhi-NCR or just it due to incomplete dataset.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('zomato.csv',encoding='latin-1')

data = data[data["Country Code"] == 1]

grouped_data = data.copy()

NCR_cities = ["New Delhi" , "Ghaziabad" , "Noida" , "Gurgaon" , "Faridabad"]

def changeName(cityname):
    if cityname in NCR_cities:
        return "NCR"
    else:
        return "Rest_Of_India"
```

```
grouped_data["City"] = grouped_data["City"].apply(changeName)


NCR_cuisines = set()
Rest_of_India_cuisines = set()


grouped_data["City_and_Cuisine"] = grouped_data["City"]+"///"+grouped_data["Cuisines"]


def findCuisine(CityCuisine):
    city = CityCuisine.split("///")[0]
    cuisine = CityCuisine.split("///")[1].split(", ")


    if city == "NCR":
        for i in cuisine:
            NCR_cuisines.add(i)
    else:
        for i in cuisine:
            Rest_of_India_cuisines.add(i)
```

```
grouped_data["City_and_Cuisine"].apply(findCuisine)

cuisines_not_in_NCR = []

for i in Rest_of_India_cuisines:
    if i not in NCR_cuisines:
        cuisines_not_in_NCR.append(i)

for i in cuisines_not_in_NCR:
    print(i)
```

Source Code – Checking using Zomato-API

```
import requests

r = requests.get('https://developers.zomato.com/api/v2.1/cuisines' ,headers = {'user-key' : '7f51a93a442e47390afafde1af103eaf'},
                 params = {"city_id" : 1})

ans = r.json()


cuisines = ans["cuisines"]

cuisines_in_delhi = []


for i in cuisines:
    cuisines_in_delhi.append(i["cuisine"]["cuisine_name"])
```

```
cuisines_not_in_NCR = ['BBQ','Cajun','Malwani','German']


for i in cuisines_not_in_NCR:
    if i in cuisines_in_delhi:
        print(i+" is actually served in NCR.")
```

Explanation

- First the necessary libraries were included for the analysis.
- Then the given dataset/csv was loaded and the appropriate encoding which was latin-1 in this was specified.
- Since the analysis was supposed to be done for India so only the data from India was kept and for this we filtered using the Country Code which is 1 for India.
- A copy of the data was made so that the original data can be used again if needed.
- Since in the given question we have to plot a graph between number of restaurants in Delhi NCR vs Rest of India , a list of cities under Delhi NCR was made . The cities were New Delhi , Ghaziabad , Noida , Gurgaon and Faridabad .
- Now the question is specifically for Delhi-NCR vs Rest of India so the cities which come under Delhi-NCR and are part of the NCR_cities list were grouped together and the ones which were not part of this list were also grouped together.

- For grouping the names of the cities were changed to NCR or Rest_Of_India accordingly. For this a function was made named changeName which changed the city name for the City column entries of the data.
- The changeName function was then applied to the City column of the data and this results in changed names of cities which will help us in classifying the cities as Delhi-NCR or Rest_of_India.
- Now two sets were made named NCR_cuisines and Rest_of_India_cuisines so as to get a list of all the cuisines in these two regions. Sets were made due to their property of keeping only unique elements and thus avoiding repetition.
- Then a new column was made named City_and_Cuisine which helped in keeping the City and Cuisine together in one entry. The City name and cuisine name were separated by “///” so as to extract each when needed.
- Now the sets are filled and for this a function is made named findCuisine which takes in entries from the City_and_Cuisine column and splits it into city name and cuisine name and appropriately adds the cuisine in the right set.
- The above function was then applied to the City_and_Cuisine column and the two sets of cuisines in both NCR and Rest of India were formed.
- Now a list is made named cuisines_not_in_NCR which will keep all the cuisines that are served in rest of India and not served in Delhi.For this we loop through cuisines in Rest of India and if the cuisine is not present in NCR cuisines we add them to this list.

Explanation – Checking using Zomato API

- Previously we found cuisines not served in NCR which are served in rest of India according to the given dataset.
- Now we checked if the results we got were actually true or not.
- The requests library was included so as to make the get requests to the Zomato api .
- We then find the cuisines that are served in Delhi using the Zomato API.
- Finally we again loop through cuisines not served in NCR from previous results and check against cuisines served in Delhi using Zomato API.
- We find that some of the cuisines which were considered not being served are actually served in Delhi.Results discussed later.

Results

- Using the given dataset we found out that the following cuisines are not served in Delhi-NCR but are served in Rest of India :-
  - German
  - BBQ
  - Malwani
  - Cajun
- We then check using the Zomato-API and find out that BBQ and Malwani are actually served in restaurants in Delhi.
- From this we can conclude that the dataset is incomplete and the results may vary.

1.3) Find the top 10 cuisines served by maximum number of restaurants in Delhi NCR and rest of India.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('zomato.csv',encoding='latin-1')
data = data[data["Country Code"] == 1]
data_copy = data.copy()

cuisines_count = {}

def cuisinecount(cuisine):
    l = cuisine.split(", ")
    for i in l:
        if i in cuisines_count.keys():
            cuisines_count[i]+=1
        else:
            cuisines_count[i] = 1
```

```
data_copy["Cuisines"].apply(cuisinecount)


cuisines_count = {key: value for key, value in sorted(cuisines_count.items(), key=lambda item: item[1],reverse =
True)}


cuisine = list(cuisines_count.keys())[:10]

count = list(cuisines_count.values())[:10]


print("Top 10 cuisines in Delhi-NCR and Rest of India are:- \n")

for i in list(cuisines_count.keys())[:10]:

    print(i)


plt.bar(cuisine,count)

plt.xticks(rotation=90)

plt.xlabel("Cuisines",size = 20)

plt.ylabel("Count", size = 20)

plt.show()
```

Explanation

- First the necessary libraries were included for the analysis.
- Then the given dataset/csv was loaded and the appropriate encoding which was latin-1 in this was specified.
- Since the analysis was supposed to be done for India so only the data from India was kept and for this we filtered using the Country Code which is 1 for India.
- A copy of the data was made so that the original data can be used again if needed.
- Then a dictionary was made named cuisines_count which would keep a count of a particular cuisine being served in a restaurant.

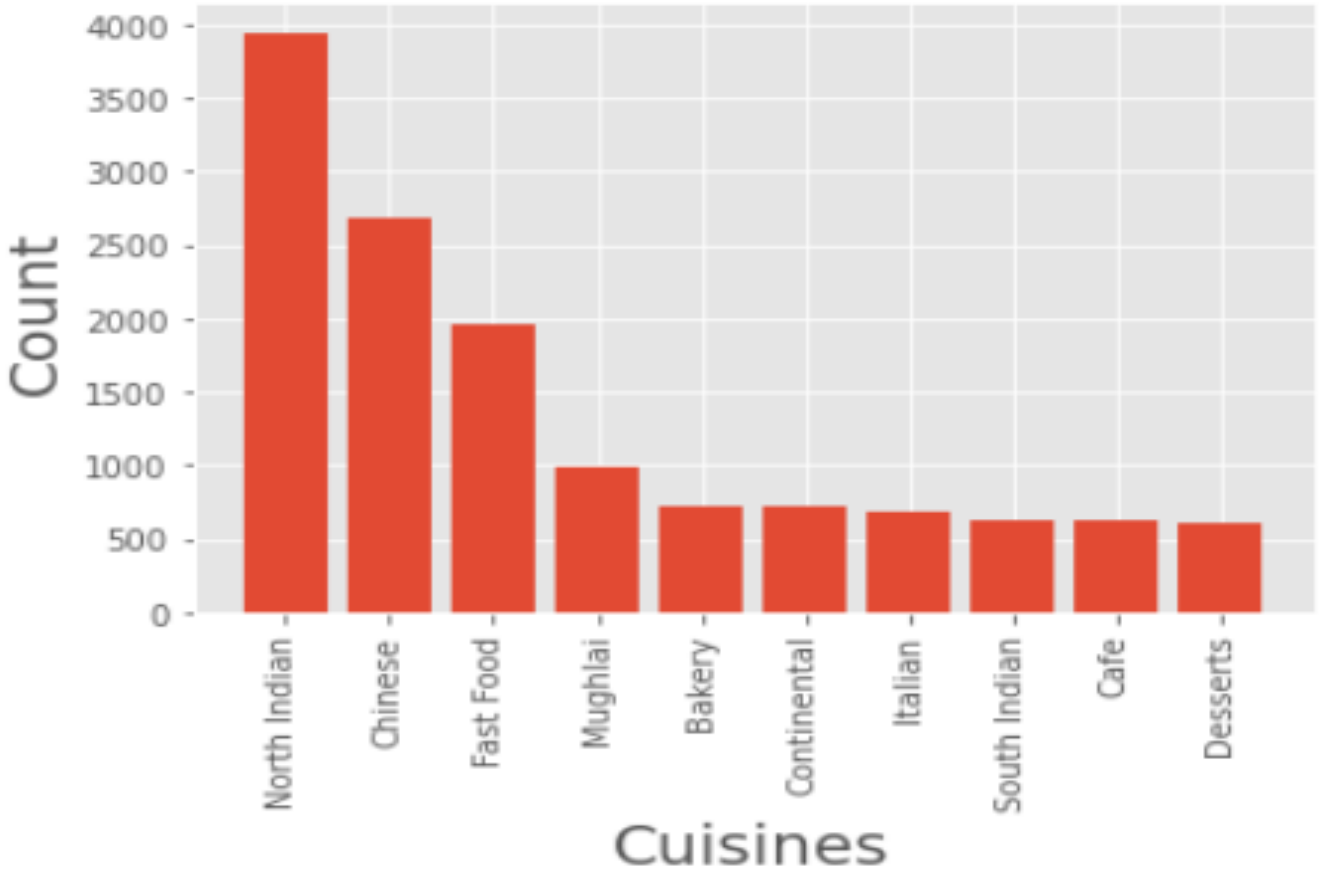
- A function named cuisinecount was made which takes in a cuisine from each of the entires of the Cuisines column of the data . Since a restaurant my offer multiple cuisines the cuisines are split first. Then the dictionary is formed to keep count of the cuisine being served at the restaurant.
- The above function is applied to the Cuisines column of the data and now we have a dictionary which gives how many restaurants serve a particular cuisine.
- We sort the dictionary on the basisi of value i.e. the count.
- Now two lists are made namely cuisine and count which store the top 10 most served cuisines and their count respectively.
- The top 10 cuisines and their count was printed.
- A bar graph was made between the cuisines and their count using the two lists formed in one of the previous steps.

Results

Top 10 cuisines in Delhi-NCR and Rest of India in order are:-

- North Indian
- Chinese
- Fast Food
- Mughlai
- Bakery
- Continental
- Italian
- South Indian
- Cafe
- Desserts

Bar graph


4000
3500
3000
2500
2000
1500
1000
500
0
Count
North Indian
Chinese
Fast Food
Mughlai
Bakery
Continental
Italian
South Indian
Cafe
Desserts
Cuisines

The top-10 cuisines served by maximum restaurants in Delhi-NCR and Rest of India were found and it found that North Indian is the most served cuisine and leads by a huge bargin , followed by Chinese and Fast Food.

1.4.) Write a short detailed analysis of how cuisine served is different from Delhi NCR to Rest of India.

Plot the suitable graph to explain your inference.

Source Code

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('zomato.csv' , encoding = 'latin-1')

data = data[data["Country Code"] == 1]

data_copy = data.copy()

NCR_cities = ["New Delhi" , "Ghaziabad" , "Noida" , "Gurgaon" , "Faridabad"]

def groupCity(cityname):
    if cityname in NCR_cities:
        return "NCR"
    else:
        return "Rest_Of_India"
```

```
data_copy["City"] = data_copy["City"].apply(groupCity)

NCR_cuisines = {}
Rest_of_India_cuisines = {}

data_copy["City_and_Cuisine"] = data_copy["City"]+"///"+data_copy["Cuisines"]

def findCuisine(CityCuisine):
    city = CityCuisine.split("///")[0]
    cuisine = CityCuisine.split("///")[1].split(", ")

    if city == "NCR":
        for i in cuisine:
            if i in NCR_cuisines.keys():
                NCR_cuisines[i] += 1
            else:
                NCR_cuisines[i] = 1
```

```
else:
        for i in cuisine:
            if i in Rest_of_India_cuisines.keys():
                Rest_of_India_cuisines[i] += 1
            else:
                Rest_of_India_cuisines[i] = 1


data_copy["City_and_Cuisine"].apply(findCuisine)


NCR_cuisines = {key: value for key, value in sorted(NCR_cuisines.items(),
key=lambda item: item[1],reverse = True)}

Rest_of_India_cuisines = {key: value for key, value in
sorted(Rest_of_India_cuisines.items(), key=lambda item: item[1],reverse = True)}
```

```
print(str(len(NCR_cuisines))+" cuisines are served in Delhi-NCR .")
print()
print(str(len(Rest_of_India_cuisines))+" cuisines are served in Rest of India .")
print()
top10_NCR_cuisines = list(NCR_cuisines.keys())[:10]
top10_NCR_cuisines_count = list(NCR_cuisines.values())[:10]

print("Top 10 Cuisines served in Delhi along with number of restaurants :- \n")
for i in top10_NCR_cuisines:
    print(i," : ",NCR_cuisines[i],end=" ")
    print()
print()
```

```
top10_RestOfIndia_cuisines = list(Rest_of_India_cuisines.keys())[:10]
top10_RestOfIndia_cuisines_count = list(Rest_of_India_cuisines.values())[:10]


print("Top 10 Cuisines served in Rest of India along with number of restaurants :- \n")
for i in top10_RestOfIndia_cuisines:
    print(i," : ",Rest_of_India_cuisines[i],end = " ")
    print()
print()
plt.style.use('ggplot')
plt.title("Cuisine vs Number of reataurants in Delhi-NCR" , size = 20)
plt.bar(top10_NCR_cuisines , top10_NCR_cuisines_count , color = 'red')
plt.xticks(rotation = 90)
plt.xlabel('Cuisine' ,size = 20)
plt.ylabel('Number of restaurants',size = 15)
plt.show()
```

```
plt.style.use('ggplot')
plt.title("Cuisine vs Number of reataurants in Rest of India" , size = 20)
plt.bar(top10_RestOfIndia_cuisines , top10_RestOfIndia_cuisines_count , color = 'b')
plt.xticks(rotation = 90)
plt.xlabel('Cuisine' ,size = 20)
plt.ylabel('Number of restaurants',size = 15)
plt.show()
print("Cuisines served in Delhi-NCR but not in Rest of India are:- \n")
for i in NCR_cuisines.keys() :
    if i not in Rest_of_India_cuisines.keys() :
        print(i)
print("\nCuisines served in Rest of India but not in Delhi-NCR are:- \n")
for i in Rest_of_India_cuisines.keys() :
    if i not in NCR_cuisines.keys() :
        print(i)
```

Explanation

- First the necessary libraries were included for the analysis.
- Then the given dataset/csv was loaded and the appropriate encoding which was latin-1 in this was specified.
- Since the analysis was supposed to be done for India so only the data from India was kept and for this we filtered using the Country Code which is 1 for India.
- A copy of the data was made so that the original data can be used again if needed.
- Since in the given question we have to plot a graph between number of restaurants in Delhi NCR vs Rest of India , a list of cities under Delhi NCR was made . The cities were New Delhi , Ghaziabad , Noida , Gurgaon and Faridabad .
- Now the question is specifically for Delhi-NCR vs Rest of India so the cities which come under Delhi-NCR and are part of the NCR_cities list were grouped together and the ones which were not part of this list were also grouped together.

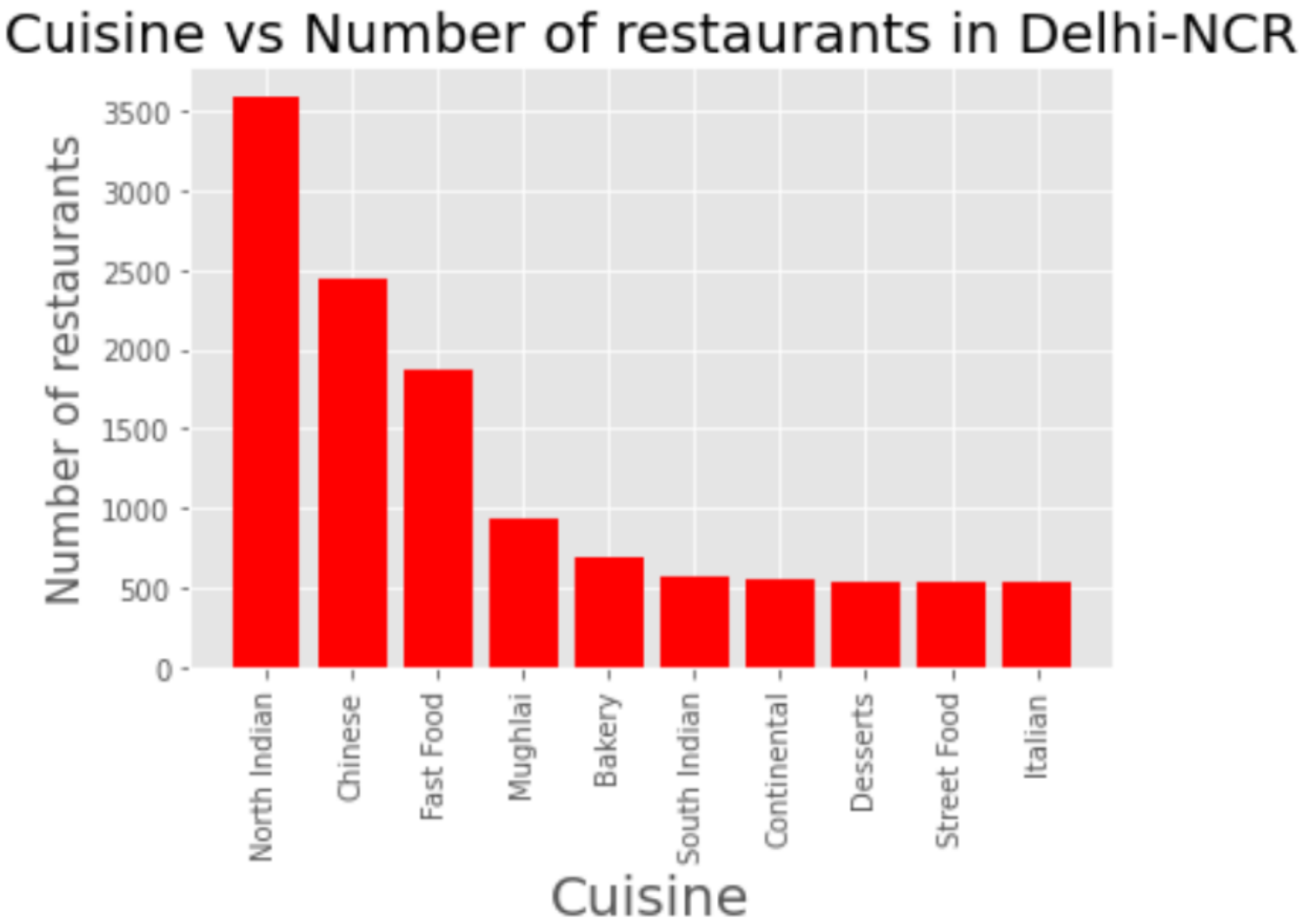
- For grouping the names of the cities were changed to NCR or Rest_Of_India accordingly. For this a function was made named groupCity which changed the city name for the City column entries of the data.
- The groupCity function was then applied to the City column of the data and this results in changed names of cities which will help us in classifying the cities as Delhi-NCR or Rest_of_India.
- Then two dictionaries were made named NCR_cuisines and Rest_of_India_cuisines .
- We then made a new column named City_and_Cuisine which stored the city name and cuisine as once entry separated by “///”.
- Then a function named findCuisine was made which helped in making the two dictionaries. It takes an entry from the City_and_Cuisine column and splits it into the city name and the cuisine and then checks if this city is from NCR or Rest of India and correspondingly adds cuisine as a key to the right dictionary and the count as value.
- The above function is then applied to the City_and_Cuisine column of the data and this results in the two dictionaries which store thr cuisines of NCR and Rest of India respectively and their corresponding counts.
- The two dictionaries were then sorted based on value.

- The number of cuisines served in Delhi NCR and Rest of India were printed.
- Two lists were then made which store the Top-10 cuisines of NCR and their count. The top 10 NCR cuisines were printed with their count.
- Two lists were then made which store the Top-10 cuisines of Rest of India and their count. The top 10 Rest of India cuisines were printed with their count.
- Two bar graphs were then plotted first for cuisine vs number of restaurants in Delhi-NCR and the other for Rest of India.
- The cuisines served in Delhi-NCR but not in Rest of India were printed.
- The cuisines served in Rest of India but not in Delhi-NCR were also printed.
- All the above were determined using the given dataset.

Results

- From the analysis it was observed that
  - 86 cuisines are served in Delhi-NCR .
  - 70 cuisines are served in Rest of India .
- Top 10 cuisines served in Delhi-NCR along with the number of restaurants are:-

Cuisine	Number of restaurants
North Indian	3597
Chinese	2448
Fast Food	1886
Mughlai	933
Bakery	697
South Indian	569
Continental	547
Desserts	542
Street Food	538
Italian	535


Cuisine vs Number of restaurants in Delhi-NCR
Number of restaurants
3500
3000
2500
2000
1500
1000
500
0
North Indian
Chinese
Fast Food
Mughlai
Bakery
South Indian
Continental
Desserts
Street Food
Italian
Cuisine

- Top 10 cuisines served in Rest of India along with the number of restaurants are:-

Cuisine	Number of restaurants
North Indian	349
Chinese	242
Continental	177
Italian	147
Cafe	136
Fast Food	97
South Indian	62
Mughlai	59
Desserts	55
Mexican	50


Cuisine vs Number of restaurants in Rest of India
Number of restaurants
0
50
100
150
200
250
300
350
North Indian
Chinese
Continental
Italian
Cafe
Fast Food
South Indian
Mughlai
Desserts
Mexican
Cuisine

Cuisines served in Delhi-NCR but not in Rest of India are:-

1.)Raw Meats

2.)Sushi

3.)Kashmiri

4.)Afghani

5.)Nepalese

6.)Turkish

7.)Naga

8.)Bihari

9.)Moroccan

10.)Assamese

11.)Iranian

12.)South American

13.)Drinks Only

14.)Pakistani

15.)Oriya

16.)Belgian

17.)Persian

18.)Cuisine Varies

19.)Deli

20.)Sri Lankan

Cuisines served in Rest of India but not in Delhi-NCR are:-

1.)German

2.)Malwani

3.)BBQ

4.)Cajun

From the above results it can be concluded that in both Delhi-NCR and Rest of India North Indian is the most popular cuisine but the number of restaurants serving this in Delhi-NCR vs Rest of India is significantly different with more than 3500 restaurants serving North Indian in Delhi-NCR vs just around 350 in Rest of India.

The Chinese cuisine is the second most popular cuisine in both the regions with the same trend as before of Delhi-NCR having significantly more restaurants serving the Chinese cuisine than Rest of India.

The third most popular cuisine in Delhi-NCR is Fast Food whereas it is Continental in Rest of India.