

## LEVEL 1 - TASK 1: TOP CUISINES

-- 1:1 Determine the top three most common cuisines in the dataset

-- 1:2 Calculate the percentage of restaurants that serve each of the top cuisines

1:1 DETERMINE THE TOP THREE MOST COMMON CUISINES IN THE DATASET.

```
#import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#import data
dataset = pd.read_csv("Dataset.csv")
```

```
#display the first few rows to understand its structure
dataset.head()
```

	Restaurant ID	Restaurant Name	Country Code	
City \				
0	6317637	Le Petit Souffle	162	Makati
City				
1	6304287	Izakaya Kikufuji	162	Makati
City				
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong
City				
3	6318506	Ooma	162	Mandaluyong
City				
4	6314302	Sambo Kojin	162	Mandaluyong
City				

	Address \
0	Third Floor, Century City Mall, Kalayaan Avenu...
1	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...
2	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...
3	Third Floor, Mega Fashion Hall, SM Megamall, O...
4	Third Floor, Mega Atrium, SM Megamall, Ortigas...

	Locality \
0	Century City Mall, Poblacion, Makati City
1	Little Tokyo, Legaspi Village, Makati City
2	Edsa Shangri-La, Ortigas, Mandaluyong City
3	SM Megamall, Ortigas, Mandaluyong City
4	SM Megamall, Ortigas, Mandaluyong City

	Locality Verbose	Longitude
Latitude \		
0	Century City Mall, Poblacion, Makati City, Mak...	121.027535
		14.565443

```

1 Little Tokyo, Legaspi Village, Makati City, Ma... 121.014101
14.553708
2 Edsa Shangri-La, Ortigas, Mandaluyong City, Ma... 121.056831
14.581404
3 SM Megamall, Ortigas, Mandaluyong City, Mandal... 121.056475
14.585318
4 SM Megamall, Ortigas, Mandaluyong City, Mandal... 121.057508
14.584450

```

```

Cuisines ... Currency Has Table
booking \
0 French, Japanese, Desserts ... Botswana Pula(P)
Yes
1 Japanese ... Botswana Pula(P)
Yes
2 Seafood, Asian, Filipino, Indian ... Botswana Pula(P)
Yes
3 Japanese, Sushi ... Botswana Pula(P)
No
4 Japanese, Korean ... Botswana Pula(P)
Yes

```

```

Has Online delivery Is delivering now Switch to order menu Price
range \
0 No No No
3
1 No No No
3
2 No No No
4
3 No No No
4
4 No No No
4

```

```

Aggregate rating Rating color Rating text Votes
0 4.8 Dark Green Excellent 314
1 4.5 Dark Green Excellent 591
2 4.4 Green Very Good 270
3 4.9 Dark Green Excellent 365
4 4.8 Dark Green Excellent 229

```

```
[5 rows x 21 columns]
```

```
#check database shape (rows and column)
```

```
dataset.shape
```

```
(9551, 21)
```

```
#check dataset information
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 9551 entries, 0 to 9550
```

```
Data columns (total 21 columns):
```

#	Column	Non-Null Count	Dtype
0	Restaurant ID	9551 non-null	int64
1	Restaurant Name	9551 non-null	object
2	Country Code	9551 non-null	int64
3	City	9551 non-null	object
4	Address	9551 non-null	object
5	Locality	9551 non-null	object
6	Locality Verbose	9551 non-null	object
7	Longitude	9551 non-null	float64
8	Latitude	9551 non-null	float64
9	Cuisines	9542 non-null	object
10	Average Cost for two	9551 non-null	int64
11	Currency	9551 non-null	object
12	Has Table booking	9551 non-null	object
13	Has Online delivery	9551 non-null	object
14	Is delivering now	9551 non-null	object
15	Switch to order menu	9551 non-null	object
16	Price range	9551 non-null	int64
17	Aggregate rating	9551 non-null	float64
18	Rating color	9551 non-null	object
19	Rating text	9551 non-null	object
20	Votes	9551 non-null	int64

```
dtypes: float64(3), int64(5), object(13)
```

```
memory usage: 1.5+ MB
```

```
#check dataset column names
```

```
dataset.columns
```

```
Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City',  
      'Address',  
      'Locality', 'Locality Verbose', 'Longitude', 'Latitude',  
      'Cuisines',  
      'Average Cost for two', 'Currency', 'Has Table booking',  
      'Has Online delivery', 'Is delivering now', 'Switch to order  
menu',  
      'Price range', 'Aggregate rating', 'Rating color', 'Rating  
text',  
      'Votes'],  
      dtype='object')
```

Data preprocessing

```
#check for null values
pd.isnull(dataset).sum()
```

```
Restaurant ID      0
Restaurant Name     0
Country Code       0
City               0
Address            0
Locality           0
Locality Verbose   0
Longitude          0
Latitude           0
Cuisines           9
Average Cost for two 0
Currency           0
Has Table booking  0
Has Online delivery 0
Is delivering now   0
Switch to order menu 0
Price range        0
Aggregate rating    0
Rating color       0
Rating text        0
Votes              0
dtype: int64
```

```
#Drop all null values
dataset.dropna(inplace=True)
```

```
#check Database
dataset.shape
```

```
(9542, 21)
```

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 9542 entries, 0 to 9550
```

```
Data columns (total 21 columns):
```

#	Column	Non-Null Count	Dtype
0	Restaurant ID	9542 non-null	int64
1	Restaurant Name	9542 non-null	object
2	Country Code	9542 non-null	int64
3	City	9542 non-null	object
4	Address	9542 non-null	object
5	Locality	9542 non-null	object
6	Locality Verbose	9542 non-null	object
7	Longitude	9542 non-null	float64
8	Latitude	9542 non-null	float64
9	Cuisines	9542 non-null	object

```

10 Average Cost for two 9542 non-null int64
11 Currency             9542 non-null object
12 Has Table booking    9542 non-null object
13 Has Online delivery  9542 non-null object
14 Is delivering now     9542 non-null object
15 Switch to order menu 9542 non-null object
16 Price range          9542 non-null int64
17 Aggregate rating     9542 non-null float64
18 Rating color         9542 non-null object
19 Rating text          9542 non-null object
20 Votes                9542 non-null int64

```

```
dtypes: float64(3), int64(5), object(13)
```

```
memory usage: 1.6+ MB
```

```
#check description
```

```
dataset[['Average Cost for two', 'Price range', 'Aggregate rating',
'Votes']].describe()
```

	Average Cost for two	Price range	Aggregate rating
Votes			
count	9542.000000	9542.000000	9542.000000
mean	1200.326137	1.804968	2.665238
std	156.772060	0.905563	1.516588
min	0.000000	1.000000	0.000000
25%	250.000000	1.000000	2.500000
50%	400.000000	2.000000	3.200000
75%	700.000000	2.000000	3.700000
max	800000.000000	4.000000	4.900000

```
#top three common cuisines in the dataset
```

```
cuisine_count =
dataset['Cuisines'].str.split(',').explode().str.strip().value_counts(
)
```

```
#str.split(',')- split the cuisine in the column into lists
```

```
#explode()- expand these lists into separate rows
```

```
#value_counts()- counts thoccurnece of cuisines
```

```
#determine the top 3 cuisine in the dataset
```

```
top_three_cuisines = cuisine_count.head(3)
```

```
print(top_three_cuisines)
```

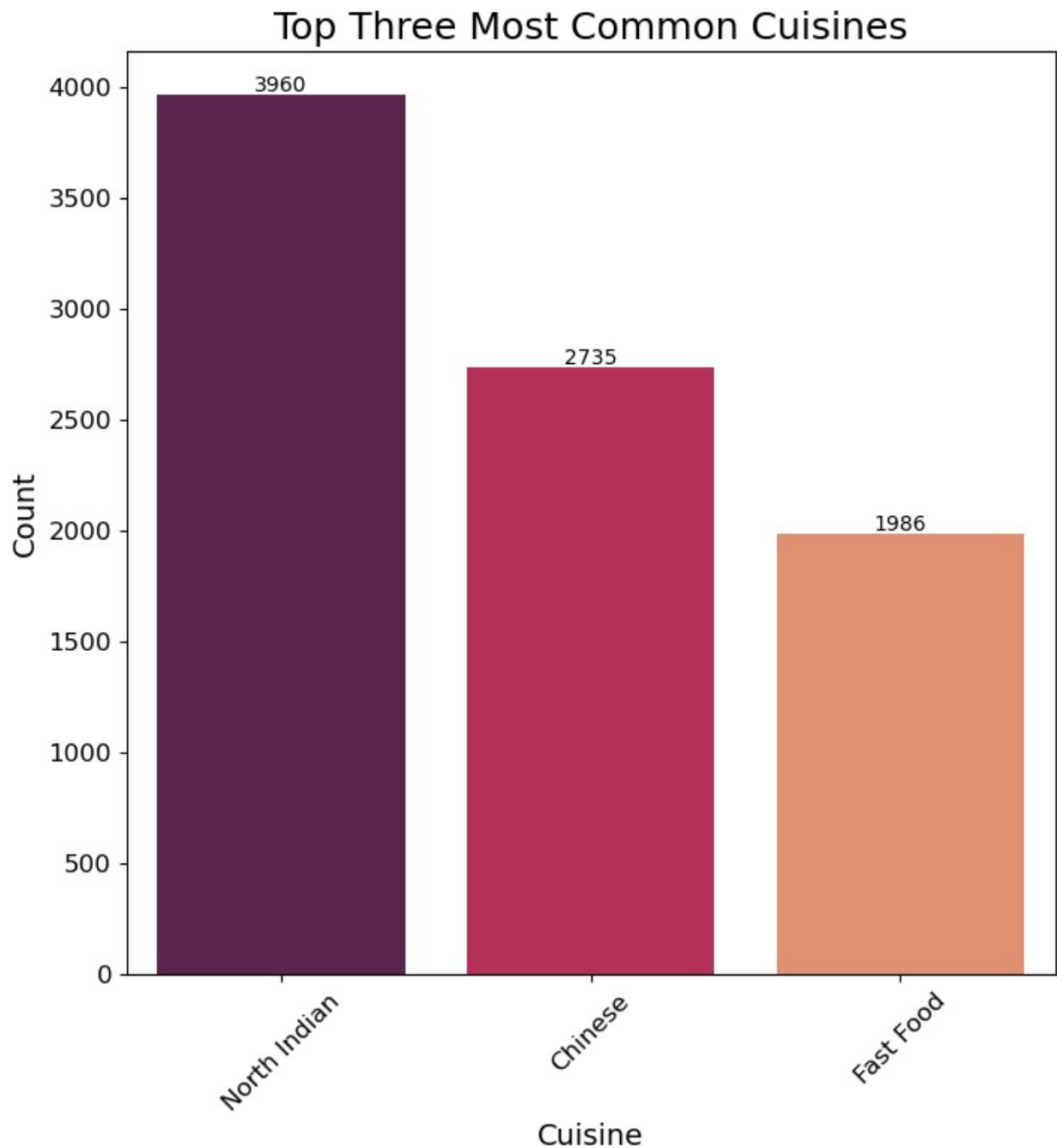
```
Cuisines
North Indian    3960
Chinese         2735
Fast Food       1986
Name: count, dtype: int64
```

```
fig, Ax= plt.subplots(figsize=(8,8))
sns.barplot(
    x=top_three_cuisines.index,
    y=top_three_cuisines.values,
    ax=Ax,
    palette="rocket",
)
for i,value in enumerate(top_three_cuisines.values):
    Ax.text(i,value + 15, #add offset above bar
            f'{value}', #display count
            ha="center",
            fontsize=10,
            color='black')
plt.title('Top Three Most Common Cuisines', fontsize=18)
plt.xlabel('Cuisine', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

```
C:\Users\Dimpi\AppData\Local\Temp\ipykernel_7952\1972996104.py:2:
FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(
```



1:2 CALCULATE THE PERCENTAGE OF RESTAURANTS THAT SERVE EACH OF THE TOP CUISINES.

```
#calculate the percentage  
total_restaurants = len(dataset)  
top_cuisines_percentage = (top_three_cuisines/ total_restaurants *  
100).round(2)  
  
#combine result into a dataframe
```

```
top_cuisines_dataset= pd.DataFrame({
    'Cuisine': top_three_cuisines.index,
    'Count': top_three_cuisines.values,
    'Percentage (%)': top_cuisines_percentage.values
})
```

```
print(top_cuisines_dataset)
```

	Cuisine	Count	Percentage (%)
0	North Indian	3960	41.50
1	Chinese	2735	28.66
2	Fast Food	1986	20.81

```
fig,ax = plt.subplots(figsize=(8,8))
ab=sns.barplot(x=top_cuisines_percentage.index,
y=top_cuisines_percentage.values, palette="rocket")
for i, value in enumerate(top_cuisines_percentage.values):
    ax.text(i, value + 1, # add offset above bar
            f'{value}', # display count
            ha='center',
            fontsize=10,
            color='black'
    )
plt.title('Percentage of Restasurants Serving Top Three Cuisines')
```

C:\Users\Dimpi\AppData\Local\Temp\ipykernel\_7952\1127148451.py:2:  
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
ab=sns.barplot(x=top_cuisines_percentage.index,
y=top_cuisines_percentage.values, palette="rocket")
```

```
Text(0.5, 1.0, 'Percentage of Restasurants Serving Top Three
Cuisines')
```



