COGNIFYZ TECHNOLOGY

IMPORTING ESSENTIAL LIBRARIES

In [3]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [7]:
```
from nltk.sentiment import SentimentIntensityAnalyzer
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from collections import Counter
```

In [9]:
```
import nltk
```

In [25]:
```
fp='interndata.csv'
idf=pd.read_csv(fp)
idf.head(5)
```

Out[25]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Cer P Ma
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Lit Ma
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Sh Man C
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	M Man
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	M Man

5 rows × 21 columns

In [21]:
```
idf.describe()
```

Out[21]:

	Restaurant ID	Country Code	Longitude	Latitude	Average Cost for two	P
count	9.551000e+03	9551.000000	9551.000000	9551.000000	9551.000000	9
mean	9.051128e+06	18.365616	64.126574	25.854381	1199.210763	
std	8.791521e+06	56.750546	41.467058	11.007935	16121.183073	
min	5.300000e+01	1.000000	-157.948486	-41.330428	0.000000	
25%	3.019625e+05	1.000000	77.081343	28.478713	250.000000	
50%	6.004089e+06	1.000000	77.191964	28.570469	400.000000	
75%	1.835229e+07	1.000000	77.282006	28.642758	700.000000	
max	1.850065e+07	216.000000	174.832089	55.976980	800000.000000	

No NaN values

LEVEL - 03

1. Analyze the text reviews to identify the most common positive and negative keywords.

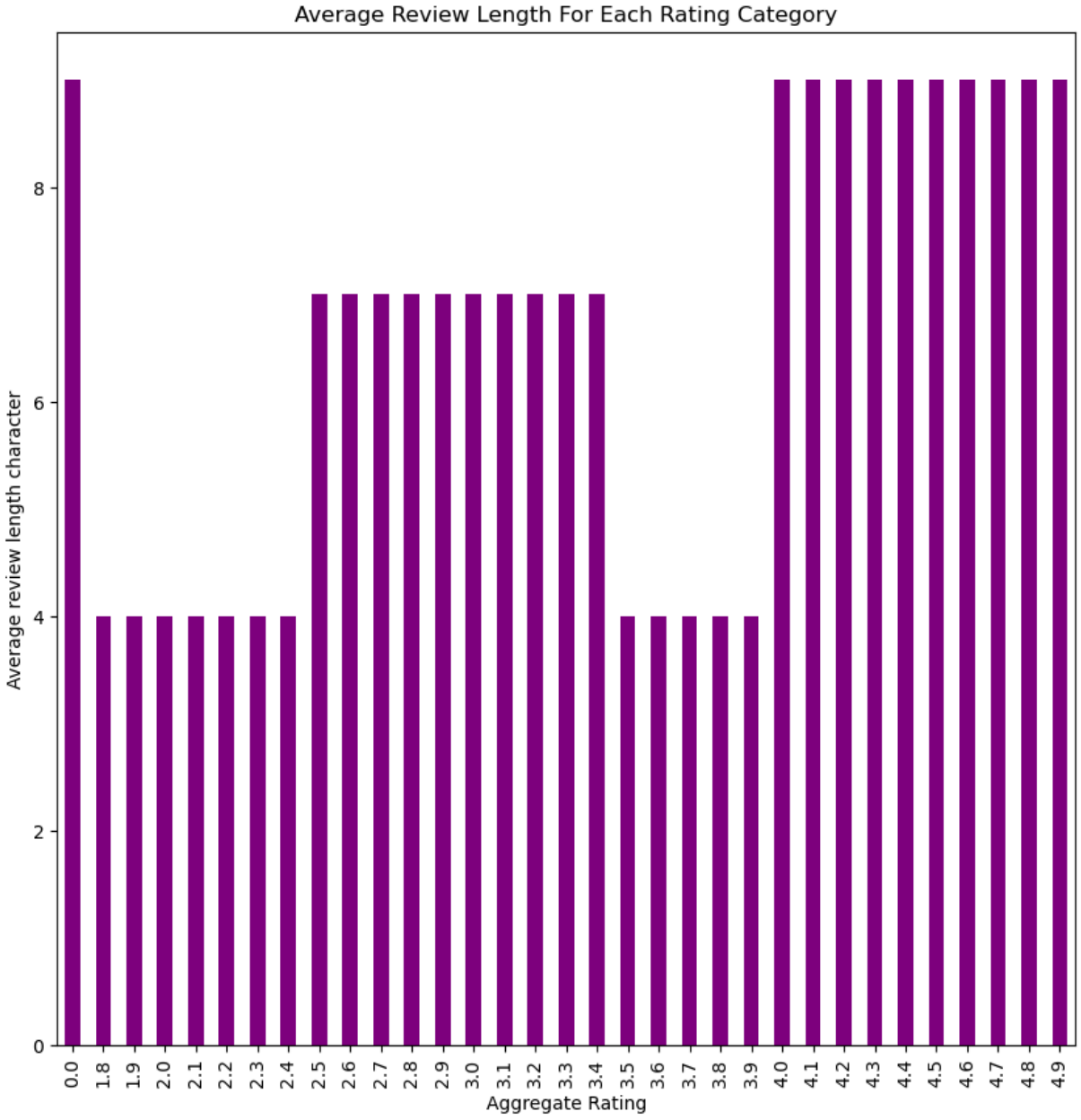
In [28]:
```
rating_texts=idf['Rating text']
rating_texts
```

Out[28]:
```
0       Excellent
1       Excellent
2       Very Good
3       Excellent
4       Excellent
          ...
9546    Very Good
9547    Very Good
9548         Good
9549    Very Good
9550    Very Good
Name: Rating text, Length: 9551, dtype: object
```

In [42]:
```
df_analyze=idf[['Rating text','Aggregate rating']].copy()
df_analyze['Review Length']=df_analyze['Rating text'].apply(lambda x: len(st
ave_rev_len=df_analyze.groupby('Aggregate rating')['Review Length'].mean()
```

In [46]:
```
plt.figure(figsize=(10,10))
ave_rev_len.plot(kind='bar',color='purple')
plt.title('Average Review Length For Each Rating Category')
```

```
plt.xlabel('Aggregate Rating')
plt.ylabel('Average review length character')
plt.show()
```



2.Identify the restaurants with the highest and lowest number of votes.

In [49]:
```
df_votes=idf[['Votes','Restaurant Name']]
df_votes
```

Out[49]:

	Votes	Restaurant Name
0	314	Le Petit Souffle
1	591	Izakaya Kikufuji
2	270	Heat - Edsa Shangri-La
3	365	Ooma
4	229	Sambo Kojin
...	...	...
9546	788	Naml\ Gurme
9547	1034	Ceviz A��ac\
9548	661	Huqqa
9549	901	A���k Kahve
9550	591	Walter's Coffee Roastery

9551 rows × 2 columns

In [51]:
```
print('Restaurant with highest Voyytes:')
max_votes=df_votes.loc[df_votes['Votes'].idxmax()]
max_votes
```

```
Restaurant with highest Voyytes:
```

Out[51]:
```
Votes              10934
Restaurant Name     Toit
Name: 728, dtype: object
```

In [53]:
```
print('Resaurant with lowest Votes:')
min_votes=df_votes.loc[df_votes['Votes'].idxmin()]
min_votes
```

```
Resaurant with lowest Votes:
```

Out[53]:
```
Votes                             0
Restaurant Name    Cantinho da Gula
Name: 69, dtype: object
```

Analyze if there is a correlation between the number of votes and the rating of a restaurant.

In [56]:
```
df_analysis = idf[['Votes','Aggregate rating']].copy()
df_analysis
```

Out[56]:

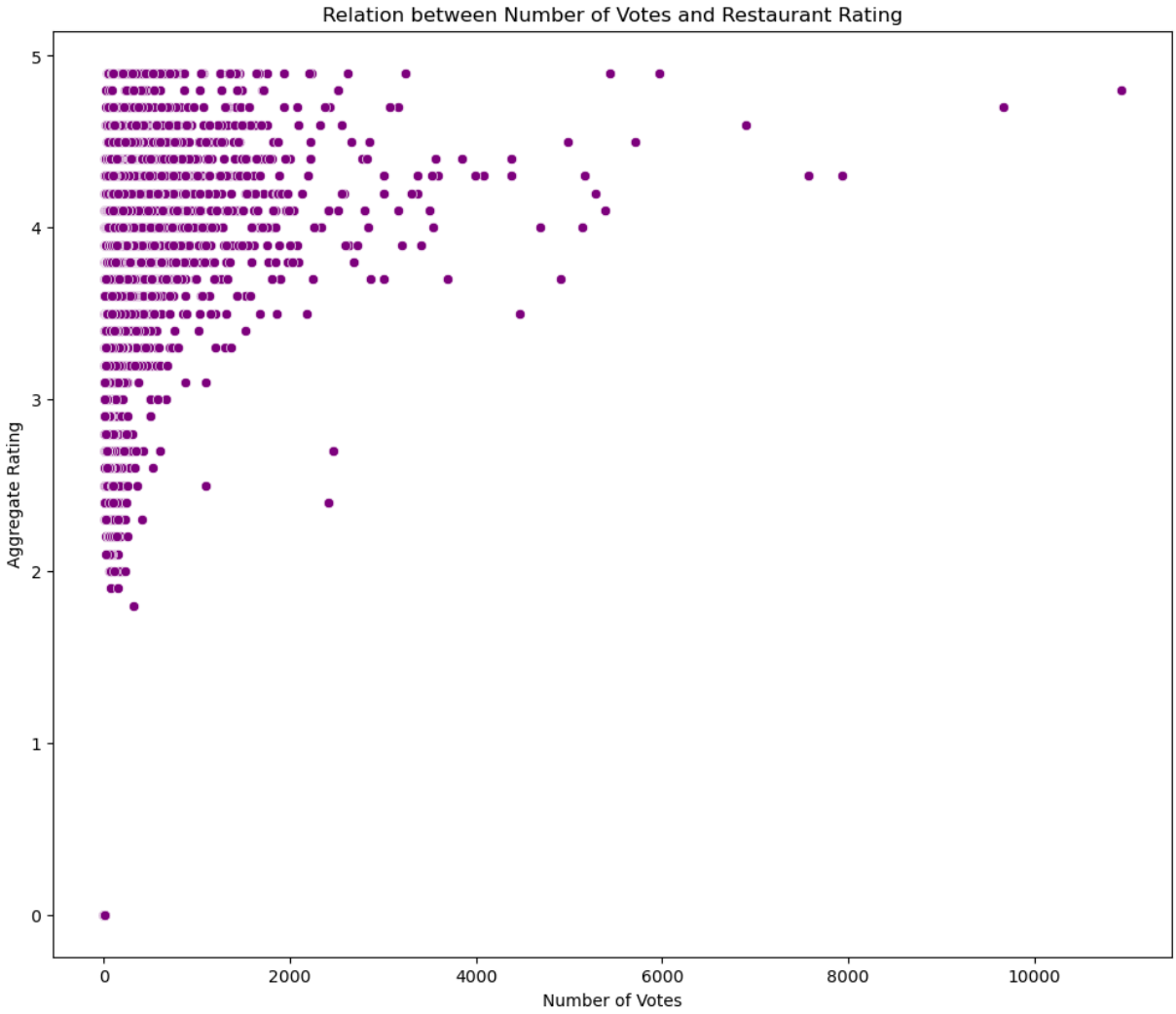
	Votes	Aggregate rating
0	314	4.8
1	591	4.5
2	270	4.4
3	365	4.9
4	229	4.8
...	...	...
9546	788	4.1
9547	1034	4.2
9548	661	3.7
9549	901	4.0
9550	591	4.0

9551 rows × 2 columns

In [58]:
```
corr=df_analysis['Votes'].corr(df_analysis["Aggregate rating"])
corr
```

Out[58]:
```
0.3136905841954117
```

In [66]:
```
plt.figure(figsize=(12,10))
sns.scatterplot(x='Votes',y='Aggregate rating',data=df_analysis , color = 'p
plt.title('Relation between Number of Votes and Restaurant Rating')
plt.xlabel("Number of Votes")
plt.ylabel('Aggregate Rating')
plt.show()
```



3. Price Range vs. Online Delivery and Table Booking

Analyze if there is a relationship between the price range and the availability of online delivery and table booking

In [70]:
```
idf.head(1)
```

Out[70]:

	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Lo
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Century City Mall, Poblacion, Makati City, Mak...	121

1 rows × 21 columns

In [72]:
```
idf_analysis=idf[['Price range','Has Online delivery','Has Table booking']].
idf_analysis['Has Online delivery']=idf_analysis['Has Online delivery'].map(
idf_analysis['Has table booking']=idf_analysis['Has Table booking'].map({'Ye
```

In [74]:
```
summary_table=pd.pivot_table(idf_analysis,index='Price range',values=['Has C
```

In [76]:
```
print('Summary Table:')
summary_table
```

```
Summary Table:
```

Out[76]:

Price range	Has Online delivery	Has table booking
1	701	1
2	1286	239
3	411	644
4	53	274

In [78]:
```
summary_table
```

Out[78]:

Price range	Has Online delivery	Has table booking
1	701	1
2	1286	239
3	411	644
4	53	274

In [80]:
```
idf.head(5)
```

Out[80]:

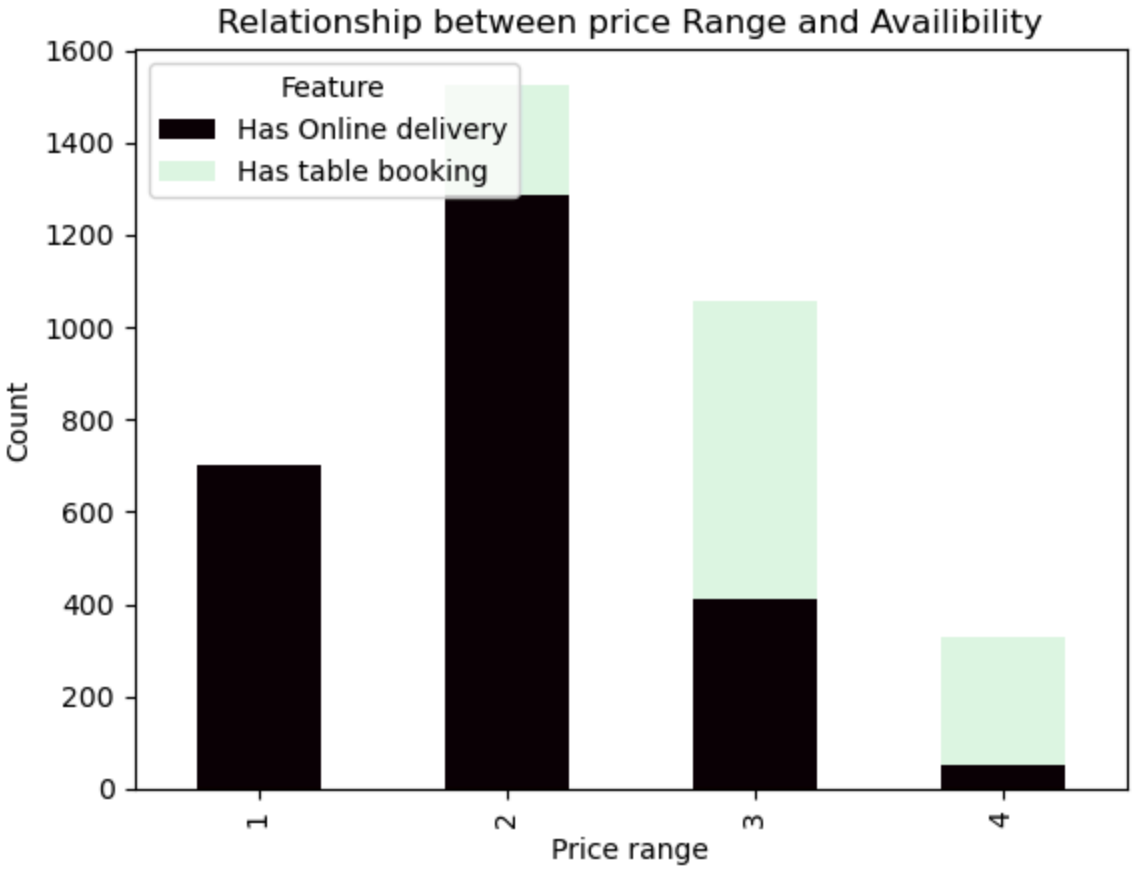
	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	
0	6317637	Le Petit Souffle	162	Makati City	Third Floor, Century City Mall, Kalayaan Avenu...	Century City Mall, Poblacion, Makati City	Cer P Ma
1	6304287	Izakaya Kikufuji	162	Makati City	Little Tokyo, 2277 Chino Roces Avenue, Legaspi...	Little Tokyo, Legaspi Village, Makati City	Lit Ma
2	6300002	Heat - Edsa Shangri-La	162	Mandaluyong City	Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...	Edsa Shangri-La, Ortigas, Mandaluyong City	Sh Man C
3	6318506	Ooma	162	Mandaluyong City	Third Floor, Mega Fashion Hall, SM Megamall, O...	SM Megamall, Ortigas, Mandaluyong City	M Man
4	6314302	Sambo Kojin	162	Mandaluyong City	Third Floor, Mega Atrium, SM Megamall, Ortigas...	SM Megamall, Ortigas, Mandaluyong City	M Man

5 rows × 21 columns

In [109…

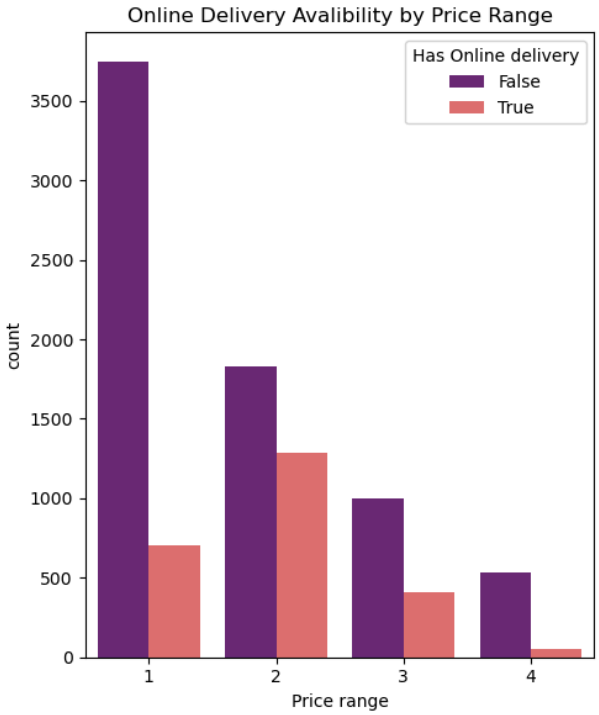
```
plt.figure(figsize=(10,8))
summary_table.plot(kind='bar',stacked=True,colormap='mako')
plt.title('Relationship between price Range and Availibility')
plt.xlabel('Price range')
plt.ylabel('Count')
plt.legend(title='Feature',loc='upper left')
plt.show()
```

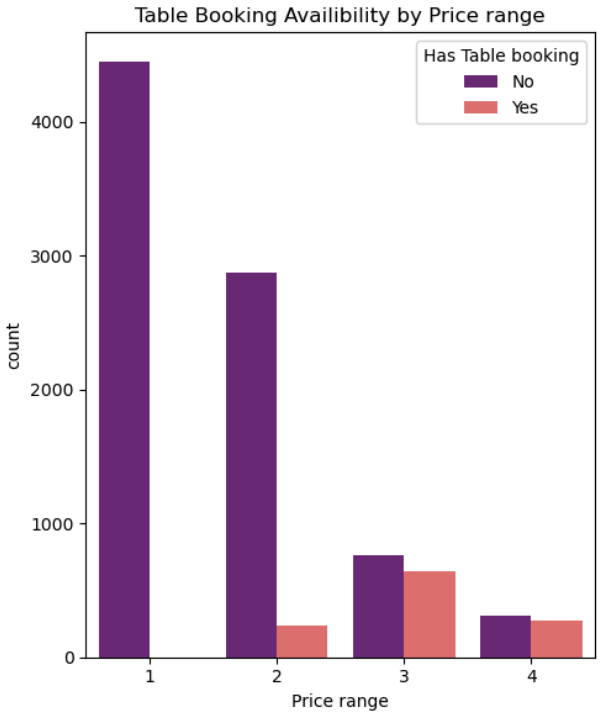
```
<Figure size 1000x800 with 0 Axes>
```



Determine if higher-priced restaurants are more likely to offer these services

In [99]:
```
plt.figure(figsize=(10,6))

plt.subplot(1,2,1)

sns.countplot(x='Price range' , hue='Has Online delivery' , data=idf_analysi
plt.title('Online Delivery Avalibility by Price Range')

plt.subplot(1,2,2)
sns.countplot(x='Price range', hue='Has Table booking', data=idf_analysis,pa
plt.title('Table Booking Availibility by Price range')

plt.tight_layout()
plt.show()
```





In [ ]: