LEVEL 3 - TASK 2: RESTAURANT REVIEWS

--2:1 Analyze the text reviews to identify the most common positive and negative keywords.

--2:2 Calculate the average length of reviews and explore if there is a relationship between review length and rating.

2:1 ANALYZE THE TEXT REVIEWS TO IDENTIFY THE MOST COMMON POSITIVE AND NEGATIVE KEYWORDS.

```python
#import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

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

#check data
dataset.head(10)
```

```
   Restaurant ID                         Restaurant Name  Country
Code  \
0        6317637                         Le Petit Souffle
162
1        6304287                         Izakaya Kikufuji
162
2        6300002                      Heat - Edsa Shangri-La
162
3        6318506                                     Ooma
162
4        6314302                              Sambo Kojin
162
5       18189371                             Din Tai Fung
162
6        6300781                               Buffet 101
162
7        6301290                                  Vikings
162
8        6300010  Spiral - Sofitel Philippine Plaza Manila
162
9        6314987                                 Locavore
162


             City                                        Address
\
0      Makati City  Third Floor, Century City Mall, Kalayaan Avenu...
```

```
1         Makati City  Little Tokyo, 2277 Chino Roces Avenue, Legaspi...

2    Mandaluyong City  Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...

3    Mandaluyong City  Third Floor, Mega Fashion Hall, SM Megamall, O...

4    Mandaluyong City  Third Floor, Mega Atrium, SM Megamall, Ortigas...

5    Mandaluyong City  Ground Floor, Mega Fashion Hall, SM Megamall, ...

6         Pasay City  Building K, SM By The Bay, Sunset Boulevard, M...

7         Pasay City  Building B, By The Bay, Seaside Boulevard, Mal...

8         Pasay City  Plaza Level, Sofitel Philippine Plaza Manila, ...

9         Pasig City  Brixton Technology Center, 10 Brixton Street, ...


                                          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
5            SM Megamall, Ortigas, Mandaluyong City
6   SM by the Bay, Mall of Asia Complex, Pasay City
7   SM by the Bay, Mall of Asia Complex, Pasay City
8       Sofitel Philippine Plaza Manila, Pasay City
9                                          Kapitolyo

                                    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
5  SM Megamall, Ortigas, Mandaluyong City, Mandal...  121.056314
14.583764
6  SM by the Bay, Mall of Asia Complex, Pasay Cit...  120.979667
14.531333
7  SM by the Bay, Mall of Asia Complex, Pasay Cit...  120.979333
14.540000
8  Sofitel Philippine Plaza Manila, Pasay City, P...  120.980090
```

```
14.552990
9                                    Kapitolyo, Pasig City  121.056532
14.572041

                              Cuisines  ...          Currency  \
0          French, Japanese, Desserts  ...  Botswana Pula(P)
1                            Japanese  ...  Botswana Pula(P)
2    Seafood, Asian, Filipino, Indian  ...  Botswana Pula(P)
3                     Japanese, Sushi  ...  Botswana Pula(P)
4                    Japanese, Korean  ...  Botswana Pula(P)
5                             Chinese  ...  Botswana Pula(P)
6                     Asian, European  ...  Botswana Pula(P)
7  Seafood, Filipino, Asian, European  ...  Botswana Pula(P)
8            European, Asian, Indian  ...  Botswana Pula(P)
9                            Filipino  ...  Botswana Pula(P)

  Has Table booking Has Online delivery Is delivering now  \
0               Yes                  No                No
1               Yes                  No                No
2               Yes                  No                No
3                No                  No                No
4               Yes                  No                No
5                No                  No                No
6               Yes                  No                No
7               Yes                  No                No
8               Yes                  No                No
9               Yes                  No                No

  Switch to order menu Price range  Aggregate rating  Rating color  \
0                   No           3               4.8    Dark Green
1                   No           3               4.5    Dark Green
2                   No           4               4.4         Green
3                   No           4               4.9    Dark Green
4                   No           4               4.8    Dark Green
5                   No           3               4.4         Green
6                   No           4               4.0         Green
7                   No           4               4.2         Green
8                   No           4               4.9    Dark Green
9                   No           3               4.8    Dark Green

  Rating text Votes
0   Excellent   314
1   Excellent   591
2   Very Good   270
3   Excellent   365
4   Excellent   229
5   Very Good   336
6   Very Good   520
7   Very Good   677
8   Excellent   621
```

```
9    Excellent    532

[10 rows x 21 columns]
```

```python
#check database shape
dataset.shape
```

```
(9551, 21)
```

```python
#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
```

```python
#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 of data
```
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
9542.000000
mean            1200.326137     1.804968          2.665238
156.772060
std            16128.743876     0.905563          1.516588
430.203324
min                0.000000     1.000000          0.000000
0.000000
25%              250.000000     1.000000          2.500000
5.000000
50%              400.000000     2.000000          3.200000
31.000000
75%              700.000000     2.000000          3.700000
130.000000
max           800000.000000     4.000000          4.900000
10934.000000
```

```
pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\dimpi\anaconda3\
lib\site-packages (1.9.4)
Requirement already satisfied: numpy>=1.6.1 in c:\users\dimpi\
```

```
anaconda3\lib\site-packages (from wordcloud) (1.26.4)
Requirement already satisfied: pillow in c:\users\dimpi\anaconda3\lib\
site-packages (from wordcloud) (10.4.0)
Requirement already satisfied: matplotlib in c:\users\dimpi\anaconda3\
lib\site-packages (from wordcloud) (3.9.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (24.1)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\dimpi\
anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\dimpi\anaconda3\
lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud)
(1.16.0)
Note: you may need to restart the kernel to use updated packages.

pip install Counter

Requirement already satisfied: Counter in c:\users\dimpi\anaconda3\
lib\site-packages (1.0.0)
Note: you may need to restart the kernel to use updated packages.
```

```python
from wordcloud import WordCloud
from collections import Counter

#categorise reviews into positive,negative or neutral
positive_reviews = dataset[dataset["Rating
text"].isin(["Excellent","Very Good", "Good"])]
neutral_reviews = dataset[dataset["Rating text"]=="Average"]
negative_reviews = dataset[dataset["Rating
text"].isin(["Poor","Terrible"])]

#extract words from rating text
positive_words = " ".join(positive_reviews["Rating
text"]).lower().split()
neutral_words = " ".join(neutral_reviews["Rating
text"]).lower().split()
negative_words = " ".join(negative_reviews["Rating
text"]).lower().split()

#get the most common words
positive_common = Counter(positive_words).most_common(10)
```

```
neutral_common = Counter(neutral_words).most_common(10)
negative_common = Counter(negative_words).most_common(10)

#generate word clouds
fig, axes = plt.subplots(1,3,figsize=(18,6))
wordcloud_positive =
WordCloud(background_color="white").generate("".join(positive_words))
wordcloud_neutral =
WordCloud(background_color="white").generate("".join(neutral_words))
wordcloud_negative =
WordCloud(background_color="white").generate("".join(negative_words))

axes[0].imshow(wordcloud_positive, interpolation="bilinear")
axes[0].set_title("Positive Reviews", fontsize=14)
axes[0].axis("off")

axes[1].imshow(wordcloud_neutral, interpolation="bilinear")
axes[1].set_title("Neutral Reviews", fontsize=14)
axes[1].axis("off")

axes[2].imshow(wordcloud_negative, interpolation="bilinear")
axes[2].set_title("Negative Reviews", fontsize=14)
axes[2].axis("off")

plt.show()

#save the visualization
fig.savefig("wordcloud_visualization.png",
dpi=300,bbox_inches='tight')


#display common words in each category
positive_common, neutral_common, negative_common
```

Positive Reviews   Neutral Reviews   Negative Reviews

```
([('good', 3174), ('very', 1078), ('excellent', 300)],
 [('average', 3734)],
 [('poor', 186)])
```

2:2 CALCULATE THE AVERAGE LENGTH OF REVIEWS AND EXPLORE IF THERE IS A
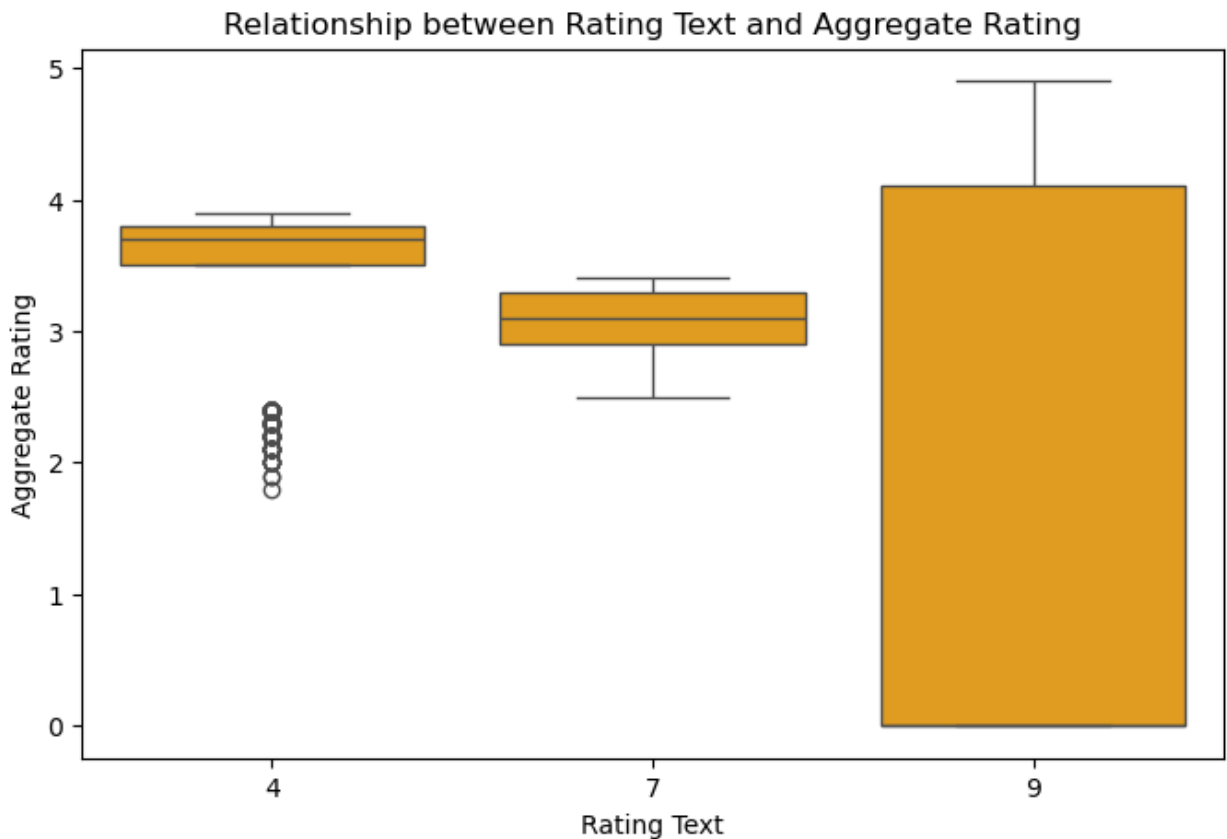RELATIONSHIP BETWEEN REVIEW LENGTH AND RATING.

```python
#visualize the relationship between rating text and aggregate rating
plt.figure(figsize=(8,5))
sns.boxplot(x="Rating text", y="Aggregate rating",
data=dataset,color="orange")
plt.title("Relationship between Rating Text and Aggregate Rating")
plt.xlabel("Rating Text")
plt.ylabel("Aggregate Rating")
plt.show()

#calculate correlation between rating text and aggregate rating
correlation= dataset["Rating text"].corr(dataset["Aggregate rating"])

#return correlation value
correlation
```


Relationship between Rating Text and Aggregate Rating

-0.47908961921650234