TABLE OF CONTENTS

Title

The application of text mining and sentiment analysis on 30 randomly selected hotels and restaurants in beach areas from the provided dataset.

Introduction

Reviews are a great way to measure or understand the sentiments of people regarding a service or product. Reviews can be described as a way by which a person seeks to describe, criticize or appraise goods and services that were utilized.

Sentiment Analysis is the scrutiny of user reviews to get a general sentiment (positive or negative) of the user towards goods and services. This can be beneficial to both the sellers and buyers. For the sellers, sentiment analysis shows how a product or service is accepted by the consumers and also shows areas within which improvements can be carried out. For buyers, sentiment analysis helps them understand the gains or risks involved in buying a product or using a service. (Mehta & Pandya, 2020)

Studies have shown that about 94% of consumers read online reviews, and 92% used a service or purchased a product based on the influence of reviews. (Inc, 2021)

The main objective of this experiment is to determine the general mood or sentiment of people who have used the services of the hotels/restaurant based in beach locations.

Datasets

The dataset used shows the reviews by tourists about accommodations in hotels and restaurants. The dataset contains 53644 rows and 5 columns. The columns are the ID, Review Date, Location, Hotel/Restaurant name, and Review.

Explanation and preparation of datasets

The initial attempt to read the .csv dataset into a variable met an error because the default utf-8 encoding was not suitable to decrypt the dataset.

```
dataset = pd.read_csv('tourist_accommodation_reviews.csv')
dataset.info()
dataset.head()
```

```
   1042     def _failover_to_python(self):

~\anaconda3\lib\site-packages\pandas\io\parsers\c_parser_wrapper.py in __init__(self, src, **kwds)
     67         kwds["dtype"] = ensure_dtype_objs(kwds.get("dtype", None))
     68         try:
---> 69             self._reader = parsers.TextReader(self.handles.handle, **kwds)
     70         except Exception:
     71             self.handles.close()

~\anaconda3\lib\site-packages\pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader.__cinit__()

~\anaconda3\lib\site-packages\pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader._get_header()

~\anaconda3\lib\site-packages\pandas\_libs\parsers.pyx in pandas._libs.parsers.TextReader._tokenize_rows()

~\anaconda3\lib\site-packages\pandas\_libs\parsers.pyx in pandas._libs.parsers.raise_parser_error()

UnicodeDecodeError: 'utf-8' codec can't decode byte 0x92 in position 2455: invalid start byte
```

Changing the encoder to ‘cp850’ solved this problem, and the dataset was loaded into a ‘dataset’ variable.

A look into the dataset shows that there are no missing values and it has 53644 rows and 5 columns. See below:

```
#encoder 'cp1252' Western Europe encoding to read the characters
dataset = pd.read_csv('tourist_accommodation_reviews.csv', encoding='cp850')
dataset.info()
dataset.head()
```

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ID                    53644 non-null  object
 1   Review Date           53644 non-null  object
 2   Location              53644 non-null  object
 3   Hotel/Restaurant name  53644 non-null  object
 4   Review                53644 non-null  object
dtypes: object(5)
memory usage: 2.0+ MB
```

	ID	Review Date	Location	Hotel/Restaurant name	Review
0	rn579778340	Reviewed 1 week ago	Kathu	Thong Dee The Kathu Brasserie	Just been for sunday roast lamb and beef truly...
1	rn576350875	Reviewed 3 weeks ago	Kathu	Thong Dee The Kathu Brasserie	Quietly set off the main road, nice atmosphere...
2	rn574921678	Reviewed 4 weeks ago	Kathu	Thong Dee The Kathu Brasserie	I made a reservation for a birthday two days i...
3	rn572905503	Reviewed April 12, 2018	Kathu	Thong Dee The Kathu Brasserie	We visit here regularly and never fail to be i...
4	rn572364712	Reviewed April 10, 2018	Kathu	Thong Dee The Kathu Brasserie	Visited this wonderful place on my travels and...

Focusing on the location column, a count of the different locations was made and the result is as shown:

```
In [9]: dataset.Location.value_counts() # to view the counts(frequency) of each location
```

```
Out[9]: Patong              16403
        Karon                5826
        Kata Beach           5752
        Rawai                3811
        Choeng Thale         3378
        Phuket Town          3356
        Kamala               3162
        Mai Khao             2372
        Cape Panwa           1500
        Chalong              1287
        Thalang District     1177
        Kathu                1078
        Nai Yang              996
        Nai Harn              881
        Bang Tao Beach        600
        Karon Beach           397
        Wichit                395
        Talat Yai             300
        Koh Kaew              293
        Kata Noi Beach        200
        Pa Khlok              100
        Ratsada                98
        Talat Nuea             97
        Nai Thon               94
        Sakhu                  91
        Name: Location, dtype: int64
```

Hotel selection

The 30 hotels and restaurants selected were done based on **beach locations**. To achieve this, the locations were first put in a list, then the dataset was filtered based on that list using the '.isin' function. See the code & result below:

```
#Selecting hotels based on beach location
beach_locations = [' Kata Beach',' Bang Tao Beach',' Karon Beach',' Kata Noi Beach']
beach_hotels_rest = dataset[dataset.Location.isin(beach_locations)]
beach_hotels_rest
```

	ID	Review Date	Location	Hotel/Restaurant name	Review
100	rn581307988	Reviewed yesterday	Kata Beach	Odysseus Greek Organic Restaurant	Food was tasty and fresh. Fast service. The ow...
101	rn580977661	Reviewed 2 days ago	Kata Beach	Odysseus Greek Organic Restaurant	Great variety of Greek dishes and fantastic se...
102	rn580827047	Reviewed 3 days ago	Kata Beach	Odysseus Greek Organic Restaurant	We had an excellent culinary experience at thi...
103	rn580517333	Reviewed 5 days ago	Kata Beach	Odysseus Greek Organic Restaurant	Amazing service and food! Highly recommend if ...
104	rn580035619	Reviewed 1 week ago	Kata Beach	Odysseus Greek Organic Restaurant	This restaurant is tucked away near the square...
...	...	...	...	...	...
53050	rn106057445	Reviewed April 29, 2011	Kata Beach	Mali Seafood Restaurant & Bar	Large range with western choices and good loca...
53051	rn101870163	Reviewed March 28, 2011	Kata Beach	Mali Seafood Restaurant & Bar	was there today. food is average I would say. ...
53052	rn95346942	Reviewed February 2, 2011	Kata Beach	Mali Seafood Restaurant & Bar	Mali is a great, cheap friendly restaurant wit...
53053	rn90299395	Reviewed December 20, 2010	Kata Beach	Mali Seafood Restaurant & Bar	We dined here in December, 2010. The food was ...
53054	rn89302857	Reviewed December 7, 2010	Kata Beach	Mali Seafood Restaurant & Bar	We stayed up the road in the sugar palm grand ...

6949 rows × 5 columns

The count of the filtered hotels and the list showing them were taken:

```
In [47]: #Number of unique hotels/restaurant
         len(beach_hotels_rest['Hotel/Restaurant name'].unique())

Out[47]: 71

In [82]: beach_hotels_rest['Hotel/Restaurant name'].unique()

Out[82]: array(['Odysseus Greek Organic Restaurant', 'The Tavern',
                'EAT. bar & grill', 'Istanbul Turkish Restaurant',
                'Katatürk Turkish Restaurant', 'Red Duck Restaurant',
                'Pooh and Friends', 'The Family Restaurant',
                'Pomodoro Pizza Restaurant', 'En Vogue Restaurant',
                'Burger House Kata Beach', 'Palm Square',
                'Autogrill Risto Bar Pizza', 'Kata On Fire Bar and Grill',
                'New York Burger Co.', 'Sabai Corner', 'Eightfold Restaurant',
                'Riverside Restaurant', 'Red Chair Restaurant',
                'MaMa Jin Restaurant', 'Curry Delight Indian Restaurant',
                "Mom Tri's Kitchen at Villa Royale",
                'Sugar & Spice Restaurant at Dome Resort', 'Kampong Kata Hill',
                'The Kitchen Restaurant Kata Beach & Thai Cooking Class',
                'Chai Thaifood Restaurant', 'Red Corner', 'Two Chefs Kata Center',
                '44 Thaikitchen "KATA FOOD COURT"', 'Horn Grill Steak and Seafood',
                'Shakers', 'Peony Cafe & Restaurant', 'Two Chefs Kata Beach',
                'Kwong Shop Seafood', 'Leonardo Davinci',
                'The Boathouse Restaurant', 'Madras Cafe',
                'Yorkshire Hotel Restaurant',
                'On The Rocks - Marina Phuket Resort',
                'Siam Smile Wine & Restaurant', 'Bella Vista Restaurant',
                '+39 Italian Street Food', 'Southern Fried Rice',
                'Catch Beach Club', 'The Bistro', 'Veranda',
                'Sawasdee Thai Cuisine', 'Rugantino',
                'Wine Connection Bar & Grill - Kata Beach', 'Red Snapper',
                'Ska Bar', "Pim's Place", 'The Ship Inn', 'Sugar Cane Restaurant',
                'Baan Chom View', 'Kata B-B-Q', 'Lobster & Prawn Restaurant',
                'Tiger Bar', 'Restaurant Mama Kata (Seafood)', 'After Beach Bar',
                'No. 24 Bar & Restaurant', 'Buffalo Steak House - Kata Plaza',
                'Chekhoff Restaurant and Bar', 'Brasserie Phuket',
                'Sorrento pizzeria', 'Cairo Restaurant', 'Outdoor Restaurant',
                "Natalie's Restaurant", 'Coconut-garden', 'Dada Yura Restaurant',
                'Mali Seafood Restaurant & Bar'], dtype=object)
```

30 random hotels/restaurants were selected from this list using the NumPy ‘random.choice’ function the random function was also seeded so that the results are reproducible. ‘replace=False’ was specified so that there will be no repetition in the random selection. The codes & results are shown below:

```
In [64]: np.random.seed(42)
         select30 = np.random.choice(beach_hotels_rest['Hotel/Restaurant name'].unique(),30,replace=False)
         select30
```

```
Out[64]: array(['Sugar & Spice Restaurant at Dome Resort',
                'Odysseus Greek Organic Restaurant', 'Red Snapper',
                'Kataturk Turkish Restaurant', 'Kata B-B-Q',
                'Red Chair Restaurant', 'Burger House Kata Beach',
                'Kwong Shop Seafood', 'Sawasdee Thai Cuisine',
                'Autogrill Risto Bar Pizza', 'Peony Cafe & Restaurant',
                'En Vogue Restaurant', 'Buffalo Steak House - Kata Plaza',
                'Red Duck Restaurant', 'Sugar Cane Restaurant', 'Shakers',
                'After Beach Bar', 'The Boathouse Restaurant',
                'Restaurant Mama Kata (Seafood)', 'Southern Fried Rice',
                'Eightfold Restaurant', 'Leonardo Davinci', 'The Bistro',
                '44 Thaikitchen "KATA FOOD COURT"', 'The Family Restaurant',
                'Baan Chom View', 'Ska Bar', 'Veranda', 'Rugantino',
                'MaMa Jin Restaurant'], dtype=object)
```

The first (initial) dataset was then filtered against these 30 hotels to get all the rows that included these 30 hotels/restaurants. This was done using the 'isin' function. Next, the 'hotels/restaurant name' and the 'reviews' columns were selected while the rest dropped. These are the columns needed in this analysis. 'hotels_rest' is the variable name for the final dataset table achieved. See below:

```
In [80]: hotels_rest_30 = dataset[dataset['Hotel/Restaurant name'].isin(select30)]
         hotels_rest=hotels_rest_30.iloc[:,[3,4]].reset_index(drop=True)
         hotels_rest
```

Out[80]:

	Hotel/Restaurant name	Review
0	Odysseus Greek Organic Restaurant	Food was tasty and fresh. Fast service. The ow...
1	Odysseus Greek Organic Restaurant	Great variety of Greek dishes and fantastic se...
2	Odysseus Greek Organic Restaurant	We had an excellent culinary experience at thi...
3	Odysseus Greek Organic Restaurant	Amazing service and food! Highly recommend if ...
4	Odysseus Greek Organic Restaurant	This restaurant is tucked away near the square...
...	...	...
3068	Buffalo Steak House - Kata Plaza	not amazing bit expensive for what it is, woul...
3069	Buffalo Steak House - Kata Plaza	I went here on a week night. Since they took A...
3070	Buffalo Steak House - Kata Plaza	We came here by mistake!\r\n\nWe read the mixed ...
3071	Buffalo Steak House - Kata Plaza	Pros:\r\nAtmosphere is good, free salad, manag...
3072	Buffalo Steak House - Kata Plaza	I normally dont post reviews unless I have a b...

3073 rows × 2 columns

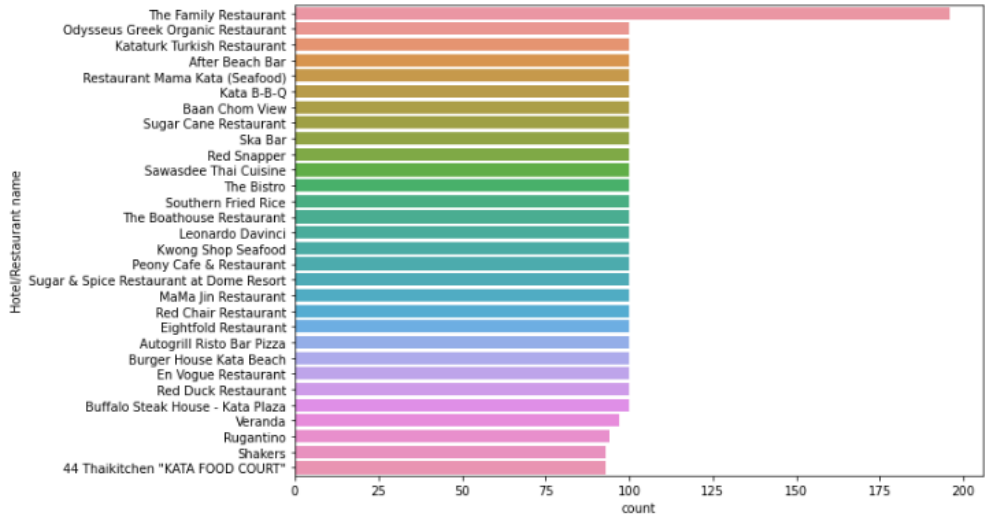
A count was taken to know the number of reviews for each of the 30 selected hotels.

```
In [91]: hotels_rest['Hotel/Restaurant name'].value_counts()
```

```
Out[91]: The Family Restaurant                        196
         Odysseus Greek Organic Restaurant            100
         Kataturk Turkish Restaurant                  100
         After Beach Bar                              100
         Restaurant Mama Kata (Seafood)               100
         Kata B-B-Q                                   100
         Baan Chom View                               100
         Sugar Cane Restaurant                        100
         Ska Bar                                      100
         Red Snapper                                  100
         Sawasdee Thai Cuisine                        100
         The Bistro                                   100
         Southern Fried Rice                          100
         The Boathouse Restaurant                     100
         Leonardo Davinci                             100
         Kwong Shop Seafood                           100
         Peony Cafe & Restaurant                      100
         Sugar & Spice Restaurant at Dome Resort      100
         MaMa Jin Restaurant                          100
         Red Chair Restaurant                         100
         Eightfold Restaurant                         100
         Autogrill Risto Bar Pizza                    100
         Burger House Kata Beach                      100
         En Vogue Restaurant                          100
         Red Duck Restaurant                          100
         Buffalo Steak House - Kata Plaza             100
         Veranda                                       97
         Rugantino                                     94
         Shakers                                       93
         44 Thaikitchen "KATA FOOD COURT"              93
         Name: Hotel/Restaurant name, dtype: int64
```

This count above was plotted using a horizontally aligned bar chart:

```
plt.figure(figsize=(10,7))
sns.countplot(y='Hotel/Restaurant name', data=hotels_rest, order=hotels_rest['Hotel/Restaurant name'].value_counts().index)
plt.show()
```



Data Cleansing

A quick look using the '.head' function shows that there are some characters that the 'cp850' encoder used in this analysis couldn't decrypt and were hence unreadable by Python.

These anomalies (underlined in red below) were eliminated using Regular Expression (re), A function of this package is the matching of strings.

```
hotels_rest.head(50)
```

20	Odysseus Greek Organic Restaurant	I regret I wait till my last night in Kala to ...
21	Odysseus Greek Organic Restaurant	This restaurant is amazing. Dimitri the owner ...
22	Odysseus Greek Organic Restaurant	Top top top!! Was looking for a Thai restauran...
23	Odysseus Greek Organic Restaurant	We were welcomed right away at this restaurant...
24	Odysseus Greek Organic Restaurant	The service and owner were lovely and welcomin...
25	Odysseus Greek Organic Restaurant	The food is excellent. Lovely place with Greek...
26	Odysseus Greek Organic Restaurant	Everything was perfect! The food was very deli...
27	Odysseus Greek Organic Restaurant	Authentic Greek cuisine. Excellent food. We ha...
28	Odysseus Greek Organic Restaurant	ItÆs a shame this restaurant is not on the Mai...
29	Odysseus Greek Organic Restaurant	Greek cuisine! The best cuisine!! Very good re...
30	Odysseus Greek Organic Restaurant	Had high expectations with all the good review...
31	Odysseus Greek Organic Restaurant	Absolutely delicious- had the hommus and pita ...
32	Odysseus Greek Organic Restaurant	Family wanted something a little different for...

Lemmatizing and Stop-Words

This is the process of reducing a word back to its root origin. For example, 'jogging', 'jogger', and 'jogged', when lemmatized, are all reduced to 'jog'. This is done to be able to group words so they can be analysed together.

An alternative to this is 'stemming', which is a method that reduces words by chopping them off to a certain point to get one uniform word. Consider words like 'continue', 'continuation', when stemmed, they may be grouped under the word 'continu'. (Balakrishnan & Lloyd-Yemoh, 2014)

Stop-words are words that are too commonly used and might not have a lot of influence on the sentiment. They are the words that the program will drop from the text.

```
def lemmatizer(text):
    return [word.lemma_ for word in english(text)]

def data_preprocess(text):
    text = str(text)
    text = re.findall('[a-zA-Z]+', text) #Extract alphabets and filter out numbers/symbols
    text = ' '.join([x for x in text if len(x)>1]) #Join elements in text and separate by string
    text = text.lower()
    text = ' '.join([word for word in text.split() if word not in STOP_WORDS])
    lemmatized = lemmatizer(text)
    cleaned = ' '.join(lemmatized)
    return cleaned
```

The image above shows a function being defined from a package called Spacy, to lemmatize the review texts. The image above also shows another function, which was defined to carry out all the data-cleansing preprocessing steps, which can be outlined sequentially as follows:

1. Converts all the characters in each review text to string
2. Uses regular expressions to select only the alphabet as this is enough to give a clear sentiment
3. Filters off strings with characters less than 1, and then join the remaining strings with space
4. Converts all text characters to lowercase.
5. All the words in the text that are in the stop-words are dropped
6. Then the text is lemmatized and finally joined (lemmatizing splits the words by space).
7. Finally, the cleaned and processed text is returned.

The figure below shows the cleaned text alongside the original text

```
hotels_rest.loc[:,'Review_processed'] = hotels_rest.loc[:,'Review'].apply(data_preprocess)
hotels_rest
```

	Hotel/Restaurant name	Review	Review_processed
0	Odysseus Greek Organic Restaurant	Food was tasty and fresh. Fast service. The ow...	food tasty fresh fast service owner super frie...
1	Odysseus Greek Organic Restaurant	Great variety of Greek dishes and fantastic se...	great variety greek dish fantastic service sta...
2	Odysseus Greek Organic Restaurant	We had an excellent culinary experience at thi...	excellent culinary experience new restaurant p...
3	Odysseus Greek Organic Restaurant	Amazing service and food! Highly recommend if ...	amazing service food highly recommend look gre...
4	Odysseus Greek Organic Restaurant	This restaurant is tucked away near the square...	restaurant tuck away near square go trip advis...
...	...	...	...
3068	Buffalo Steak House - Kata Plaza	not amazing bit expensive for what it is, woul...	amazing bit expensive wouldn recommend service...
3069	Buffalo Steak House - Kata Plaza	I went here on a week night. Since they took A...	go week night take american express currency e...
3070	Buffalo Steak House - Kata Plaza	We came here by mistake!\r\nWe read the mixed ...	come mistake read mixed review want avoid plac...
3071	Buffalo Steak House - Kata Plaza	Pros:\r\nAtmosphere is good, free salad, manag...	pro atmosphere good free salad manager willing...
3072	Buffalo Steak House - Kata Plaza	I normally dont post reviews unless I have a b...	normally do not post review average experience...

To see the effect clearly, 2 examples were cleaned and the results are shown below:

```
for example in [100, 200]:
    print ('Original\n****************')
    print(hotels_rest.Review.iloc[example])
    print('_________________\nCleaned\n****************')
    print(hotels_rest.Review_processed.iloc[example])
    print('')
```

```
Original
****************
In our 10 day honeymoon while we stayed in the Vijit Resort at Hawa´ we have visited this restaurant atleast 5 times, that's ho
w great it is.
We come from the Netherlands and have problems with the food hygiÙne in Thailand. This is the only...More

_________________
Cleaned
****************
day honeymoon stay vijit resort hawa visit restaurant atleast time great come netherlands problem food hygi ne thailand

Original
****************
I have been eating here for many years and the food is always top quality. I prefer the Israeli food, but the Thai is also grea
t. Friendly staff. Highly recommended.

_________________
Cleaned
****************
eat year food quality prefer israeli food thai great friendly staff highly recommend
```

Implementation in Python

Brief Description of Sentiment Analysis

Sentiment analysis as a machine learning method is a type of natural language processing that is used to analyse how users feel about certain a product, service, or topic.

With the advent of Social Media like Facebook, Twitter, Reddit, etc. the importance of sentiment analysis cannot be overstated. For example, companies can gauge the level of acceptance of their product in the market, political candidates can assess how the public feels about them, Service providers can tell where to improve, and the list goes on.

Sentiment analysis with all its 'pros' also has its cons, as natural language is complex. A word can have different sentiments depending on the context that the algorithm might not pick up on, people can also use sarcasm and irony that the algorithm might process literally. (Vinodhini & Chandrasekaran, 2012)

Application & Explanation of the algorithm

Count Vectorizer/Bag-of-words

The first step taken in this implementation is getting the Bag-of-words. This is simply text converted into vectors, which are numbers (in form of a sparse matrix) that can be processed by machine learning models. Which is then arranged in a DataFrame table format with all the words as columns and the frequency of the appearance of each word recorded under the respective columns for each different review. See the output below:

```
words_df = pd.DataFrame(data_cv.toarray(), columns=cv.get_feature_names_out())
words_df.head()
```

	aaa	aaannndd	aback	abc	abd	abf	able	abound	abroad	abrupt	...	yum	yummie	yummmmmy	yummy	zero	zinger	zlatan	zoom	zucchini	zufried
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	

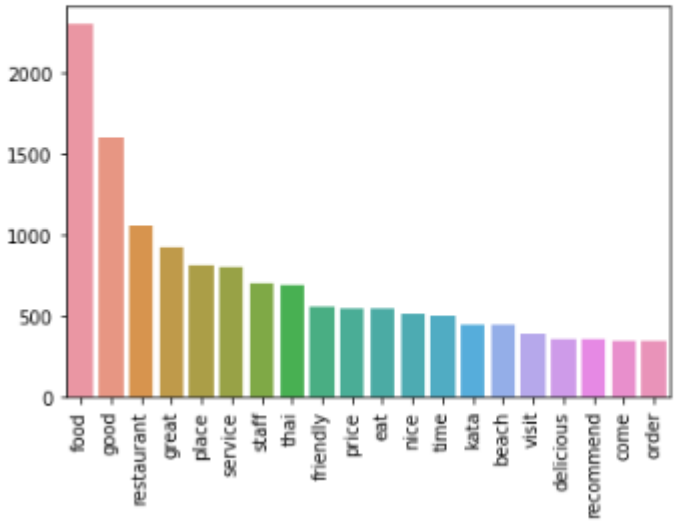
5 rows × 4518 columns

Word distribution and Word-Cloud

From the bag-of words derived, a barplot can be plotted alongside a word-cloud for better visuals.

The figure below shows the top 20 most frequent words:

```
#To view the top 20 most frequent words in the reviews
freq_20 = words_df.sum(axis=0).sort_values(ascending=False)[:20]
sns.barplot(x=freq_20.index, y=freq_20.values)
plt.xticks(rotation=90)
plt.show()
```



See the code for the word-cloud and the result visualised below:

```
fulltext=' '.join(word for word in hotels_rest.Review_processed)
```

```
def cloud_plot(word_cloud):
    plt.figure(figsize=(17,10))
    plt.imshow(word_cloud, interpolation='bilinear')
    plt.axis('off')
    plt.show()
```

```
cloud_com = np.array(Image.open("cloud1.png"))
wc=WordCloud(background_color='white',
             colormap='Dark2',
             random_state=42,
             collocations=True,
             mask=cloud_com).generate(fulltext)
cloud_plot(wc)
```

Sentiment Classification

With the use of the TextBlob package and function, the polarity and subjectivity of the reviews were derived. The TextBlob function gives its polarity results as numbers ranging from -1 to +1, while its subjectivity result is from 0 to 1.

Subjectivity measures the degree to which the review is personalized and gives an opinion or point of view.

Polarity is the degree to which a comment is positive, negative, or neutral. (Satapathy, Pardesh, & Cambria, 2022)

For this report, this classification was set within the ranges shown below

- Positive: polarity >= 0.05
- Negative: polarity <= -0.05
- Neutral: -0.05 < polarity < 0.05

See the codes and the result as shown below:

```
polarity = lambda review: TextBlob(review).sentiment.polarity
subjectivity = lambda review: TextBlob(review).sentiment.subjectivity
```

```
hotels_rest.loc[:,'Polarity'] = hotels_rest.loc[:, 'Review_processed'].apply(polarity)
hotels_rest.loc[:,'Subjectivity'] = hotels_rest.loc[:, 'Review_processed'].apply(subjectivity)
```

```
hotels_rest['Sentiment'] = 'Neutral'
hotels_rest['Sentiment'] = np.where(hotels_rest['Polarity'] >= 0.05, 'positive', hotels_rest['Sentiment'])
hotels_rest['Sentiment'] = np.where(hotels_rest['Polarity'] <= -0.05, 'negative', hotels_rest['Sentiment'])
```

```
hotels_rest
```

	Hotel/Restaurant name	Review	Review_processed	Polarity	Subjectivity	Sentiment
0	Odysseus Greek Organic Restaurant	Food was tasty and fresh. Fast service. The ow...	food tasty fresh fast service owner super frie...	0.361389	0.592778	positive
1	Odysseus Greek Organic Restaurant	Great variety of Greek dishes and fantastic se...	great variety greek dish fantastic service sta...	0.429167	0.583333	positive
2	Odysseus Greek Organic Restaurant	We had an excellent culinary experience at thi...	excellent culinary experience new restaurant p...	0.427189	0.491246	positive
3	Odysseus Greek Organic Restaurant	Amazing service and food! Highly recommend if ...	amazing service food highly recommend look gre...	0.315000	0.510000	positive
4	Odysseus Greek Organic Restaurant	This restaurant is tucked away near the square...	restaurant tuck away near square go trip advis...	0.450000	0.525000	positive
...	...	...	...	...	...	...
3068	Buffalo Steak House - Kata Plaza	not amazing bit expensive for what it is, woul...	amazing bit expensive wouldn recommend service...	-0.091667	0.541667	negative
3069	Buffalo Steak House - Kata Plaza	I went here on a week night. Since they took A...	go week night take american express currency e...	0.350000	0.300000	positive
3070	Buffalo Steak House - Kata Plaza	We came here by mistake!\r\nWe read the mixed ...	come mistake read mixed review want avoid plac...	-0.250000	0.500000	negative
3071	Buffalo Steak House - Kata Plaza	Pros:\r\nAtmosphere is good, free salad, manag...	pro atmosphere good free salad manager willing...	-0.037500	0.727083	Neutral
3072	Buffalo Steak House - Kata Plaza	I normally dont post reviews unless I have a b...	normally do not post review average experience...	0.043750	0.512500	Neutral

Visualisation of the results

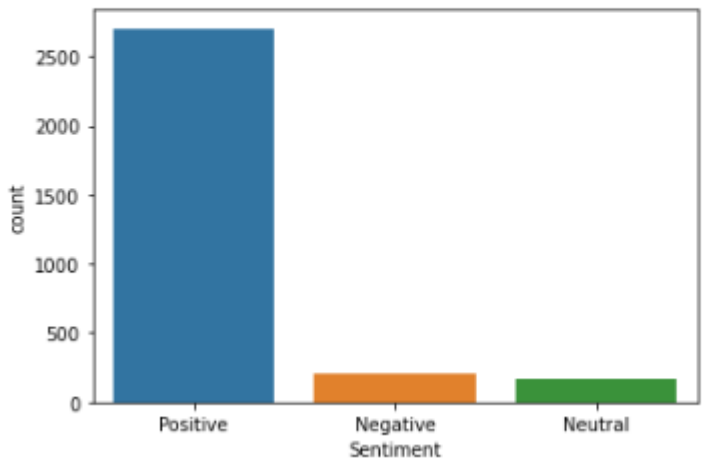
First, the number of positive, neutral, and negative reviews is shown numerically and plotted visually below:

```
hotels_rest.Sentiment.value_counts()
```

```
Positive    2702
Negative     206
Neutral      165
Name: Sentiment, dtype: int64
```

```
sent_viz = sns.countplot(data=hotels_rest, x=hotels_rest.Sentiment)
sent_viz
```

```
<AxesSubplot:xlabel='Sentiment', ylabel='count'>
```



Then, the code to get the number of reviews for each hotel/restaurant, and the review count for the hotels are shown below:

```
new_list = []
def hotel_sentiment(df):
    hotels_list = df['Hotel/Restaurant name'].unique()
    for hotels in hotels_list:
        new_df = df[df['Hotel/Restaurant name'] == hotels]
        counts = new_df['Sentiment'].value_counts()
        new_list.append(counts)
        print(hotels)
        print(counts)
        print('')
hotel_sentiment(hotels_rest)
```

```
Odysseus Greek Organic Restaurant
Positive    93
Negative     4
Neutral      3
Name: Sentiment, dtype: int64

Kataturk Turkish Restaurant
Positive    88
Neutral      8
Negative     4
Name: Sentiment, dtype: int64

The Family Restaurant
Positive    178
Negative      9
Neutral       9
Name: Sentiment, dtype: int64

Red Duck Restaurant
Positive    94
Negative     5
Neutral      1
Name: Sentiment, dtype: int64

En Vogue Restaurant
Positive    95
Negative     3
Neutral      2
Name: Sentiment, dtype: int64

Burger House Kata Beach
Positive    91
Negative     5
Neutral      4
Name: Sentiment, dtype: int64

Autogrill Risto Bar Pizza
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

Eightfold Restaurant
Positive    91
Neutral      5
Negative     4
Name: Sentiment, dtype: int64

Red Chair Restaurant
Positive    87
Neutral     10
Negative     3
Name: Sentiment, dtype: int64

MaMa Jin Restaurant
Positive    91
Neutral      7
Negative     2
Name: Sentiment, dtype: int64

Sugar & Spice Restaurant at Dome Resort
Positive    93
Negative     5
Neutral      2
Name: Sentiment, dtype: int64

44 Thaikitchen "KATA FOOD COURT"
Positive    88
Negative     3
Neutral      2
Name: Sentiment, dtype: int64

Shakers
Positive    85
Neutral      4
Negative     4
Name: Sentiment, dtype: int64

Peony Cafe & Restaurant
Positive    90
Negative     8
Neutral      2
Name: Sentiment, dtype: int64

Kwong Shop Seafood
Positive    87
Negative     8
Neutral      5
Name: Sentiment, dtype: int64

Leonardo Davinci
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

The Boathouse Restaurant
Positive    95
Neutral      3
Negative     2
Name: Sentiment, dtype: int64

Southern Fried Rice
Positive    92
Neutral      4
Negative     4
Name: Sentiment, dtype: int64
```

```
The Bistro
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

Veranda
Positive    67
Negative    21
Neutral      9
Name: Sentiment, dtype: int64

Sawasdee Thai Cuisine
Positive    94
Negative     3
Neutral      3
Name: Sentiment, dtype: int64

Rugantino
Positive    77
Negative     9
Neutral      8
Name: Sentiment, dtype: int64

Red Snapper
Positive    78
Negative    13
Neutral      9
Name: Sentiment, dtype: int64

Ska Bar
Positive    87
Neutral      8
Negative     5
Name: Sentiment, dtype: int64

Sugar Cane Restaurant
Positive    94
Negative     3
Neutral      3
Name: Sentiment, dtype: int64

Baan Chom View
Positive    83
Neutral      9
Negative     8
Name: Sentiment, dtype: int64

Kata B-B-Q
Positive    79
Negative    16
Neutral      5
Name: Sentiment, dtype: int64

Restaurant Mama Kata (Seafood)
Positive    80
Negative    14
Neutral      6
Name: Sentiment, dtype: int64

After Beach Bar
Positive    86
Negative    11
Neutral      3
Name: Sentiment, dtype: int64

Buffalo Steak House - Kata Plaza
Positive    69
Neutral     19
Negative    12
Name: Sentiment, dtype: int64
```

Code & result to get the tabular form:

```python
def sentiments(df):
    list_of_dicts = []
    hotels = df['Hotel/Restaurant name'].unique()
    for hotel in hotels:
        new_df = df[df['Hotel/Restaurant name'] == hotel]
        counts = new_df['Sentiment'].value_counts(normalize=True).round(2)

        try:
            positive=counts['Positive']*100
        except KeyError:
            positive=0

        try:
            negative=counts['Negative']*100
        except KeyError:
            negative=0

        try:
            neutral=counts['Neutral']*100
        except KeyError:
            neutral=0

        df_1={
            'Hotel/Restaurant name': hotel,
            'Positive(%)': positive,
            'Neutral(%)': neutral,
            'Negative(%)': negative
        }
        list_of_dicts.append(df_1)

    df_func = pd.DataFrame(list_of_dicts)
    return df_func
```

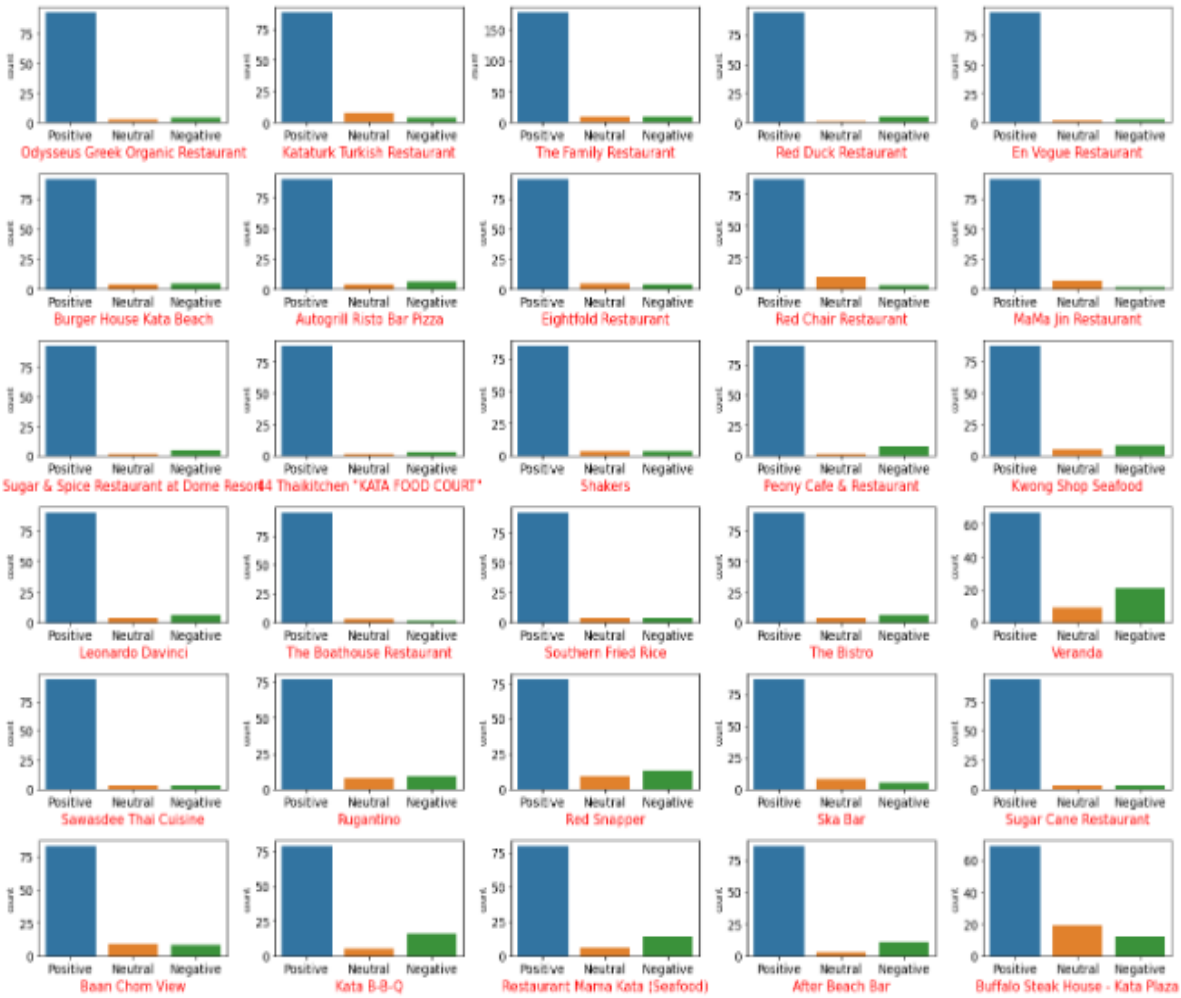
```python
sentiment_df = sentiments(hotels_rest)
sentiment_df
```

	Hotel/Restaurant name	Positive(%)	Neutral(%)	Negative(%)
0	Odysseus Greek Organic Restaurant	93.0	3.0	4.0
1	Kataturk Turkish Restaurant	88.0	8.0	4.0
2	The Family Restaurant	91.0	5.0	5.0
3	Red Duck Restaurant	94.0	1.0	5.0
4	En Vogue Restaurant	95.0	2.0	3.0
5	Burger House Kata Beach	91.0	4.0	5.0
6	Autogrill Risto Bar Pizza	90.0	4.0	6.0
7	Eightfold Restaurant	91.0	5.0	4.0
8	Red Chair Restaurant	87.0	10.0	3.0
9	MaMa Jin Restaurant	91.0	7.0	2.0
10	Sugar & Spice Restaurant at Dome Resort	93.0	2.0	5.0
11	44 Thaikitchen "KATA FOOD COURT"	95.0	2.0	3.0
12	Shakers	91.0	4.0	4.0
13	Peony Cafe & Restaurant	90.0	2.0	8.0
14	Kwong Shop Seafood	87.0	5.0	8.0
15	Leonardo Davinci	90.0	4.0	6.0
16	The Boathouse Restaurant	95.0	3.0	2.0
17	Southern Fried Rice	92.0	4.0	4.0
18	The Bistro	90.0	4.0	6.0
19	Veranda	69.0	9.0	22.0
20	Sawasdee Thai Cuisine	94.0	3.0	3.0
21	Rugantino	82.0	9.0	10.0
22	Red Snapper	78.0	9.0	13.0
23	Ska Bar	87.0	8.0	5.0
24	Sugar Cane Restaurant	94.0	3.0	3.0
25	Baan Chom View	83.0	9.0	8.0
26	Kata B-B-Q	79.0	5.0	16.0
27	Restaurant Mama Kata (Seafood)	80.0	6.0	14.0
28	After Beach Bar	86.0	3.0	11.0
29	Buffalo Steak House - Kata Plaza	69.0	19.0	12.0

Finally, a plot showing the distribution of reviews in all 30 hotels is shown:

```
hotels_list = hotels_rest['Hotel/Restaurant name'].unique().tolist()
series_list = [pd.Series(new_list[i], name=hotels_list[i]) for i in range(30)]
final_rest = pd.concat([x for x in series_list], axis=1)
final_rest
```

```
plt.figure(figsize=(20,15))
order = ['Positive', 'Neutral', 'Negative']
for i in range(30):
    plt.subplot(6,5,i+1)
    plt.tight_layout()
    plt.tick_params(axis='both', which='major', labelsize=14)
    plt.xlabel('xlabel', fontsize=14, color='red')
    sns.countplot(data=final_rest, x=final_rest.columns.tolist()[i], order = order)
```



Relevant Literature

(Cambria, Schuller, Xia, & Havasi, 2013) in their journal article share the general opinion that sentiment analysis has evolved through the years and can now be regarded as a different branch of natural language processing (NLP). Even though it involves some aspects of NLP, it differs in that it does not require a deep understanding of the text as it focuses mainly on semantics and word inference.

It involves the collection of opinions and reviews online and does well to rank and filter off unopinionated reviews.

This AI method can be slow as a result of operations that involves summarization and auto-categorization but it is still very effective in its simple yet operative functionality.

This experiment aligns perfectly with this school of thought and effectively shows how resourceful the algorithm is.

Results analysis and discussion

Performance metric used

To gauge the performance, a manual look at samples of the classified sentiments of the analysed reviews was done and the results proved satisfactory. This is explained further in the next section

Presentation of results

A look at the results of the sentiment analysis of random samples of the reviews.

- **Random positive reviews:**

```
for i in [29, 35, 103]:
    print(hotels_rest[hotels_rest.Sentiment == 'Positive']['Review'][i])
    print('')
    print('**********************************************')
```

```
Greek cuisine! The best cuisine!! Very good restaurant! The food was very nice and tasty! Recommended!

**********************************************
The staff is very friendly and the food is beyond belief, fresh every ingredient is coming from a selected area!
Recommended!

**********************************************
Ate here on my last day in Phuket and have been craving it ever since. The food was sensational and the service was just as imp
ressive. Highly recommended.

**********************************************
```

It can be seen above that these samples show positive sentiments in the reviews.

- **Random neutral reviews:**

```
for i in [111, 68, 45]:
    print(hotels_rest[hotels_rest.Sentiment == 'Neutral']['Review'][i])
    print('')
    print('**********************************************')
```

```
We lived in Turkey for 5 years and have missed so much about this lovely country, from the food, to the history, to the warmth
of the Turkish people.
We went to K. Ataturk because a friend recommended it and were NOT disappointed. The menu...More

**********************************************
Was expecting big things and left totally disappointed the food was ok the timing of the food was terrible the owner seemed hig
h on drugs the food i had was ok but the other quests in my party were far from Happy don't know where...More

**********************************************
We went there (2 couples), and I felt like going home to my favourite restaurant. Everyone very friendly and Chef Dimitiris cam
e out to give us his recommendations. Pretty much the service you would pay in Europe for a really expensive restaurant. Also h
e invited...More
```

These reviews above show no real bias, so were rightly classified as neutral by the experiment.

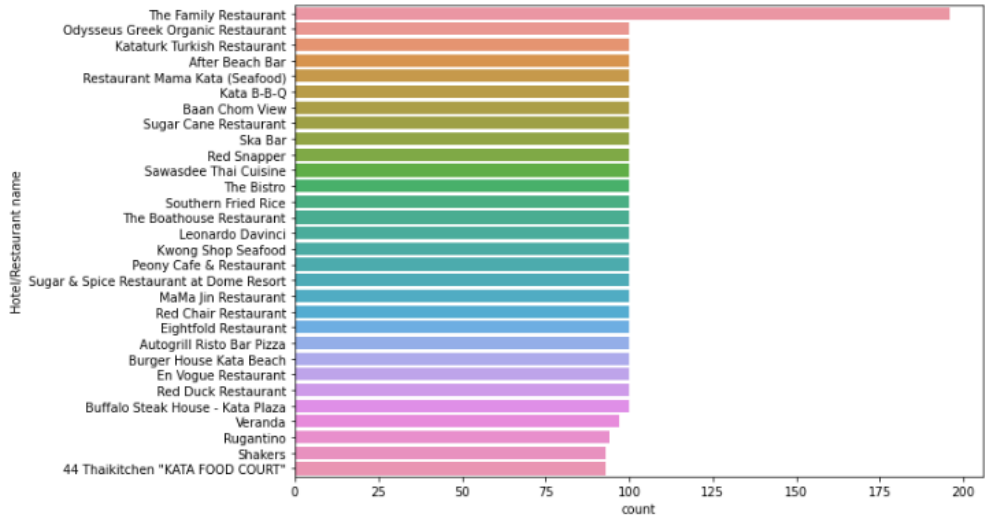
- **Random negative reviews:**

```
for i in [495, 133, 817]:
    print(hotels_rest[hotels_rest.Sentiment == 'Negative']['Review'][i])
    print('')
    print('**********************************************')
```

```
Unfortunately my review is impeded with the fact we waited so long for food. The restaurant was not busy but we were told at th
e end that they were short staffed. Each meal came out in at different times with the wait from the first...More

**********************************************
After asking specifically if the wine and beers were served ice cold we were assured but then served with room temperature drin
ks.Bad vibes immediately so after sending the beers back and once again being served with the same we decided to leave.
The owner followed...More

**********************************************
Very overpriced. Portion is extremely small, not enough to feed an adult. Red curry is best out of 5 dishes we ordered. Disappo
inted upon arrival, I guess the overwhelming reviews on TripAdvisor are based on the fact that they have burgers for people who
prefer...More
```

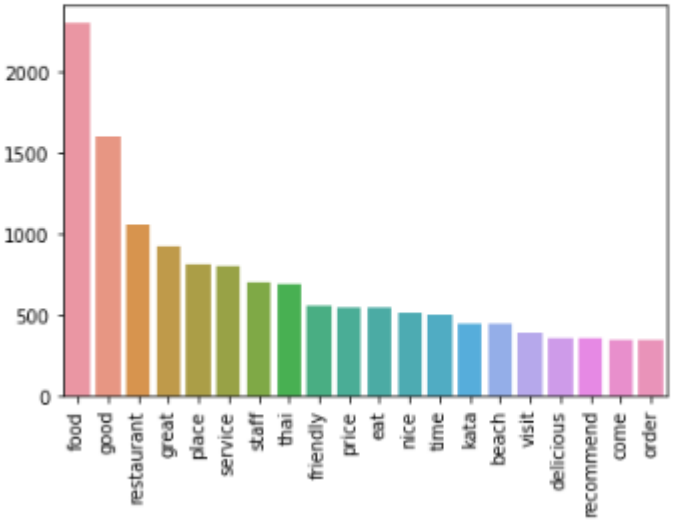
The review count distribution of the 30 hotels selected based on beach is as seen below:

```
plt.figure(figsize=(10,7))
sns.countplot(y='Hotel/Restaurant name', data=hotels_rest, order=hotels_rest['Hotel/Restaurant name'].value_counts().index)
plt.show()
```



The most frequent words in the reviews of the 30 hotels selected are plotted and visualised using a barplot and the WordCloud below:

```
#To view the top 20 most frequent words in the reviews
freq_20 = words_df.sum(axis=0).sort_values(ascending=False)[:20]
sns.barplot(x=freq_20.index, y=freq_20.values)
plt.xticks(rotation=90)
plt.show()
```



The difference between the original reviews and the reviews that underwent pre-processing (number & symbols filtering, and lemmatization) can be seen below:

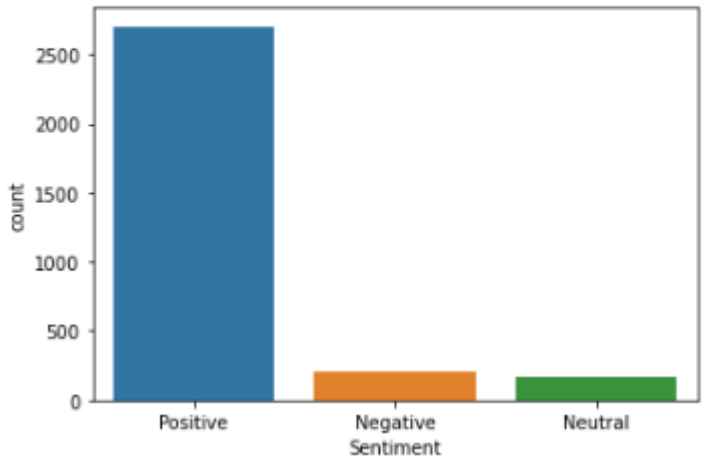
```
for example in [100, 200]:
    print ('Original\n**********')
    print(hotels_rest.Review.iloc[example])
    print('_____\nCleaned\n**********')
    print(hotels_rest.Review_processed.iloc[example])
    print('')
```

```
Original
**********
In our 10 day honeymoon while we stayed in the Vijit Resort at Hawa´ we have visited this restaurant atleast 5 times, that's ho
w great it is.
We come from the Netherlands and have problems with the food hygiÙne in Thailand. This is the only...More
_____
Cleaned
**********
day honeymoon stay vijit resort hawa visit restaurant atleast time great come netherlands problem food hygi ne thailand

Original
**********
I have been eating here for many years and the food is always top quality. I prefer the Israeli food, but the Thai is also grea
t. Friendly staff. Highly recommended.
_____
Cleaned
**********
eat year food quality prefer israeli food thai great friendly staff highly recommend
```

Distribution of the sentiment classification of all the reviews is seen below:

```
sent_viz = sns.countplot(data=hotels_rest, x=hotels_rest.Sentiment)
sent_viz
```

```
<AxesSubplot:xlabel='Sentiment', ylabel='count'>
```

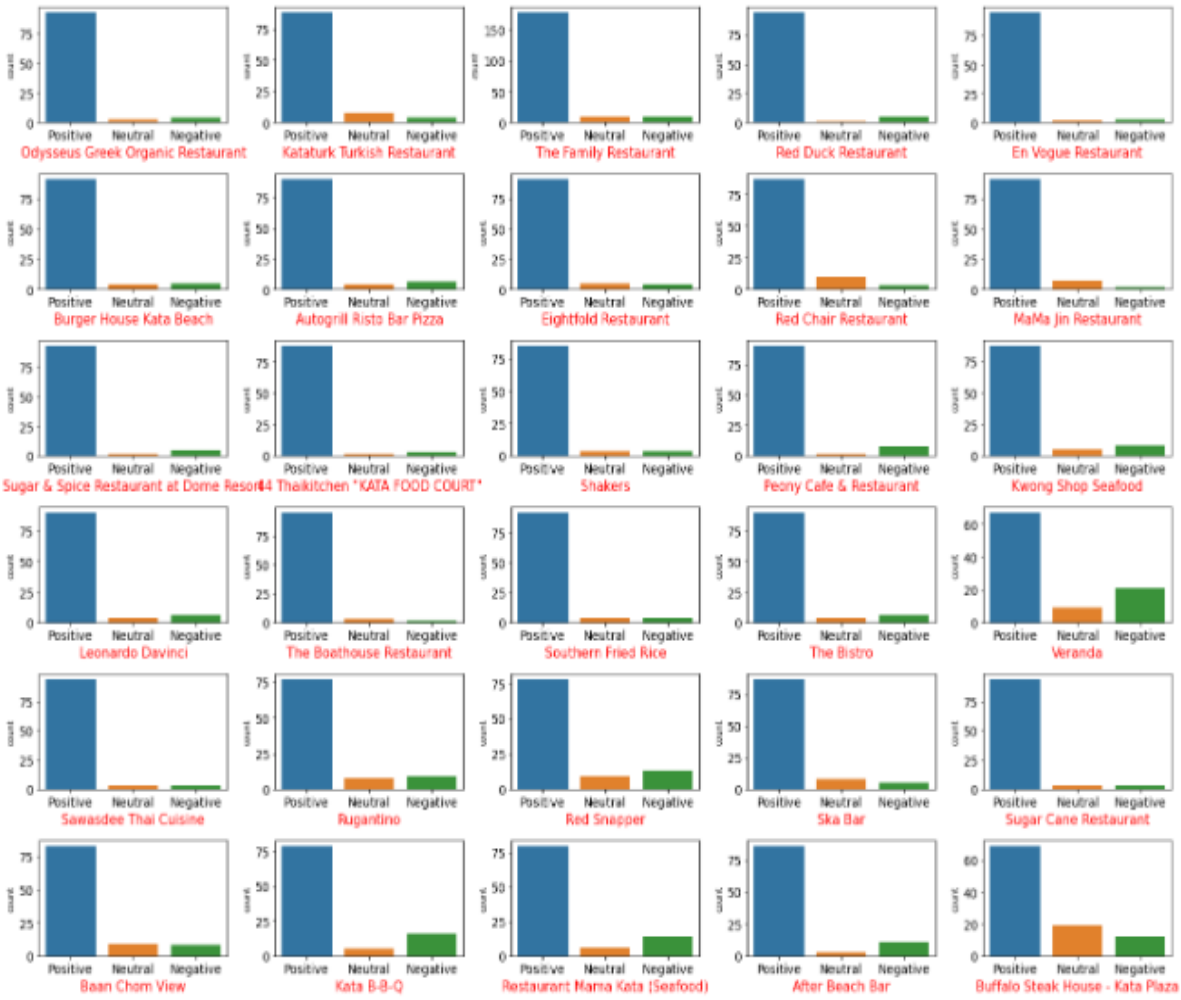


A list of the sentiment distribution for each of the 30 hotels is shown below:

```
Odysseus Greek Organic Restaurant
Positive    93
Negative     4
Neutral      3
Name: Sentiment, dtype: int64

Kataturk Turkish Restaurant
Positive    88
Neutral      8
Negative     4
Name: Sentiment, dtype: int64

The Family Restaurant
Positive    178
Negative      9
Neutral       9
Name: Sentiment, dtype: int64

Red Duck Restaurant
Positive    94
Negative     5
Neutral      1
Name: Sentiment, dtype: int64

En Vogue Restaurant
Positive    95
Negative     3
Neutral      2
Name: Sentiment, dtype: int64

Burger House Kata Beach
Positive    91
Negative     5
Neutral      4
Name: Sentiment, dtype: int64

Autogrill Risto Bar Pizza
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

Eightfold Restaurant
Positive    91
Neutral      5
Negative     4
Name: Sentiment, dtype: int64

Red Chair Restaurant
Positive    87
Neutral     10
Negative     3
Name: Sentiment, dtype: int64

MaMa Jin Restaurant
Positive    91
Neutral      7
Negative     2
Name: Sentiment, dtype: int64

Sugar & Spice Restaurant at Dome Resort
Positive    93
Negative     5
Neutral      2
Name: Sentiment, dtype: int64

44 Thaikitchen "KATA FOOD COURT"
Positive    88
Negative     3
Neutral      2
Name: Sentiment, dtype: int64

Shakers
Positive    85
Neutral      4
Negative     4
Name: Sentiment, dtype: int64

Peony Cafe & Restaurant
Positive    90
Negative     8
Neutral      2
Name: Sentiment, dtype: int64

Kwong Shop Seafood
Positive    87
Negative     8
Neutral      5
Name: Sentiment, dtype: int64

Leonardo Davinci
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

The Boathouse Restaurant
Positive    95
Neutral      3
Negative     2
Name: Sentiment, dtype: int64

Southern Fried Rice
Positive    92
Neutral      4
Negative     4
Name: Sentiment, dtype: int64
```

```
The Bistro
Positive    90
Negative     6
Neutral      4
Name: Sentiment, dtype: int64

Veranda
Positive    67
Negative    21
Neutral      9
Name: Sentiment, dtype: int64

Sawasdee Thai Cuisine
Positive    94
Negative     3
Neutral      3
Name: Sentiment, dtype: int64

Rugantino
Positive    77
Negative     9
Neutral      8
Name: Sentiment, dtype: int64

Red Snapper
Positive    78
Negative    13
Neutral      9
Name: Sentiment, dtype: int64

Ska Bar
Positive    87
Neutral      8
Negative     5
Name: Sentiment, dtype: int64

Sugar Cane Restaurant
Positive    94
Negative     3
Neutral      3
Name: Sentiment, dtype: int64

Baan Chom View
Positive    83
Neutral      9
Negative     8
Name: Sentiment, dtype: int64

Kata B-B-Q
Positive    79
Negative    16
Neutral      5
Name: Sentiment, dtype: int64

Restaurant Mama Kata (Seafood)
Positive    80
Negative    14
Neutral      6
Name: Sentiment, dtype: int64

After Beach Bar
Positive    86
Negative    11
Neutral      3
Name: Sentiment, dtype: int64

Buffalo Steak House - Kata Plaza
Positive    69
Neutral     19
Negative    12
Name: Sentiment, dtype: int64
```

A graphical representation of the sentiment distribution for each of the 30 hotels is shown below:



Discussion of result

From the results shown above, it can be seen that the majority of the reviews had positive sentiments. 87.93% of the reviews were positive, while 6.7% of the reviews were negative. 5.37% were neither positive nor negative, so they were classified as neutral.

Ethical, Legal, and professional considerations

Some considerations made in the

1. The dataset poses no risks to individuals or organisations by making sure the URL names were excluded.
2. The data involved was collected from legitimate, publicly available means.

All parties involved in the collection of the dataset were duly listed

Conculsion

Based on this report of the experiment just concluded, it can be said that the majority of the time, the hotels/restaurants in the beach locations provide services above average. This means the objective of mining the overall sentiment of the reviews has been decisively met.

In conclusion, even though AI sentiment analysis is prone to some errors as it mines information from very complex unstructured data, it remains a powerful tool that can be used by individuals or organisations to sift through large amounts of text, get the general mood, sentiment, & opinion of the public and greatly improve their products or services accordingly.