In [7]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

In [8]:
```python
df = pd.read_csv('tripadvisor_hotel_reviews.csv')
df.head()
```

Out[8]:

	Review	Rating
0	nice hotel expensive parking got good deal sta...	4
1	ok nothing special charge diamond member hilto...	2
2	nice rooms not 4* experience hotel monaco seat...	3
3	unique, great stay, wonderful time hotel monac...	5
4	great stay great stay, went seahawk game aweso...	5

In [9]:
```python
df['Rating'].value_counts()
```
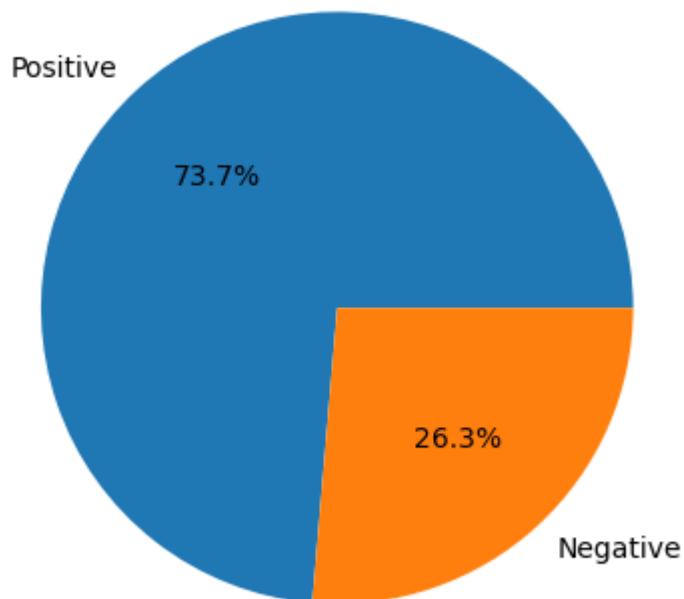
Out[9]:
```
5    9054
4    6039
3    2184
2    1793
1    1421
Name: Rating, dtype: int64
```

In [10]:
```python
def rating(rating):
    if rating>3 and rating<=5:
        return "Positive"
    if rating>0 and rating<=3:
        return "Negative"
```

In [11]:
```python
df['Rating'] = df['Rating'].apply(rating)
```

In [14]:
```python
plt.pie(df['Rating'].value_counts(),
        labels=df['Rating'].unique().tolist(),
        autopct='%1.1f%%')
plt.show()
```


In [15]:
```python
lenght = len(df['Review'][0])
print(f'Lenght of a sample review: {lenght}')
```

Lenght of a sample review: 593

In [16]:
```python
df['Lenght'] = df['Review'].str.len()
df.head()
```

Out[16]:

	Review	Rating	Lenght
0	nice hotel expensive parking got good deal sta...	Positive	593
1	ok nothing special charge diamond member hilto...	Negative	1689
2	nice rooms not 4* experience hotel monaco seat...	Negative	1427
3	unique, great stay, wonderful time hotel monac...	Positive	600
4	great stay great stay, went seahawk game aweso...	Positive	1281

In [17]:
```python
word_count = df['Review'][0].split()
print(f'Word count in a samplereview: {len(word_count)}')
```

Word count in a samplereview: 87

In [18]:
```python
def word_count(review):
    review_list = review.split()
    return len(review_list)
```

In [19]:
```python
df['Word_count'] = df['Review'].apply(word_count)
df.head()
```

Out[19]:

	Review	Rating	Lenght	Word_count
0	nice hotel expensive parking got good deal sta...	Positive	593	87
1	ok nothing special charge diamond member hilto...	Negative	1689	250
2	nice rooms not 4* experience hotel monaco seat...	Negative	1427	217
3	unique, great stay, wonderful time hotel monac...	Positive	600	89
4	great stay great stay, went seahawk game aweso...	Positive	1281	191

In [ ]:
```python
df['mean_word_lenght'] = df['Review'].map(lambda rev: np.mean([len(word)
                                                        for word in r
df.head()
```

In [ ]: