```python
In [1]:   import pandas as pd
          import matplotlib.pyplot as plt
          import numpy as np
```

```python
In [5]:   df = pd.read_csv(r"C:\Users\Vaish\Downloads\datasets-main\datasets-main\tripadvisor
          df.head()
```

Out[5]:

	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

```python
In [9]:   df['Rating'].value_counts()
```

```
Out[9]:   Rating
          5    9054
          4    6039
          3    2184
          2    1793
          1    1421
          Name: count, dtype: int64
```

```python
In [25]:  # rating 4, 5 => Positive; 1, 2, 3 => Negative

          def ratings(rating):

              if rating>3 and rating<=5:

                  return "Positive"

              if rating>0 and rating<=3:

                  return "Negative"
```

```python
In [27]:  df['Rating'] = df['Rating'].apply(ratings)
```
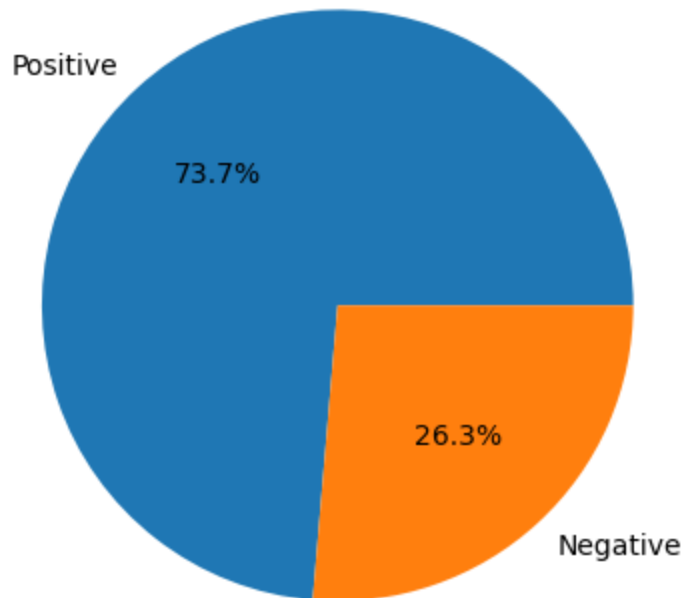
```python
In [29]:  plt.pie(df['Rating'].value_counts(),

                  labels=df['Rating'].unique().tolist(),

                  autopct='%1.1f%%')

          plt.show()
```

Positive 73.7%
Negative 26.3%

```
In [39]: lenght = len(df['Review'][0])

         print(f'Length of a sample review: {lenght}')
```

Length of a sample review: 593

```
In [33]: df['Length'] = df['Review'].str.len()

         df.head()
```

Out[33]:

	Review	Rating	Length
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 [41]: word_count = df['Review'][0].split()

         print(f'Word count in a sample review: {len(word_count)}')
```

Word count in a sample review: 87

```
In [49]: def word_count(review):

             review_list = review.split()
```

```
    return len(review_list)
```

In [47]:
```python
df['Word_count'] = df['Review'].apply(word_count)

df.head()
```

Out[47]:

	Review	Rating	Length	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 [51]:
```python
df['mean_word_length'] = df['Review'].map(lambda rev: np.mean([len(word) for word i

df.head()
```

Out[51]:

	Review	Rating	Length	Word_count	mean_word_length
0	nice hotel expensive parking got good deal sta...	Positive	593	87	5.804598
1	ok nothing special charge diamond member hilto...	Negative	1689	250	5.752000
2	nice rooms not 4* experience hotel monaco seat...	Negative	1427	217	5.571429
3	unique, great stay, wonderful time hotel monac...	Positive	600	89	5.730337
4	great stay great stay, went seahawk game aweso...	Positive	1281	191	5.701571

In [53]:
```python
from nltk import tokenize

np.mean([len(sent) for sent in tokenize.sent_tokenize(df['Review'][0])])
```

Out[53]: 591.0

In [55]:
```python
df['mean_sent_length'] = df['Review'].map(lambda rev: np.mean([len(sent)

                            for sent in tokenize.sent_tokenize(rev)]))

df.head()
```

Out[55]:

	Review	Rating	Length	Word_count	mean_word_length	mean_sent_length
0	nice hotel expensive parking got good deal sta...	Positive	593	87	5.804598	591.0
1	ok nothing special charge diamond member hilto...	Negative	1689	250	5.752000	1687.0
2	nice rooms not 4* experience hotel monaco seat...	Negative	1427	217	5.571429	712.0
3	unique, great stay, wonderful time hotel monac...	Positive	600	89	5.730337	598.0
4	great stay great stay, went seahawk game aweso...	Positive	1281	191	5.701571	1279.0

In [ ]:

In [ ]:

In [ ]:

In [ ]: