

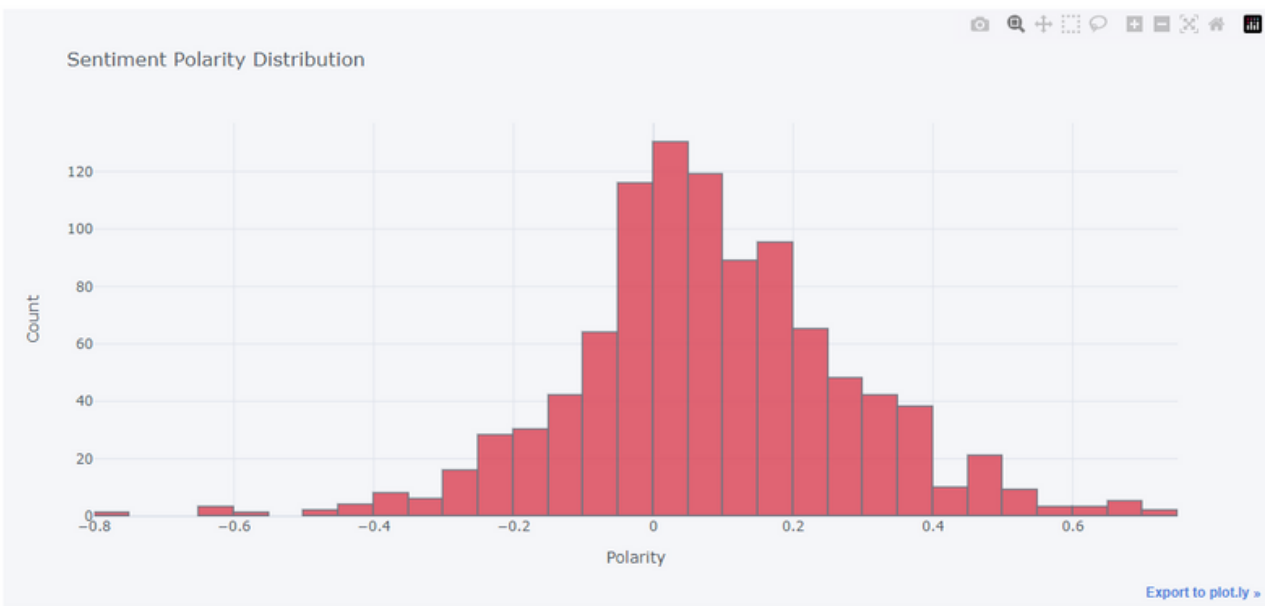
Skytrax

Airline Review

Data Analysis

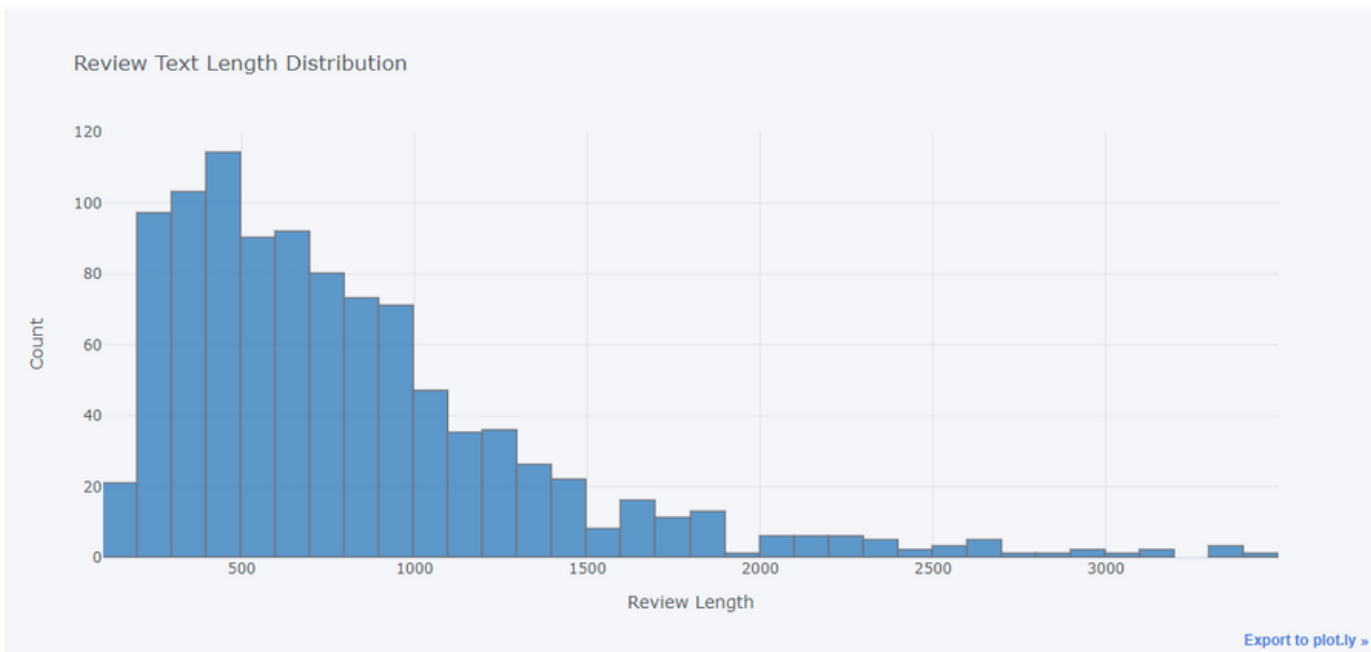
Praneesh Sharma

```
In [21]: df['polarity'].iplot(kind = 'hist', colors = 'red',
                        xTitle = 'Polarity', yTitle='Count',
                        title = 'Sentiment Polarity Distribution')
```

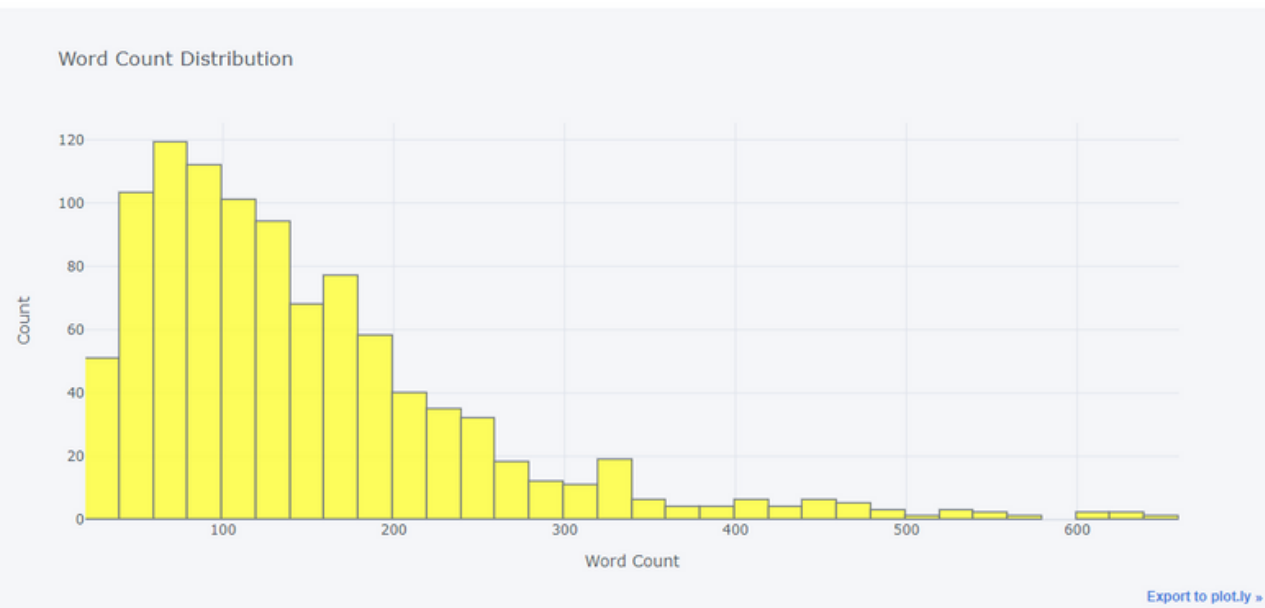


The polarity of major reviews is nearly 0, so the reviews are mostly neutral.

```
df['review_len'].iplot(kind = 'hist', colors = 'blue',
                      xTitle = 'Review Length', yTitle = 'Count',
                      title = 'Review Text Length Distribution')
```



```
In [23]: df['word_count'].iplot(kind = 'hist', colors='yellow',
                        xTitle = 'Word Count', yTitle = 'Count',
                        title = 'Word Count Distribution')
```



Most of the reviews have an average word count of 60-70 words.

```
28]: df1 = pd.DataFrame(words, columns = ['Unigram', 'Frequency'])
df1 = df1.set_index('Unigram')
df1.iplot(kind='bar', xTitle='Unigram', yTitle='Count', title='Top 20 words')
```

