In [3]:

```
import numpy as np
import pandas as pd
import re
import nltk
import spacy
import string
import matplotlib.pyplot as plt
```

In [7]:

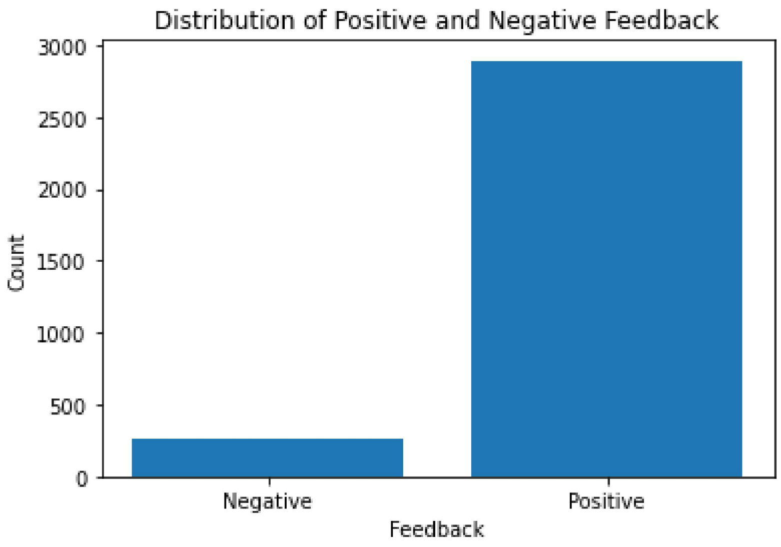
```
alexa=pd.read_csv('Alexa-Dataset.csv')
alexa.head()
```

Out[7]:

	rating	date	variation	verified_reviews	feedback
0	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
1	5	31-Jul-18	Charcoal Fabric	Loved it!	1
2	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer...	1
3	5	31-Jul-18	Charcoal Fabric	I have had a lot of fun with this thing. My 4 ...	1
4	5	31-Jul-18	Charcoal Fabric	Music	1

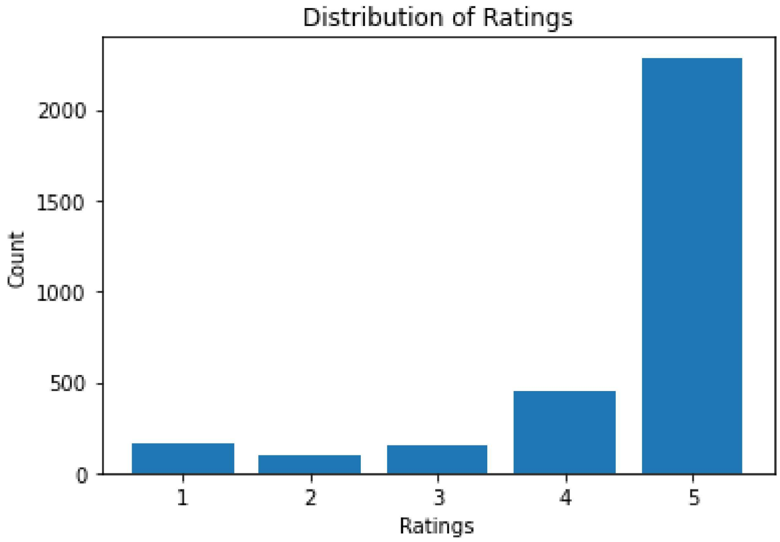
In [8]:

```
feedback_counts = alexa['feedback'].value_counts()
plt.bar(feedback_counts.index, feedback_counts.values)
plt.xlabel('Feedback')
plt.ylabel('Count')
plt.title('Distribution of Positive and Negative Feedback')
plt.xticks([0, 1], ['Negative', 'Positive'])  # Labeling x-axis
plt.show()
```



In [9]:

```
ratings_counts = alexa['rating'].value_counts().sort_index()
plt.bar(ratings_counts.index, ratings_counts.values)
plt.xlabel('Ratings')
plt.ylabel('Count')
plt.title('Distribution of Ratings')
plt.xticks(ratings_counts.index)
plt.show()
```



In [10]:

```
alexa['verified_reviews'] = alexa['verified_reviews'].str.lower()
alexa['verified_reviews']
```

Out[10]:

```
0                                           love my echo!
1                                               loved it!
2       sometimes while playing a game, you can answer...
3       i have had a lot of fun with this thing. my 4 ...
4                                                   music
                              ...
3145    perfect for kids, adults and everyone in betwe...
3146    listening to music, searching locations, check...
3147    i do love these things, i have them running my...
3148    only complaint i have is that the sound qualit...
3149                                                 good
Name: verified_reviews, Length: 3150, dtype: object
```

In [15]:

```
def remove_punctuation(text):
    punctuationfree="".join([i for i in text if i not in string.punctuation])
    return punctuationfree
#storing the puntuation free text
alexa['cleaned_reviews']= alexa['verified_reviews'].apply(lambda text: remove_punctuation(text))
alexa['cleaned_reviews']
```

Out[15]:

```
0                                            love my echo
1                                                loved it
2       sometimes while playing a game you can answer ...
3       i have had a lot of fun with this thing my 4 y...
4                                                   music
                              ...
3145      perfect for kids adults and everyone in between
3146    listening to music searching locations checkin...
3147    i do love these things i have them running my ...
3148    only complaint i have is that the sound qualit...
3149                                                 good
Name: cleaned_reviews, Length: 3150, dtype: object
```

In [17]:

```
def remove_emoticons(text):
    emoji_pattern = re.compile("["
                               u"\U0001F600-\U0001F64F"  # emoticons
                               u"\U0001F300-\U0001F5FF"  # symbols & pictographs
                               u"\U0001F680-\U0001F6FF"  # transport & map symbols
                               u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
                               u"\U00002702-\U000027B0"
                               u"\U000024C2-\U0001F251"
                               "]+", flags=re.UNICODE)
    return emoji_pattern.sub(r'', text)

alexa['cleaned_reviews'] = alexa['cleaned_reviews'].apply(remove_emoticons)
alexa['cleaned_reviews']
```

Out[17]:

```
0                                            love my echo
1                                                loved it
2       sometimes while playing a game you can answer ...
3       i have had a lot of fun with this thing my 4 y...
4                                                   music
                              ...                        
3145      perfect for kids adults and everyone in between
3146    listening to music searching locations checkin...
3147    i do love these things i have them running my ...
3148    only complaint i have is that the sound qualit...
3149                                                 good
Name: cleaned_reviews, Length: 3150, dtype: object
```

In [21]:

```
import re
def tokenization(text):
    tokens = re.split('W+',text)
    return tokens
#applying function to the column
alexa['msg_tokenied']= alexa['cleaned_reviews'].apply(lambda x: tokenization(x))
alexa['msg_tokenied'].head()
```

Out[21]:

```
0                                       [love my echo]
1                                           [loved it]
2    [sometimes while playing a game you can answer...
3    [i have had a lot of fun with this thing my 4 ...
4                                              [music]
Name: msg_tokenied, dtype: object
```

In [22]:

```
from nltk.corpus import stopwords
nltk.download('stopwords')
", ".join(stopwords.words('english'))
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\parth\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[22]:

```
"i, me, my, myself, we, our, ours, ourselves, you, you're, you've, you'll, you'd, your, yours, you
rself, yourselves, he, him, his, himself, she, she's, her, hers, herself, it, it's, its, itself, t
hey, them, their, theirs, themselves, what, which, who, whom, this, that, that'll, these, those, a
m, is, are, was, were, be, been, being, have, has, had, having, do, does, did, doing, a, an, the,
and, but, if, or, because, as, until, while, of, at, by, for, with, about, against, between, into,
through, during, before, after, above, below, to, from, up, down, in, out, on, off, over, under, a
gain, further, then, once, here, there, when, where, why, how, all, any, both, each, few, more, mo
st, other, some, such, no, nor, not, only, own, same, so, than, too, very, s, t, can, will, just,
don, don't, should, should've, now, d, ll, m, o, re, ve, y, ain, aren, aren't, couldn, couldn't, d
idn, didn't, doesn, doesn't, hadn, hadn't, hasn, hasn't, haven, haven't, isn, isn't, ma, mightn, m
ightn't, mustn, mustn't, needn, needn't, shan, shan't, shouldn, shouldn't, wasn, wasn't, weren, we
ren't, won, won't, wouldn, wouldn't"
```

In [23]:

```
STOPWORDS = set(stopwords.words('english'))
def remove_stopwords(text):
    return " ".join([word for word in str(text).split() if word not in STOPWORDS])

alexa["text_wo_stop"] = alexa["msg_tokenied"].apply(lambda text: remove_stopwords(text))
alexa.head()
```

Out[23]:

	rating	date	variation	verified_reviews	feedback	cleaned_punctuation	cleaned_reviews	tokenized_reviews	msg_
0	5	31-Jul-18	Charcoal Fabric	love my echo!	1	love my echo	love my echo	[love, my, echo]	[love
1	5	31-Jul-18	Charcoal Fabric	loved it!	1	loved it	loved it	[loved, it]	
2	4	31-Jul-18	Walnut Finish	sometimes while playing a game, you can answer...	1	sometimes while playing a game you can answer ...	sometimes while playing a game you can answer ...	[sometimes, while, playing, a, game, you, can,...	[s whi a g car
3	5	31-Jul-18	Charcoal Fabric	i have had a lot of fun with this thing. my 4 ...	1	i have had a lot of fun with this thing my 4 y...	i have had a lot of fun with this thing my 4 y...	[i, have, had, a, lot, of, fun, with, this, th...	[i h lot o this
4	5	31-Jul-18	Charcoal Fabric	music	1	music	music	[music]	

In [25]:

```
from nltk.stem.porter import PorterStemmer

stemmer = PorterStemmer()
def stem_words(text):
    return " ".join([stemmer.stem(word) for word in text.split()])

alexa["text_stemmed"] = alexa["verified_reviews"].apply(lambda text: stem_words(text))
alexa["text_stemmed"]
```

Out[25]:

```
0                                            love my echo!
1                                                 love it!
2       sometim while play a game, you can answer a qu...
3       i have had a lot of fun with thi thing. my 4 y...
4                                                    music
                              ...
3145     perfect for kids, adult and everyon in between!!
3146    listen to music, search locations, check time,...
3147    i do love these things, i have them run my ent...
3148    onli complaint i have is that the sound qualit...
3149                                                  good
Name: text_stemmed, Length: 3150, dtype: object
```

In [26]:

```
from nltk.stem import WordNetLemmatizer
nltk.download('wordnet')
lemmatizer = WordNetLemmatizer()
def lemmatize_words(text):
    return " ".join([lemmatizer.lemmatize(word) for word in text.split()])

alexa["text_lemmatized"] = alexa["text_wo_stop"].apply(lambda text: lemmatize_words(text))
alexa.head()
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\parth\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[26]:

	rating	date	variation	verified_reviews	feedback	cleaned_punctuation	cleaned_reviews	tokenized_reviews	msg_
0	5	31-Jul-18	Charcoal Fabric	love my echo!	1	love my echo	love my echo	[love, my, echo]	[love
1	5	31-Jul-18	Charcoal Fabric	loved it!	1	loved it	loved it	[loved, it]	
2	4	31-Jul-18	Walnut Finish	sometimes while playing a game, you can answer...	1	sometimes while playing a game you can answer ...	sometimes while playing a game you can answer ...	[sometimes, while, playing, a, game, you, can,...	[s whi a g car
3	5	31-Jul-18	Charcoal Fabric	i have had a lot of fun with this thing. my 4 ...	1	i have had a lot of fun with this thing my 4 y...	i have had a lot of fun with this thing my 4 y...	[i, have, had, a, lot, of, fun, with, this, th...	[i h lot o this
4	5	31-Jul-18	Charcoal Fabric	music	1	music	music	[music]	

In [29]:
```python
alexa.head()
```

Out[29]:

	rating	variation	verified_reviews	feedback	cleaned_punctuation	cleaned_reviews	tokenized_reviews	msg_tokeni
0	5	Charcoal Fabric	love my echo!	1	love my echo	love my echo	[love, my, echo]	[love my ech
1	5	Charcoal Fabric	loved it!	1	loved it	loved it	[loved, it]	[loved
2	4	Walnut Finish	sometimes while playing a game, you can answer...	1	sometimes while playing a game you can answer ...	sometimes while playing a game you can answer ...	[sometimes, while, playing, a, game, you, can,...	[sometim while playi a game y can answe
3	5	Charcoal Fabric	i have had a lot of fun with this thing. my 4 ...	1	i have had a lot of fun with this thing my 4 y...	i have had a lot of fun with this thing my 4 y...	[i, have, had, a, lot, of, fun, with, this, th...	[i have had lot of fun w this thing r 4
4	5	Charcoal Fabric	music	1	music	music	[music]	[mus

In [30]:
```python
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer()

bow_matrix = vectorizer.fit_transform(alexa['cleaned_reviews'])

bow_array = bow_matrix.toarray()

feature_names = vectorizer.get_feature_names_out()
```

```python
bow_df = pd.DataFrame(bow_array, columns=feature_names)

bow_df
```

Out[30]:

	072318	10	100	1000	100x	1010	1030pm	11	1100sf	1220	...	yr	yrs	yup	zero	zigbee	zonkedout
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
2	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	...	1	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3145	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3146	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3147	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3148	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
3149	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0

3150 rows × 4485 columns

In [32]:

```python
from sklearn.feature_extraction.text import TfidfVectorizer

tfidf_vectorizer = TfidfVectorizer()
tfidf_matrix = tfidf_vectorizer.fit_transform(alexa['cleaned_reviews'])
tfidf_array = tfidf_matrix.toarray()
tfidf_feature_names = tfidf_vectorizer.get_feature_names_out()
tfidf_df = pd.DataFrame(tfidf_array, columns=tfidf_feature_names)
```

In [35]:

```python
# Print the number of samples (documents)
print("Number of Samples (Documents):", len(alexa))

# Print the number of features (words)
print("Number of Features (Words):", len(tfidf_feature_names))
```

```
Number of Samples (Documents): 3150
Number of Features (Words): 4485
```

In [ ]: