VIEWER'S REVIEWS SENTIMENT ANALYSIS ON IMDB'S MADAM WEB MOVIE

Group members:

1. Muhammad Khairin Asnawi bin Rosli (IS01082068)
2. Mohd Taufiq Eizaz bin M Jamaludin (IS01080910)
3. Haliki Bachar Djimet (IS01081569)
4. Ayman Fikry bin Asmajuda (IS01081779)

Group Project (Part 2)

PREPARED FOR: TS. NUR LAILA BTE AB GHANI

Link to Video Presentation

TOPIC OUTLINE:

Data Gathering

Text Preprocessing

Modelling and Evaluation,

Results Visualization

Key Findings

DATA GATHERING

DATA GATHERING

1. Method used for Data Gathering:
    - The method used is web scraping for data extraction
2. Target Data Source:
    - Viewer reviews for the movie "Madam Web" on IMDb
3. Tool Utilized:
    - Python package BeautifulSoup
4. Process:
    - Parsed and collected textual data from IMDb user reviews section
5. Outcome:
    - Collected a thousand plus reviews on IMDB
    - These were separated by name, score,title of their review and comment underneath

TEXT PREPROCESSING

Data preprocessing steps:-

- Data Selection:
  - Selected relevant columns: Score and Text
- Data Cleaning:
  - Removed duplicate rows
  - Cleaned text by removing HTML tags, URLs, special characters, digits, and extra whitespaces
  - Converted text to lowercase
- Handling Missing Values:
  - Checked for missing values
  - Removed rows with missing 'Score' values
  - Tokenization & Stopword Removal:
  - Tokenized text into individual words using NLTK
  - Removed stop words
- Lemmatization:
  - Applied lemmatization to tokens using WordNet Lemmatizer
- Reconstruction:
  - Joined lemmatized tokens back into sentences

MODELLING AND EVALUATION

MODELLING : TEXTBLOB

```
In [26]: from textblob import TextBlob
         blob = TextBlob(mystring)
         blob.sentiment.polarity

Out[26]: 0.025581905683177322

In [27]: blob.sentiment.subjectivity

Out[27]: 0.5279996622967936

In [28]: blob.words

Out[28]: WordList(['well', 'guess', 'movie', 'sony', 'universe', 'im', 'shocked', 'profitablethe', 'story', 'place', 'basically', 'mad
         ame', 'web', 'paramedic', 'develops', 'power', 'see', 'future', 'current', 'threat', 'try', 'change', 'future', 'teaming', 'p
         owerful', 'character', 'yet', 'develop', 'full', 'potential', 'must', 'shield', 'change', 'future', 'protect', 'themtrying',
         'give', 'little', 'story', 'get', 'told', 'away', 'opening', 'vaguewhat', 'liked', 'one', 'two', 'good', 'scene', 'didnt', 'm
         ind', 'end', 'really', 'liked', 'lot', 'camera', 'work', 'cool', 'looking', 'shot', 'decent', 'scorebut', 'movie', 'bad', 'ma
         ke', 'morbius', 'look', 'like', 'batman', 'acting', 'lacking', 'actor', 'look', 'bored', 'feel', 'script', 'bad', 'beyond',
         'silly', 'line', 'story', 'chop', 'change', 'much', 'neck', 'hurt', 'bad', 'cgi', 'boring', 'actionand', 'people', 'harp', 'd
         ont', 'like', 'woman', 'lead', 'crap', 'film', 'enjoyed', 'marvel', 'black', 'widow', 'decent', 'movie', 'regardless', 'lea
         d', 'poorextremely', 'disappointing', 'first', 'impressionthe', 'opening', 'madame', 'web', 'immersed', 'mystique', 'ambitio
         n', 'enhanced', 'great', 'musical', 'score', 'yet', 'movie', 'progress', 'becomes', 'mixed', 'bag', 'rollercoaster', 'emotion
         sthe', 'positivemusical', 'score', 'sound', 'design', 'standout', 'feature', 'madame', 'web', 'musical', 'score', 'soundtrac
         k', 'enhances', 'emotional', 'depth', 'film', 'also', 'succeeds', 'creating', 'moment', 'genuine', 'goosebump', 'sound', 'des
         ign', 'really', 'enhances', 'impact', 'carefully', 'using', 'sound', 'pan', 'bass', 'reverse', 'effect', 'creates', 'sense',
         'tension', 'immersion', 'one', 'film', 'strongest', 'featuresthe', 'mixedvisual', 'effect', 'costume', 'design', 'visual', 'e
         ffect', 'madame', 'web', 'visually', 'appealing', 'costume', 'design', 'like', 'caseys', 'red', 'jacket', 'three', 'spider',
         'girl', 'villain', 'outfit', 'standing', 'particularly', 'stunning', 'contribution', 'film', 'aesthetic', 'often', 'shown',
         'film', 'since', 'film', 'direction', 'focus', 'different', 'disappointedacting', 'direction', 'although', 'thought', 'casey
         s', 'character', 'development', 'ability', 'see', 'world', 'different', 'perspective', 'great', 'didnt', 'like', 'visually',
         'presented', 'way', 'edited', 'poorly', 'done', 'editor', 'could', 'use', 'distinct', 'color', 'grading', 'feature', 'one',
```

MODELLING : LEXICON-BASED APPROACH

In [29]:
```
# Download the VADER lexicon
nltk.download('vader_lexicon')

# Define a function to assign sentiment labels based on the 'Score' column
def assign_sentiment(score):

    if score >= 7:
        return 'Positive'
    elif score <= 3:
        return 'Negative'
    else:
        return 'Neutral'

# Assign sentiment labels based on the 'Score' column
data['Sentiment'] = data['Score'].apply(assign_sentiment)

# Initialize the sentiment analyzer
sid = SentimentIntensityAnalyzer()

# Calculate sentiment scores for each review
data['Lexicon_Sentiment'] = data['Preprocessed_Text'].apply(lambda x: sid.polarity_scores(x)['compound'])

# Map sentiment scores to labels
data['Lexicon_Sentiment_Label'] = data['Lexicon_Sentiment'].apply(lambda x: 'Positive' if x > 0 else ('Negative' if x < 0 else 'N

# Evaluate the lexicon-based approach
lexicon_accuracy = accuracy_score(data['Sentiment'], data['Lexicon_Sentiment_Label'])
print("Accuracy of the Lexicon-based Approach:", lexicon_accuracy)
```

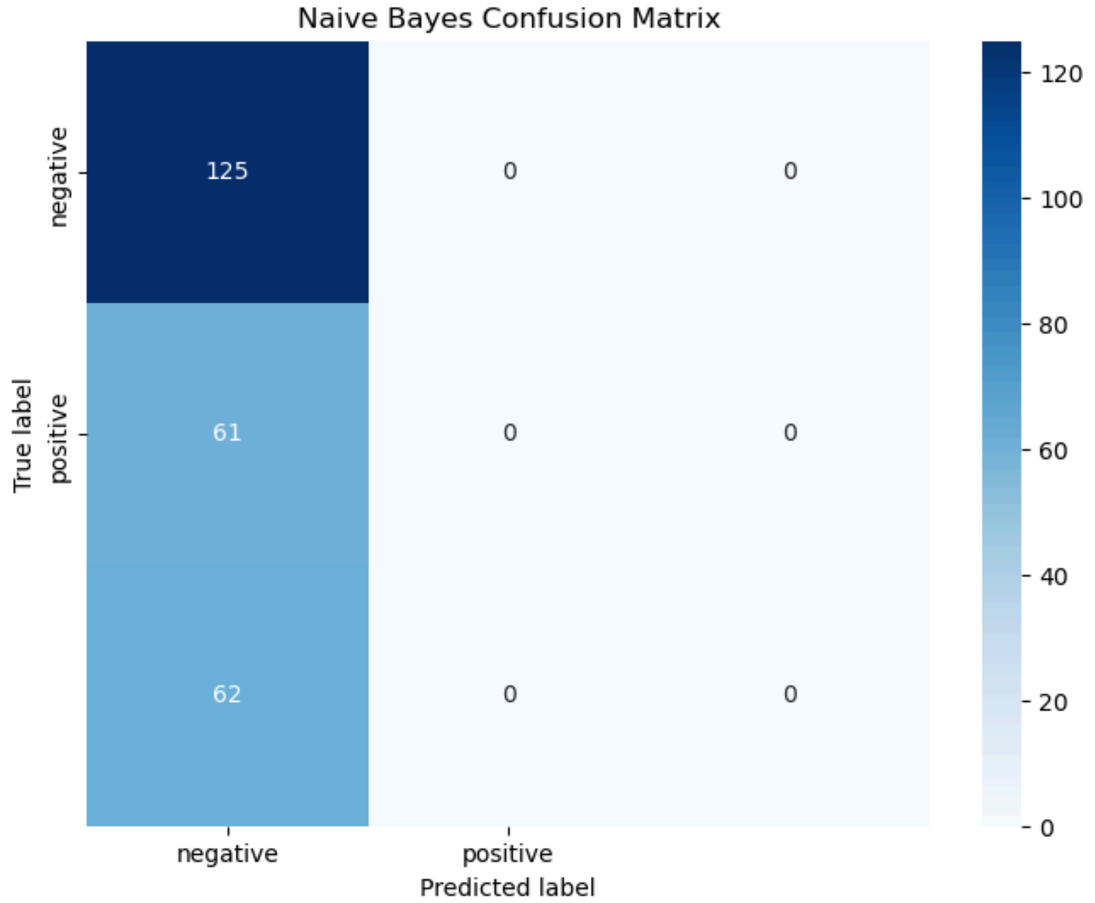
```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\Administrator\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

```
Accuracy of the Lexicon-based Approach: 0.4727822580645161
```

```
X_train_tfidf shape: (793, 10753)
X_test_tfidf shape: (199, 10753)
Naive Bayes Classifier Accuracy: 0.49748743718592964
              precision    recall  f1-score   support

    Negative       0.50      1.00      0.66        99
     Neutral       0.00      0.00      0.00        50
    Positive       0.00      0.00      0.00        50

    accuracy                           0.50       199
   macro avg       0.17      0.33      0.22       199
weighted avg       0.25      0.50      0.33       199

Support Vector Machine (SVM) Classifier Accuracy: 0.7185929648241206
              precision    recall  f1-score   support

    Negative       0.73      0.97      0.83        99
     Neutral       0.55      0.32      0.41        50
    Positive       0.79      0.62      0.70        50

    accuracy                           0.72       199
   macro avg       0.69      0.64      0.65       199
weighted avg       0.70      0.72      0.69       199
```

MODELLING : MACHINE-LEARNING BASED APPROACH

MODELLING : NB CONFUSION MATRIX

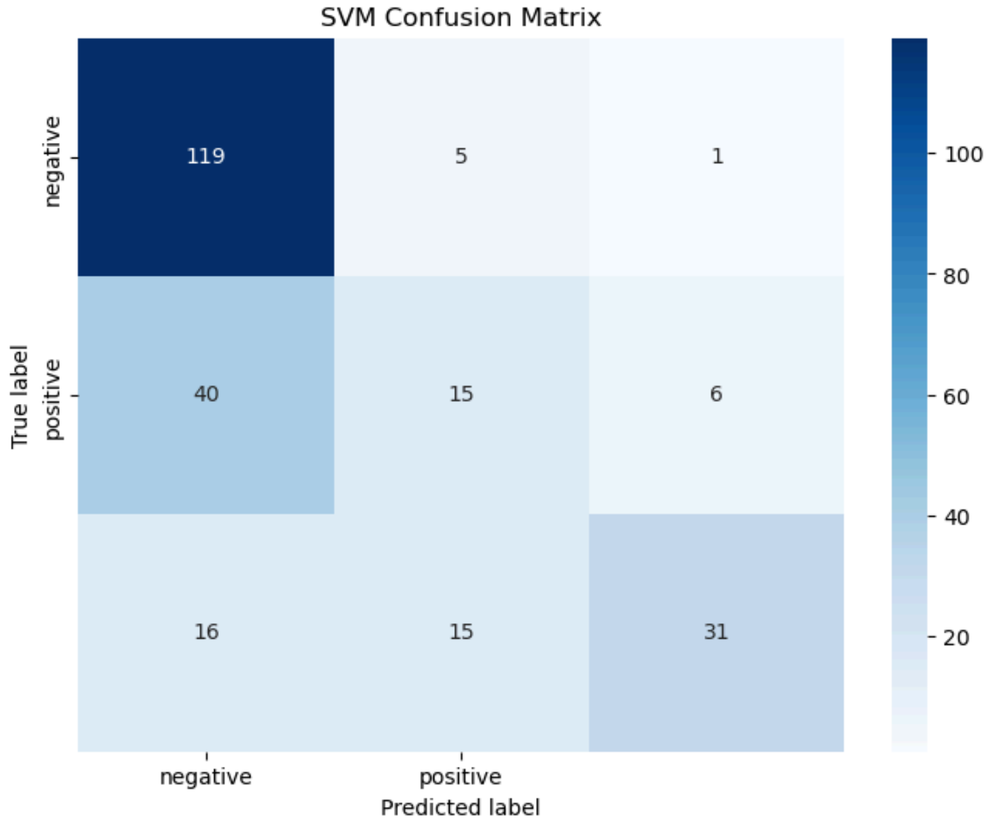


Naive Bayes Classifier Report:

	precision	recall	f1-score	support
Negative	0.50	1.00	0.67	125
Neutral	0.00	0.00	0.00	61
Positive	0.00	0.00	0.00	62
accuracy			0.50	248
macro avg	0.17	0.33	0.22	248
weighted avg	0.25	0.50	0.34	248

Naive Bayes Error Rate: 0.50

MODELLING : SVM CONFUSION MATRIX



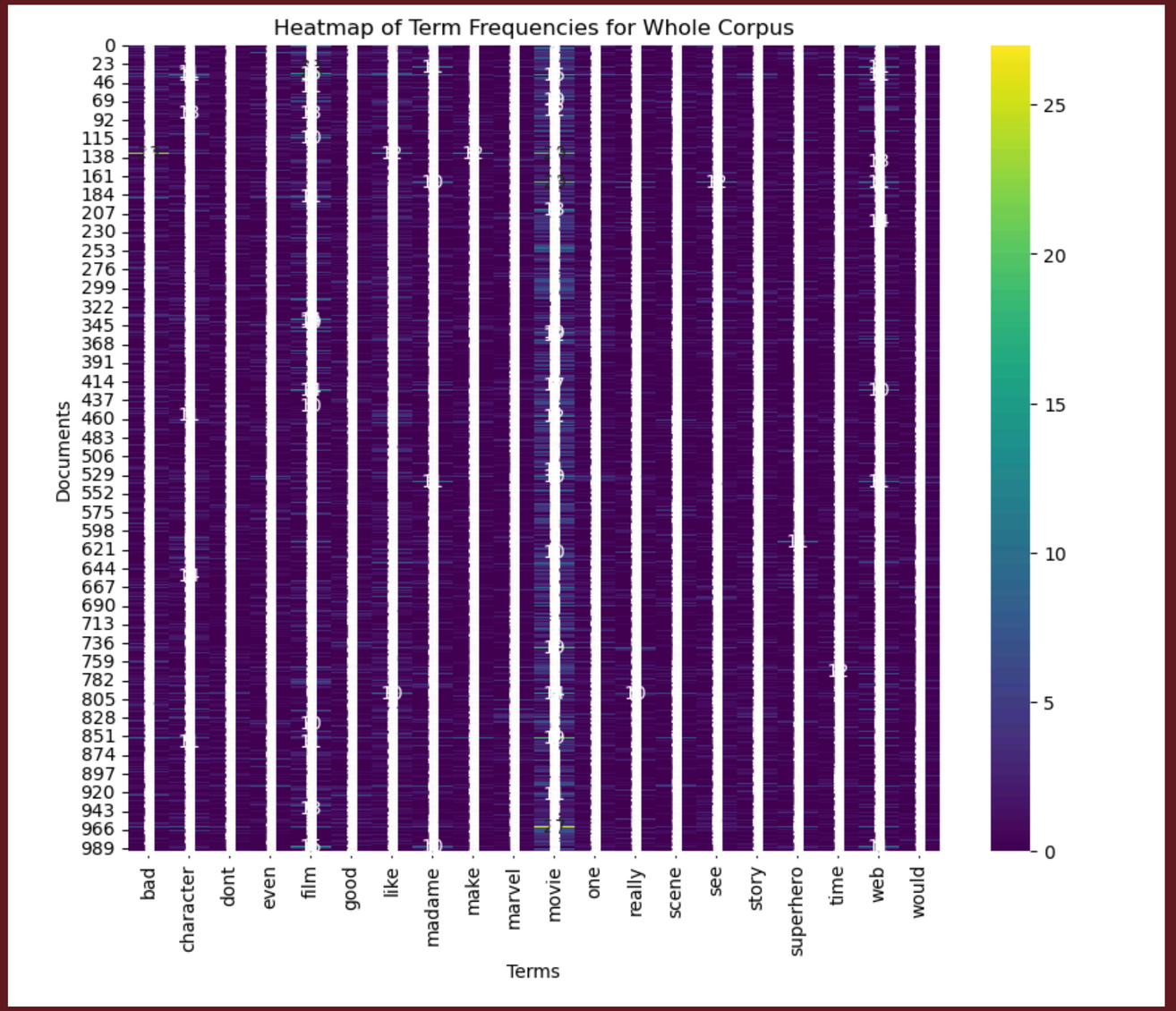
SVM Classifier Report:

	precision	recall	f1-score	support
Negative	0.68	0.95	0.79	125
Neutral	0.43	0.25	0.31	61
Positive	0.82	0.50	0.62	62
accuracy			0.67	248
macro avg	0.64	0.57	0.58	248
weighted avg	0.65	0.67	0.63	248

SVM Error Rate: 0.33

RESULTS VISUALIZATION

HEATMAP

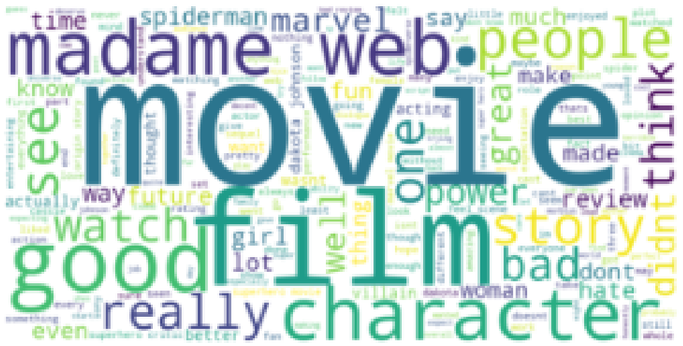

Heatmap of Term Frequencies for Whole Corpus
Documents
Terms
bad
character
dont
even
film
good
like
madame
make
marvel
movie
one
really
scene
see
story
superhero
time
web
would
0
23
46
69
92
115
138
161
184
207
230
253
276
299
322
345
368
391
414
437
460
483
506
529
552
575
598
621
644
667
690
713
736
759
782
805
828
851
874
897
920
943
966
989
0
5
10
15
20
25

WORDCLOUD

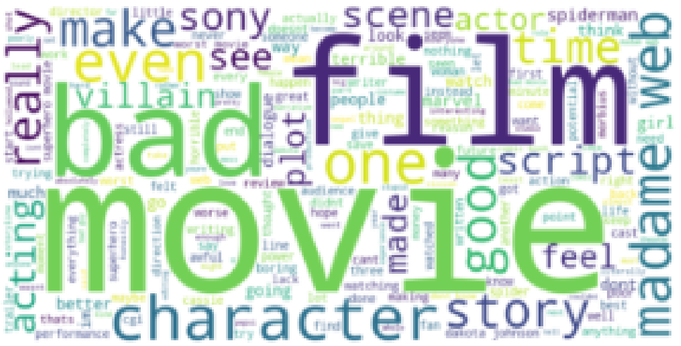

Word Cloud for Whole Corpus
something
without
seen
review
much
villain
plot
future
madame web
character
movie
scene
fun
writing
cant
think
girl
story
action
nothing
go
seem
know
felt
thats
end
bit
best
start
watching
thought
superhero movie
want
come
great
im
part
time
better
spider
work
sony
first
spiderman
power
point
line
watch
maybe
little
cast
cassie
dont
feel
writer
web
see
people
make
superhero
good
give
say
thing
dakota johnson
script
anything
way
look
got
film
bad
made
didnt
least
woman
never
even
acting
need
well
dialogue
marvel
isnt
many
still
one
lot
hope
going
performance
really
moment
actor
everything
making
wasnt
actually
terrible
right

WORLDCLOUD DIVIDED BY SENTIMENTS

Positive Sentiment

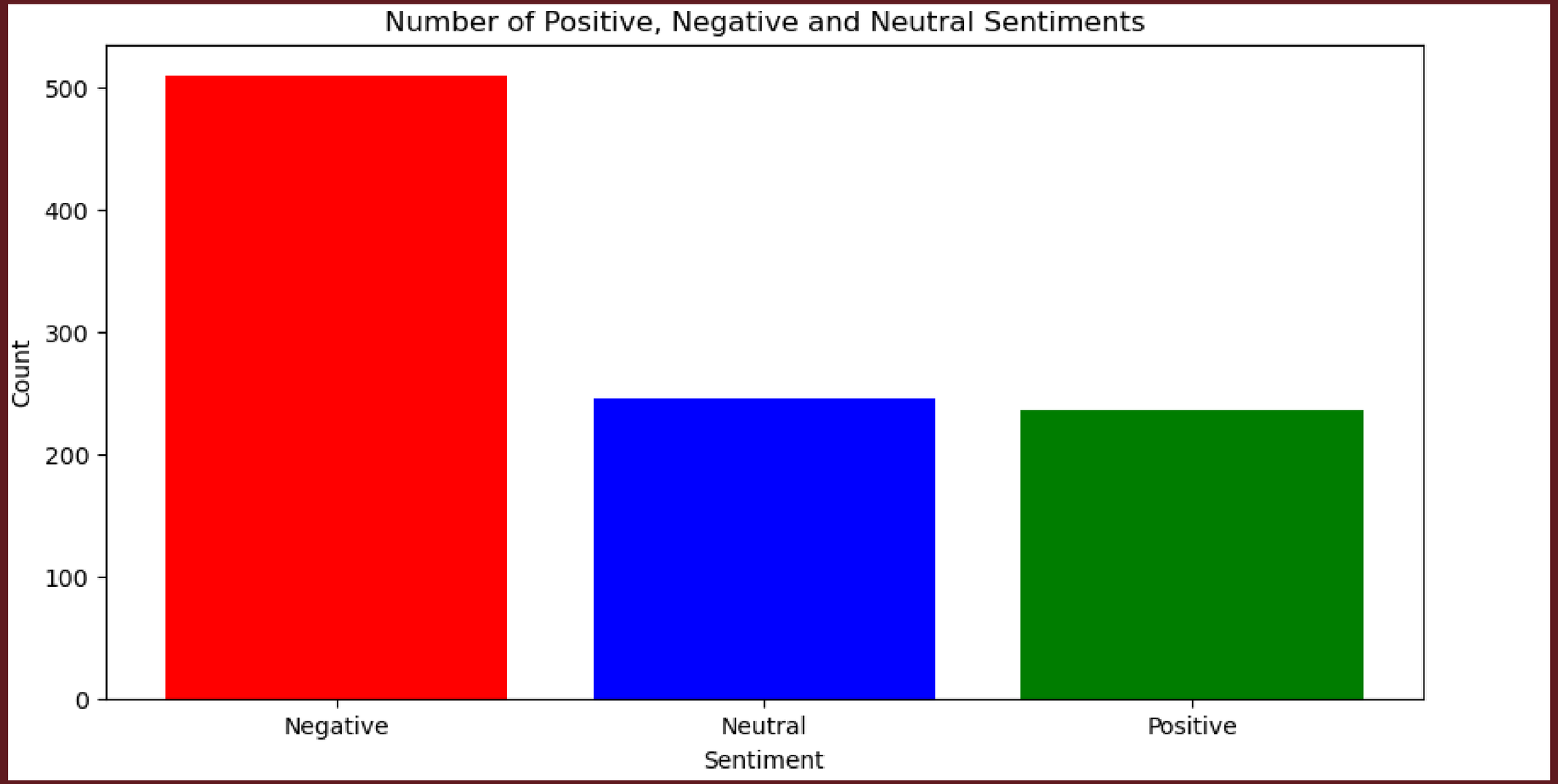


Negative Sentiment



Neutral Sentiment

BAR CHART COUNT OF DIFFERENT SENTIMENTS



KEY FINDINGS

- Data Gathering:

Method: Web scraping
Tool: BeautifulSoup
Source: IMDb user reviews for "Madam Web"
Outcome: Collected 1000+ reviews

- Text Preprocessing:

Data Selection: Selected Score and Text columns
Data Cleaning: Removed duplicates, cleaned HTML tags, URLs, special characters, digits, and converted text to lowercase
Handling Missing Values: Removed rows with missing scores
Tokenization & Stopword Removal: Tokenized text and removed stop words using NLTK
Lemmatization: Applied lemmatization and reconstructed sentences

- Modeling & Evaluation:

Models Used: Various sentiment analysis models
Evaluation: Performed cross-validation and measured accuracy

- Results:

Heat Map: Displayed sentiment distribution across reviews

THANK YOU FOR LISTENING