PHASE 5 PROJECT SUBMISSION

PROJECT TITLE:

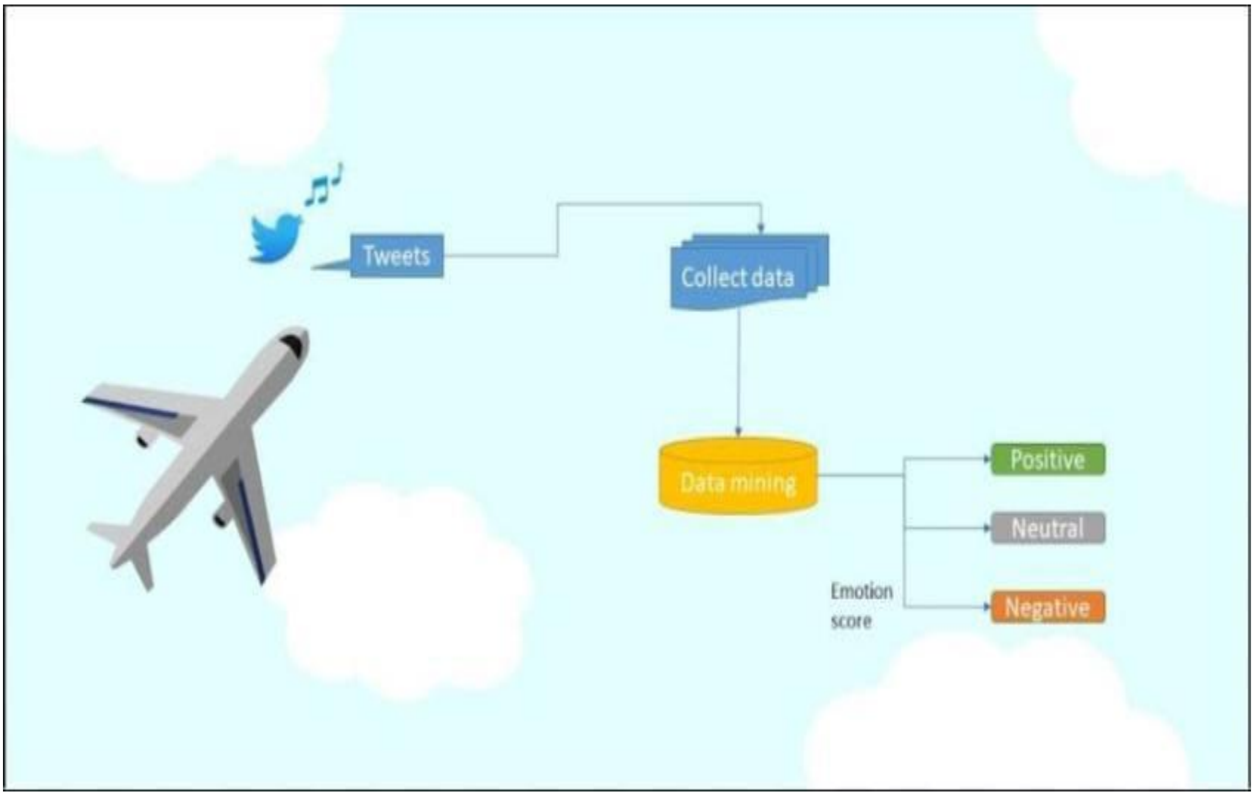
SENTIMENT ANALYSIS FOR MARKETING

PHASE 5:

PROJECT DOCUMENTATION & SUBMISSION

TOPIC:

COMPLETE PROJECT DOCUMENTATION AND PREPARE FOR SUBMISSION



INTRODUCTION:

➢ Sentiment analysis is a powerful tool in the realm of US airlines marketing, enabling companies to gain valuable insights into how customers perceive their services.

➢ In a highly competitive industry where customer experience is paramount, understanding the sentiment behind customer feedback is critical.

➢ Sentiment analysis involves the use of natural language processing and machine learning techniques to classify and quantify customer comments and reviews, helping airlines discern whether the sentiment is positive, negative, or neutral.

➢ This analysis goes beyond merely gauging customer satisfaction; it provides airlines with the ability to pinpoint specific pain points and areas of excellence, offering actionable insights for marketing campaigns, service improvements, and crisis management.

➢ By harnessing the sentiments expressed in social media, customer reviews, and other textual data, US airlines can tailor their marketing strategies to enhance customer satisfaction, strengthen brand loyalty, and stay ahead in a dynamic and competitive industry.

➢ Furthermore, sentiment analysis equips airlines with the tools to detect and respond to emerging trends, ensuring that their marketing efforts remain relevant and responsive evolving customer sentiment.

Dataset Link:

https://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment

GIVEN DATASET:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	tweet_id	airline_se	airline_se	negativer	negativer	airline	airline_se	name	negativer	retweet_c	text	tweet_coord	tweet_created	tweet_locatior	user_timezone		
2	5.7E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica What @dhepburn sa		24-02-2015 11:35		Eastern Time (US & Canada)		
3	5.7E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plus you've added c		24-02-2015 11:15		Pacific Time (US & Canada)		
4	5.7E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I didn't today... Mus		24-02-2015 11:15	Lets Play	Central Time (US & Canada)		
5	5.7E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's really aggressive		24-02-2015 11:15		Pacific Time (US & Canada)		
6	5.7E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and it's a really big b		24-02-2015 11:14		Pacific Time (US & Canada)		
7	5.7E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@VirginAmeri		24-02-2015 11:14		Pacific Time (US & Canada)		
8	5.7E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes, nearly every tir		24-02-2015 11:13	San Francisco (	Pacific Time (US & Canada)		
9	5.7E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Really missed a prin		24-02-2015 11:12	Los Angeles	Pacific Time (US & Canada)		
10	5.7E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Well, I didn'tâ€¦but		24-02-2015 11:11	San Diego	Pacific Time (US & Canada)		
11	5.7E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it was amazing, and		24-02-2015 10:53	Los Angeles	Eastern Time (US & Canada)		
12	5.7E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did you know that s		24-02-2015 10:48	1/1 loner squa	Eastern Time (US & Canada)		
13	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I <3 pretty graphi		24-02-2015 10:30	NYC	America/New_York		
14	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica This is such a great c		24-02-2015 10:30	NYC	America/New_York		
15	5.7E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @virginmedia I'm fl		24-02-2015 10:21		Eastern Time (US & Canada)		
16	5.7E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Thanks!		24-02-2015 10:15	San Francisco,	Pacific Time (US & Canada)		
17	5.7E+17	negative	0.6842	Late Flight	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFO-PDX schedule i		24-02-2015 10:01	palo alto, ca	Pacific Time (US & Canada)		
18	5.7E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So excited for my fi		24-02-2015 09:42	west covina	Pacific Time (US & Canada)		
19	5.7E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I flew from NYC to S		24-02-2015 09:39	this place calle	Eastern Time (US & Canada)		
20	5.7E+17	positive	1			Virgin America		thebrandiray		0	I â¤ï¸ flying @VirginAmerica. â˜ºï¸ðŸ‘		24-02-2015 09:15	Somewhere ce	Atlantic Time (Canada)		
21	5.7E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you know what wou		24-02-2015 09:04	Boston | Walth	Quito		
22	5.7E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica why are your first fa		24-02-2015 08:55				
23	5.7E+17	positive	1			Virgin America		DT_Les		0	@VirginAmeric	[40.74804263, -73.992	24-02-2015 08:49				
24	5.7E+17	positive	1			Virgin America		ElvinaBeck		0	@VirginAmerica I love the hipster in		24-02-2015 08:30	Los Angeles	Pacific Time (US & Canada)		
25	5.7E+17	neutral	1			Virgin America		rjlynch21086		0	@VirginAmerica will you be making		24-02-2015 08:27	Boston, MA	Eastern Time (US & Canada)		
26	5.7E+17	negative	1	Customer	0.3557	Virgin America		ayeevickiee		0	@VirginAmerica you guys messed up		24-02-2015 08:18	714	Mountain Time (US & Canada)		
27	5.7E+17	negative	1	Customer	1	Virgin America		Leora13		0	@VirginAmerica status match progra		24-02-2015 07:49				

14641 Rows*17 Columns

Tools and software commonly used in the process for sentiment analysis for us airlines marketing

1.Python:

A versatile programming language used for natural language processing and sentiment analysis.

2. NLTK (Natural Language Toolkit):

A Python library for working with human language data, including sentiment analysis.

3. spaCy:

Another Python library for NLP, known for its speed and efficiency in text processing.

4. VADER Sentiment Analysis:

A lexicon and rule-based sentiment analysis tool for Python.

5. TextBlob:

A simple Python library for processing textual data, including sentiment analysis.

6. IBM Watson NLU:

A cloud-based service offering sentiment analysis capabilities.

7. Google Cloud Natural Language API:

Google's cloud-based NLP API, which includes sentiment analysis.

8. Microsoft Azure Text Analytics:

Azure's cloud service for text analytics, offering sentiment analysis features.

9. Tableau:

Data visualization software used to create interactive dashboards for sentiment analysis results.

10. Power BI:

Microsoft's business analytics service for creating reports and dashboards.

11. RapidMiner:

An integrated data science platform that includes sentiment analysis capabilities.

12. Brandwatch:

A social listening and analytics platform for monitoring social media mentions and sentiments.

13. Hootsuite:

A social media management tool that provides sentiment analysis features.

14. Sprout Social:

Another social media management and analytics platform.

15. SurveyMonkey:

An online survey platform used for collecting customer feedback.

16. Qualtrics:

A comprehensive experience management platform, including survey and feedback analysis.

17. Salesforce:

A leading CRM software that can be used for managing customer interactions and feedback.

18. HubSpot:

An inbound marketing, sales, and customer service platform with CRM features.

19. Mention:

A social media and web monitoring tool for tracking brand mentions and sentiment.

20. Lexalytics:

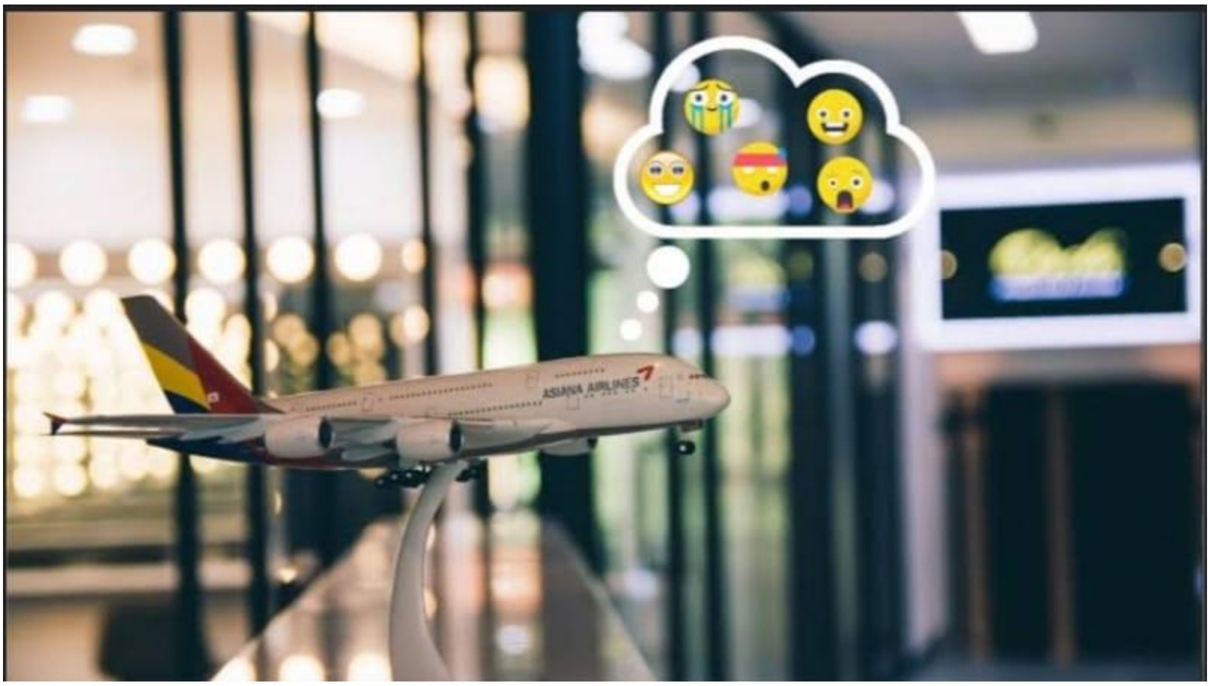
A specialized sentiment analysis software focused on business applications.

21. Semantria:

A sentiment analysis and text analytics solution for various industries.

22. Clarabridge:

A customer experience management platform with sentiment analysis capabilities.



Design Thinking Document

Project Objective:

Enhance the customer experience and brand reputation of US airlines through sentiment analysis of customer feedback and social media mentions.

Empathize

User Research

- Conduct user interviews and surveys to understand the needs and pain points of airline customers.
- Collect and analyze existing customer feedback data and social media conversations.

Define

- Identify common themes and sentiments in customer feedback.
- Define key customer personas and their sentiments.
- Create a problem statement: "How might we improve the overall sentiment of customers towards US airlines?"

Ideate

Brainstorming

- Organize brainstorming sessions with cross-functional teams.
- Generate creative solutions to address identified issues.

Prototyping

- Develop a prototype sentiment analysis system using Python and NLTK.
- Utilize a sample dataset for initial testing.

Prototype

Testing

- Gather feedback from internal team members and pilot users.
- Iterate on the prototype based on the feedback received.

Test

Data Collection

- Collect real-time customer feedback and social media data using monitoring tools.
- Preprocess the data to remove noise and irrelevant information.

Sentiment Analysis

- Utilize machine learning models and NLP techniques to perform sentiment analysis.
- Categorize sentiment as positive, negative, or neutral.

Visualization

- Use data visualization tools to create dashboards presenting sentiment trends.
- Share these visualizations with the marketing and customer service teams.

Implement

Actionable Insights

- Provide actionable insights to marketing and customer service teams based on sentiment analysis.
- Identify areas for improvement and recommend strategies to enhance the customer experience.

Continuous Monitoring

- Set up automated sentiment monitoring processes.
- Regularly review and update sentiment analysis models for accuracy.

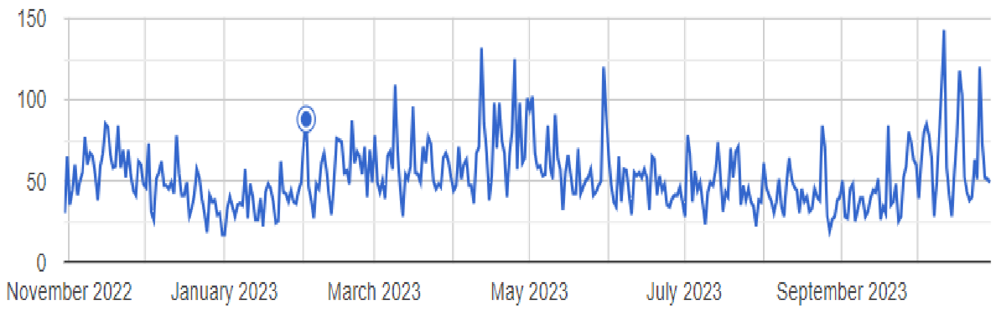
Learn

Performance Evaluation

- Continuously assess the impact of sentiment analysis on improving customer satisfaction and brand reputation.

Feedback Loop

- Gather feedback from marketing, customer service, and customers.
- Adapt strategies based on the feedback receive.



DESIGN INTO INNOVATION

Incorporate Emerging Technologies

1. AI and Machine Learning

Advanced Models: Develop state-of-the-art sentiment analysis models using deep learning algorithms like LSTM and BERT for enhanced accuracy.

Real-time Analysis: Implement AI-driven real-time sentiment analysis to react to customer sentiments instantly.

2. Voice Analysis

Call Center Data: Utilize voice analysis tools to extract sentiments from customer service calls, providing insights into customer experiences and issues.

In-flight Interaction: Integrate voice sentiment analysis into in-flight interactions, allowing airlines to monitor passenger emotions during the journey.

3. Sentiment Detection in Images

Visual Sentiment: Implement image recognition and sentiment analysis to analyze customer-posted images on social media for non-textual sentiments.

Emotion Recognition: Detect facial expressions in passenger photos and videos to understand in-the-moment sentiments.

Leverage Data Sources

4. IoT Data

In-flight Sensors: Utilize data from in-flight sensors and IoT devices to gather real-time passenger feedback and emotional data.

Aircraft Health Data: Integrate aircraft health data to assess the impact of flight conditions on passenger sentiment.

5. Blockchain for Data Security

Data Security: Utilize blockchain technology to secure customer data while maintaining transparency, ensuring data privacy and security compliance.

Anonymized Feedback: Enable passengers to provide feedback anonymously, enhancing honesty and compliance with privacy regulations.

6. Open Data Collaboration

Industry Partnerships: Collaborate with airports, hotels, and other travel-related companies to access a broader array of data for sentiment analysis.

Synergy Opportunities: Leverage data sharing to enhance the customer journey and overall travel experience.

Personalization and Predictive Analytics

7. Customer Journey Mapping

Journey Analysis: Create comprehensive customer journey maps to understand touchpoints, emotions, and pain points along the travel experience.

Micro-Personalization: Use micro-segmentation to tailor services, offers, and communication based on passengers' emotional state during their journey.

8. Predictive Insights

Sentiment Forecasting: Develop predictive models that forecast customer sentiment trends based on historical and real-time data, allowing proactive interventions.

Behavioral Predictors: Identify behavioral patterns that signal changes in passenger sentiment and respond accordingly.

Augmented Reality (AR) and Virtual Reality (VR)

9. AR/VR Customer Feedback

Immersive Feedback: Allow passengers to provide feedback in immersive AR/VR environments, capturing richer sentiments and suggestions.

Virtual Cabin Experiences: Develop VR-based simulations to receive feedback on in-flight services, cabin layout, and entertainment options.

10. VR-Based Training

Employee Empathy: Train customer service teams using VR simulations to enhance empathy, communication skills, and conflict resolution.

Virtual Crisis Management: Simulate crisis situations to train staff on managing and mitigating customer sentiment during emergencies.

Ethical AI and Privacy Compliance

11. Data Ethics

Data Guidelines: Develop clear ethical guidelines for data collection and usage, ensuring compliance with data privacy regulations (e.g., GDPR).

Transparency: Make AI algorithms transparent to passengers to build trust and transparency.

12. Algorithm Transparency

Explainable AI: Invest in explainable AI techniques to ensure transparency and to explain AI-driven decisions to passengers.

Trust Assurance: Establish trust through transparency, allowing customers to understand how their data is used.

Real-time Sentiment Management

13. Real-time Feedback Loop

Immediate Response: Implement a real-time feedback loop to address issues as they arise, improving customer satisfaction instantly.

Crisis Management: Use real-time sentiment analysis to detect and manage potential PR crises effectively.

14. Social and Environmental Responsibility

Sustainability Metrics: Evaluate sentiment around airlines' environmental initiatives and incorporate feedback into sustainability strategies.

Community Engagement: Foster positive sentiment by actively engaging in community projects and sharing the results with passengers.

LOAD THE CLEAN DATASET:

#import the required library

```
import pandas as pd
import numpy as np
import nltk
import re
import os
import random
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import CountVectorizer
```

Clean the data set

Program

```
clean_data = pd.read_csv('../input/twitter-airline-sentiment/T
weets.csv')
clean_data.head()
```

Output:

	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidenc
0	570306133677760513	neutral	1.0000	NaN	NaN
1	570301130888122368	positive	0.3486	NaN	0.0000
2	570301083672813571	neutral	0.6837	NaN	NaN
3	570301031407624196	negative	1.0000	Bad Flight	0.7033
4	570300817074462722	negative	1.0000	Can't Tell	1.0000

```
clean_data.info()
```

OUTPUT

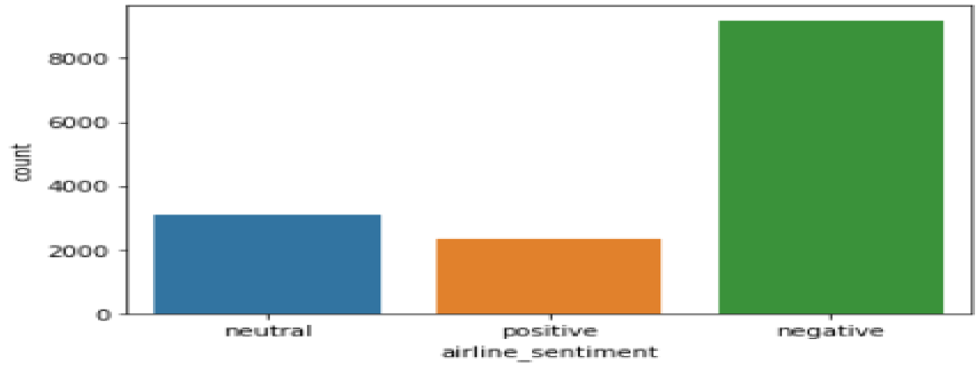
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14640 entries, 0 to 14639
Data columns (total 15 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   tweet_id                      14640 non-null  int64
 1   airline_sentiment             14640 non-null  object
 2   airline_sentiment_confidence  14640 non-null  float64
 3   negativereason                9178 non-null   object
 4   negativereason_confidence     10522 non-null  float64
 5   airline                       14640 non-null  object
 6   airline_sentiment_gold        40 non-null     object
 7   name                          14640 non-null  object
 8   negativereason_gold           32 non-null     object
 9   retweet_count                 14640 non-null  int64
 10  text                          14640 non-null  object
 11  tweet_coord                   1019 non-null   object
 12  tweet_created                 14640 non-null  object
 13  tweet_location                9907 non-null   object
 14  user_timezone                 9820 non-null   object
dtypes: float64(2), int64(2), object(11)
memory usage: 1.7+ MB
```

PROGRAM

```
sns.countplot(x = "airline_sentiment", data = clean_data)
```

OUTPUT

```
<AxesSubplot:xlabel='airline_sentiment', ylabel='count'>
```



TEXT PRE-PROCESSING:

Text preprocessing steps include a few essential tasks to further clean the available text data. It includes tasks like:-

1.Stop-Word Removal:

In English words like a, an, the, as, in, on, etc. are considered as stop-words so according to our requirements we can remove them to reduce vocabulary size as these words don't have some specific meaning

2. Lower Casing:

Convert all words into the lower case because the upper or lower case may not make a difference for the problem. And we are reducing vocabulary size by doing so.

3. Stemming:

Stemming refers to the process of removing suffixes and reducing a word to some base form such that all different variants of that word can be represented by the same form (e.g., “walk” and “walking” are both reduced to “walk”).

4. Tokenization:

NLP software typically analyzes text by breaking it up into words (tokens) and sentences.

Pre-processing of the text is not the main objective of this notebook that's why I am just covering a few basic steps in a brief

```
# First of all let's drop the columns which we don't required

waste_col = ['tweet_id', 'airline_sentiment_confidence',
       'negativereason', 'negativereason_confidence', 'airline
       'airline_sentiment_gold', 'name', 'negativereason_gold'
       'retweet_count', 'tweet_coord', 'tweet_created',
       'tweet_location', 'user_timezone']

data = clean_data.drop(waste_col, axis = 1)
```

```
data.head()
```

OUTPUT:

	airline_sentiment	text
0	neutral	@VirginAmerica What @dhepburn said.
1	positive	@VirginAmerica plus you've added commercials t...
2	neutral	@VirginAmerica I didn't today... Must mean I n...
3	negative	@VirginAmerica it's really aggressive to blast...
4	negative	@VirginAmerica and it's a really big bad thing...

PROGRAM:

```
def sentiment(x):
    if x == 'positive':
        return 1
    elif x == 'negative':
        return -1
    else:
        return 0
```

```
nltk.download('stopwords')
```

```python
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer
from nltk.tokenize import RegexpTokenizer

stopwords = stopwords.words('english')
stemmer = SnowballStemmer('english')
tokenizer = RegexpTokenizer(r'\w+')
# As this dataset is fetched from twitter so it has lots of peop
le tag in tweets
# we will remove them
tags = r"@\w*"


def preprocess_text(sentence, stem = False):

    sentence = [re.sub(tags, "", sentence)]
    text = []
    for word in sentence:

        if word not in stopwords:

            if stem:
                text.append(stemmer.stem(word).lower())
            else:
                text.append(word.lower())
    return tokenizer.tokenize(" ".join(text))
```

OUTPUT:

```
[nltk_data] Downloading package stopwords to /usr/share/nltk_d
ata...
[nltk_data]   Package stopwords is already up-to-date!
```

PROGRAM:

```python
print(f"Orignal Text : {data.text[11]}")
print()
print(f"Preprocessed Text : {preprocess_text(data.text[11])}")
```

```
Orignal Text : @VirginAmerica I &lt;3 pretty graphics. so much
better than minimal iconography. :D
```

```
Preprocessed Text : ['i', 'lt', '3', 'pretty', 'graphics', 'so
', 'much', 'better', 'than', 'minimal', 'iconography', 'd']
```

```
data.text = data.text.map(preprocess_text)
data.head()
```

OUTPUT:

	airline_sentiment	text
0	neutral	[what, said]
1	positive	[plus, you, ve, added, commercials, to, the, e...
2	neutral	[i, didn, t, today, must, mean, i, need, to, t...
3	negative	[it, s, really, aggressive, to, blast, obnoxio...
4	negative	[and, it, s, a, really, big, bad, thing, about...

VISUALIZATION FOR SENTIMENT ANALYSIS:

Visualization plays a crucial role in sentiment analysis for US airlines marketing, as it helps transform complex sentiment data into clear, actionable insights. Here are some key aspects of visualization in sentiment analysis for this specific context:

1. Dashboard Creation:

Sentiment Overview: Develop a comprehensive dashboard that provides an overview of sentiment trends over time, including positive, negative, and neutral sentiments.

Customer Sentiment Segmentation: Create visual representations of sentiment data segmented by different factors such as flight routes, customer demographics, and travel classes.

2. Temporal Analysis:

Time-Series Plots: Use line charts to illustrate sentiment trends over time, which can help identify seasonal patterns or the impact of specific events on sentiment.

Heatmaps: Utilize heatmaps to visualize sentiment fluctuations during specific hours of the day or days of the week.

3. Geospatial Analysis:

Sentiment Maps: Create geographical maps with color-coded sentiment markers to show regional variations in customer sentiment.

Airport-Specific Sentiment: Visualize sentiment data for each airport or destination, allowing airlines to tailor marketing efforts.

4. Customer Persona Analysis:

Sentiment by Customer Segments: Generate bar charts or pie charts to depict sentiment for different customer personas, highlighting their unique experiences.

Customer Sentiment Journey: Create journey maps that track sentiment from booking to post-trip, identifying key touchpoints.

5. Social Media Monitoring:

Word Clouds: Display word clouds to highlight the most frequently mentioned terms in social media conversations, helping airlines understand popular topics.

Hashtag Analysis: Track and visualize sentiment associated with specific hashtags used in airline-related posts.

6. Feedback Source Visualization:

Source Attribution: Differentiate sentiment by the source of feedback, such as surveys, social media, or customer service interactions.

Channel-specific Sentiment: Illustrate how sentiment varies across platforms like Twitter, Facebook, Instagram, or review websites.

7. Competitor Benchmarking:

Competitor Sentiment Comparison: Use side-by-side bar charts to compare sentiment with competitors, identifying strengths and weaknesses.

Market Share Analysis: Visualize the correlation between sentiment trends and market share in the airline industry.

8. Sentiment Impact Analysis:

Revenue Impact: Create visualizations that show the correlation between sentiment improvements and changes in revenue.

Customer Retention: Visualize how sentiment influences customer retention and loyalty, helping airlines understand the business impact.

9. Real-time Sentiment Monitoring:

Live Feeds: Implement live sentiment feeds on dashboards to enable immediate response to emerging sentiment issues or crises.

Alerts and Notifications: Use visual indicators and notifications to alert teams when sentiment thresholds are reached.

10. Social Media Engagement:

Response Heatmap: Visualize response rates and sentiment changes after airline interactions on social media, highlighting successful engagements.

Virality Tracking: Monitor the virality of positive customer experiences by visualizing the spread of positive sentiment across social networks.

11. Feedback Sentiment Analysis:

Voice of the Customer (VoC): Create visualizations that consolidate feedback from various sources to present a holistic view of customer sentiment.

Topic Modeling: Utilize topic modeling techniques to visualize the most prominent themes and their associated sentiments.

12. Storytelling Visuals:

Narrative Infographics: Develop infographics that tell a compelling story about sentiment changes, improvements, and their impact on the airline's brand.

Emotional Journey Maps: Create visual journey maps that reflect passengers' emotional experiences throughout their travel.

PROGRAM

```
# Visualizing Word2vec Word Embedding

keys = ['India', 'good', 'friday', 'science', 'Twitter', 'mast
ers', 'computer', 'election', 'costly',
        'learning', 'finance', 'machine', 'android', 'peace',
'nature', 'war']

words_clusters = []
embeddings_clusters = []

for word in keys:

    words = []
```

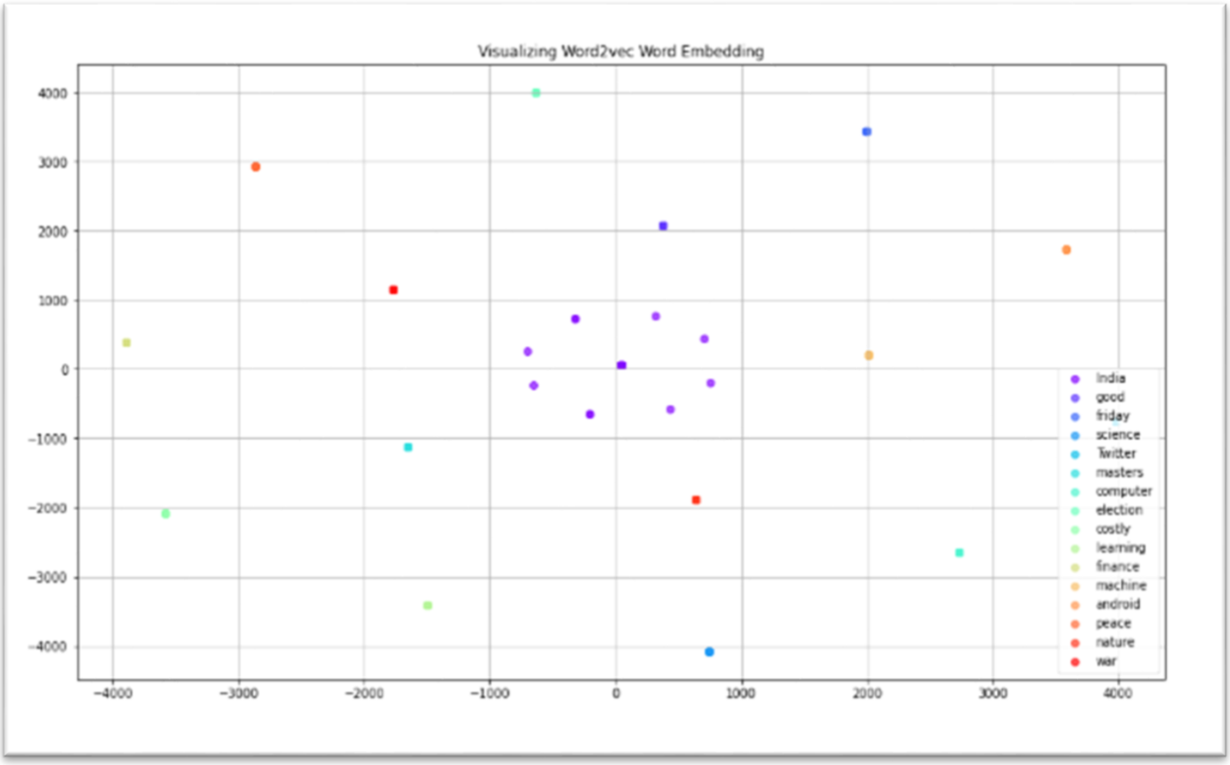
```
    embeddings = []

    for similar_word, _ in Word2VecModel.most_similar(word, to
pn = 30):
        words.append(similar_word)
        embeddings.append(Word2VecModel[word])
    words_clusters.append(words)
    embeddings_clusters.append(embeddings)
```

```
from sklearn.manifold import TSNE

embedding_array = np.array(embeddings_clusters)
n, m, k = embedding_array.shape

tsne_2d_model = TSNE(perplexity = 15, n_components = 2, n_iter
= 4000, random_state = 11, init = 'pca')
tsne_embeddings = np.array(tsne_2d_model.fit_transform(embeddi
ng_array.reshape(n * m, k))).reshape(n, m, 2)
```

```
import matplotlib.pyplot as plt
import matplotlib.cm as cm
%matplotlib inline

def plot_most_similar_words(labels, embedding_cluster, word_cl
uster, title):

    colors = cm.rainbow(np.linspace(0, 1, len(labels)))
    plt.figure(figsize = (16,9))
    for label, embeddings, words, color in zip(labels, embeddi
ng_cluster, word_cluster, colors):
        x = embeddings[:, 0]
        y = embeddings[:, 1]
        plt.scatter(x, y, c=color, alpha=0.7, label=label)
    plt.legend(loc = 4)
    plt.title(title)
    plt.grid(True)
```

```
    plt.show()
```

```
linkcode
plot_most_similar_words(keys, tsne_embeddings, words_clusters, "Visualizing Word2vec Word Embedding")
```

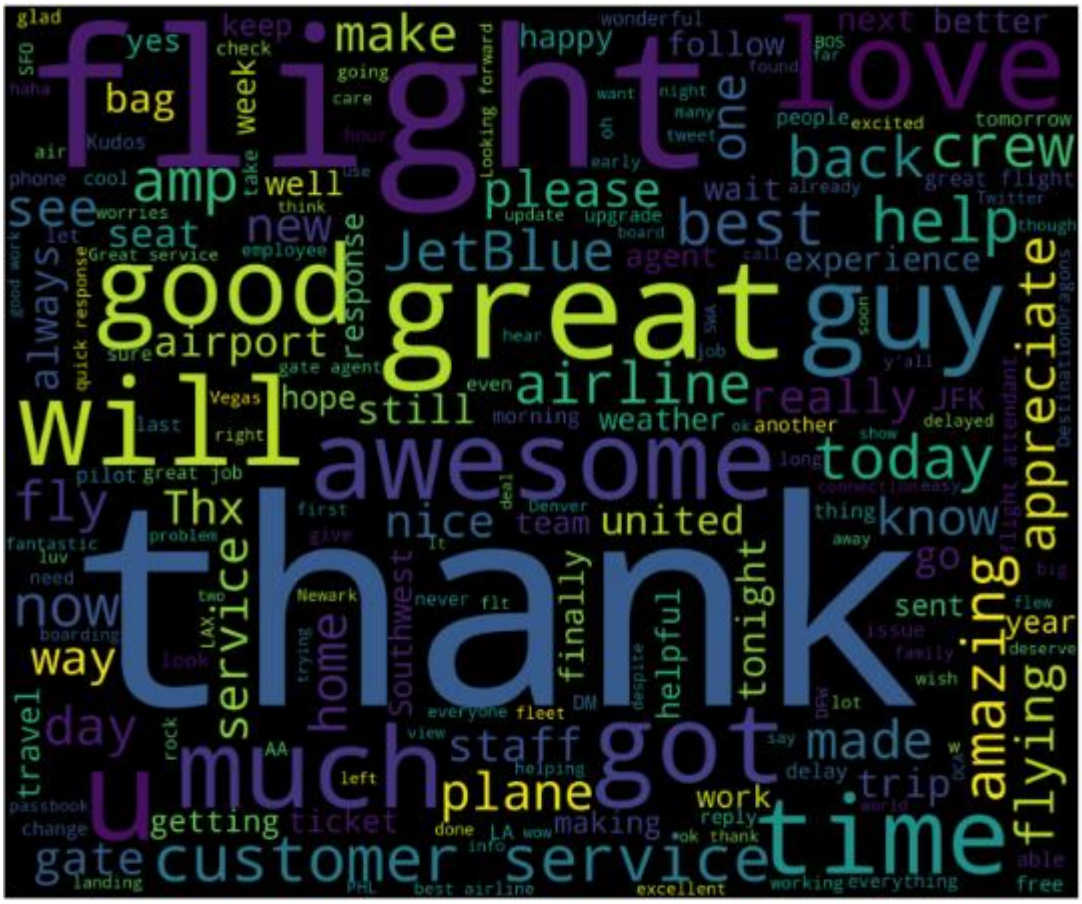


Wordcloud for positive reasons

PROGRAM:

```
from wordcloud import WordCloud,STOPWORDS
new_df=data[data['airline_sentiment']=='positive']
words = ' '.join(new_df['text'])
cleaned_word = " ".join([word for word in words.split()
                            if 'http' not in word
                                and not word.startswith('@')
                                and word != 'RT'
                            ])
wordcloud = WordCloud(stopwords=STOPWORDS,
                      background_color='black',
                      width=3000,
                      height=2500
                     ).generate(cleaned_word)
plt.figure(1,figsize=(12, 12))
plt.imshow(wordcloud)
```

```
plt.axis('off')
plt.show()
```

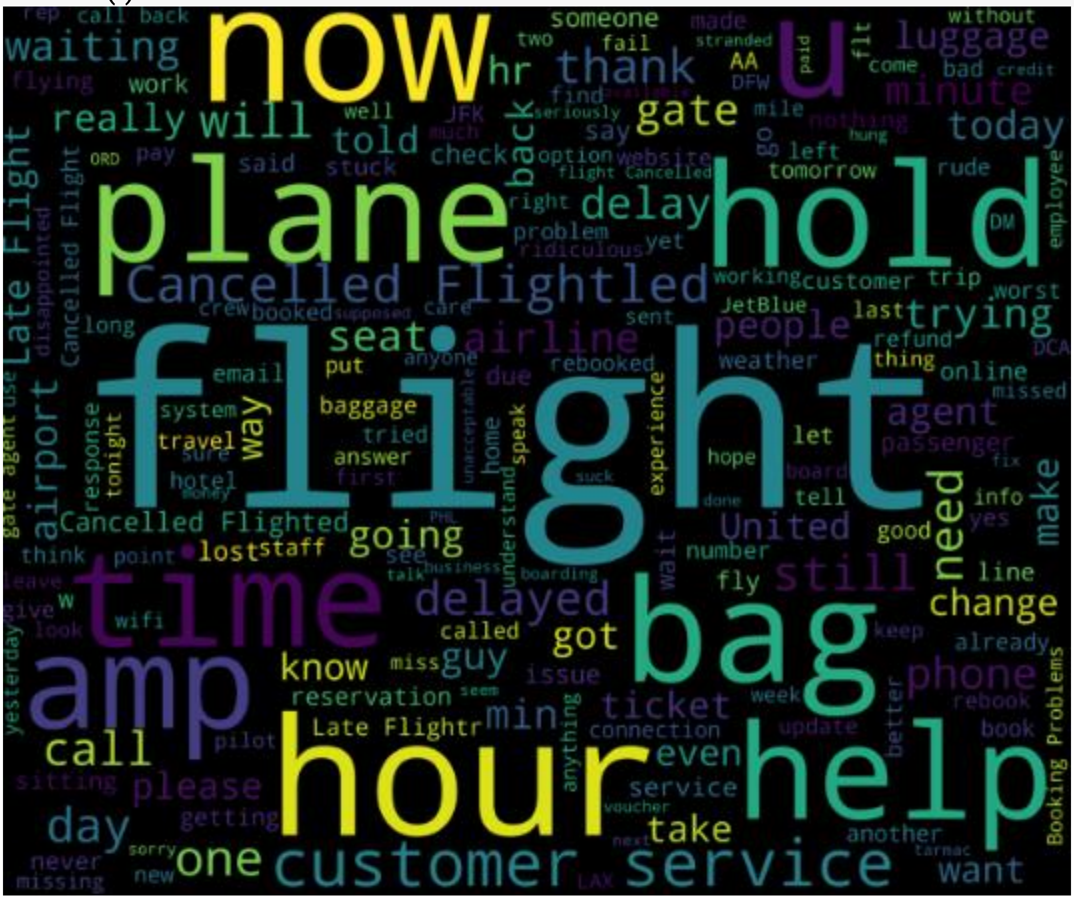


Wordcloud for Negative sentiments of tweets

PROGRAM:

```
new_df=data[data['airline_sentiment']=='negative']
words = ' '.join(new_df['text'])
cleaned_word = " ".join([word for word in words.split()
                            if 'http' not in word
                                and not word.startswith('@')
                                and word != 'RT'
                            ])
wordcloud = WordCloud(stopwords=STOPWORDS,
                      background_color='black',
                      width=3000,
                      height=2500
                     ).generate(cleaned_word)
```

```python
plt.figure(1,figsize=(12, 12))
plt.imshow(wordcloud)
plt.axis('off')
plt.show()
```



MODEL SELECTION:

Choose an appropriate machine learning model for your regression task. Common choices include:

- Linear Regression
- Decision Trees
- Random Forest
- Gradient Boosting (e.g., XGBoost or LightGBM)
- Neural Networks (Deep Learning)

PROGRAM:

```
cls = [LogisticRegression(),
       MultinomialNB(),
       DecisionTreeClassifier(),
       RandomForestClassifier(n_estimators=200),
       KNeighborsClassifier(n_neighbors = 5)]

cls_name = []
```

```
lbl_actual = test_df.airline_sentiment
i = 0
accuracy = []
for cl in cls:
    model = cl.fit(train_tfidf_model,train_df.airline_sentimen
t)
    lbl_pred = model.predict(test_tfidf_model)
    a = (100*accuracy_score(lbl_pred, lbl_actual))
    a = round(a,2)
    accuracy.append(a)
    cls_name.append(cl.__class__.__name__)
    print ("{}  Accuracy Score : {}%".format(cls_name[i],a))
    print ( classification_report(lbl_pred, lbl_actual))
    i +=1
```

OUTPUT:

```
LogisticRegression  Accuracy Score : 79.1%
              precision    recall  f1-score   support

    negative       0.93      0.81      0.87      3232
     neutral       0.48      0.66      0.56       648
    positive       0.60      0.81      0.69       512

    accuracy                           0.79      4392
   macro avg       0.67      0.76      0.71      4392
weighted avg       0.83      0.79      0.80      4392

MultinomialNB  Accuracy Score : 69.69%
              precision    recall  f1-score   support

    negative       0.99      0.69      0.81      4081
```

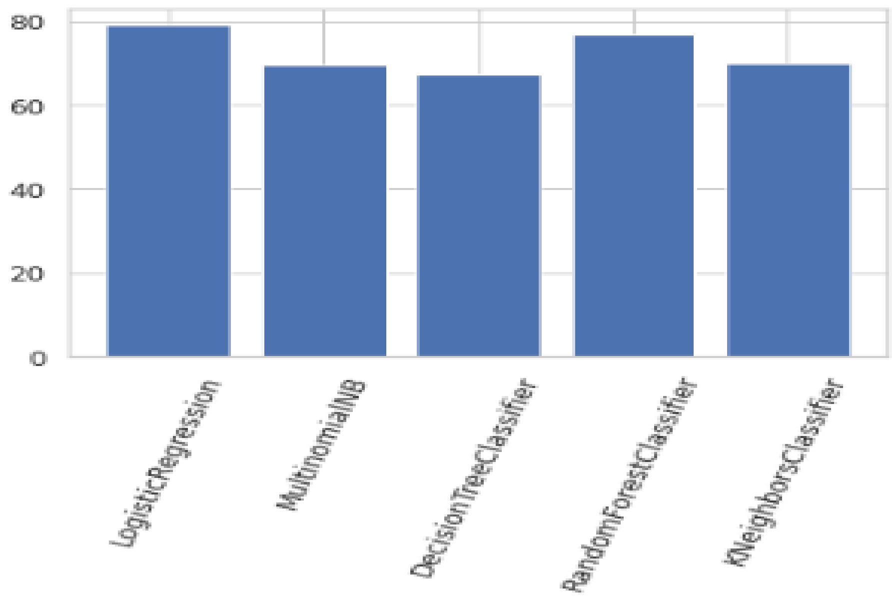
```
     neutral       0.15      0.78      0.26       174
    positive       0.18      0.93      0.31       137

    accuracy                           0.70      4392
   macro avg       0.44      0.80      0.46      4392
weighted avg       0.94      0.70      0.77      4392

DecisionTreeClassifier  Accuracy Score : 67.42%
              precision    recall  f1-score   support

    negative       0.79      0.78      0.79      2841
     neutral       0.40      0.41      0.40       879
    positive       0.55      0.57      0.56       672

    accuracy                           0.67      4392
   macro avg       0.58      0.58      0.58      4392
weighted avg       0.68      0.67      0.67      4392

RandomForestClassifier  Accuracy Score : 76.78%
              precision    recall  f1-score   support

    negative       0.94      0.79      0.86      3378
     neutral       0.38      0.63      0.48       535
    positive       0.54      0.79      0.64       479

    accuracy                           0.77      4392
   macro avg       0.62      0.74      0.66      4392
weighted avg       0.83      0.77      0.79      4392

KNeighborsClassifier  Accuracy Score : 69.83%
              precision    recall  f1-score   support

    negative       0.82      0.80      0.81      2866
     neutral       0.46      0.42      0.44       960
    positive       0.53      0.65      0.58       566

    accuracy                           0.70      4392
   macro avg       0.60      0.62      0.61      4392
weighted avg       0.70      0.70      0.70      4392
```

PROGRAM:

```
plt.bar(cls_name, accuracy)
plt.xticks(rotation=70)
```

OUTPUT:

```
([0, 1, 2, 3, 4], <a list of 5 Text major ticklabel objects>)
```




```
# Save to csv

lg_model = LogisticRegression().fit(train_tfidf_model,train_df
.airline_sentiment)
lg_lbl_pred = model.predict(test_tfidf_model)
```

In [43]:

```
linkcode
lg_lbl_pred_df = pd.DataFrame({'tweet_id': test_df.tweet_id,
                              'text' : test_df.text,
                              'lg_reg' : lg_lbl_pred})
lg_lbl_pred_df.head()
```

OUTPUT:

	tweet_id	text	lg_reg
4794	569731104070115329	@SouthwestAir you're my early frontrunner for ...	positive
10480	569263373092823040	@USAirways how is it that my flt to EWR was Ca...	negative
8067	568818669024907264	@JetBlue what is going on with your BDL to DCA...	negative
8880	567775864679456768	@JetBlue do they have to depart from Washingto...	negative
8292	568526521910079488	@JetBlue I can probably find some of them. Are...	neutral

FEATURE ENGINEERING:

Feature selection is a critical step in sentiment analysis for US airlines marketing. The choice of features affects the accuracy and performance of sentiment analysis models. Here are common features and considerations for feature selection in this context:

1. Text Features:

Bag of Words (BoW):

Represents text as a collection of unique words, often with term frequency (TF) or term frequency-inverse document frequency (TF-IDF) weighting.

Bag of words (BOW) feature extraction

```
from sklearn.feature_extraction.text import CountVectorizer
#from sklearn.feature_extraction.text import TfidfVectorizer

vocabulary_size = 5000

# Tweets have already been preprocessed hence dummy function wil
l be passed in
# to preprocessor & tokenizer step
count_vector = CountVectorizer(max_features=vocabulary_size,
#                               ngram_range=(1,2),    # unigram
and bigram
                                preprocessor=lambda x: x,
                               tokenizer=lambda x: x)
#tfidf_vector = TfidfVectorizer(lowercase=True, stop_words='engl
ish')

# Fit the training data
X_train = count_vector.fit_transform(X_train).toarray()

# Transform testing data
X_test = count_vector.transform(X_test).toarray()
```

```
#import sklearn.preprocessing as pr

# Normalize BoW features in training and test set
#X_train = pr.normalize(X_train, axis=1)
#X_test = pr.normalize(X_test, axis=1)
```

linkcode

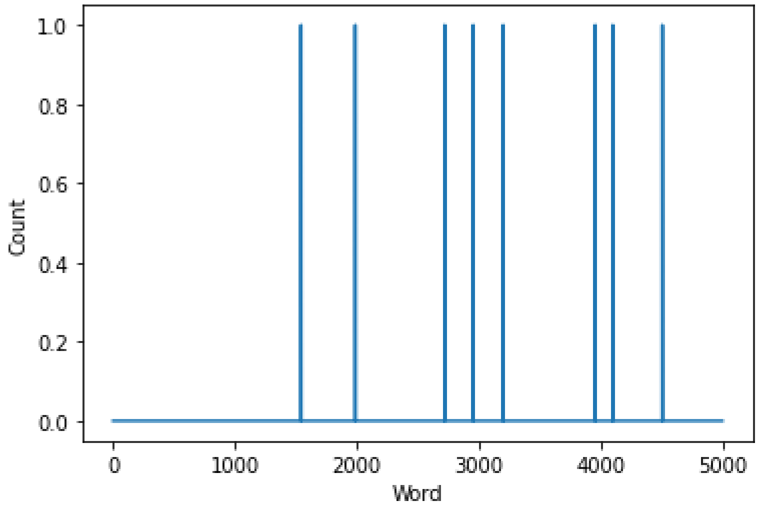
```
# print first 200 words/tokens
print(count_vector.get_feature_names()[0:200])
```

OUTPUT:

```
['0', '000', '1', '10', '100', '1000', '10000', '10th', '11', '111', '1
145', '11th', '12', '1200', '12000', '125', '12k', '130', '140', '14000
', '15', '150', '1500', '1500000', '1520', '157200000', '15l', '15lac',
'15lakh', '18', '180', '19', '1947', '1958', '1962', '1969', '1971', '1
980', '1984', '1998', '1st', '2', '20', '200', '2000', '2002', '2004',
'2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015',
'2016', '2017', '2018', '2019', '2020', '2022', '2024', '2029', '21st',
'23', '23rd', '24', '247', '24x7', '25', '250', '2611', '26th', '272',
'27th', '280319', '282', '28th', '2cr', '2day', '2nd', '3', '30', '300'
, '3000', '30000', '300km', '31st', '350', '35a', '370', '3rd', '4', '4
0', '400', '4000', '45', '456', '4th', '5', '50', '500', '5000', '50000
', '5th', '5year', '5yr', '6', '60', '600', '6000', '6th', '7', '72000'
, '7200000', '72k', '7th', '8', '80', '800', '8020', '9', '90', '9000',
'aa', 'aadhaar', 'aadhar', 'aadmi', 'aag', 'aaj', 'aalo', 'aam', 'aan',
'aap', 'aapk', 'aapko', 'aapl', 'aapn', 'aay', 'aaya', 'aayega', 'aayog
', 'abandon', 'abdul', 'abdullah', 'abe', 'abhi', 'abhinandan', 'abhisa
r', 'abhiyan', 'abil', 'abki', 'abl', 'abolish', 'abroad', 'abscond', '
absolut', 'absurd', 'abt', 'abus', 'abv', 'academ', 'acc', 'accept', 'a
ccess', 'acch', 'accha', 'accid', 'accident', 'accommod', 'accompani',
'accomplish', 'accord', 'accordingli', 'account', 'accumul', 'accur', '
accus', 'ach', 'acha', 'acheiv', 'achh', 'achiev', 'achiv', 'acknowledg
', 'acquir', 'acquit', 'acronym', 'across', 'act', 'action', 'activ', '
activist', 'actor', 'actress', 'actual', 'ad', 'adani', 'add']
```

```
# Plot the BoW feature vector
plt.plot(X_train[2,:])
plt.xlabel('Word')
plt.ylabel('Count')
plt.show()
```

OUTPUT:



Word Embeddings:

Utilizes pre-trained word vectors like Word2Vec or GloVe to capture word semantics and context.

N-grams:

Includes sequences of words (bigrams, trigrams) to capture context and phrases.

Stop Words:

Deciding whether to include or exclude common stop words depends on the specific analysis.

2. Lexicon-Based Features:

Sentiment Lexicons:

Integrates sentiment lexicons such as the AFINN lexicon or SentiWordNet to score individual words for sentiment.

Emotion Lexicons:

Utilizes emotion lexicons to extract emotional features from text.

Negation Handling:

Identifies negations in text and reverses the sentiment of subsequent words.

3. Part-of-Speech (POS) Features:

POS Tagging:

Extracts POS information to understand the role of words in a sentence and how they influence sentiment.

4. Syntactic Features:

Dependency Parsing:

Analyzes grammatical relationships between words to understand sentence structure.

Parse Tree Features:

Incorporates features based on parse trees to capture linguistic structure.

5. Semantic Features:

Named Entity Recognition (NER):

Identifies named entities (e.g., airline names, locations) in text and considers their impact on sentiment.

Word Sense Disambiguation:

Resolves word sense ambiguities to improve sentiment analysis accuracy.

6. Domain-Specific Features:

Airlines-Related Keywords:

Includes domain-specific terms and airline-specific keywords that may influence sentiment.

Aspect-Based Features:

Analyzes aspects such as flight experience, customer service, and pricing separately to understand sentiment at a granular level.

7. Emoticons and Emoji Features:

Emoticon Analysis:

Considers the presence and sentiment of emoticons in text.

Emoji Analysis:

Analyzes the sentiment conveyed by emojis, which are commonly used in social media posts.

8. Social Media Features:

Hashtags:

Considers the sentiment associated with hashtags in social media posts.

Mentions and Retweets:

Incorporates engagement metrics to gauge the impact of a post on sentiment.

9. Text Length and Structure:

Text Length:

Evaluates sentiment in relation to the length of the text (short vs. long).

Sentence Structure:

Analyzes sentiment in complex or compound sentences compared to simple ones.

10. Sentiment Scores:

Overall Sentiment Scores:

Calculates the overall sentiment score for a document or text.

Sentence-Level Sentiment:

Analyzes sentiment at the sentence level to understand variations within a document.

11. Sentiment Shift Detection:

Sentiment Transitions:

Identifies shifts in sentiment within a document or conversation.

12.User Profile Features:

User Information:

Considers user profiles, previous sentiments, and influence of users in social media sentiment analysis.

13.Temporal Features:

Time-Based Analysis:

Incorporates temporal features to understand how sentiment changes over time.

Tokenizing & Padding

```
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences

max_words = 5000
max_len=50

def tokenize_pad_sequences(text):
    '''
    This function tokenize the input text into sequnences of int
ergers and then
    pad each sequence to the same length
    '''
    # Text tokenization
    tokenizer = Tokenizer(num_words=max_words, lower=True, spl
it=' ')
    tokenizer.fit_on_texts(text)
    # Transforms text to a sequence of integers
    X = tokenizer.texts_to_sequences(text)
    # Pad sequences to the same length
    X = pad_sequences(X, padding='post', maxlen=max_len)
    # return sequences
    return X, tokenizer

print('Before Tokenization & Padding \n', df['clean_text'][0])
X, tokenizer = tokenize_pad_sequences(df['clean_text'])
print('After Tokenization & Padding \n', X[0])
```

OUTPUT:

```
Before Tokenization & Padding
 when modi promised “minimum government maximum governance” ex
pected him begin the difficult job reforming the state why doe
s take years get justice state should and not business and sho
uld exit psus and temples
After Tokenization & Padding
```

```
 [  41    1  349   73 1911 1180   44 2465    2 1259  219    2
236   32
  165  102   53   55 1184  236   50    3    6  533    3   50 3
833    3
 3077    0    0    0    0    0    0    0    0    0    0    0
0    0
    0    0    0    0    0    0    0    0]
```

Saving tokenized data

```
import pickle

# saving
with open('tokenizer.pickle', 'wb') as handle:
    pickle.dump(tokenizer, handle, protocol=pickle.HIGHEST_PRO
TOCOL)

# loading
with open('tokenizer.pickle', 'rb') as handle:
    tokenizer = pickle.load(handle)
```

Train & Test Split

```
y = pd.get_dummies(df['category'])
X_train, X_test, y_train, y_test = train_test_split(X, y, test
_size=0.2, random_state=1)
X_train, X_val, y_train, y_val = train_test_split(X_train, y_t
rain, test_size=0.25, random_state=1)
print('Train Set ->', X_train.shape, y_train.shape)
print('Validation Set ->', X_val.shape, y_val.shape)
print('Test Set ->', X_test.shape, y_test.shape)
```

OUTPUT:

```
Train Set -> (109397, 50) (109397, 3)
Validation Set -> (36466, 50) (36466, 3)
Test Set -> (36466, 50) (36466, 3)
```

MODEL EVALUATION:

Evaluating a sentiment analysis model for US airlines marketing is crucial to ensure its accuracy and effectiveness. Here are some common evaluation metrics and methods for assessing the performance of a sentiment analysis model:

1. Confusion Matrix:

Calculate true positives, true negatives, false positives, and false negatives to understand how well the model classifies sentiments.

2. Accuracy:

Measure the overall correctness of sentiment predictions by dividing the sum of true positives and true negatives by the total number of predictions.

3. Precision:

Assess the proportion of true positive predictions among all positive predictions, helping to determine how well the model identifies positive sentiments.

4. Recall (Sensitivity):

Calculate the ratio of true positives to the total number of actual positive instances, indicating the model's ability to capture positive sentiments.

5. F1-Score:

The harmonic mean of precision and recall, providing a balanced evaluation metric that considers both false positives and false negatives.

Model Accuracy & Loss

```
# Evaluate model on the test set
loss, accuracy, precision, recall = model.evaluate(X_test, y_t
est, verbose=0)
# Print metrics
print('')
print('Accuracy  : {:.4f}'.format(accuracy))
print('Precision : {:.4f}'.format(precision))
print('Recall    : {:.4f}'.format(recall))
print('F1 Score  : {:.4f}'.format(f1_score(precision, recall))
)
```

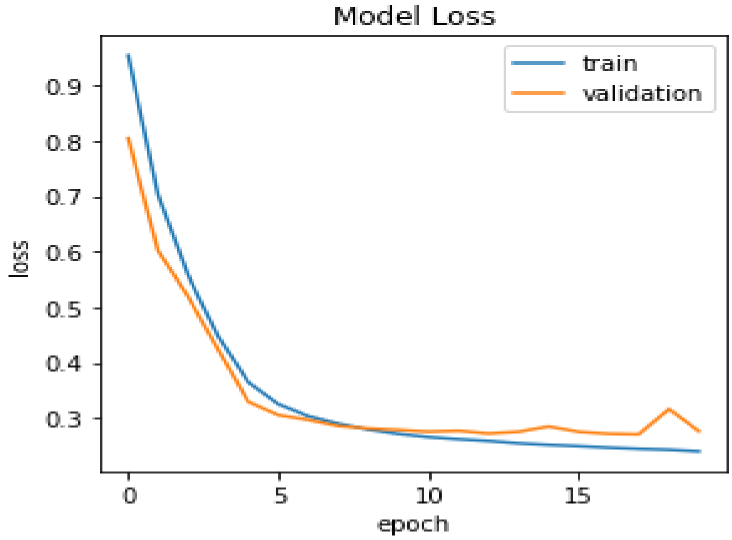
OUTPUT:

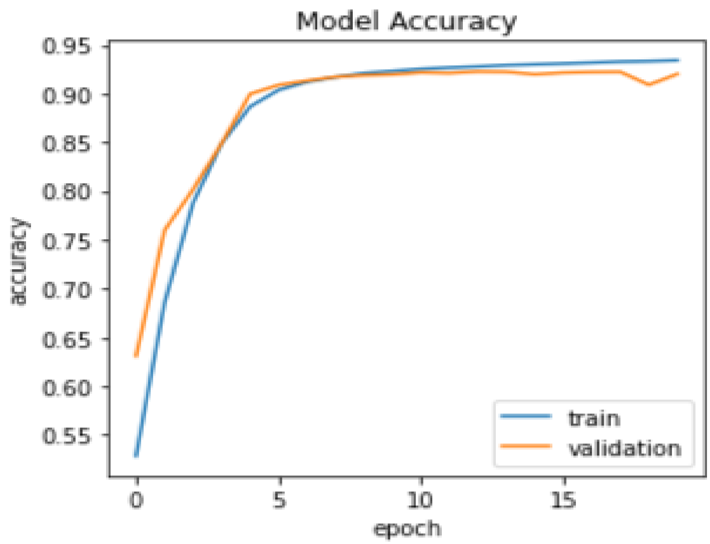
```
Accuracy  : 0.9152
Precision : 0.9175
Recall    : 0.9127
F1 Score  : 0.9151
```

PROGRAM:

```
def plot_training_hist(history):
    '''Function to plot history for accuracy and loss'''

    fig, ax = plt.subplots(1, 2, figsize=(10,4))
    # first plot
    ax[0].plot(history.history['accuracy'])
    ax[0].plot(history.history['val_accuracy'])
    ax[0].set_title('Model Accuracy')
    ax[0].set_xlabel('epoch')
    ax[0].set_ylabel('accuracy')
    ax[0].legend(['train', 'validation'], loc='best')
    # second plot
    ax[1].plot(history.history['loss'])
    ax[1].plot(history.history['val_loss'])
    ax[1].set_title('Model Loss')
    ax[1].set_xlabel('epoch')
    ax[1].set_ylabel('loss')
    ax[1].legend(['train', 'validation'], loc='best')

plot_training_hist(history)
```

OUTPUT:



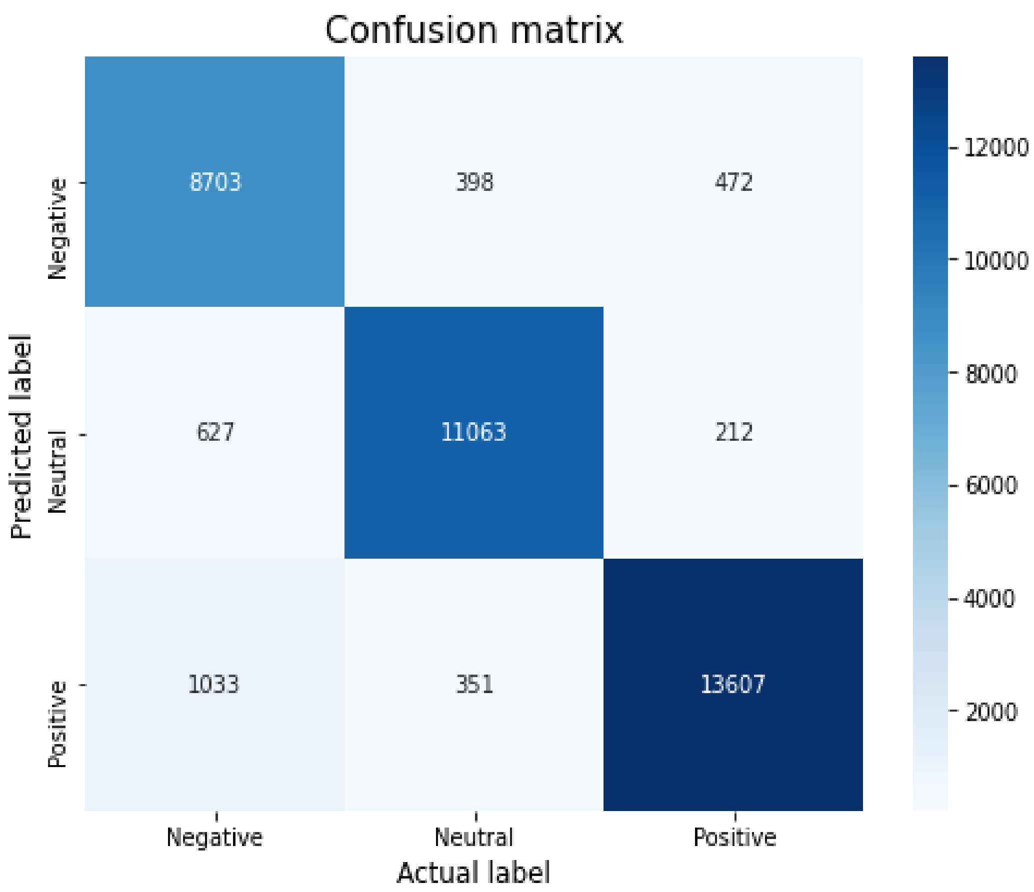


Model Confusion Matrix

```
from sklearn.metrics import confusion_matrix

def plot_confusion_matrix(model, X_test, y_test):
    '''Function to plot confusion matrix for the passed model a
nd the data'''

    sentiment_classes = ['Negative', 'Neutral', 'Positive']
    # use model to do the prediction
    y_pred = model.predict(X_test)
    # compute confusion matrix
```

```python
    cm = confusion_matrix(np.argmax(np.array(y_test),axis=1), np.argmax(y_pred, axis=1))
    # plot confusion matrix
    plt.figure(figsize=(8,6))
    sns.heatmap(cm, cmap=plt.cm.Blues, annot=True, fmt='d',
                xticklabels=sentiment_classes,
                yticklabels=sentiment_classes)
    plt.title('Confusion matrix', fontsize=16)
    plt.xlabel('Actual label', fontsize=12)
    plt.ylabel('Predicted label', fontsize=12)

plot_confusion_matrix(model, X_test, y_test)
```

OUTPUT:



ADVANTAGES

Sentiment analysis offers several advantages for US airlines marketing:

1.Customer Insights:

It provides valuable insights into customer opinions, emotions, and preferences, helping airlines understand what passengers like and dislike about their services.

2.Brand Monitoring:

Airlines can monitor their brand reputation in real time and address negative sentiment promptly, preventing potential PR crises.

3.Competitor Benchmarking:

Sentiment analysis allows airlines to compare their performance with competitors, identifying areas where they excel or need improvement.

4.Customer Feedback Utilization:

Airlines can use sentiment analysis to harness customer feedback from various sources, such as social media, reviews, and surveys, to improve their services.

5.Personalization:

Understanding passenger sentiments enables airlines to offer personalized services, tailored offers, and targeted marketing campaigns, increasing customer satisfaction.

6.Operational Improvements:

Airlines can identify operational issues and bottlenecks that affect passenger sentiment, leading to process enhancements and cost savings.

7.Crisis Management:

By monitoring sentiment in real time, airlines can quickly respond to negative events or issues, minimizing the impact on their brand.

8.Product Development:

Sentiment analysis helps airlines identify opportunities for new services or product enhancements based on customer demand.

9.Enhanced Marketing Strategies:

Airlines can create more effective marketing campaigns by aligning their messaging with customer sentiment and preferences.

10.Predictive Insights:

Sentiment trends can be used to make predictions about future customer behavior and sentiment shifts, allowing proactive actions.

11.Customer Retention:

Addressing negative sentiment and providing excellent customer service can improve customer loyalty and retention rates.

12.Improved Crisis Preparedness:

Airlines can develop proactive strategies and crisis management plans based on past sentiment data, reducing the impact of potential crises.

13.Cost Reduction:

By identifying and addressing issues that lead to customer dissatisfaction, airlines can reduce customer service costs associated with handling complaints.

14.Enhanced Public Relations:

Airlines can build a positive public image by addressing issues highlighted in sentiment analysis, demonstrating their commitment to customer satisfaction.

15.Feedback Loop:

Sentiment analysis creates a feedback loop that allows airlines to continuously learn from customer opinions, make data-driven decisions, and iterate on their services.

16.Efficient Resource Allocation:

Airlines can allocate resources more efficiently by focusing on areas that have the greatest impact on customer sentiment.

17.Customer-Centric Culture:

Promoting a customer-centric culture within the organization by constantly emphasizing the importance of customer feedback and sentiment.

18.Sustainability Initiatives:

Airlines can gauge customer sentiment regarding environmental efforts and make data-driven decisions to enhance their sustainability programs.

19.Proactive Marketing:

By detecting positive sentiment, airlines can amplify positive feedback, leveraging it for marketing purposes and building customer advocacy.

DISADVANTAGES:

While sentiment analysis provides many benefits, it also has some drawbacks when applied to US airlines marketing:

1.Ambiguity and Context:

Sentiment analysis may struggle with understanding context, humor, sarcasm, and sarcasm, leading to misinterpretations of sentiment.

2.Negation and Contrast:

It may misinterpret sentiment when negation or contrasting phrases are used, leading to inaccurate results.

3.Language Variability:

Different dialects and slang can challenge sentiment analysis, causing it to misclassify sentiments, particularly in a diverse country like the United States.

4.Emotion Complexity:

Human emotions are multifaceted, and sentiment analysis might oversimplify them by categorizing them as merely positive, negative, or neutral.

5.Data Bias:

Sentiment analysis can be influenced by biased or unrepresentative data, which can lead to skewed results.

6.Data Source Limitations:

The analysis heavily relies on available data sources, and if certain channels or sources are not included, it may not provide a comprehensive view of sentiment.

7.Imbalanced Data:

Some sentiments may be more prevalent than others, leading to imbalanced datasets and potentially skewing the analysis.

8.Subjectivity:

Sentiment analysis tools often rely on predefined sentiment lexicons, which can be subjective and may not accurately reflect the sentiment of certain phrases or words.

9.Sensitivity to Data Noise:

Noisy data, such as misspellings, typos, or irrelevant information, can reduce the accuracy of sentiment analysis.

10.Human vs. Machine Error:

Even advanced sentiment analysis models can make errors, and human reviewers may need to validate results, introducing potential human bias.

11.Privacy Concerns:

Handling customer data for sentiment analysis requires ensuring data privacy and security compliance, which can be challenging.

12.Real-time Challenges:

Real-time sentiment analysis may not always be immediate, and delays can impact an airline's ability to respond promptly to issues.

13.Cost:

Developing and maintaining sentiment analysis tools and systems can be expensive, and smaller airlines may face budget constraints.

14.Over-Reliance on Data:

Relying too heavily on sentiment analysis can lead to neglecting other aspects of customer feedback and market research.

15.Inability to Understand Unstructured Data:

Sentiment analysis may not effectively handle unstructured data, such as qualitative feedback, which can provide valuable insights.

16.No Human Nuance:

It lacks the human nuance and emotional intelligence required to fully understand and empathize with customers.

17.Inconsistent Terminology:

Customers may use different terminology to express similar sentiments, making it challenging to recognize patterns.

18.Misclassification:

Sentiment analysis tools may misclassify sentiments, leading to incorrect conclusions.

CONCLUSION:

Enhancing US Airlines Marketing with Sentiment Analysis

Sentiment analysis has become an indispensable tool for US airlines marketing, offering a powerful means to understand, engage with, and adapt to customer sentiment. In an industry where customer satisfaction is paramount and competition is fierce, sentiment analysis serves as a strategic ally, providing valuable insights and shaping marketing strategies to enhance passenger experiences and brand reputation.

Real-time Insights: By monitoring sentiment in real time, airlines can detect and respond to emerging issues swiftly. This proactive approach not only prevents potential crises but also showcases the airline's commitment to passenger satisfaction.

Personalization:

Sentiment analysis empowers airlines to personalize services, offers, and communication, aligning them with passenger preferences and emotions. This not only drives customer satisfaction but also fosters brand loyalty.

Operational Efficiency:

Operational improvements driven by sentiment analysis reduce costs and enhance efficiency. Understanding issues affecting passenger sentiment allows airlines to allocate resources effectively and streamline processes.

Crisis Management:

In an industry susceptible to sudden crises, sentiment analysis provides a vital tool for managing and mitigating issues that could tarnish a brand's reputation. Airlines can respond swiftly, minimizing damage and maintaining trust.

Data-Driven Decision Making:

The data-driven approach enabled by sentiment analysis ensures that marketing decisions are grounded in customer feedback and sentiment. This leads to more effective strategies and resource allocation.

Brand Reputation Management:

Proactively addressing issues identified through sentiment analysis bolsters brand reputation. By demonstrating a commitment to customer satisfaction, airlines can protect and enhance their image.

Customer Loyalty:

Addressing negative sentiment and providing excellent customer service drives customer loyalty and retention. In a highly competitive market, this is a critical advantage.

Feedback Loop:

Sentiment analysis creates a feedback loop that enables airlines to continually learn from customer opinions, make data-driven decisions, and iteratively enhance their services.

In the dynamic and customer-centric world of US airlines marketing, sentiment analysis is not just a tool; it is a catalyst for growth, resilience, and lasting positive sentiment. As airlines continue to embrace the insights provided by sentiment analysis, they position themselves to thrive in a competitive industry, delivering exceptional experiences and building strong brand reputations.