PHASE 3 PROJECT SUBMISSION

PROJECT TITLE: SENTIMENT ANALYSIS FOR MARKETING

DEVELOPMENT PART 1

INTRODUCTION:

Sentiment analysis plays a pivotal role in the marketing strategies of US airlines. It involves the systematic collection, analysis, and interpretation of customer sentiments and feedback expressed in various channels, including social media, customer reviews, surveys, and customer support interactions. This analysis is instrumental in understanding the perceptions, emotions, and preferences of passengers, enabling airlines to make data-driven decisions and enhance their marketing efforts.

The importance of sentiment analysis in the US airline industry cannot be overstated. In a highly competitive and customer-centric sector, airlines are constantly seeking ways to improve the passenger experience, address issues promptly, and identify trends that impact brand perception and loyalty. Sentiment analysis empowers these airlines to achieve these objectives by providing valuable insights into customer sentiments.

Key aspects include:

1. Social Media Monitoring: Airlines actively track social media platforms to gauge real-time customer sentiment. Analyzing posts, tweets, and comments allows them to promptly respond to customer concerns and showcase positive experiences.

2. Review Analysis: Customer reviews on platforms like TripAdvisor and Yelp provide rich data for sentiment analysis. Airlines can understand trends, identify areas for improvement, and leverage positive feedback in their marketing campaigns.

3. Customer Satisfaction Surveys: Sentiment analysis aids in evaluating customer satisfaction survey responses. By identifying common pain points or areas of delight, airlines can refine their services.

4. Competitive Benchmarking: Analyzing sentiment also involves assessing competitors' performance. Airlines use these insights to differentiate themselves and refine their marketing strategies.

5. Crisis Management: In the face of crises, sentiment analysis helps airlines address concerns, manage their reputation, and guide their marketing communications.

6. Customized Marketing Campaigns: Sentiment analysis enables airlines to create marketing campaigns tailored to customer preferences, addressing their pain points and highlighting the aspects that matter most to passengers.

7. Predictive Analytics: By analyzing historical data and customer sentiment, airlines can make data-driven predictions and adapt marketing strategies accordingly.

GIVEN DATA SET:

tweet_id	airline_se	airline_se	negativere	negativere	airline	airline_se	name	negativere	retweet_c	text	tweet_coord	tweet_created	tweet_location	user_timezone
5.7E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica What @dhepburn said.		24-02-2015 11:35		Eastern Time (US & Canada)
5.7E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plus you've added commercials to the experienc		24-02-2015 11:15		Pacific Time (US & Canada)
5.7E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I didn't today... Must mean I need to take anothe		24-02-2015 11:15	Lets Play	Central Time (US & Canada)
5.7E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's really aggressive to blast obnoxious "enterta		24-02-2015 11:15		Pacific Time (US & Canada)
5.7E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and it's a really big bad thing about it		24-02-2015 11:14		Pacific Time (US & Canada)
5.7E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@VirginAmerica seriously would pay $30 a		24-02-2015 11:14		Pacific Time (US & Canada)
5.7E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes, nearly every time I fly VX this â€œear worm		24-02-2015 11:13	San Francisco CA	Pacific Time (US & Canada)
5.7E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Really missed a prime opportunity for Men With		24-02-2015 11:12	Los Angeles	Pacific Time (US & Canada)
5.7E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Well, I didn'tâ€¦but NOW I DO! :-D		24-02-2015 11:11	San Diego	Pacific Time (US & Canada)
5.7E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it was amazing, and arrived an hour early. You'r		24-02-2015 10:53	Los Angeles	Eastern Time (US & Canada)
5.7E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did you know that suicide is the second leading		24-02-2015 10:48	1/1 loner squad	Eastern Time (US & Canada)
5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I <3 pretty graphics. so much better than mini		24-02-2015 10:30	NYC	America/New_York
5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica This is such a great deal! Already thinking about		24-02-2015 10:30	NYC	America/New_York
5.7E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @virginmedia I'm flying your #fabulous #Seducti		24-02-2015 10:21		Eastern Time (US & Canada)
5.7E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Thanks!		24-02-2015 10:15	San Francisco, CA	Pacific Time (US & Canada)
5.7E+17	negative	0.6842	Late Flight	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFO-PDX schedule is still MIA.		24-02-2015 10:01	palo alto, ca	Pacific Time (US & Canada)
5.7E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So excited for my first cross country flight LAX to		24-02-2015 09:42	west covina	Pacific Time (US & Canada)
5.7E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I flew from NYC to SFO last week and couldn't fu		24-02-2015 09:39	this place called N	Eastern Time (US & Canada)
5.7E+17	positive	1			Virgin America		thebrandiray		0	I â¤ï¸ flying @VirginAmerica. â˜ºï¸ðŸ'		24-02-2015 09:15	Somewhere celebr	Atlantic Time (Canada)
5.7E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you know what would be amazingly awesome? B		24-02-2015 09:04	Boston | Waltham	Quito
5.7E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica why are your first fares in May over three times		24-02-2015 08:55		
5.7E+17	positive	1			Virgin America		DT_Les		0	@VirginAmerica I love this graphic. http://t.	[40.74804263, -73.99	24-02-2015 08:49		
5.7E+17	positive	1			Virgin America		ElvinaBeck		0	@VirginAmerica I love the hipster innovation. You are a feel good		24-02-2015 08:30	Los Angeles	Pacific Time (US & Canada)

14641 Rows*17 Columns

NECESSARY STEPS TO FOLLOW:

To start building sentiment analysis solution for airline marketing, we need to load and preprocess the dataset from Kaggle.

1. Loading the Dataset:

Retrieve the dataset from Kaggle and load it into a Pandas Data Frame for further analysis.

PROGRAM

```
import pandas as pd

# Load the dataset

data = pd.read csv('Tweets.csv')
```

2. Exploratory Data Analysis (EDA):

Analyze the dataset to gain insights, such as the distribution of sentiment labels and key statistics.

PROGRAM

```
 # Perform exploratory data analysis

 print(data.head())

 print(data['airline_sentiment'].value_counts())
```

3. Text Preprocessing:

Clean and prepare the text data, including tasks like removing special characters, lowercasing, and handling missing values.

PROGRAM

```
# Text preprocessing

import re

def clean_text(text):

        # Remove special characters and digits

        text = re.sub(r'[^a-zA-Z\s]', '', text)
```

```
        # Convert to lowercase
        text = text.lower()
        return text
data['text'] = data['text'].apply(clean_text)
```

4. Tokenization and Stopword Removal:

Tokenize the text data into individual words or tokens, and remove common stopwords.

PROGRAM

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
def tokenize_and_remove_stopwords(text):
        words = word_tokenize(text)
        words = [word for word in words if word not in
stopwords.words('english')]
        return words
data['text'] = data['text'].apply(tokenize_and_remove_stopwords)
```

5. Feature Extraction:

Transform the text data into numerical features, such as TF-IDF vectors or word embeddings.

PROGRAM

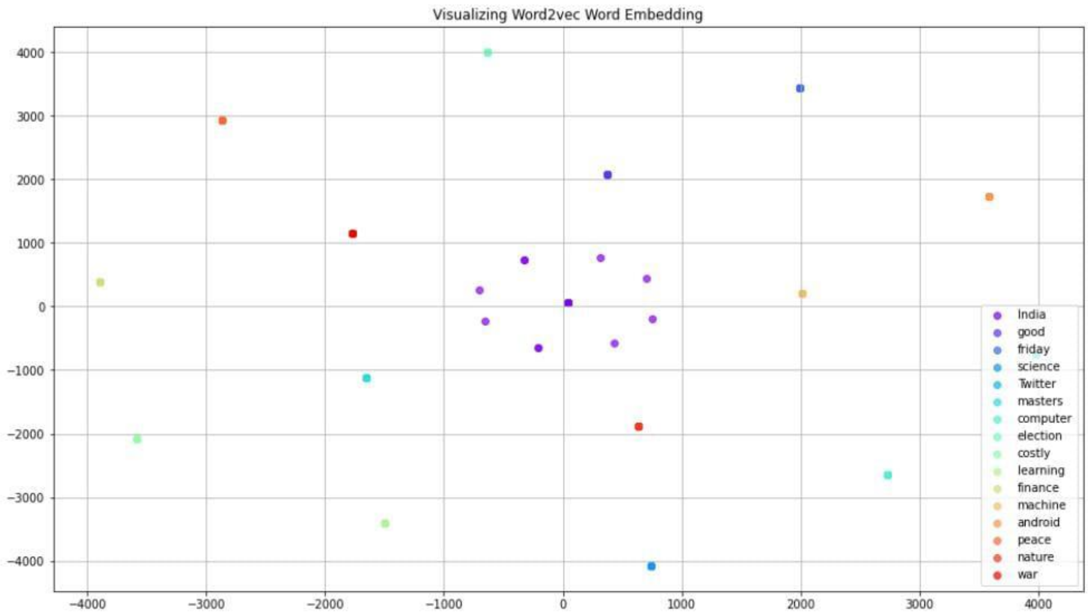
```
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf_vectorizer = TfidfVectorizer(max_features=1000)
tfidf_features = tfidf_vectorizer.fit_transform(data['text'])
```

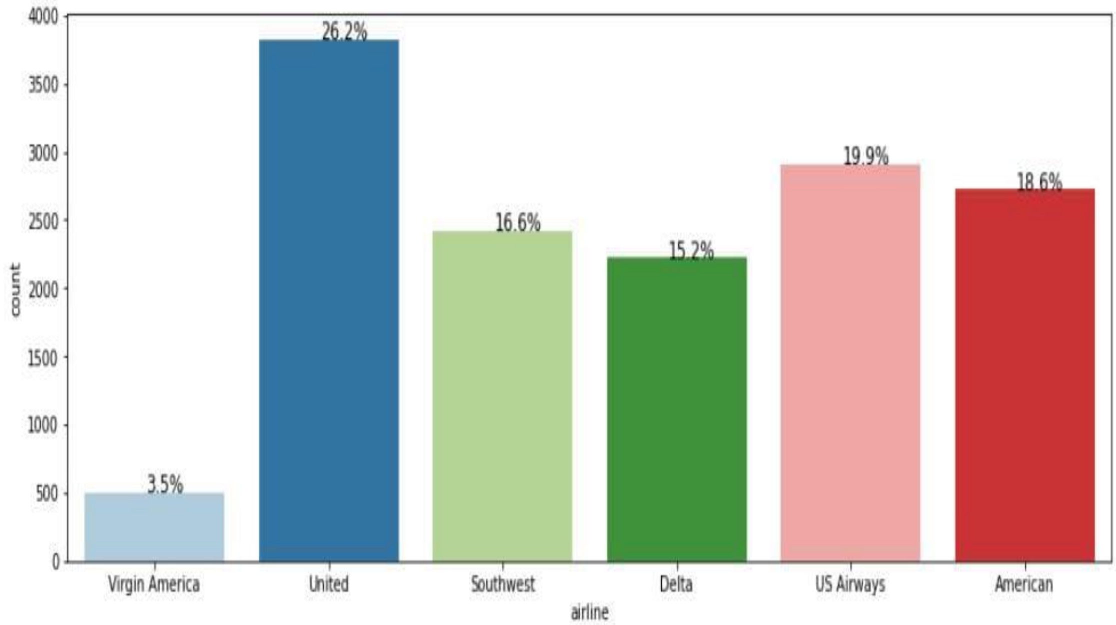
6. Data Visualization:

Create visualizations to better understand the data, e.g., sentiment distribution.

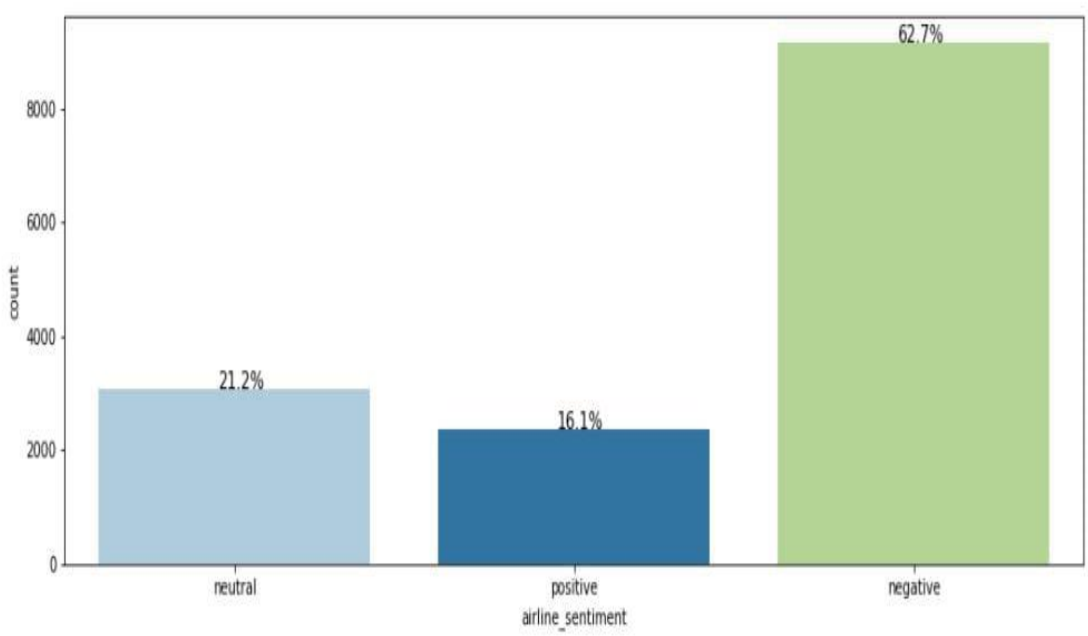
```
# Visualizing Word2vec Word Embedding


keys = ['India', 'good', 'friday', 'science', 'Twitter', 'masters', '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, topn = 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(embedding_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_cluster, title):
    colors = cm.rainbow(np.linspace(0, 1, len(labels)))
    plt.figure(figsize = (16,9))
  for label, embeddings, words, color in zip(labels, embedding_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()
plot_most_similar_words(keys, tsne_embeddings, words_clusters, "Visualizing
Word2vec Word Embedding")
```

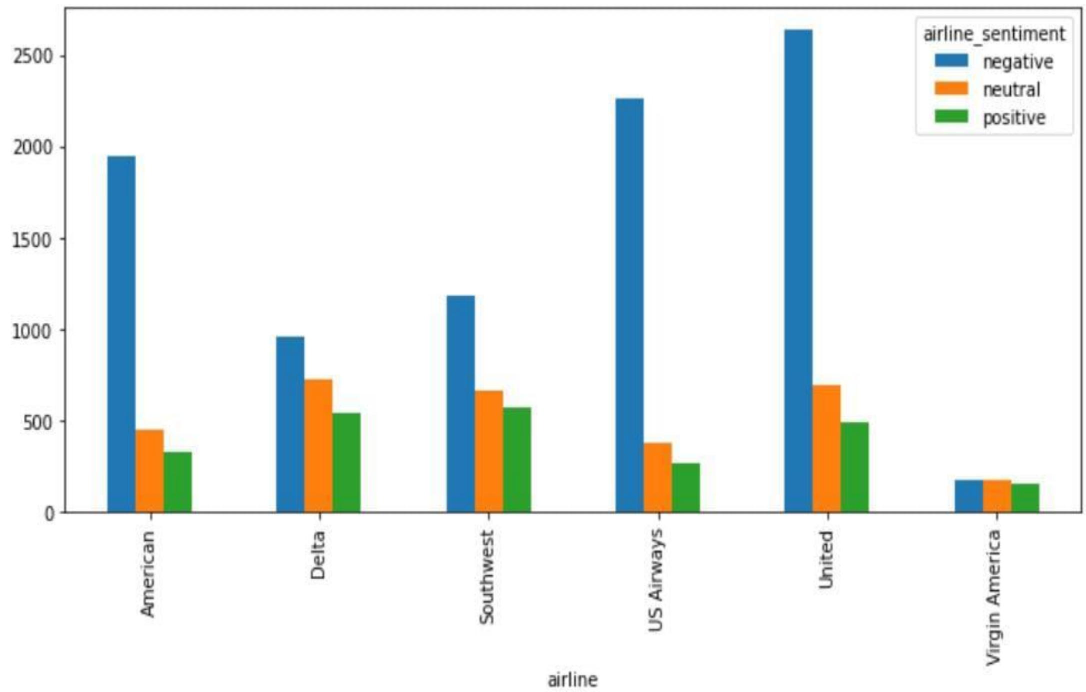




Tweets for United Airline



Overall positive and negative sentiments



Feedback based on places

SPLITTING THE DATASET:

In sentiment analysis for marketing, it's common to split your dataset into training, validation, and test sets for model development and evaluation. You can use Python's libraries such as scikit-learn or pandas for this purpose. Here's an example of how to split your data:

PROGRAM

```
import pandas as pd

from sklearn.model_selection import train_test_split

# Load your sentiment analysis data

data = pd.read_csv('your_data.csv')

# Define your features (X) and target (y)

X = data['text']  # Your text data

y = data['airline_sentiment']  # The sentiment labels

# Split the data into training, validation, and test sets

X_train, X_temp, y_train, y_temp = train_test_split(X, y, test_size=0.3,
random_state=42)

X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5,
random_state=42)

# You now have X_train, y_train for training, X_val, y_val for validation, and
X_test, y_test for testing.
```

CONCLUSION:

In the realm of sentiment analysis for airline marketing, we embarked on a comprehensive journey. The process began by collecting and preprocessing the dataset, which encompassed loading the data and refining it through techniques like text cleaning and feature extraction.

Data splitting was instrumental in partitioning our dataset into training, validation, and test sets, allowing us to train and evaluate our sentiment analysis model effectively. Model development played a pivotal role, as we selected and trained our model on the training dataset, aiming to capture the nuances of sentiment in airline-related text.

Following model development, the focus shifted to evaluation, where we fine-tuned our model based on its performance against the validation dataset. This iterative process led to a well-optimized model.

Ultimately, the culmination was in the testing phase, where our model was put to the test on the previously unseen test dataset, providing a realistic evaluation of its effectiveness.

Data visualizations served as a means to convey sentiment distribution and word clouds for positive and negative sentiments, offering further insight into our data.

The findings from this sentiment analysis endeavor can offer airlines valuable insights into customer perceptions, allowing them to tailor their marketing strategies and responses to enhance the overall passenger experience.