```
from google.colab import files
uploaded = files.upload()
```

Choose files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

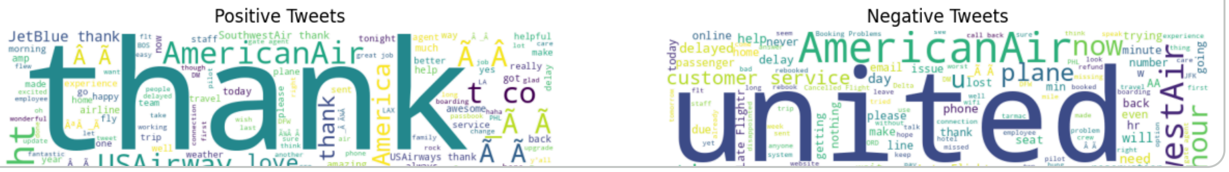
```
Saving archive.zip to archive.zip
```

```
:ets
 = " ".join(df[df['airline_sentiment'] == 'positive']['text'])
 = WordCloud(width=800, height=400, background_color='white').generate(positive_

:ets
 = " ".join(df[df['airline_sentiment'] == 'negative']['text'])
 = WordCloud(width=800, height=400, background_color='white').generate(negative_

/ side
.t.subplots(1, 2, figsize=(15,6))
v(wordcloud_pos)
.tle("Positive Tweets")
'off")

v(wordcloud_neg)
.tle("Negative Tweets")
'off")
```



```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, confusion_matrix, classification_re

# Keep only positive & negative tweets for clarity
df_model = df[df['airline_sentiment'].isin(['positive', 'negative'])]

# Split data
X_train, X_test, y_train, y_test = train_test_split(df_model['text'], df_model

# Convert text to numbers
vectorizer = CountVectorizer(stop_words='english')
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)
```

```
model = MultinomialNB()
model.fit(X_train_vec, y_train)

# Predict
y_pred = model.predict(X_test_vec)
```

Double-click (or enter) to edit

Double-click (or enter) to edit

```
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy: 0.9042875703767865

Confusion Matrix:
 [[1815   47]
 [ 174  273]]

Classification Report:
               precision    recall  f1-score   support

    negative       0.91      0.97      0.94      1862
    positive       0.85      0.61      0.71       447

    accuracy                           0.90      2309
   macro avg       0.88      0.79      0.83      2309
weighted avg       0.90      0.90      0.90      2309
```

```
import zipfile
import os

# unzip the file
with zipfile.ZipFile("archive.zip", 'r') as zip_ref:
    zip_ref.extractall("dataset")

# check what files are inside
os.listdir("dataset")
```

```
['Airline-Sentiment-2-w-AA.csv']
```

```
import pandas as pd

# Load the CSV file (use the exact file name)
df = pd.read_csv("dataset/Airline-Sentiment-2-w-AA.csv", encoding='latin1')

# Show first 5 rows
df.head()
```

	index	_unit_id	_golden	_unit_state	_trusted_judgments	_last_judgme
0	0	681448150	False	finalized	3	2/25/1
1	1	681448153	False	finalized	3	2/25/1
2	2	681448156	False	finalized	3	2/25/15
3	3	681448158	False	finalized	3	2/25/1
4	4	681448159	False	finalized	3	2/25/1

5 rows × 21 columns

```
# Check all column names
print(df.columns.tolist())

# Check missing values
print("\nMissing values per column:")
print(df.isnull().sum())

# Check data types
print("\nData types:")
print(df.dtypes)
```

```
['index', '_unit_id', '_golden', '_unit_state', '_trusted_judgments', '_l

Missing values per column:
index                               0
_unit_id                            0
_golden                             0
_unit_state                         0
_trusted_judgments                  0
_last_judgment_at                  56
airline_sentiment                   0
airline_sentiment:confidence        0
negativereason                   5462
negativereason:confidence        4118
airline                             0
airline_sentiment_gold          14600
name                                0
negativereason_gold             14608
```

```
retweet_count                       0
text                                0
tweet_coord                     13621
tweet_created                       0
tweet_id                            0
tweet_location                   4733
user_timezone                    4820
dtype: int64

Data types:
index                             int64
_unit_id                          int64
_golden                            bool
_unit_state                      object
_trusted_judgments                int64
_last_judgment_at                object
airline_sentiment                object
airline_sentiment:confidence    float64
negativereason                   object
negativereason:confidence       float64
airline                          object
airline_sentiment_gold           object
name                             object
negativereason_gold              object
retweet_count                     int64
text                             object
tweet_coord                      object
tweet_created                    object
tweet_id                        float64
tweet_location                   object
user_timezone                    object
dtype: object
```

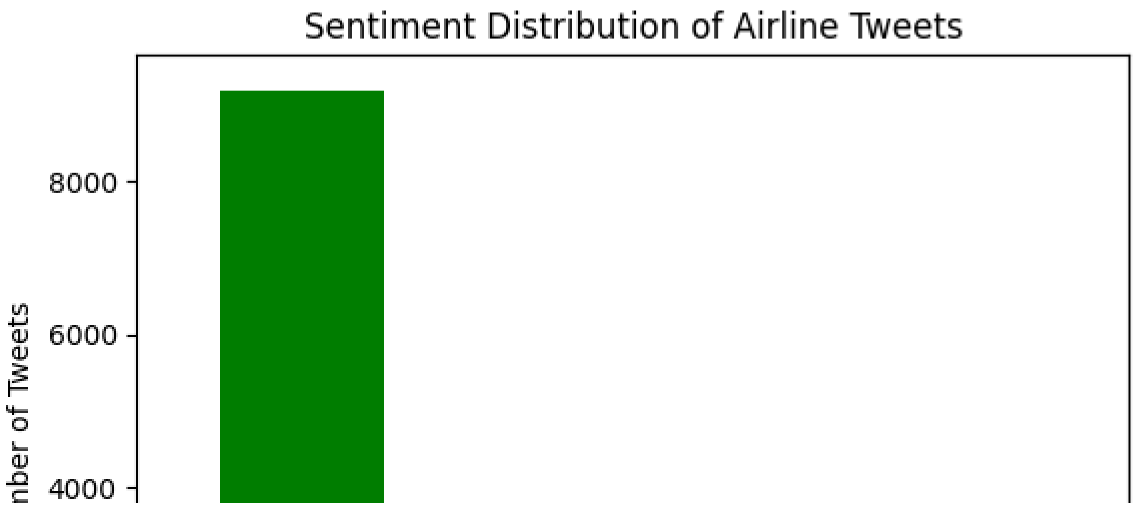
```
# Keep only text and sentiment columns (adjust names if needed)
df = df[['airline_sentiment', 'text']]

# Check the cleaned data
df.head()
```

	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...

```
import matplotlib.pyplot as plt

df['airline_sentiment'].value_counts().plot(kind='bar', color=['green','gray',
plt.title("Sentiment Distribution of Airline Tweets")
plt.xlabel("Sentiment Type")
plt.ylabel("Number of Tweets")
plt.show()
```

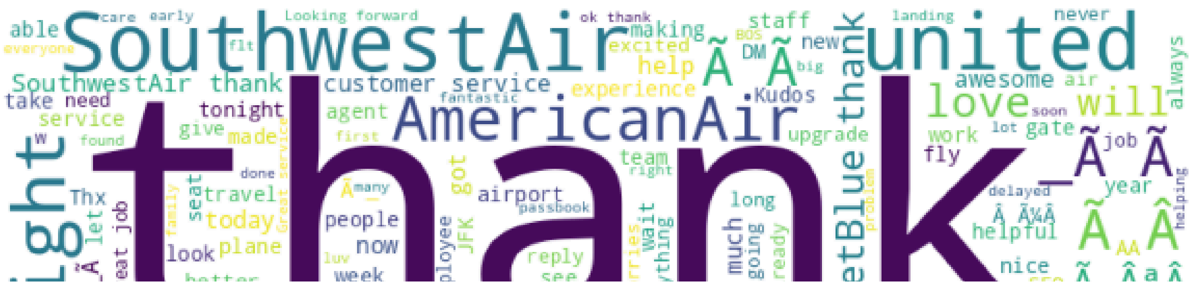


```
from wordcloud import WordCloud

# For positive tweets
positive_text = " ".join(df[df['airline_sentiment']=='positive']['text'])
wordcloud_pos = WordCloud(width=800, height=400, background_color='white').gene

plt.figure(figsize=(8,4))
plt.imshow(wordcloud_pos)
plt.axis('off')
plt.title("Word Cloud - Positive Tweets")
plt.show()
```

Word Cloud - Positive Tweets


thank
SouthwestAir
united
AmericanAir
JetBlue thank
flight
love
will
SouthwestAir thank
customer service
awesome
able
care
early
Looking forward
ok thank
staff
landing
never
everyone
flt
making
excited
BOS
DM
new
big
help
experience
Kudos
air
always
take
need
service
tonight
agent
fantastic
soon
gate
upgrade
work
lot
job
found
give
made
first
team
fly
right
done
airport
long
delayed
year
helpful
Thx
let
travel
seat
many
people
passbook
today
plane
look
now
reply
much
going
ready
nice
helping
AA
great job
week
better
see
JFK
got
family
Great service
luv
problem
employee
Wait