PROGRAM:

import csv

import matplotlib.pyplot as plt

# Read the CSV file

with open('Tweets.csv', 'r', encoding='utf-8') as f:

reader = csv.reader(f)

data = list(reader)

# Extract the data columns

tweet\_ids = [row[0] for row in data]

airline\_sentiments = [row[1] for row in data]

airline\_sentiment\_confidences = [row[2] for row in data]

negative\_reasons = [row[3] for row in data]

negative\_reason\_confidences = [row[4] for row in data]

airlines = [row[5] for row in data]

airline\_sentiment\_golds = [row[6] for row in data]

airline\_names = [row[7] for row in data]

negative\_reason\_golds = [row[8] for row in data]

# Try to convert the retweet\_count to an integer. If the conversion fails, set the retweet\_count to 0.

retweet\_counts = []

for row in data:

try:

retweet\_counts.append(int(row[9]))

except ValueError:

retweet\_counts.append(0)

texts = [row[10] for row in data]

tweet\_coords = [row[11] for row in data]

tweet\_createds = [row[12] for row in data]

tweet\_locations = [row[13] for row in data]

user\_timezones = [row[14] for row in data]

# Create a dictionary to store the data by airline

airline\_data = {}

for i in range(len(airline\_names)):

airline = airline\_names[i]

if airline not in airline\_data:

airline\_data[airline] = []

airline\_data[airline].append({

'tweet\_id': tweet\_ids[i],

'airline\_sentiment': airline\_sentiments[i],

'airline\_sentiment\_confidence': airline\_sentiment\_confidences[i],

'negative\_reason': negative\_reasons[i],

'negative\_reason\_confidence': negative\_reason\_confidences[i],

'airline\_sentiment\_gold': airline\_sentiment\_golds[i],

'negative\_reason\_gold': negative\_reason\_golds[i],

'retweet\_count': retweet\_counts[i],

'text': texts[i],

'tweet\_coord': tweet\_coords[i],

'tweet\_created': tweet\_createds[i],

'tweet\_location': tweet\_locations[i],

'user\_timezone': user\_timezones[i]

})

# Create a bar chart showing the number of tweets for each airline

plt.figure(figsize=(10, 6))

plt.bar(airline\_names, [len(airline\_data[airline]) for airline in airline\_names])

plt.xlabel('Airline')

plt.ylabel('Number of tweets')

plt.title('Number of tweets for each airline in the Tweets')

plt.show()

# Create a pie chart showing the percentage of tweets with each airline sentiment

airline\_sentiment\_counts = {}

for airline in airline\_data:

for tweet in airline\_data[airline]:

airline\_sentiment = tweet['airline\_sentiment']

if airline\_sentiment not in airline\_sentiment\_counts:

airline\_sentiment\_counts[airline\_sentiment] = 0

airline\_sentiment\_counts[airline\_sentiment] += 1

plt.figure(figsize=(10, 6))

plt.pie(list(airline\_sentiment\_counts.values()), labels=list(airline\_sentiment\_counts.keys()), autopct='%1.1f%%')

plt.title('Percentage of tweets with each airline sentiment')

plt.show()

# Create a scatter plot showing the relationship between