Theory Activity

Name: Ritesh Kumar

PRN: 2024010236

Div : CS1

Roll No.: CS1-63

Problem 1: Find the top 3 airlines with the highest average sentiment score.

Solution:

```
sentiment_map = {'positive': 1, 'neutral': 0, 'negative': -1}
df['sentiment_score'] = df['airline_sentiment'].map(sentiment_map)
avg_sentiment                                                                    =
df.groupby('airline')['sentiment_score'].mean().sort_values(ascending=False).head(3)
print(avg_sentiment)
```

Problem 2: Find which airline has the most balanced sentiment distribution (least var

Solution:

```
variance_sentiment = df.groupby('airline')['sentiment_score'].var().sort_values()
print(variance_sentiment)
```

Problem 3: Identify the airline with the maximum number of tweets without negative r

Solution:

```
no_neg_reason = df[df['negativereason'].isnull()]
airline_no_neg_reason = no_neg_reason['airline'].value_counts().idxmax()
print(airline_no_neg_reason)
```

Problem 4: Find the day with the most number of positive tweets.

Solution:

```
positive_tweets = df[df['airline_sentiment'] == 'positive']
positive_tweets['date'] = pd.to_datetime(positive_tweets['tweet_created']).dt.date
most_positive_day = positive_tweets['date'].value_counts().idxmax()
print(most_positive_day)
```

Problem 5: Calculate the average retweet count for each sentiment.

Solution:

```
avg_retweets_sentiment = df.groupby('airline_sentiment')['retweet_count'].mean()
print(avg_retweets_sentiment)
```

Problem 6: Determine which negative reason is most common for United Airlines.

Solution:

```
united_neg = df[(df['airline'] == 'United') & (df['airline_sentiment'] ==

'negative')] most_common_reason = united_neg['negativereason'].value_counts().idxmax()

print(most_common_reason)
```

Problem 7: Find out which airline received the least number of negative tweets.

Solution:

```
neg_tweets_count = df[df['airline_sentiment'] == 'negative'].groupby('airline').size()
least_neg_airline = neg_tweets_count.idxmin() print(least_neg_airline)
```

Problem 8: Calculate the proportion of tweets that are replies.

Solution:

```
reply_proportion = df['in_reply_to_status_id'].notnull().mean()
print(reply_proportion)
```

Problem 9: Find the user who posted the longest tweet.

Solution:

```
df['tweet_length'] = df['text'].str.len()
longest_tweet_user = df.loc[df['tweet_length'].idxmax(), 'user']
print(longest_tweet_user)
```

Problem 10: Find the most common tweet hour for American Airlines.

Solution:

```
aa_df = df[df['airline'] == 'American']
aa_df['tweet_created'] = pd.to_datetime(aa_df['tweet_created'])
common_hour = aa_df['tweet_created'].dt.hour.value_counts().idxmax()
print(common_hour)
```

Problem 11: Calculate the standard deviation of tweet lengths.

Solution:

```
tweet_length_std = df['tweet_length'].std()
print(tweet_length_std)
```

Problem 12: List the top 5 users who posted positive tweets.

Solution:

```
positive_users = df[df['airline_sentiment'] == 'positive']['user'].value_counts().head(5)
print(positive_users)
```

Problem 13: Find how many tweets mention 'delay' or 'delayed'.

Solution:

```
delay_mentions = df['text'].str.contains('delay|delayed', case=False, na=False).sum()
print(delay_mentions)
```

Problem 14: Get the number of neutral tweets mentioning 'customer service'.

Solution:

```
neutral_service = df[(df['airline_sentiment'] == 'neutral') & (df['text'].str.contains('customer service', case=False, na=False))]
print(neutral_service.shape[0])
```

Problem 15: Identify the airline most praised for 'on time' flights.

Solution:

```
on_time_praise = df[df['text'].str.contains('on time', case=False, na=False)]
airline_on_time = on_time_praise['airline'].value_counts().idxmax()
print(airline_on_time)
```

Problem 16: Find how many tweets were created on weekends.

Solution:

```
df['tweet_created'] = pd.to_datetime(df['tweet_created'])
weekend_tweets = df[df['tweet_created'].dt.dayofweek >= 5].shape[0]
print(weekend_tweets)
```

Problem 17: Determine the airline whose tweets received the maximum average numb

Solution:

```
max_retweet_airline = df.groupby('airline')['retweet_count'].mean().idxmax()
print(max_retweet_airline)
```

Problem 18: Find the correlation between sentiment score and retweet count.

Solution:

```
correlation_sr = df['sentiment_score'].corr(df['retweet_count'])
print(correlation_sr)
```

Problem 19: Find the median tweet length for each sentiment category.

Solution:

```
median_length_sentiment = df.groupby('airline_sentiment')['tweet_length'].median()
print(median_length_sentiment)
```

Problem 20: Find the airline with the maximum number of positive tweets containing

Solution:

```
positive_thank = df[(df['airline_sentiment'] == 'positive') &
(df['text'].str.contains('thank you', case=False, na=False))]
airline_thank = positive_thank['airline'].value_counts().idxmax()
print(airline_thank)
```