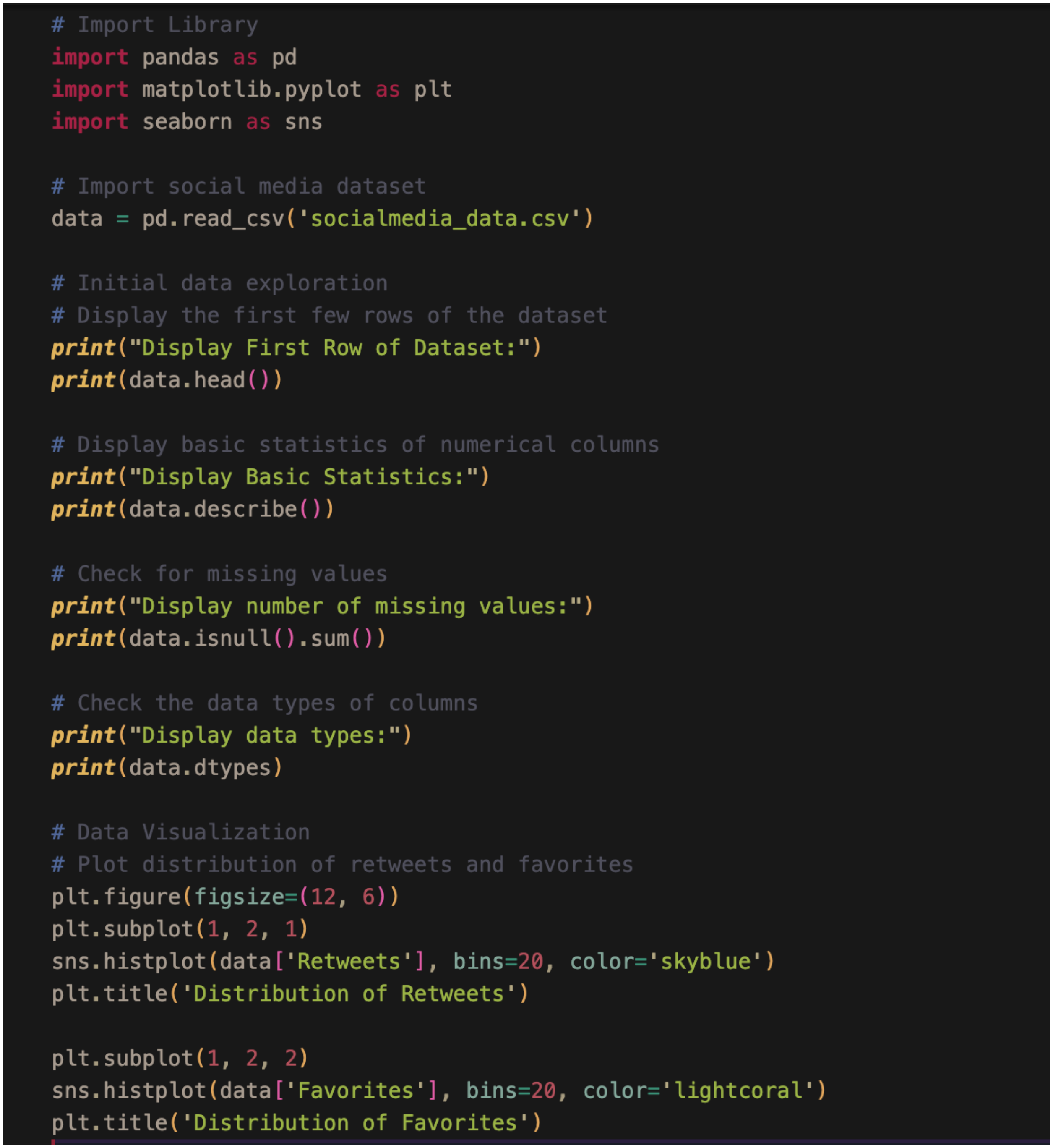
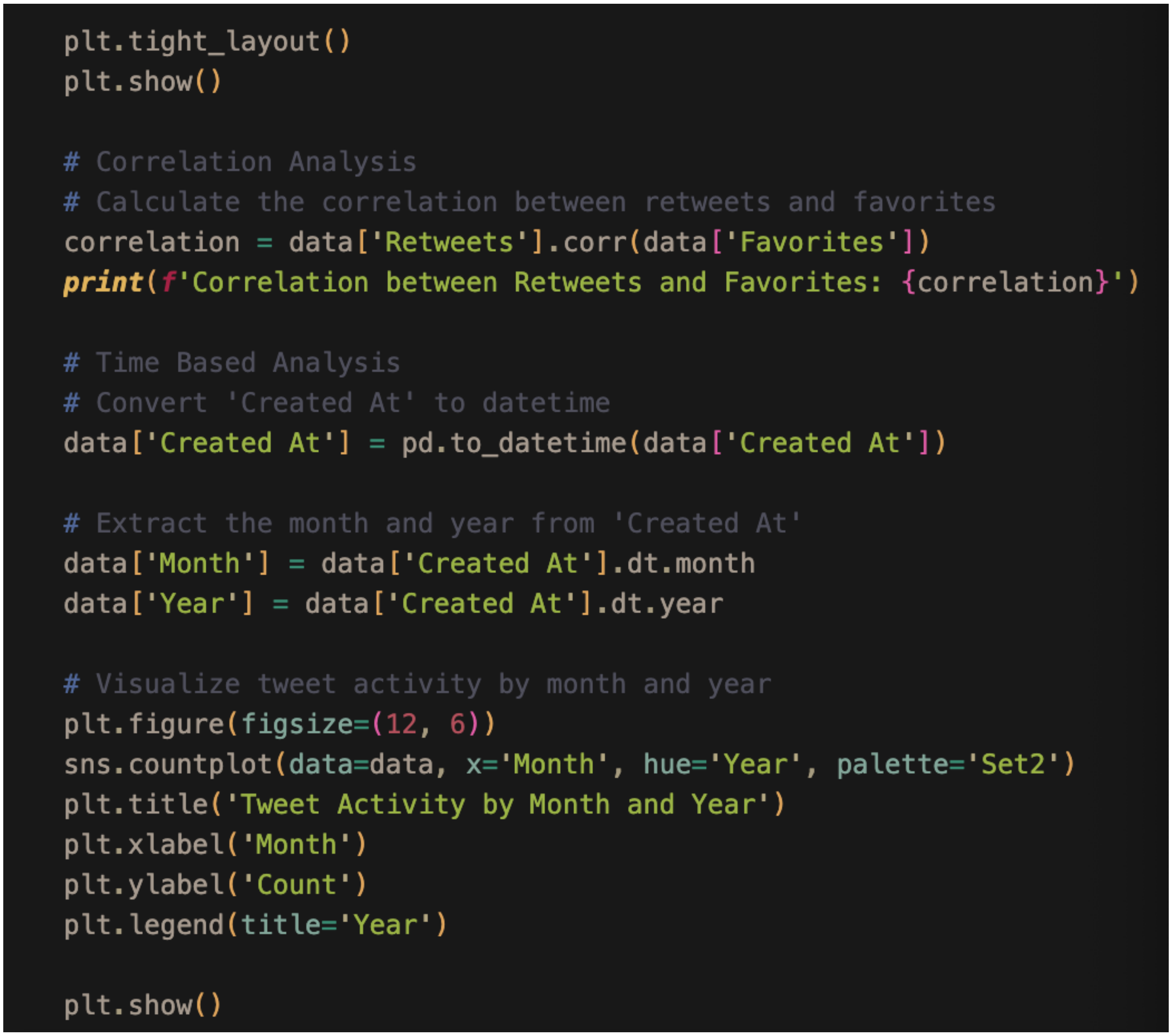
T8D2 Discussion

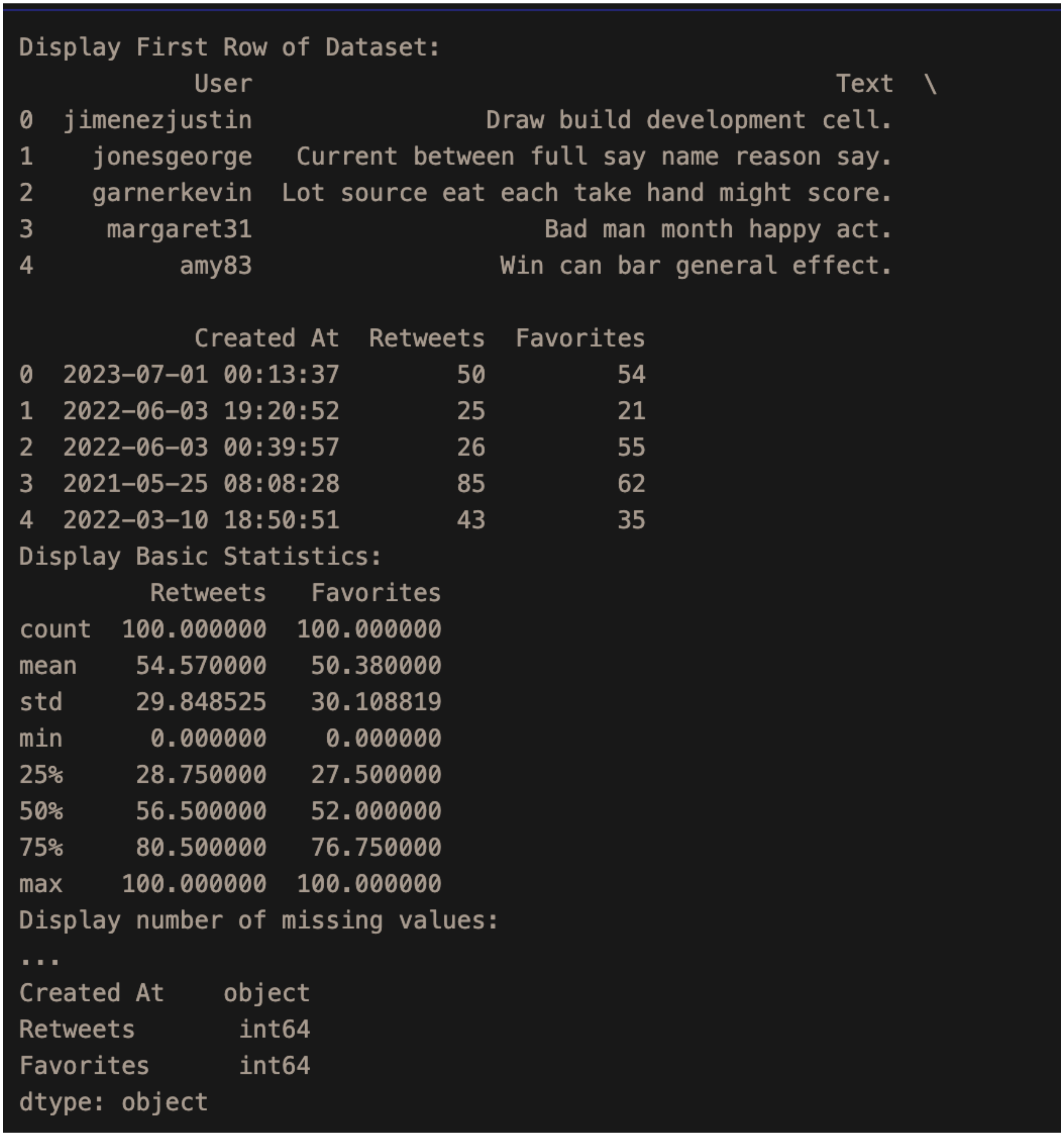
November 6, 2023



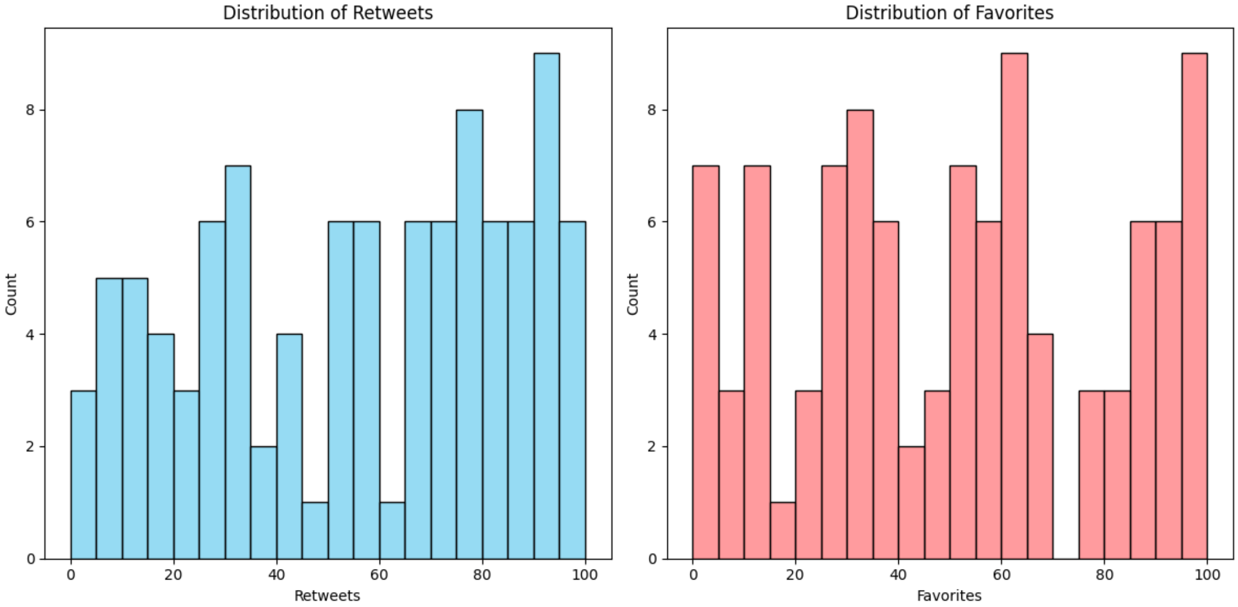
```python
# Import Library
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Import social media dataset
data = pd.read_csv('socialmedia_data.csv')

# Initial data exploration
# Display the first few rows of the dataset
print("Display First Row of Dataset:")
print(data.head())

# Display basic statistics of numerical columns
print("Display Basic Statistics:")
print(data.describe())

# Check for missing values
print("Display number of missing values:")
print(data.isnull().sum())

# Check the data types of columns
print("Display data types:")
print(data.dtypes)

# Data Visualization
# Plot distribution of retweets and favorites
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
sns.histplot(data['Retweets'], bins=20, color='skyblue')
plt.title('Distribution of Retweets')

plt.subplot(1, 2, 2)
sns.histplot(data['Favorites'], bins=20, color='lightcoral')
plt.title('Distribution of Favorites')
```



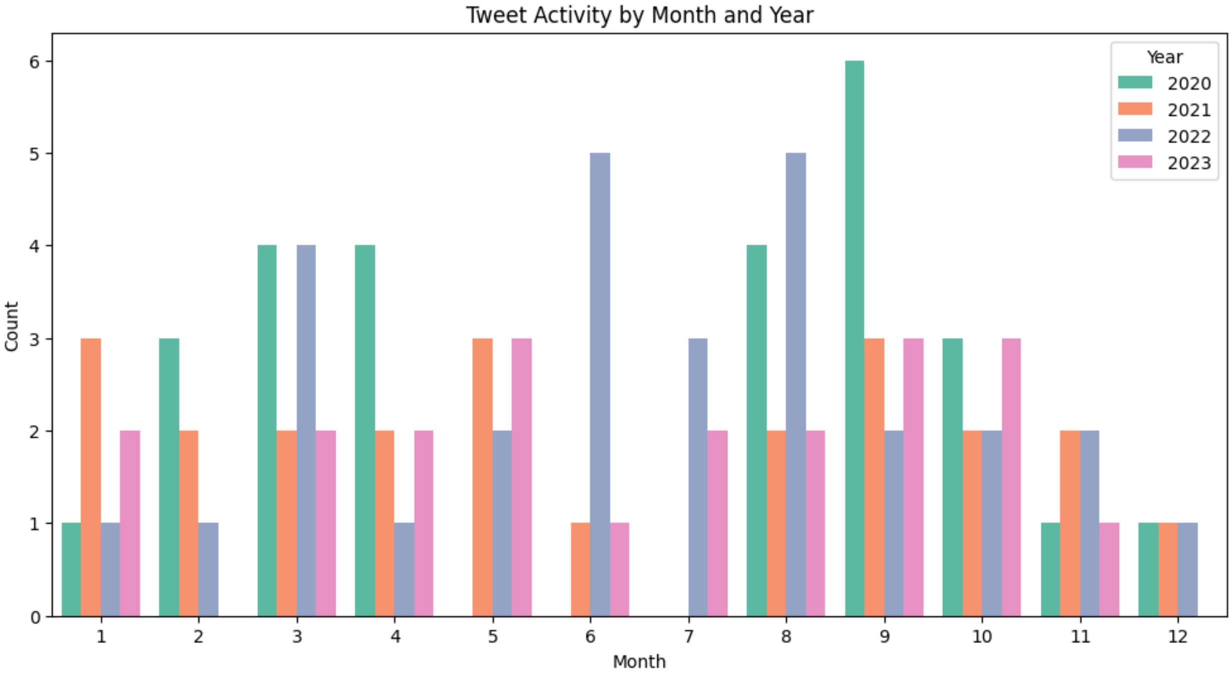
```
plt.tight_layout()
plt.show()

# Correlation Analysis
# Calculate the correlation between retweets and favorites
correlation = data['Retweets'].corr(data['Favorites'])
print(f'Correlation between Retweets and Favorites: {correlation}')

# Time Based Analysis
# Convert 'Created At' to datetime
data['Created At'] = pd.to_datetime(data['Created At'])

# Extract the month and year from 'Created At'
data['Month'] = data['Created At'].dt.month
data['Year'] = data['Created At'].dt.year

# Visualize tweet activity by month and year
plt.figure(figsize=(12, 6))
sns.countplot(data=data, x='Month', hue='Year', palette='Set2')
plt.title('Tweet Activity by Month and Year')
plt.xlabel('Month')
plt.ylabel('Count')
plt.legend(title='Year')

plt.show()
```



```
Display First Row of Dataset:
            User                                        Text  \
0  jimenezjustin                Draw build development cell.   
1    jonesgeorge   Current between full say name reason say.   
2    garnerkevin  Lot source eat each take hand might score.   
3     margaret31                    Bad man month happy act.   
4          amy83                 Win can bar general effect.   

            Created At  Retweets  Favorites  
0  2023-07-01 00:13:37        50         54  
1  2022-06-03 19:20:52        25         21  
2  2022-06-03 00:39:57        26         55  
3  2021-05-25 08:08:28        85         62  
4  2022-03-10 18:50:51        43         35  
Display Basic Statistics:
         Retweets   Favorites
count  100.000000  100.000000
mean    54.570000   50.380000
std     29.848525   30.108819
min      0.000000    0.000000
25%     28.750000   27.500000
50%     56.500000   52.000000
75%     80.500000   76.750000
max    100.000000  100.000000
Display number of missing values:
...
Created At    object
Retweets       int64
Favorites      int64
dtype: object
```



```
Correlation between Retweets and Favorites: 0.07537611865206571
```


Tweet Activity by Month and Year
Count
Month
Year