In [*]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


while True:

    print("Enter 1 to see the analysis of Reviews given by Customers")
    print("Enter 2 to see the analysis of different payment methods used by th
    print("Enter 3 to see the analysis of Top Consumer States of India")
    print("Enter 4 to see the analysis of Top Consumer Cities of India")
    print("Enter 5 to see the analysis of Top Selling Product Categories")
    print("Enter 6 to see the analysis of Reviews for All Product Categories")
    print("Enter 7 to see the analysis of Number of Orders Per Month Per Year"
    print("Enter 8 to see the analysis of Reviews for Number of Orders Per Mon
    print("Enter 9 to see the analysis of Number of Orders Across Parts of a D
    print("Enter 10 to see the Full Report")
    choice = input("Enter the number to see the analysis of your choice: ")



    if choice == "1":

        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)
        value_counts = df['stars'].value_counts()
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(x=value_counts.index, y=value_counts.values)
        plt.xlabel('Star Rating')
        plt.ylabel('Count')
        plt.title('Distribution of Star Ratings')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "2":


        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        value_counts = df['Payment Method'].value_counts()
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(x=value_counts.index, y=value_counts.values)
        plt.xlabel('Payment Method')
        plt.ylabel('Count')
        plt.title('Distribution of Payment Methods')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "3":
                ##Load the Excel sheet into a DataFrame
        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        state_customers = df['Shipping State'].value_counts()
        plt.figure(figsize=(10, 6))
```

```
        ax = sns.barplot(x=state_customers.index, y=state_customers.values)
        plt.xlabel('State')
        plt.ylabel('Number of Customers')
        plt.title('Number of Customers by State')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "4":
        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        city_counts = df['Shipping City'].value_counts().head(10)

        # Plot a bar chart of the top 10 cities with the most customers
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(x=city_counts.index, y=city_counts.values)
        plt.xlabel('City')
        plt.ylabel('Number of Customers')
        plt.title('Top 10 Cities with the Most Customers')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "5":

        # Load the Excel sheet into a DataFrame
        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)

        # Select the top 10 categories
        top_categories = df['category'].value_counts().head(10)
        plt.figure(figsize=(10, 6))
        sns.countplot(data=df, x='category', order=top_categories.index, palet
        plt.xlabel('Category')
        plt.ylabel('Count')
        plt.title('Count of Top 10 Categories')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "6":

        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)
        category_counts = df['category'].value_counts()
        top_category = category_counts.index[0]
        plt.figure(figsize=(8, 6))
        category_counts.plot(kind='bar')
        plt.xlabel('Category')
        plt.ylabel('Count')
        plt.title(f'Top Category: {top_category}')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()

    elif choice == "7":
```

```
        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        df['Order Date and Time Stamp'] = pd.to_datetime(df['Order Date and Ti
        df['Year'] = df['Order Date and Time Stamp'].dt.year
        df['Month'] = df['Order Date and Time Stamp'].dt.month
        orders_per_month = df.groupby(['Year', 'Month']).size().reset_index(na
        plt.figure(figsize=(10, 6))
        colors = ['red', 'blue', 'green', 'orange', 'purple']
        years = orders_per_month['Year'].unique()

        for i, year in enumerate(years):
            year_data = orders_per_month[orders_per_month['Year'] == year]
            plt.plot(year_data['Month'], year_data['Number of Orders'], marker

        plt.xlabel('Month')
        plt.ylabel('Number of Orders')
        plt.title('Number of Orders per Year and Month')
        months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep
        plt.xticks(range(1, 13), labels=months)

        plt.legend([str(year) for year in years])
        plt.grid(True)
        plt.tight_layout()
        plt.show()

    elif choice == "8":

        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)
        df['status'].fillna('not reviewed', inplace=True)
        status_counts = df['status'].value_counts()
        plt.figure(figsize=(8, 6))
        plt.pie(status_counts, labels=status_counts.index, autopct='%1.1f%%',
        plt.title('Review Status')
        plt.axis('equal')
        plt.show()
        plt.figure(figsize=(8, 6))
        status_counts.plot(kind='bar')
        plt.xlabel('Review Status')
        plt.ylabel('Count')
        plt.title('Review Status')
        plt.xticks(rotation=0)
        plt.show()
        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)
        star_ratings = ['5.0 star rating', '4.9 star rating', '4.6 star rating
                        '4.0 star rating', '3.0 star rating', '2.3 star rating
                        '4.7 star rating', '3.3 star rating', '4.2 star rating
        filtered_df = df[df['stars'].isin(star_ratings)]
        rating_counts = filtered_df.groupby(['product_name', 'stars']).size().
        for rating in star_ratings:
            products = rating_counts[rating]
            top_product = products.idxmax()
            print("Product with the highest count of", rating, ":", top_produc
        plt.figure(figsize=(8, 8))
        explode = [0.1] * len(star_ratings)
        labels = [f"{rating}\n({rating_counts[rating].idxmax()})" for rating i
```

```
        plt.pie(rating_counts.sum(), labels=labels, autopct='%1.1f%%', explode
        plt.title('Distribution of Star Ratings for Products')
        plt.axis('equal')
        plt.show()

    elif choice== "9":


        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        df['Order Date and Time Stamp'] = pd.to_datetime(df['Order Date and Ti
        df['Order Date'] = df['Order Date and Time Stamp'].dt.date
        order_counts = df.groupby('Order Date').size()
        plt.figure(figsize=(12, 6))
        order_counts.plot(kind='bar', color='skyblue')
        plt.xlabel('Order Date')
        plt.ylabel('Order Count')
        plt.title('Order Count for Each Day')
        plt.xticks(rotation=45)
        plt.tight_layout()
        plt.show()

    elif choice == "10":



        file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
        df = pd.read_csv(file_path)
        value_counts = df['stars'].value_counts()
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(x=value_counts.index, y=value_counts.values)
        plt.xlabel('Star Rating')
        plt.ylabel('Count')
        plt.title('Distribution of Star Ratings')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()



        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        value_counts = df['Payment Method'].value_counts()
        plt.figure(figsize=(10, 6))
        ax = sns.barplot(x=value_counts.index, y=value_counts.values)
        plt.xlabel('Payment Method')
        plt.ylabel('Count')
        plt.title('Distribution of Payment Methods')
        plt.xticks(rotation=45, ha='right')
        plt.tight_layout()
        plt.show()



        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
```

```
state_customers = df['Shipping State'].value_counts()
plt.figure(figsize=(10, 6))
ax = sns.barplot(x=state_customers.index, y=state_customers.values)
plt.xlabel('State')
plt.ylabel('Number of Customers')
plt.title('Number of Customers by State')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

```
file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
df = pd.read_csv(file_path)
city_counts = df['Shipping City'].value_counts().head(10)
plt.figure(figsize=(10, 6))
ax = sns.barplot(x=city_counts.index, y=city_counts.values)
plt.xlabel('City')
plt.ylabel('Number of Customers')
plt.title('Top 10 Cities with the Most Customers')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

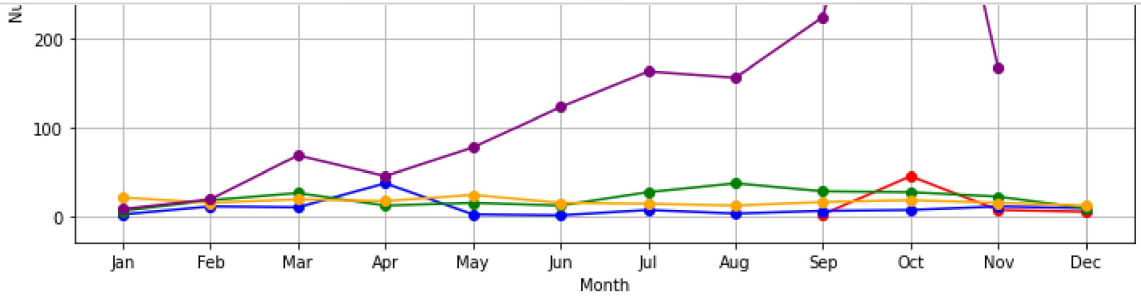
```
file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
df = pd.read_csv(file_path)
top_categories = df['category'].value_counts().head(10)
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='category', order=top_categories.index, palet
plt.xlabel('Category')
plt.ylabel('Count')
plt.title('Count of Top 10 Categories')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

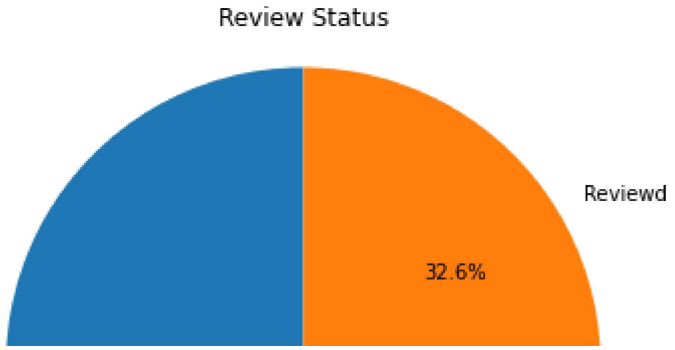
```
file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
df = pd.read_csv(file_path)
category_counts = df['category'].value_counts()
top_category = category_counts.index[0]
plt.figure(figsize=(8, 6))
category_counts.plot(kind='bar')
plt.xlabel('Category')
plt.ylabel('Count')
plt.title(f'Top Category: {top_category}')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

```
file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
df = pd.read_csv(file_path)
df['Order Date and Time Stamp'] = pd.to_datetime(df['Order Date and Ti
df['Year'] = df['Order Date and Time Stamp'].dt.year
df['Month'] = df['Order Date and Time Stamp'].dt.month
```

```
orders_per_month = df.groupby(['Year', 'Month']).size().reset_index(na
plt.figure(figsize=(10, 6))
colors = ['red', 'blue', 'green', 'orange', 'purple']
years = orders_per_month['Year'].unique()

for i, year in enumerate(years):
    year_data = orders_per_month[orders_per_month['Year'] == year]
    plt.plot(year_data['Month'], year_data['Number of Orders'], marker

plt.xlabel('Month')
plt.ylabel('Number of Orders')
plt.title('Number of Orders per Year and Month')


months = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Sep
plt.xticks(range(1, 13), labels=months)

plt.legend([str(year) for year in years])
plt.grid(True)
plt.tight_layout()
plt.show()



file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
df = pd.read_csv(file_path)
df['status'].fillna('not reviewed', inplace=True)
status_counts = df['status'].value_counts()
plt.figure(figsize=(8, 6))
plt.pie(status_counts, labels=status_counts.index, autopct='%1.1f%%',
plt.title('Review Status')
plt.axis('equal')
plt.show()


plt.figure(figsize=(8, 6))
status_counts.plot(kind='bar')
plt.xlabel('Review Status')
plt.ylabel('Count')
plt.title('Review Status')
plt.xticks(rotation=0)
plt.show()


file_path = r"C:\Users\hp\Downloads\review_dataset.csv"
df = pd.read_csv(file_path)
star_ratings = ['5.0 star rating', '4.9 star rating', '4.6 star rating
                '4.0 star rating', '3.0 star rating', '2.3 star rating
                '4.7 star rating', '3.3 star rating', '4.2 star rating


filtered_df = df[df['stars'].isin(star_ratings)]

rating_counts = filtered_df.groupby(['product_name', 'stars']).size().

for rating in star_ratings:
```

```
            products = rating_counts[rating]
            top_product = products.idxmax()
            print("Product with the highest count of", rating, ":", top_produc

        plt.figure(figsize=(8, 8))
        explode = [0.1] * len(star_ratings)
        labels = [f"{rating}\n({rating_counts[rating].idxmax()})" for rating i

        plt.pie(rating_counts.sum(), labels=labels, autopct='%1.1f%%', explode
        plt.title('Distribution of Star Ratings for Products')
        plt.axis('equal')
        plt.show()

        file_path = r"C:\Users\hp\Downloads\orders_2016-2020_Dataset.csv"
        df = pd.read_csv(file_path)
        df['Order Date and Time Stamp'] = pd.to_datetime(df['Order Date and Ti
        df['Order Date'] = df['Order Date and Time Stamp'].dt.date
        order_counts = df.groupby('Order Date').size()
        plt.figure(figsize=(12, 6))
        order_counts.plot(kind='bar', color='skyblue')
        plt.xlabel('Order Date')
        plt.ylabel('Order Count')
        plt.title('Order Count for Each Day')
        plt.xticks(rotation=45)
        plt.tight_layout()
        plt.show()


    else:
        print("Invalid choice. Please enter a number from 1 to 10.")
```





In [ ]:

In [ ]: