

```

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

# Download the dataset from Kaggle (specific instructions may vary)
# Assuming you have the 'olist_orders_dataset.csv' file downloaded

# Read the CSV file
df = pd.read_csv('input/olist_orders_dataset.csv')

# **Data Overview**

print(df.shape) # Print the shape (rows, columns) of the DataFrame
print(df.info()) # Print data types and missing value information

# **Customer Analysis (assuming customer_id is unique identifier):**

# Example: Order frequency per customer
customer_orders = df.groupby('customer_id')['order_id'].count()
sns.displot(customer_orders) # Histogram of number of orders per customer
plt.xlabel('Number of Orders')
plt.ylabel('Number of Customers')
plt.title('Distribution of Orders per Customer')
plt.show()

# Example: Average order value by customer location (assuming 'customer_state' exists)
avg_order_value_state = df.groupby('customer_state')['freight_value', 'price'].sum()
avg_order_value_state['average_order_value'] = avg_order_value_state['price'] + avg_order_value_state['freight_value']
avg_order_value_state.drop(['price', 'freight_value'], axis=1, inplace=True)
sns.barplot(x=avg_order_value_state.index, y=avg_order_value_state['average_order_value'])
plt.xlabel('Customer State')
plt.ylabel('Average Order Value')
plt.title('Average Order Value by Customer Location')
plt.xticks(rotation=45) # Rotate state names for readability
plt.show()

# **Product Analysis:**

# Example: Top selling products by category
top_selling_products = df.groupby('product_category_name')['order_id'].count().nlargest(10)
sns.barplot(x=top_selling_products.index, y=top_selling_products.values)
plt.xlabel('Product Category')
plt.ylabel('Number of Orders')
plt.title('Top 10 Selling Product Categories')
plt.xticks(rotation=45) # Rotate category names for readability
plt.show()

# Example: Price distribution by product category
sns.boxplot(
    x="product_category_name",
    y="price",
    showmeans=True,
    data=df
)
plt.xlabel('Product Category')
plt.ylabel('Price')
plt.title('Distribution of Price by Product Category')
plt.xticks(rotation=45) # Rotate category names for readability
plt.show()

# **Sales Analysis (assuming 'order_approved_at' exists):**

# Example: Daily sales trend
df['date'] = pd.to_datetime(df['order_approved_at']).dt.date # Convert order approval date to date format
daily_sales = df.groupby('date')['freight_value', 'price'].sum()
daily_sales['total_sales'] = daily_sales['price'] + daily_sales['freight_value']
daily_sales.drop(['price', 'freight_value'], axis=1, inplace=True)
plt.figure(figsize=(12, 6))
plt.plot(daily_sales.index, daily_sales['total_sales'])
plt.xlabel('Date')
plt.ylabel('Total Sales')
plt.title('Daily Sales Trend')
plt.xticks(rotation=45) # Rotate dates for readability (optional)
plt.show()

```