```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Load data

customer_df = pd.read_csv('QVI_purchase_behaviour.csv')

transaction_df = pd.read_excel('QVI_transaction_data.xlsx')

# Preprocessing

transaction_df["DATE"] = pd.to_datetime("1899-12-30") + pd.to_timedelta(transaction_df["DATE"], unit="D")

transaction_df.columns = transaction_df.columns.str.strip()

customer_df.columns = customer_df.columns.str.strip()

# Feature extraction

transaction_df["BRAND"] = transaction_df["PROD_NAME"].str.split().str[0]

transaction_df["PACK_SIZE"] = transaction_df["PROD_NAME"].str.extract(r'(\d+)(?=g)')[0].astype(float)

# Merge

merged_df = pd.merge(transaction_df, customer_df, on="LYLTY_CARD_NBR", how="inner")

# Remove outliers (unusually large quantity purchases)

cleaned_df = merged_df[merged_df["PROD_QTY"] <= 10]

# Metrics

segment_sales = cleaned_df.groupby(["LIFESTAGE", "PREMIUM_CUSTOMER"])["TOT_SALES"].sum().reset_index()

avg_sales = cleaned_df.groupby(["LIFESTAGE", "PREMIUM_CUSTOMER"])["TOT_SALES"].mean().reset_index(name="AVG_SALE")

brand_sales = cleaned_df.groupby("BRAND")["TOT_SALES"].sum().reset_index().sort_values(by="TOT_SALES", ascending=False)
```

```
pack_sales = cleaned_df.groupby("PACK_SIZE")["TOT_SALES"].sum().reset_index().sort_values(by="TOT_SALES", ascending=False)

# Monthly trend

cleaned_df["MONTH"] = cleaned_df["DATE"].dt.to_period("M").astype(str)

monthly_sales = cleaned_df.groupby("MONTH")["TOT_SALES"].sum().reset_index()


# Plots

plt.figure(figsize=(12, 6))

sns.barplot(data=segment_sales, x="LIFESTAGE", y="TOT_SALES", hue="PREMIUM_CUSTOMER")

plt.title("Total Sales by Customer Segment")

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()


plt.figure(figsize=(12, 6))

sns.barplot(data=brand_sales.head(10), x="BRAND", y="TOT_SALES")

plt.title("Top 10 Brands by Sales")

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()


plt.figure(figsize=(12, 6))

sns.barplot(data=pack_sales.head(10), x="PACK_SIZE", y="TOT_SALES")

plt.title("Top 10 Pack Sizes by Sales")
```

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


plt.figure(figsize=(12, 6))
sns.lineplot(data=monthly_sales, x="MONTH", y="TOT_SALES", marker="o")
plt.title("Monthly Sales Trend")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```