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

# Load the data
df = pd.read_csv("QVI_data.csv")

# Convert date column
df["DATE"] = pd.to_datetime(df["DATE"])

# Add month and year
df["MONTH"] = df["DATE"].dt.to_period("M")

# Trial store list
trial_stores = [77, 86, 88]

# Step 1: Aggregate metrics per store per month
monthly_metrics = df.groupby(["STORE_NBR", "MONTH"]).agg(
    TOTAL_SALES=('TOT_SALES', 'sum'),
    NUM_CUSTOMERS=('LYLTY_CARD_NBR', 'nunique'),
    TRANSACTIONS=('TXN_ID', 'nunique')
).reset_index()
```

```
monthly_metrics["AVG_TXN_PER_CUSTOMER"] =
monthly_metrics["TRANSACTIONS"] /
monthly_metrics["NUM_CUSTOMERS"]


# Step 2: Define function to calculate similarity between stores
def calculate_similarity(trial_store, metric_df,
metric="TOTAL_SALES"):
    trial_data = metric_df[metric_df["STORE_NBR"] == trial_store]
    control_scores = {}
    for store in metric_df["STORE_NBR"].unique():
        if store == trial_store:
            continue
        control_data = metric_df[metric_df["STORE_NBR"] == store]
        merged = pd.merge(trial_data, control_data, on="MONTH",
suffixes=('_TRIAL', '_CONTROL'))
        corr =
merged[f"{metric}_TRIAL"].corr(merged[f"{metric}_CONTROL"])
        control_scores[store] = corr
    return sorted(control_scores.items(), key=lambda x: -x[1])[:3]  #
Top 3 similar stores


# Example: Find control store candidates for store 77
similar_stores_77 = calculate_similarity(77, monthly_metrics)
print("Top control store candidates for Store 77:", similar_stores_77)
```

```
# Step 3: Plot sales comparison between trial & control
def plot_trial_vs_control(trial_store, control_store,
metric="TOTAL_SALES"):
    trial = monthly_metrics[monthly_metrics["STORE_NBR"] ==
trial_store]
    control = monthly_metrics[monthly_metrics["STORE_NBR"] ==
control_store]
    combined = pd.merge(trial, control, on="MONTH",
suffixes=('_TRIAL', '_CONTROL'))

    plt.figure(figsize=(12,6))
    plt.plot(combined["MONTH"].astype(str),
combined[f"{metric}_TRIAL"], label=f"Trial Store {trial_store}")
    plt.plot(combined["MONTH"].astype(str),
combined[f"{metric}_CONTROL"], label=f"Control Store
{control_store}")
    plt.xticks(rotation=45)
    plt.title(f"{metric} Comparison: Store {trial_store} vs
{control_store}")
    plt.legend()
    plt.tight_layout()
    plt.show()

# Plot example
plot_trial_vs_control(77, similar_stores_77[0][0])
```