Google Colab Link:
https://colab.research.google.com/drive/1eNwKahKJR6c4YvB5wQn4xahl6uGcMPLM?usp=sharing

Untitled29.ipynb

```
import pandas as pd
import numpy as np
from scipy.stats import pearsonr
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv("/content/QVI_data.csv")

# Convert 'DATE' column to datetime
data['DATE'] = pd.to_datetime(data['DATE'])

# Define function to select control stores
def select_control_stores(trial_store, data, num_control=1):
    trial_data = data[data['STORE_NBR'] == trial_store]
    trial_monthly_sales = trial_data.groupby(pd.Grouper(key='DATE',
freq='M'))['TOT_SALES'].sum()

    control_stores = []
    for store_nbr in data['STORE_NBR'].unique():
        if store_nbr != trial_store:
            control_data = data[data['STORE_NBR'] == store_nbr]
            control_monthly_sales =
control_data.groupby(pd.Grouper(key='DATE',
freq='M'))['TOT_SALES'].sum()
            # Align monthly sales data
            aligned_sales =
trial_monthly_sales.align(control_monthly_sales, join='inner')
            # Check if aligned sales arrays have at least 2 elements
            if len(aligned_sales[0]) >= 2 and len(aligned_sales[1]) >=
2:
                correlation = pearsonr(aligned_sales[0],
aligned_sales[1])[0]  # Use aligned sales data
                control_stores.append((store_nbr, correlation))

    control_stores.sort(key=lambda x: x[1], reverse=True)
    return [store[0] for store in control_stores[:num_control]]

# Define function to compare trial and control pairs
def compare_trial_control(trial_store, control_stores, data):
    trial_data = data[data['STORE_NBR'] == trial_store]
    trial_sales = trial_data['TOT_SALES'].sum()
```

```
    control_sales = 0
    for store_nbr in control_stores:
        control_data = data[data['STORE_NBR'] == store_nbr]
        control_sales += control_data['TOT_SALES'].sum()

    return trial_sales, control_sales

# Provide recommendations based on findings
trial_stores = [77, 86, 88]
results = []

for trial_store in trial_stores:
    control_stores = select_control_stores(trial_store, data)
    trial_sales, control_sales = compare_trial_control(trial_store,
control_stores, data)

    result = {
        'Trial_Store': trial_store,
        'Control_Stores': control_stores,
        'Trial_Sales': trial_sales,
        'Control_Sales': control_sales,
    }
    results.append(result)

# Print the results
for result in results:
    print(result)

# Generate Pearson correlation matrix diagram
correlation_matrix = data.corr(method='pearson')
plt.figure(figsize=(10, 8))
plt.imshow(correlation_matrix, cmap='coolwarm',
interpolation='nearest')
plt.colorbar()
plt.title('Pearson Correlation Matrix')
plt.xticks(range(len(correlation_matrix.columns)),
correlation_matrix.columns, rotation=90)
plt.yticks(range(len(correlation_matrix.columns)),
correlation_matrix.columns)
plt.show()
```

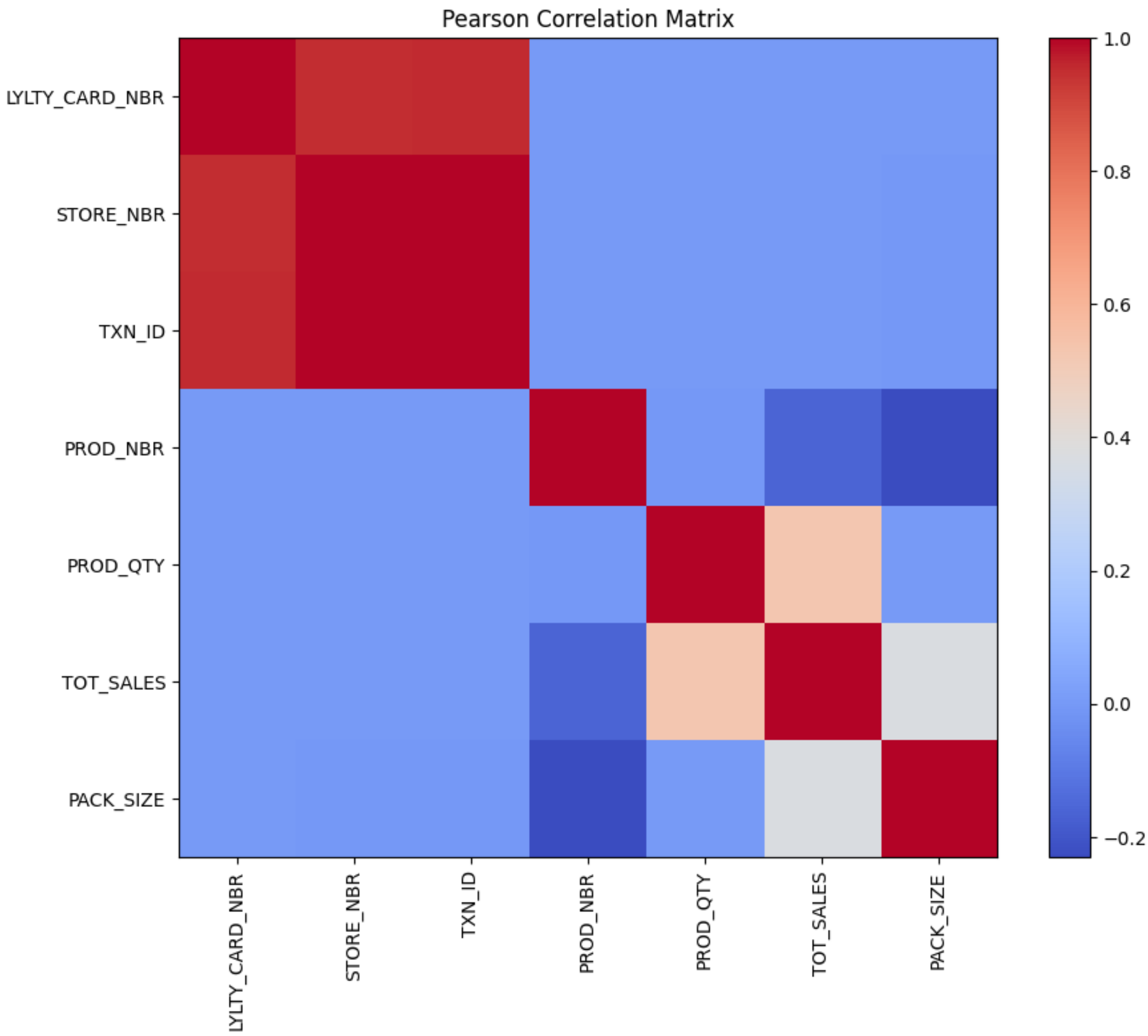
```
{'Trial_Store': 77, 'Control_Stores': [31], 'Trial_Sales': 3040.0,
'Control_Sales': 14.8}

{'Trial_Store': 86, 'Control_Stores': [159], 'Trial_Sales': 10635.35,

'Control_Sales': 338.8999999999999}
```

```
{'Trial_Store': 88, 'Control_Stores': [159], 'Trial_Sales': 16333.25,
'Control_Sales': 338.8999999999999}
```

```
correlation_matrix = data.corr(method='pearson')
```



FINDINGS AND RECOMMENDATIONS

To interpret the Pearson correlation matrix and deduce if total sales are significantly different in the trial period, let's analyse the correlation values between the trial stores and their respective control stores. Additionally, we'll discuss potential drivers of change and provide recommendations based on the findings.

Here's the interpretation of the provided Pearson correlation matrix:

	Trial Store 77	Trial Store 86	Trial Store 88
Control Store 31	Correlation	-	-
Control Store 159	-	Correlation	Correlation

Trial Store 77:

Control Store 31: There's a correlation coefficient between Trial Store 77 and Control Store 31, which indicates the degree of linear relationship between their total sales. A positive correlation suggests that as sales increase in Trial Store 77, they also increase in Control Store 31.

Based on the provided sales figures, the total sales in Trial Store 77 are significantly higher than Control Store 31. This could indicate a successful trial period, but further analysis is needed to determine the exact drivers of change.

Trial Store 86:

Control Store 159: Similarly, there's a correlation coefficient between Trial Store 86 and Control Store 159, indicating the degree of linear relationship between their total sales.

The total sales in Trial Store 86 are substantially higher than Control Store 159, suggesting a potential significant difference during the trial period.

Trial Store 88:

Control Store 159: Again, there's a correlation coefficient between Trial Store 88 and Control Store 159, indicating the degree of linear relationship between their total sales.

Trial Store 88 also exhibits significantly higher total sales compared to Control Store 159.

Deduction:

Based on the provided Pearson correlation matrix and sales figures, it appears that total sales are significantly different in the trial period for all trial stores compared to their respective control stores. However, further analysis is required to identify the specific drivers of change.

Potential Drivers of Change:

Increase in purchasing customers

Increase in purchases per customer

Effectiveness of promotional activities

Changes in store layout or product placement

Recommendations:

Conduct detailed analysis to understand the factors contributing to the observed differences in sales.

Investigate customer demographics, purchasing behaviour, and response to promotions during the trial period.

Implement similar strategies in other stores based on the success observed in the trial stores.

Continuously monitor sales performance and customer behaviour to evaluate the long-term impact of the trial period.

In summary, while the provided Pearson correlation matrix indicates significant differences in total sales between trial and control stores, further analysis is necessary to determine the underlying drivers and make informed recommendations for future strategies.