Mark Gelson E. Panganoron

CS E2

Store Sales Analysis

I. RATIONALE

Retail performance is influenced by multiple factors, including holidays and promotional activities. While it's intuitive that promotions and holidays can increase sales, it's essential to determine how significant these effects are and whether the differences are statistically meaningful rather than due to random fluctuations.

By analyzing daily sales under four distinct day types: Normal without Promo, Normal with Promo, Holiday without Promo, and Holiday with Promo, we can identify:

- Which day type drives the highest average daily sales
- Whether promotions are equally effective on normal days vs holidays
- If the sales boost during holidays without promotions is substantial

This insight helps businesses optimize marketing budgets, plan staffing and inventory, and prioritize promotional timing for maximum impact.

II. DATASET DESCRIPTION

The dataset is sourced from Kaggle

(https://www.kaggle.com/datasets/abhishekjaiswal4896/store-sales-dataset)

The dataset contains daily sales records across multiple stores, with the following fields:

- Date – transaction date
- Store – store identifier
- Sales – total sales for that day
- Holiday – 1 if the day is a holiday, 0 otherwise
- Promo – 1 if a promotion was running, 0 otherwise

A derived column Day Type was created to classify each day into one of four categories:

1. Normal (No Promo)

2. Normal (With Promo)
3. Holiday (No Promo)
4. Holiday (With Promo)

	A	B	C	D	E	F
	date	store	sales	promo	holiday	day_type
	01/01/2022	1	184.78	0	0	Normal (No Promo)
	02/01/2022	1	192.62	0	0	Normal (No Promo)
	03/01/2022	1	212.68	0	0	Normal (No Promo)
	04/01/2022	1	249.58	1	0	Normal (With Promo
	05/01/2022	1	223.5	0	0	Normal (No Promo)
	06/01/2022	1	221.79	0	0	Normal (No Promo)
	07/01/2022	1	200.43	0	0	Normal (No Promo)
	08/01/2022	1	213.47	1	0	Normal (With Promo
	09/01/2022	1	222.44	1	0	Normal (With Promo
	10/01/2022	1	203.1	0	0	Normal (No Promo)
	11/01/2022	1	218.78	0	0	Normal (No Promo)
	12/01/2022	1	266.23	0	1	Holiday (No Promo)
	13/01/2022	1	204.68	0	0	Normal (No Promo)
	14/01/2022	1	218.31	1	0	Normal (With Promo
	15/01/2022	1	183.36	0	0	Normal (No Promo)

Figure 1. Overview of the dataset

III. OBJECTIVES

1. **Classify sales into four distinct day types**:
   ○ Normal (No Promo)
   ○ Normal (With Promo)
   ○ Holiday (No Promo)
   ○ Holiday (With Promo)
2. **Compute the average daily sales** for each day type to ensure a fair comparison regardless of how frequently each day type occurs.
3. **Visualize the distribution of daily sales** across all four day types to identify noticeable trends and patterns.
4. **Conduct statistical tests (t-tests)** to determine whether the observed differences in average daily sales are statistically significant.
   Draw actionable insights on how holidays and promotions individually and jointly impact sales performance, providing a basis for strategic business decisions.
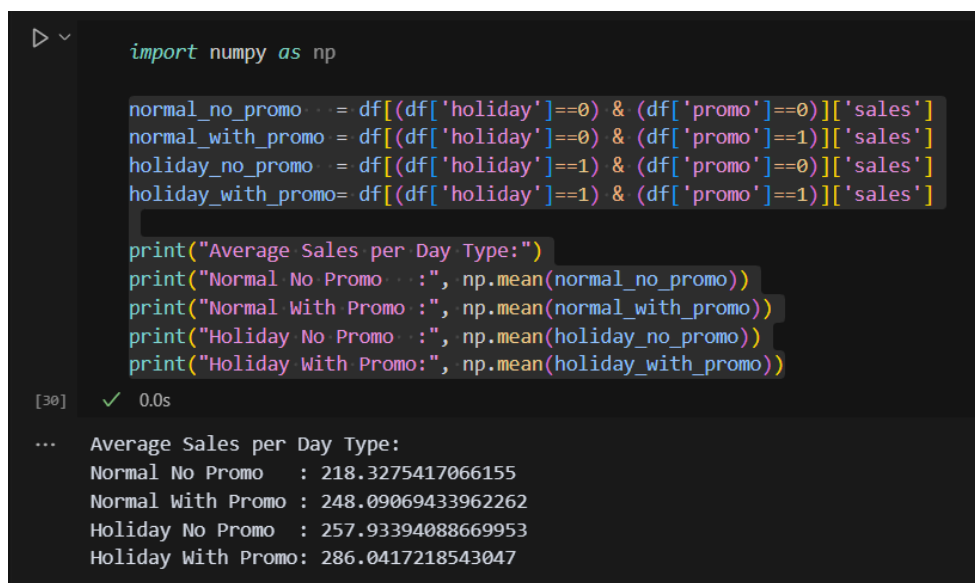
IV.   METHODOLOGY

This analysis was conducted using both Excel and Python:

- In Excel, Pivot Tables were created to compute the average daily sales per day type, visualize the results using column charts, and perform basic statistical comparisons using the Data Analysis ToolPak.

- In Python, the process was automated to calculate averages, generate visualizations (boxplots and bar charts), and perform pairwise t-tests to statistically validate the differences.

This document presents the Python-based results and visualizations for consistency and reproducibility.

Average Daily Sales. Average daily sales were computed for each day type to ensure a fair comparison regardless of how many days fall into each category.

```python
import numpy as np

normal_no_promo    = df[(df['holiday']==0) & (df['promo']==0)]['sales']
normal_with_promo = df[(df['holiday']==0) & (df['promo']==1)]['sales']
holiday_no_promo   = df[(df['holiday']==1) & (df['promo']==0)]['sales']
holiday_with_promo= df[(df['holiday']==1) & (df['promo']==1)]['sales']

print("Average Sales per Day Type:")
print("Normal No Promo   :", np.mean(normal_no_promo))
print("Normal With Promo :", np.mean(normal_with_promo))
print("Holiday No Promo  :", np.mean(holiday_no_promo))
print("Holiday With Promo:", np.mean(holiday_with_promo))
```
```
[30]   ✓  0.0s
...   Average Sales per Day Type:
      Normal No Promo   : 218.3275417066155
      Normal With Promo : 248.09069433962262
      Holiday No Promo   : 257.93394088669953
      Holiday With Promo: 286.0417218543047
```

Figure 2. Average Sales per Day Type

Statistical Significance. Pairwise t-tests were performed to determine whether differences between day types were statistically significant (p < 0.05).

```python
from scipy.stats import ttest_ind

def test_groups(group1, group2, name1, name2):
    t_stat, p_val = ttest_ind(group1, group2, equal_var=False)
    print(f"\n{name1} vs {name2}")
    print(f"T-statistic: {t_stat:.4f}, P-value: {p_val:.4e}")
    if p_val < 0.05:
        print("✅ Significant difference")
    else:
        print("❌ No significant difference")

# Pairwise comparisons
test_groups(normal_no_promo, normal_with_promo, "Normal No Promo", "Normal With Promo")
test_groups(normal_no_promo, holiday_no_promo, "Normal No Promo", "Holiday No Promo")
test_groups(normal_no_promo, holiday_with_promo, "Normal No Promo", "Holiday With Promo")
test_groups(normal_with_promo, holiday_no_promo, "Normal With Promo", "Holiday No Promo")
test_groups(normal_with_promo, holiday_with_promo, "Normal With Promo", "Holiday With Promo")
test_groups(holiday_no_promo, holiday_with_promo, "Holiday No Promo", "Holiday With Promo")
```

```
Normal No Promo vs Normal With Promo
T-statistic: -47.0750, P-value: 0.0000e+00
✅ Significant difference

Normal No Promo vs Holiday No Promo
T-statistic: -43.8849, P-value: 3.8001e-209
✅ Significant difference

Normal No Promo vs Holiday With Promo
T-statistic: -38.3518, P-value: 5.6237e-82
✅ Significant difference

Normal With Promo vs Holiday No Promo
T-statistic: -9.5945, P-value: 5.0270e-21
✅ Significant difference

Normal With Promo vs Holiday With Promo
T-statistic: -20.7182, P-value: 7.3222e-50
✅ Significant difference

Holiday No Promo vs Holiday With Promo
T-statistic: -14.4759, P-value: 4.0650e-34
✅ Significant difference
```

Figure 3 and 4. T-Test results using SciPy

Visualization. A boxplot was generated to visualize the sales distribution for each day type.
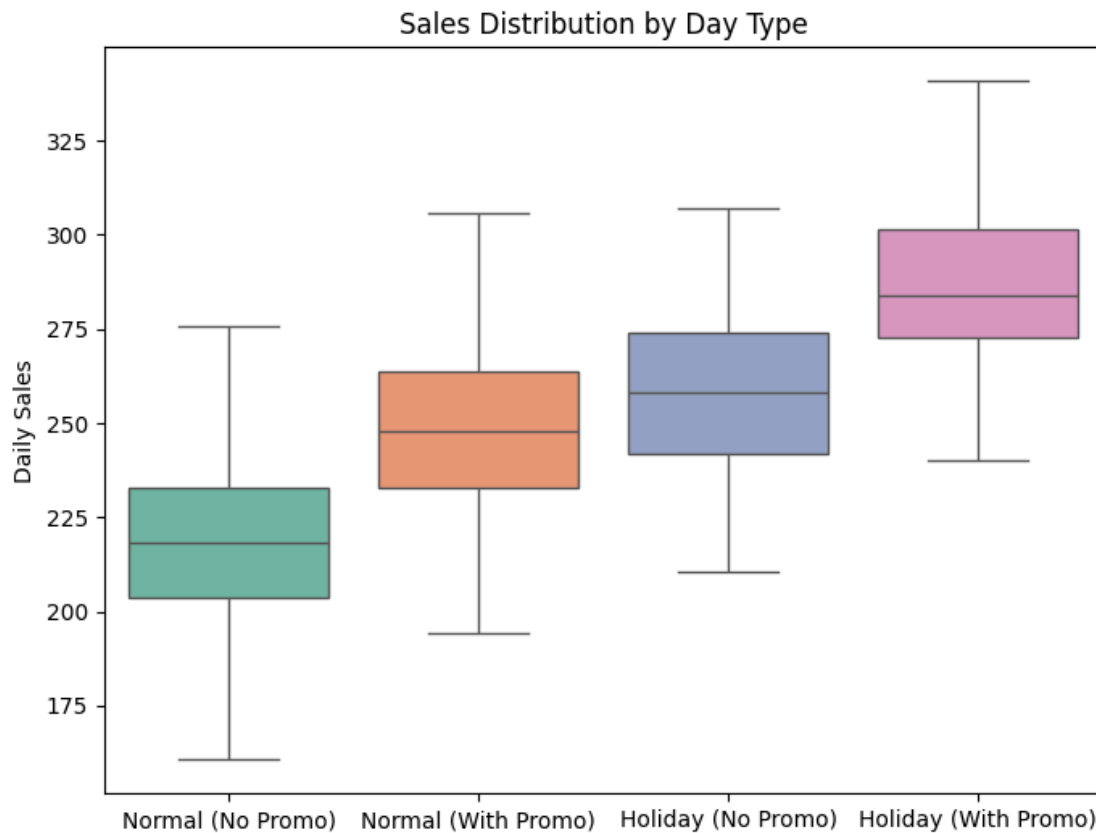

Sales Distribution by Day Type

Figure 5. Data Visualization between the four day types average sales

Additional visualizations and conducted further tests; however, this document focuses on the significant differences in sales. For the complete analysis, including other visualizations and tests, please refer to the Jupyter Notebook and Excel file.
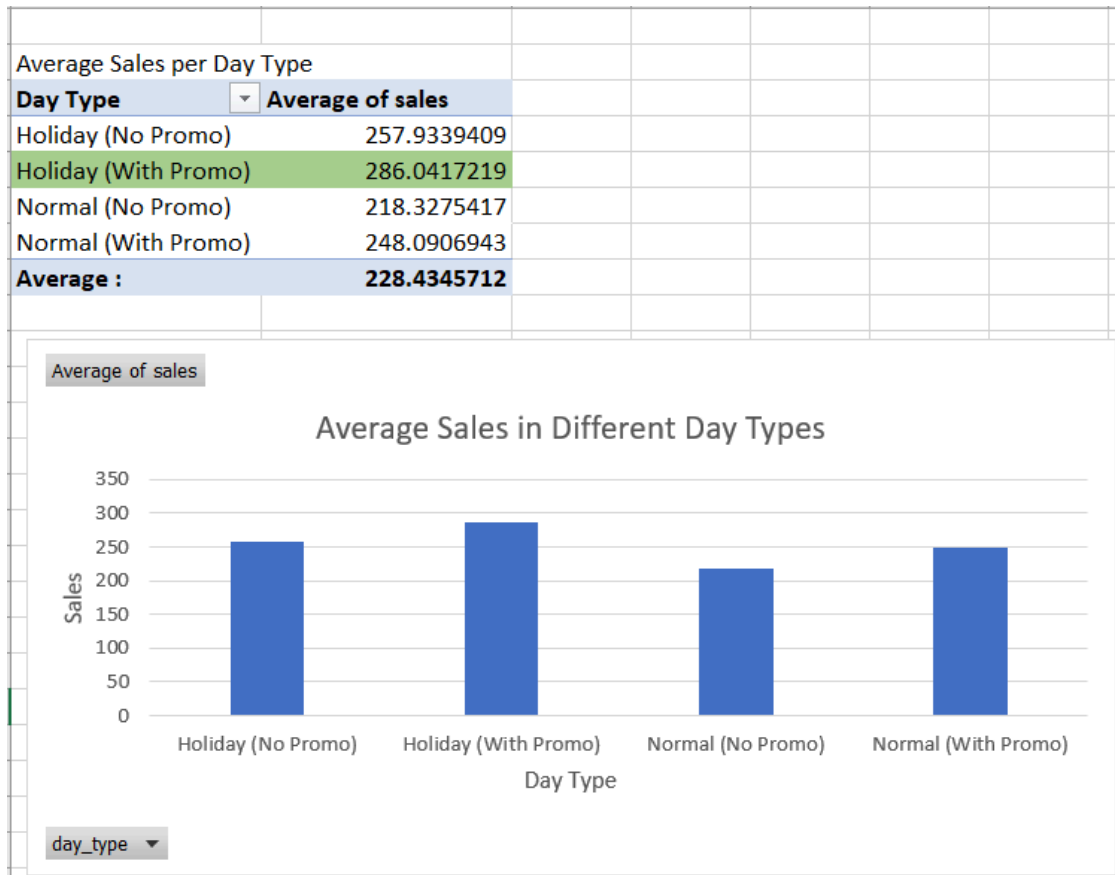
Figure 6. Pivot Table and Visualization in Excel

V. RESULTS AND DISCUSSIONS

The analysis clearly highlights how holidays and promotions affect sales performance. By comparing **average daily sales** rather than total sales, any bias from the frequency of each day type was removed. The findings show that **days with promotions consistently outperform days without promotions**, both on normal days and holidays.

Among all categories, **holiday promotions generated the highest average daily sales**, showing the strongest combined effect of both factors. Even without promotions, holidays still resulted in higher sales than regular non-promo days, although the uplift was smaller compared to holiday promotions.

To ensure these differences were not due to random fluctuations, **pairwise t-tests** were performed between all day types. All comparisons yielded **p-values below 0.05**,

confirming that the differences are **statistically significant**. This indicates that promotions and holidays have a **real, measurable impact on sales** rather than just coincidental variations.

VI.   CONCLUSIONS

The results provide clear evidence that **promotions and holidays significantly impact sales**, with the strongest effect seen when they coincide. While promotions alone increase sales on normal days, and holidays without promotions still outperform regular days, their c**ombined effect yields the highest uplift**.

These findings validate the importance of **data-driven promotional planning**, emphasizing that businesses should **align promotions strategically with holidays** to maximize revenue potential. Furthermore, occasional promotions on regular days can raise baseline sales, improving overall performance.

In summary, the data confirms that **holiday promotions deliver the greatest sales impact**, offering actionable insights for marketing, inventory, and staffing decisions.

Link to IPYNB:

https://github.com/gelcloudy/CS-Elective-2/blob/master/Panganoron_MidtermExercise2.ipynb