

IST 652 – SCRIPTING FOR DATA ANALYSIS

INSIGHTS FROM LOCAL BOOKSTORE SALES DATA

BOOKS END

BOOKSHOP

Leveraging Data Analytics for Business Growth

By Kishan Rathor

Under the guidance of - Professor Angela Ramnarine Rieks

INTRODUCTION



- Established in the early 1980s, Books End stands as Syracuse's oldest used bookstore, boasting a legacy of community trust and a diverse collection spanning over three decades.
- Patrick McGrath, a friend and the new owner as of August 2023, is determined to intertwine the bookstore's storied past with a data-informed future.
- Embracing modernity, their goal is to harness cutting-edge data analytics to refine business strategies, enhance customer engagement, and catalyze growth, ensuring that Books End remains a cornerstone for book lovers in Syracuse for years to come.



BUSINESS QUESTIONS

- How can we tailor promotional strategies to encourage the use of payment methods that are less frequently used?
- How could staffing and inventory management be adjusted to accommodate the weekly sales cycle?
- What are the potential causes of sales peaks and troughs throughout the period?
- How do monthly sales trends align with marketing efforts and inventory changes?
- How can holiday marketing strategies be improved based on sales on Holidays?
- How can we use the variability in quantity and total collected to optimize pricing strategies and inventory stocking?
- Can promotional activities or special events be scheduled during lower-activity times to increase sales?

A LOOK AT THE DATA SET

	A	B	C	D	E	F	G	H	I	J
1	Date	Time	Item	Qty	SKU	Gross Sales	Discounts	Net Sales	Tax	Unit
2	11/16/23	14:44:31	Custom Amount	1		\$3.00	\$0.00	\$3.00	\$0.24	ea
3	11/16/23	14:44:31	Custom Amount	1		\$3.00	\$0.00	\$3.00	\$0.24	ea
4	11/16/23	14:44:31	364131	1	Turkeys, Pilgrims, and Indian Corn:	\$4.00	\$0.00	\$4.00	\$0.32	ea
5	11/16/23	14:44:31	458418	1	Child Life in Colonial Days	\$20.00	\$0.00	\$20.00	\$1.60	ea
6	11/16/23	14:44:31	Custom Amount	3		\$9.00	\$0.00	\$9.00	\$0.72	ea
7	11/16/23	14:09:05	Custom Amount	1		\$5.00	\$0.00	\$5.00	\$0.40	ea
8	11/16/23	14:09:05	Custom Amount	1		\$0.50	\$0.00	\$0.50	\$0.04	ea
9	11/16/23	13:43:47	Custom Amount	1		\$3.50	\$0.00	\$3.50	\$0.28	ea
10	11/16/23	13:43:47	458452	1	Look At Me Now and Here I Am:	\$5.00	\$0.00	\$5.00	\$0.40	ea
11	11/16/23	13:43:47	Custom Amount	1		\$0.50	\$0.00	\$0.50	\$0.04	ea
12	11/16/23	13:43:47	Custom Amount	1		\$1.99	\$0.00	\$1.99	\$0.16	ea
13	11/16/23	13:15:29	467181	1	The New York State Lunatic Asylum At	\$10.00	\$0.00	\$10.00	\$0.80	ea
14	11/16/23	13:06:19	463741	1	The Motivation Manifesto:	\$5.00	\$0.00	\$5.00	\$0.40	ea
15	11/16/23	13:06:19	439258	1	The Astonishing Power of Emotions:	\$5.00	\$0.00	\$5.00	\$0.40	ea
16	11/16/23	13:06:19	Custom Amount	1		\$20.00	\$0.00	\$20.00	\$1.60	ea
17	11/16/23	13:02:06	Custom Amount	1		\$4.00	\$0.00	\$4.00	\$0.32	ea
18	11/16/23	12:37:18	466720	1	The Venus Hottentot	\$5.00	\$0.00	\$5.00	\$0.40	ea
19	11/16/23	12:37:18	453477	1	Straight James / Gay James	\$4.50	\$0.00	\$4.50	\$0.36	ea
20	11/16/23	12:37:18	455108	1	Memories of My Melancholy Whores	\$7.50	\$0.00	\$7.50	\$0.60	ea
21	11/16/23	12:37:18	Custom Amount	1		\$4.50	\$0.00	\$4.50	\$0.36	ea
22	11/16/23	12:23:34	371116	1	A Conspiracy of Faith:	\$5.00	\$0.00	\$5.00	\$0.40	ea
23	11/16/23	12:23:34	460444	1	In a House of Lies	\$6.00	\$0.00	\$6.00	\$0.48	ea
24	11/16/23	12:23:34	461658	1	The Jealousy Man and Other Stories	\$10.00	\$0.00	\$10.00	\$0.80	ea
25	11/16/23	11:14:28	Custom Amount	1		\$7.50	\$0.00	\$7.50	\$0.60	ea
26	11/16/23	11:14:28	466926	1	Synchronicity:	\$7.00	\$0.00	\$7.00	\$0.56	ea
27	11/16/23	11:14:28	342488	1	Light :	\$5.00	\$0.00	\$5.00	\$0.40	ea
28	11/16/23	11:14:28	342485	1	Light.	\$5.00	\$0.00	\$5.00	\$0.40	ea
29	11/16/23	11:13:13	Custom Amount	1		\$5.00	\$0.00	\$5.00	\$0.40	ea
30	11/16/23	11:13:13	Custom Amount	1		\$4.00	\$0.00	\$4.00	\$0.32	ea
31	11/15/23	17:44:10	390984	1	Swing Time	\$6.00	\$0.00	\$6.00	\$0.48	ea
32	11/15/23	17:44:10	460105	1	You Say to Brick: the Life of Louis Kahr	\$12.50	\$0.00	\$12.50	\$1.00	ea
33	11/15/23	17:44:10	448212	1	Re-Architecture:	\$10.00	\$0.00	\$10.00	\$0.80	ea
34	11/15/23	17:44:10	53545	1	Buildings Reborn	\$4.00	\$0.00	\$4.00	\$0.32	ea
35	11/15/23	17:44:10	459364	1	Estate Gardens of California	\$10.00	\$0.00	\$10.00	\$0.80	ea
36	11/15/23	17:44:10	Custom Amount	1		\$1.99	\$0.00	\$1.99	\$0.16	ea
37	11/15/23	17:44:10	Custom Amount	1		\$4.00	\$0.00	\$4.00	\$0.32	ea
38	11/15/23	17:44:10	Custom Amount	1		\$2.00	\$0.00	\$2.00	\$0.16	ea
39	11/15/23	17:02:58	Custom Amount	1		\$4.00	\$0.00	\$4.00	\$0.32	ea
40	11/15/23	17:02:58	Custom Amount	1		\$3.00	\$0.00	\$3.00	\$0.24	ea
41	11/15/23	17:02:58	465352	1	Hellboy and the B. P. R. D:	\$5.00	\$0.00	\$5.00	\$0.40	ea
42	11/15/23	17:02:58	465354	1	Hellboy and the B. P. R. D. :	\$5.00	\$0.00	\$5.00	\$0.40	ea
43	11/15/23	17:02:58	466654	1	The Fall, Volume 1	\$6.00	\$0.00	\$6.00	\$0.48	ea
44	11/15/23	17:02:58	414076	1	Gustav, P. I.	\$6.00	\$0.00	\$6.00	\$0.48	ea
45	11/15/23	16:14:33	227103	1	The Master:	\$6.00	\$0.00	\$6.00	\$0.48	ea
46	11/15/23	15:49:54	Custom Amount	1		\$3.00	\$0.00	\$3.00	\$0.24	ea

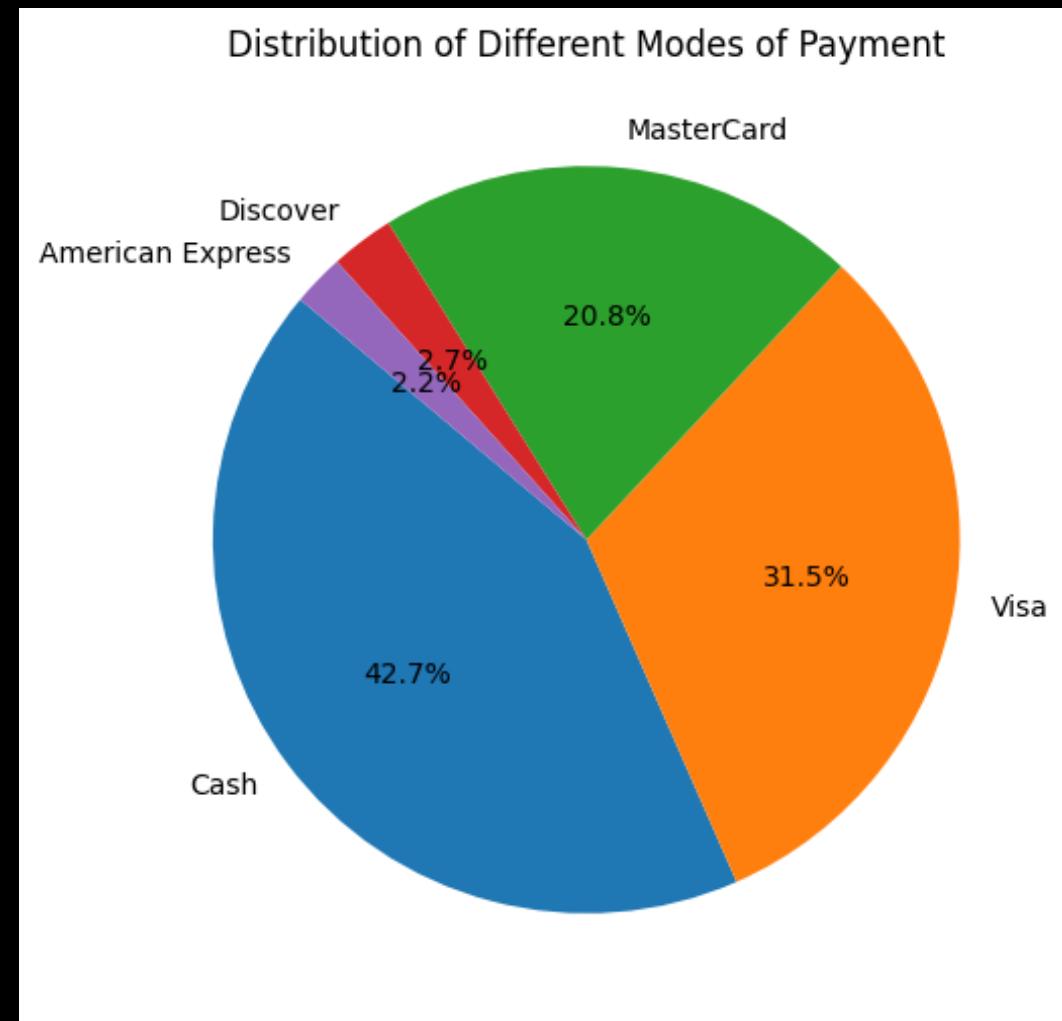
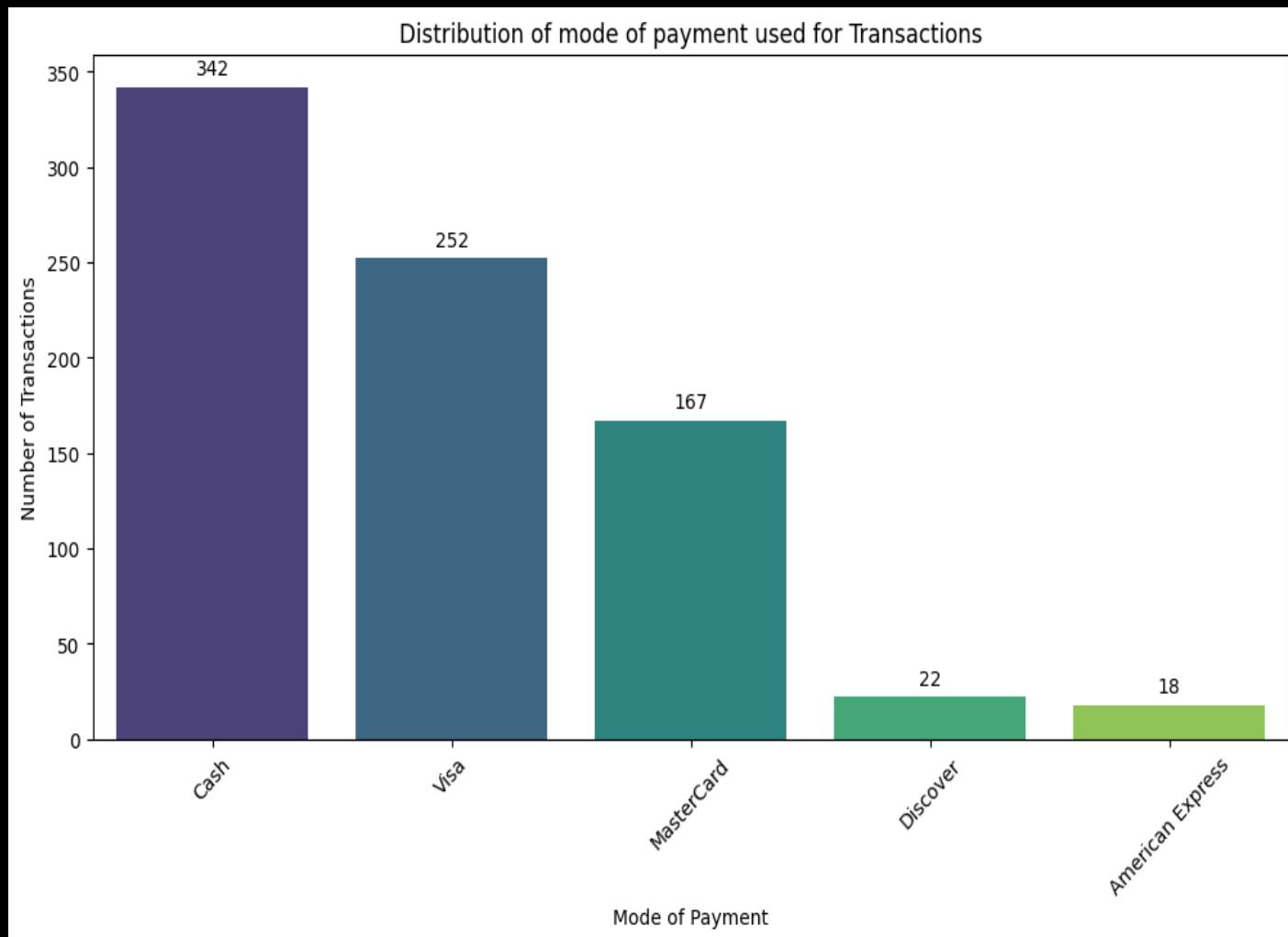
- **Dataset Name:** New Data
- **Entry Count:** 4448 transactions, indexed from 0 to 4447.
- **Data Types:** Mixed types including dates (datetime64[ns]), strings (object), integers (int64), and floating-point numbers (float64).
- **Key Columns:**
 - Date: Date of transaction.
 - Time: Time of transaction.
 - Item: Description or name of the item sold.
 - Qty: Quantity of items sold in the transaction.
 - SKU: Stock Keeping Unit or unique identifier for items.
 - Gross Sales: Total sales amount before discounts.
 - Discounts: Discounts applied to the sales.
 - Net Sales: Sales after discounts, before tax.
 - Tax: Tax amount applied to the net sales.
 - Unit: Unit of measurement or sale.
- **Usage:** Utilized to assess current sales trends, inventory management, and the effectiveness of pricing strategies.

Date	Time	Gross Sales	Discounts	Service Charges	Net Sales	Gift Card Sales	Tax	Total Collected	Card	Cash	Other Tender Type	Other Tender Note	Fees	Net Total	Card Brand	Description	Initial Fees	Adjustment Fees	Fee Plan	Fee Percentage Rate	Fee Fixed Rate	
2023-09-28	13:20:34	\$10.00	\$0.00	\$0.00	\$10.00	\$0.00	\$0.80	\$10.80	\$10.80	\$0.00			-\$0.37	\$10.43	Discover		447728	-\$0.37	\$0.00	Blended Rate	2.5	\$0.10
2023-09-28	10:45:29	\$47.50	\$0.00	\$0.00	\$47.50	\$0.00	\$3.80	\$51.30	\$0.00	\$51.30			\$0.00	\$51.30		Custom Amount, Custom Amount, Custom Amount, Custom Amount, 460572, Custom Amount, 442400, Custom Amount, 423942	\$0.00	\$0.00				
2023-09-28	10:17:10	\$6.00	\$0.00	\$0.00	\$6.00	\$0.00	\$0.48	\$6.48	\$0.00	\$6.48			\$0.00	\$6.48			442446	\$0.00	\$0.00			
2023-09-27	15:52:00	\$183.53	\$0.00	\$0.00	\$183.53	\$0.00	\$14.68	\$198.21	\$0.00	\$198.21			\$0.00	\$198.21		444947, 417218, 463760, 459484, 448267, 411822, 78857, Custom Amount, 412028, 128495, 407625, 460500, Custom Amount	\$0.00	\$0.00				
2023-09-27	14:32:14	\$16.00	\$0.00	\$0.00	\$16.00	\$0.00	\$1.28	\$17.28	\$17.28	\$0.00			-\$0.53	\$16.75	American Express	463530, 292665, 413966		-\$0.53	\$0.00	Blended Rate	2.5	\$0.10
2023-09-27	14:30:45	\$1.50	\$0.00	\$0.00	\$1.50	\$0.00	\$0.12	\$1.62	\$0.00	\$1.62			\$0.00	\$1.62			391809	\$0.00	\$0.00			
2023-09-27	13:48:59	\$9.00	\$0.00	\$0.00	\$9.00	\$0.00	\$0.72	\$9.72	\$0.00	\$9.72			\$0.00	\$9.72		Custom Amount		\$0.00	\$0.00			
2023-09-27	13:22:29	\$6.00	\$0.00	\$0.00	\$6.00	\$0.00	\$0.48	\$6.48	\$6.48	\$0.00			-\$0.26	\$6.22	MasterCard		462063	-\$0.26	\$0.00	Blended Rate	2.5	\$0.10
2023-09-27	11:40:41	\$15.50	\$0.00	\$0.00	\$15.50	\$0.00	\$1.24	\$16.74	\$0.00	\$16.74			\$0.00	\$16.74		462646, 440825, Custom Amount		\$0.00	\$0.00			
2023-09-27	11:39:21	\$44.00	\$0.00	\$0.00	\$44.00	\$0.00	\$3.52	\$47.52	\$47.52	\$0.00			-\$1.29	\$46.23	Visa	Custom Amount, 2 x 148123, 454781		-\$1.29	\$0.00	Blended Rate	2.5	\$0.10
2023-09-27	10:26:13	\$27.50	\$0.00	\$0.00	\$27.50	\$0.00	\$2.20	\$29.70	\$29.70	\$0.00			-\$0.84	\$28.86	Visa	Custom Amount		-\$0.84	\$0.00	Blended Rate	2.5	\$0.10
2023-09-27	10:04:53	\$9.00	\$0.00	\$0.00	\$9.00	\$0.00	\$0.72	\$9.72	\$0.00	\$9.72			\$0.00	\$9.72		Custom Amount		\$0.00	\$0.00			
2023-09-26	17:36:38	\$37.50	\$0.00	\$0.00	\$37.50	\$0.00	\$0.00	\$37.50	\$37.50	\$0.00			-\$1.04	\$36.46	Visa	457223, 449773, 448780, 463034, Custom Amount, 453202		-\$1.04	\$0.00	Blended Rate	2.5	\$0.10
2023-09-26	17:25:07	\$10.00	\$0.00	\$0.00	\$10.00	\$0.00	\$0.80	\$10.80	\$0.00	\$10.80			\$0.00	\$10.80		Custom Amount, Custom Amount, Custom Amount		\$0.00	\$0.00			
2023-09-26	16:14:03	\$5.00	\$0.00	\$0.00	\$5.00	\$0.00	\$0.40	\$5.40	\$0.00	\$5.40			\$0.00	\$5.40			152701	\$0.00	\$0.00			
2023-09-26	14:41:26	\$6.00	\$0.00	\$0.00	\$6.00	\$0.00	\$0.48	\$6.48	\$6.48	\$0.00			-\$0.26	\$6.22	Visa		299412	-\$0.26	\$0.00	Blended Rate	2.5	\$0.10
2023-09-26	12:23:09	\$17.00	\$0.00	\$0.00	\$17.00	\$0.00	\$1.36	\$18.36	\$18.36	\$0.00			-\$0.56	\$17.80	American Express	463245, 463275		-\$0.56	\$0.00	Blended Rate	2.5	\$0.10
2023-09-26	11:03:45	\$10.00	\$0.00	\$0.00	\$10.00	\$0.00	\$0.80	\$10.80	\$10.80	\$0.00			-\$0.37	\$10.43	MasterCard	467482, 425098		-\$0.37	\$0.00	Blended Rate	2.5	\$0.10
2023-09-26	10:32:48	\$8.00	\$0.00	\$0.00	\$8.00	\$0.00	\$0.64	\$8.64	\$0.00	\$8.64			\$0.00	\$8.64		Custom Amount		\$0.00	\$0.00			
2023-09-25	16:54:27	\$8.00	\$0.00	\$0.00	\$8.00	\$0.00	\$0.64	\$8.64	\$0.00	\$8.64			\$0.00	\$8.64		Custom Amount, Custom Amount, Custom Amount, Custom Amount		\$0.00	\$0.00			
2023-09-25	16:46:11	\$4.50	\$0.00	\$0.00	\$4.50	\$0.00	\$0.36	\$4.86	\$0.00	\$4.86			\$0.00	\$4.86		Custom Amount		\$0.00	\$0.00			
2023-09-25	16:31:16	\$7.00	\$0.00	\$0.00	\$7.00	\$0.00	\$0.56	\$7.56	\$7.56	\$0.00			-\$0.29	\$7.27	Visa	Custom Amount, Custom Amount		-\$0.29	\$0.00	Blended Rate	2.5	\$0.10
2023-09-25	15:42:25	\$10.00	\$0.00	\$0.00	\$10.00	\$0.00	\$0.80	\$10.80	\$0.00	\$10.80			\$0.00	\$10.80		466075, 466076		\$0.00	\$0.00			
2023-09-25	14:58:14	\$5.50	\$0.00	\$0.00	\$5.50	\$0.00	\$0.36	\$5.86	\$5.86	\$0.00			-\$0.25	\$5.61	American Express	Bookmark (Regular), Custom Amount		-\$0.25	\$0.00	Blended Rate	2.5	\$0.10
2023-09-25	14:37:12	\$31.00	\$0.00	\$0.00	\$31.00	\$0.00	\$2.32	\$33.32	\$33.32	\$0.00			-\$0.93	\$32.39	MasterCard	465401, Custom Amount, 2 x Bookmark (Regular), Custom Amount, Custom Amount, Custom Amount - JMM		-\$0.93	\$0.00	Blended Rate	2.5	\$0.10
2023-09-25	14:15:56	\$7.20	\$0.00	\$0.00	\$7.20	\$0.00	\$0.58	\$7.78	\$7.78	\$0.00			-\$0.29	\$7.49	Visa	Custom Amount		-\$0.29	\$0.00	Blended Rate	2.5	\$0.10
2023-09-25	13:36:36	\$23.50	\$0.00	\$0.00	\$23.50	\$0.00	\$1.88	\$25.38	\$0.00	\$25.38			\$0.00	\$25.38		320630, 320636, Custom Amount		\$0.00	\$0.00			
2023-09-25	13:24:45	\$28.50	\$0.00	\$0.00	\$28.50	\$0.00	\$2.28	\$30.78	\$0.00	\$30.78			\$0.00	\$30.78		Custom Amount, 451267		\$0.00	\$0.00			
2023-09-25	13:09:52	\$5.00	\$0.00	\$0.00	\$5.00	\$0.00	\$0.40	\$5.40	\$5.40	\$0.00			-\$0.24	\$5.16	Visa		464534	-\$0.24	\$0.00	Blended Rate	2.5	\$0.10
2023-09-25	12:53:11	\$15.50	\$0.00	\$0.00	\$15.50	\$0.00	\$1.24	\$16.74	\$16.74	\$0.00			-\$0.52	\$16.22	American Express	428780, 231750, 465348		-\$0.52	\$0.00	Blended Rate	2.5	\$0.10

- Dataset Name:** Old Data
- Entry Count:** 801 transactions, indexed from 0 to 800.
- Data Type:** Primarily categorical and monetary values, with most fields as strings (object) and one numeric (float64).
- Key Columns Retained:**
 - Date: Transaction dates.
 - Time: Transaction times.
 - Net Sales: Sales after discounts, excluding tax.
 - Total Collected: Final amount collected, including tax.
 - Card Brand: Type of card used, if applicable.
- Usage:** Analyzed for insights into historical sales patterns, customer payment methods, and overall business performance before the implementation of new strategies.

DATA CLEANING AND PREPARATION

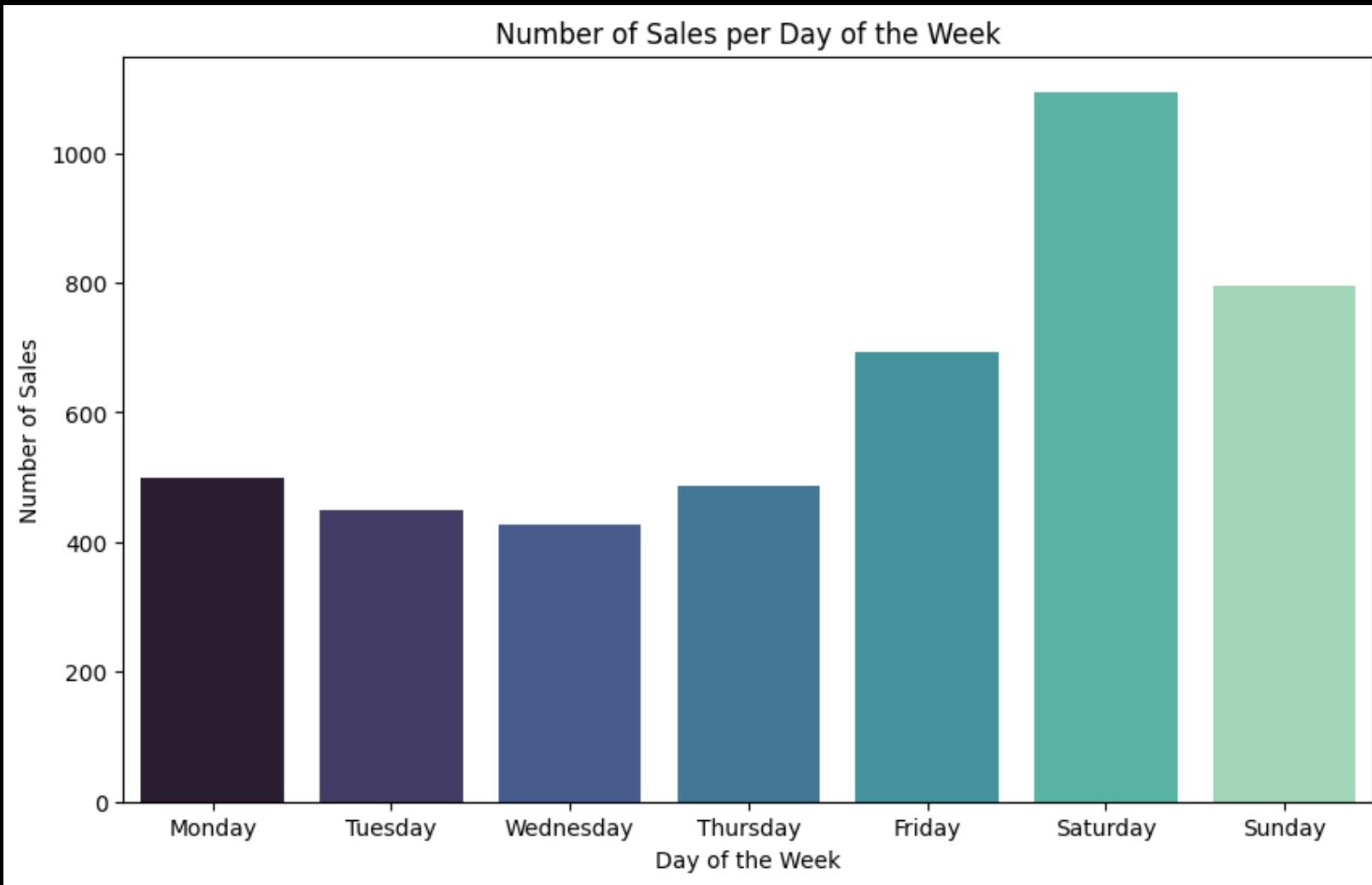
PREFERRED MODE OF PAYMENT



PREFERRED MODE OF PAYMENT

- I noticed varying usage of different payment methods.
- Determined the most popular payment methods and identify opportunities.
- Analyzed transaction data for payment trends.
- As you can see, cash and Visa are the most common payment methods, with over 300 and 250 transactions, respectively.
- This suggests a customer preference that we can leverage by introducing cash-back or loyalty programs, especially for Visa users, to increase retention and sales."

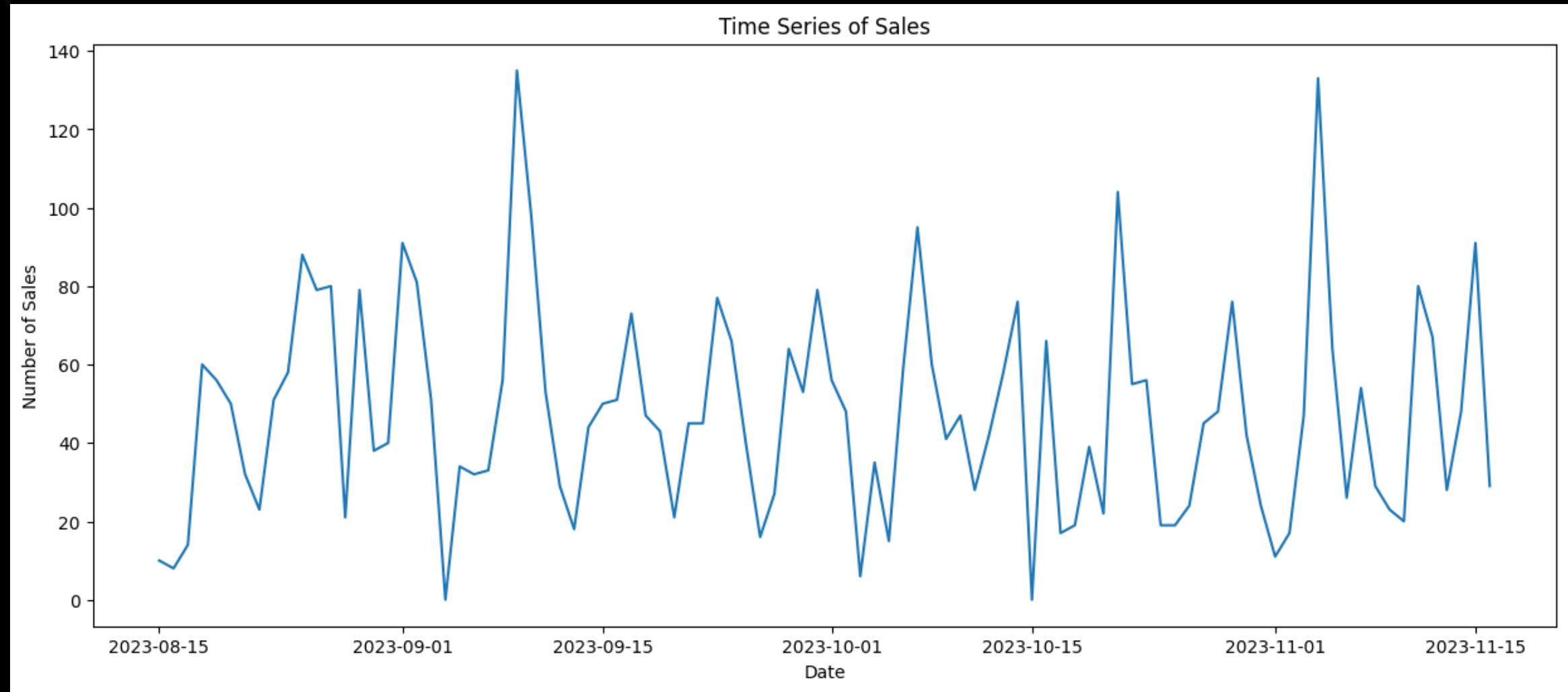
SALES PER DAY OF THE WEEK



SALES PER DAY OF THE WEEK

- Sales fluctuate throughout the week.
- Understood which days drive the most traffic and sales.
- Compile daily sales data to identify patterns.
- Saturday shows the highest number of sales, peaking near 1000 transactions, while Monday has the lowest.
- To capitalize on this, we could offer weekend promotions or events to further boost high-traffic days and introduce special Monday deals to improve sales on slower days.

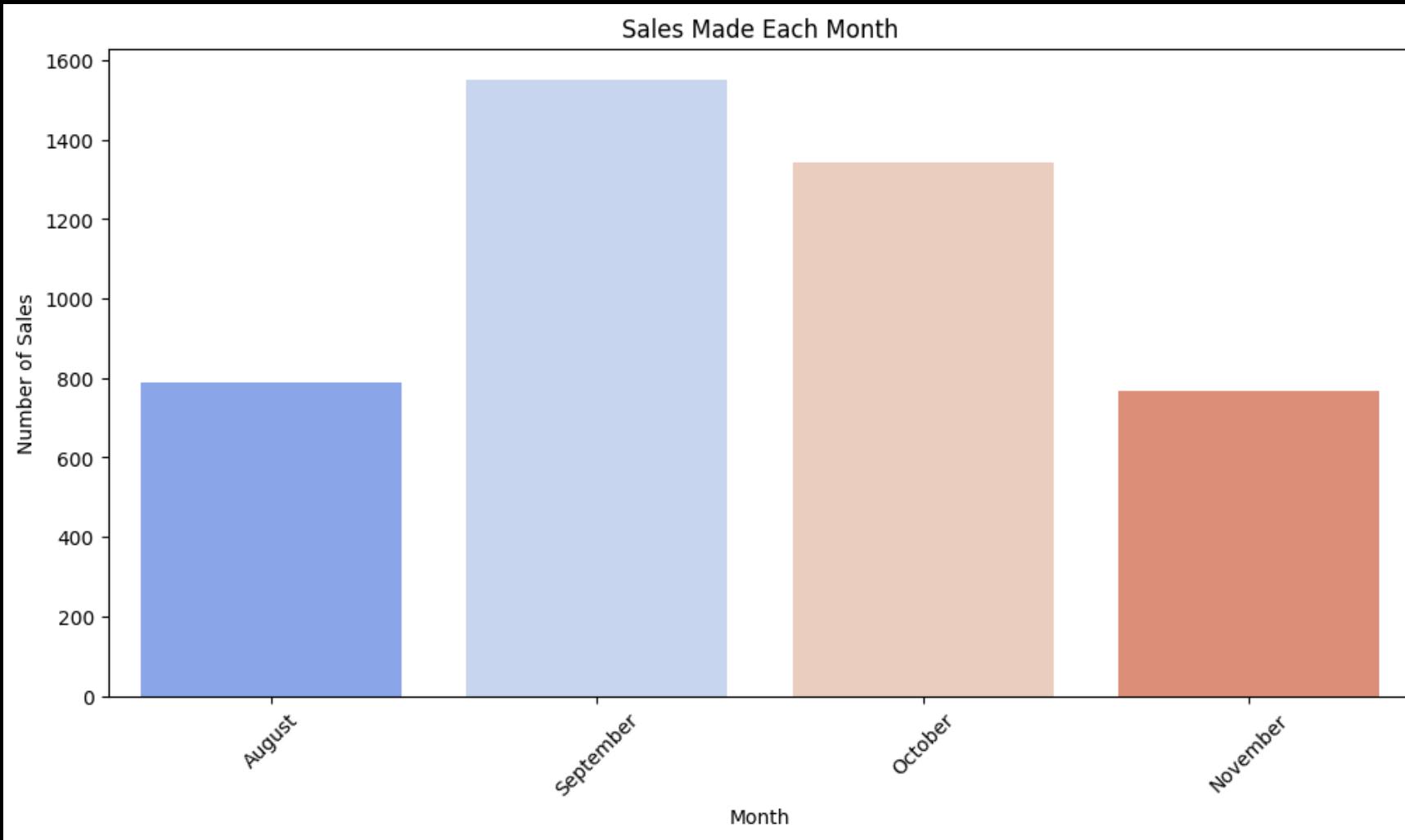
TIME SERIES ANALYSIS OF SALES



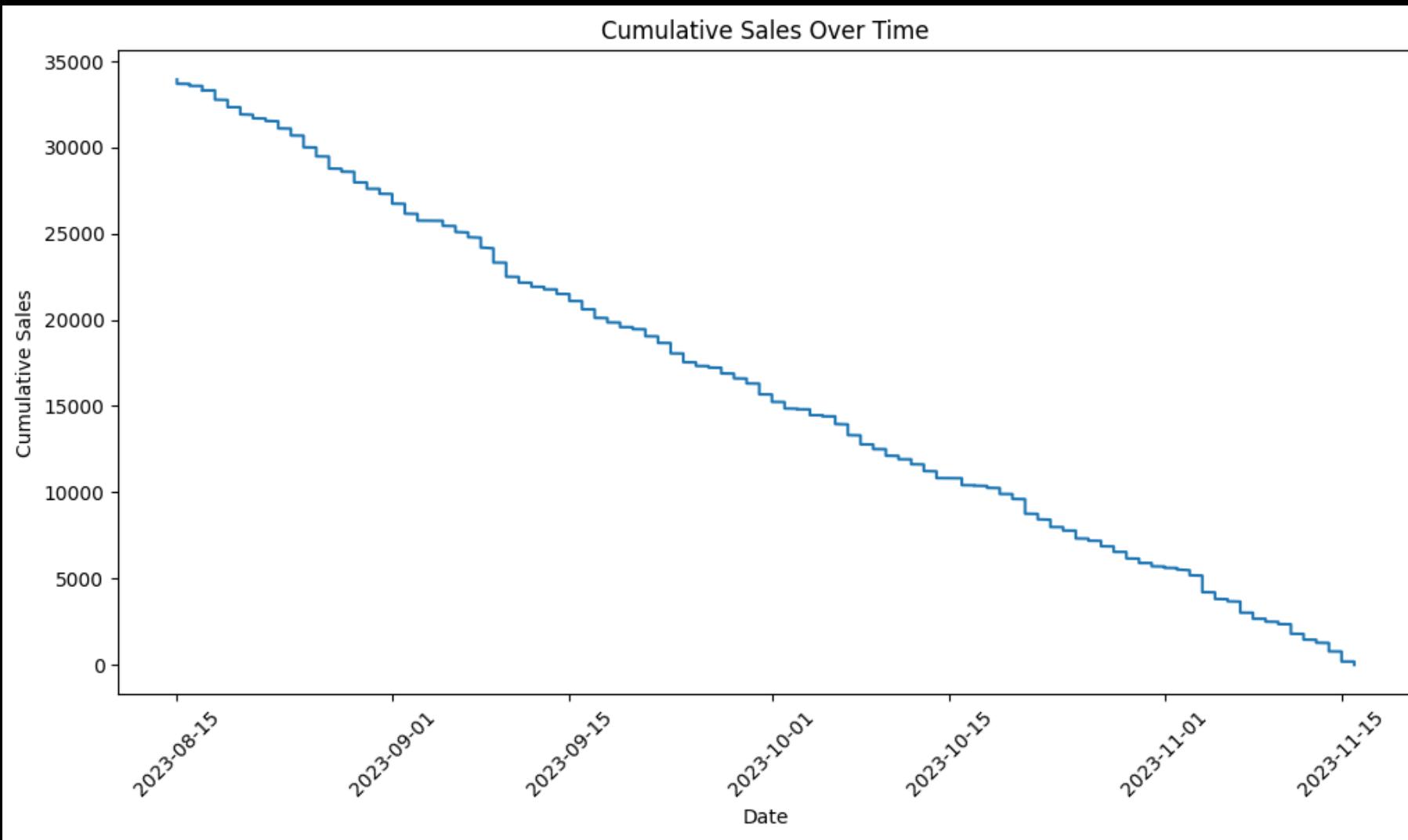
TIME SERIES ANALYSIS OF SALES

- Sales vary day-to-day and over time.
- Identified any long-term trends or patterns in sales.
- Create a time series plot of daily sales.
- The time series graph here exhibits consistent peaks and troughs. Notably, sales spike significantly around mid-week.
- This could be utilized to schedule high-impact marketing campaigns during these peak periods to maximize reach and conversion.

MONTHLY SALES DISTRIBUTION



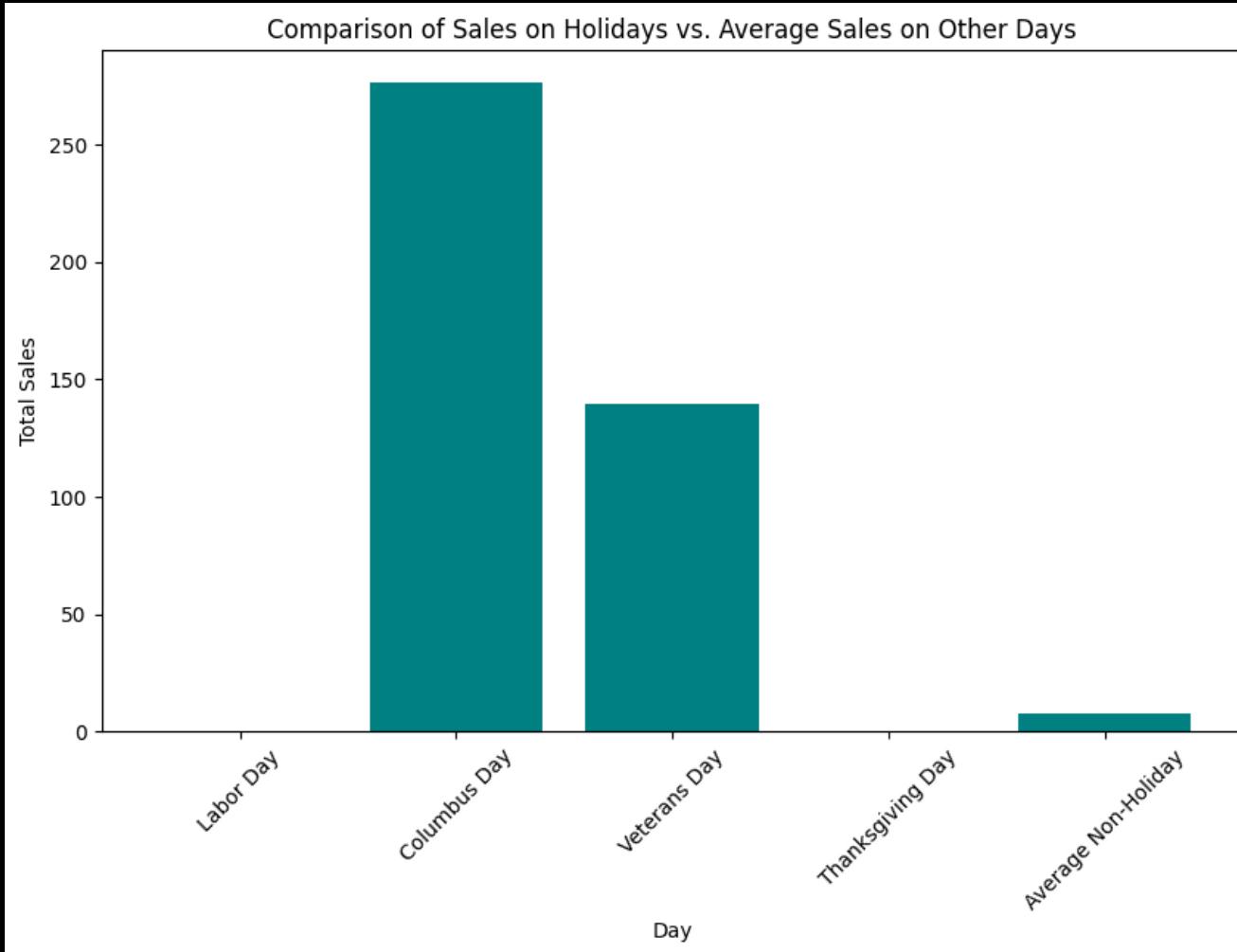
CUMMULATIVE SALES OVER TIME



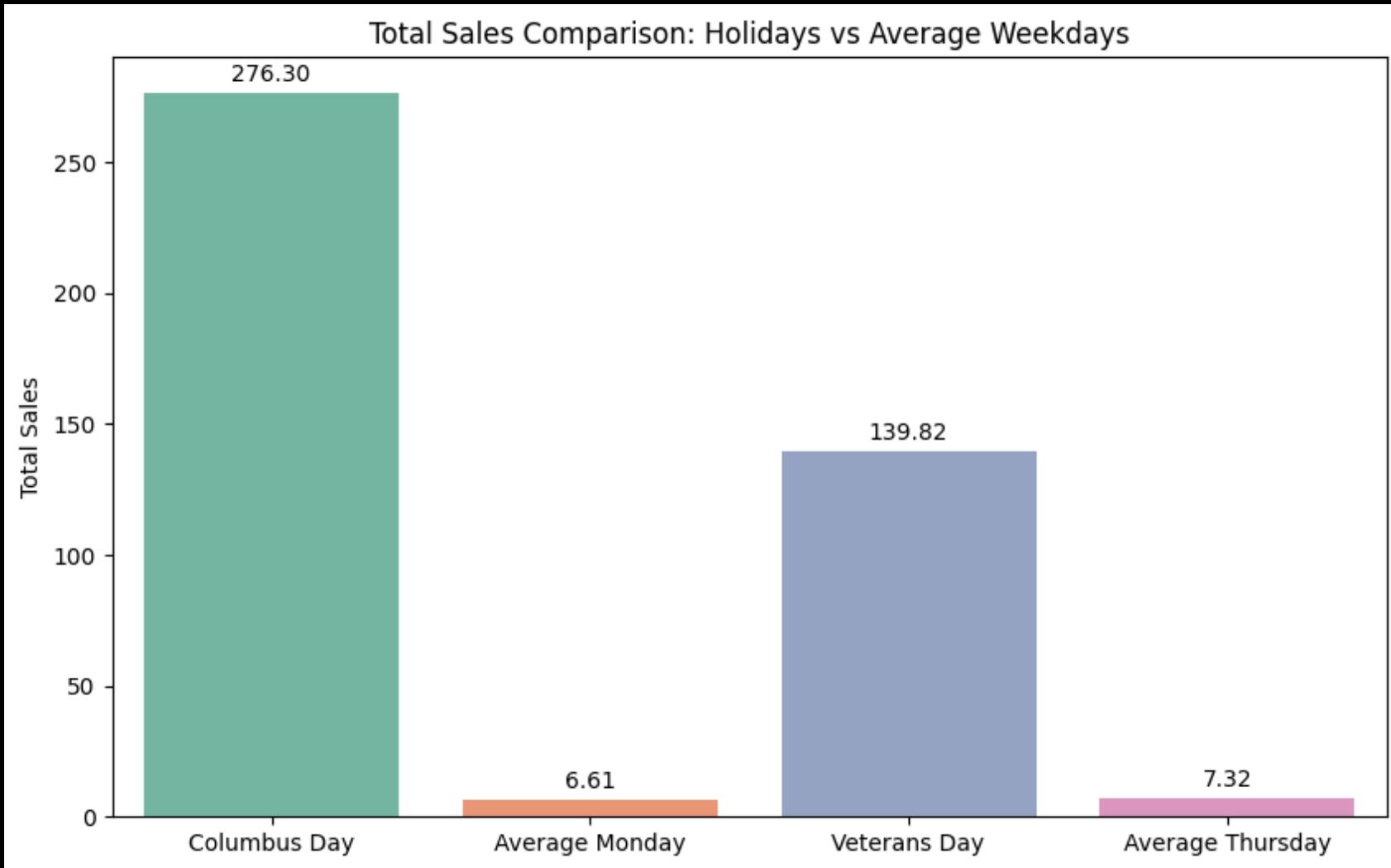
MONTHLY SALES DISTRIBUTION

- Monthly sales figures are crucial for understanding seasonal trends.
- Analyzed how sales figures change from month to month.
- Visualized monthly sales data.
- September experienced the highest sales with over 1500 transactions, while there's a downward trend into November, but we only had data for half of the November.
- This indicates a need for strategic planning to address the sales dip post-September, potentially through targeted marketing and holiday sales promotions."

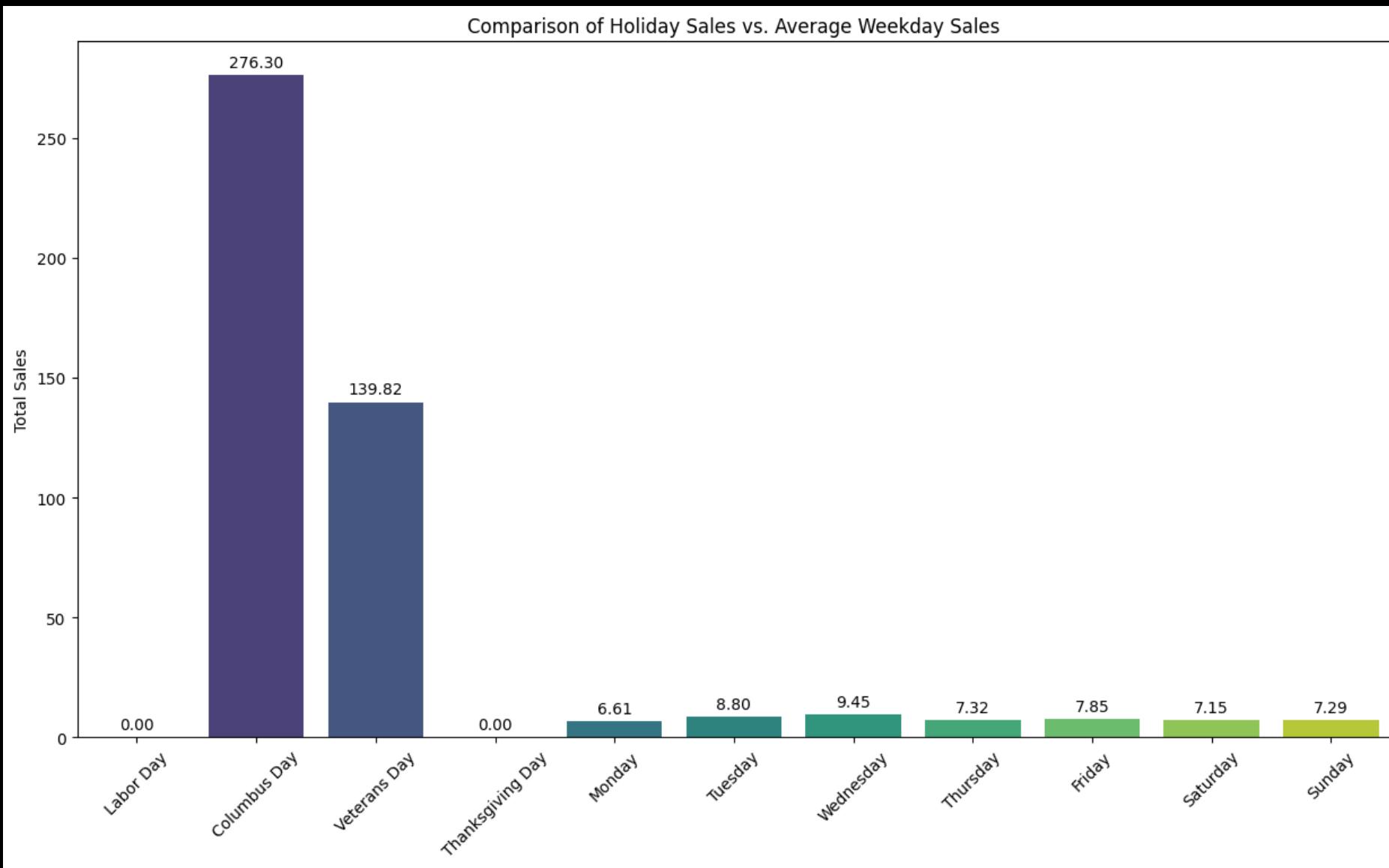
HOLIDAY SALES V/S AVG OTHER DAY SALES



HOLIDAY VS. WEEKDAY SALES PERFORMANCE



HOLIDAY VS. WEEKDAY SALES PERFORMANCE

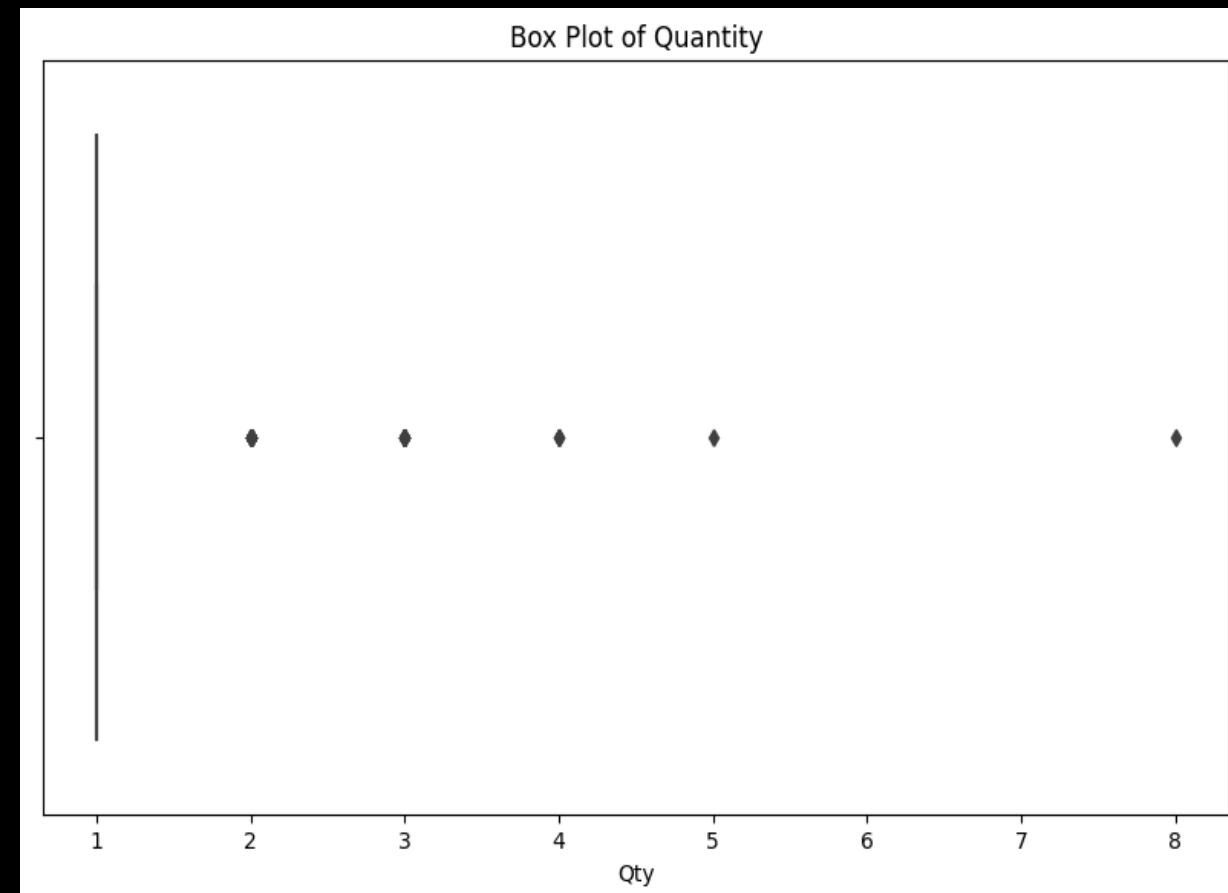


HOLIDAY VS. WEEKDAY SALES PERFORMANCE

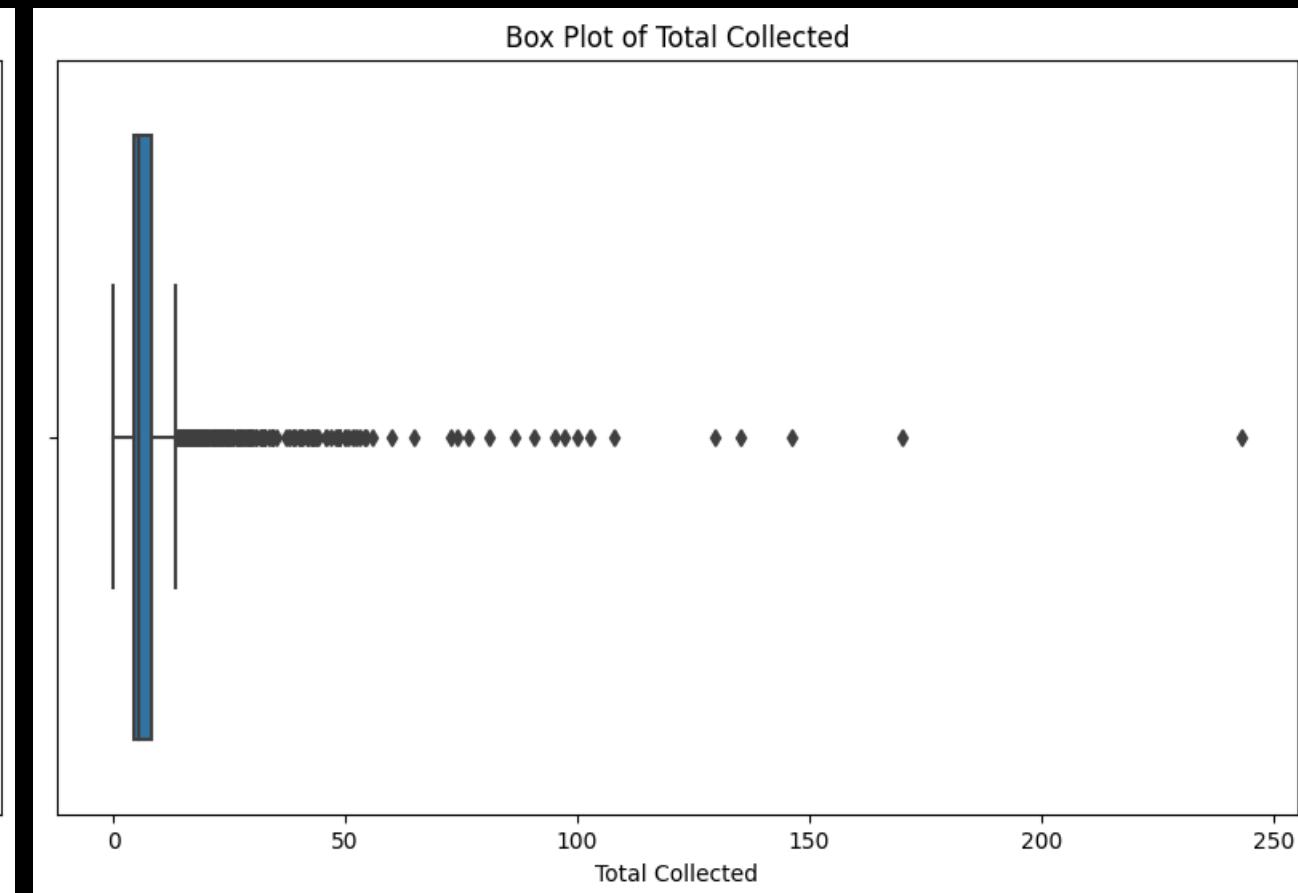
- Holidays may influence sales differently.
- Compared holiday sales to average sales on those days of the week.
- Plotted sales data for holidays and compare them to non-holiday averages.
- Columbus Day sales were notably high, almost 40x the average Monday's sales, suggesting holidays are significant drivers of traffic.
- Veterans Day outperformed the average Thursday by a wide margin.
- These insights could inform future holiday-specific marketing efforts and inventory planning."

VARIABILITY IN TRANSACTION METRICS

Box Plot of Quantity



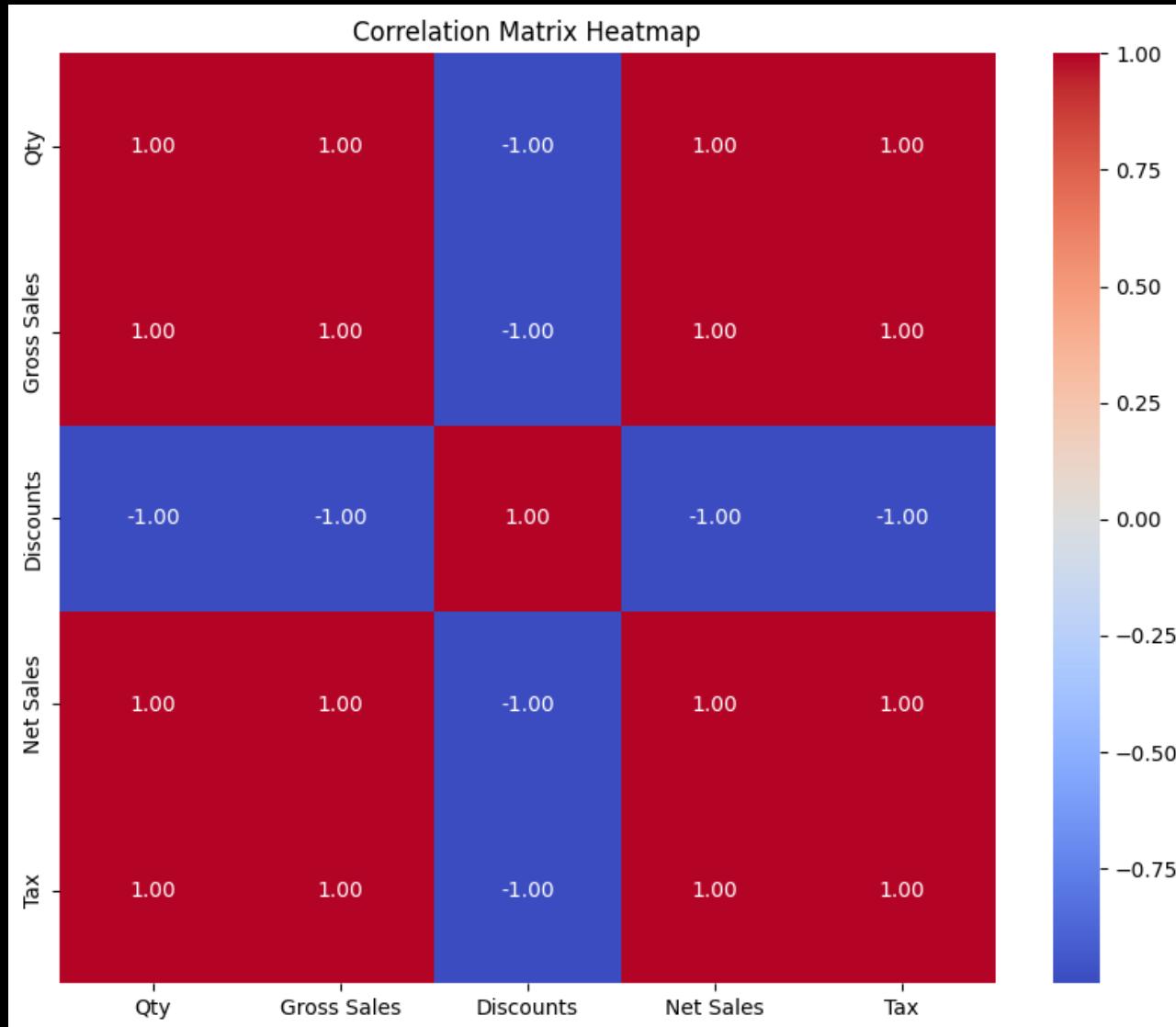
Box Plot of Total Collected



VARIABILITY IN TRANSACTION METRICS

- Transaction sizes and values vary.
- Understood the distribution and range of transaction sizes and values.
- Created box plots for 'Quantity' and 'Total Collected'.
- The box plots reveal that while most transactions involve 1-2 items, we have outliers with up to 8 items sold.
- Similarly, most sales collect under \$50, but outliers go up to \$250.
- This suggests the opportunity for bundling products and creating tiered pricing strategies to cater to diverse customer spending behaviors."

CORRELATION MATRIX HEATMAP



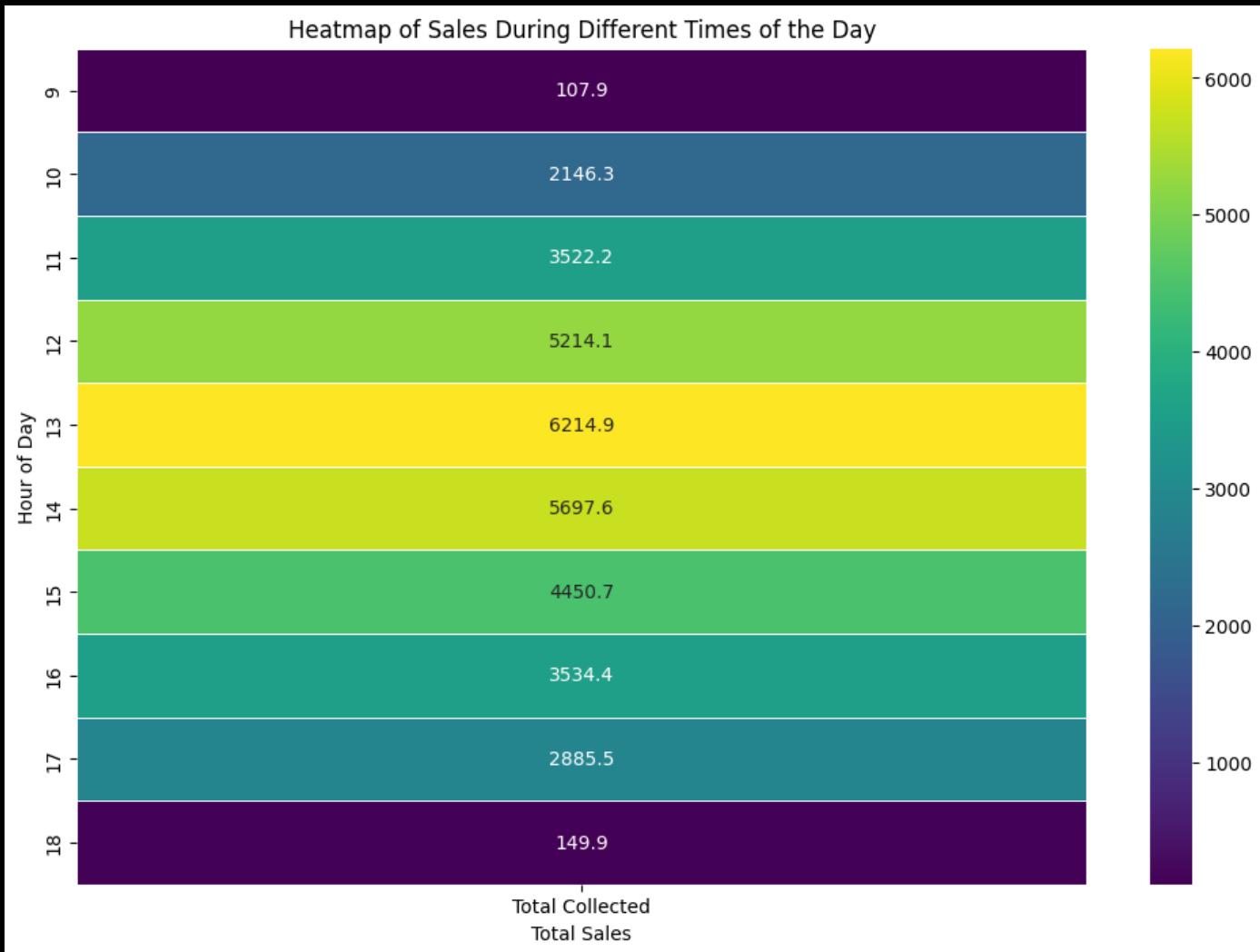
Strong Positive Relationship:

- Between Sales and Taxes: As sales figures (both gross and net) increase, the amount of tax collected also increases.
- Sales and Quantity: Higher quantities sold lead to higher sales amounts.

Negative Impact of Discounts:

- Reduction in Sales: An increase in discounts leads to a decrease in net sales, indicating that heavier discounts may reduce overall revenue.

SALES DURING DIFFERENT TIME OF THE DAY



SALES DURING DIFFERENT TIME OF THE DAY

- Identified peak sales times can optimize operations.
- Determined the busiest hours for sales.
- Generated a heatmap of sales throughout the day.
- Our heatmap shows that midday, around 12-3 PM, is our busiest time, indicating when we should have the most staff on hand.
- Conversely, we could explore promotional activities during our slowest hours, like early mornings and late evenings, to increase sales during those times.

CONCLUSION

- Analyzed transaction data from Books End to discern customer purchasing behavior and sales trends.
- Identified cash and Visa as predominant payment methods, indicating potential areas for strategic partnerships and rewards programs.
- Discovered that Saturday is the peak day for sales, suggesting opportunities for weekend-specific promotions.
- Recognized a need for intervention in October and November due to a slowing in sales momentum.
- Recommended actionable strategies such as targeted marketing during peak sales periods and special incentives on slower business days.
- Proposed continuous data monitoring to adapt strategies and maintain a competitive edge in the market.

FUTURE SCOPE

- Predictive Analytics: Utilizing machine learning to forecast future sales trends and customer behavior.
- Customer Segmentation: Targeted marketing based on genres sold
- Inventory Optimization: Applying data analytics to optimize stock levels, reduce overstock, and avoid stockouts based on genres sold
- Chat bot: Building a Chat bot using prompt engineering for specific queries of customers
- Integration with Other Data Sources: Combining sales data with external data sources like market trends, demographic information, or economic indicators to gain a more comprehensive business insight.

THANK YOU!