



# Retail Data Analytics

Sales & Store Performance Analysis

End-to-End Data Analytics Project

--- By Preeti Prajapati



# The Business Challenge

## Key Questions

- Why do sales vary across stores and departments?
- How do holidays and promotions affect revenue?
- Which stores are high-performing or underperforming?
- How do economic factors impact sales?

## The Goal

Analyze retail data to understand store performance, customer behavior, and factors influencing sales. Provide clear insights to improve business decisions using data-driven analysis.

# Dataset & Tools

## Three Data Tables

**Sales:** Store, Department, Date, Weekly\_Sales, IsHoliday

**Features:** Temperature, Fuel\_Price, CPI, Unemployment,  
Markdowns

**Stores:** Store Type and Size

## Analysis Tools

**Excel:** Data cleaning & validation

**SQL:** Data verification & complex queries

**Python:** Exploratory Data Analysis

**Power BI:** Dashboard visualization

# Data Preparation Process

## Excel Cleaning



- Removed duplicates
- Handled missing values in Markdown columns
- Cleaned date formats
- Corrected data types

## Python Processing



- Handled missing values
- Removed duplicates
- Corrected data types
- Merged all three tables

## SQL Verification



- Row count checks
- Duplicate checks
- Null value check
- Data type validation

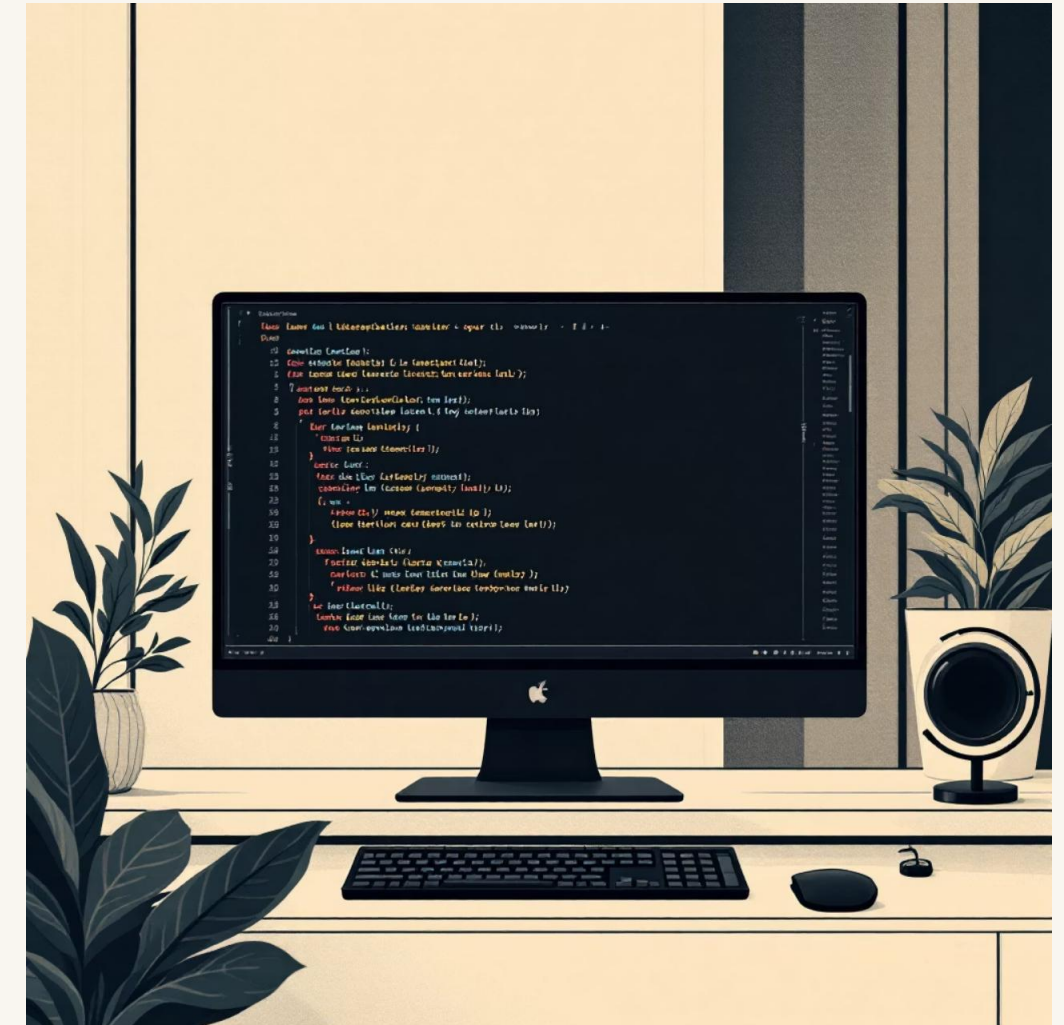


# SQL Analysis

## Key SQL Operations

- Joined Sales, Features, and Stores tables
- Calculated total revenue by store
- Compared Holiday vs Non-Holiday sales
- Ranked top 10 stores by revenue used by window functions
- Analyze impact of economic factors ( CPI & Unemployment )
- Used CTE functions to find impact of temperature on sales

📄 **Top Finding:** Store 10 on 11/26/2010 generated \$693,099 in weekly sales—the highest single-week performance



- Join all three tables

	store	Date	dept	weekly_sales	type	size	temperature	fuel_price	Unemployment
▶	1	2010-02-05	1	24924.50	A	151315	42.31	2.57	8.11
	1	2010-02-12	1	46039.49	A	151315	38.51	2.55	8.11
	1	2010-02-19	1	41595.55	A	151315	39.93	2.51	8.11
	1	2010-02-26	1	19403.54	A	151315	46.63	2.56	8.11
	1	2010-03-05	1	21827.90	A	151315	46.50	2.63	8.11
	1	2010-03-12	1	21043.39	A	151315	57.79	2.67	8.11
	1	2010-03-19	1	22136.64	A	151315	54.58	2.72	8.11
	1	2010-03-26	1	26229.21	A	151315	51.45	2.73	8.11
	1	2010-04-02	1	57258.43	A	151315	62.27	2.72	7.81
	1	2010-04-09	1	42960.91	A	151315	65.86	2.77	7.81
	1	2010-04-16	1	17596.96	A	151315	66.32	2.81	7.81
	1	2010-04-23	1	16145.35	A	151315	64.84	2.80	7.81
	1	2010-04-30	1	16555.11	A	151315	67.41	2.78	7.81
	1	2010-05-07	1	17413.94	A	151315	72.55	2.84	7.81
	1	2010-05-14	1	18926.74	A	151315	74.78	2.85	7.81
	1	2010-05-21	1	14773.04	A	151315	76.44	2.83	7.81
	1	2010-05-28	1	15580.43	A	151315	80.44	2.76	7.81
	1	2010-06-04	1	17558.09	A	151315	80.69	2.71	7.81
	1	2010-06-11	1	16637.62	A	151315	80.43	2.67	7.81
	1	2010-06-18	1	16216.27	A	151315	84.11	2.64	7.81
	1	2010-06-25	1	16328.72	A	151315	84.34	2.65	7.81
	1	2010-07-02	1	16333.14	A	151315	80.91	2.67	7.79

- Ranked top stores by sales

	store	date	dept	weekly_sales	sales_rank
▶	10	2010-11-26	72	693099.36	1
	10	2010-11-26	72	693099.36	1
	35	2011-11-25	72	649770.18	2
	35	2011-11-25	72	649770.18	2
	10	2011-11-25	72	630999.19	3
	10	2011-11-25	72	630999.19	3
	35	2010-11-26	72	627962.93	4
	35	2010-11-26	72	627962.93	4
	14	2010-11-26	72	474330.10	5
	14	2010-11-26	72	474330.10	5
	20	2010-11-26	72	422306.25	6
	20	2010-11-26	72	422306.25	6
	27	2010-11-26	72	420586.57	7
	27	2010-11-26	72	420586.57	7
	10	2010-12-24	7	406988.63	8
	10	2010-12-24	7	406988.63	8
	10	2010-12-24	72	404245.03	9
	10	2010-12-24	72	404245.03	9
	22	2010-11-26	72	393705.20	10
	22	2010-11-26	72	393705.20	10
	20	2011-11-25	72	392023.02	11
	20	2011-11-25	72	392023.02	11

- Impact of economic factors on sales

	cpi	unemployment	total_sales	avg_sales
▶	126.06	6.70	3039980.72	11343.21
	126.06	7.37	8028203.00	28268.32
	126.06	7.80	1164112.36	6192.09
	126.06	7.95	7510370.20	26077.67
	126.06	9.20	10184306.68	21576.92
	126.06	9.50	1061470.04	6031.08
	126.06	9.82	3804834.60	13397.30
	126.06	14.18	9953412.84	13235.92
	126.08	6.70	3135718.80	11700.44
	126.08	7.37	7988724.36	28129.31
	126.08	7.80	1138962.00	5811.03
	126.08	7.95	7989590.52	27361.61
	126.08	9.20	9936544.76	21231.93
	126.08	9.50	922077.96	5239.08
	126.08	9.82	3940611.76	13875.39
	126.08	14.18	9893938.68	13227.19
	126.09	6.70	6391540.24	12014.17
	126.09	7.37	15024170.00	26083.63
	126.09	7.80	2359838.28	6145.41
	126.09	7.95	15509598.88	26557.53
	126.09	9.20	18889356.92	19841.76
	126.09	9.50	1947284.56	5532.06

- Total sales by temperature

	store	Total_sales	Temprerature
▶	1	889611235.40	2795565.60
	2	1101529763.92	2789053.76
	3	230346940.28	2577017.56
	4	1198175813.52	2554723.12
	5	181902755.60	2491053.48
	6	895024522.56	2844373.52
	7	326393100.56	1546410.72
	8	519804724.52	2471669.44
	9	311156875.96	2399289.12
	10	1086470855.56	2976819.52
	11	775851147.20	2916459.32
	12	577148920.60	2722851.92
	13	1146070815.20	2245753.24
	14	1155999645.36	2320918.28
	15	356534735.68	2049527.04
	16	297009701.60	1700001.60
	17	511128555.32	1827992.28
	18	620458936.84	2101607.40
	19	826539448.40	2118737.28
	20	1205591169.84	2261441.52
	21	432471515.68	2635796.12
	22	588302594.28	2125883.04



# Python EDA Discoveries

1

## Store Types

Three store types identified: A, B, and C with varying performance levels

2

## Departments

78 unique departments analyzed across all stores

3

## Correlations

Store size shows 0.24 correlation with weekly sales—larger stores perform better

4

## Monthly Trends

December averages \$19,356 in weekly sales—highest of any month



- Find correlation

```
df= df1.select_dtypes(include=['int64','float64'])
corr = df.corr()
print(corr)
```

	Store	Dept	Weekly_Sales	IsHoliday_x	Temperature	\
Store	1.000000	0.024004	-0.085195	-0.000548	-0.050097	
Dept	0.024004	1.000000	0.148032	0.000916	0.004437	
Weekly_Sales	-0.085195	0.148032	1.000000	0.012774	-0.002312	
IsHoliday_x	-0.000548	0.000916	0.012774	1.000000	-0.155949	
Temperature	-0.050097	0.004437	-0.002312	-0.155949	1.000000	
Fuel_Price	0.065240	0.003569	-0.000123	-0.078011	0.143762	
Markdown1	-0.059844	0.001494	0.047172	-0.003521	-0.026415	
Markdown2	-0.033829	0.000587	0.020716	0.207604	-0.179672	
Markdown3	-0.020331	0.001475	0.038562	0.266471	-0.056026	
Markdown4	-0.042724	0.001937	0.037467	0.011565	-0.050281	
Markdown5	-0.012452	0.002668	0.050465	-0.015235	-0.014752	
CPI	-0.211088	-0.007477	-0.020921	-0.001943	0.182110	
Unemployment	0.208566	0.007841	-0.025871	0.010447	0.096780	
IsHoliday_y	-0.000548	0.000916	0.012774	1.000000	-0.155949	
Size	-0.182881	-0.002966	0.243828	0.000593	-0.058313	

	Fuel_Price	Markdown1	Markdown2	Markdown3	Markdown4	\
Store	0.065240	-0.059844	-0.033829	-0.020331	-0.042724	
Dept	0.003569	0.001494	0.000587	0.001475	0.001937	
Weekly_Sales	-0.000123	0.047172	0.020716	0.038562	0.037467	
IsHoliday_x	-0.078011	-0.003521	0.207604	0.266471	0.011565	
Temperature	0.143762	-0.026415	-0.179672	-0.056026	-0.050281	
Fuel_Price	1.000000	0.297224	0.029318	0.018789	0.166720	
Markdown1	0.297224	1.000000	0.174868	-0.014411	0.838904	
Markdown2	0.029318	0.174868	1.000000	-0.006080	0.113250	
Markdown3	0.018789	-0.014411	-0.006080	1.000000	-0.012020	

- Monthly sales trend

```
df1['month'] = df1['Date'].dt.month
monthly_sales = df1.groupby('month')['Weekly_Sales'].mean()
print(monthly_sales)
```

month	
1	14126.075111
2	16008.779217
3	15416.657597
4	15650.338357
5	15776.337202
6	16326.137002
7	15861.419650
8	16062.516933
9	15095.886154
10	15243.855576
11	17491.031424
12	19355.702141

Name: Weekly\_Sales, dtype: float64

- Summary overview

```
print(sales.describe())
print(features.describe())
print(features.describe())
```

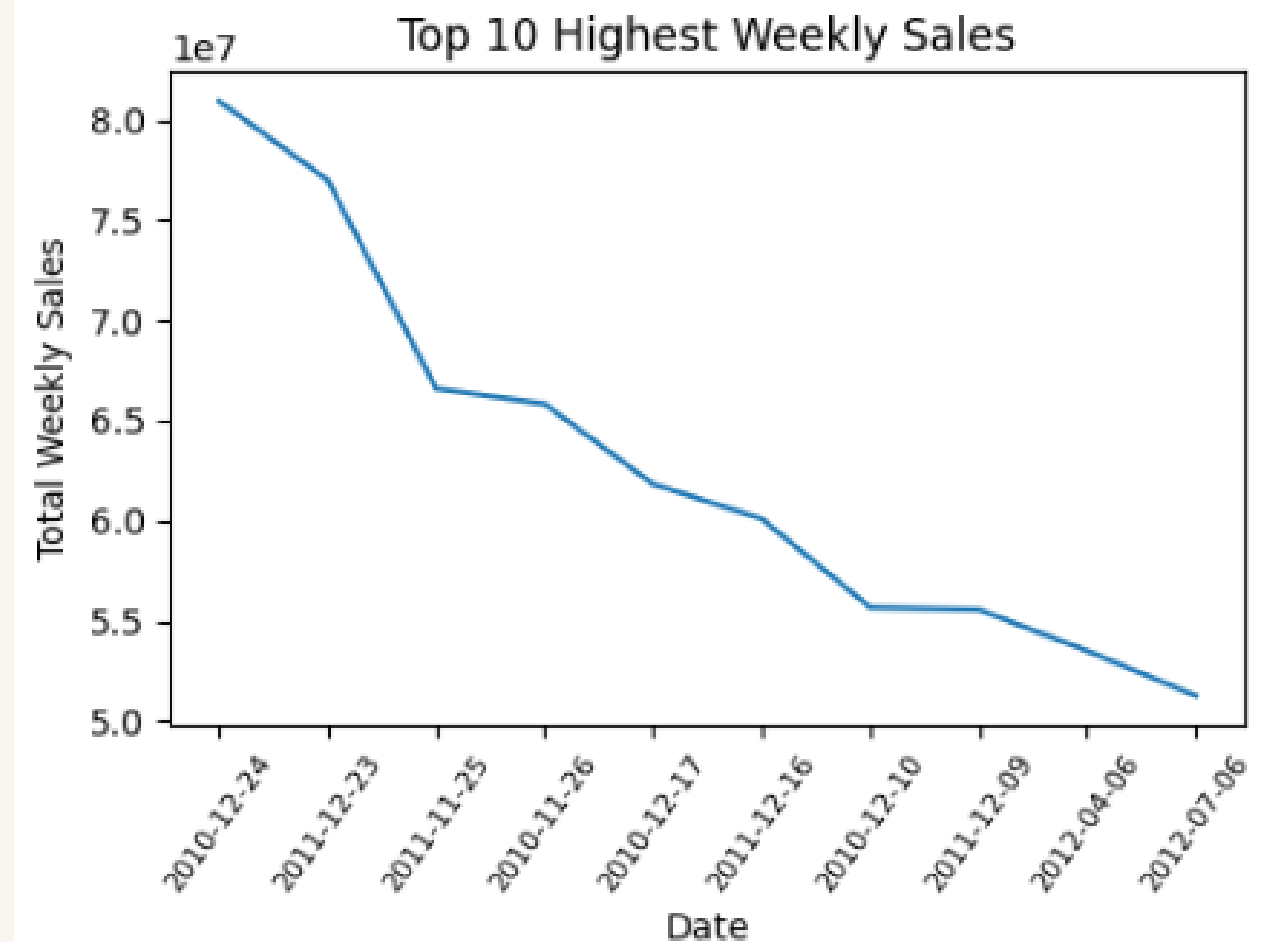
	Store	Dept	Date \
count	421570.000000	421570.000000	421570
mean	22.200546	44.260317	2011-06-18 08:30:31.963375104
min	1.000000	1.000000	2010-02-05 00:00:00
25%	11.000000	18.000000	2010-10-08 00:00:00
50%	22.000000	37.000000	2011-06-17 00:00:00
75%	33.000000	74.000000	2012-02-24 00:00:00
max	45.000000	99.000000	2012-10-26 00:00:00
std	12.785297	30.492054	NaN

	Weekly_Sales	IsHoliday
count	421570.000000	421570.000000
mean	15981.258123	0.070358
min	-4988.940000	0.000000
25%	2079.650000	0.000000
50%	7612.030000	0.000000
75%	20205.852500	0.000000
max	693099.360000	1.000000
std	22711.183519	0.255750

	Store	Date	Temperature	Fuel_Price \
count	8190.000000	8190	8190.000000	8190.000000
mean	23.000000	2011-10-31 12:00:00	59.356198	3.406497
min	1.000000	2010-02-05 00:00:00	-7.290000	2.470000
25%	12.000000	2010-12-17 00:00:00	45.902500	3.040000

- Top ten weekly sales

```
plt.figure(figsize=(5,4))
plt.plot(top10_weeks.index.astype(str), top10_weeks.values)
plt.title("Top 10 Highest Weekly Sales")
plt.xlabel("Date")
plt.ylabel("Total Weekly Sales")
plt.xticks(rotation=55,fontsize=8)
plt.tight_layout()
plt.show()
```

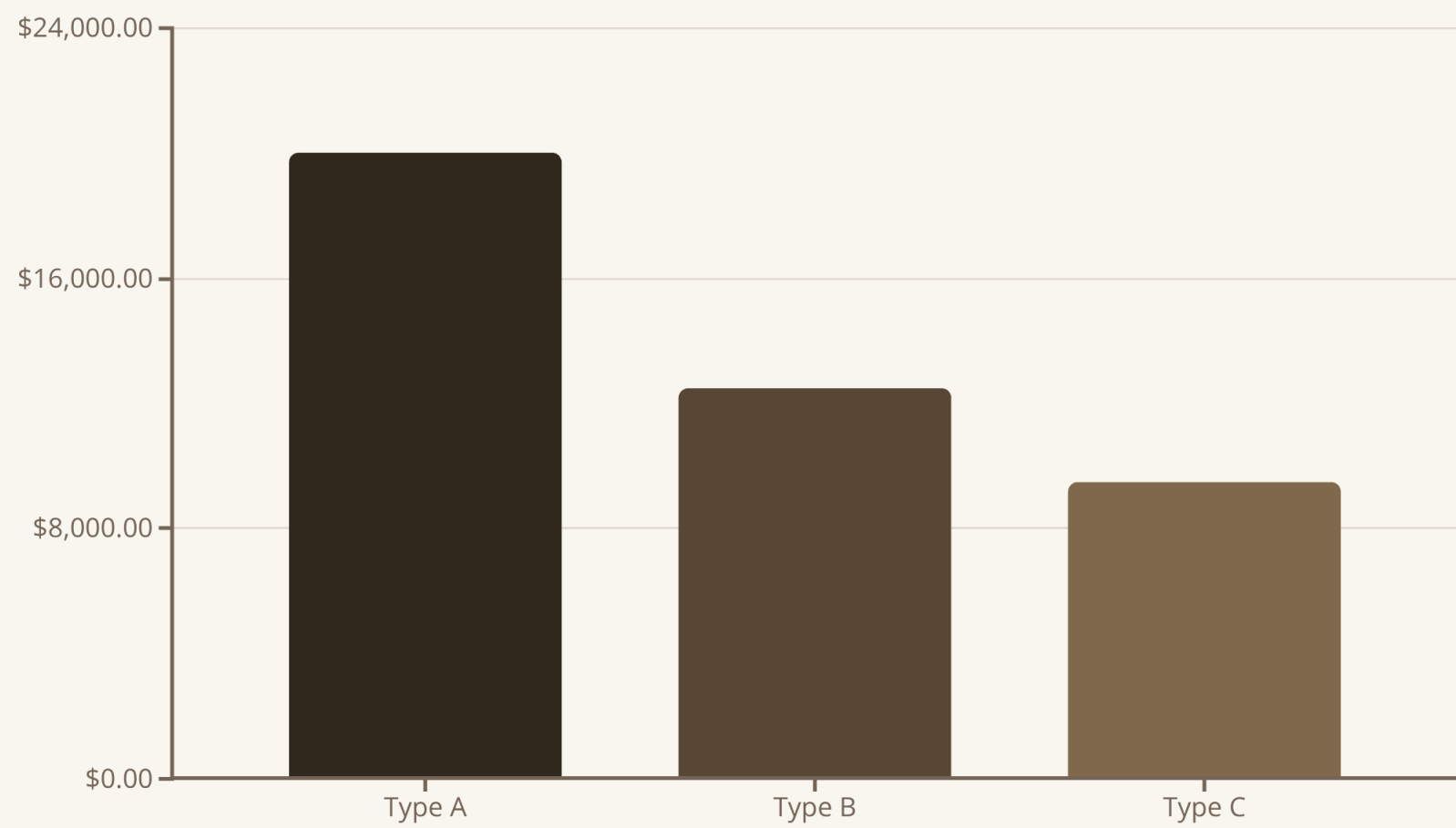


# Store Performance by Type

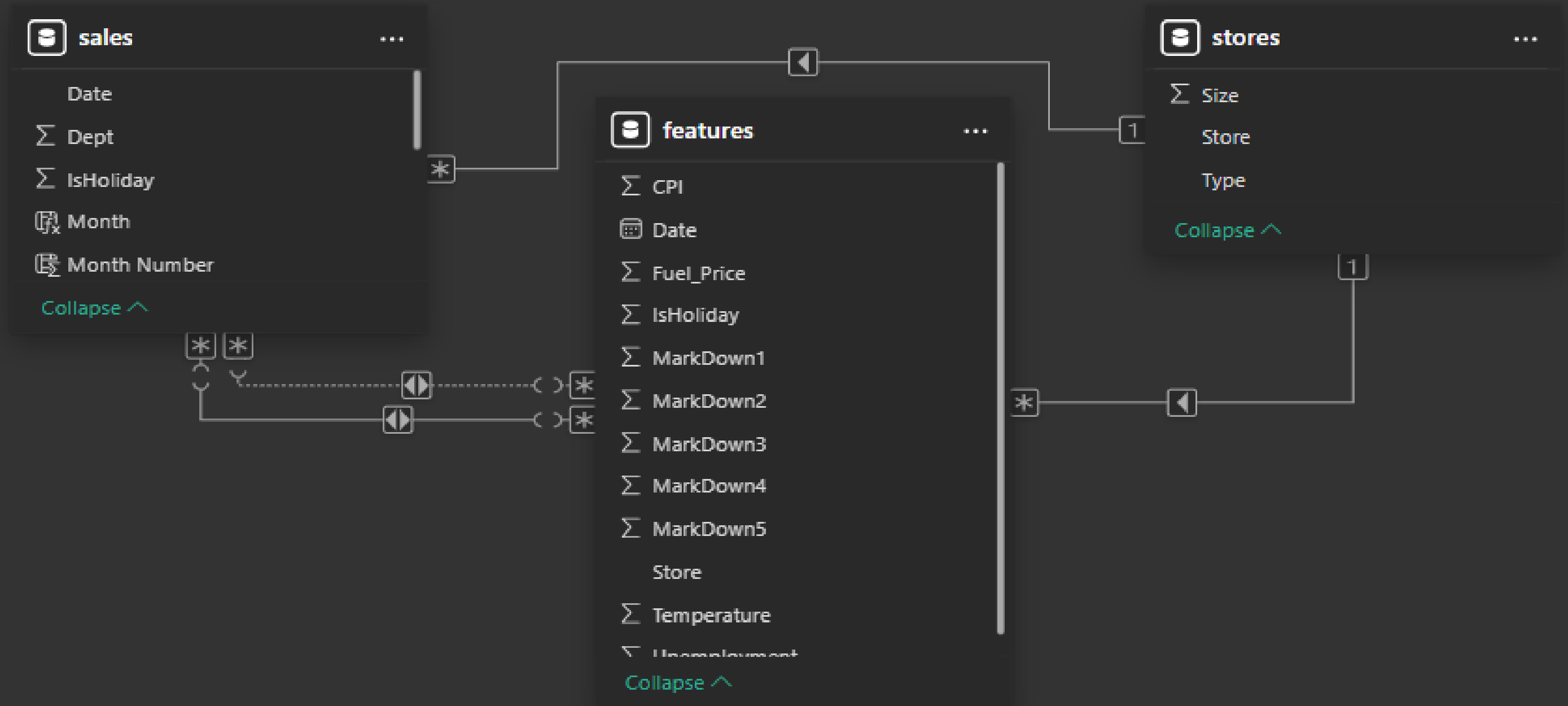
## Clear Winner

Type A stores generate the highest average weekly sales at approximately \$20,000, followed by Type B at \$12,500 and Type C at \$9,500.

Store size directly correlates with revenue generation.



# Data Modeling





# Power BI Dashboard Insights

422K    \$6.74B    \$15.98K

Total Records

Comprehensive dataset  
analyzed

Total Sales

Revenue across all  
stores

Average Sales

Per week performance

\$505M

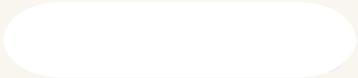
Holiday Sales

Peak period revenue

\$6.23B

Non-Holiday Sales

Strong Sales





# Power BI Dashboard

Retail  
Data  
Analytics

Year

○ 2010

○ 2011

○ 2012

422K

Total Records

6.74bn

Total Sales

15.98K

Avg Sales

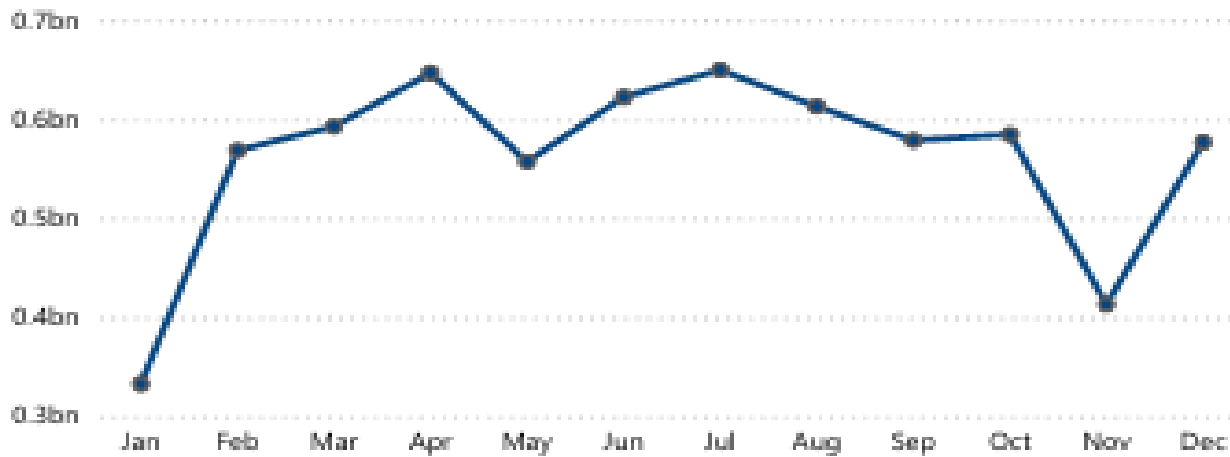
505.30M

Holiday Sales

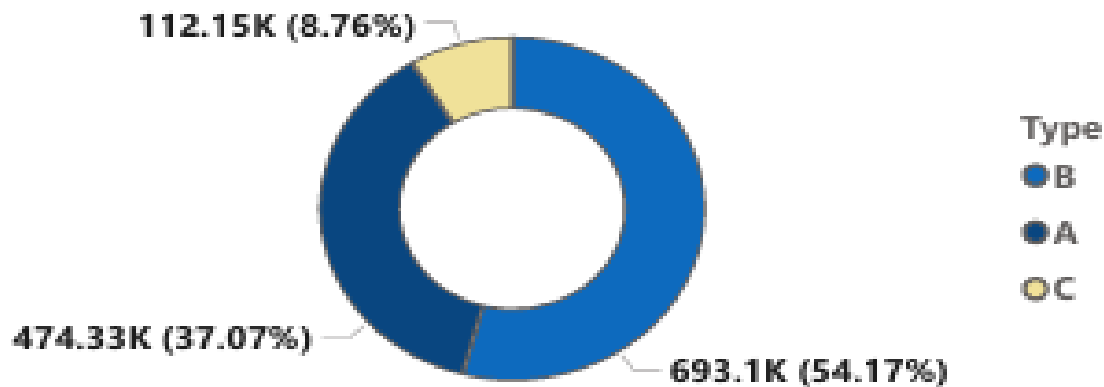
6.23bn

Non-Holiday Sales

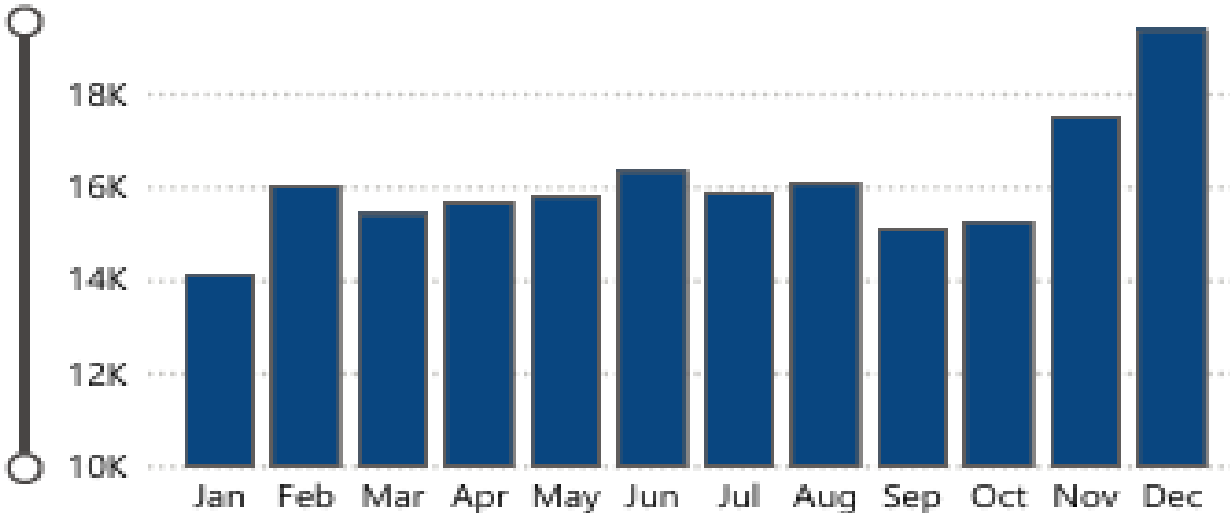
Total Sales by Month



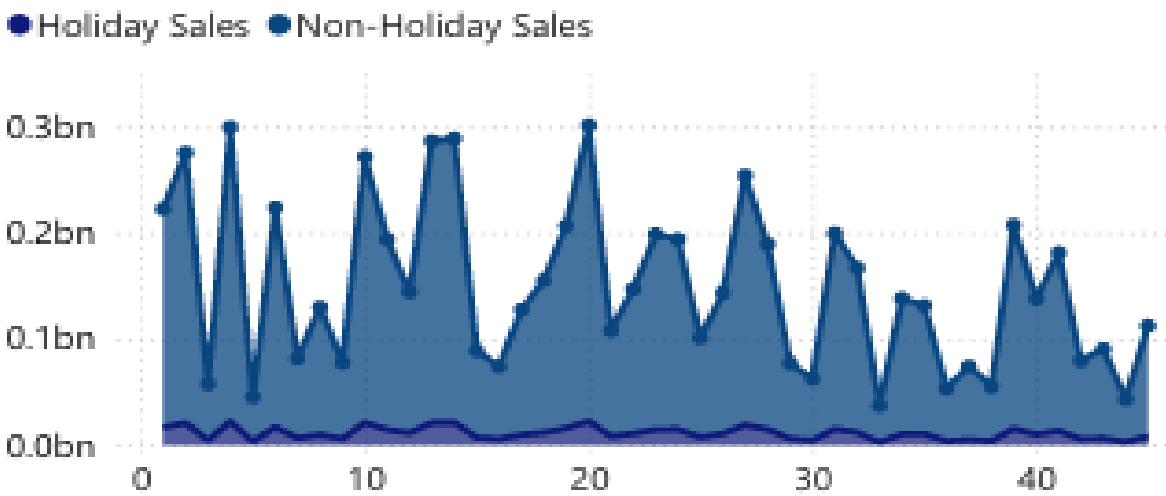
Max of Weekly\_Sales by Type



Avg Sales by Month



Holiday Sales and Non-Holiday Sales by Store



# Critical Findings



## Holiday Impact

Holiday weeks show higher per-week sales, but non-holiday periods generate more total revenue due to frequency.  
Target holidays with strong promotions.



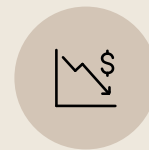
## Markdown Effect

Markdown promotions significantly boost weekly sales, especially during holidays. Stores with higher markdown usage show better uplift.



## Seasonal Patterns

November-December peak with highest sales. February-April dip presents opportunity for targeted campaigns.



## Economic Factors

High CPI and fuel prices slightly reduce spending.  
Temperature has mild influence on weekly sales patterns.



# Strategic Recommendations

01

## Promotions & Marketing

Increase markdown promotions during holiday weeks. Introduce targeted offers during low-season months (Feb-Apr). Provide store-level discount strategies for underperforming locations.

03

## Store Strategy

Expand characteristics of successful Type A stores. Consider upsizing or optimizing smaller store formats to improve performance.

02

## Inventory Planning

Increase inventory in November-December. Allocate more stock to high-performing stores. Improve staff planning and operations in low-performing stores.

04

## External Monitoring

Track fuel price and CPI for demand forecasting. Adjust pricing strategies when macro-economic conditions worsen to maintain competitiveness.

# Conclusion

This end-to-end Retail Data Analytics project provides a complete analysis of weekly sales, store performance, holiday impact, economic factors, and markdown influence.

Using Excel, SQL, Python, and Power BI, meaningful insights were generated that can support a retail business in decision-making related to planning, promotions, marketing, and store operations.

The project successfully demonstrates strong data analytics skills across data cleaning, EDA, SQL, and dashboarding—making it suitable for portfolio and interview presentations.





Thank You