Walmart Sales Forecasting

PRESENTATION

By Mahasweta Bhunia

INTRODUCTION

Hello I am Mahasweta. In this project, I analyzed Walmart's sales data from 2010 to 2012 across 45 stores to identify key sales trends and patterns. Using SQL, I explored store performance, the impact of holidays and economic factors, and the effect of discounts on sales to get the insights that can help to optimize revenue forecasting and business decisions.

ABOUT THE DATASET

- I downloaded the dataset from Kaggle.
- It has 4 tables including Features, Stores, Test and Train.

PROJECT GOALS

The Goal is to analyze historical sales data, identify key factors affecting sales, forecast future sales trends using SQL.

QUESTIONS

- Which store has highest total sales?
- Which type of stores have higher average sales ?
- Which department generate the most revenue ?
- Do sales increase during holiday weeks ?
- How do sales change each month ?
- Does the fluctuation in the fuel prices increase sales?
- Which store type is most efficient in revenue generation ?
- Do discounts increase sales?
- Which top 5 departments have the highest variation in sales?
- Which top 10 stores have sales consistently decreasing over time?

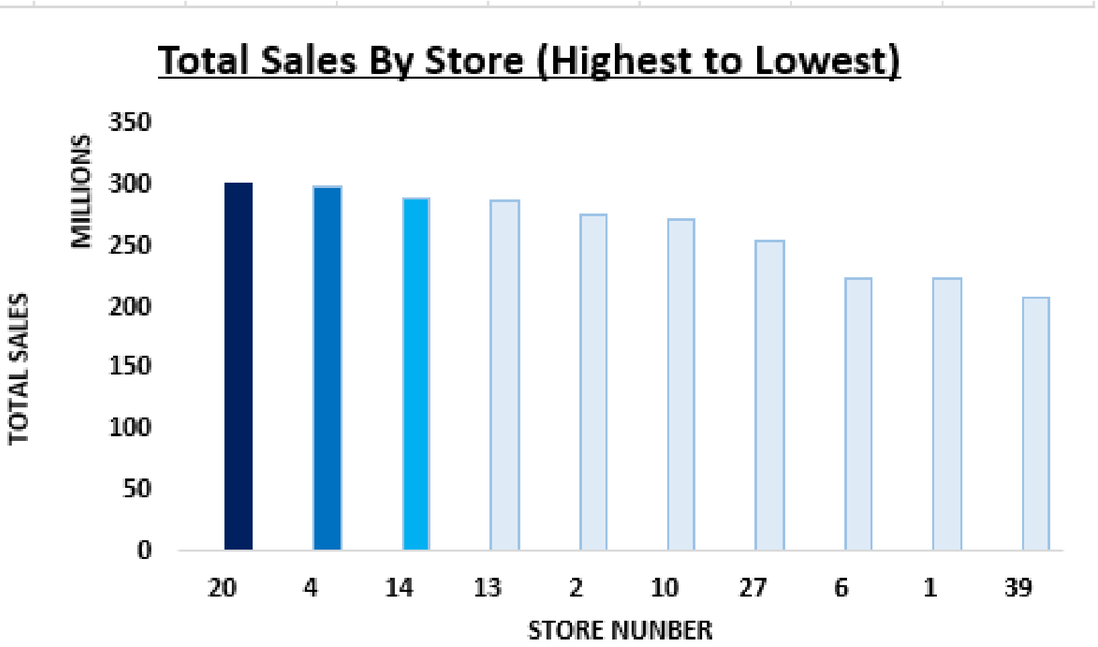
Q-1 WHICH TOP 10 STORES HAS HIGHEST TOTAL SALES?

```sql
/* Which top 10 stores has highest total sales?*/
select Store, round(sum(Weekly_Sales), 2) as Total_Sales
from train
group by Store
order by Total_Sales desc
limit 10;
```

WHICH TOP 10 STORES HAS HIGHEST TOTAL SALES?

OUTPUT

Store	Total_Sales
20	301397792.46
4	299543953.38
14	288999911.34
13	286517703.8
2	275382440.98
10	271617713.89
27	253855916.88
6	223756130.64
1	222402808.85
39	207445542.47



INSIGHTS

Store 20 leads with the highest total sales of $301.4M, followed closely by Store 4 at $299.5M. Stores 14, 13, and 2 also show strong performance, each exceeding $275M in sales. While Store 39, the lowest in the top 10, still generates $207.4M, there is a noticeable decline compared to the top performers.

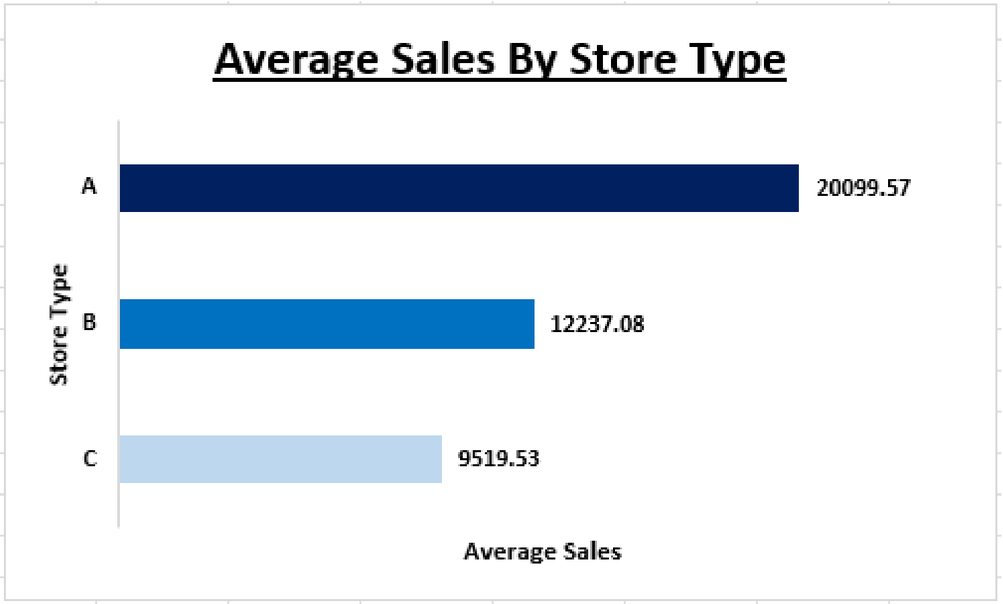
Q-2 WHICH TYPE OF STORES HAVE HIGHER AVERAGE SALES?

```sql
/*Which type of stores have higher average sales?*/
select s.Type, round(avg(t.Weekly_Sales), 2) as Avg_Sales
from train as t
join stores as s
on s.Store = t.Store
group by s.Type;
```

WHICH TYPE OF STORES HAVE HIGHER AVERAGE SALES?

OUTPUT

	Type	Avg_Sales
▶	A	20099.57
	B	12237.08
	C	9519.53



INSIGHTS

Store Type A has the highest average sales at 20,099.57, significantly outperforming Store Types B and C, which have average sales of 12,237.08 and 9,519.53, respectively. This suggests that Store Type A is the most efficient in driving revenue per transaction, potentially due to factors like larger store size, better product variety, or higher customer footfall. The noticeable gap between store types highlights opportunities to analyze operational strategies and identify what makes Type A stores more successful.

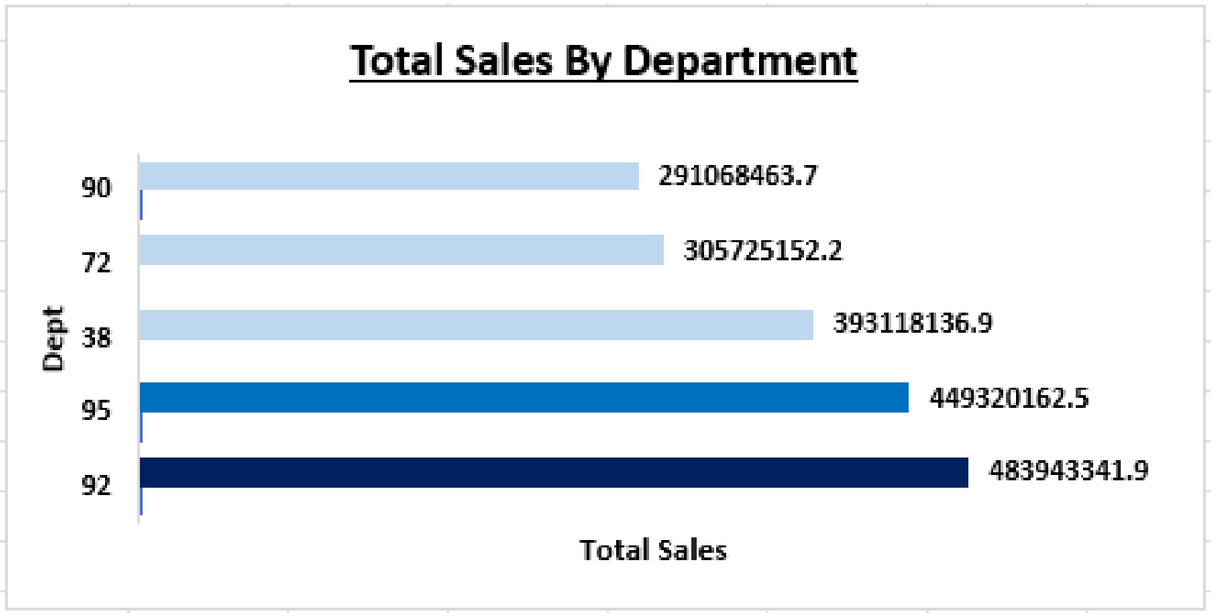
Q-3 WHICH DEPARTMENT GENERATE THE MOST REVENUE?

```sql
/* Which department generate the most revenue?*/
select Dept, sum(Weekly_Sales) as Total_Sales
from train
group by Dept
order by Total_Sales desc
limit 5;
```

WHICH DEPARTMENT GENERATE THE MOST REVENUE?

OUTPUT

	Dept	Total_Sales
▶	92	483943341.8699997
	95	449320162.5200014
	38	393118136.9200009
	72	305725152.21000063
	90	291068463.68000007



INSIGHTS

Department 92 generates the highest revenue at approximately $483.9M, followed closely by Department 95 with $449.3M. Department 38 also performs well with $393.1M, while Departments 72 and 90 contribute $305.7M and $291.1M, respectively. The significant revenue differences between departments suggest that certain product categories drive higher sales, possibly due to customer demand, pricing strategies, or seasonal trends..

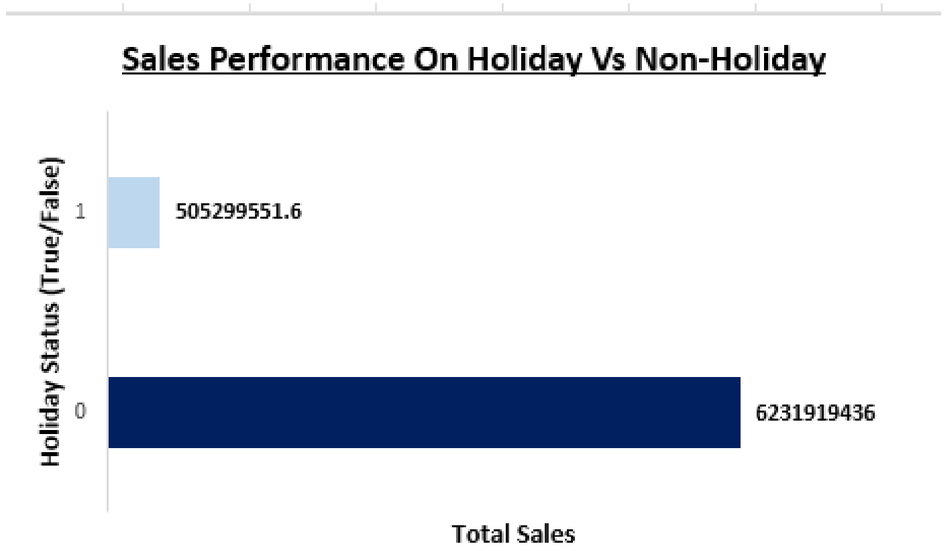
Q-4 DO SALES INCREASE DURING HOLIDAY WEEKS?

```sql
/*Do sales increase during holiday weeks?*/
select IsHoliday, sum(Weekly_Sales) as Total_Sales
from train
group by IsHoliday;
```

DO SALES INCREASE DURING HOLIDAY WEEKS?

OUTPUT

IsHoliday	Total_Sales
FALSE	6231919435.550058
TRUE	505299551.56000257



INSIGHTS

Sales during non-holiday weeks are higher than holiday weeks, with total sales reaching approximately $623.2M compared to $505.3M during holiday periods. This suggests that while holidays may drive increased consumer spending in specific categories, overall sales performance remains stronger on regular weeks. Factors such as promotional strategies, seasonal demand variations, and shopping behaviors could influence these trends, indicating the need for targeted marketing efforts during holiday seasons to maximize sales.

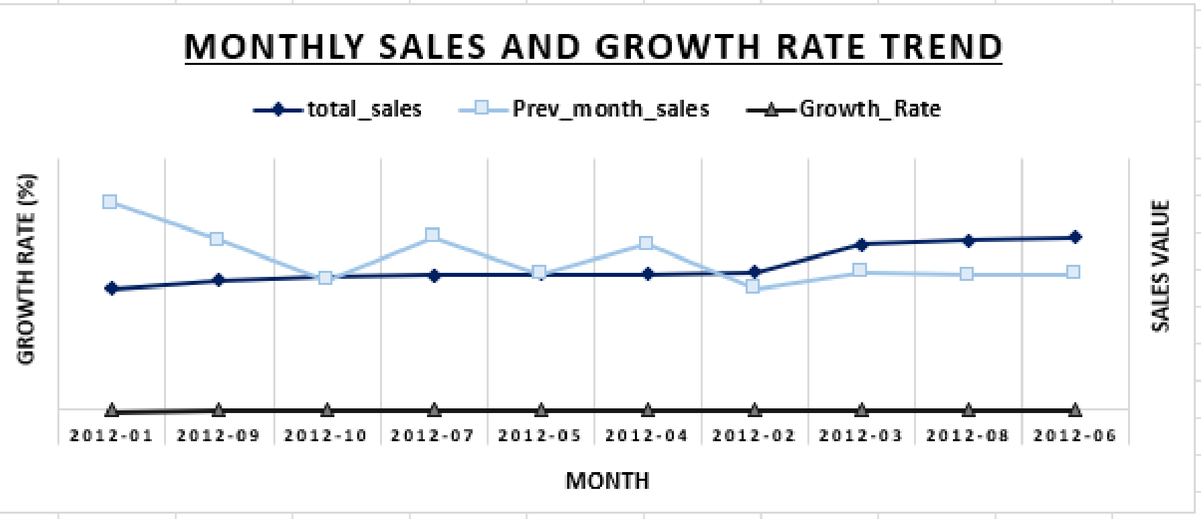
Q-5 HOW DO SALES CHANGE EACH MONTH?

```sql
select date_format(date, '%Y-%m') as month,
       sum(Weekly_Sales) as total_sales,
       lag(sum(Weekly_Sales)) over (order by date_format(date, '%Y-%m')) as Prev_month_sales,
       round((sum(Weekly_Sales) - lag(sum(Weekly_Sales)) over (order by date_format(date, '%Y-%m')))/
       lag(sum(Weekly_Sales)) over (order by date_format(date, '%Y-%m')) * 100, 2) as Growth_Rate
       from train
       group by month
       order by month desc
       limit 10;
```

HOW DO SALES CHANGE EACH MONTH?

OUTPUT

month	total_sales	Prev_month_sales	Growth_Rate
2012-10	184361680.42000026	180645544.46999907	2.06
2012-09	180645544.46999907	236850765.67999858	-23.73
2012-08	236850765.67999858	187509452.40000063	26.31
2012-07	187509452.40000063	240610329.29000032	-22.07
2012-06	240610329.29000032	188766479.4500001	27.46
2012-05	188766479.4500001	188920905.94999975	-0.08
2012-04	188920905.94999975	231509650.49000058	-18.4
2012-03	231509650.49000058	192063579.5399994	20.54
2012-02	192063579.5399994	168894471.65999863	13.72
2012-01	168894471.65999863	288078102.4799982	-41.37



INSIGHTS

Sales growth demonstrates significant monthly volatility, ranging from a peak of +27.46% to a sharp decline of -41.37%. The highest growth rates are observed in June at 27.46% and August at 26.31%, suggesting potential seasonal peaks during these months. Conversely, January experiences a substantial growth drop of -41.37%, indicating a period requiring strategic focus. These fluctuations underscore the need for businesses to adopt adaptable strategies to navigate seasonal variations and optimize growth.

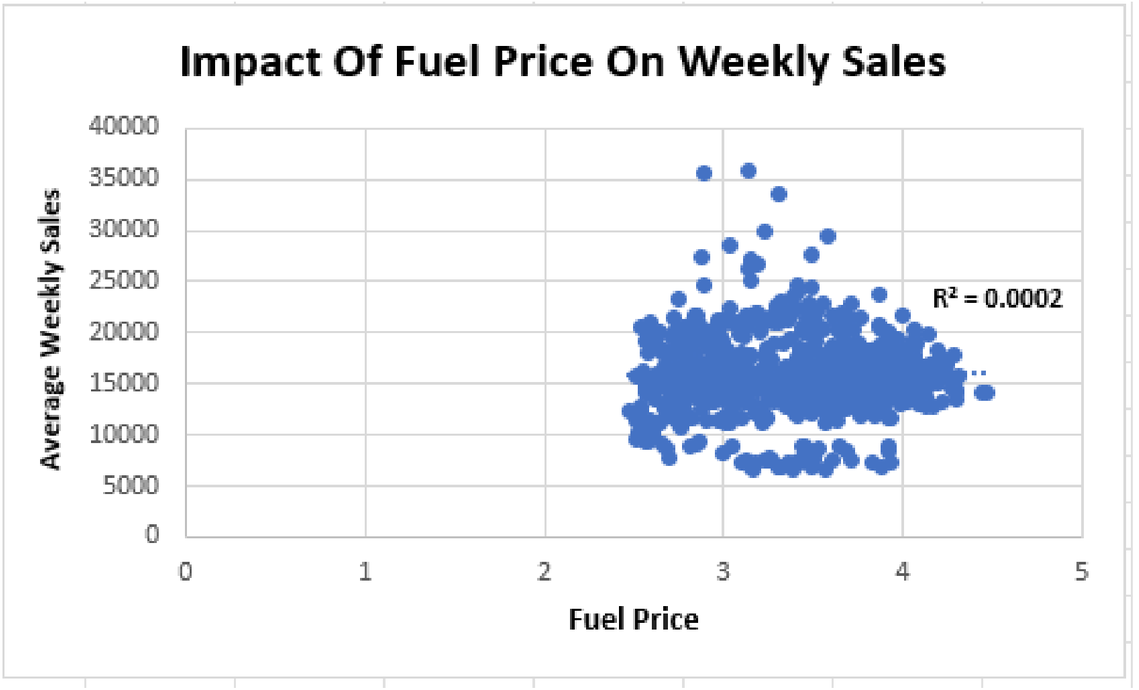
Q-6 DOES THE FLUCTUATION IN FUEL PRICES IMPACT WEEKLY SALES?

```sql
/*Does the fluctuation in fuel prices impact weekly sales?*/
select f.Fuel_Price, round(avg(t.Weekly_Sales), 2) as Weekly_Sales
 from train as t
 join features as f
 on f.Store = t.Store and t.Date = f.Date
 group by f.Fuel_Price
 order by f.Fuel_Price
 limit 10;
```

DOES THE FLUCTUATION IN FUEL PRICES IMPACT WEEKLY SALES?

OUTPUT

	Fuel_Price	Weekly_Sales
▶	2.472	12375.82
	2.513	9654.92
	2.514	15685.86
	2.52	11474.86
	2.533	10268.92
	2.539	12690.91
	2.54	20638.25
	2.542	10044.92
	2.545	12303.86
	2.548	16153.45



INSIGHTS

This dataset contains 893 rows, and a sample is displayed. The R^2 value is 0.0002, which is extremely low, indicating that fuel price fluctuations have almost no impact on weekly sales. The scatter plot further supports this, showing no clear trend between fuel price and sales. This suggests that other factors, such as promotions, seasonality, or economic conditions, might have a greater influence on weekly sales

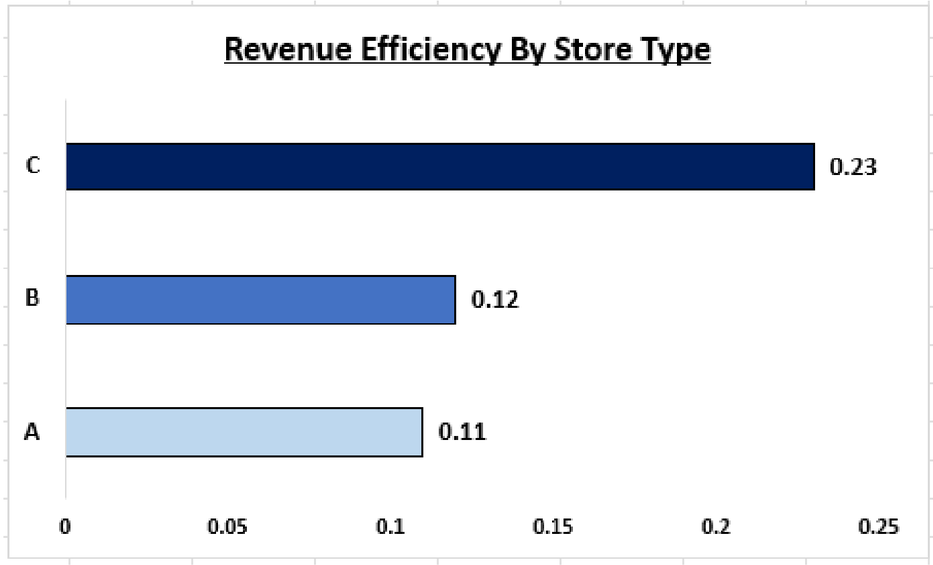
Q-7 WHICH STORE TYPE IS MOST EFFICIENT IN REVENUE GENERATION?

```sql
select s.Type, round(sum(t.Weekly_Sales)/sum(s.Size), 2) as revenue_per_sqr
from train as t
join stores as s
on s.Store = t.Store
group by s.Type
order by revenue_per_sqr desc;
```

WHICH STORE TYPE IS MOST EFFICIENT IN REVENUE GENERATION?

OUTPUT

	Store_Type	Revenue_per_Sqr
▶	C	0.23
	B	0.12
	A	0.11



INSIGHTS

Store Type C is the most efficient in revenue generation, with the highest revenue per square foot (0.23). Store Type B follows at 0.12, while Store Type A is the least efficient at 0.11. This suggests that Store Type C utilizes its space best, and Walmart should consider expanding this format. Store Type A's lower efficiency may indicate inefficiencies in space utilization or lower sales density, requiring further investigation.

Q- 8 DO DISCOUNTS INCREASE SALES?

```sql
select f.MarkDown1, round(avg(t.Weekly_Sales),2) as Avg_Sales
from train as t
join features as f
on t.Store = f.Store and f.Date = t.Date
group by f.MarkDown1
order by Avg_Sales desc;
```

DO DISCOUNTS INCREASE SALES?

INSIGHTS

Based on the analysis, the majority of sales occur at a markdown value of 0, indicating that discounts are not always the primary driver of sales. There is no strong correlation observed between increasing markdown values and higher sales, suggesting that other factors such as demand patterns, store performance, or seasonal trends may play a more significant role in influencing sales.

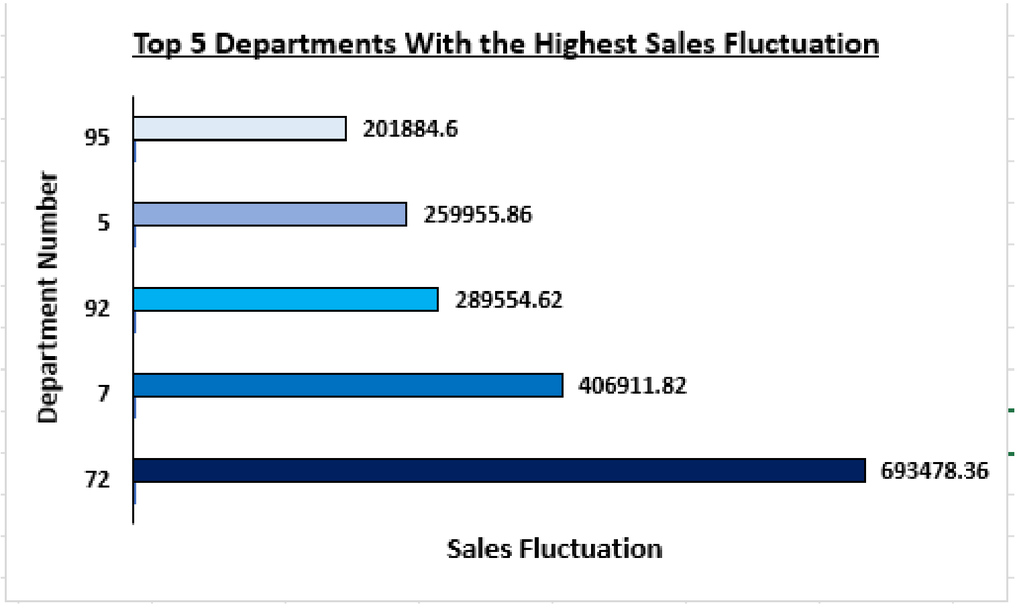
Q-9 WHICH TOP 5 DEPARTMENTS HAVE THE HIGHEST VARIATION IN SALES?

```sql
select Dept, max(Weekly_Sales) - min(Weekly_Sales) as Sales_Fluctuation
from train
group by Dept
order by Sales_Fluctuation desc
limit 5;
```

Q-9 WHICH TOP 5 DEPARTMENTS HAVE THE HIGHEST VARIATION IN SALES?

OUTPUT

	Dept	Sales_Fluctuation
▶	72	693478.36
	7	406911.82
	92	289554.62
	5	259955.86000000002
	95	201884.6



INSIGHTS

The analysis reveals that Department 72 experiences the highest sales fluctuation (693,478.36), significantly surpassing other departments. Departments 7, 92, 5, and 95 also show notable variations, indicating potential factors like seasonal demand shifts, promotional impacts, or inconsistent customer preferences. Such high fluctuations can lead to inventory mismanagement and revenue instability. To address this, Walmart should closely monitor these departments, optimize stock levels, and refine pricing strategies to minimize unpredictability and enhance overall sales stability..

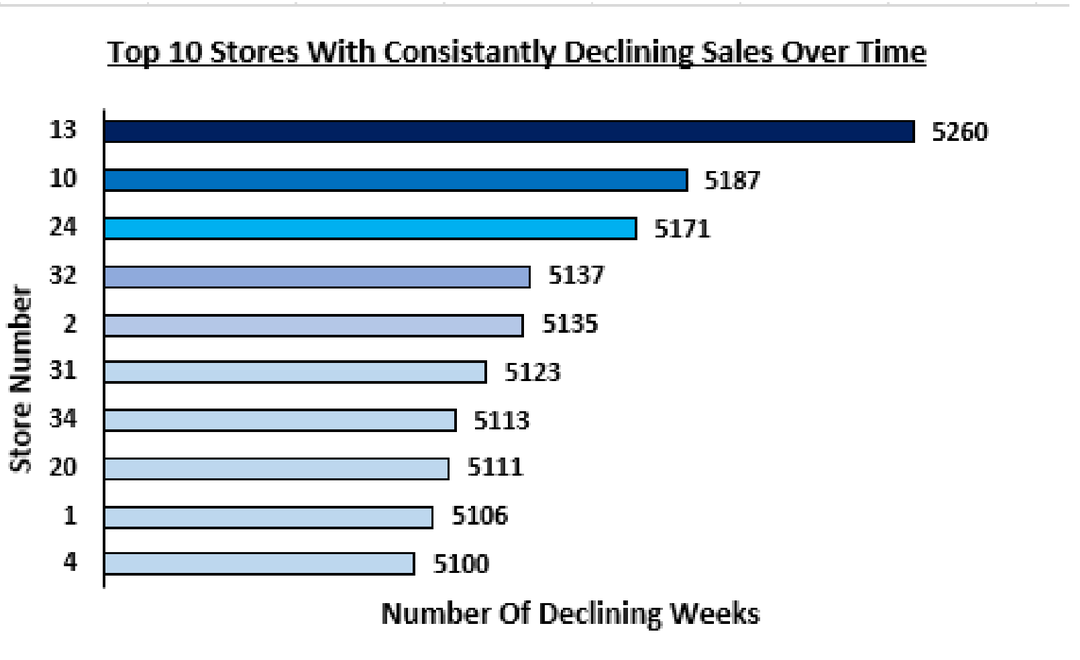
WHICH STORES HAVE SALES CONSISTANTLY DECREASING OVER TIME?

```sql
with Sales_Data as
(select Store, Date, Weekly_Sales,
    lag(Weekly_Sales,1) over (partition by Store order by Date) as Prev_Week_Sales
    from train)
select Store, count(*) as Declining_Weeks
from Sales_Data
where Weekly_Sales < Prev_Week_Sales
group by Store
order by Declining_Weeks desc
limit 10;
```

Q-10 WHICH STORES HAVE SALES CONSISTANTLY DECREASING OVER TIME?

OUTPUT

	Store	Declining_Weeks
▶	13	5260
	10	5187
	24	5171
	32	5137
	2	5135
	31	5123
	34	5113
	20	5111
	1	5106
	4	5100



INSIGHTS

Store 13 leads with the most weeks of consistently declining sales at 5260. The top ten stores experiencing this trend, including 10, 24, 32, 2, 31, 34, 20, 1, and 4, all show a significant number of declining weeks, each exceeding 5100 and ranging downwards to 5100 weeks of consistent sales decrease.

CONCLUSION

- The analysis provided valuable insights into store performance, department-wise revenue, and seasonal trends.
- Holiday weeks and markdowns positively impact sales, confirming the importance of strategic discounting.
- Certain stores are underperforming, indicating a need for operational improvements.
- The findings help Walmart optimize inventory, pricing, and promotions to drive better sales performance.

THANK YOU