DATA DIVERS
WALMART ANALYSIS

SUBTITLE GOES HERE

DATE 2023

INTRODUCTION

Data Divers is a specialist data analytics agency hired by the US grocery conglomerate Walmart to provide some insights based on historical data to help their sales and marketing team.

We were given the a dataset that covers the period 4 February 2010 to 1 November 2012, including monthly sales figures, national holidays, and inflation rate.

TEAM

Data Divers is a new data analytics agency that is providing bespoke solutions for a range of clients.

Adnan Hussain

Data Analysis & Exploration

Sanduni Kaushalya

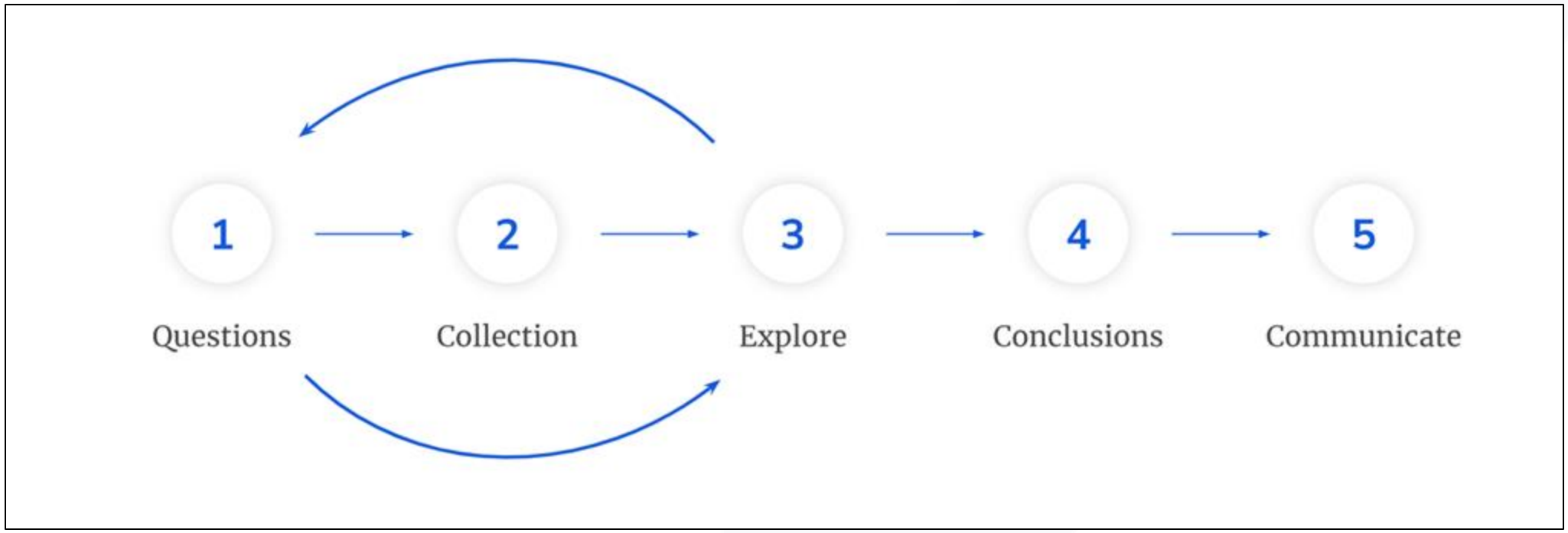
Data Cleaning & Analysis

Charlie Macnaughton

Visualisation & Communication

METHODOLOGY

Our team followed a tried and tested methodology based on their experience at the prestigious BrainStation Academy in London under the instruction of Amazon's Rianna Beaton and Iwoca's Angela Huang.

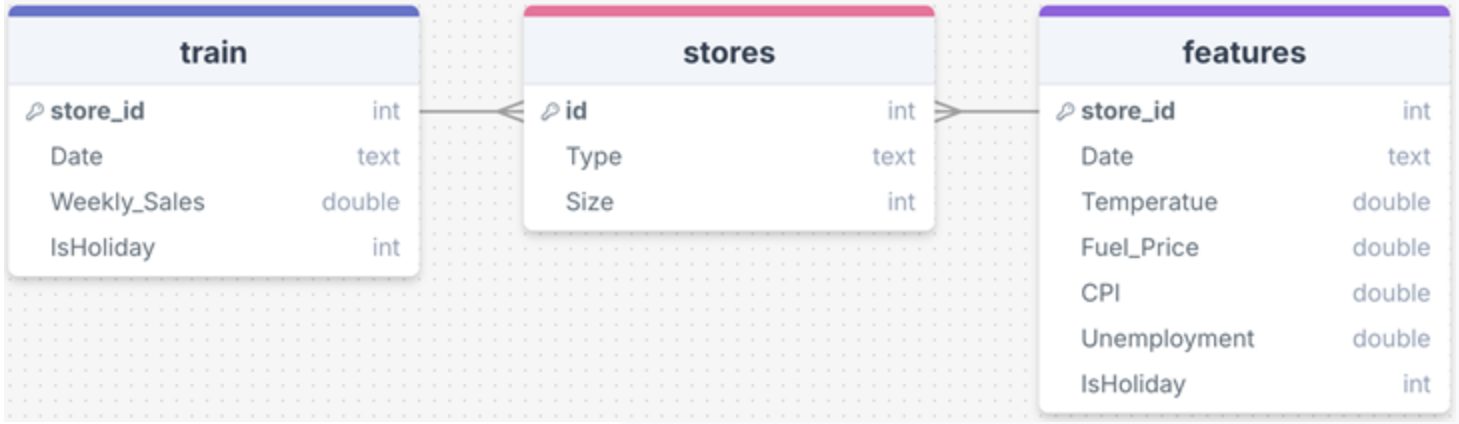


DATA
SELECTION & CLEANING

SUBTITLE GOES HERE

DATE 2023

DATA SELECTION



- This dataset contains valuable information regarding weekly sales of different Walmart stores in multiple areas across the USA
- Contains aspect of human behaviour and potential marketing ideas.
- Notable columns include Weekly_Sales, Date, Type and Temperature.

DATA CLEANING

01 REMOVAL OF UNNECESSARY COLUMNS

02 HANDLING 'N/A' VALUES

03 STANDARDISING HOLIDAY VALUES

04 STANDARDISING NUMERIC DATA

05 CONSISTENT DATA TYPES ACROSS COLUMNS

DATA ANALYSIS

- A combination of aggregate and window functions were used to calculate averages, quartiles, assemble foreign keys etc.
- CTEs were used where appropriate.
- STR_TO_DATE was required to change `Date` from type text to datetime.
- Quartile analysis was used instead of averages to more accurately highlight extremities.

```sql
-- UNDERPERFORMING MONTHS (BELOW 25TH PERCENTILE)--
WITH monthly_sales AS (
    SELECT
        -- Date is of type text which needs to be converted to type datetime to extract data from Date.
        MONTH(STR_TO_DATE(Date, "%d/%m/%Y")) AS month_num,
        MONTHNAME(STR_TO_DATE(Date, "%d/%m/%Y")) AS month,
        CASE
            WHEN MONTH(STR_TO_DATE(Date, "%d/%m/%Y")) IN (12,  1,  2) THEN 1
            WHEN MONTH(STR_TO_DATE(Date, "%d/%m/%Y")) IN ( 3,  4,  5) THEN 2
            WHEN MONTH(STR_TO_DATE(Date, "%d/%m/%Y")) IN ( 6,  7,  8) THEN 3
            WHEN MONTH(STR_TO_DATE(Date, "%d/%m/%Y")) IN ( 9, 10, 11) THEN 4
        END AS season_num,
        ROUND(SUM(Weekly_Sales), 2) AS Monthly_Sales,
        ROUND(PERCENT_RANK() OVER (ORDER BY ROUND(SUM(Weekly_Sales), 2)), 4) AS quartile
    FROM train t
    GROUP BY 1, 2, 3
    ORDER BY 4 DESC, 1
)
SELECT
    month_num,
    month,
    season_num,
    Monthly_Sales
FROM monthly_sales
-- Only gives months that are underperforming.
WHERE quartile < 0.25;
```

QUESTIONS TO CONSIDER

DO CERTAIN TYPES OF STORES PERFORM BETTER THAN OTHERS?

IS THERE A CORRELATION BETWEEN TEMPERATURE AND SALES?

ARE ANY MONTHS SEASONS CONSISTENTLY OUTPERFORMING OR UNDERPERFORMING COMPARED TO THE AVERAGE MONTH/ SEASON?

WHICH HOLIDAYS PERFORM THE BEST FOR A GIVEN WEEK AND WHY?

DOES THE SIZE OF THE STORE IMPACT SALES?

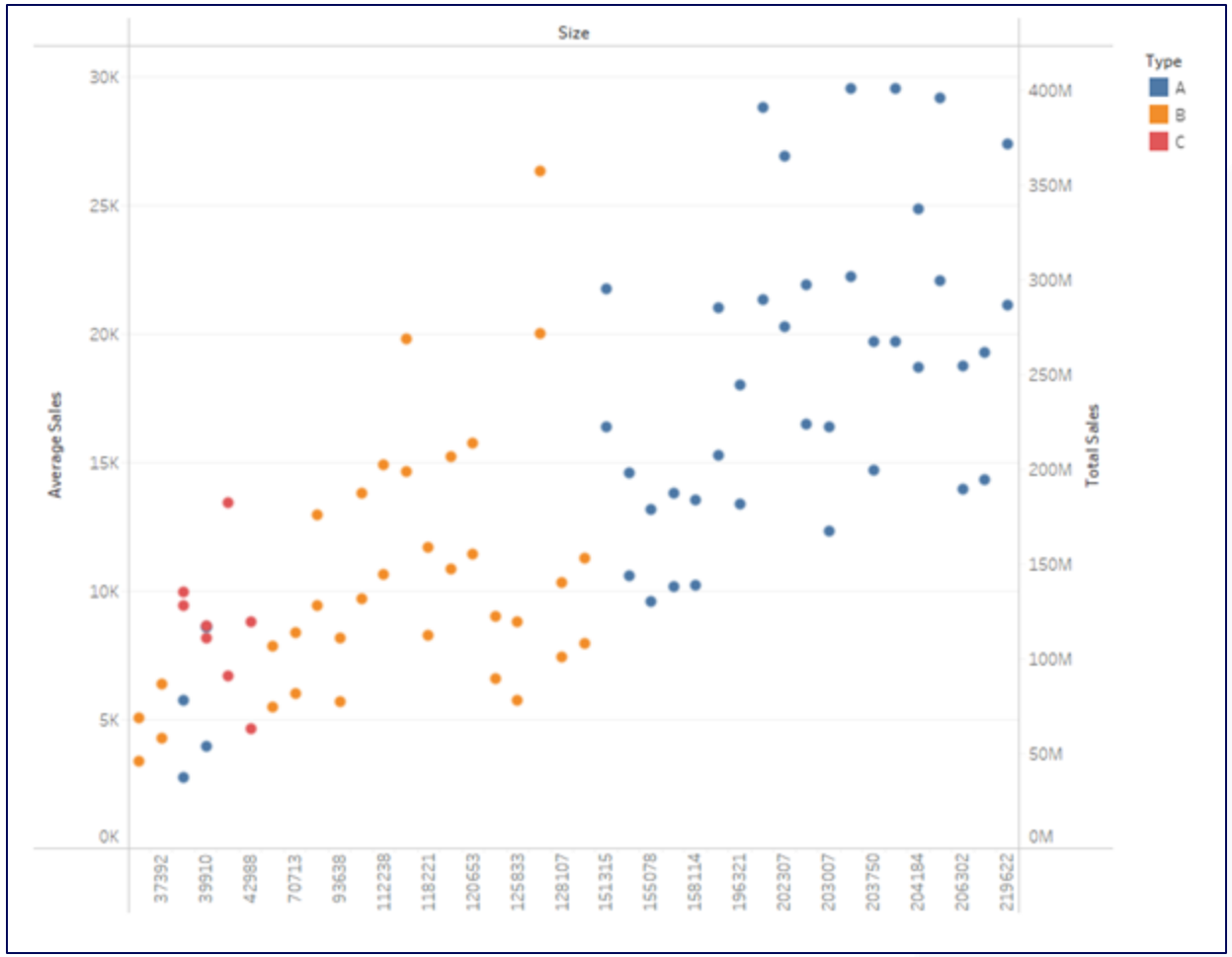
IS THERE ADDITIONAL INFORMATION/DATA THAT IS NEEDED TO MAKE FURTHER DECISIONS?

ANALYSIS & INSIGHTS
SALES ASSESSMENT

SUBTITLE GOES HERE

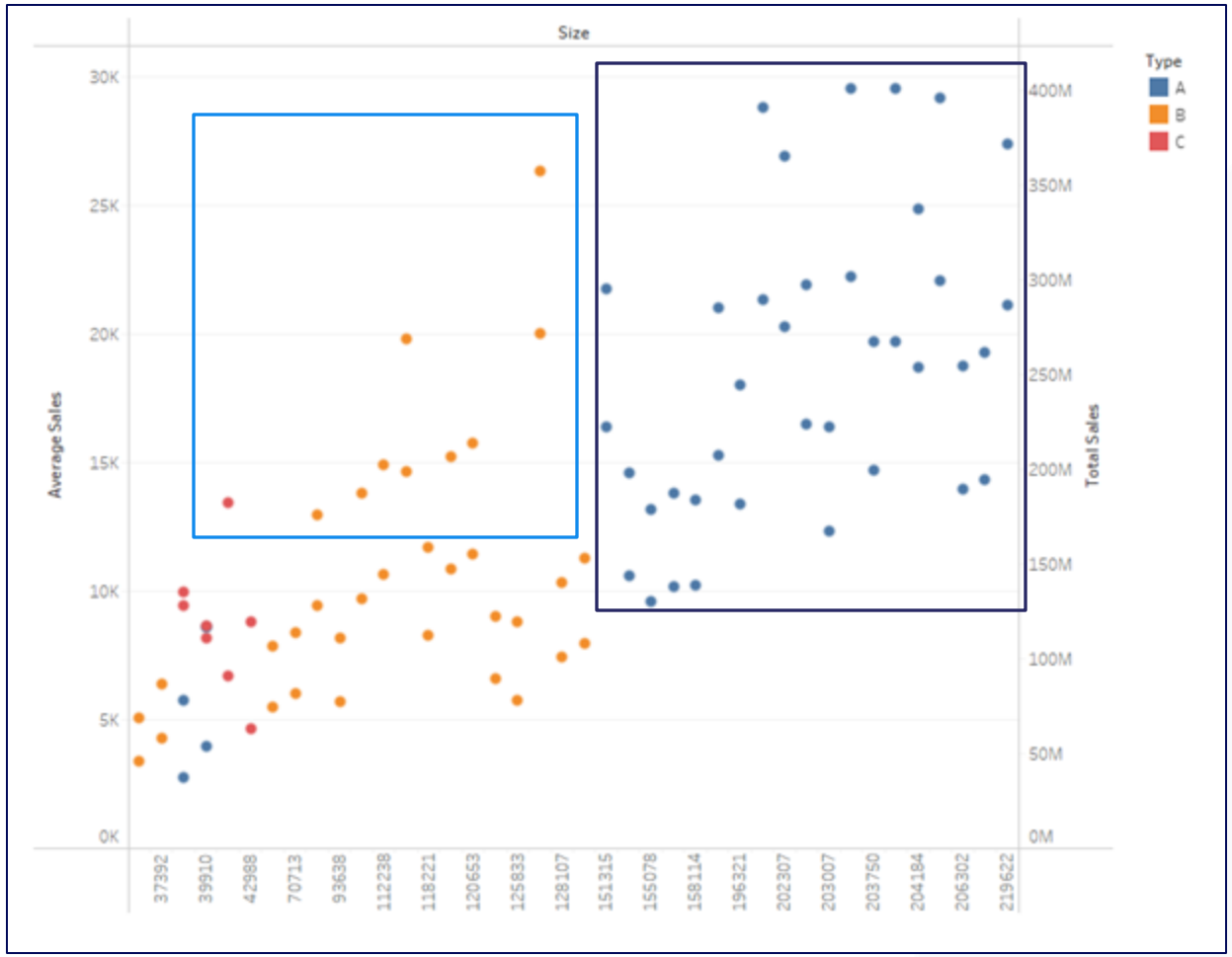
DATE 2023

STORE TYPE & SALES



- Data included three types of store A,B,C.
- Y Axis = Average Sales
- X-Axis = Store Size (per sq ft)
- **Type A (blue)** experienced the most sales, with an average of 206m, and is usually the largest at 175K sq ft per store.
- **Type B (orange)** has a size of 101K sq ft with average sales of 117m per store.
- **C (red)** tended to be the smallest at 41K sq ft with average sales of 101m per store.

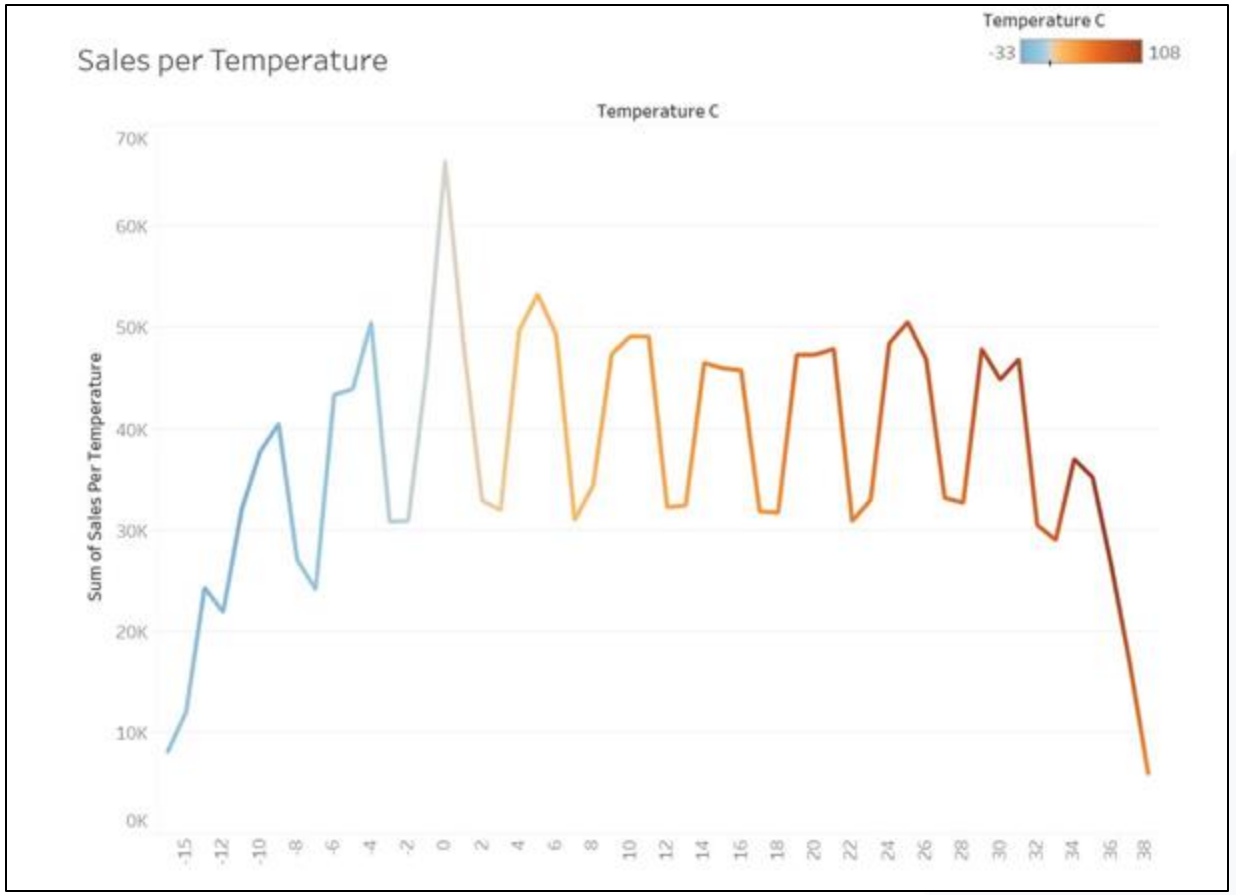
STORE TYPE & SALES [2]



Main hypothesis was that:
Greater Size = Greater Sales

- We can see that some Type B stores (light blue) see higher sales despite smaller sizes compared to Type A.
- Type C also maintained average sales per store similar to Type B despite significant difference in store size. Further investigation required.
- Although the size of stores clearly leads to more sales, does this necessarily lead to greater profitability? Need more data.

TEMPERATURE & SALES



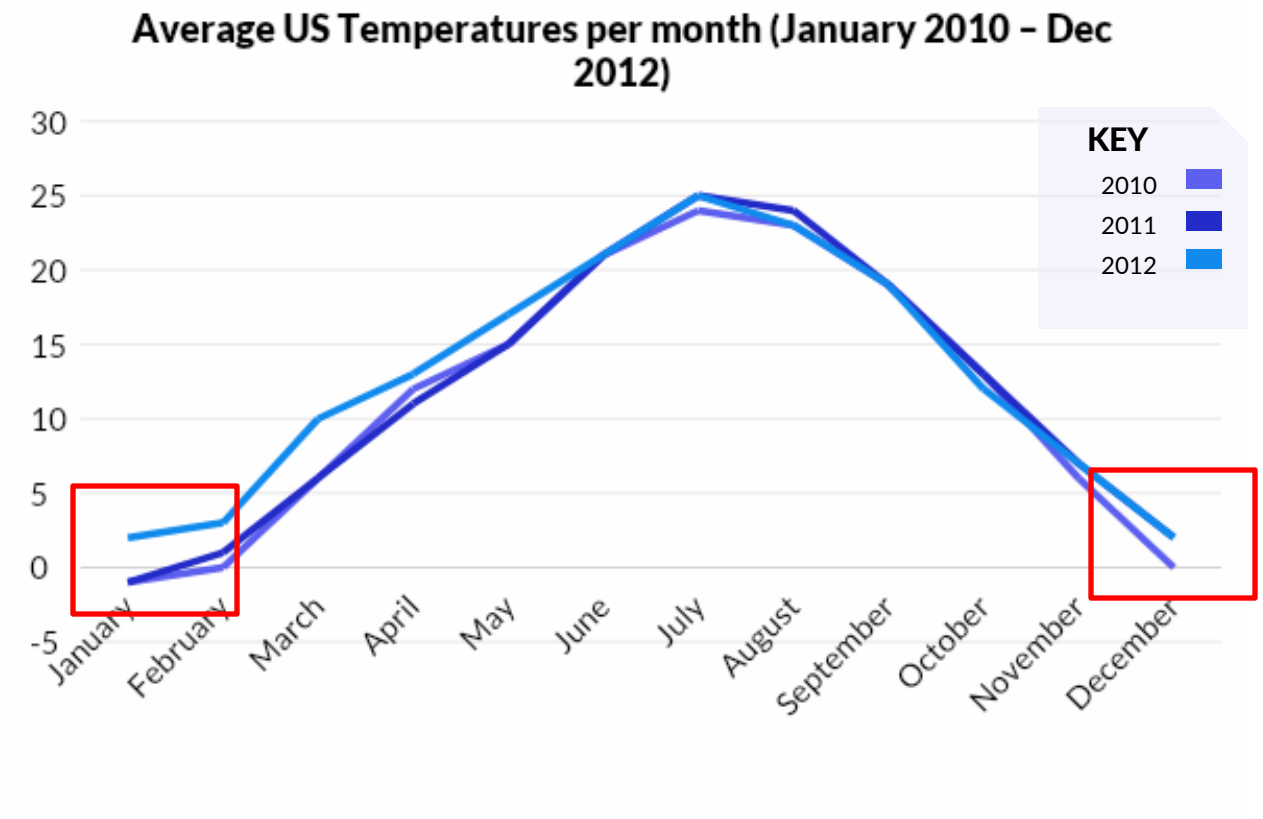
- Highest sales occur at 0 degrees Celsius (32 degrees Fahrenheit).

  One could argue that customers spend more at this temperature. However, we feel that this correlation does not necessarily suggest causation.

- We decided that further analysis would be required to better understand purchasing patterns and consumer behaviour.

- Fundamentally, we want to understand whether temperature is a factor in determining higher overall sales.

TEMPERATURES & SALES [2]

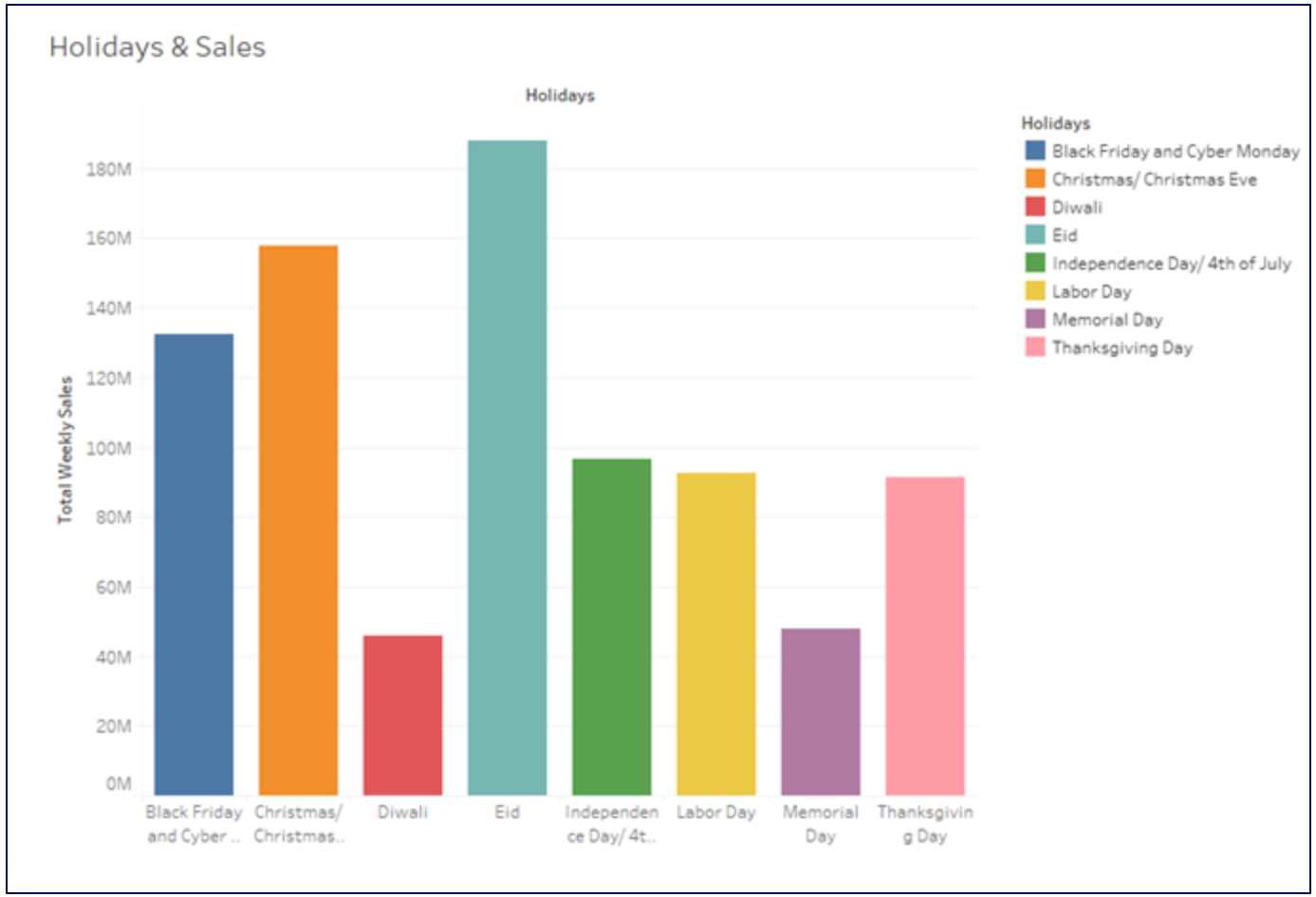


Source: US Average Temperatures by Month (2010-2012), National Centers for Environmental Information

Average temperatures across the US are around the 0 degrees Celsius mark (32 degrees Fahrenheit).

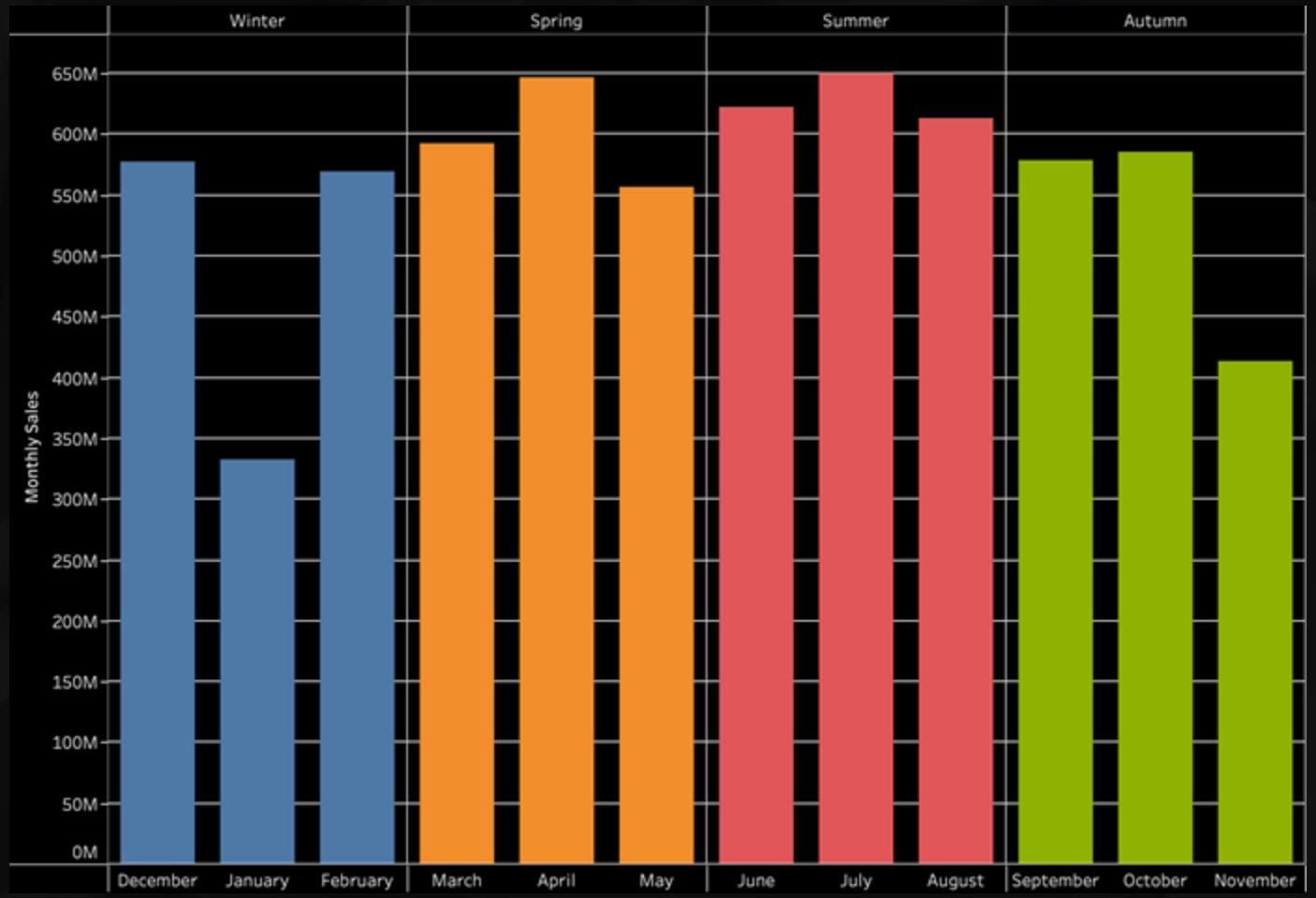
	2010	2011	2012
January	-1	-1	2
February	0	1	3
March	6	6	10
April	12	11	13
May	15	15	17
June	21	21	21
July	24	25	25
August	23	24	23
September	19	19	19
October	13	13	12
November	6	7	7
December	0	2	2

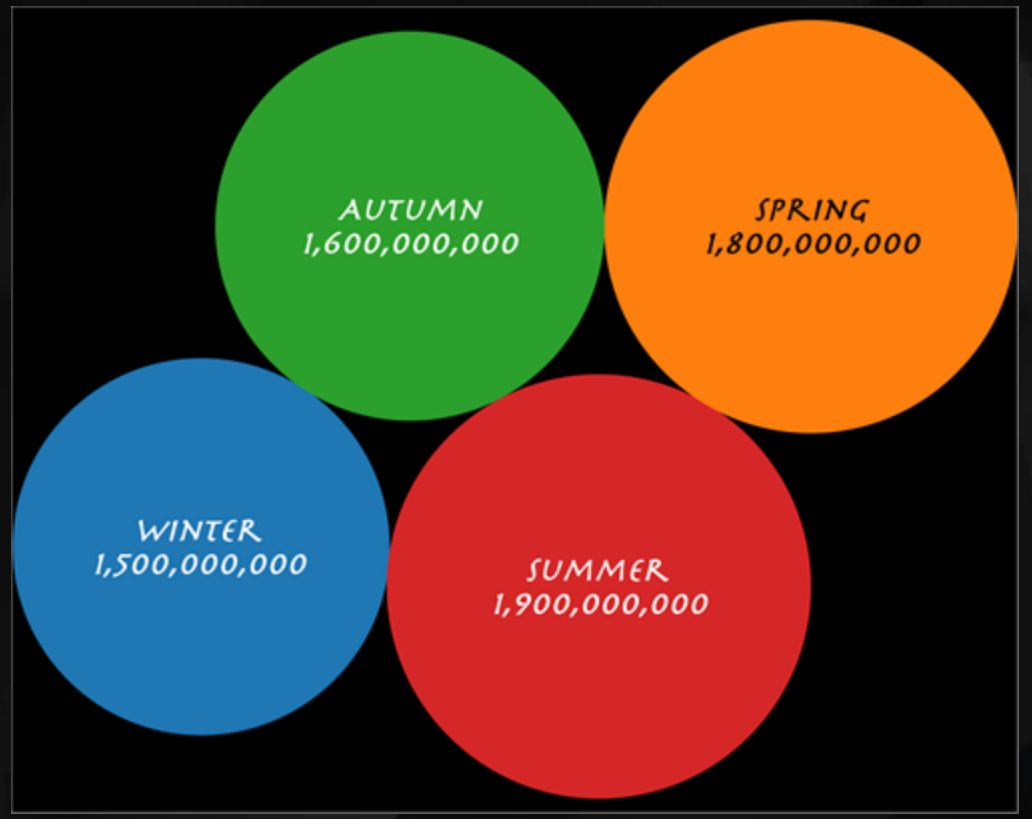
HOLIDAYS & SALES



- With this graph, we can see that instead of temperature being the factor on the number of sales, it is that holidays that drive sales.
- In the winter months which typically are around 0 degrees Celsius across the United States, this when the major holidays of Christmas and Thanksgiving, which typically involve buying gifts, food, and drink etc.

MONTHLY & SEASONAL SALES INSIGHTS





CONCLUSIONS
RECOMMENDATIONS

SUBTITLE GOES HERE

DATE 2023

KEY RESULTS

Type A | 64%

Type A stores contribute the most to sales with averages of 20K.

5% | 0.3bn

January average sales

6% | 0.4bn

November average sales

19% | 1.3bn

Sales contribution of underperforming months

29% | 1.9bn

Sales contribution of overperforming months

0.6bn

Average Monthly Sales

NEXT STEPS & RECOMMENDATIONS

Based on our assessment of the provided data, we would suggest the following to Walmart:

RECOMMENDATIONS

- **Early-Year Promotions:** Focus on strong promotional activities from January to April to capitalize on the natural sales increase.
- **Holiday Focus:** Enhance holiday marketing campaigns in December to boost sales further, leveraging the slight recovery seen after the November dip.
- **Promotional Strategies:** Targeted promotions during the January-April growth period and the mid-year stable period can maximize sales, while strategic discounts in November could help mitigate the sales dip.

BUSINESS IMPLICATIONS

- **Inventory Planning:** Walmart should focus on ramping up inventory before peak sales periods like April and maintain steady supply during the mid-year months to prevent stockouts.
- **Sustain Mid-Year Sales:** Maintain consistent inventory and targeted marketing during May to August to sustain the steady demand.
- **Address November Decline:** Investigate the reasons behind the November sales drop and consider strategies like early holiday promotions or special events to boost sales.

APPENDIX
LEARNINGS & LIMITATIONS

SUBTITLE GOES HERE

DATE 2023

LEARNINGS & LIMITATIONS

- **Type and Sales:** It was not clear what type represented (A,B,C), it seems to relate to size but the
- **Lower Sales Months:** No available data on profit or revenue. Despite lower sales, those quieter months of January and November could actually perform well for Walmart.
- **No Use of Inflation & Unemployment:** Not enough supporting data to generate meaningful insights.
- **Original Dataset:** Walgreen dataset was the original dataset but was scrapped due to the small sales time period covered (only a few minutes).
- **Blank cells:** caused problems during SQL analysis and had to be changed to NULL in Excel.

QUESTIONS?

SUBTITLE GOES HERE

DATE 2023