# DUNDER MIFFLIN REVIEW

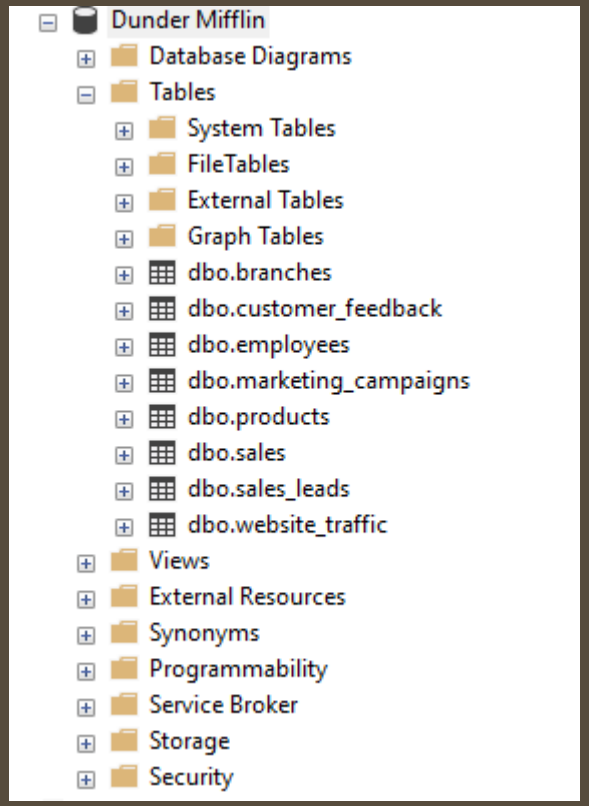
July 2025

# THE DATA

# AI GENERATED DATA

Created prompt to generate sales, marketing, and CRM data for the month of July 2025 for the fictional paper company Dunder Mifflin.

Prompt created eight CSV files. These were checked and edited in Excel before being saved and uploaded into MS SQL Server.



# CUSTOMER SATISFACTION

# FIRST STEPS

Queried the full table for Customer Satisfaction*

SELECT * FROM [Dunder Mifflin].[dbo].[customer_feedback]

Noticed that comments fall into one of six general categories (support/service, cost, shipping, product, website) and either a positive or negative tone.

I created a derived column to group comments into these categories.

* all data is for the month of July 2025

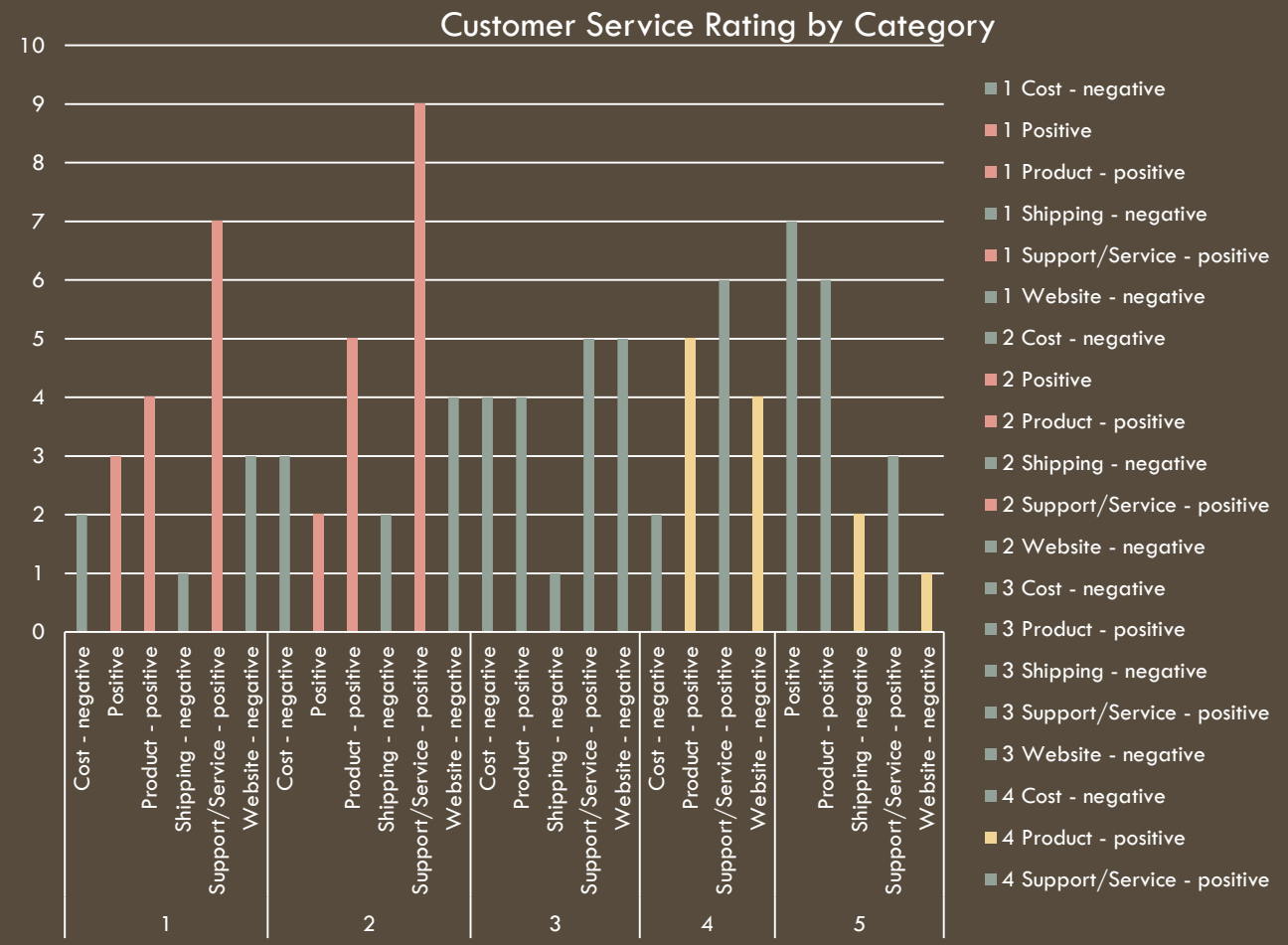
```
WITH Satisfaction_Category AS (
SELECT Satisfaction_1_5,
Comments,
    CASE
    WHEN Comments = 'Support was helpful.' THEN 'Support/Service - positive'
    WHEN Comments LIKE 'Great service!' THEN 'Support/Service - positive'
    WHEN Comments = 'Too expensive.' THEN  'Cost - negative'
    WHEN Comments LIKE '%shipping%delayed%' THEN 'Shipping - negative'
    WHEN Comments = 'Loved the product!' THEN  'Product - positive'
    WHEN Comments = 'Excellent paper quality.' THEN  'Product - positive'
    WHEN Comments LIKE '%website%slow%' THEN  'Website - negative'
    WHEN Comments = 'Will order again.' THEN 'Positive'
    ELSE 'Other'
 END AS Category
FROM [Dunder Mifflin].[dbo].[customer_feedback]),


)

SELECT * FROM Satisfaction_Category
```

# VISUALISATION

Extracted the Satisfaction_by_Category data to Excel.

Created a bar chart from a pivot table in order to visualize the customer service rating by category.

In mulitple cases, customers chose a low rating (1 or 2) when giving positive feedback, while in a few cases left negative feedback and scored service high (4 or 5).



# ONE POSSIBLE SOLUTION

Using SQL, I create a derived column in which low ratings with positive comments were rated higher i.e. a rating of 1 or 2 was converted into a rating of 5 or 4 respectively, and a high rating with a negative comment was converted to a  (core data was not altered).

Created a stored procedure as this may be called again mulitple times in future.

```
CREATE PROCEDURE Satisfaction_by_Category
AS
BEGIN

 WITH Satisfaction_Category AS (
 SELECT Satisfaction_1_5,
Comments,
    CASE
      WHEN Comments = 'Support was helpful.' THEN  'Support/Service - positive'
      WHEN Comments LIKE 'Great service!' THEN  'Support/Service - positive'
      WHEN Comments = 'Too expensive.' THEN  'Cost - negative'
      WHEN Comments LIKE '%shipping%delayed%' THEN  'Shipping - negative'
      WHEN Comments = 'Loved the product!' THEN  'Product - positive'
      WHEN Comments = 'Excellent paper quality.' THEN  'Product - positive'
      WHEN Comments LIKE '%website%slow%' THEN  'Website - negative'
      WHEN Comments = 'Will order again.' THEN 'Positive'
      ELSE 'Other'
END AS Category
FROM [Dunder Mifflin].[dbo].[customer_feedback]),

Satisfaction_Adjusted AS (
  SELECT Satisfaction_1_5,
    Comments,
Category,
    CASE
      WHEN Category LIKE '%positive%' AND Satisfaction_1_5 = 1 THEN 5
      WHEN Category LIKE '%positive%' AND Satisfaction_1_5 = 2 THEN 4
      WHEN Category LIKE '%negative%' AND Satisfaction_1_5 = 5 THEN 1
      WHEN Category LIKE '%negative%' AND Satisfaction_1_5 = 4 THEN 2
      ELSE Satisfaction_1_5
END AS Satisfaction_Adjusted_Positive_Negative
FROM Satisfaction_Category
)

SELECT * FROM Satisfaction_Adjusted

END;
```

# VISUALISATION 2

Again extracted data to Excel.

From pivot table created bar chart.

Conclusions:

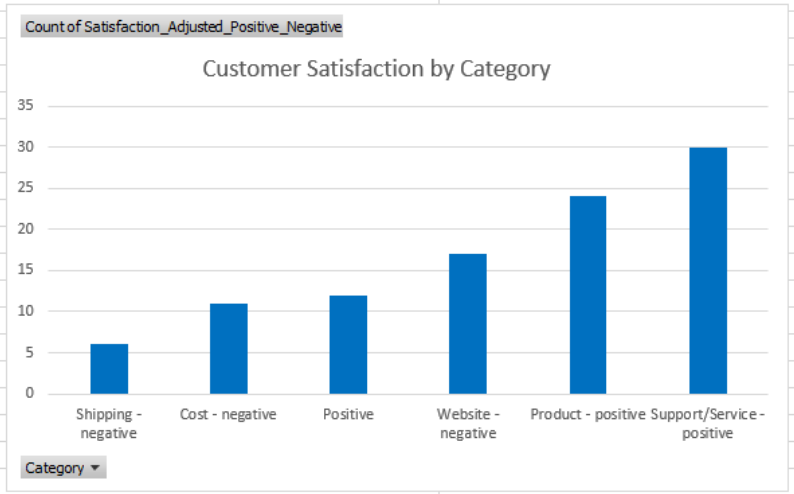
Overall customer ratings in the month of July were positive, however, there were multiple issues with the DM website, as well as some concerns about pricing and shipping delay.

Most customers were satisfied with the quality of DM products as well as DM support and service.

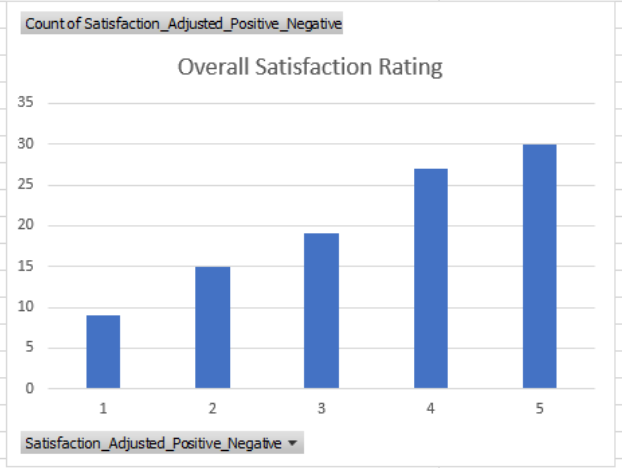
Recommendations:

- look into reasons website is slow and prioritise fixing it.
- (In longer project, would look at how website traffic compares to pervious months and how important is having a website to our customers / how do we compare to our competitors in this regard?)
- look into shipping delays
- double click into which products customers object to the cost and how to overcome objections so perceived value is higher than cost.

| Row Labels | Count of Satisfaction_Adjusted_Positive_Negative |
|---|---|
| Shipping - negative | 6 |
| Cost - negative | 11 |
| Positive | 12 |
| Website - negative | 17 |
| Product - positive | 24 |
| Support/Service - positive | 30 |
| **Grand Total** | **100** |




| Row Labels | Count of Satisfaction_Adjusted_Positive_Negative |
|---|---|
| 1 | 9 |
| 2 | 15 |
| 3 | 19 |
| 4 | 27 |
| 5 | 30 |
| **Grand Total** | **100** |



# CONVERSIONS

# WHAT IS THE SOURCE WITH MOST / FEWEST CONVERSIONS?

Context: There are 6 marketing sources (e.g. YouTube) with one possible medium (e.g. video ad) per source.

**Google CPC** has the **highest** total conversions while **Facebook** - Paid Social and **YouTube** - Video Ads have the **fewest** conversions.

```
SELECT DISTINCT
       [Source]
      ,[Medium]
      ,SUM([Conversions]) AS Total_Conversions
FROM [Dunder Mifflin].[dbo].[website_traffic]
GROUP BY [Source], [Medium]
ORDER BY SUM([Conversions]) DESC
```

| | Source | Medium | Total_Conversions |
|---|---|---|---|
| 1 | Google | CPC | 1409 |
| 2 | direct | none | 1212 |
| 3 | NYTimes | Display | 1148 |
| 4 | Email | Email | 1147 |
| 5 | YouTube | Video Ads | 926 |
| 6 | Facebook | Paid Social | 712 |

- Further investigation: Double click into these values. e.g. What is considered a conversion - what is the lifetime value for different conversions.
- What is the cost of CPC vs sales, Paid Social (Facebook) and Video Ads (YouTube) vs sales?
- Why are Facebook and YouTube ads underperforming?

# WHAT IS THE SOURCE WITH MOST / FEWEST SESSIONS AND CLICKS, AND HOW DO SESSIONS AND CLICKS RELATE TO CONVERSIONS?

Why do Fb and YT have such low engagement as well as low conversion. What are ways to improve this?

Is it possible that our target customers are congregating more on other social media sites?

| | Source | Total_Sessions | Total_Clicks | Total_Conversions |
|---|---|---|---|---|
| 1 | Google | 31684 | 29149 | 1409 |
| 2 | direct | 28893 | 29163 | 1212 |
| 3 | NYTimes | 23232 | 23289 | 1148 |
| 4 | Email | 28387 | 24252 | 1147 |
| 5 | YouTube | 17727 | 16443 | 926 |
| 6 | Facebook | 20999 | 16653 | 712 |

```
SELECT
      [Source]
      ,SUM([Sessions]) AS Total_Sessions
      ,SUM([Clicks]) AS Total_Clicks
      ,SUM([Conversions]) AS Total_Conversions

  FROM [Dunder Mifflin].[dbo].[website_traffic]
  GROUP BY [Source]
  ORDER BY SUM([Conversions]) DESC, SUM([Clicks]) DESC, SUM([Sessions])  DESC
```

# WHAT IS THE SOURCE WITH MOST / FEWEST SESSIONS AND CLICKS PER CONVERSION ?

Facebook has the highest number of sessions per conversions at 29, while YouTube has the fewest sessions and clicks per conversion at 19 and 17 respectively.

Therefore YouTube might be being underutilised, as it seems to be the most efficient source for converting, but has some of the fewest conversions.

DM should also assess their Fb campaign to understand why sessions and clicks per conversions are higher than other sources.

```
SELECT
      [Source]
      ,SUM([Sessions])/SUM([Conversions]) AS Sessions_per_Conversion
      ,SUM([Clicks])/SUM([Conversions]) AS Clicks_per_Conversion

  FROM [Dunder Mifflin].[dbo].[website_traffic]
  GROUP BY [Source]
  ORDER BY SUM([Clicks])/SUM([Conversions]) DESC
```

| | Source | Sessions_per_Conversion | Clicks_per_Conversion |
|---|---|---|---|
| 1 | direct | 23 | 24 |
| 2 | Facebook | 29 | 23 |
| 3 | Email | 24 | 21 |
| 4 | Google | 22 | 20 |
| 5 | NYTimes | 20 | 20 |
| 6 | YouTube | 19 | 17 |

# WHAT IS THE REGIONS WITH MOST / FEWEST SESSIONS AND CLICKS PER CONVERSION ?

Sessions and Clicks per ConVersion by Region

- The Southeast, along with being the region with highest number of conversions, is also the region with fewest clicks and sessions per conversion (18 and 17 respectively).
- In future DM may consider focusing more on other regions, as currently all of DM's branches are located in the NE.
- There is also a large number of conversions from Region Unknown, meaning we should try to find ways to identify these conversions if at all possible in future, e.g. if conversion = purchase of paper, where is the shipping address located?

```
SELECT
    [Region]
    ,SUM([Sessions]) AS Total_Sessions
    ,SUM([Clicks]) AS Total_Clicks
    ,SUM([Conversions]) AS Total_Conversions

FROM [Dunder Mifflin].[dbo].[website_traffic]
GROUP BY [Region]
ORDER BY SUM([Conversions]) DESC
```

| | Region | Total_Sessions | Total_Clicks | Total_Conversions |
|---|---|---|---|---|
| 1 | Southeast | 28368 | 26645 | 1505 |
| 2 | West | 37505 | 30696 | 1475 |
| 3 | Midwest | 32456 | 25654 | 1315 |
| 4 | Northeast | 25463 | 27986 | 1171 |
| 5 | Unknown | 27130 | 27968 | 1088 |

```
SELECT
    [Region]
    ,SUM([Sessions])/SUM([Conversions]) AS Sessions_per_Conversion
    ,SUM([Clicks])/SUM([Conversions]) AS Clicks_per_Conversion

FROM [Dunder Mifflin].[dbo].[website_traffic]
GROUP BY [Region]
ORDER BY SUM([Clicks])/SUM([Conversions])
```

| | Region | Sessions_per_Conversion | Clicks_per_Conversion |
|---|---|---|---|
| 1 | Southeast | 18 | 17 |
| 2 | Midwest | 24 | 19 |
| 3 | West | 25 | 20 |
| 4 | Northeast | 21 | 23 |
| 5 | Unknown | 24 | 25 |

# WHAT IS THE CAMPAIGN WITH MOST / FEWEST CONVERSIONS AND THE HIGHEST / LOWEST AVERAGE COST PER CONVERSION?

```
SELECT
      M.[Campaign_Name]
      ,SUM(W.[Conversions]) AS Total_Conversions
      ,M.[Budget]
      ,LEFT((M.[Budget]*1.00)/(SUM(W.[Conversions])*1.00),4) AS Avg_Cost_per_Conversion

  FROM [Dunder Mifflin].[dbo].[website_traffic] W
    JOIN [Dunder Mifflin].[dbo].[marketing_campaigns] M
          ON W.Campaign_ID = M.Campaign_ID
  GROUP BY  M.[Campaign_Name], M.[Budget]
  ORDER BY SUM([Conversions]) DESC
```

BY CAMPAIGN

| | Campaign_Name | Total_Conversions | Budget | Avg_Cost_per_Conversion |
|---|---|---|---|---|
| 1 | Recycled Awareness | 1403 | 6000 | 4.27 |
| 2 | Summer Blowout | 1217 | 4000 | 3.28 |
| 3 | Shred it All | 1087 | 7000 | 6.43 |
| 4 | Back to Work Promo | 903 | 8000 | 8.85 |
| 5 | Spring Sale 2025 | 817 | 5000 | 6.11 |

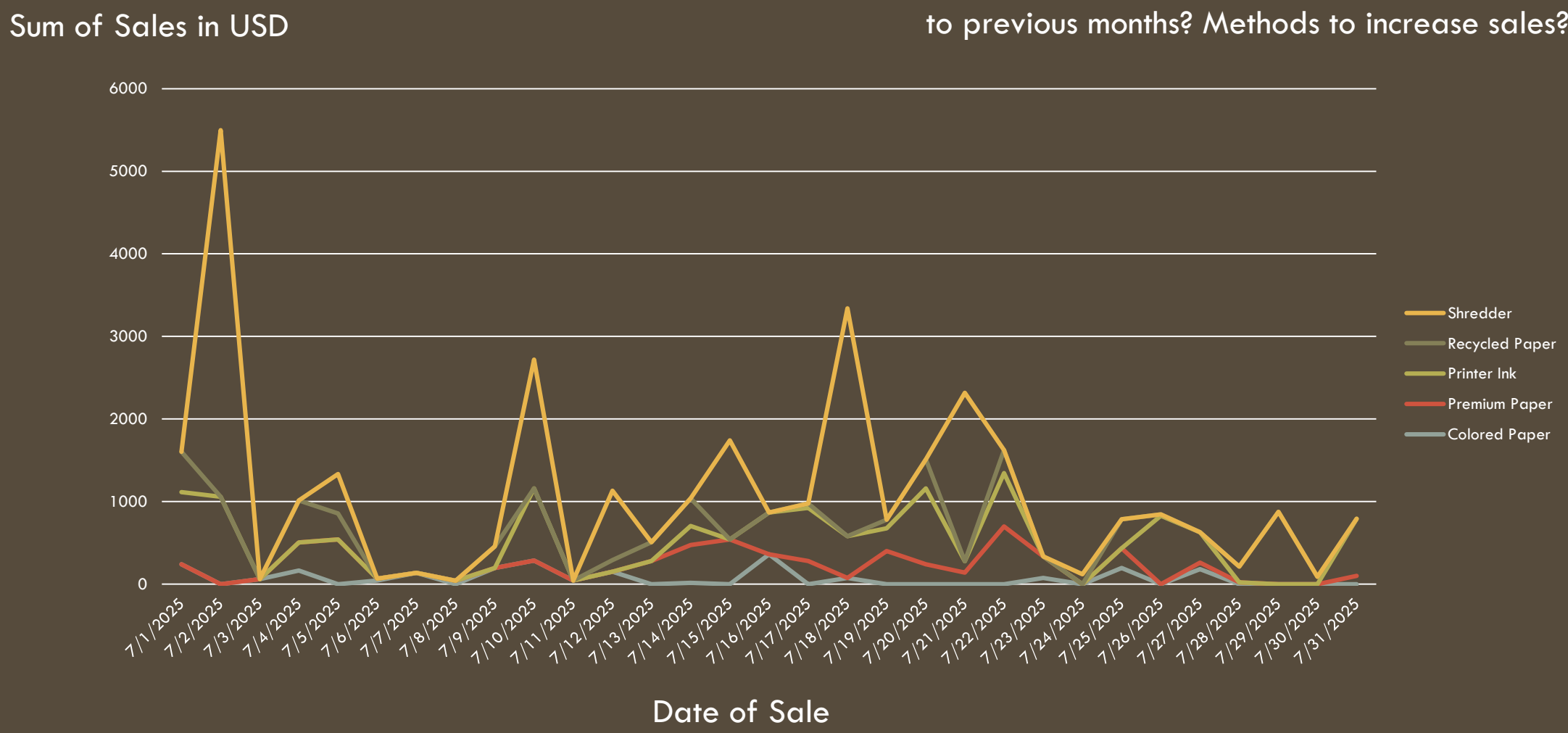
- Best performing campaigns were the "Recycled Awareness" campaign with a total of  1403 conversions at an avg cost of 4.27USD per conversion
- AND the Summer Blowout campaign which resulted in the second highest number of conversions at 1217 and had the lowest avg cost per conversion at 3.28 USD.
- The highest avg cost per conversion was 8.85USD for the Back to Work Promo which only produced a total of 903 conversions.
- Should assess both the success of the top two campaigns and the high cost and low conversion of the Back to Work Promo, which coincidentally had the highest budget.

# PRODUCT SALES

# JULY SALES AT A GLANCE

Note:

- There were several spikes for Shredder sales. Does this correlate in any way to a promotion, if not what could other causes be?
- Colored paper sales were low throughout the month of July, why might this be the case? How does this compare to previous months? Methods to increase sales?



# WHAT IS THE BEST / WORST - PERFORMING CATEGORY OF PRODUCT?

By Sales in USD, Equipment is the highest earning product category at 13.4k USD, while Paper is the best-selling at 644 units sold.

| | Category | Quantity_Sold | Sales_USD |
|---|---|---|---|
| 1 | Equipment | 112 | 13440 |
| 2 | Paper | 644 | 11480 |
| 3 | Supplies | 186 | 8554 |

```
SELECT
      P.[Category]
      ,SUM(S.[Quantity]) AS Quantity_Sold
      ,ROUND(SUM(S.[Sale_Amount]), 0) AS Sales_USD

  FROM [Dunder Mifflin].[dbo].[sales] S
    JOIN [Dunder Mifflin].[dbo].[products] P
        ON S.[Product_ID] = P.[Product_ID]
GROUP BY  P.[Category]
ORDER BY SUM(S.[Sale_Amount]) DESC
```

# WHAT IS THE BEST / WORST - PERFORMING PRODUCT?

The best performing by sales is the Shredder at 13.4k USD while Colored Paper is the lowest performing product both in terms of quantity sold (148 units) and sales (2.2k USD)

```
SELECT
      P.[Name]
      ,SUM(S.[Quantity]) AS Quantity_Sold
      ,ROUND(SUM(S.[Sale_Amount]), 0) AS Sales_USD

  FROM [Dunder Mifflin].[dbo].[sales] S
    JOIN [Dunder Mifflin].[dbo].[products] P
        ON S.[Product_ID] = P.[Product_ID]
GROUP BY  P.[Name]
ORDER BY SUM(S.[Sale_Amount]) DESC
```

| | Name | Quantity_Sold | Sales_USD |
|---|---|---|---|
| 1 | Shredder | 112 | 13440 |
| 2 | Printer Ink | 186 | 8554 |
| 3 | Premium Paper | 234 | 4678 |
| 4 | Recycled Paper | 262 | 4582 |
| 5 | Colored Paper | 148 | 2220 |

# WHO IS THE BEST / WORST - PERFORMING SALESPERSON?

Jim Halpert is top Salesman at 17,019 USD while Ryan Howard is worst-performing Salesman for July at 3,515 USD.

How does July compare to other months for Ryan? Does he need further training?

Note: Several Salespeople are missing – are they on holiday? Is the data on them missing or is there an error? This needs to be addressed.

| | Salesperson | Total_Product_Sold | Total_Sales_USD |
|---|---|---|---|
| 1 | Jim Halpert | 428 | 17018.98 |
| 2 | Phillis Vance | 175 | 6852.87 |
| 3 | Michael Scott | 166 | 6087.25 |
| 4 | Ryan Howard | 173 | 3515.08 |

```
SELECT E.[Name] AS Salesperson
      ,SUM(S.[Quantity]) AS Total_Product_Sold
      ,ROUND(SUM(S.[Sale_Amount]), 2) AS Total_Sales_USD

 FROM [Dunder Mifflin].[dbo].[sales] S
    JOIN [Dunder Mifflin].[dbo].[products] P
        ON S.[Product_ID] = P.[Product_ID]
    JOIN [Dunder Mifflin].[dbo].[sales_leads] L
        ON S.[Lead_ID] = L.[Lead_ID]
    JOIN [Dunder Mifflin].[dbo].[employees] E
        ON L.[Assigned_To] = E.[Employee_ID]

 GROUP BY E.[Name]
 ORDER BY SUM(S.[Sale_Amount]) DESC
```

# POSSIBLE FURTHER QUERIES AND SUGGESTIONS

Do sales per product category correlate with profit e.g. is the shredder bringing in the highest profit as well as the most sales? – if so recommend to push sales of shredders more.

Why are sales for colored paper so low and how to improve them?

How can we improve sales for supplies?

How do sales trends for July compare to the rest of this year? How do they compare to July of last year?

# METHODS

MS SQL Server
Excel