

In this document, I provide the detailed methods and approaches I used to solve tasks in Excel.

### Task 1 Data Preprocessing and Cleaning

In the Social Media Analytics Dataset workbook, there are three worksheets: Posts, Engagement Summary, and Campaigns Metadata.

#### SubTask 1

➤ **Remove any duplicate post entries.**

I selected all the data and tables, then went to the "Data" tab. Under the Data Tools section, I chose "Remove Duplicates." I repeated this process for each of the other worksheets. Thankfully, there were no duplicates found anywhere in the entire set of worksheets.

➤ **Standardise date formats and platform names.**

In the "Date" column, all the dates were initially in text format, so I used the "Text to Columns" function to convert them. First, I selected the entire Date column and went to the "Data" tab. Under the Data Tools section, I clicked on "Text to Columns." This opened a dialog box where I selected "Delimited" and clicked "Next." In the "Delimiters" section, the default option was "Tab," so I unchecked that box and clicked "Next."

In the "Column Data Format" section, I selected "Date" and then clicked "Finish." This conversion changed the text into actual date formats.

Next, to standardise the date format to (yyyy-mm-dd), I selected the Date column (excluding the header). Then, I right-clicked, chose "Format Cells," selected "Date" under the Category menu, and picked "YYYY-MM-DD" from the Type menu before clicking "OK."

I repeated this process for all date entries throughout the table.

To standardise the platform names, I first applied a filter to determine the number of unique platform names in the worksheet. Fortunately, all platforms are consistently named as Instagram, Facebook, Twitter, and YouTube..

➤ **Ensure numeric columns like Likes, Reach, Impressions, and Ad Spend have proper formatting.**

I selected all number columns and converted them to Numerical Format with zero decimal places. Additionally, I formatted the Ad Spend and Total Budget columns in Indian Currency.

➤ **Split hashtags into individual tags.**

To split the hashtags into individual columns, I utilised the “Text to Columns” function. First, I selected the Hashtag\_Used column, then went to the Data tab and clicked on Text to Columns. I chose the Delimited option and selected a comma as the delimiter since the existing hashtags were separated by commas. After that, I clicked on Finish. Finally, I renamed the new columns to reflect their content, labelling them as Hashtag 1, Hashtag 2, and so on.

## Task 2 Engagement Analysis

Data Sheet: Posts

- **Calculate the average engagement rate per platform: Engagement Rate = (Likes + Shares + Comments) / Impressions**

I created a new column labelled "Engagement Rate," calculated using the formula  $\text{=(F2 + G2 + H2) / I2}$ , as specified in the task. I then established a Pivot Table to display the average engagement rate for each platform. In the Pivot Table fields, I placed "Platform" in the Rows and "Engagement Rate" in the Values section, setting the value field to display the average. I also sorted the data from largest to smallest.

- **Identify the top 10 posts with the highest engagement.**

To identify the top 10 posts, I sorted the Engagement Rate column in descending order. After that, I highlighted the top 10 cells in engagement using conditional formatting, specifically the Top 10 Items feature for highlighting..

- **Create a pivot table that shows:**

- **Total Likes, Shares, and Comments by Content Type and Platform.**

I have created a Pivot Table for this analysis. In the Pivot Table field, I added the Platform column to the Rows, followed by the Content Type column. This layout is clear and makes it easy to understand which content types are associated with each platform. In the Values section, I included the Like, Share, and Comments metrics, and I set the default value field to calculate the Total using the SUM function.

- **Average Clicks per post by Hashtag.**

To combine my split hashtag columns into rows, I used the Unpivot function in Power Query. Here's how I did it:

1. I selected the Posts Table and navigated to the Data tab.
2. Under the Get & Transform section, I clicked on 'From Table/Range.' This opened the Power Query Editor with the loaded Posts Table.
3. I selected all the Hashtag columns, then went to the Transform tab and chose 'Unpivot Columns' from the options.
4. This resulted in two new columns: the first is labelled 'Attribute' and the second is 'Values' (the main column). Since I only needed the values, I removed the 'Attribute' column and renamed the 'Values' column to 'Hashtags.'
5. Next, I went to the Home tab and clicked on 'Close & Load,' which created a new sheet in the workbook.
6. I then used this new sheet to create a Pivot Table. In the Pivot Table field, I placed 'Hashtags' in the Rows section, and set the Values field to display the Average.

This process effectively transformed my split hashtag columns into a cohesive row format for analysis.

➤ **Highlight top-performing hashtags using a formula-based ranking.**

I created a new column called "Rank" next to my pivot table that shows the Average Clicks per Post by Hashtag. To make the ranking interactive, I used the following formula: “=RANK.EQ(N5, \$N\$5:\$N\$10, 0)”. Additionally, I applied formula-based conditional formatting to highlight the top five performing hashtags in the pivot table. The formula for the conditional formatting is: “=RANK.EQ(\$N5, \$N\$5:\$N\$10) <= 5”.

## Task 3 Social Media Platform analysis

Data Sheet: Engagement Summary

➤ **Identify which Platform generates the highest engagement.**

I created a Pivot Table and organised it by placing the Platform in the Rows and the Engagement Rate in the Values, setting the value field to Average. This allows me to calculate the true average engagement rate for each platform. To identify which platform has the highest engagement, I sorted the Average Engagement Rate column from largest to smallest.

➤ **Compare Growth rates of followers across different platforms.**

Firstly, I created two new columns: the first is "Net Follower" and the second is "Net Growth Rate."

Net Follower: For this column, I calculated the Net Follower using a simple formula: New Followers minus Unfollowers. In my case, the formula is "**=C2-D2**".

Net Growth Rate: I calculated the Net Growth Rate using Net Followers and Total Followers. The formula I used is Net Followers divided by Total Followers, which in my case is "**=H2/E2**".

After that, I changed the format of the Net Growth Rate to Percentage, as I need this metric for comparison purposes, and it must be presented as a percentage.

Next, I created a Pivot Table for comparison using the new columns. In the Pivot Table, I placed "Platform" in the Rows field and "Net Growth Rate" in the Values field, setting the value field to SUM for the entire comparison. To enhance the comparison, I also inserted a Pivot Chart based on the Pivot Table.

➤ **Visualise the engagement by ad spend across the Platform.**

In this task, I started with a blank scatter plot and manually created a data series for each platform. First, I sorted the data in ascending order by platform (A to Z). I then applied a filter to focus on a specific platform, for example, Instagram.

Next, I selected the blank chart and went to the Chart Design tab. I clicked on "Select Data," then clicked on "Add" to create a new data series. I named the series "Instagram." For the series X values, I selected the Ad Spend data, and for the Y values, I selected the Engagement Rate. After that, I clicked "OK."

I repeated the same process for the other platforms. Finally, I enabled the Axis Title, Chart Title, Gridlines, and Legend on the chart.

## Task 4 Hashtags and Content Strategy

Dats Sheet: Posts

### ➤ **Identify**

#### ○ **Most frequently used hashtags.**

To analyse the data, I utilised the Hashtags Analysis Table created in Task 2, which includes the average clicks per post by hashtag. I generated a pivot table using the hashtag analysis, where I placed the hashtags in the Rows and the Post ID in the Values, setting it to count the frequency of each hashtag. Finally, I sorted the pivot table from largest to smallest to identify the most frequently used hashtags.

#### ○ **Average performance (likes, comments) of posts containing each hashtag.**

For this analysis, I utilised the Hashtags Analysis Table. I created a Pivot Table in the Pivot Table field, placing Hashtags in Rows and averaging Likes and Comments in the values.

### ➤ **Compare the performance of videos vs. images vs. carousels across platforms.**

I utilised the Posts data sheet to create a Pivot Table. In this table, I placed the Platform in the Rows, the Content Type in the Columns, and set the Engagement Rate in the Values section, adjusting it to calculate the Average. The filter only displays Videos, Images, and Carousels. Additionally, I ranked the platforms to identify which one performs the best.

### ➤ **Recommend which content types should be prioritised on which platform.**

In the text box, I have provided the recommendations.

## Task 5 Campaign Effectiveness

Data Sheets: Campaign Metadata, Posts, and Engagement Summary

### ➤ Calculate:

- **Total and average impressions, likes, and clicks per campaign.**

To accomplish this, I created a Pivot Table.

1. Open the Posts sheet and select the Posts Table.
2. Go to the "Insert" tab and choose "Pivot Table."
3. In the Pivot Table Field, drag "Campaign Name" into the Rows section.
4. Place "Impressions," "Likes," and "Clicks" into the Values section and set them to calculate the SUM.
5. Repeat by adding "Impressions," "Likes," and "Clicks" again into the Values section, but this time set them to calculate the AVERAGE.

### ➤ Engagement uplift during vs. before the campaign period.

I have created a table in the Campaign Result Sheet. In this table, I copied the first three columns from the Campaign Metadata, which are Campaign Name, Start Date, and End Date, to form the Engagement Uplift Table. Additionally, I have included several other columns: Campaign Days, Pre-Campaign Engagement, Engagement During, Engagement Before, Daily Engagement During, Daily Engagement Before, and Engagement Uplift %. Let's discuss each of these columns below.

**Campaign Days:** This information is crucial for my calculations, as I need to know the total number of days each campaign runs. I calculated this using a straightforward formula: Start Date - End Date + 1. The addition of +1 ensures that both the start and end dates are included in the count. In my case, the formula is: **"=C2 - B2 + 1."**

**Pre-Campaign:** It is also important to calculate data before the campaign begins. To do this, I have used a simple logic: Start Date minus Campaign Days. In my case, the formula is **"=B2 - D2."**

In the Posts Sheet, I need to create a new column labelled "Engagement." I have calculated Engagement using the following formula: Likes + Shares + Comments.

**Engagement During:** This is the most important column for this task. I have calculated it using the following logic and formula: **"=SUMIFS(Posts!\$L:\$L, Posts!\$C:\$C, ">=" & B2, Posts!\$C:\$C, "<=" & C2)"**.

In this formula, the SUMIFS function is applied based on specific conditions. The first condition checks that the date in the Posts sheet is greater than or equal to the Start Date. The second

condition verifies that the date in the Posts sheet is less than or equal to the End Date. If both conditions are true, it calculates the total engagement during the campaign period.

**Engagement Before:** The same logic applies here, but with some modifications. The formula used is: “=SUMIFS(Posts!\$L:\$L, Posts!\$C:\$C, ">=" & (B2-30), Posts!\$C:\$C, "<=" & B2)”

In this case, the logic remains consistent with the previous one. The first condition checks that the date in my posts sheet is greater than or equal to (Start Date - 30), while the second condition ensures that the date in the Posts Sheet is less than or equal to the Start Date. If both conditions are true, the formula calculates the total engagement during the campaign period. The 30-day window is considered to capture any engagement leading up to the start of the campaign.

**Daily Engagement During:** This metric is crucial for understanding engagement uplift, as each campaign has a different duration. To calculate this, I used a simple formula: “=F2/D2”. This formula represents Engagement During divided by Campaign Days.

**Daily Engagement Before:** This metric is crucial for boosting engagement. Since each campaign lasts a different number of days, I calculated it using a simple formula: “=G2/D2”, where G2 represents Engagement Before and D2 represents the number of campaign days.

**Engagement Uplift:** This column displays the engagement uplift for this task. I calculated it using the following formula: “=(H2-I2)/H2\*100”. The logic behind this formula is to find the change in daily engagement: (Daily Engagement During - Daily Engagement Before) divided by Daily Engagement During, multiplied by 100 to express the result as a percentage.

➤ **Provide insights into:**

- **Which campaign had the highest ROI in terms of engagement vs. budget.**

To create the ROI Table, I copied two columns: the campaign name and Total Budget from the Campaign Metadata. In this new table, I have included three columns: Total Engagement, ROI, and Rank.

**Total Engagement:** I calculated this using the following formula: “=SUMIFS(Posts!\$L\$2:\$L\$301, Posts!\$M\$2:\$M\$301, "=" & A17)”.

In this formula, I used the SUMIFS function because I have a specific condition for the calculation. The condition is that the campaign name in my Posts sheet must match exactly. If



this condition is met, the formula calculates the total engagement for that particular campaign name.

**ROI:** It is calculated using the formula "**=C17/B17.**" This represents the Total Budget divided by Total Engagement, illustrating the Return on Investment.

**Rank:** To determine which campaign had the highest ROI regarding engagement versus budget, I calculated a ranking using a formula and sorted it in ascending order.

- **Which campaign had the strongest follower growth impact.**

I copied four columns from the Engagement Uplift Table: Campaign Name, Start Date, End Date, and Campaign Days. Using this information, I created a new table called the Growth Impact Table. In this table, I added two additional columns: Follower Growth and Rank.

**Follower Growth:** This is important for the task requirements, so I calculated it using the following formula: **'=SUMIFS('Engagement Summary'!\$H:\$H, 'Engagement Summary'!\$A:\$A, ">=" & B32, 'Engagement Summary'!\$A:\$A, "<=" & C32)'**.

In this formula, the logic is based on the SUMIFS function, which performs calculations based on specified conditions. The first condition checks if the date in the Engagement Summary sheet is greater than or equal to the Start Date. The second condition checks if the date in the Engagement Summary sheet is less than or equal to the End Date. If both conditions are true, it calculates the total follower growth during the campaign period.

**Rank:** To identify which campaign had the greatest impact on follower growth, I have calculated a ranking based on a specific formula and sorted the results in ascending order.

## Task 6 Follower Retention & Audience Loyalty

Data Sheets: Engagement Summary

➤ **Visualise weekly follower growth trends by platform using line charts.**

First, I need to calculate the Net Followers.

I have created a Pivot Table in which I placed the week start date in the rows.

Additionally, I created two groups based on the week start date: Years and Months, to enhance interactivity. I added the platform in the columns and set the Net Followers in the values section to be summed.

I also created a Pivot Line and Markers chart, enabling the legend, axis titles, and chart title for clarity. Furthermore, I added a slicer for the platform, allowing me to filter and view data for one or two platforms, which improves interactivity.

➤ **Identify the week with the highest net follower gain.**

I have created a Pivot Table. In the Pivot Table field, I added the week start date to the rows and set the Net followers to the values field, using the SUM function. Additionally, I created a ranking column to sort and identify the weeks with the highest net follower gain. I also applied formula-based highlighting to the rows.

➤ **Create a moving average chart to smooth out follower growth trends.**

I have created a Pivot Table where I placed the Week Start Date in the rows. Additionally, I organised the dates into two groups: Years and Months based on the Week Start Date to enhance interactivity. In the Values section, I set the Net Followers to SUM.

Furthermore, I built a Pivot Line Chart and enabled the Legend, Axis Title, and Chart Title. A key feature of the chart is the Trend Line, with the trend type set to Moving Average over a four-period interval, which represents the monthly Moving Average. Lastly, I created a Slicer for the Platform, allowing me to filter the data by a specific platform, making the chart more interactive.

➤ **Determine the correlation between ad spend and follower growth using scatter plots and basic correlation calculations.**

I selected "Ad Spend" and "Net Follower" data. Then, I went to the "Insert" tab and chose "Scatter Plot" under the chart options, which created a scatter plot. Next, I added a trend line by selecting "More Options" and choosing the linear type, which is set by default. I checked the box to display R-squared values on the chart, and now I have a linear trend line along with the correlation value. Additionally, I calculated the basic correlation between Ad Spend and Net Followers using the CORREL function.