



Performance Marketing Analysis Documentation

Data Analysis Track Final Project



May 17, 2025

Digital Egypt Pioneers Initiative - DEPI

EYOUTH

List of Contents

1. [What is our project about?](#_What_is_our)   
   → Page: 2
2. [Project Overview](#_Project_Overview)  
   → Page: 3

1. [Data Exploration in SQL](#_Data_Exploration_in)   
   → Page: 7
2. [Preprocessing and Visualization in Excel](#_Preprocessing_and_Visualization)  
   → Page: 8
3. [Cleaning and Visualization in Python](#_Cleaning_and_Visualization)  
   → Page: 10
4. [Visualization in Power BI](#_Visualization_in_Power)  
   → Page: 12
5. [Insights and Recommendations](#_Insights_and_Recommendations)   
   → Page: 14

# **What is our project about?**

The objective of this project is to clean, analyse, and interpret performance marketing data to gain a deeper understanding of key trends and patterns. Through data visualization, we aim to identify significant insights that reveal the factors influencing marketing performance.

The final dashboard will serve as a comprehensive tool to communicate these findings effectively, highlighting correlations, anomalies, and emerging trends. Ultimately, the insights derived from this analysis will be used to make data-driven observations and provide strategic business recommendations to optimize marketing efforts.

Performance Marketing is a new and emerging field within the marketing industry and is predicted to be the biggest type of marketing in the future. Data and projects based on this concept are still quite limited despite the increasing demand for it by companies.

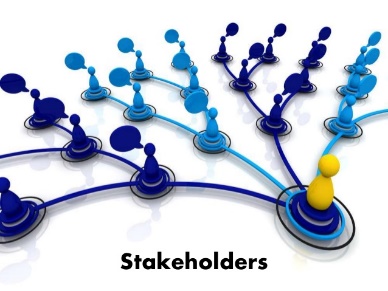
MRHM is a performance marketing company that focuses on elevating marketing performance. Our customers are companies that want to launch online advertising campaigns, and we help them optimize on their campaigns to achieve the highest profit possible.

# **Project Overview**

Team Members

* 1. Analysis and Leader: Malak Waheed
  2. Preprocessing: Hager Mostafa
  3. Visualization: Rahma Mahmoud
  4. Forecasting: Mariam Ahmed

Who is our target audience?

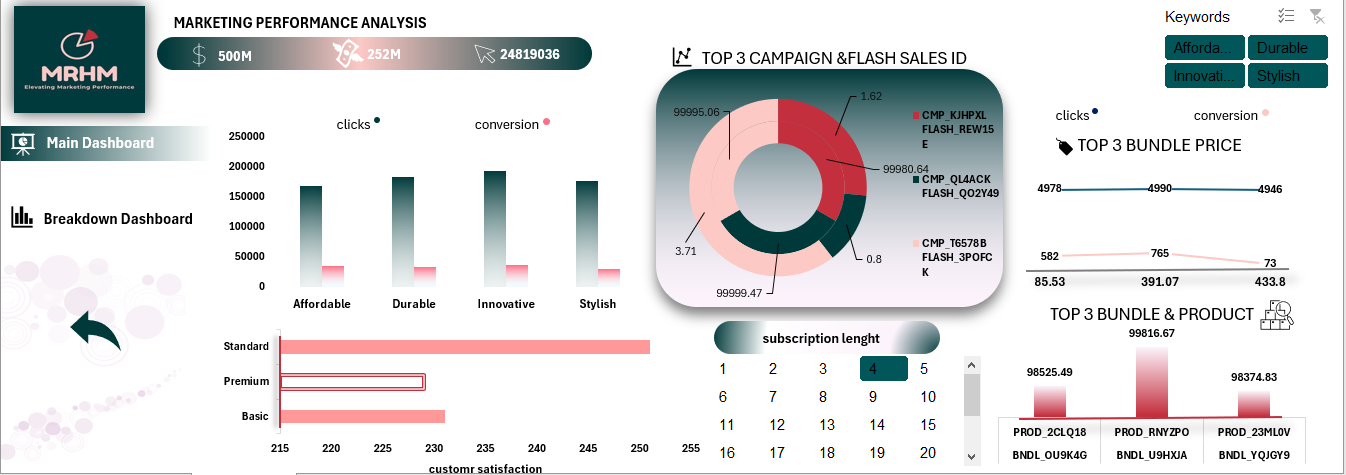
Our target audience are stakeholders and other senior employees in managerial positions within our company.

The dashboards created are easy to understand and focus on the overall picture of our data rather than the day-to-day transactions and records.

This means that the dashboards and insights derived from its focus on trends and patterns that affect the business decisions of the company.

Structure of our Data

Our database is a non-relational database that originally contained 17 columns and 10000 rows. Through cleaning and preprocessing we have added multiple columns and checked the data for any errors or inconsistencies. Key entities include Order ID, Customer ID and Product ID. The data is stored on both SQL database and Excel files.

Project Wireframe and Design

The picture above shows one of the Dashboards we created, this one is the main dashboard in our excel file. All dashboards contain multiple graphs of different types that allow for easier understanding and clear representation of trends.

There are slicers which allow the user to have more specific and specialized insights. Buttons allow for easy navigation, whether that be in moving to a different page in the dashboard or clearing any selections made.

Business Questions about our Data

1. What is the Sum of revenue, budget and average of ROI by subscription tier?
2. What is the Sum of clicks and average conversion by common keywords?
3. What are top 3 campaign and flash sales id by revenues and ROI?
4. What are the clicks, conversion by bundle price?
5. Do customers with high subscription tier have highest satisfaction post refund?
6. What are the top 5 discount level by revenue generated and sold units?
7. What are top 3 bundle and product by revenue?
8. What are the top 10 campaigns by revenue?
9. What are the campaigns with the highest ROI?
10. What are the campaigns with the highest conversion rate?
11. Analyze the impact of discount level on units sold
12. Get the top 10 bundles by units sold.
13. Get the top 10 bundles by revenue.

Tools and Technologies Used

* Python: preprocessing, cleaning and visualization
* Excel: cleaning, pivot tables, dashboard
* SQL: For querying and managing relational databases
* Jupyter Notebook : As the development environment for interactive coding and collaboration.
* GitHub: For version control and team collaboration
* Power BI: interactive dashboard
* Tableau: interactive dashboard

Project’s Key Deliverables

* Documentation is in 2 formats:
  1. Word document
  2. PowerPoint presentation
* Timeline for deliverables:
  1. Data Cleaning and Preprocessing (27 February 2025)
  2. Analysis and Insights (13 March 2025)
  3. Forecasting Questions (27 March 2025)
  4. Visualization and Reporting (10 April 2025)
* Deliverables:
  1. Python code (Jupyter Notebook)
  2. Excel File and Dashboard
  3. SQL Query File
  4. Power BI Dashboard
  5. There is also the GitHub link for the project: [GitHub-Link](https://github.com/Malak-Waheid/Performance-Marketing-Data-Analysis)

# **Data Exploration in SQL**

We have written 27 SQL Queries to explore the data and the trends in it. Each query answers a different business question and the Queries are divided into 3 categories.

Categories of SQL Queries

1. Basic Insights: these queries are the first ones since we are still exploring our data and still understanding what each part means and how all the rows relate to each other. Business Questions asked include: Count total number of campaigns
2. Top Performers: here the business questions revolve around us investigating who are top performers are in several rows and for different metrics and KPIS. Business Questions asked include: Get 5 campaigns with the highest ROI
3. Categorization Examples: here the business questions combine the previous two by investigating data from a new perspective: categories. Data is grouped on different criteria and the Business Questions asked include: Categorize discount level into low, moderate, and high

# **Preprocessing and Visualization in Excel**

New Columns Created

enhanced the dataset by adding two new columns to provide deeper insights:

1- Conversion Rate:

The formula used is =[@Conversions] /[@Clicks]

This allowed me to express the conversion efficiency as a percentage

2-Cost per Conversion:

The formula for this metric is =[@Budget] / [@Conversions]

] This metric helped in assessing the financial performance of the conversion process

.Additionally, I refined the formatting of the dataset’s columns to improve clarity. For example, I formatted the Revenue column to include a dollar sign (e.g., $1,234.56) for clear financial representation. The Conversion Rate column was formatted to display as a percentage (e.g., 25.00%) to align with standard reporting practices.

Preprocessing

In the analysis phase, created pivot tables to summarize key metrics, each paired with a chart for visualization:

1 Revenue, budget, ROI by Subscription: Pivot table showing Rev. , Budget , and ROI per subscription, visualized with a Combo

2 Clicks and Conversions by Top Keywords: Pivot table for Clicks and Conversions of the most common keywords, shown in a Column Chart.

3 Top 3 Campaigns by Rev. and ROI: Pivot table for the top 3 campaigns by Rev. and ROI, visualized with a Donut Chart.

4 Satisfaction Refund by Subscription tier: Pivot table for Satisfaction Refund per subscription, displayed in a Horizontal Bar Chart.

5 Top 5 Discounts by Rev. and Sales units: Pivot table for the top 5 discounts by Rev. and Sales units, shown in a Line Chart

6 Top 3 Bundles/Products by Rev. : Pivot table for the top 3 bundles/products by Rev. , visualized in a Column Chart.

7 Top 3 Bundle Prices by Clicks and Conversions: Pivot table for the top 3 bundle prices by Clicks and Conversions, displayed in a Line

Analysis

The data was exceptionally clean, meaning it was free from common issues such as missing values, inconsistencies, or formatting errors that often require preprocessing. Additionally, there were no duplicate entries, which ensured the integrity and reliability of the dataset. When I utilized the pivot query to explore and summarize the data, I found that it was already well-structured and organized

# **Cleaning and Visualization in Python**

Our cleaning

-Checks for missing values in the Data. The result is 0, meaning there are no missing values.2.

-Checks for duplicate rows. The result is 0, confirming no duplicates exist.

-Lists the column names of the Data, which include the metrics

-Displays the structure of the Data, confirming it’s a Pandas Data with the listed columns and their data types

.Summary of Operations: The code verifies that the dataset is clean (no missing values or duplicates) and provides an overview of its structure and columns, preparing it for further analysis.

Preprocessing

used pivot tables to summarize metrics like Revenue, ROI, Clicks, and Conversions by campaigns, subscriptions, and bundles. I also used cross tabs to compare categories like Discount\_Level vs. Common\_Keywords for Units\_Sold.

Pivot Table vs. Cross Tab: Pivot tables are flexible for aggregating numerical data (e.g., sum of Revenue) across indices, while cross tabs are simpler for comparing two categorical variables (e.g., sales by discount and keyword).

This approach revealed key trends and top performers, guiding data-driven decisions.

Analysis and visualization

* Cross-checked statistical distributions
* Validated business logic consistency
* Ensured temporal coherence (if time-based data)
* Confirmed key metric calculations
* Cross-checked statistical distributions
* Validated business logic consistency
* Ensured temporal coherence (if time-based data)
* Confirmed key metric calculations

# **Visualization in Power BI**

Power BI Components

* Home Page: This page is the first one that you see when you open Power BI and it shows the company logo and links to the dashboards.
* Who are we page: A page that explains the company and what services we offer.
* Dashboards: We have two main dashboards: Customer & Subscription Behavior and also Marketing & Sales Performance. Each dashboard focuses on a different part of the data.
* Recommendations Page: The final page that shows insights and recommendations based on our dashboards

Power BI Features Used

* Tooltip: Three gauge graphs appear when hovering over graphs that shows comparison between our chosen category and the full data.
* Drill down Graph: this allows for more detail and therefore more comprehensive understating of our data.
* Buttons:
  + Light/ Dark mode buttons: The page colors changes when you press on it
  + Navigation buttons: the buttons allow you to easily navigate between pages in your dashboard
  + Clear filters button: allows us to easily clear all filters on the page so we can see how it originally was.

KPI and DAX

1. Total Revenue = SUM(marketing\_and\_product\_performan[Revenue\_Generated])
2. Average ROI = AVERAGE(marketing\_and\_product\_performan[ROI])
3. Average Customer Satisfaction = AVERAGE(marketing\_and\_product\_performan[Customer\_Satisfaction\_Post\_Refund])
4. Total Units Sold = SUM(marketing\_and\_product\_performan[Units\_Sold])
5. Total Budget = SUM(marketing\_and\_product\_performan[Budget])
6. Average Conversion Rate = AVERAGE(marketing\_and\_product\_performan[Conversion\_rate])
7. Average Discount Level = AVERAGE(marketing\_and\_product\_performan[Discount\_Level])
8. Average CPC = AVERAGE(marketing\_and\_product\_performan[Cost\_per\_conversion])
9. Number of Campaigns = COUNT(marketing\_and\_product\_performan[Campaign\_ID])

# **Insights and Recommendations**

**1.** **Reallocate Budget Based on ROI**

* **Why**: ROI ranges from **0.5 to 5.0**, with a **median of 2.75**, showing significant variation. Some campaigns yield much better returns for similar budgets.
* **What to do**: Identify and invest more in high-ROI campaigns, and audit or scale down low-ROI ones.

**2. Address Low Conversion Efficiency**

* **Why**: The **average conversion rate is only 0.61**, but some campaigns with **very high clicks (e.g., 4946)** achieve **very few conversions (e.g., 73)**.
* **What to do**: Audit underperforming campaigns with high clicks but low conversion rates to improve landing pages or targeting.

**3. Optimize Product Bundling**

* **Why**: The correlation between Bundle Price and Units Sold is almost zero (**-0.0025**), suggesting price is not a major driver.
* **What to do**: Focus on the **value perception and relevance** of the bundles rather than just price.

**4. Tailor Offers by Subscription Tier**

* **Why**: Average subscription length is **about the same** (~18 months) across all tiers.
* **What to do**: Use tier-based promotions to upsell from **Basic** and **Standard** to **Premium**, since all stay loyal for a similar time.

**5. Use High-Converting Flash Sale Templates**

* **Why**: Some Flash Sale IDs consistently led to **999 conversions**, which is the maximum seen.
* **What to do**: Replicate the structure and timing of these successful flash sales in future campaigns.

**6. Improve Post-Refund Satisfaction**

* **Why**: Satisfaction is **evenly spread** between ratings of 1 to 4, with no one score clearly dominant.
* **What to do**: Investigate pain points in the refund process and improve post-sale support to shift scores toward 5.

**7. Double Down on Winning Keywords**

* **Why**: The most frequent keywords are:
  + "Affordable" (2559 times),
  + "Stylish" (2514),
  + "Innovative" (2491).
* **What to do**: Prioritize these terms in ads and SEO; they likely align well with customer expectations and drive engagement.