Facebook Page Metrics Analysis

Capstone Project Report

Prepared by Innocentia

Table of Contents

1. Introduction

This capstone project demonstrates an end-to-end data analysis of Facebook Page Posts data for a cosmetics brand.

The project satisfies Workplace Module Capstone Project Guidelines by showing the entire data science process-from loading to business recommendations.

This document provides a professional analysis for management, explaining both technical approach and business value.

2. Business Objective

The core business problem is to improve user engagement on Facebook.

By understanding drivers of interactions (likes, comments, shares), the business can:

- Increase organic reach

- Better allocate marketing budget

- Enhance brand awareness

Questions:

- Which types of posts generate more engagement?

- Does Paid promotion increase interactions?

- How can content strategy be improved?

3. Dataset Overview

Source: UCI Machine Learning Repository

~500 posts from a cosmetics brand page

Features:

- Page total likes

- Post Type (Photo, Link, Status, Video)

- Category

- Paid / Unpaid indicator

- Post month/day/hour

- Likes, Comments, Shares

- Total Interactions

This dataset captures real marketing activity for practical analysis.

4. Methodology

Structured workflow:

- Data Collection & Loading

- Data Cleaning & Preprocessing

- Exploratory Data Analysis (EDA)

- Statistical Testing

- Visualisation

- Insights and Recommendations

Ensures reproducibility and clarity.

5. Data Collection & Loading

Data loaded from CSV using pandas.

- Verified shape, columns, types

- Described variables

Example:

df = pd.read_csv('FacebookMetrics.csv')

df.info()

df.describe()

Outcome: dataset ready for analysis.

6. Data Cleaning & Preprocessing

Aligned with guidelines:

- Handling missing values: drop rows

- Removing duplicates

- Correct data types: Paid as 0/1

- Outlier removal with IQR

Example:

df = df.dropna().drop_duplicates()

df['Paid'] = df['Paid'].astype(int)

IQR filtering applied

Outcome: clean, reliable dataset.

7. Exploratory Data Analysis (EDA)

Uncovered patterns and relationships:

- Descriptive stats: mean, median, std

- Distribution: Histogram showed skew

- Post Type Counts: Photos most common

- Boxplots: Photos and Videos higher median interactions

- Correlation Heatmap: Likes, Comments, Shares strongly correlated

Example:

sns.histplot(df['Total Interactions'])

sns.countplot(x='Type', data=df)

sns.boxplot(x='Type', y='Total Interactions', data=df)

sns.heatmap(corr, annot=True)

Outcome: clear understanding of drivers.

8. Statistical Analysis

Hypothesis: Paid promotion increases Total Interactions.

Method: Independent two-sample t-test

Result: P-value > 0.05

Conclusion: No significant difference between Paid and Unpaid posts.

Example:

from scipy.stats import ttest_ind

t_stat, p_value = ttest_ind(paid, unpaid)

Outcome: Paid promotion effect is questionable.

9. Insights

- Post Type is a major driver of engagement

  Photos and Videos outperform Links/Status

- Paid promotion showed no significant lift

- Engagement is uneven; small % of posts drive most interactions

- Need for testing and refining strategy

10. Recommendations

- Prioritise high-quality Photos and Videos

- A/B test Paid promotion to validate ROI

- Monitor Likes, Shares, Comments continuously

- Invest in creative strategy

- Explore predictive modelling in future phases

11. Conclusion

Delivered end-to-end analysis aligned with Capstone guidelines:

- Data loading, understanding

- Cleaning and preprocessing

- EDA with visualisation

- Statistical testing

- Actionable business recommendations

Outcome: robust analysis for data-driven decisions.

12. Future Work

- Time series forecasting for post timing

- Predictive models for engagement

- Segmentation for personalised strategies

- Integration with other marketing channels