A DATA ANALYSIS
PROJECT REPORT
ON
SOCIAL MEDIA
ENGAGEMENT ANALYSIS

SUBMITTED BY

MEKHA MOLLY RAJU(**st90763)**

GUIDED BY:

PAVLYUK DMITRY

Transport And Telecommunication Institute Riga, Latvia

CERTIFICATE

This is certify that the project titled "**Social Media Engagement Analysis Using Python".** Submitted by Mekha Molly Raju (st90673) is a Bonafide record of work carried out under my guidance and supervision in partial fulfilment of the requirements for the award of the degree.

Guide Signature:

Date: 11/01/2026

DECLARATION

I hereby declare that this project titled **"Social Media Engagement Analysis Using Python"** is my original work and has not been submitted previously for any degree or diploma. All sources of information used have been duly acknowledged.

Place: Riga , Latvia

Date: 11/01/2026

Signature: Mekha Molly Raju

ACKNOWLEDGEMENT

I express my sincere gratitude to my project guide for their valuable guidance and support. I also thank my department faculty, friends, and family members for their encouragement throughout the completion of this project.

ABSTRACT

Social media platforms generate large volumes of engagement data such as platform, timestamp, day_of_week and mentions. Analysing this data helps organizations understand audience behaviour and optimize content strategy. This project focuses on analysing social media engagement data using python libraries such as Pandas, Numpy, Matplotlib, and seaborn. The analysis identifies high-performing content types, optimal posting times, and engagement trends. The results provide actionable insights that can help improve social media performance and audience interaction.

CONTENTS

1. INTRODUCTION

Social Media Engagement refers to the interactions users have with content published on social media platforms. These interactions include platform, timestamp, day_of_week and mentions. High engagement indicates strong audience interest and effective content strategy. With the increasing use of social media for marketing and communication, analysing engagement data has become essential for businesses and content creators.

2. PROBLEM STATEMENT

Many organizations struggle to identify which type of social media content performs best and the optimal time to post in order to maximize audience engagement. Without data-driven analysis, content strategies may not deliver desired results.

3. OBJECTIVES

- To analyse social media engagement data
- To identify high-performance content types
- To determine the best day and time for posting
- To calculate engagement metrics such as engagement rate
- To visualize engagement trends

4. SCOPE OF THE PROJECTS

The projects focuses on analysing social media engagement data collected from a sample dataset. The analysis includes engagement metrices such as platform, timestamp, day_of_week and mentions. The project does not involve real-time data or platform APIs.

5. LITERATURE REVIEW

Previous studies show that data analytics plays a crucial role in social media marketing by helping organizations understand user behaviour. Engagement-based analysis improve content reach and customer interaction. Several researchers have emphasized the importance of timing and content type in maximizing engagement.

6. TOOLS AND TECHNOLOGIES USED

- Python Programming language
- Pandas
- Numpy
- Matplotlib
- Seaborn
- Jupyter Notebook

7. DATASET DESCRIPTION

The dataset used in this project consists of social media post idea

Column Name	Description
Post_id	Unique identifier for each post
timestamp	Time of posting
day_of_week	Day of posting
Platform	Which social media platfrom is used
User_id	Unique identifier for each users
Location	Location of the post
Language	Language used for each content
Text_content	Description of the contents
Hashtags	Hashtags of each posts
Mentions	Mentions of other users,locations of each posts
Comments_count	Number of comments of ach posts
Impressions	Impressions of people in each posts
Engagement_rate	Rate of engagement
Brand_name	Brand name used in each posts
Product_name	Product name used in each posts
Compaign_name	Name of compaign
Compaign_phase	Phase of compaign
User_past_sentiment_avg	Avg of user pasr sentiment
User_engagement_growth	Growth of user engagement
Buzz_change_rate	Change rate of buzz

8. METHODOLOGY

a. Data collection from CSV dataset
b. Data cleaning and preprocessing
c. Feature engineering(total engagement, engagement rate)
d. Exploratory Data Analysis
e. Data Visualization

9. DATA ANALYSIS AND RESULTS

The data analysis revealed patterns in engagement based on platforms, day_of_week, and timestamp. Visualizations were used to compare engagement metrics across different categories.

```
[26] ✓ 0s
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[27] ✓ 0s
df=pd.read_csv('/content/sample_data/Social Media Engagement Dataset.csv')
df.head()
```

	post_id	timestamp	day_of_week	platform	user_id	location	language	text_content	hashtags	mentions	...	comments_count	impr
0	kcqbs6hxybia	2024-12-09 11:26:15	Monday	Instagram	user_52nwb0a6	Melbourne, Australia	pt	Just tried the Chromebook from Google. Best pu...	#Food	NaN	...	701	
1	vkmervg4ioos	2024-07-28 19:59:26	Sunday	Twitter	user_ucryct98	Tokyo, Japan	ru	Just saw an ad for Microsoft Surface Laptop du...	#MustHave, #Food	@CustomerService, @BrandCEO	...	359	
2	memhx4o1x6yu	2024-11-23 14:00:12	Saturday	Reddit	user_7rrev126	Beijing, China	ru	What's your opinion about Nike's Epic React? ...	#Promo, #Food, #Trending	NaN	...	643	
		2024-09-				Lagos		Bummed out	#Reviews	@StyleGuide			

```
[28] ✓ 0s
df.isnull().sum()
df.dropna(inplace=True)

[29] ✓ 0s
def count_items(text):
    if pd.isna(text) or not isinstance(text, str) or text.strip() == '':
        return 0
    return len(text.split(', '))

df['mentions_count'] = df['mentions'].apply(count_items)
df['hashtags_count'] = df['hashtags'].apply(count_items)
df['total_engagement'] = df['mentions_count'] + df['comments_count'] + df['hashtags_count']
df['engagement_rate'] = (df['total_engagement'] / df['impressions']) * 100

[30] ✓ 0s
df.groupby('platform')['total_engagement'].mean().sort_values(ascending=False)
```

platform	total_engagement
YouTube	516.758831
Twitter	514.840100
Reddit	511.001871

[31] ✓ 0s

```
df.groupby('day_of_week')['total_engagement'].mean()
```

day_of_week	total_engagement
Friday	504.992181
Monday	504.075809
Saturday	499.388601
Sunday	509.667832
Thursday	514.539894
Tuesday	525.191489
Wednesday	507.099490

dtype: float64

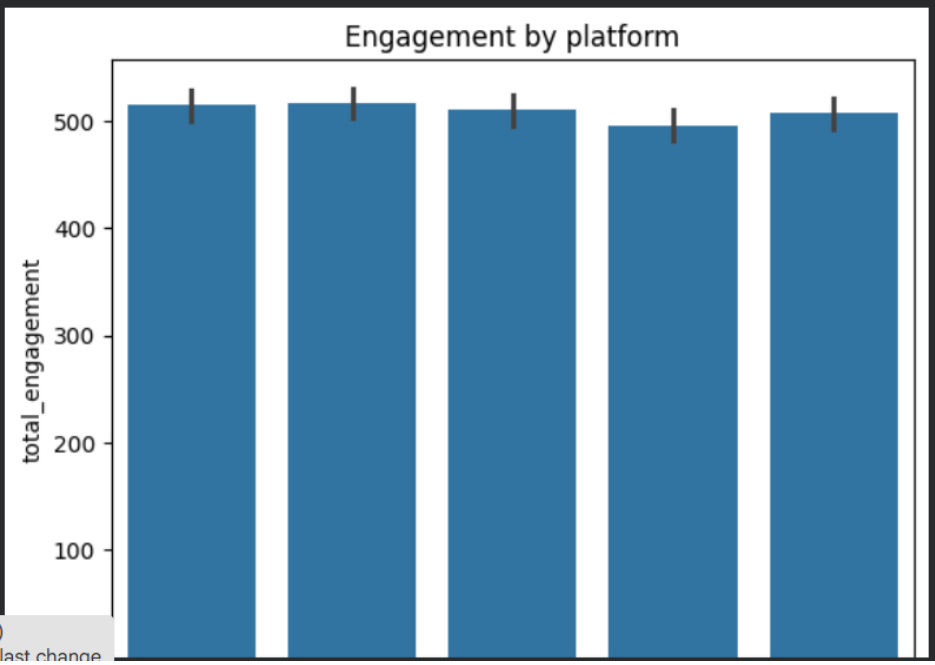
[32] ✓ 0s

```
df.groupby('timestamp')['total_engagement'].mean()
```

timestamp	total_engagement
2024-05-01 01:04:04	559.0
2024-05-01 01:35:53	816.0
2024-05-01 02:24:32	287.0
2024-05-01 04:57:30	646.0
2024-05-01 06:08:47	397.0
...	...
2025-04-30 15:54:26	673.0
2025-04-30 16:47:42	570.0
2025-04-30 17:13:20	140.0
2025-04-30 21:21:15	891.0
2025-04-30 21:57:10	598.0

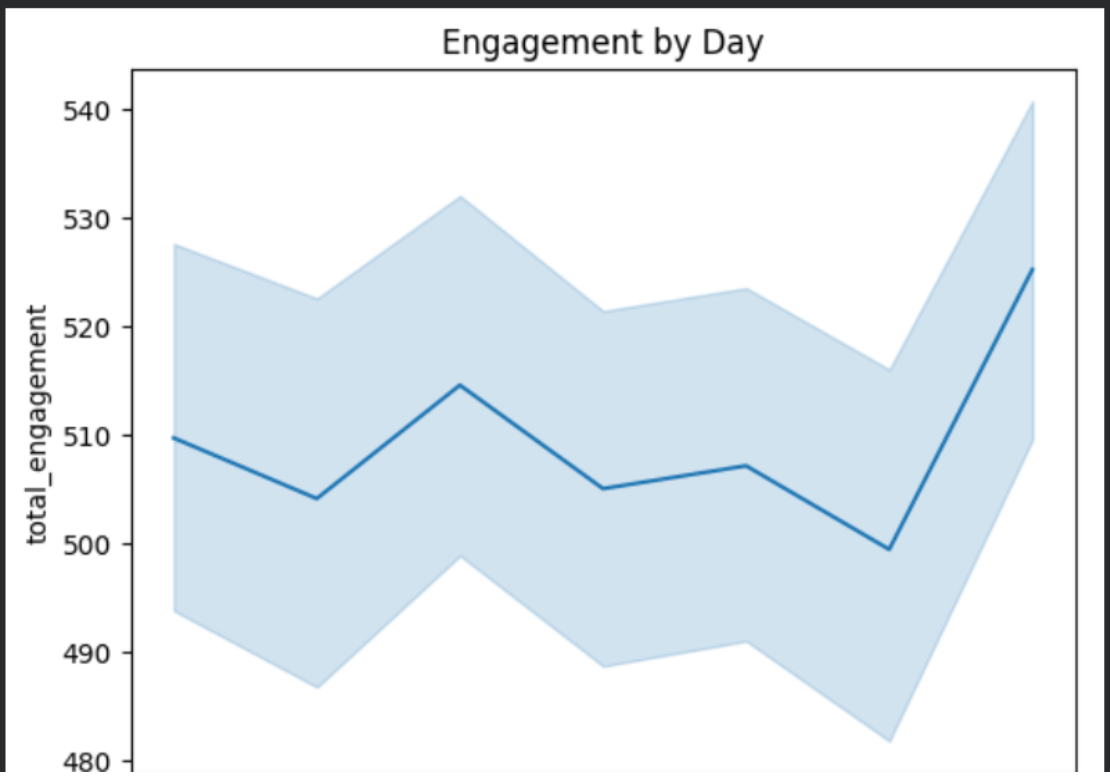
8058 rows × 1 columns

```
sns.barplot(x='platform', y='total_engagement', data=df)
plt.title("Engagement by platform")
plt.show()
```



l+Enter)
d since last change

```
sns.lineplot(x='day_of_week', y='total_engagement', data=df)
plt.title("Engagement by Day")
plt.show()
```



10. KEY FINDINGS

- Video and reel posts generate higher engagement than text posts
- Evening posting hours show higher interaction
- Fridays and weekends have better engagement rates

11. CONCLUSION

The projects successfully analysed social media engagement data and achieved its objectives. The findings can help improve content planning and posting strategies. Data analytics proves to be a powerful tool in understanding audience behaviour.

12. FUTURE ENHANCEMENT

- Use real-time data through social media APIs
- Apply machine learning models for engagement prediction
- Perform sentiment analysis on comments

13. LIMITATIONS

- Dataset is limited and simulated
- No real-time data analysis
- Platform-specific factors not included

14. REFERENCES

- Kaggle Datasets
- Python Documentation
- Research articles on social media analytics

15. APPENDIX

Python code :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('/content/sample_data/Social Media Engagement Dataset.csv')
df.head()
df.isnull().sum()
df.dropna(inplace=True)
def count_items(text):
    if pd.isna(text) or not isinstance(text, str) or text.strip() == '':
        return 0
    return len(text.split(', '))


df['mentions_count'] = df['mentions'].apply(count_items)
df['hashtags_count'] = df['hashtags'].apply(count_items)
df['total_engagement'] = df['mentions_count'] + df['comments_count'] +
df['hashtags_count']
df['engagement_rate'] = (df['total_engagement'] / df['impressions']) * 100
df.groupby('platform')['total_engagement'].mean().sort_values(ascending=False
)
df.groupby('day_of_week')['total_engagement'].mean()


df.groupby('timestamp')['total_engagement'].mean()
sns.barplot(x='platform', y='total_engagement', data=df)
plt.title("Engagement by platform")
plt.show()
```

```
sns.lineplot(x='day_of_week', y='total_engagement', data=df)
plt.title("Engagement by Day")
plt.show()
```