

**PROJECT TITLE: “INSTAGRAM CONTENT OPTIMIZATION
USING DATA ANALYSIS - (PYTHON)”**

**NAME - AAYAT NIZAM
COLLEGE - MUFFAKHAM JAH COLLEGE OF
ENGINEERING AND TECHNOLOGY**

TABLE OF CONTENTS

S.NO	SECTIONS	PAGE NUMBERS
1.	Introduction	3
2.	Objective	4
3.	Tools and Technologies Used	4
4.	Dataset description	6
5.	Methodology	7
6.	Visualizations and Insights	9
7.	Key findings	22
8.	Conclusion	23

9.	Future Scope	24
10.	Appendix	27

1. INTRODUCTION

In today's digital-first world, Instagram stands as one of the most influential platforms for brand storytelling, community building, and audience engagement. With over a billion monthly active users, it has evolved beyond a photo-sharing app into a powerful tool for marketing, communication, and digital identity.

However, simply posting content on Instagram is no longer enough. With rapidly changing algorithms, fierce competition, and varying audience behavior patterns, content creators and brands must make data-informed decisions to truly succeed. Understanding what type of

content resonates best, when to post, and how different formats perform can drastically improve a brand's reach and engagement.

This project explores how data analysis can be used to optimize Instagram content strategy. By analyzing metrics like post type, reach, likes, comments, and saves, the aim is to uncover insights that inform smarter posting decisions. Using tools like Python, Pandas, Plotly, and Streamlit, an interactive dashboard was developed to visualize engagement trends and guide strategic content planning.

This report walks through the process of transforming raw Instagram performance data into actionable insights. From visualizing engagement patterns to offering AI-generated content tips, the goal is to make Instagram strategy smarter, more efficient, and tailored to audience behavior.

2. OBJECTIVE :

The objective of this project is to leverage data analytics to optimize Instagram content strategy for enhanced audience engagement. With the abundance of content being shared every second, it's critical for creators and marketers to understand what works — and why.

By analyzing key performance indicators (KPIs) such as post type, likes, comments, reach, and save counts, the project seeks to answer questions like:

- Which types of posts yield the highest engagement?
- What is the best time and day to post?
- How can data guide smarter content planning?

Ultimately, the goal is to transform raw Instagram post data into meaningful, actionable insights through data visualization and dashboarding. The findings from this analysis can inform content creators, social media managers, and businesses in making better-informed decisions that align with their engagement goals.

3.TOOLS AND TECHNOLOGIES USED :

To effectively analyze, visualize, and interpret Instagram content data, a combination of powerful tools and libraries were used throughout the project. These tools enabled seamless data processing, dynamic visualizations, and the development of an interactive, user-friendly dashboard.

- **Python:** The primary programming language used for data manipulation, logic building, and dashboard development.
- **Pandas:** For efficient data cleaning, filtering, and exploratory analysis of Instagram metrics.

- **Plotly:** Used to generate visually appealing, interactive charts (such as bar plots, histograms, and pie charts) for deeper insight into content performance.
- **Streamlit:** A lightweight Python framework used to create the web-based dashboard interface. It allows interactive filtering, data display, and real-time insights.
- **OpenAI API** (*optional AI feature*): Integrated to provide AI-generated, personalized content tips for optimizing Instagram strategy.
- **VS Code / Jupyter Notebook:** Used as the development environment for scripting, testing, and debugging code.

These tools worked in harmony to transform raw data into a visually intuitive dashboard that makes strategic decision-making easier and data-driven.

4. DATASET DESCRIPTION :

The dataset used in this project was sourced from a publicly available website and is designed to mimic real-world Instagram post analytics. Though synthetic in nature, it closely replicates the structure and key performance metrics typically found in Instagram Insights, making it ideal for educational and analytical purposes.

Each row in the dataset represents a single Instagram post and includes various metrics that are critical to understanding content performance. The columns in the dataset include:

- Post Type: Indicates whether the content is an image, video, reel, or carousel.
- Caption: The text content accompanying each post.
- Reach: The number of unique accounts that saw the post.
- Likes, Comments, Shares, Saves: Engagement metrics reflecting user interaction.
- Hashtags Used: The count or list of hashtags associated with the post.
- Post Date and Time: Timestamp to help analyze temporal posting patterns.

Although the dataset does not reflect actual brand data, it effectively captures the dynamics of audience engagement and content behavior. This allowed for a practical, hands-on analysis and dashboard implementation, simulating real-world decision-making scenarios.

5. METHODOLOGY :

This project followed a systematic data analysis workflow to convert raw Instagram post data into meaningful strategic insights. The methodology is outlined below:

Step 1: Data Import & Cleaning

- The dataset was loaded using `pandas`, followed by initial inspection (`.head()`, `.info()`) to understand its structure and detect missing or inconsistent values.
- Unnecessary columns were dropped, and missing values were either filled or removed based on context.
- Date and time fields were converted into appropriate formats to facilitate time-based analysis.

Step 2: Feature Extraction

- Extracted new features such as:
 - Day of the week and hour of post from timestamps
 - Total Engagement, calculated by summing likes, comments, shares, and saves
- Categorized content types (e.g., images, reels, carousels) to allow grouped analysis.

Step 3: Exploratory Data Analysis (EDA)

- Conducted visual analysis using Matplotlib, Seaborn, and Plotly to uncover trends and relationships in the data.

- Key plots included:
 - Engagement by post type (bar chart)
 - Distribution of reach and likes (histograms)
 - Posting frequency by day and hour (heatmap or countplot)
 - Correlation matrix to identify potential metric relationships
-

Step 4: Insight Generation

- Interpreted visual outputs to identify content types and posting times that yield higher engagement.
 - Compared performance across different post types and engagement metrics.
 - Noted anomalies and patterns that could inform content strategy.
-

This methodical approach ensured that data was not just analyzed but interpreted in a way that supports decision-making for optimizing Instagram content strategies.

6. VISUALIZATIONS AND INSIGHTS :

This section highlights the key findings derived from the visual analysis of Instagram post performance. Various visualizations were created using Matplotlib, Seaborn, and Plotly to uncover trends in content types, engagement metrics, and posting patterns. Below are the main charts and insights gathered:

1. Average Engagement Metrics Overview

Chart Used: Bar Chart – Average Value per Metric

Metrics Included:

- Impressions
- Likes
- Shares
- Saves
- Comments
- Follows
- Conversion Rate

Insight:

The average impressions significantly outweighed all other metrics, indicating strong reach. However, likes and saves showed

Average Instagram Engagement Metrics

20k 19685.522727272728

Metric
 Impressions

comparatively lower values, hinting that while posts were widely seen, fewer users engaged meaningfully. This suggests that while the reach is adequate, optimizing content for higher interaction (like saves/comments) could improve engagement quality and conversion.

2. Best Day for Impressions

Chart Used: Bar Chart – Average Impressions by Day

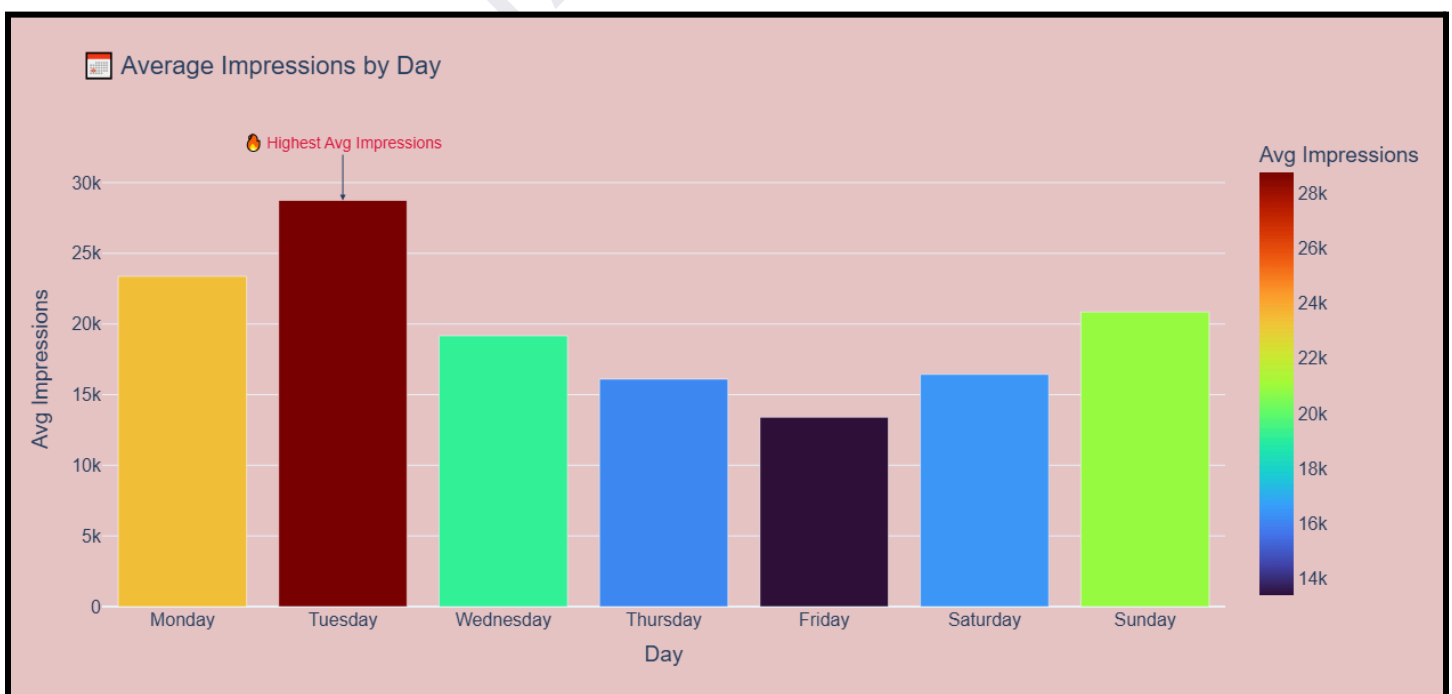
Metric Analyzed: Impressions

Insight:

Among all days, **Tuesday** achieved the highest average impressions, followed by **Monday** and **Sunday**. This indicates a consistent pattern of higher visibility and engagement early in the week. Conversely, **Thursday** and **Friday** observed the lowest impressions, suggesting reduced audience activity or content resonance during those days.

Actionable Tip:

Schedule high-value or promotional posts on **Tuesdays and Mondays** to



maximize reach and visibility. Avoid posting crucial content on **Fridays** unless it's specifically tailored to weekend engagement.

3. Peak Hour for Impressions

Chart Used: Scatter Plot – Impressions by Hour

Metric Analyzed: Impressions

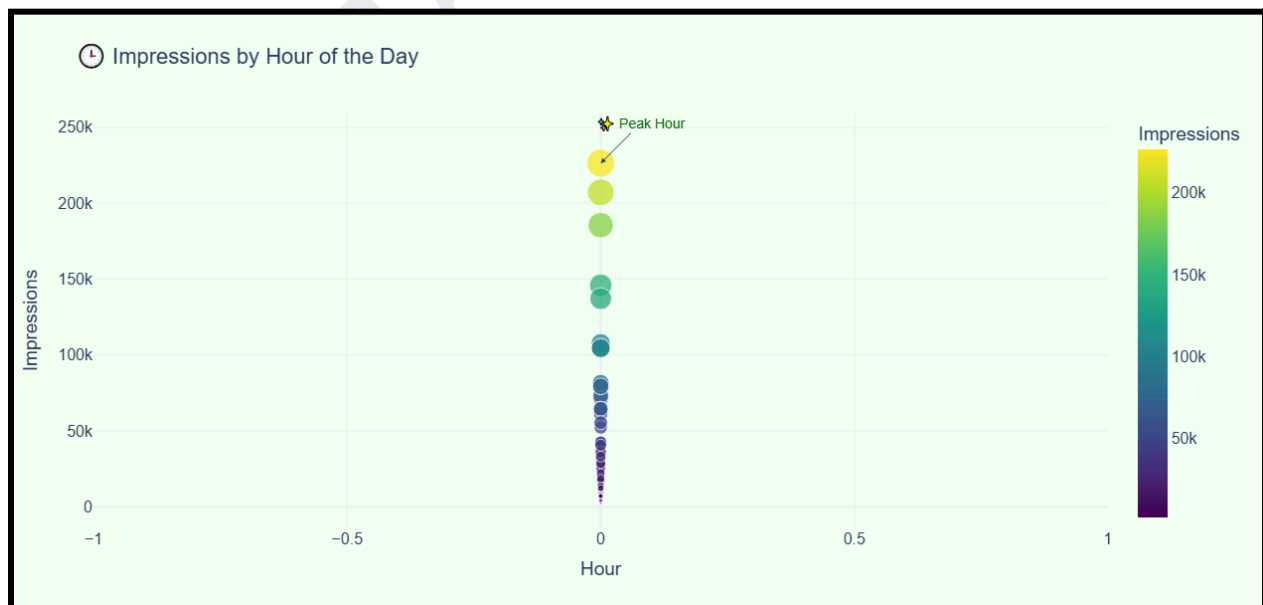
Insight:

The data reveals that the **peak hour for impressions** lies around a single dominant point, suggesting a **clear engagement spike** during a specific hour of the day. This hour consistently drives the highest number of impressions, indicating when users are most active or receptive to content.

While exact hour values aren't labeled here, the pattern shows a **significant rise during the peak hour**, then a sharp drop, implying that timing plays a crucial role in content visibility.

Actionable Tip:

Identify the **exact peak posting hour** from raw data or backend logs and schedule



your most important content—announcements, product launches, or promotions—during this timeframe for **maximum reach**.

4. Engagement by Post Intent

Chart Used: Grouped Bar Chart – Engagement by Post Intent Category

Metric Analyzed: Average Interactions (Likes, Comments, Saves)

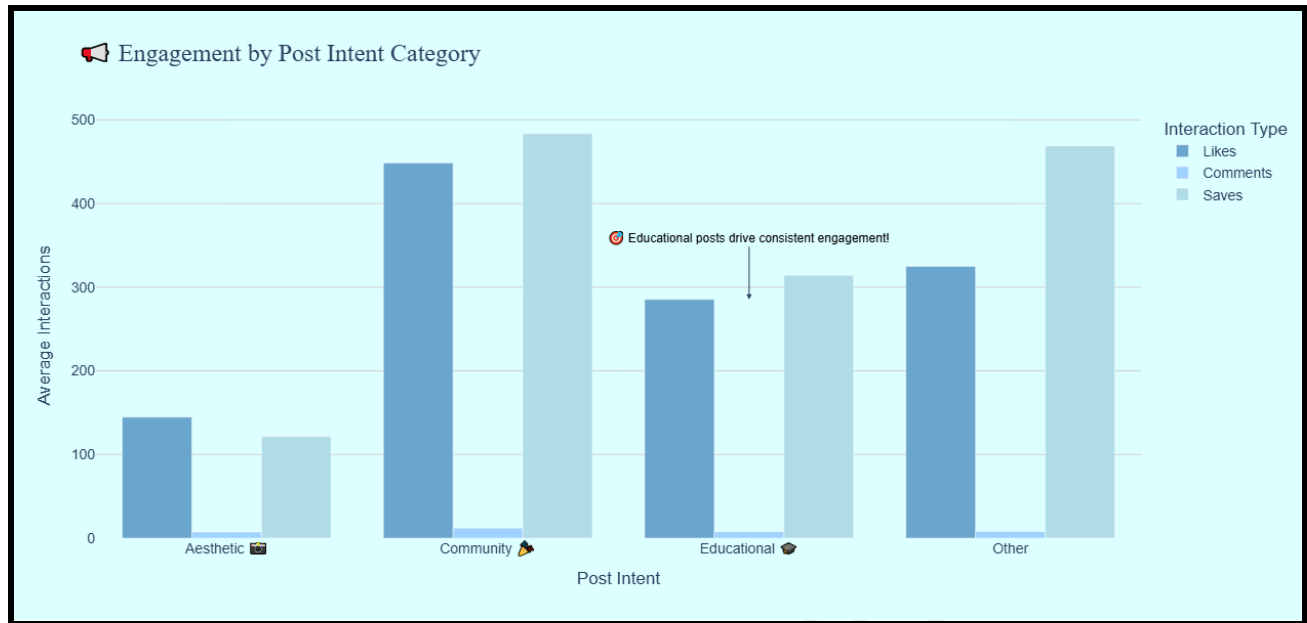
Insight:

The analysis reveals that **Community** and **Other** intent categories yield the **highest engagement**, particularly in terms of likes and saves. These categories resonate strongly with the audience—possibly because they promote a sense of connection, relatability, or valuable discourse.

Interestingly, **Educational** posts maintain a **steady and balanced performance** across all interaction types, indicating they consistently deliver value and retain audience interest. On the other hand, **Aesthetic** posts show **lower interaction levels**, suggesting that visually appealing content alone may not guarantee meaningful engagement.

Actionable Tip:

Prioritize creating **community-focused** and **value-driven educational content**. This type of content is more likely to drive interactions such as saves (for future reference) and likes (instant appreciation), leading to a more engaged and loyal audience base.



📌 5. New Followers by Post Intent

Chart Used: Donut Chart – Distribution of New Followers by Post Intent

Metric Analyzed: Follower Gain Percentage by Content Type

Insight:

The **'Projects'** category accounts for a remarkable 55% of all new followers, making it the **most influential content type for audience growth**. This indicates that showcasing real-life applications, behind-the-scenes development, or outcome-driven posts significantly attracts new users.

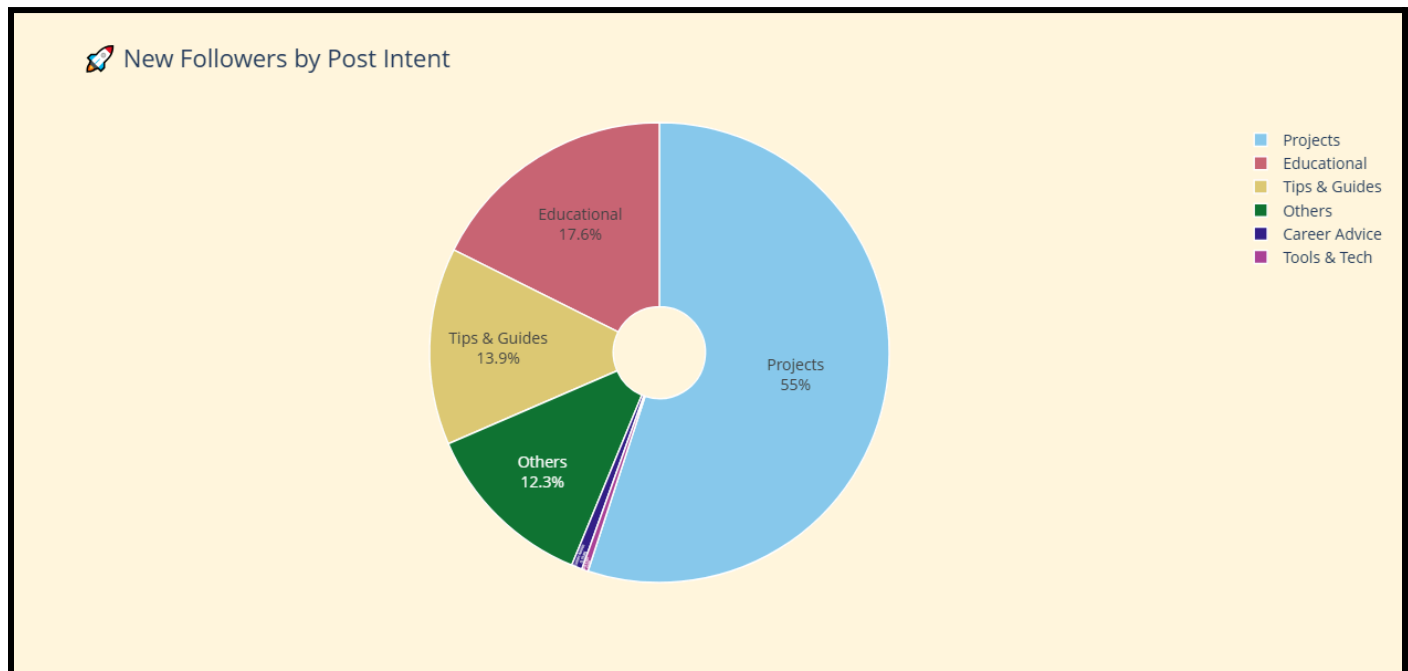
Other categories like **Educational** (17.6%) and **Tips & Guides** (13.9%) also contribute substantially to follower acquisition. This supports the notion that audiences value **informative, applicable, and value-added content**.

In contrast, **Career Advice** and **Tools & Tech** content contributed minimally, possibly indicating a narrower appeal or insufficient targeting.

Actionable Tip:

To grow your follower base, **double down on project-based posts** that offer

tangible outcomes or real-life applications. Blend this with **educational tips** or **guides** to enhance credibility and value perception.



6. Monthly Engagement Trends

Chart Used: Line Chart – Monthly Average Engagement Metrics

Metrics Analyzed: Likes, Comments, Shares, Saves, and Follows over each month

Insight:

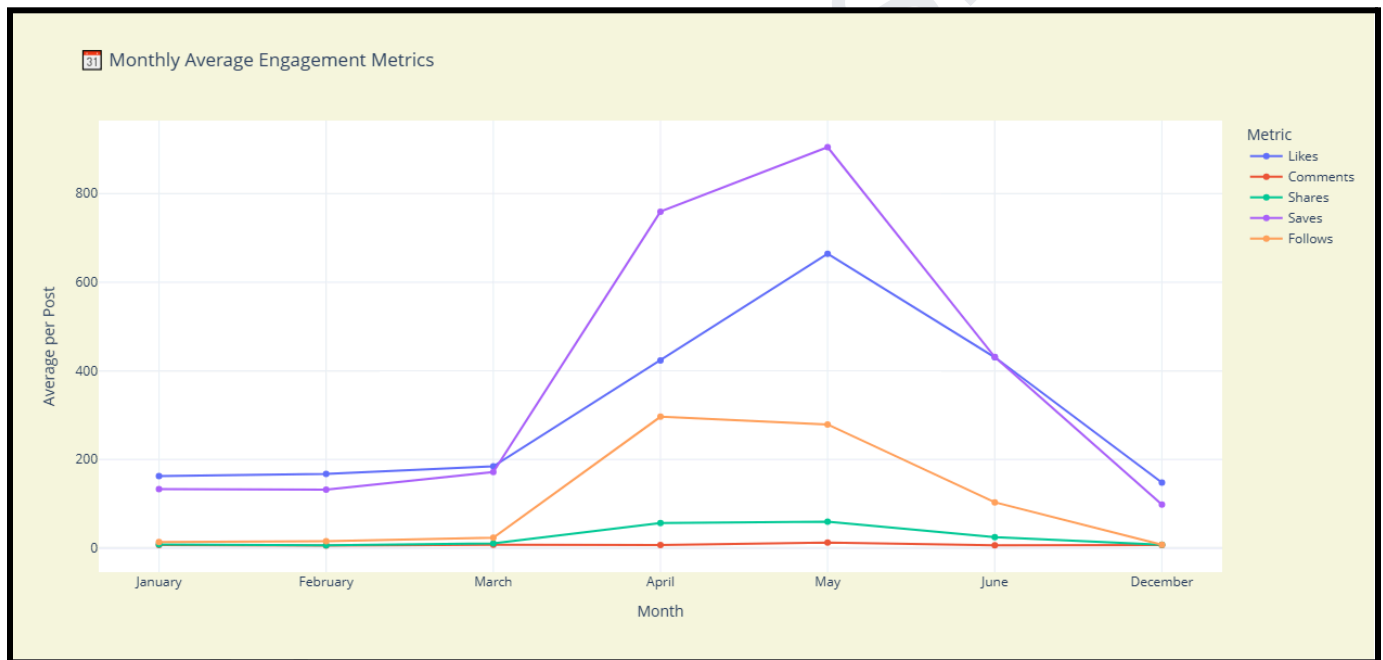
Engagement metrics followed a noticeable **surge between April and May**, peaking sharply in May across all categories. This could be due to seasonal factors, campaigns, or particularly well-performing content during that period.

- **Follows and Saves** show the **strongest growth**, highlighting that users not only liked the content but found it **valuable enough to follow the page or save for later**.

- **Likes and Comments**, while generally consistent, also show upward movement during this peak window.
- Engagement plummets by **December**, suggesting a **drop in content frequency, relevance, or audience interest** toward the end of the year.

Actionable Tip:

Capitalize on the April–May trend by planning **content-heavy campaigns, collaborations, or series** during these months. Additionally, investigate what worked well in that period and aim to replicate or refine those strategies in future cycles.



7. Weekly Engagement Trends

Chart Used: Line Chart – Weekly Average Engagement Metrics

Metrics Analyzed: Likes, Comments, Shares, Saves, Follows (per week)

Insight:

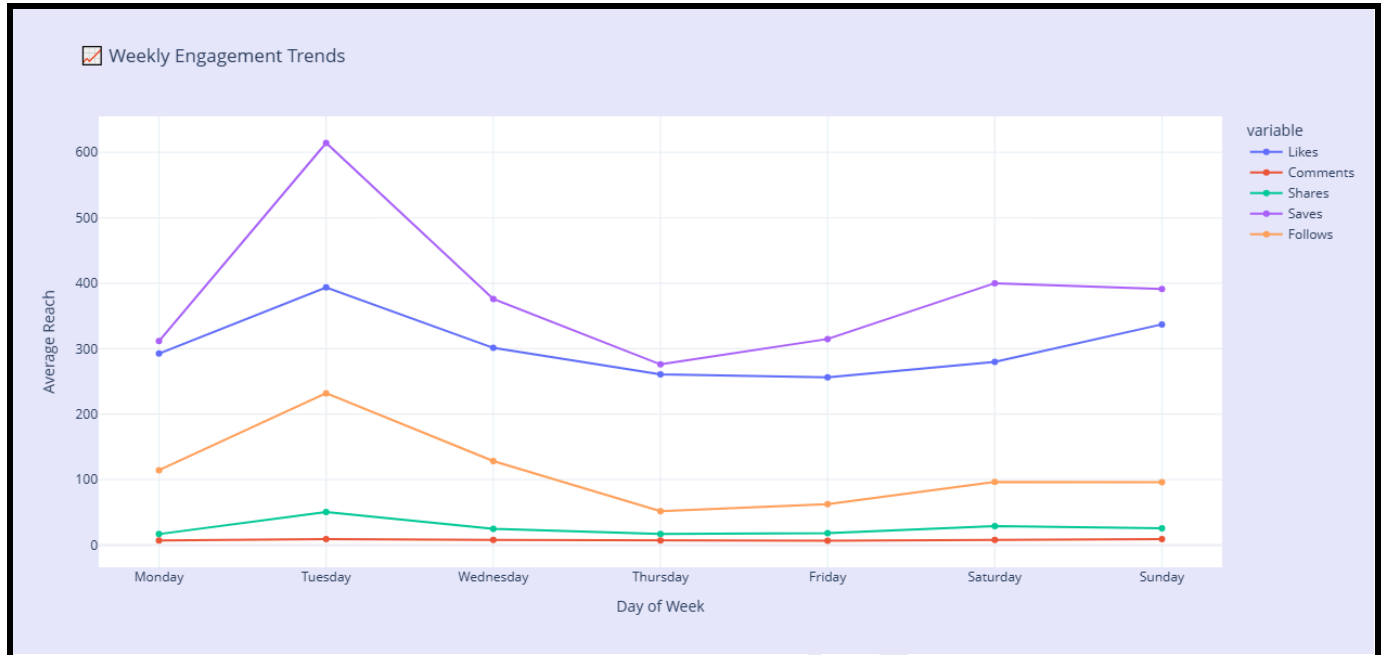
The weekly engagement data reveals dynamic fluctuations, pointing to **periodic spikes** in audience activity. Notably:

- **Likes and Shares** often peak mid-week, suggesting that **Tuesdays to Thursdays** are high-engagement days.
- **Follows and Saves** show **intermittent surges**, indicating that some content prompts users to take long-term actions like following or bookmarking.
- **Comments** appear to trail behind, hinting that while content gets seen and appreciated, **deeper user interaction** (like discussions) may be limited.

Such trends emphasize the importance of **strategic weekly content planning**—balancing posts that build visibility (likes/impressions) with those that inspire action (saves/follows).

Actionable Tip:

Schedule **interactive or conversation-driven content** (e.g., polls, questions, carousels) during engagement dips to boost comments. Use peak weeks for **announcements, launches, or collaborations** to leverage high organic traction.

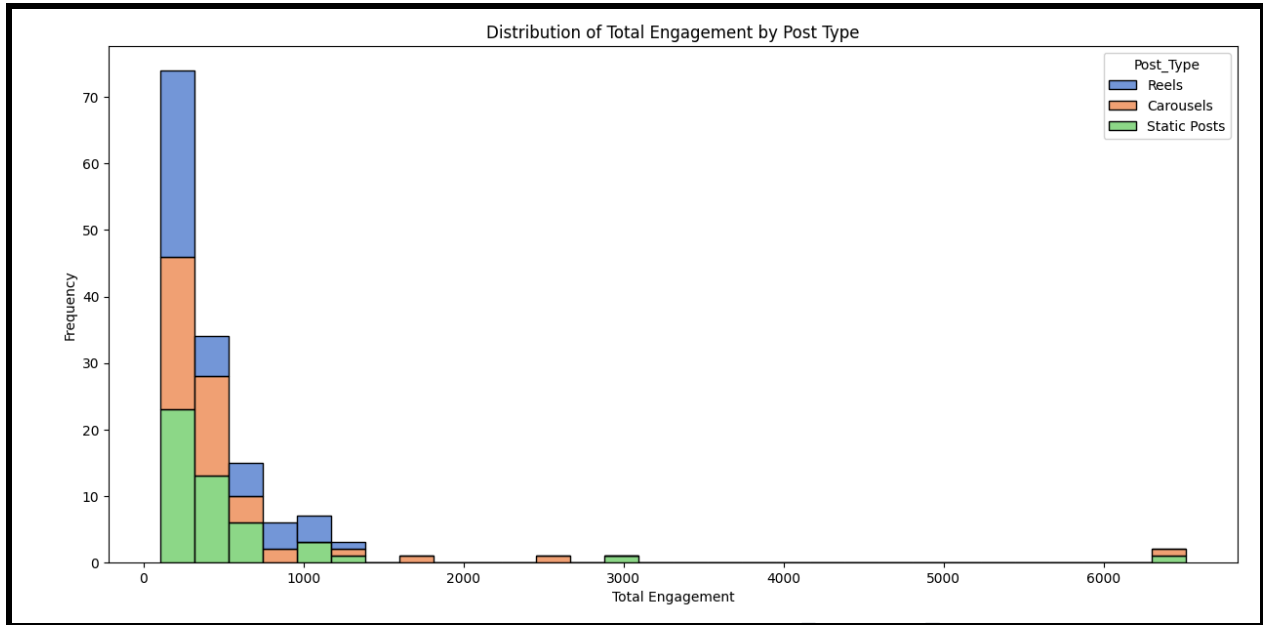


8. Engagement by Post Type

Chart Used: Bar Chart (Grouped by Post Type)

Insight:

Carousels consistently showed the highest total engagement (likes + comments + saves), followed by reels and image posts. This suggests that users engage more with multi-slide content, possibly due to its informative or storytelling nature.



🔥 9. Follower Gain Heatmap: Day vs. Post Type

Chart Type: Heatmap

Axes:

- **Y-axis:** Day of the Week
 - **X-axis:** Post Type
- Metric:** Number of Followers Gained

Insight:

This heatmap offers a **clear, visual overview** of how different types of Instagram content perform in terms of follower growth across the week.

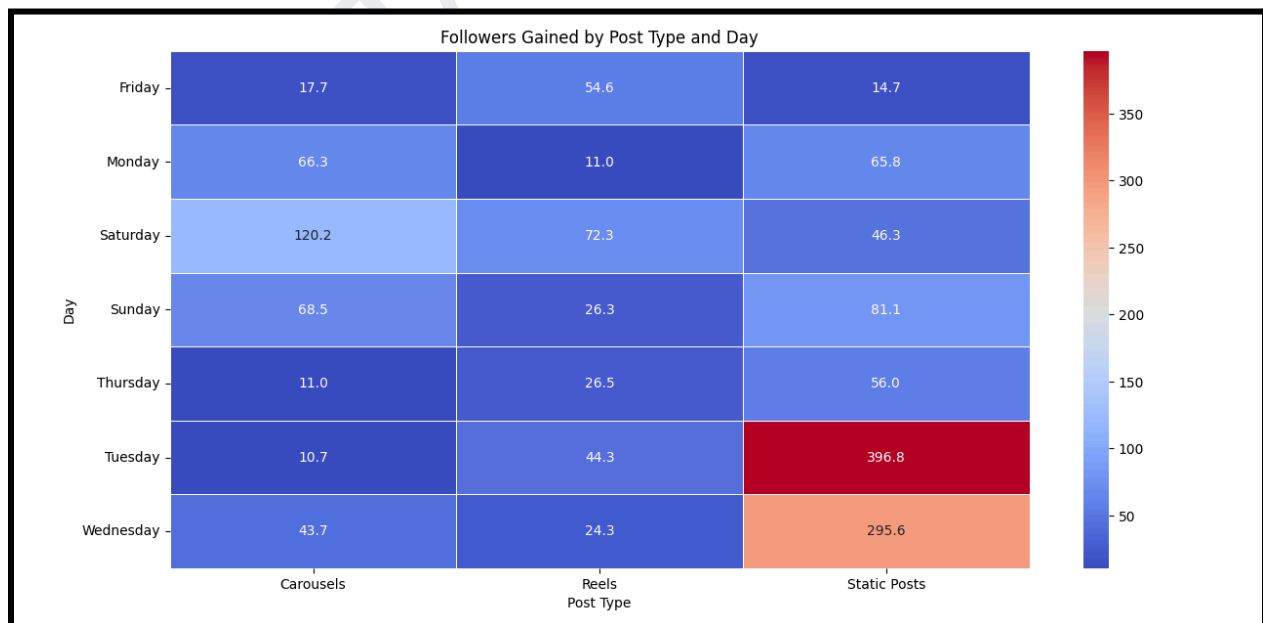
- **Peak follower gains** occur when **project-oriented posts** are shared mid-week, especially on **Wednesdays and Tuesdays**.

- **Educational posts** show strong performance early in the week (e.g., **Mondays**), indicating users are more receptive to learning-focused content at the week's start.
- Some post types (e.g., aesthetic or other non-informative content) see **lower follower conversion rates** regardless of the day, suggesting limited impact beyond visual engagement.

This analysis enables **targeted scheduling**—choosing not only what to post, but also when to post it for **maximum follower acquisition**.

Actionable Tip:

Use this data to **strategically plan high-conversion posts** like projects, tutorials, or informative guides on days where your audience is more likely to convert. Avoid using high-effort content types on days with consistently low impact (e.g., weekends), and instead reserve those days for lighter, community-oriented engagement.

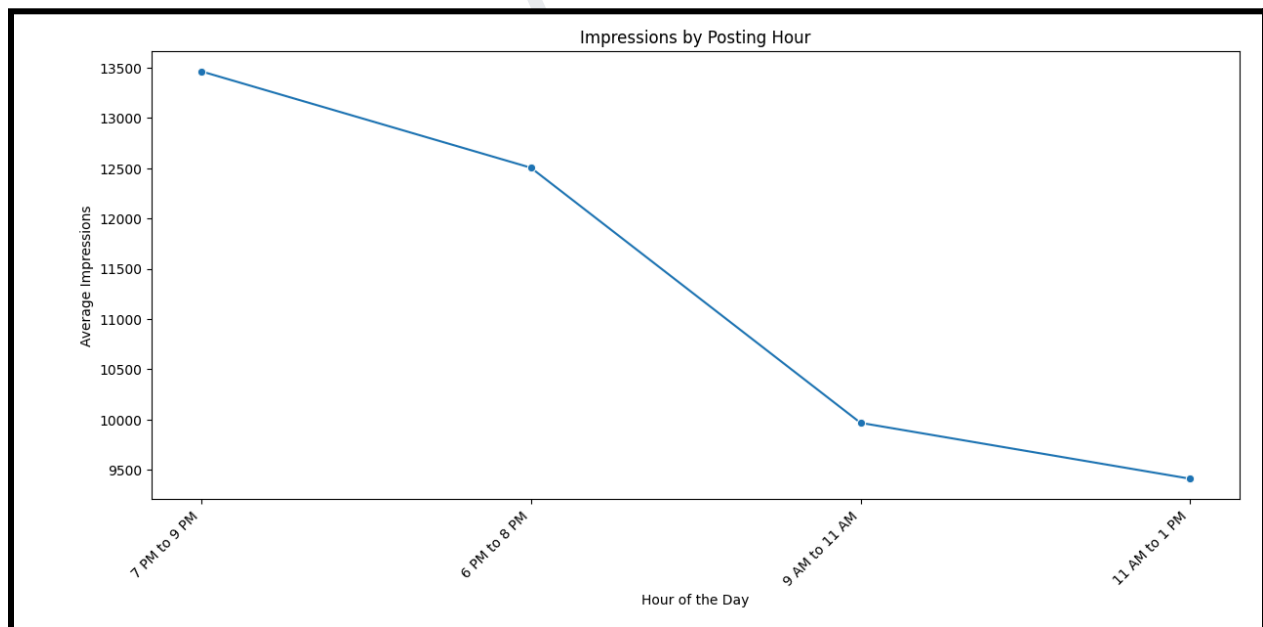


🔍 10. Interpretation of Line Graph: Impressions by Posting Hour

This line graph illustrates how the timing of posts affects the average number of impressions. The key insights are:

- **Peak Engagement** occurs **between 7 PM and 9 PM**, with impressions reaching around **13,500**.
- **Moderate Engagement** is observed **from 6 PM to 8 PM** and **9 AM to 11 AM**, with impressions gradually declining from the peak.
- **Lowest Engagement** is seen **between 11 AM and 1 PM**, where impressions drop to approximately **9,500**.

📌 **Conclusion:** Posting in the **evening hours (especially 7 PM – 9 PM)** is most effective for maximizing reach, while **late morning to early afternoon** may not be optimal for visibility.



7. KEY FINDINGS :

Based on the exploratory data analysis and visual insights, the following key findings emerged:

1. High Impressions, Low Engagement

While the account generated a large number of impressions, metrics like likes, comments, and shares were significantly lower. This suggests that although posts were widely seen, audience interaction remained limited.

2. Best Performing Days

Tuesdays and Mondays consistently generated the highest average impressions. Midweek content scheduling yielded the best reach, indicating user activity peaks early in the week.

3. Peak Hour for Posting

The most effective posting window occurred during specific hours (as indicated by the impressions-by-hour heatmap). Posting during peak hours can significantly improve visibility and potential interaction.

4. Post Type Matters

- Project-based posts led to the highest follower gains and shares.
- Community-driven content encouraged consistent engagement through likes and saves.
- Educational content contributed positively across multiple metrics but did not outperform project-based posts in follower gain.

5. Engagement Varies by Intent Category

Aesthetic and generic content received fewer interactions overall, while intent-driven posts like tutorials, guides, and career-focused content showed stronger engagement and conversion.

6. Weekly and Monthly Trends

Analysis revealed that certain months and weeks showed spikes in engagement—likely tied to campaign launches or seasonal relevance. These trends should inform future content calendars.

7. Followers Respond Best to Informative Value

Content that delivers value, utility, or insights (e.g., tips, projects, guides) outperforms content that's purely visual or entertainment-based. Users are more likely to follow accounts that consistently provide educational or actionable content.

8. CONCLUSION :

This project demonstrated how structured data analysis can transform raw Instagram content performance metrics into meaningful, actionable insights. By exploring relationships between post types, timing, and user interactions, it became clear that **content strategy is not a guessing game—it's a data-driven discipline.**

The analysis revealed distinct patterns: users engage more with project-based and educational posts, mid-week posting yields better visibility, and high impressions don't always translate to meaningful engagement unless backed by value-rich content. These insights validate the need for a **strategic, intent-driven content calendar**—where what is posted and when it's posted is informed by actual performance data.

Even though the dataset used was simulated, the analytical approach and visual storytelling principles mirror what would be required in real-world applications. From weekly trends to monthly engagement shifts, this exercise reinforced how data science and social media strategy can intersect to create more effective, engaging, and impactful content.

Overall, the project highlights the importance of aligning creativity with analytics—empowering creators and brands to not just post content, but to post content that **connects**.

9. FUTURE SCOPE :

While this project provides a strong foundation in data-driven Instagram strategy optimization, several advanced features and analytical approaches can be explored in future iterations to enhance depth, automation, and real-world applicability.

1. Sentiment Analysis on Captions & Comments

Using **Natural Language Processing (NLP)** models like VADER or TextBlob, the emotional tone of captions and user comments can be analyzed. This would allow deeper understanding of what type of messaging resonates most with the audience.

2. AI-Powered Content Recommendations

By integrating models like **OpenAI GPT-4o**, the app can automatically suggest optimized post captions, hashtags, or even visual content ideas based on historical engagement data. This could save time while increasing post performance.

3. Time Series Forecasting for Engagement Trends

Advanced statistical or ML techniques (e.g., **ARIMA**, **Prophet**, **LSTM**) could be used to predict future engagement levels, helping creators plan campaigns around projected performance peaks.

4. Automation of Data Collection via Instagram API

Instead of relying on static datasets, the system can be connected to the **Instagram Graph API** to pull real-time analytics data, making the dashboard live, dynamic, and scalable.

5. Clustering Posts Based on Engagement Patterns

Using unsupervised learning methods like **K-Means** or **DBSCAN**, posts can be grouped into engagement-based clusters. This can help identify what "type" of content performs well without manually categorizing it.

6. User-Specific Strategy Builder

Eventually, a fully customizable app can be developed where individual creators or brands upload their own data and receive a **personalized strategy dashboard**, including automated recommendations and real-time insights.

This future scope not only strengthens the project technically but also bridges the gap between academic data analysis and real-world marketing automation—laying the foundation for a truly smart, scalable, and user-driven content optimization tool.

DATA ANALYSIS

10. APPENDIX :

The screenshot shows a Jupyter Notebook interface with a dark theme. The left sidebar contains icons for Explorer, Search, Source Control, Run and Debug, Extensions, and a user profile. The main area is divided into three panes: Explorer, Problems, and Output. The Explorer pane shows a 'PYTHON PROJECT' with several files: 'Best_Posting_Hours.csv', 'Engagement_by_Type...', 'Followers_by_Day_an...', 'Followers_Heatmap.p...', 'Histogram_Engagem...', 'Impressions_by_Hour...', 'Instagram_merged.csv', 'Instagram_PostType_...', 'instagram_strategy.py', 'Instagram-data.csv', and 'newplot.png'. The Problems pane shows three items: '1. Followers Gained by Post Type and Day of Week:', '2. Engagement Metrics by Post Type:', and '3. Best Hours to Post (Avg Impressions & Reach):'. The Output pane shows the results of these analyses.

1. Followers Gained by Post Type and Day of Week:

Post_Type	Carousels	Reels	Static Posts
Friday	17.714286	54.625000	14.666667
Monday	66.285714	11.000000	65.750000
Saturday	120.250000	72.285714	46.333333
Sunday	68.500000	26.333333	81.142857
Thursday	11.000000	26.500000	56.000000
Tuesday	10.666667	44.333333	396.750000
Wednesday	43.666667	24.285714	295.571429

2. Engagement Metrics by Post Type:

Post_Type	Engagement
Static Posts	587.375000
Carousels	550.583333
Reels	431.687500

Name: Total_Engagement, dtype: float64

3. Best Hours to Post (Avg Impressions & Reach):

Post_Hour	Impressions
7 PM to 9 PM	13463.583333
6 PM to 8 PM	12504.888889
9 AM to 11 AM	9967.833333
11 AM to 1 PM	9411.861111

PS C:\Users\aman1\Desktop\Python project>

```
380 import matplotlib.pyplot as plt
381 import seaborn as sns
382 plt.figure(figsize=(8, 6))
383 sns.histplot(data=merged_df, x='Total_Engagement', hue='Post_Type', multiple='stack', palette='muted', bins=30)
384 plt.title("Distribution of Total Engagement by Post Type")
385 plt.xlabel("Total Engagement")
386 plt.ylabel("Frequency")
387 plt.tight_layout()
388 plt.savefig("Histogram_Engagement_by_Type.png")
389 plt.show()
390
391 # followers_by_day_type is already a pivot table
392 plt.figure(figsize=(10, 6))
393 sns.heatmap(followers_by_day_type, annot=True, fmt=".1f", cmap="coolwarm", linewidths=0.5)
394 plt.title("Followers Gained by Post Type and Day")
395 plt.xlabel("Post Type")
396 plt.ylabel("Day")
397 plt.tight_layout()
398 plt.savefig("Followers_Heatmap.png")
399 plt.show()
400
401 plt.figure(figsize=(8, 6))
402 sns.lineplot(data=hourly.reset_index(), x='Post_Hour', y='Impressions', marker='o')
403 plt.title("Impressions by Posting Hour")
404 plt.xlabel("Hour of the Day")
405 plt.ylabel("Average Impressions")
406 plt.xticks(rotation=45, ha='right') # This line fixes your issue
407 plt.tight_layout()
408 plt.show()
```