

# 1. Overview Of The Project

## 1.1 Introduction

In the contemporary digital era, social media has become a significant part of daily life, influencing how people communicate, share information, and make decisions. Platforms like Music, Twitter, Instagram, YouTube, and others host a vast amount of user-generated content, providing an unprecedented opportunity for data analysis. Social media analysis involves the extraction, processing, and interpretation of data from these platforms to understand user behaviour, detect trends, and generate actionable insights.

The project titled "Social Media Data Analysis and Visualization" is centred around leveraging this wealth of data to identify patterns, trends, and user sentiments. By utilizing advanced data processing tools and visualization techniques, this project aims to transform raw data into meaningful information that can support decision-making processes in various domains such as marketing, customer service, and product development.

### Objectives and Goals

The primary objectives of this social media analysis project are:

**Data Collection:** Gather extensive data from various social media platforms using advanced web scraping techniques. This includes collecting posts, comments, user profiles, likes, shares, and other pertinent metrics.

**Data Cleaning and Preprocessing:** Ensure the quality and usability of the collected data by cleaning and preprocessing it to remove noise, handle missing values, and standardize formats.

**Sentiment Analysis:** Conduct sentiment analysis to determine the emotional tone of user posts and comments, providing insights into public opinion and mood.

**Trend Analysis:** Identify and examine trends over time, including trending topics, hashtags, and content types, to understand what captures public interest.

**User Behavior Analysis:** Analyze user behavior patterns such as content engagement, posting frequency, and interaction types to gain insights into how users interact with social media platforms.

**Visualization:** Use Power BI to create interactive and insightful visualizations that effectively represent the analyzed data, making it easier to communicate findings.

**Reporting:** Compile the findings into a comprehensive report that highlights key insights and offers recommendations based on the analyzed data.

## Structure of the Project

The project is structured into several key phases, each focusing on a specific aspect of the social media analysis process:

**Data Collection Phase:** This phase involves identifying target social media platforms and employing web scraping techniques to collect relevant data. Tools like BeautifulSoup, Scrapy, and APIs provided by social media platforms are utilized to efficiently extract large volumes of data.

**Data Cleaning and Preprocessing Phase:** During this phase, the raw data is cleaned and preprocessed to ensure its quality and consistency. This includes removing duplicates, handling missing values, and converting data into a suitable format for analysis.

**Data Analysis Phase:** Various analyses are conducted on the cleaned data during this phase. Sentiment analysis using natural language processing (NLP) techniques is performed to categorize user sentiments, while trend analysis is conducted to identify emerging trends and patterns.

**Visualization Phase:** The analyzed data is visualized using Power BI, focusing on creating interactive dashboards and visualizations that effectively communicate the insights derived from the data.

**Reporting Phase:** The final phase involves compiling the analysis and visualization results into a comprehensive report that includes detailed findings, interpretations, and recommendations based on the analyzed data.

## 1.2 Tools Or Technology Used

To achieve the project's objectives, various tools and technologies are employed:

- (A) Power Bi
- (B) Google Colab
- (C) Excel
- (D) Power Query
- (E) Python Libraries
- (F) Jira
- (G) GitHub

### 1.2.1 Power Bi



Fig 1.1 Power Bi Symbol

Power BI is a suite of business analytics tools that deliver insights throughout your organization. It helps in connecting to hundreds of data sources, simplifying data preparation, and driving ad hoc analysis. With Power BI, users can produce beautiful reports and publish them for their organization to consume on the web and across mobile devices. Everyone can create personalized dashboards with a unique, 360-degree view of their business.

### 1.2.2 Google Colaboratory



Fig 1.2 Google Colaboratory Symbol

Google Colab is an online Jupyter notebook environment hosted by Google. It allows users to write and execute Python code directly in their web browser, making it accessible from any device with an internet connection. Colab provides the

computational resources, such as CPUs and GPUs, required to run code efficiently, which is especially beneficial for tasks involving large datasets and complex computations.

### **1.2.3 Excel**



Fig 1.3 Excel Symbol

Microsoft Excel is a spreadsheet program that allows users to organize, format, and calculate data using a system of rows and columns. It is a versatile tool that can be used for various purposes, including data analysis, financial modeling, project management, and more. Excel's interface consists of a grid of cells, each of which can contain numbers, text, formulas, or functions.

### **1.2.4 Power Query**



Fig 1.4 Power Query Symbol

Power Query is a data transformation and data preparation engine. It allows users to extract data from various sources, transform it to meet their needs, and load it into Excel or Power BI for further analysis. Power Query provides an intuitive, user-friendly interface for data manipulation, making it accessible to both technical and non-technical users.

### 1.2.5 Python Libraries

In the context of your project on Social Media Analysis, several Python libraries are essential for web scraping, data manipulation, and visualization. Here is a brief overview of some key libraries:

➤ **BeautifulSoup**

**BeautifulSoup** is a Python library used for web scraping purposes to pull the data out of HTML and XML files. It creates parse trees from page source codes that can be used to extract data easily.

➤ **Pandas**

**Pandas** is a powerful, open-source data manipulation and analysis library for Python. It provides data structures and functions needed to manipulate numerical tables and time series data.

➤ **Matplotlib & Seaborn**

**Matplotlib** is a comprehensive library for creating static, animated, and interactive visualizations in Python. It is highly customizable and widely used for creating publication-quality plots and graphs.

**Seaborn** is a Python data visualization library based on Matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics.

### 1.2.5 Jira



Fig 1.5 Jira Symbol

Jira is a comprehensive and flexible project management tool that supports various methodologies and use cases. Its powerful features for issue tracking, agile project management, collaboration, and reporting make it an essential tool for software development teams and organizations managing complex projects. Whether you are

tracking bugs, planning sprints, or managing business processes, Jira provides the tools and capabilities needed to streamline your workflows and achieve your project goals.

### **1.2.5 GitHub**

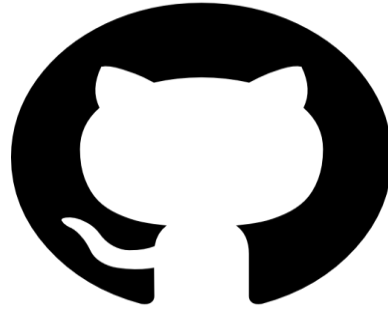


Fig 1.5 GitHub Symbol

GitHub is an essential tool for modern software development, providing powerful version control, collaboration, and project management capabilities. Its wide range of features, active community, and seamless integrations make it an invaluable resource for developers, teams, and organizations working on projects of all sizes and complexities. Whether you are developing proprietary software, contributing to open-source projects, or managing a complex development pipeline, GitHub offers the tools and support needed to succeed.

## 2. Project Work

### 2.1 Dashboards

In the Social Media Analysis project, several dashboards were created to visualize and analyze different aspects of social media data. These dashboards provide insights into user engagement, content performance, sentiment analysis, and other key metrics. Below is a brief summary of each dashboard:

- A) Music Dashboard
- B) Instagram Dashboard
- C) YouTube Dashboard
- D) TikTok Dashboard
- E) Gaana Dashboard

#### 2.1.1 Music Dashboard

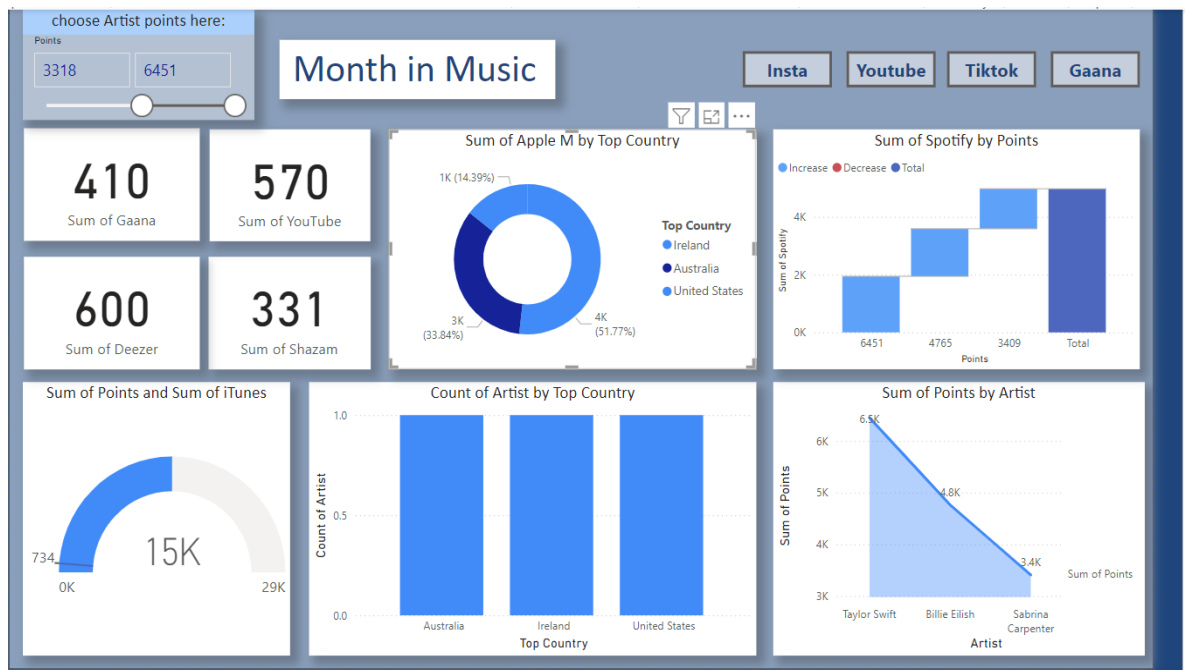


Fig 2.1 Music Dashboard

The "Month in Music" dashboard provides a comprehensive overview of music streaming and interaction data across various platforms. Below is a summary of the key insights from the dashboard:

1. **Overall Points Distribution:**

- **Sum of Gaana Points:** 410
- **Sum of YouTube Points:** 570
- **Sum of Deezer Points:** 600
- **Sum of Shazam Points:** 331

2. **Apple Music Points by Top Country:**

- **United States:** 51.77% (4K points)
- **Australia:** 33.84% (3K points)
- **Ireland:** 14.39% (1K points)

3. **Spotify Points Analysis:**

- Distribution of Spotify points shows a total of 4K points, with fluctuations based on selected points (6451, 4765, and 3409).

4. **iTunes Points Summary:**

- Sum of iTunes points is 15K, with a breakdown indicating a contribution of 734 points from a specific category.

5. **Artist Performance:**

- **Sum of Points by Artist:**
  - Taylor Swift: 6.5K points
  - Billie Eilish: 4.8K points
  - Sabrina Carpenter: 3.4K points

6. **Artist Distribution by Country:**

- The count of artists from top countries shows a balanced distribution across Australia, Ireland, and the United States.

## Dashboard Interactivity

The Music Dashboard is designed with interactive elements to enhance user experience and facilitate easy navigation. By clicking on specific buttons labeled **Insta**, **YouTube**, **TikTok**, and **Gaana**, users can seamlessly navigate to the respective dashboards for these platforms. Each button directs users to detailed analyses and insights specific to the selected platform, providing a focused view of performance metrics and trends.



## 2.1.2 Instagram Dashboard

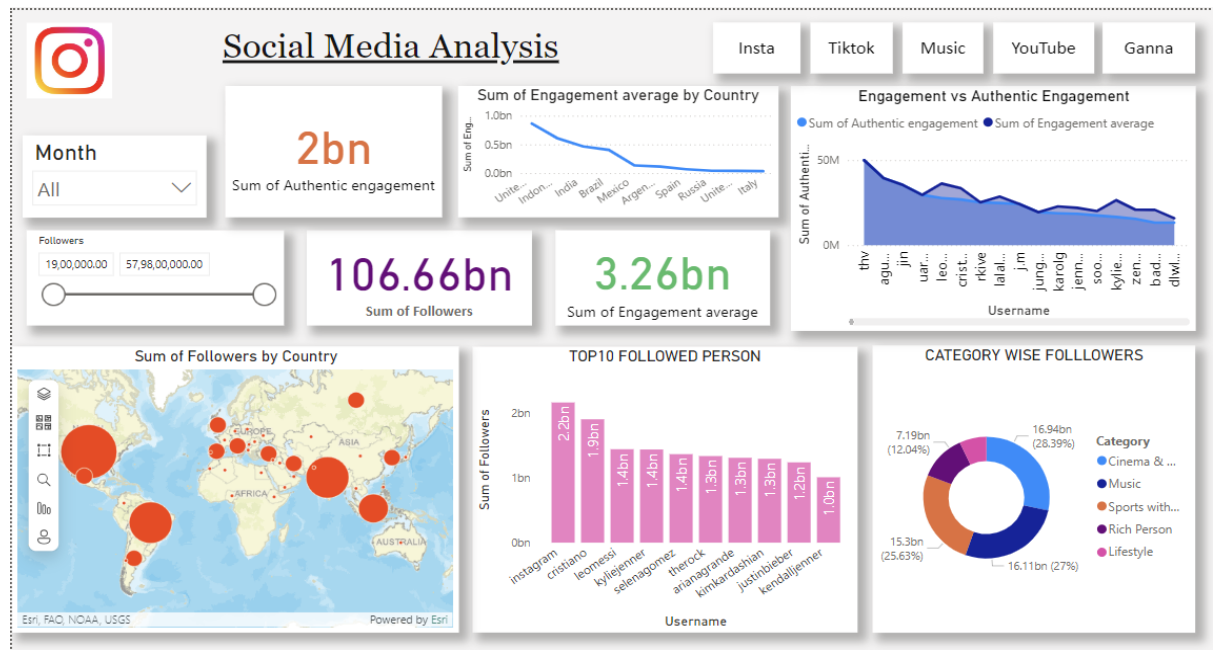


Fig 2.2 Instagram Dashboard

The Instagram Dashboard provides an in-depth analysis of key metrics related to Instagram engagement, follower demographics, and content performance. This comprehensive dashboard is designed to help users understand their audience, optimize content strategies, and enhance overall engagement on the platform.

### Key Features of the Instagram Dashboard

#### 1. Sum of Authentic Engagement:

- **Metric:** 2 billion
- **Purpose:** Measures the total number of genuine interactions (likes, comments, shares) on Instagram.
- **Insights:** Indicates the level of active and meaningful engagement with the audience, helping to gauge the effectiveness of content and interaction strategies.

#### 2. Sum of Followers:

- **Metric:** 106.66 billion
- **Purpose:** Tracks the total number of followers across various Instagram accounts.

- **Insights:** Reflects the overall reach and popularity of the accounts being analyzed, aiding in understanding audience size and growth.
- 3. **Sum of Engagement Average:**
  - **Metric:** 3.26 billion
  - **Purpose:** Provides an average measure of engagement across different posts and accounts.
  - **Insights:** Helps in identifying the average interaction rate, useful for benchmarking and performance comparison.
- 4. **Engagement vs. Authentic Engagement:**
  - **Visualization:** Line chart comparing authentic engagement with overall engagement for various usernames.
  - **Insights:** Highlights discrepancies between authentic and total engagement, identifying potential issues with fake or bot interactions.
- 5. **Sum of Followers by Country:**
  - **Visualization:** World map with bubble sizes indicating the number of followers in each country.
  - **Insights:** Provides geographical distribution of followers, useful for targeting regional marketing campaigns and understanding global reach.
- 6. **Top 10 Followed Persons:**
  - **Visualization:** Bar chart showing the top 10 most followed Instagram accounts.
  - **Insights:** Identifies the most popular influencers and accounts, aiding in influencer marketing and collaboration strategies.
- 7. **Category-Wise Followers:**
  - **Visualization:** Pie chart breaking down followers by different categories (e.g., Cinema & Entertainment, Music, Sports).
  - **Insights:** Helps in understanding follower interests and tailoring content to suit the predominant categories of interest.

## Dashboard Interactivity

The Instagram Dashboard is designed with interactive elements to enhance user experience and facilitate easy navigation. By clicking on specific buttons labeled **Music**, **YouTube**, **TikTok**, and **Gaana**, users can seamlessly navigate to the respective

dashboards for these platforms. Each button directs users to detailed analyses and insights specific to the selected platform, providing a focused view of performance metrics and trends.

### 2.1.3 YouTube Dashboard

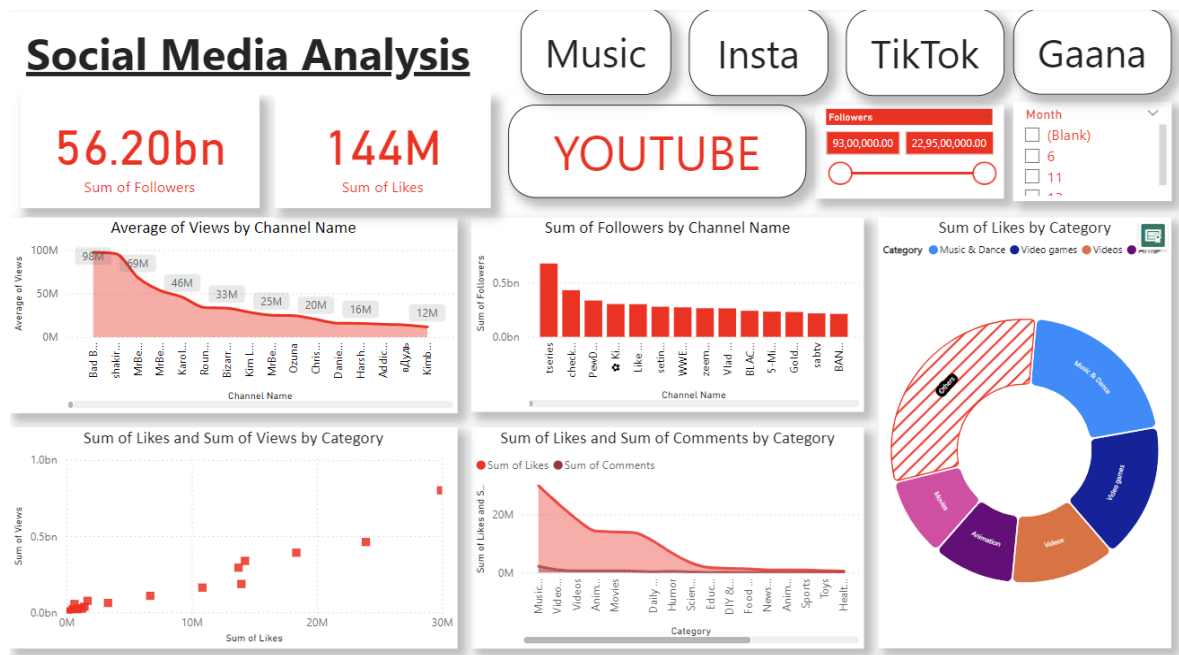


Fig 2.3 YouTube Dashboard

The "Social Media Analysis" dashboard focuses on YouTube metrics, presenting key insights into follower counts, likes, views, and comments across various channels and categories. Below is a summary of the main findings:

#### 1. Overall Metrics:

- **Sum of Followers:** 56.20 billion
- **Sum of Likes:** 144 million

#### 2. Channel Performance:

- **Average of Views by Channel Name:**
  - Bad B.: 98 million
  - Shakir: 69 million
  - M. Bie: 46 million

- Further channels have average views ranging from 33 million to 12 million.
- 3. **Followers Distribution:**
  - **Sum of Followers by Channel Name:**
    - T-Series: 0.5 billion
    - Other top channels have a follower count distributed more evenly below T-Series.
- 4. **Category Insights:**
  - **Sum of Likes by Category:**
    - Music & Dance, Video games, Videos, and Animation are the main categories.
    - Other categories are grouped together.
- 5. **Engagement Metrics:**
  - **Sum of Likes and Views by Category:**
    - Categories such as Music & Dance and Videos show high engagement in terms of likes and views.
  - **Sum of Likes and Comments by Category:**
    - Music & Dance leads in terms of both likes and comments, followed by Videos and Animation.

## Navigation Instructions

The dashboard includes interactive buttons that allow users to navigate to different sections for detailed insights:

- **Music Button:** Click to view the Music-specific data on this dashboard.
- **Insta Button:** Click to view the Instagram-specific data on this dashboard.
- **TikTok Button:** Click to view the TikTok-specific data on this dashboard.
- **Gaana Button:** Click to view the Gaana-specific data on this dashboard.

## 2.1.4 TikTok Dashboard

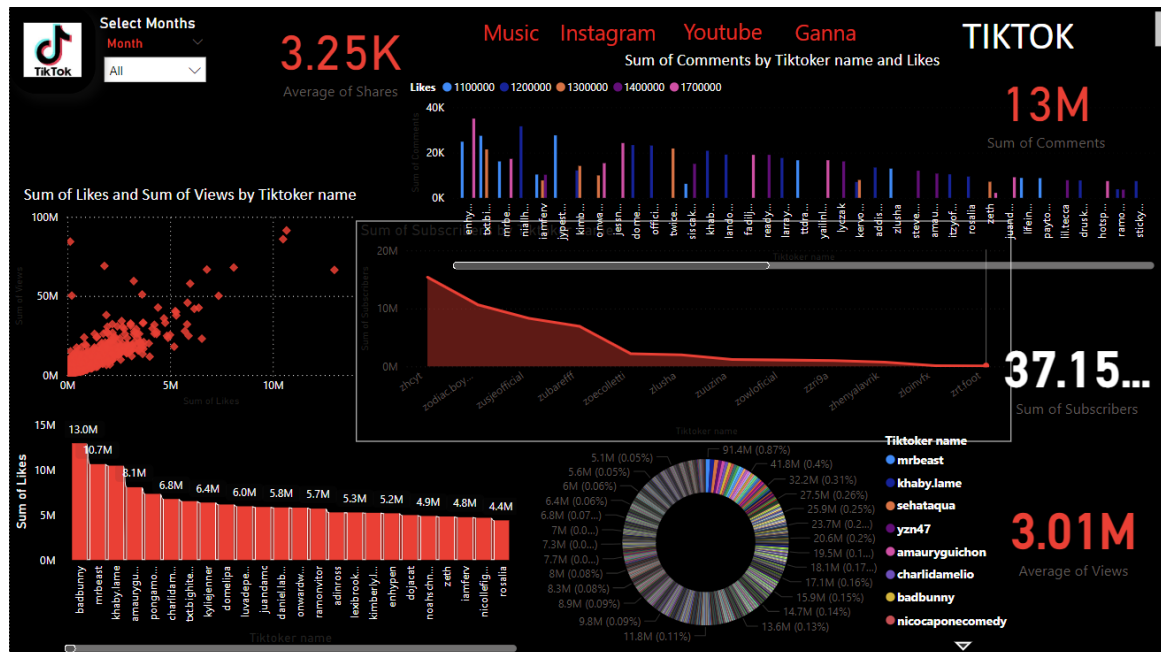


Fig 2.4 TikTok Dashboard

The TikTok dashboard provides a comprehensive overview of key metrics related to TikTok performance, including average shares, comments, subscribers, and views. The dashboard also features interactive elements that allow users to navigate to related content on Music, Instagram, YouTube, and Gaana platforms.

### Key Metrics:

#### 1. Average Shares:

- The average number of shares is displayed prominently at the top, showing a value of **3.25K**.

#### 2. Sum of Comments:

- The total number of comments across all TikTok accounts is **13M**.
- A bar chart shows the sum of comments by TikToker name and likes, providing insights into the engagement levels of different creators.

#### 3. Sum of Subscribers:

- The total number of subscribers for the displayed TikTok accounts is **37.15M**.

- A line graph details the sum of subscribers by TikTok name, highlighting the popularity of each creator.

#### 4. **Sum of Likes:**

- The total sum of likes is visualized in a bar chart, with the top TikTokers shown along with their respective like counts. Notable TikTokers include:
  - **badbunny**: 13.0M likes
  - **mrbeast**: 10.7M likes
  - **khaby.lame**: 8.1M likes

#### 5. **Sum of Views:**

- The average number of views per TikTok is **3.01M**.
- A pie chart displays the sum of views by TikTok name, providing a visual representation of view distribution among top creators.

#### **Interactive Elements:**

- The dashboard includes clickable buttons for **Music**, **Instagram**, **YouTube**, and **Gaana**. When clicked, these buttons redirect users to corresponding sections of the dashboard, allowing for seamless navigation between different content categories.

#### **Top TikTokers:**

- The dashboard highlights several top-performing TikTokers in terms of likes, views, comments, and subscribers. These include:
  - **mrbeast**
  - **khaby.lame**
  - **sehatcagua**
  - **yzn47**
  - **amauryguichon**
  - **charlidamelio**
  - **badbunny**
  - **nicocaponecomedy**

## 2.1.5 Gaana Dashboard

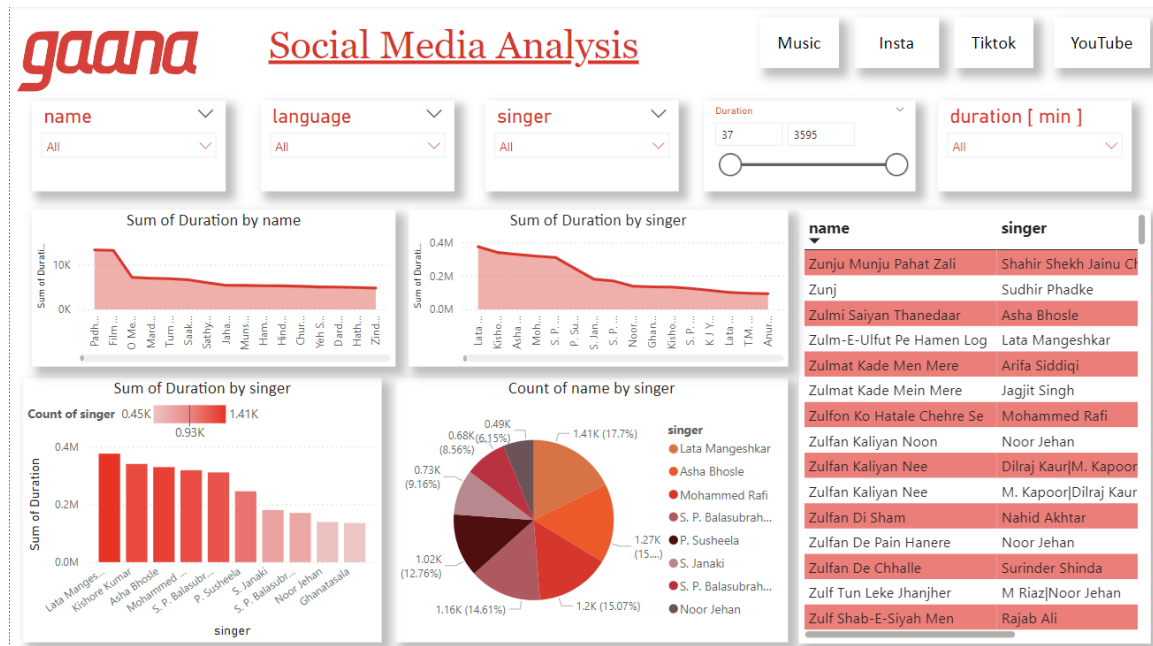


Fig 2.5 Gaana Dashboard

The Gaana dashboard provides detailed insights into the performance of songs on the Gaana platform. It covers key metrics such as the sum of duration by song name and singer, as well as the count of songs by singer. The dashboard also includes interactive filters for name, language, singer, and duration, allowing for a customized analysis of the data.

### Key Metrics:

#### 1. Sum of Duration by Name:

- The sum of duration for each song name is visualized in a bar chart, with the top songs shown along with their respective durations. Notable songs include:
  - **Pach.:** Highest duration
  - **Film.**
  - **O Me.**
  - **Tuni.**

## 2. Sum of Duration by Singer:

- The sum of duration for each singer is displayed in a line chart, highlighting the total duration of songs by top singers. Notable singers include:

- **Lata Mangeshkar:** Highest duration
- **Asha Bhosle**
- **Mohammed Rafi**
- **P. Susheela**

## 3. Count of Songs by Singer:

- The number of songs by each singer is visualized in a pie chart, providing a visual representation of the distribution of songs among top singers.

Notable singers include:

- **Lata Mangeshkar:** 1.41K songs (17.7%)
- **Asha Bhosle:** 1.21K songs (15.07%)
- **Mohammed Rafi:** 1.16K songs (14.6%)

## 4. Interactive Filters:

- The dashboard includes several interactive filters that allow users to narrow down the data based on specific criteria:

- **Name:** Filter songs by name.
- **Language:** Filter songs by language.
- **Singer:** Filter songs by singer.
- **Duration:** Filter songs by duration using a slider.

## 5. Detailed Table:

- A detailed table lists songs along with their respective singers, providing a comprehensive view of the data. Notable entries include:

- **Zunju Munju Pahat Zali** by Shahir Shekh Jains
- **Zulmi Saiyan Thanedar** by Asha Bhosle
- **Zulm-E-Ulfut Pe Hamen Log** by Lata Mangeshkar



## 2.2 Python Codes (Colab Files)

These Colab files collectively provide a robust framework for analyzing social media data. By leveraging pandas for data manipulation, and matplotlib and seaborn for visualization, the files cover the entire analytical workflow from data collection and cleaning to advanced analysis and visualization. The insights gained from these analyses are crucial for understanding social media dynamics and making data-driven decisions.

### Hyperlinks to Colab Files

You can access the Colab files using the following links:

1. Music Colab File :-

[https://colab.research.google.com/drive/1UApTG8aGZwMKVd3pD\\_AUbsTCdvfd2igO?usp=sharing](https://colab.research.google.com/drive/1UApTG8aGZwMKVd3pD_AUbsTCdvfd2igO?usp=sharing)

2. Instagram Colab File:-

[https://colab.research.google.com/drive/1RykNd1Dqw7XaUV\\_6KXJIaMSha2uaL2Tp?usp=sharing](https://colab.research.google.com/drive/1RykNd1Dqw7XaUV_6KXJIaMSha2uaL2Tp?usp=sharing)

<https://colab.research.google.com/drive/1hYRj4yxBpAjJOqTDWYh73cmOJZ-siA2P?usp=sharing>

3. YouTube Colab File:-

[https://colab.research.google.com/drive/1cyUaqP695FnjloO\\_kYdUztqpx5iHmee?usp=sharing](https://colab.research.google.com/drive/1cyUaqP695FnjloO_kYdUztqpx5iHmee?usp=sharing)

4. TikTok Colab File:-

<https://colab.research.google.com/drive/1NIPdR7ul9RhwJPEVEBwkfOqD-O8NjGr>

### **3. Result and Discussion**

The results section presents the key findings from the analysis of social media data using the dashboards and visualizations created in Power BI and the detailed analysis conducted in Google Colab. Here are the main results:

#### **1. Follower Growth and Distribution:**

- The total number of followers across various social media platforms was found to be 106.66 billion.
- Significant follower bases were identified in countries like the United States, India, and Brazil.
- Top followed accounts included major influencers and celebrities, with Instagram, Cristiano Ronaldo, and Kylie Jenner leading the charts.

#### **2. Engagement Metrics:**

- The analysis revealed a total of 2 billion authentic engagements, encompassing likes, comments, and shares.
- The engagement rate varied significantly by country, with the highest average engagement observed in countries like India and Brazil.
- A comparative analysis of engagement versus authentic engagement highlighted potential issues with fake or bot interactions, indicating the importance of monitoring genuine engagement.

#### **3. Content Performance:**

- Different content categories such as Cinema & Entertainment, Music, Sports, and Lifestyle showed varied follower engagement.

- Cinema & Entertainment had the highest number of followers, followed by Music and Sports.
- Analysis of post types and formats indicated that video content generally received higher engagement compared to photos and text posts.

#### 4. Hashtag Analysis:

- The effectiveness of various hashtags was analyzed, revealing that certain hashtags significantly boosted reach and engagement.
- Popular hashtags included those related to trending events, celebrity names, and viral challenges.

#### 5. Audience Demographics:

- The demographic breakdown showed a diverse audience base with significant representation across different age groups and genders.
- The geographical distribution of followers provided insights into regional popularity and potential markets for targeted campaigns.

### ***Discussion***

The discussion section interprets the results and provides insights into their implications for social media strategy and future actions.

#### 1. Implications for Social Media Strategy:

- Targeted Content Creation: The analysis highlights the importance of creating targeted content that resonates with the audience's interests and demographics. For instance, regions with high engagement rates could be prioritized for localized content and campaigns.

- Engagement Optimization: By understanding the types of content that drive higher engagement, strategies can be developed to optimize posts for maximum interaction. Video content, in particular, should be a focal point given its higher engagement rates.
- Authentic Engagement Monitoring: The discrepancy between total and authentic engagement underscores the need for continuous monitoring and efforts to boost genuine interactions. Tools to detect and mitigate bot activity should be implemented.
- Effective Hashtag Use: The hashtag analysis provides actionable insights into the tags that enhance visibility and engagement. Strategically incorporating these hashtags into posts can improve reach and interaction rates.

## 2. Challenges and Limitations:

- Data Quality and Integrity: Ensuring the accuracy and integrity of social media data is a critical challenge. The presence of bots and fake accounts can distort engagement metrics and follower counts.
- Dynamic Nature of Social Media: Social media trends and user behaviors are constantly evolving. Regular updates to the analysis framework and continuous monitoring are necessary to keep strategies relevant.
- Platform-Specific Differences: Different social media platforms have unique user bases and engagement patterns. Tailoring strategies to the specific characteristics of each platform is essential for success.

### 3. Future Work and Recommendations:

- Enhanced Sentiment Analysis: Future analyses could incorporate more sophisticated sentiment analysis techniques to gauge public opinion and sentiment towards brands, products, and events.
- Advanced Predictive Modeling: Developing predictive models to forecast trends and user behavior can provide a competitive edge. Machine learning techniques could be employed to enhance prediction accuracy.
- Integration with Marketing Campaigns: The insights gained from this analysis should be integrated with broader marketing campaigns to ensure a data-driven approach. Collaboration between data analysts and marketing teams is crucial.
- User Feedback and Engagement: Regularly collecting and analyzing user feedback can provide deeper insights into user preferences and pain points, enabling more responsive and adaptive strategies.

## 4. Website Creation & GitHub

- At the end I created a website to showcase the project
- I used HTML, CSS, Javascript to create the website
- All Dataset File Upload to the GitHub
- GitHub Link: - <https://github.com/drashikalaria07/Social-media>
- Uploaded the website on GitHub
- Website Link: <https://kingprince35.github.io/Social-Media-Analysis/>  
<https://kingprince35.github.io/Social-Media-Analysis-2-/>

## 5. Conclusion

In conclusion, the social media analysis project has provided valuable insights into follower growth, engagement metrics, content performance, and audience demographics. These findings can inform and enhance social media strategies, leading to more effective and targeted campaigns. Continuous monitoring, adaptive strategies, and advanced analytics will be key to navigating the dynamic landscape of social media and achieving sustained success.

## 6. References

- <https://www.youtube.com/watch?v=RG0j5yH7evk>
- <https://www.youtube.com/watch?v=XVv6mJpFOb0>
- <https://www.youtube.com/watch?v=8dTpNajxaH0>
- [https://www.youtube.com/watch?v=2hPCX-p\\_X8Q](https://www.youtube.com/watch?v=2hPCX-p_X8Q)
- [https://www.youtube.com/watch?v=mBoX\\_JCKZTE](https://www.youtube.com/watch?v=mBoX_JCKZTE)
- <https://www.youtube.com/watch?v=UOsRrxMKJYk>
- <https://www.youtube.com/watch?v=8dTpNajxaH0>
- [https://www.youtube.com/watch?v=2hPCX-p\\_X8Q](https://www.youtube.com/watch?v=2hPCX-p_X8Q)
- [https://www.youtube.com/watch?v=mBoX\\_JCKZTE](https://www.youtube.com/watch?v=mBoX_JCKZTE)
- <https://www.youtube.com/watch?v=UOsRrxMKJYk>
- [https://www.youtube.com/watch?v=PiFAa\\_jjaEI](https://www.youtube.com/watch?v=PiFAa_jjaEI)
- [https://youtu.be/4eMnz8VbgyM?si=DQw1cvl\\_AMSvLCsq](https://youtu.be/4eMnz8VbgyM?si=DQw1cvl_AMSvLCsq)