

# GOVERNMENT COLLEGE OF ENGINEERING [IRTT] ERODE-638316

**Computer Science and Engineering**



# NAAN MUDHALVAN

**TEAM ID: NM2023TMID11499**

***SB8040 - Professional Readiness for Innovation, Employment and Entrepreneurship***

***By Naan Mudhalvan scheme – May 2023***

# PROJECT DOMAIN: DATA ANALYTICS

**PROJECT TITLE: Dissecting the Digital Landscape: A Comprehensive Analysis of Social Media.**

|  |  |
| --- | --- |
| **TEAM MEMBERS** | |
| **REGISTER NUMBER** | **NAME** |
| **731120104034** | **PAUL PANDIYAN K** |
| **731120104016** | **GOKUL P** |
| **731120104009** | **DHARANITHARAN R** |
| **731120104008** | **DHANAPAL M** |

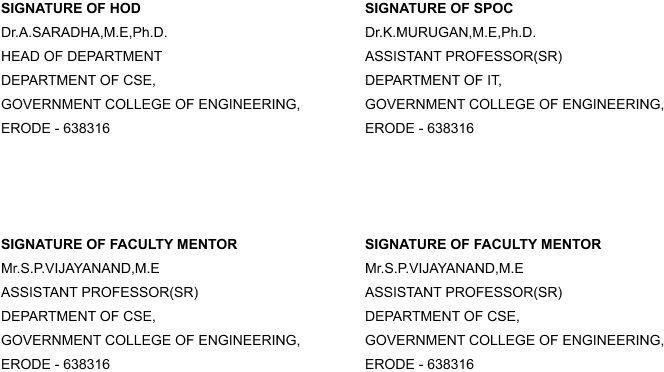
GOVERNMENT COLLEGE OF ENGINEERING [IRTT] ERODE-638316



BONAFIDE CERTIFICATE

Certified that this project report **“Dissecting the Digital Landscape: A Comprehensive Analysis of social media “**is the Bonafide work of **“PAUL PANDIYAN K (731120104034), GOKUL P (731120104016), DHARANITHARAN R**

**(731120104009), DHANAPAL M (731120104008) “**who carried out the project work under my supervision.



**Project Report Format**

1. **INTRODUCTION**
   1. Project Overview
   2. Purpose
2. **IDEATION & PROPOSED SOLUTION**
   1. Problem Statement Definition
   2. Empathy Map Canvas
   3. Ideation & Brainstorming
   4. Proposed Solution
3. **REQUIREMENT ANALYSIS**
   1. Functional requirement
   2. Non-Functional requirements
4. **PROJECT DESIGN**
   1. Data Flow Diagrams
   2. Solution & Technical Architecture
   3. User Stories
5. **CODING & SOLUTION (Explain the features added in the project along with code)**
   1. Feature 1
   2. Feature 2
   3. Database Schema (if Applicable)
6. **RESULTS**
   1. Performance Metrics
7. **ADVANTAGES & DISADVANTAGES**
8. **CONCLUSION**
9. **FUTURE SCOPE**
10. **APPENDIX**

Source Code , GitHub & Project Video Demo Link

# INTRODUCTION

* 1. **PROJECT OVERVIEW**

The project aims to develop a comprehensive analysis system for social media platforms. It involves collecting data from various social media sources, creating data modules for processing and analyzing the collected data, and providing visualizations and dashboards for better insights. The system will be built using modern technologies and will be embedded in a Flask-based web application.

Key Features:

* + - Data Collection: Implement mechanisms to collect data from popular social media platforms, such as Twitter, Facebook, Instagram, etc. This includes fetching posts, comments, user profiles, and other relevant information.
    - Data Modules: Develop modules for data processing, cleaning, and structuring. Apply techniques for data normalization, filtering, and feature extraction to prepare the data for analysis.
    - Visualization and Dashboard: Create visually appealing and interactive visualizations to represent the analyzed social media data. Develop a dashboard that provides users with a

comprehensive overview of social media trends, sentiment analysis, user engagement, and other relevant metrics.

* + - Embedded Flask Application: Integrate the social media analysis system into a Flask-based web application. Develop user authentication and access control mechanisms to ensure secure access to the system. Enable users to interact with the visualizations and customize the analysis based on their preferences.

# PURPOSE

* + - Data Analysis: The project seeks to analyze social media data from various platforms to uncover patterns, trends, and valuable information. By applying data analytics techniques, the project aims to provide users with a comprehensive understanding of social media dynamics, user behavior, sentiment analysis, and engagement metrics.
    - Decision-Making Support: The project aims to support decision-making processes by providing data-driven insights and visualizations. Users can leverage the analysis results to make informed decisions, develop

effective social media strategies, identify market trends, and assess the impact of their social media activities.

* + - Monitoring and Tracking: The project intends to monitor social media activities in real-time, enabling users to track their brand reputation, monitor competitor activities, and stay updated with the latest trends and developments in their industry. By providing timely and relevant information, the system helps users proactively respond to social media events and engage with their target audience effectively.
    - User Engagement Enhancement: The project aims to enhance user engagement by providing an interactive and user-friendly interface. Users can explore visualizations, customize analysis parameters, and interact with the data to gain deeper insights. By facilitating user engagement, the system aims to improve user satisfaction and encourage regular usage.

# IDEATION & PROPOSED SOLUTION

* 1. **PROBLEM STATEMENT DEFINITION**
     + Inadequate Tools: Existing social media analysis tools lack the capability to effectively analyse and understand social media data, leading to limited insights and decision-making support.
     + Data Volume and Complexity: The vast amount of social media data, including posts, comments, and user profiles, presents challenges in extracting meaningful information and identifying patterns due to the volume, variety, and complexity of the data.
     + Limited Analytics and Visualization: Current tools often lack advanced analytics capabilities and intuitive visualizations, restricting users' ability to explore social media trends, sentiment analysis, and user engagement metrics in a comprehensive manner.

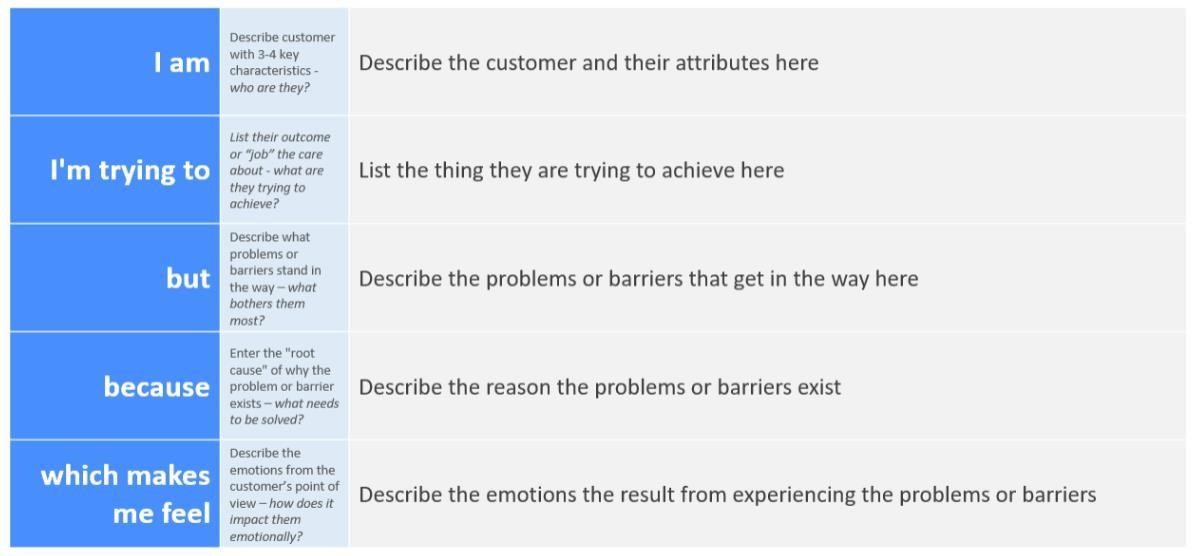
Security and Trust: Users require robust security implementations to ensure the privacy and protection of their data. The lack of secure measures in existing systems hampers users' trust and inhibits their willingness to engage with social media analysis platforms

|  |  |
| --- | --- |
| Date | 04 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Dissecting the Digital Landscape : A Comprehensive Analysis of Social Media |
| Maximum Marks |  |

# Customer Problem Statement Template:

Social media has become an essential part of our daily lives, and it has significantly impacted the way we communicate, share information, and consume content. However, despite its widespread adoption, there is still a lack of comprehensive understanding of the digital landscape and its impact on individuals, organizations, and society as a whole.

Many businesses struggle to develop effective social media strategies and fail to leverage its full potential. Therefore, there is a need for a comprehensive analysis of the social media landscape to better understand its opportunities and challenges and guide businesses in developing effective social media strategies.



**Example:**



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Problem Statement (PS)** | **I am (Customer)** | **I’m trying to** | **But** | **Because** | **Which makes me feel** |
| PS-1 | Busy professional | Manage work-life balance by staying healthy and fit | Struggle to find time and motivation to exercise regularly | Busy schedule and lack of motivation | Stressed, tired, and unhappy with body |
| PS-2 | Student | Improve learning and academic performance | Difficulty in concentrating and staying focused while studying | Affects ability to retain information and perform well in exams | Anxious, overwhelmed, and frustrated with self |

# Empathy Map Canvas

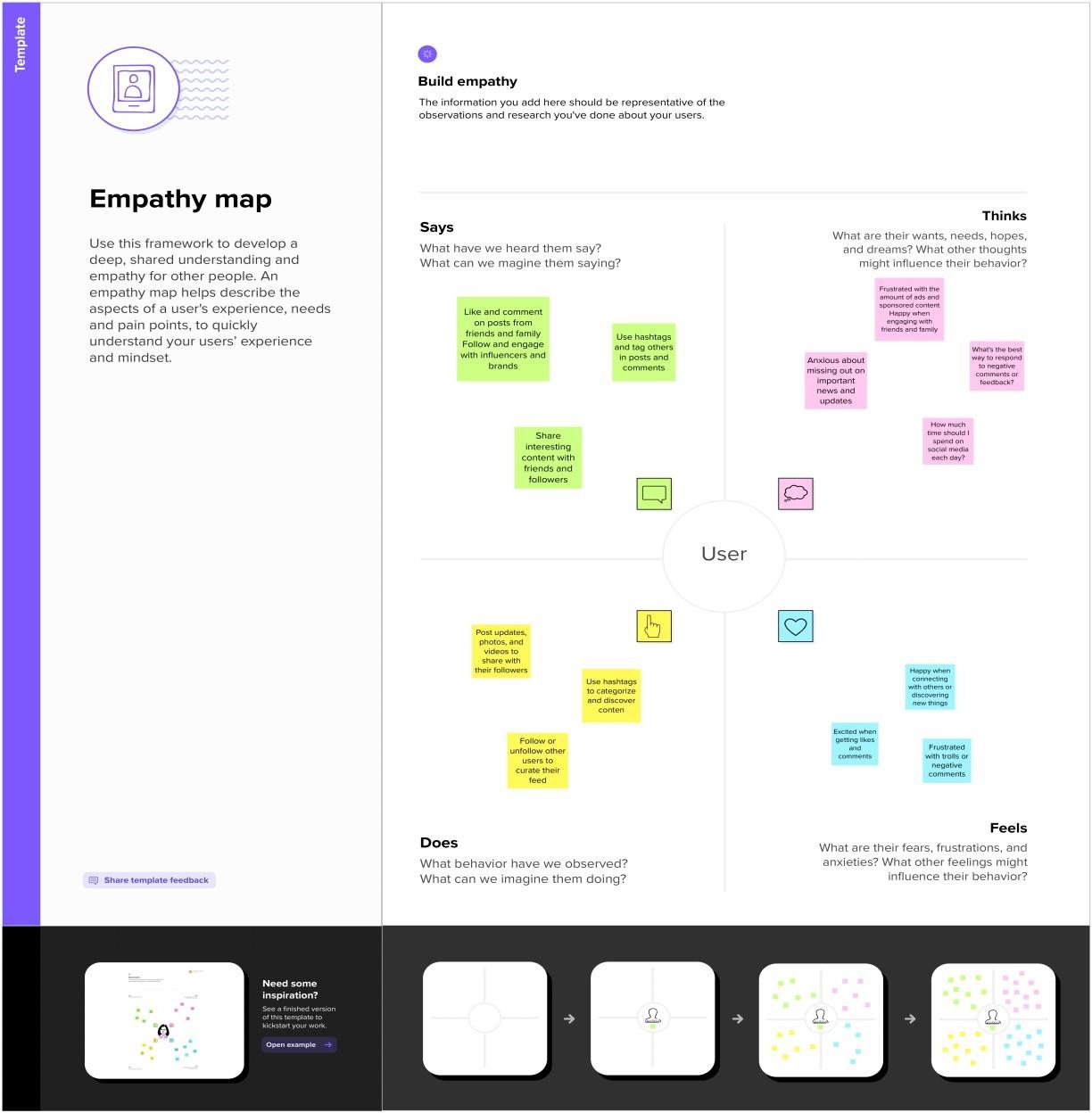
**Ideation Phase Empathize & Discover**

|  |  |
| --- | --- |
| Date | 04 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Project - Dissecting the Digital  Landscape: A Comprehensive Analysis of Social Media |
| Maximum Marks | 4 Marks |

# Empathy Map Canvas:

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user’s behaviours and attitudes.

It is a useful tool to helps teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user’s perspective along with his or her goals and challenges.





# Ideation & Brainstorming

**Ideation Phase Brainstorm & Idea Prioritization Template**

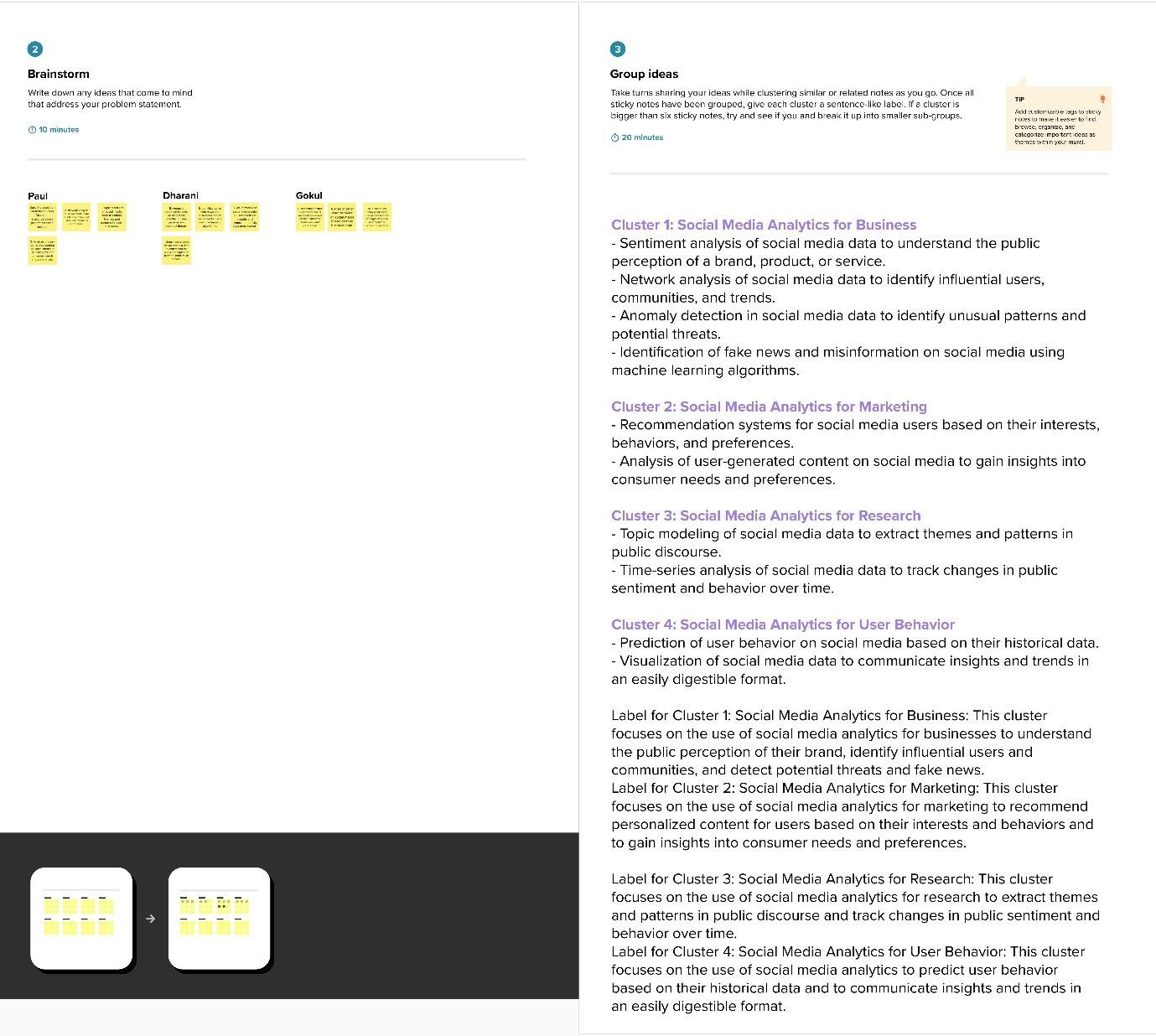
|  |  |
| --- | --- |
| Date | 04 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | A comprehensive analysis of social media |
| Maximum Marks |  |

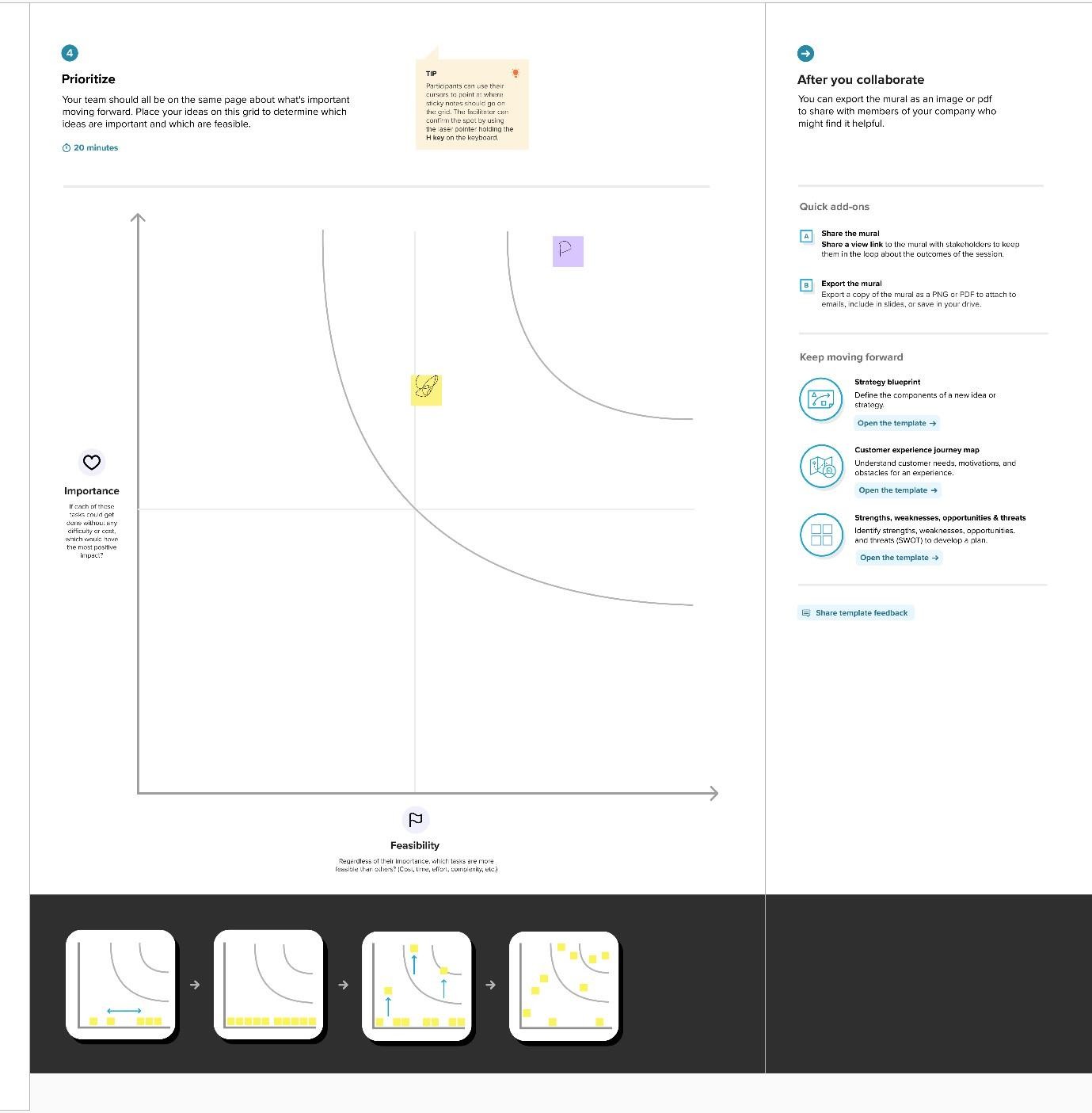
# Brainstorm & Idea Prioritization Template:

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of- the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

# Step-1: Team Gathering, Collaboration and Select the Problem Statement





* 1. **Proposed Solution**
* Data Collection Module: Implement a module that efficiently collects data from various social media platforms, leveraging their APIs or web scraping techniques. The module should handle large volumes of data and ensure data integrity and reliability.
* Data Processing and Analysis: Develop advanced data processing algorithms to clean, preprocess, and analyze social media data. Apply techniques such as natural language processing (NLP), text mining, and machine learning to extract meaningful insights, sentiment analysis, and user behavior patterns.
* Visualization and Dashboard: Create interactive visualizations and dashboards that present the analyzed data in a user-friendly and intuitive manner. Provide customizable options for users to explore different metrics, compare trends, and generate reports for further analysis.
* Integration and Scalability: Build a scalable architecture that can handle increasing data loads and accommodate future growth. Implement distributed computing technologies, cloud infrastructure, or scalable databases to ensure seamless performance and scalability.
* Security Implementations: Incorporate robust security measures to protect user data and ensure privacy. Implement authentication, encryption, and access controls to safeguard sensitive information and comply with data protection regulations.

# REQUIREMENT ANALYSIS

**Project Design Phase-II**

**Solution Requirements (Functional & Non-functional)**

|  |  |
| --- | --- |
| Date | 08 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Dissecting the Digital Landscape : A Comprehensive Analysis of Social Media |

# Functional Requirements:

Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR**  **No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | Data Collection | Collect social media posts from Twitter API Collect user profiles from Facebook  API |
| FR-2 | Data Processing | Clean and preprocess collected data |
| FR-3 | Data Analysis | Perform topic modeling on social media data |
| FR-4 | Visualization | Create interactive dashboards for data visualization |
| FR-5 | Search and Filtering | Allow filtering of data by date, location, or user |

|  |  |  |
| --- | --- | --- |
| FR-6 | Security | Implement user authentication and access controls |

# Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR**  **No.** | **Non-Functional Requirement** | **Description** |
| NFR- 1 | **Usability** | The system should ensure data confidentiality and integrity, implementing encryption and access  controls. |
| NFR- 2 | **Security** | The system should be able to handle increasing data volumes and user activity, with the ability to scale horizontally or  vertically. |
| NFR- 3 | **Reliability** | The system should have a user-friendly interface, with intuitive navigation and clear instructions for user interactions. |
| NFR- 4 | **Performance** | The system should process and analyze data efficiently, providing timely and responsive results to users. |

# PROJECT DESIGN

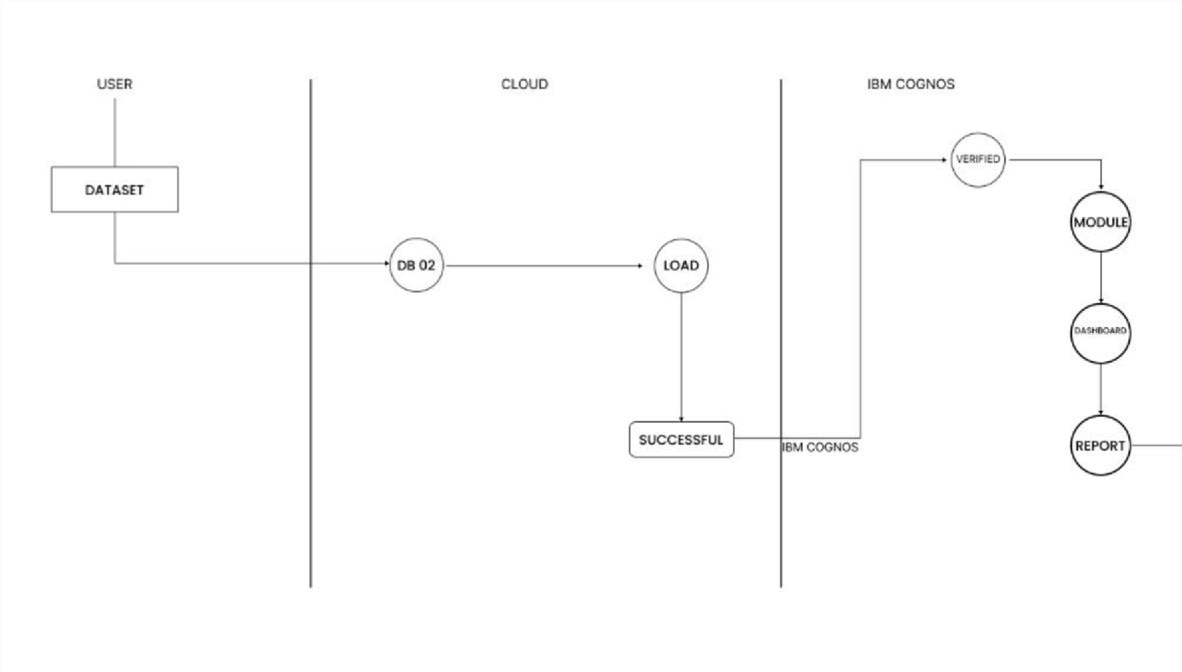
**Project Design Phase-II**

**Data Flow Diagram & User Stories**

|  |  |
| --- | --- |
| Date | 08 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Dissecting the Digital Landscape : A Comprehensive  Analysis of Social Media |

**Data Flow Diagrams:**

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



* 1. **SOLUTION & TECHNICAL ARCHUTECTURE**

**Project Design Phase-I Solution Architecture**

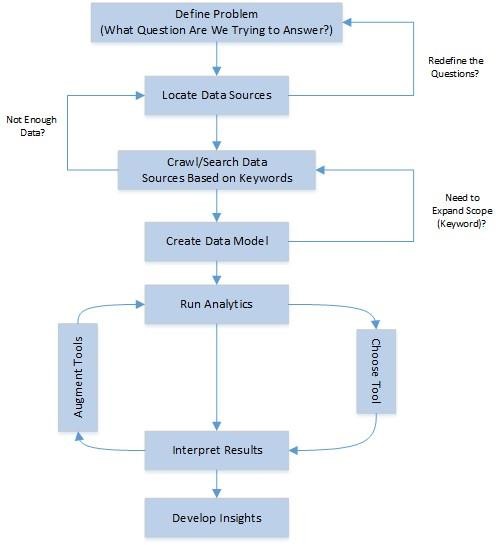
|  |  |
| --- | --- |
| Date | 05 May 2023 |

|  |  |
| --- | --- |
| Team ID | NM2023TMID11499 |
| Project Name | Project - Dissecting the Digital Landscape: A Comprehensive Analysis of Social Media |
| Maximum Marks |  |

# Solution Architecture:

Solution architecture is a complex process – with many sub- processes – that bridges the gap between business problems and technology solutions. Its goals are to:

* + - Find the best tech solution to solve existing business problems.
    - Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.
    - Define features, development phases, and solution requirements.
    - Provide specifications according to which the solution is defined, managed, and delivered.



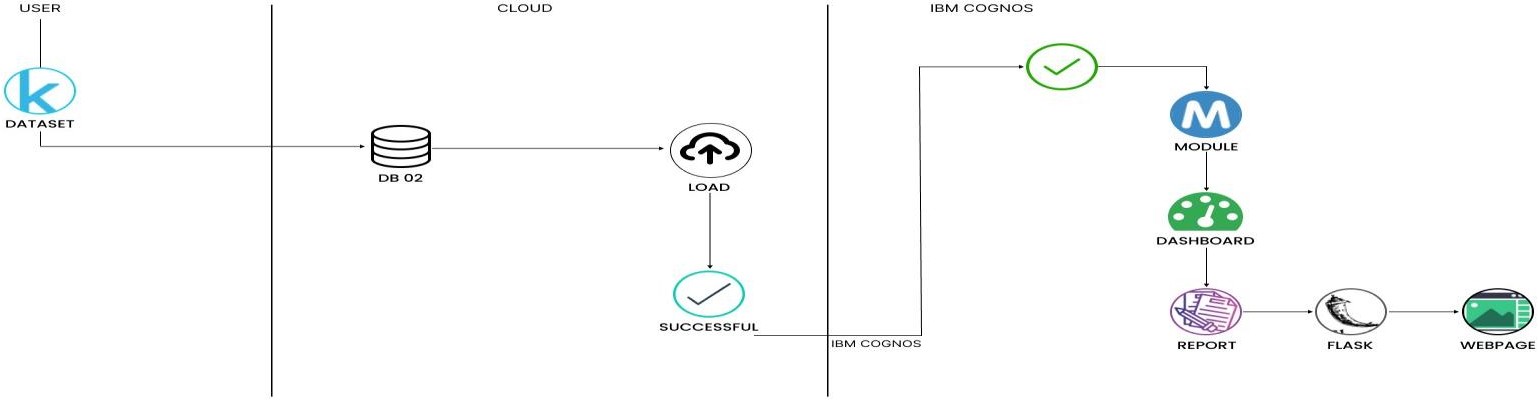
**P**

**Technology Stack (Architecture & Stack)**

|  |  |
| --- | --- |
| Date | 08 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Dissecting the Digital Landscape : A Comprehensive Analysis of Social Media |

**Technical Architecture:**

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



**Table-1 : Components & Technologies:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Component** | **Description** | **Technology** |
| 1. | Data Collection | Gathering social media data from various platforms | APIs, Web scraping |
| 2. | Data Module | Creating a module to process and manage collected data | Python, Pandas, SQL |
| 3. | Visualization | Creating visual representations of analyzed data | Matplotlib, Plotly, D3.js |
| 4. | Dashboard | Developing a web-based dashboard to present data visually | Flask, HTML, CSS,  JavaScript |
| 5. | Flask Integration | Embedding the dashboard within a Flask application | Flask, Jinja2 Templates |
| 6. | Scalability | Designing architecture for scalability and handling large  datasets | Distributed systems, Cloud infrastructure |

|  |  |  |  |
| --- | --- | --- | --- |
| 7. | Scalability | Implementing security measures  for data protection and access control | Encryption, Authentication, Authorization |

**Table-2: Application Characteristics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | Utilizing open-source frameworks for social media analysis | Apache Hadoop, Apache Spark, Python |
| 2. | Security Implementations | Implementing security measures to protect data and user privacy | Encryption, access controls, secure protocols |
| 3. | Scalable Architecture | Designing an architecture that can handle growing data volumes and  user activity | Distributed computing, cloud infrastructure |
| 4. | Availability | Ensuring the platform is accessible and available to users | Load balancing, fault tolerance mechanisms |
| 5. | Performance | Optimizing the system for efficient and fast data processing | Parallel processing, optimized algorithms |

* 1. **USER STORIES**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional Requiremen t (Epic)** | **User Story Numbe**  **r** | **User Story / Task** | **Acceptance criteria** | **Priorit y** | **Team Member** |
| End User | Data Collection | USN-1 | As an end user, I want to collect  literacy data | - The system should retrieve accurate and  upto-date data | High | GOKUL |
| Data Analyst | Data Processing | USN-2 | As a data analyst, I want to clean and preprocess  data | - The system should remove inconsistencies and errors | Mediu m | PAUL |
| Data Analyst | Data Analysis | USN-3 | As a data analyst, I want to analyze literacy trends | - The  system should identify patterns and trends in literacy  rates | High | PAUL |
| End User | Visualization | USN-4 | As an end user, I want to view visualizations of literacy  data | - The system should present visual representation s of literacy rates | High | DHANAPA L |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data Analyst | Reporting | USN-5 | As a data analyst, I want to generate reports on  literacy rates | - The system should generate comprehensive reports with  insights | Mediu m | PAUL |
| Security Specialist | Security | USN-6 | As a security specialist, I want to implement data security  measures | - The system should ensure confidentiality and integrity of literacy data | High | JOHN |
| System Administrato r | Scalability | USN-7 | As a system administrator  , I want to ensure the system can handle increasing  data | - The system should be designed to handle growing volumes of literacy data | High | JUEL |
| UI/UX  Designer | Usability | USN-8 | As a UI/UX designer, I want to create a userfriendly  interface | - The system should have an intuitive and easy-touse interface | High | KARTHI |
| Data Engineer | Performance | USN-9 | As a data engineer, I want to optimize data processing | - The  system should efficiently process and analyze  data | Mediu m | DHARANI |

# CODING & SOLUTION

* 1. **FEATURE 1**

It appears that you are working on a web application project for Twitter analysis using Flask and IBM Cognos for visualization

Flask: Flask is a web framework for Python that allows you to build web applications. It provides tools, libraries, and functionalities to handle routing, request handling, and response generation. Flask is lightweight and easy to use, making it a popular choice for developing web applications in Python.

IBM Cognos: IBM Cognos is a suite of business intelligence and performance management software products. It provides powerful tools for data visualization, reporting, and analytics. In your project, you are integrating IBM Cognos to visualize the analyzed Twitter data through dashboards, stories, and reports.

Visualization: Visualization refers to the graphical representation of data to facilitate understanding and interpretation. In the context of your project, you are using IBM Cognos to create visualizations that represent the analyzed Twitter data in a meaningful and intuitive way. Visualizations can include charts, graphs, maps, and other visual elements to convey insights and patterns effectively.

1. **PERFORMANCE METRICES**

**Project Development Phase :**

|  |  |
| --- | --- |
| Date | 18 May 2023 |
| Team ID | NM2023TMID11499 |
| Project Name | Dissecting the Digital Landscape: A Comprehensive Analysis of Social Media. |

**Model Performance Testing:**

Project team shall fill the following information in the performance testing template.

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Screenshot / Values** |
|  | Dashboard design | No of Visualizations - 7 |
|  | Data Responsiveness | Data is only Responsiveness when network is strong |
| 3. | Utilization of Data Filters | Used for Story insights |
| 4. | Effective User Story | No of Scene Added - 7 |
| 5. | Descriptive Reports | No of Visulizations - 1 |

# ADVANTAGES / DISADVANTAGES

* Real-Time Insights: Twitter analysis allows you to access real- time data and gain immediate insights into trending topics, public sentiment, and user behavior.
* Large Data Volume: Twitter generates a massive amount of data every day, providing a rich source of information for analysis. This allows you to analyze a vast amount of user-generated content and extract valuable insights.
* Customer Insights: Twitter analysis enables you to understand your target audience better. By analyzing user interactions, sentiment, and preferences, you can gain valuable customer insights and tailor your marketing strategies accordingly.
* Brand Monitoring: Twitter analysis allows you to monitor mentions of your brand, products, or services in real-time. This helps you identify customer feedback, address issues promptly, and manage your brand reputation effectively.
* omer interactions. This allows you to gain insights into their strategies and identify areas where you can differentiate and improve.

# DISADVANTAGES:

* Noisy Data: Twitter data can be noisy and contain irrelevant or spammy content. Filtering and cleaning the data to remove noise and focus on meaningful information can be challenging.
* Bias and Sample Representativeness: Twitter users may not represent the entire population, and the data may be biased towards certain demographics or user groups. This can limit the generalizability of the insights obtained from Twitter analysis.
* Data Access Limitations: Twitter API limitations and restrictions on historical data access can impact the depth and scope of analysis. Access to complete and historical Twitter data may require additional resources or paid services.
* Contextual Understanding: Twitter data often lacks the full context of conversations, making it challenging to interpret the meaning behind tweets accurately. Understanding sarcasm, irony, or nuanced sentiments can be difficult solely based on the text.
* Privacy Concerns: Analyzing Twitter data raises privacy concerns as it involves analyzing publicly available user-generated content. Ensuring compliance with privacy regulations and ethical data usage is essential.

# CONCLUSION

Twitter analysis is a powerful tool for gaining insights into user behavior, sentiment, and trends on the platform. By leveraging data analytics techniques and visualization tools like IBM Cognos, you can extract valuable information from the vast amount of user-generated content on Twitter.

Through the integration of Flask and the Cognos SDK, you have created a web application that allows users to access and interact with dashboards, stories, and reports for Twitter analysis. This provides a user-friendly interface for exploring and visualizing Twitter data, enabling businesses and individuals to make data-driven decisions and gain a competitive edge.

Twitter analysis offers numerous advantages, including real-time insights, access to large data volumes, customer insights, brand monitoring, and competitive analysis. However, there are also challenges to consider, such as noisy data, bias and sample representativeness, data access limitations, contextual understanding, and privacy concerns.

# 10. FEATURE SCOPE

* + User Engagement Metrics: Calculate and display engagement metrics such as retweets, likes, and replies for individual tweets or user profiles. Provide insights into the level of engagement and popularity of specific tweets or users.
  + Visualization Dashboards: Create interactive and visually appealing dashboards using IBM Cognos or other visualization tools. Display charts, graphs, and visual representations of the analyzed Twitter data for easy interpretation and analysis.
  + Report Generation: Allow users to generate customized reports summarizing the Twitter analysis results. Include key metrics, insights, and visualizations in the reports for easy sharing and presentation.
  + Export Functionality: Enable users to export the analyzed data, visualizations, or reports in different formats such as CSV, Excel, PDF, or image files for further analysis or sharing with stakeholders.