***Social Media Analysis***

***Data Wrangling and Visualization***

***Final report***

***A screenshot of a computer

Description automatically generated***

***Presented By: Rodi Yasser-222300414, Adham Zafer-222300447***

***Ali Ayman-222300460, Yahia Fouda-222300390***

***Presented to: Amal Mahmoud***

***Table Of Contents***

Table of Contents

[Project Objectives 2](#_Toc185469304)

[**Technical Requirements** 2](#_Toc185469305)

[Data Overview 2](#_Toc185469306)

[Sources 2](#_Toc185469307)

[Structure 2](#_Toc185469308)

[Preparation 3](#_Toc185469309)

[Dashboards Overview 3](#_Toc185469310)

[Overall Performance Metrics Dashboard 3](#_Toc185469311)

[Audience Demographics Dashboard 3](#_Toc185469312)

[Engagement Insights Dashboard 3](#_Toc185469313)

[Sentiment Analysis Dashboard 3](#_Toc185469314)

[Challenges and Solutions 3](#_Toc185469315)

[Insights and Analysis 7](#_Toc185469316)

[Overview: Overall Performance Metrics Dashboard 7](#_Toc185469317)

[Overview: Audience Demographics Dashboard 9](#_Toc185469318)

[Overview: Engagement Insights Dashboard 11](#_Toc185469319)

[Overview: Sentiment Analysis Dashboard 12](#_Toc185469320)

[Recommendations 14](#_Toc185469321)

[Conclusion 14](#_Toc185469322)

[Appendices: 15](#_Toc185469323)

[16](#_Toc185469324)

[17](#_Toc185469325)

**Executive Summary:**

This project aimed to analyze social media engagement data by building four dashboards to derive meaningful insights. The dashboards include:

* **Overall Performance Metrics Dashboard**: Provides a summary of key engagement metrics.
* **Audience Demographics Dashboard**: Highlights the distribution of audience segments.
* **Engagement Insights Dashboard**: Focuses on engagement trends by time and day.
* **Sentiment Analysis Dashboard**: Analyzes sentiment trends across posts

# **Project Objectives**

* **Analyze Engagement Trends:** Examine metrics such as likes, shares, comments, and reach across platforms.
* **Identify Audience Demographics:** Determine the distribution of engagement by age, gender, and geographic location.
* **Understand Sentiment Analysis:** Evaluate audience sentiment trends and their relationship to engagement.
* **Optimize Posting Strategies:** Use time-based data to identify the best days and times for posting content.

**Technical Requirements**

* Showcase proficiency in data processing and visualization tools.
* **Python**: Perform data cleaning, transformation, and analysis using libraries such as Pandas, NumPy, and scikit-learn.
* **SQL**: Execute complex queries, manage databases, and extract data from multiple sources.
* **Tableau**: Create interactive dashboards and visually compelling narratives from processed data.

# **Data Overview**

## Sources

Dataset obtained from Kaggle containing over 100,000 rows with multiple attributes.

## Structure

The dataset was transformed into structured SQL tables to facilitate analysis and querying.

## Preparation

* Data cleaning and preprocessing were conducted in Python to handle missing values, outliers, and inconsistencies.
* SQL was used for data manipulation and table creation, with exported tables prepared for Tableau.

# **Dashboards Overview**

## Overall Performance Metrics Dashboard

* **Purpose**: Track post performance across platforms.
* **Visuals**: Engagement bar charts, top-performing posts table, filters.
* **Findings**: Likes dominate interaction metrics, while shares have the lowest participation.

## Audience Demographics Dashboard

* **Purpose**: Provide insights into audience’s age, gender, and geographic distribution.
* **Visuals**: Bubble charts, heatmaps, pie charts.
* **Findings**: Senior adults engage the most, with equal gender distribution across posts.

## Engagement Insights Dashboard

* **Purpose**: Analyze engagement trends across times, days, and weeks.
* **Visuals**: Bar charts, line charts, heatmaps.
* **Findings**: Engagement peaks at night and on Fridays.

## Sentiment Analysis Dashboard

* **Purpose**: Evaluate sentiment trends and engagement performance by sentiment category.
* **Visuals**: Stacked bar charts, bubble charts, line charts.
* **Findings**: Neutral sentiment dominates, with positive sentiment showing significant likes.

# **Challenges and Solutions**

**1. Choosing a Windows Laptop Instead of a Mac and Ensuring MySQL Server Availability**

**Challenge:**Initially, we struggled with connecting to the MySQL server on a Mac. We decided to switch to a Windows laptop, as it was simpler to get MySQL running smoothly. On Windows, we also encountered the issue of MySQL not appearing in Services and difficulty connecting to 127.0.0.1:3306.

**How We Solved It:**

* We installed or confirmed the MySQL Server installation on Windows.
* We checked the Windows Services, located the MySQL service, and started it if it was stopped.
* We verified the login credentials and confirmed the correct configuration in my.ini.

**2. Importing CSV Files Instead of SQL Scripts**

**Challenge:** MySQL initially tried interpreting the CSV file as SQL, causing syntax errors (e.g., ERROR 1064).

**How We Solved It:**

* We used the Table Import Wizard in MySQL Workbench to import CSV files correctly.
* We specified FIELDS TERMINATED BY ',', ENCLOSED BY '"', and LINES TERMINATED BY '\n' and used IGNORE 1 ROWS to skip the header.

**3. Verifying Successful Import and Row Counts**

**Challenge**:After importing, we weren’t sure if all rows were imported or if any duplicates existed.

**How We Solved It:**

* We ran SELECT COUNT(\*) FROM processeddata; to confirm the total number of rows.
* We compared this count to the expected number.
* We checked for duplicates by grouping by Post ID and using HAVING COUNT(\*) > 1.

**4. Handling a Large Dataset**

**Challenge:** Importing all 100,000 rows at once was challenging and time-consuming.

**How We Solved It:**

* We split the 100,000 rows into five smaller CSV files.
* We imported each file one by one, verifying the row count after each import to ensure the increments matched our expectations.

**5. Ensuring Data Integrity Across Multiple Imports**

**Challenge:** We needed to ensure that each subsequent CSV import appended data to the same table, increasing the row count as expected.

**How We Solved It:**

* We used the same table and import settings for each CSV file.
* After each import, we ran SELECT COUNT(\*) to confirm that the total row count increased as expected.

**6. Designing the Database Schema for Analysis**

**Challenge:** After import, we wanted to structure our database into multiple related tables (e.g., post, platform, demographics, engagement, time, date, sentiment) to facilitate more complex queries and insights.

**How We Solved It:**

* We identified the main post table and created foreign key relationships to other tables, ensuring relational integrity.
* We wrote CREATE TABLE statements for each entity and defined PRIMARY KEY and FOREIGN KEY constraints to reflect the relationships between posts, platforms, demographics, engagement metrics, time, date, and sentiment data.

**7. Resetting and Starting Fresh**

**Challenge**: At one point, we decided to start over due to confusion or data issues.

**How We Solved It:**

* We used DROP TABLE or DROP DATABASE commands to clear out old data.
* We re-imported the cleaned CSV files and followed a systematic, step-by-step approach to ensure no errors or confusion occurred this time.

A screenshot of a computer

Description automatically generated

*After SQL the data looked like this becase it was a CSV so we could not preform anythin on excel*

A screenshot of a computer

Description automatically generated

*Congo Was Empty Beacse we had it in our dataset as congo, but in tableau*

# **Insights and Analysis**

## Overview: Overall Performance Metrics Dashboard

The Overall Performance Metrics Dashboard provides a comprehensive view of key engagement metrics across all platforms and post types. It highlights critical performance indicators such as total shares, comments, likes, average engagement rate, and total reach. Additionally, it identifies the top-performing posts by engagement rate, allowing for a clear understanding of content performance.

***The dashboard is organized into three main sections:***

* Key Metrics Summary: Displays aggregated totals for Shares, Comments, Likes, Engagement Rate, and Reach.
* Engagement Metrics Across Platforms: Visualizes total Reach, Likes, Comments, and Shares segmented by post types (Image, Link, Video).
* Top Performing Posts: A detailed table showcasing posts with the highest engagement rates, including their respective engagement metrics.

***Insights and Analysis***

**Engagement Distribution:**

* Reach significantly outperformed other metrics with 248.6M recorded, indicating a wide audience exposure across platforms.
* Likes (49.9M) dominate the interaction metrics, while Comments (25M) and Shares (10M) show relatively lower participation.
* The Average Engagement Rate (47%) provides a baseline to evaluate content success across platforms.

**Performance by Post Type:**

* Links generated the highest engagement for Likes with 16.7M, followed closely by Images and Videos.
* Images and Links also contributed significantly to Comments, while Shares remained consistent across post types at 3.3M.
* This suggests that content with visual or clickable links engages the audience more effectively.

**Top-Performing Posts:**

The table on the right lists posts ranked by engagement rates. Posts such as "Expert final position money weight I like" and "Father concern remember but certainly environment now" demonstrate high engagement, suggesting a focus on topics that resonate emotionally or contextually with the audience.

Posts with bolded or concise content tend to rank higher, possibly due to clarity and relatability.

***Overall Trends:***

* The dominance of Likes indicates that users prefer easy and quick interactions.
* A relatively lower engagement rate for Shares could indicate a need for creating more shareable or viral content.

## Overview: Audience Demographics Dashboard

The Audience Demographics Dashboard provides valuable insights into the audience's age, gender, and geographic distribution. It helps identify the most engaged audience segments and regions, offering a detailed view of where posts are most successful and who is interacting with the content.

***The dashboard is organized into the following key sections:***

* Top 5 Countries With the Most Number of Posts: A bubble chart highlighting countries with the highest post activity.
* Global Engagement Rate Map: A geographical visualization showing the average engagement rate across different countries.
* Audience Demographics: A heatmap table displaying engagement metrics segmented by age group and gender.
* Posts by Gender and Age Group: Two pie charts providing a breakdown of posts by gender and age group.

***Insights and Analysis***

**Top 5 Countries for Post Activity:**

* Congo and Korea lead in post activity, followed closely by Dominica, Uzbekistan, and Kuwait.
* These countries represent key audience bases and potential regions to focus on for targeted campaigns.

**Global Audience Engagement:**

* Engagement is highly concentrated in specific regions, as shown on the map. Africa and Asia exhibit significant engagement, with countries like Congo, Korea, and regions in Northern Africa standing out.
* Notably, there are areas marked as "unknown", highlighting the need to ensure all data is accurately captured for analysis.
* Audience Demographics by Age and Gender:
* Senior Adults exhibit the highest engagement rates across all genders, with 47.731% for "Other" and 47.106% for "Male".
* Mature Adults and Adolescent Adults show fairly balanced engagement rates, with Male and Other consistently being the top categories in their respective groups.
* This indicates that older audience segments are more likely to engage with the content.

**Posts by Gender:**

Post activity is evenly distributed across Male (33K), Female (33K), and Other (33K) genders, suggesting a balanced gender audience.

**Posts by Age Group:**

Adolescent Adults (42K posts) represent the largest segment, followed by Mature Adults (31K) and Senior Adults (27K).

This reinforces the importance of tailoring content that appeals to younger demographics, as they contribute to a larger share of the posts.

**Recommendations**

**Regional Focus:**

* Focus campaigns and content on high-performing countries like Congo, Korea, and Dominica, while investigating reasons for engagement success in these regions.
* Work to address gaps in the "unknown" geographic data to ensure full representation.

**Audience Segmentation:**

* Prioritize content that resonates with Senior Adults, as they have the highest engagement rates.
* Design campaigns targeted toward Adolescent Adults to leverage their strong post activity.

**Gender Balance:**

Maintain a balanced approach to content creation, as posts show equal participation across genders.

**Content Localization:**

Develop region-specific content tailored for the top-performing countries to maximize engagement and interaction.

## Overview: Engagement Insights Dashboard

The Engagement Insights Dashboard provides a detailed analysis of engagement trends across different times, days, and weeks. It helps pinpoint the most effective posting times to maximize audience interaction.

***The dashboard is organized into the following key sections:***

* Key metrics for Highest Engagement Time, Top Day of Engagement, and Average Engagement Rate.
* Engagement by Time of Day
* A bar chart breaking down engagement metrics (likes, comments, and shares) by specific time periods:
* Morning, Afternoon, Evening, and Night.
* Engagement Metrics by Day of the Week
* A line chart showing the engagement trends for each day of the week, identifying consistent patterns and peak days.
* Monthly Heatmap of Engagement Rate by Time of Day
* A heatmap visualization that combines time of day and month to illustrate when engagement spikes occur.

***Insights and Analysis***

**Highest Engagement Time:**

Night records the highest engagement, with 18.6M likes and notable comment activity.

**Top Day for Engagement:**

Friday leads in engagement across all metrics.

**Time-Based Trends:**

Engagement consistently rises at night and afternoon, while morning and evening show moderate activity.

**Day of the Week Trends:**

Engagement remains steady throughout the week, peaking on Friday with higher likes.

**Monthly Heatmap:**

Patterns show a gradual rise in engagement over time, especially during nights and afternoons in specific months.

**Key Takeaways:**

* Focus on nighttime and Fridays for posting to optimize engagement.
* Monitor heatmap trends to identify peak months for targeted campaigns.

## Overview: Sentiment Analysis Dashboard

The Sentiment Analysis Dashboard offers a comprehensive evaluation of audience sentiment across different post types, platforms, and time trends. It helps uncover the distribution of positive, neutral, and negative sentiments and highlights engagement performance for each sentiment category.

***The dashboard is organized into the following key sections:***

* Engagement Rate by Sentiment Category and Post Type
* A stacked bar chart showing engagement (likes, comments, and shares) segmented by Negative, Neutral, and Positive sentiments.
* Comparison of sentiment categories across post types to identify performance.
* Sentiment Distribution
* A bubble chart visualizing the proportional breakdown of sentiment:
* Neutral (50.53%), Positive (35.26%), and Negative (14.21%).
* Sentiment Trends Over Time
* A line chart tracking sentiment trends from February 2021 to February 2024, revealing changes in post sentiment.
* Highlights any irregularities, such as the sharp drop in early 2024 due to fewer posts.
* Sentiment Summary and Filters
* A sentiment summary displaying the percentage distribution of each sentiment.
* Filters for Post Type, Platform Name, and Sentiment Category for interactive exploration.
* Count of total posts analyzed (14,213 to 50,531 posts).

***Insights and Analysis***

**Sentiment Distribution:**

* Neutral sentiment dominates, accounting for 50.53%, followed by Positive (35.26%) and Negative (14.21%).
* Positive sentiment still contributes significantly, while negative sentiment is the smallest share.

**Engagement by Sentiment Category:**

* Neutral sentiment posts generate the highest engagement, with a consistent mix of likes, comments, and shares.
* Positive sentiment shows strong engagement performance, particularly with likes.
* Negative sentiment has the lowest engagement, highlighting its limited interaction.

**Sentiment Trends Over Time:**

* Neutral sentiment consistently trends the highest over time, reflecting audience preference for balanced or non-polarizing content.
* Positive sentiment maintains steady engagement levels, with occasional fluctuations.
* Negative sentiment remains relatively low but stable.
* A sharp drop in early 2024 indicates a lower volume of posts, likely affecting engagement and trend visibility.

**Key Takeaways:**

* Neutral sentiment drives the most engagement, making it ideal for maintaining audience balance and interest.
* Positive sentiment posts can enhance engagement, particularly for likes-driven content.
* Regularly monitor sentiment trends to identify fluctuations and adjust content strategies.

# **Recommendations**

1. Focus campaigns on high-performing regions like Congo, Korea, and Dominica.
2. Prioritize content targeting senior adults and adolescent adults.
3. Post during nighttime and Fridays to maximize engagement.
4. Develop content with neutral and positive sentiments to sustain interaction levels.

# **Conclusion**

The project successfully demonstrated the use of data processing and visualization tools to analyze engagement patterns and audience behavior. The insights derived can guide content strategies and improve performance metrics.

# **Appendices:**

A screenshot of a computer

Description automatically generated

*Dashboard 1: Overall Performance Metrics*

A screenshot of a computer screen

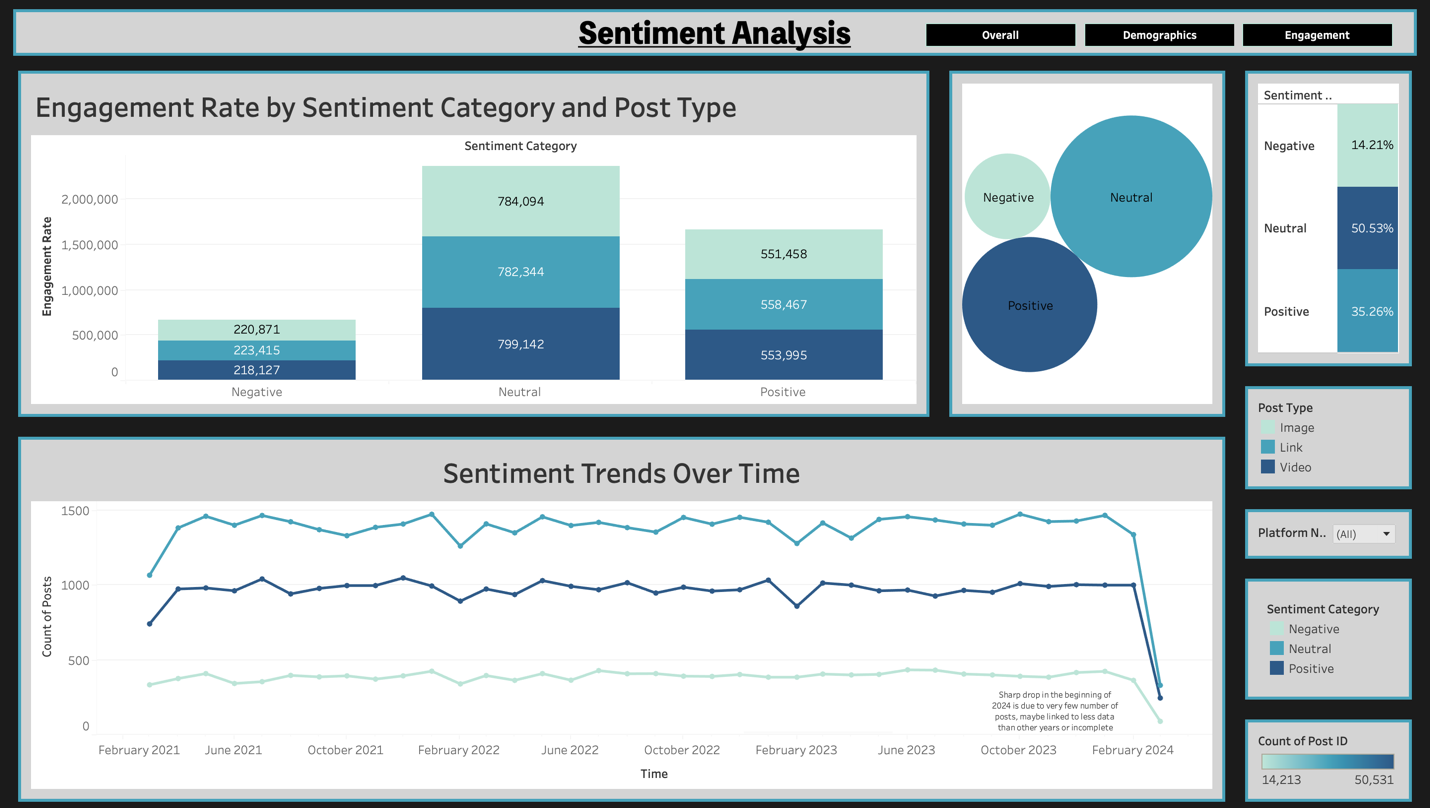
Description automatically generated

*Dashboard 2: Audience Demoghraphics Dashboard*

A screenshot of a graph

Description automatically generated

*Dashborad 3: Engagment Insights DashBoard*

******

*Dashborad 4: Sentiment Analysis DashBoard*

***Thank you!***