**Report on Social Media Analysis Project**

**Introduction**

This project analyses social media engagement data to derive actionable insights into audience behaviour, content performance, and platform-specific trends. The dataset from Kaggle includes metrics such as likes, comments, shares, and impressions across multiple social media platforms. The project showcases the power of data-driven decision-making by using Python for data wrangling, SQL for database normalisation, and Tableau for data visualisation.

**Important Links:**

**Dataset:**

<https://www.kaggle.com/datasets/aliredaelblgihy/social-media-engagement-report>

**Github:**

<https://github.com/noorhatem123/Social-Media-Analysis>

**Project Workflow**

**Step 1: Data Collection and Tools Used**

* **Dataset Source**: The dataset was sourced from Kaggle and downloaded as a CSV file.
* **Tools Used**:
  + **Python**: Data cleaning, wrangling, exploratory data analysis, and splitting the dataset.
  + **Pandas**: Data manipulation and cleaning.
  + **NumPy**: Numerical operations.
  + **Seaborn & Matplotlib**: Data visualizations.
  + **MySQL**: SQL queries for database creation and management.
  + **PHPMyAdmin**: Database management and data uploads.
  + **Tableau**: Interactive dashboard creation and visualization.
  + **GitHub**: Version control and project repository ([GitHub Link](https://github.com/noorhatem123/Social-Media-Analysis)).
  + **Google Colab**: Python development environment.
* **Step 2: Data Wrangling (Python)**

Team Members: *Ali Sherif*  
Assisted by: *Noor Hatem & Natalie Emad*

**1. Importing the Dataset**

* The dataset was loaded into a Pandas DataFrame for exploration.
* Dataset structure and summary statistics were examined.

**2. Data Cleaning**

* **Handling Missing Values**:
  + Filled missing values in the Sentiment column using the mode.
  + Mapped Sentiment values (Positive, Neutral, Negative) to numeric representations.
  + Dropped columns (Campaign ID, Influencer ID) with high missing values.
* **Fixing Data Types**:
  + Ensured Post Timestamp was correctly parsed as a datetime field.
  + Removed invalid timestamp rows.

**3. Feature Engineering**

* Extracted time-based features such as Hour, Minute, and Time Range.

**4. Outlier Detection and Removal**

* Used the Interquartile Range (IQR) method to remove outliers in numerical columns.

**5. Exploratory Data Analysis (EDA)**

Visualizations were generated to explore engagement trends:

* **Correlation Heatmap**:
  + Visualized correlations between numerical features.
* **Engagement Rate Distributions**:
  + Distribution of engagement rates across platforms, time ranges, and post types.

**6. Dataset Normalization**

The dataset was split into normalized tables:

1. sentiment: Contains sentiment mappings.
2. platforms: List of platforms.
3. audience\_demographics: Information about audience gender, age, and location.
4. post\_details: Metadata for each post.
5. engagement\_metrics: Detailed engagement statistics.

**7. Data Export**

* The cleaned and normalized tables were exported as CSV files for SQL database import.

**Step 3: Database Creation (SQL)**

Team Members: *Natalie Emad*  
Assisted by: *Noor Hatem & Ali Sherif*

**1. Database Design**

* Created a normalized database social\_media\_analytics with the following tables:
  + sentiment
  + platforms
  + audience\_demographics
  + post\_details
  + engagement\_metrics

**2. SQL Table Creation**

* Foreign key constraints were added to establish relationships between tables.

**3. Data Import**

* Imported the normalized CSV files into phpMyAdmin.

**Step 4: Data Visualization (Tableau)**

Team Member: *Noor Hatem*  
Assisted by: *Ali Sherif & Natalie Emad*

**1. Establishing Relationships**

* Relationships between tables were defined in Tableau based on foreign keys.

**2. Visualizations**

* **Platform-Specific Insights**:
  + Engagement metrics by platform.
  + Sentiment analysis across platforms.
* **Time-Based Analysis**:
  + Engagement trends by time of day.
  + Sentiment trends over time.
* **Demographic Analysis**:
  + Audience distribution by gender, age, and location.
* **Content Performance**:
  + Engagement by post type.
  + Correlation between sentiment and engagement.

**3. Dashboard Design**

* Designed interactive dashboards:
  + Home page with navigation icons for themes.
  + Filters for platform, time range, and audience demographics.
  + Hamburger menu for easy navigation between dashboards.

**Team Contributions**

* **Noor Hatem**:
  + Designed Tableau dashboards.
  + Assisted with Python and SQL code.
* **Natalie Emad:**
  + Created and normalized the SQL database.
  + Assisted with Tableau visualization and Python Code.
* **Ali Sherif**:
  + Performed data wrangling in Python.
  + Assisted with Tableau visualization and Python Code.
* **Fares Al Habashi**:
  + Prepared the final report.

**Key Insights**

**1. Platform-Specific Trends:**

* **Engagement Metrics:**
  + **Facebook** and **Twitter** exhibit the highest total engagements, with Facebook achieving notable figures in **likes** (4,362,669) and Twitter excelling in **comments** (2,273,241).
  + **Instagram** and **LinkedIn** maintain balanced engagement across all metrics, making them consistent performers.
* **Recommendation:**
  + Focus more resources on content tailored for Facebook and Twitter to maximize audience interaction.
* **Sentiment Analysis:**
  + Positive sentiments are predominant across all platforms, with **Twitter** showing the highest number of positive sentiments (6,211), followed by LinkedIn (6,176).
  + Negative sentiments are relatively low, indicating positive reception across platforms.

**2. Time-Based Trends:**

* **Engagement Trends Over Time:**
  + Engagement rates peak during **late afternoons and evenings** (4 PM to 10 PM), suggesting these are optimal times for posting content.
  + Negative sentiments show slight spikes during early morning hours.

**3. Demographic Analysis:**

* **Gender Distribution:**
  + Engagement is almost evenly distributed among **male (12,361)** and **female (12,305)** audiences, with a minor edge toward male participants.
* **Age Groups:**
  + Audiences aged **20 to 30 years old** represent the most active engagement demographic, indicating young adults are the primary target audience.
* **Geographic Trends:**
  + The highest engagement comes from **North America** and **Western Europe**, with significant contributions from countries like the **United States**, **Canada**, and **Germany**.

**4. Content Performance:**

* **Post Types:**
  + **Image posts** generate the highest engagement, with 5,933,196 likes and 2,999,545 comments, followed closely by **link posts** and **video posts**.
  + Videos drive consistent engagement across all metrics, making them highly effective for sustained attention.
* **Sentiment and Engagement Correlation:**
  + Posts with **positive sentiments** (11,862,941 likes) outperform others, indicating audiences are more responsive to uplifting or positive content.

**Conclusion**

This project successfully combined Python, SQL, and Tableau to analyze social media engagement data. By cleaning, normalizing, and visualizing the dataset, we derived actionable insights for optimizing social media strategies. The project deliverables include a cleaned dataset, normalized database, and interactive Tableau dashboards. The complete code and assets are hosted on GitHub.