Social Media Analysis

**Project Title: Social Media Post Engagement Analysis for Influencer Collaborations**

**Objective:**

The goal of this project is to analyze the latest social media posts from 2000 influencers who were provided with a collaboration opportunity by our media agency. These influencers, representing various niches such as fashion, food, tech, and sports, posted content on their profiles in different formats (images, videos, text-based posts). By assessing their engagement rates (likes, comments, and shares), we aim to identify high-performing influencers and determine the most effective locations for targeting future collaborations.

**Story:**

As a media agency offering collaboration opportunities, we provided 2000 users with the chance to promote a versatile product that aligns with various niches, including fashion, food, tech, and sports. These influencers posted about the product in the form of images, videos, and text-based posts on their profiles. Our analysis focuses on their latest posts to evaluate which influencers generate the highest engagement and identify the geographical locations with the most potential for collaboration.

By analyzing engagement metrics (likes, comments, and shares) across different content types (text, image, video), we can assess the potential impact of each influencer. This analysis helps us identify influencers who are most likely to drive brand awareness, understand which content types work best for different audiences, and pinpoint locations that yield higher engagement, enabling the agency to optimize future collaborations and location-based campaigns.

**Why Analyze This Data?**

1. **Maximizing Engagement:**
   * Social media platforms thrive on engagement metrics like likes, comments, and shares. By examining these metrics, we can determine the factors that drive higher interaction rates and tailor content strategies accordingly.
2. **Understanding Content Preferences:**
   * The dataset includes post\_type (e.g., text, image, video) and post\_category (e.g., fashion, food, sports). Analyzing these fields helps identify which content formats and topics resonate most with the audience.
3. **Optimizing Posting Time:**
   * With data on post\_hour and is\_weekend, we can uncover patterns of user activity, such as peak engagement times, enabling businesses to post content when it’s most likely to gain traction.
4. **Audience Behavior Insights:**
   * Including user demographics like gender, age, and location would enhance the analysis, revealing how different user groups interact with posts and what they value in content.
5. **Spam Identification and Content Quality:**
   * The spam\_flag column helps assess the prevalence of low-quality or irrelevant content. This is vital for maintaining the credibility and effectiveness of a social media strategy.
6. **Performance of Verified Accounts:**
   * The user\_verified column provides an opportunity to compare how verified and non-verified accounts perform. This is crucial for understanding the impact of credibility and trust on engagement.

**What Are We Analyzing?**

1. **Engagement Patterns:**
   * Key metric: engagement\_rate to determine the effectiveness of posts in sparking user interaction.
2. **Post Type Effectiveness:**
   * Do videos, images, or text posts generate the most engagement?
3. **Category Trends:**
   * Which topics (fashion, sports, technology, etc.) are most popular, and how does that vary by audience demographic?
4. **Optimal Post Timing:**
   * Is there a significant difference in engagement between weekday vs. weekend posts or between morning, afternoon, and evening?
5. **User Influence:**
   * How do factors like user\_followers and user\_verified status impact engagement?
6. **Spam and Low-Quality Content:**
   * How prevalent is spam content, and how does it affect overall engagement metrics?

**Understanding the Columns in the Dataset**

* **User ID:** Unique identifier for each user.
* **Post ID:** Unique identifier for each post analyzed.
* **Post Type:** Specifies the type of content (image, video, carousel).
* **Post Length:** Measured as the length of captions or text content accompanying the post.
* **Likes, Comments, Shares:** Core engagement metrics indicating audience interaction.
* **Engagement Rate:** Calculated metric showing interaction relative to followers.
* **Demographics (Added Manually):**
  + **Gender:** Categorized as Male, Female, or Other.
  + **Age Group:** Ranges like 18-24, 25-34, etc.
  + **Location:** General geographic regions (e.g., US, Europe).

**Key Questions the Analysis Aims to Answer**

1. What types of posts generate the highest engagement?
2. How do engagement rates vary across demographics?
3. Are there specific trends tied to the type or length of content?
4. What strategies can improve user engagement based on findings?

**Project Workflow**

1. **Data Loading and Transformation:**
   * Import the dataset into Power BI.
   * Add demographic data manually using conditional logic.
2. **Data Cleaning:**
   * Remove duplicates and irrelevant entries.
   * Handle missing values, especially in engagement metrics.
3. **Data Visualization (Dashboard Blueprint):**
   * **Page 1:** Basic Insights
   * **Page 2:** Advanced Insights
   * **Page 3:** Report of user
   * **Page 4:** Final insights
4. **Final Deliverables:**
   * A polished Power BI dashboard with actionable insights.
   * A report summarizing findings and recommendations for content strategy.