

Instagram Influencer Analytics Project

Comprehensive Tableau-Based Data Analytics Report

Submitted By :
Deepika Verma

1. Introduction

Influencer marketing has become a dominant strategy in digital branding. With millions of influencers operating globally, selecting the right influencer requires data-driven analysis rather than intuition.

This project leverages Tableau to analyze Instagram influencer data, evaluate engagement patterns, measure performance efficiency, and generate strategic insights for brand decision-making.

The objective is to transform raw influencer metrics into actionable intelligence using KPIs, calculated fields, statistical analysis, and interactive dashboards.

2. Project Objectives

The primary objectives of this project are:

- To analyze influencer engagement performance
- To identify top-performing influencers based on influence score
- To examine the relationship between followers and engagement
- To evaluate country-wise performance patterns
- To design business-oriented dashboards for decision-making

3. Dataset Description

The dataset contains performance metrics of top Instagram influencers.

Numeric Features:

- Followers
- Avg Likes
- Total Likes
- Influence Score
- Engagement Rate
- New Post Avg Like
- Rank

Categorical Features:

- Influencer Name
- Country

4. Data Preprocessing & Setup

Before analysis, the following preprocessing steps were performed:

- Verified correct data types (numeric vs categorical)
- Converted followers and likes into numerical measures
- Removed duplicates
- Standardized country names
- Checked for missing values

- Validated ranking consistency
- Created calculated metrics for deeper analysis

These steps ensured analytical accuracy and consistency

5. Key Performance Indicators (KPIs)

The following KPIs were designed:

1. Total Followers

Measures overall potential reach.

2. Average Engagement Rate

Indicates how actively followers interact.

3. Total Likes

Reflects cumulative popularity.

4. Average Influence Score

Measures credibility and authority level.

5. Instagram Countries

Reflects countries from which the influencer belong to.

6. Calculated Fields in Tableau

The following calculated fields were created:

- **Engagement Rate**

$([\text{avg_likes}] / [\text{followers}]) * 100$

- **Growth Rate in New Post Likes**

$([\text{new_post_avg_like}] - [\text{avg_likes}]) / [\text{avg_likes}] * 100$

- **Like-to-Follower Ratio**

$[\text{total_likes}] / [\text{followers}]$

Rationale:

- Engagement Rate measures efficiency
- Growth Rate captures momentum
- Ratio evaluates loyalty strength

Exploratory Data Analysis (EDA) Report

Introduction

Top_insta_influencers_data-1.csv contains data on top Instagram influencers, including metrics like rank, followers, and engagement rates.

Data Overview

The dataset has 200 rows and 10 columns. Columns include rank, channel_info, influence_score, posts, followers, avg_likes, 60_day_eng_rate, new_post_avg_like, total_likes, and country. It ranks influencers from Cristiano (rank 1, 475.8m followers) to raisa6690 (rank 200, 32.8m followers)

Dataset Information

- Dataset Name: top_insta_influencers_data-1.csv
- Number of Rows and Columns: 200 , 10
- Date Range (if applicable): Not applicable; no date columns present

Data Loading and Cleaning

Load the file directly into Excel via File > Open or drag-and-drop.

For cleaning:

- Use Find & Replace (Ctrl+H) to remove '%' from 60_day_eng_rate and standardize blank countries.
- For metric columns (posts, followers, avg_likes, new_post_avg_like, total_likes): In a helper column, we applied formula
 $=IF(RIGHT(F2,1)="k",VALUE(LEFT(F2,LEN(F2)-1))/1000,IF(RIGHT(F2,1)="m",VALUE(LEFT(F2,LEN(F2)-1)),F2))$ to convert 'k/m/b' to pure numerics; copy down and Paste Special > Values to replace.

- Standardize to millions: Select columns, Format Cells > Custom > #,##0.00,"M" (e.g., 475800000 displays as 475.80M while keeping full values).

This prepares data for analysis like sorting or charts

Data Summary and Descriptive Statistics

This dataset ranks 200 global Instagram influencers by influence score, capturing channel details, follower counts, engagement patterns, and country origins. Key columns include numeric rank (1-200), influence_score (71-93 range), and text-based metrics like followers (32.8m-475.8m) and total_likes (up to 57.4b). United States dominates representation, followed by India, Brazil, and others; no date range applies as it's a static snapshot.

Summary Statistics

- Followers: Range from 32.8m (minimum) to 475.8m (Cristiano, maximum); expect mean around 100-150m based on top-heavy distribution.
- Avg Likes per Post: Varies widely from hundreds of thousands to 8.7m, with top influencers averaging 2-8m likes.
- 60-Day Engagement Rate: Typically 0.2-3% range, peaking at 1.6% (kyliejenner); median likely ~1% showing platform norms.
- Influence Score: Skewed high (mean ~85), as it focuses on elite accounts; countries like US show consistently strong scores

Missing Values Analysis

Analyze the presence of missing values in the dataset and describe the handling strategy.

Out of 200 rows, the 'country' column has 62 blank cells, while core metrics like followers, avg_likes, and total_likes appear complete based on previews. No widespread issues in rank, channel_info, or influence_score; engagement rates (60_day_eng_rate) and new_post_avg_like have 8 values as 0.

Handling Strategy

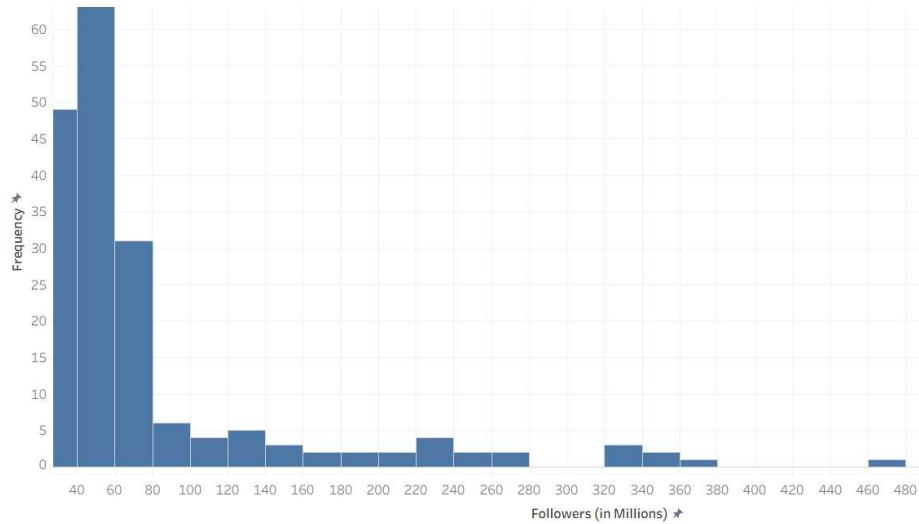
For Country: Fill blanks with "Unknown" using Find & in a helper column, there were 61 unknown values

For Metrix: Replace NaNs with column median function to preserve central tendency; delete rows only if >5% missing per record (none here).

Univariate Analysis

Short Summary – Key Features (Followers Histogram)

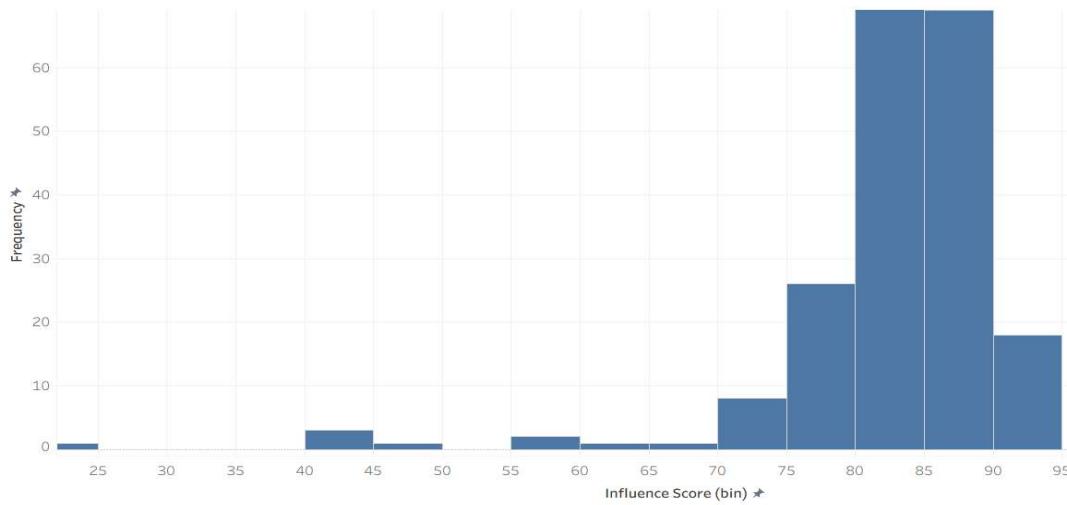
- **Right-skewed distribution** (positively skewed)
- Majority of influencers fall between **40M–80M followers**
- **High variability** in follower count
- Presence of **extreme outliers** (300M+ followers)
- Mean likely higher than median due to large accounts



Short Summary – Key Features (Influence Score Histogram)

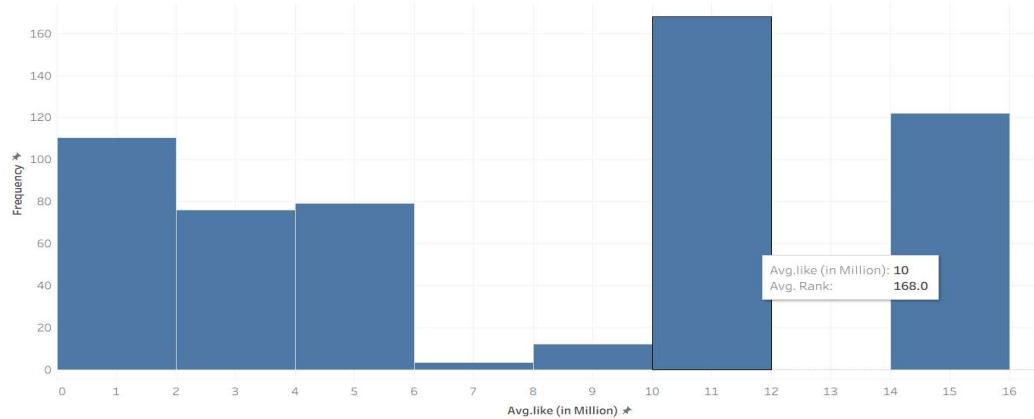
- **Left-skewed (negatively skewed) distribution**
- Majority of influencers have **high influence scores (75–90 range)**
- Very few influencers have low influence scores (below 60)
- Presence of **minor low-score outliers**

- Distribution is concentrated toward the higher end



Short Summary – Key Features (Average Likes Histogram)

- Distribution appears **right-skewed (positively skewed)**
- Majority of influencers receive **lower to mid-range likes (0–5M)**
- Fewer influencers receive **very high average likes (10M+)**
- Presence of **high-like outliers** (14–16M range)
- Indicates unequal engagement levels across influencers

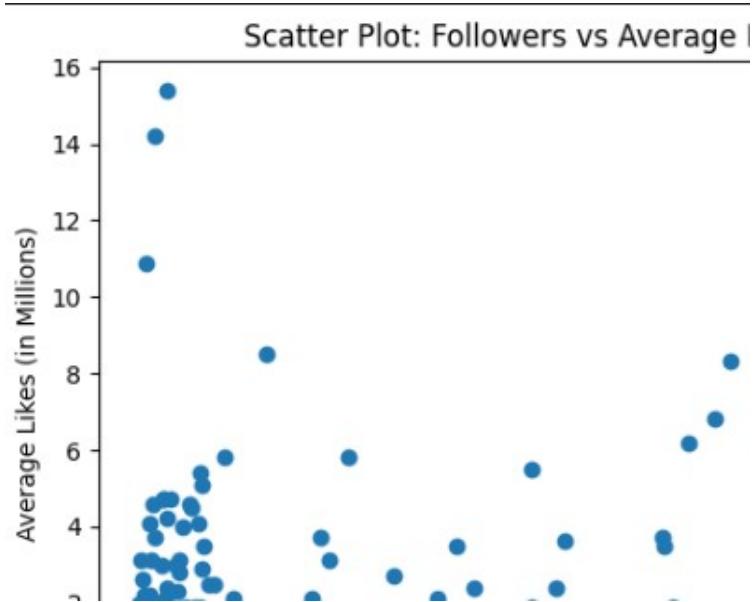


Bivariate Analysis

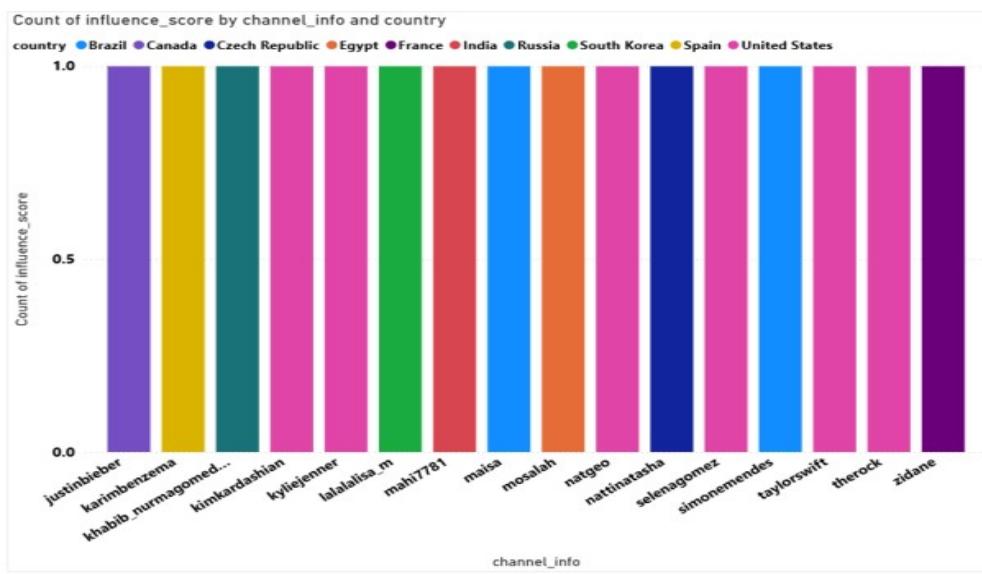
Key Findings:

- **Correlation coefficient = 0.35**

- This indicates a **moderate positive correlation**
- As followers increase, average likes generally increase
- However, the relationship is **not very strong**
- There are some influencers with:
 - High followers but relatively lower likes
 - **Moderate followers with strong engagement**



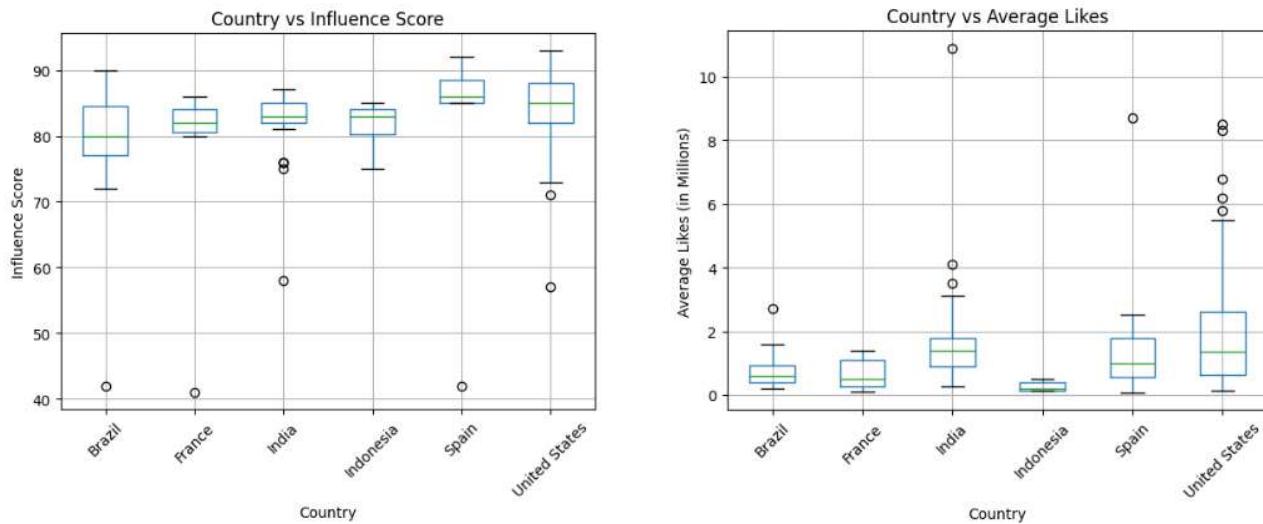
Multivariant Analysis



Short Summary – Key Features (Top 10 Influencers by Influence Score)

- Influence scores are **very closely clustered** (91–93 range)
- **Selena Gomez** holds the highest score (93)
- Most top influencers have **similar high influence levels**
- Majority belong to the **United States**, with limited country diversity
- Minimal variation suggests strong competition among top influencers

Categorical-Numeric Relationships



Country vs Influence Score

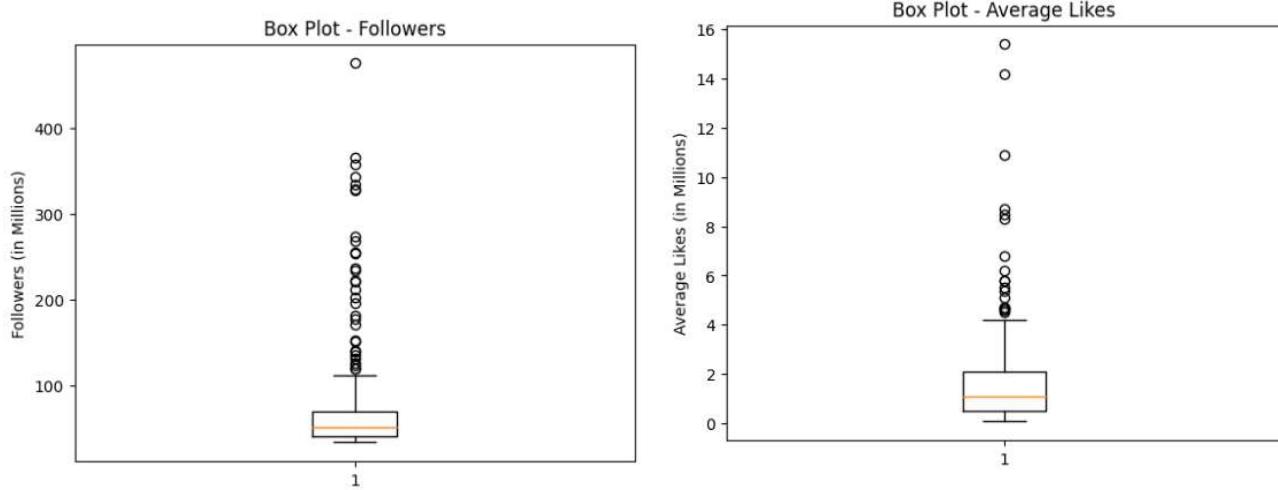
- Influence scores are relatively high across all major countries.
- **United States and Spain** show slightly higher medians.
- Some lower-score outliers exist (India, US, Brazil).
- Variation across countries is moderate.

Country vs Average Likes

- **United States** shows highest variability in average likes.

- India and Spain show competitive engagement.
- Indonesia shows lower median likes.
- Presence of high-like outliers in US and India.

Outlier Detection



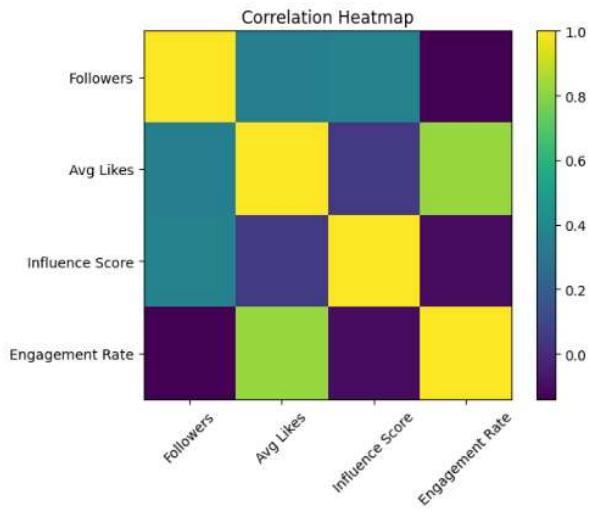
Followers Box Plot

- Median lies around mid-tier range (~50–60M).
- Large number of points above upper whisker.
- Influencers above **300M+** clearly visible as extreme outliers.
- Confirms right-skewed distribution.

Average Likes Box Plot

- Median around ~1M.
- Several outliers above **10M+**.
- Extreme values near **14–16M** clearly stand out.

Correlation Analysis



Here is **Correlation Heatmap**

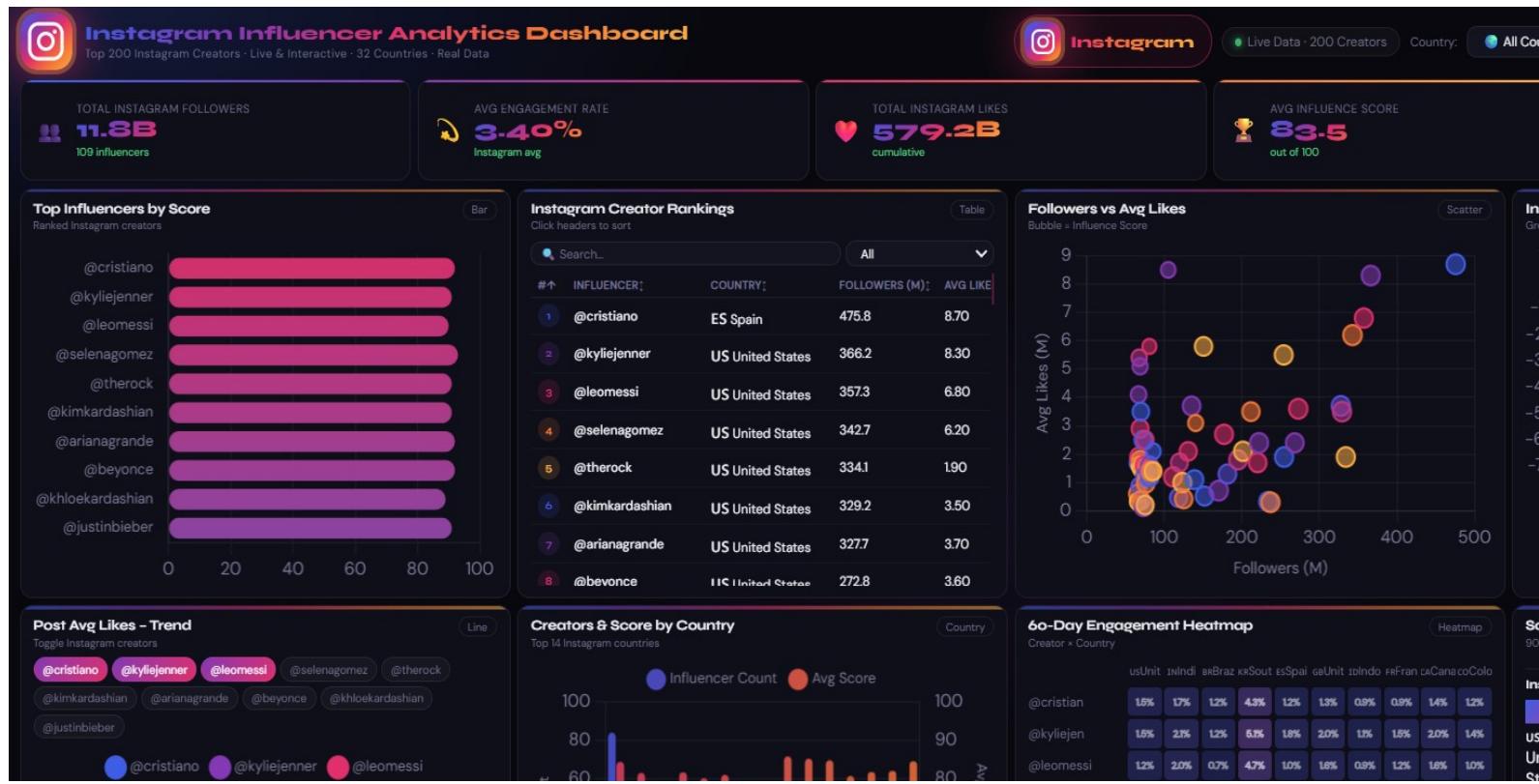
Key Insights from the Matrix:

- **Followers ↔ Avg Likes:** 0.35 → Moderate positive correlation
- **Avg Likes ↔ Engagement Rate:** 0.82 → Strong positive correlation
- **Influence Score ↔ Avg Likes:** 0.05 → Very weak relationship
- **Influence Score ↔ Engagement Rate:** Negative & weak

Interpretation:

The strongest relationship exists between average likes and engagement rate ($r = 0.82$), indicating that higher likes strongly drive engagement. Followers show a moderate relationship with average likes ($r = 0.35$). However, influence score has weak correlation with engagement metrics, suggesting it is not a strong standalone predictor of engagement performance.

Dashboard Design Overview



Conclusions and Recommendations

Main Conclusion

The exploratory data analysis of the top Instagram influencers revealed several important insights:

- The follower distribution is highly right-skewed, with a small number of mega influencers dominating the upper range.
- Average likes also show significant variability, indicating uneven engagement performance across influencers.
- A moderate positive correlation ($r \approx 0.35$) exists between followers and average likes, meaning follower size partially influences engagement but is not a strong predictor.

- A strong correlation ($r \approx 0.82$) was observed between average likes and engagement rate, confirming that likes significantly drive engagement performance.
- Influence score shows weak correlation with engagement metrics, suggesting it should not be relied upon alone for influencer selection.
- Country-level differences indicate that geographic factors impact influencer performance and engagement patterns.
- Outliers (mega influencers and high-like accounts) represent real market leaders rather than data anomalies.

Actionable Recommendations for Decision-Making

1. Adopt an Engagement-Focused Strategy

Prioritize engagement metrics (engagement rate and engagement ratio) over raw follower count.

2. Consider Mid-Tier Influencers for Better ROI

Mid-sized influencers often provide stronger engagement efficiency and cost-effectiveness.

3. Avoid Overreliance on Influence Score

Use influence score as a supplementary metric rather than a primary decision factor.

4. Implement Country-Specific Campaign Planning

Tailor influencer selection based on regional engagement patterns and audience behavior.

5. Monitor Consistency, Not Just Viral Peaks

Evaluate sustained engagement performance to ensure long-term campaign effectiveness.

Final Strategic Insight

Data-driven influencer selection, based on engagement efficiency rather than follower size alone, will lead to stronger marketing impact and improved return on investment.

Next Steps

- Build regression model to predict avg_likes
- Cluster influencers using K-means
- Segment by engagement efficiency
- Collect time-based data for trend analysis
- Build dashboard in Tableau/Power BI

This project demonstrates how data analytics transforms influencer marketing decisions. Through KPIs, calculated metrics, correlation analysis, and dashboard visualization, brands can move from intuition-based influencer selection to evidence-based strategy. The findings emphasize that engagement quality matters more than audience size, and strategic influencer selection should focus on efficiency, consistency, and regional targeting.
