Top Instagram Influencers Analysis

SQL & Tableau Data Analytics Project Report

Author: Harsh Belekar
Tools Used: Tableau, SQL, Python (Pandas, NumPy), Jupyter Notebook
Dataset: instagram_influencers_raw.csv

---

1. Executive Summary

The **Top Instagram Influencers Analysis** project aims to analyze and evaluate the performance, engagement, reach, and influence of top Instagram influencers worldwide.

Using **PostgreSQL for data analysis** and **Tableau for interactive dashboards**, this project explores:

- Influencer performance metrics
- Engagement trends
- Growth potential
- Country-wise influence distribution
- Like-to-follower efficiency

The analysis helps understand how influencers perform beyond just follower count and identifies high-impact creators based on engagement and influence score.

---

2. Project Objectives

The primary objectives of this project are:

- Analyze distribution of followers among top influencers
- Identify top influencers based on influence score
- Evaluate engagement rates across creators
- Detect influencers with high engagement but relatively low followers
- Assess growth potential using recent post performance
- Compare influence score across countries
- Build interactive dashboards for business insights

---

3. Dataset Overview

The dataset contains cleaned Instagram influencer data with the following attributes:

Column Name	Description
rank	Rank of the influencer
channel_info	Influencer username/handle
influence_score	Overall influence rating
posts	Total number of posts
followers	Total followers count
avg_likes	Average likes per post
60_day_eng_rate	Engagement rate over last 60 days
new_post_avg_like	Average likes on recent posts
total_likes	Total lifetime likes
country	Country of influencer

---

4. Tools & Technologies Used

- **PostgreSQL** – Data storage and SQL analysis
- **SQL** – Data querying and analytical insights
- **Tableau** – Data visualization and dashboard creation
- **Jupyter Notebook** – Data cleaning and preprocessing
- **Git & GitHub** – Version control and project documentation

---

5. Database Setup (PostgreSQL)

The project includes:

- Database schema creation
- Data type optimization
- Data import using COPY command
- Data cleaning adjustments (handling decimal values in numeric columns)

Tables were structured using appropriate numeric data types such as:

- BIGINT for large counts (followers, total_likes)
- NUMERIC for engagement rates
- INTEGER for rank and posts

---

6. Exploratory Data Analysis (SQL Insights)

6.1 Distribution of Followers

- Follower count is highly skewed.
- A small number of influencers dominate with billions of followers.
- Majority fall into lower follower brackets.

6.2 Top Influencers by Influence Score

- Influence scores are concentrated between 75–90.
- Only a few influencers score above 90.
- Influence score does not always correlate directly with follower count.

6.3 Engagement Rate Analysis

- High follower count does not guarantee high engagement.
- Some mid-tier influencers show exceptionally strong engagement rates.

6.4 High Engagement but Low Followers

- Several influencers with moderate followers have superior engagement efficiency.
- These accounts represent potential high ROI for brand collaborations.

---

7. Advanced Analysis

7.1 Growth Potential (New Post Performance)

By comparing:

- Average likes
- New post average likes

We identified influencers with rising engagement momentum.
Influencers with increasing new post likes demonstrate strong audience growth and content relevance.

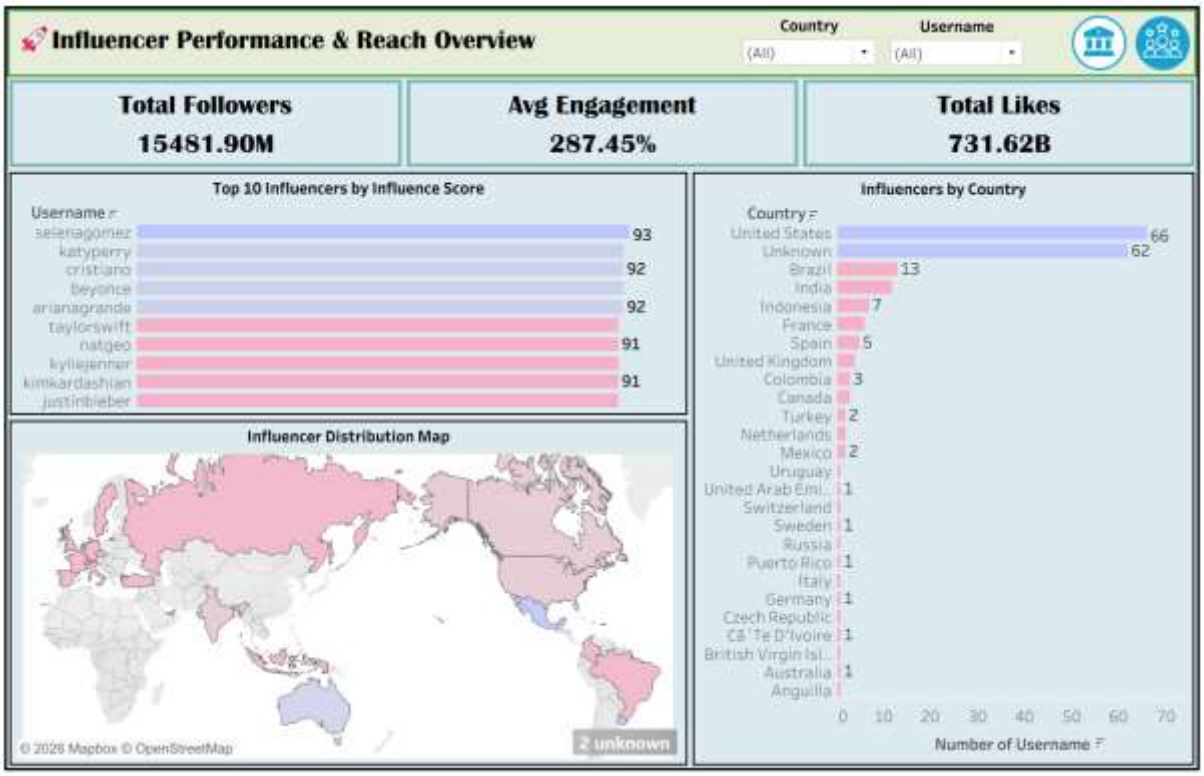
7.2 Country-wise Influence Score

- United States dominates total influence and engagement volume.
- India and Brazil show strong emerging influence presence.
- Some influencers have "Unknown" country classification but still contribute significantly to total likes.

---

8. Tableau Dashboard Analysis

Dashboard 1: Influencer Performance & Reach Overview



Key Metrics Displayed:

- Total Followers
- Total Posts
- Total Likes
- Average Influence Score

Key Visualizations:

1. Followers vs Average Likes (Scatter Plot)

- Shows correlation between reach and engagement.
- Identifies outliers with strong engagement despite lower followers.

2. Posts vs Influence Score

- Helps analyze whether posting frequency impacts influence.
- Moderate posting frequency often aligns with higher influence.

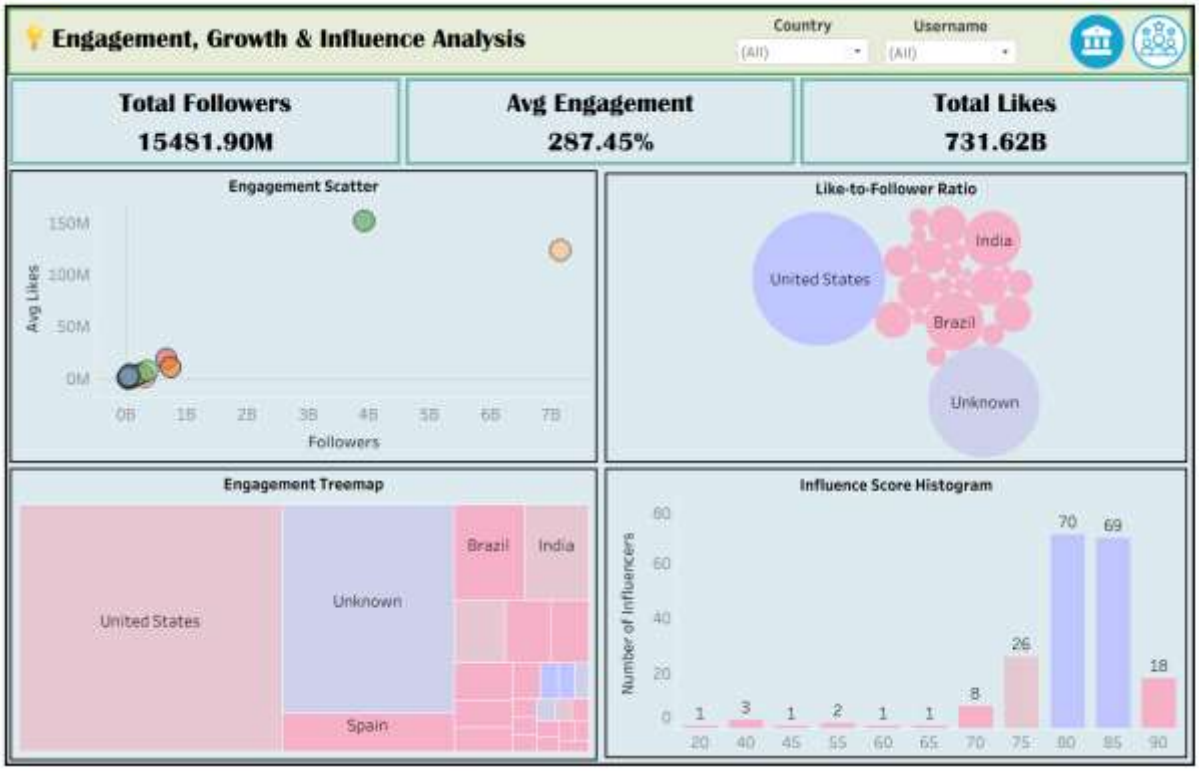
3. Country-wise Followers Distribution

- United States holds the largest share.
- Emerging markets contribute significantly to global influence.

4. Top Influencers by Total Likes

- Highlights creators generating massive engagement volume.

Dashboard 2: Engagement, Growth & Influence Analysis



KPI Summary:

- Total Followers: 15,481.90M
- Average Engagement: 287.45%
- Total Likes: 731.62B

1. Engagement Scatter Plot

- Plots Followers vs Average Likes.
- Identifies:
  - High reach, high engagement influencers
  - Low reach but strong engagement creators
- Demonstrates that influence is not purely follower-driven.

2. Like-to-Follower Ratio (Bubble Chart)

- Measures engagement efficiency.
- Countries like the United States dominate.
- India and Brazil show strong competitive ratios.
- Highlights audience loyalty and content effectiveness.

3. Engagement Treemap

- Shows country-wise contribution.
- United States leads.
- Unknown and Spain contribute notable engagement.
- Brazil and India show rising impact.

4. Influence Score Histogram

- Majority of influencers fall between 80–85 range.
- Few influencers below 60.
- Very limited ultra-high influence above 90.

This indicates a competitive but concentrated influence landscape.

---

9. Key Business Insights

1. Follower count alone is not a reliable performance metric.
2. Engagement rate and like-to-follower ratio are stronger performance indicators.
3. Emerging markets (India, Brazil) present strong growth opportunities.
4. Mid-tier influencers may offer better ROI than mega-influencers.
5. Growth potential can be predicted using new post performance trends.

---

10. Conclusion

This project successfully combines SQL-based data analysis with Tableau visualization to uncover meaningful insights about Instagram influencer performance.

Key takeaways:

- Influence is multi-dimensional.
- Engagement efficiency matters more than scale alone.
- Geographic distribution impacts influence trends.
- Data-driven influencer selection improves marketing decisions.

The project demonstrates strong skills in:

- Data cleaning
- SQL querying
- Analytical thinking
- Data storytelling
- Dashboard design
- Business insight extraction

---

11. Future Scope

- Add time-series analysis for growth tracking
- Predict influence score using machine learning
- Brand-category based segmentation
- ROI estimation model for brand collaborations
- Integration with live Instagram API data