Assignment 1 - Part II
Data Analytics Process and Interpretation

IS4116 – Business Intelligence Systems

By
M.S. Kariyawasam - Index No: 20020554

GitHub Link - Click Here

Degree of Bachelor of Science Honours in Information
Systems

University of Colombo School of Computing
35, Reid Avenue, Colombo 07,
Sri Lanka

Feb 2025

1 Introduction

Understanding the factors that drive influencer success is crucial for businesses leveraging social media marketing. This analysis explores a dataset containing information on top Instagram influencers. By applying data preprocessing, exploratory analysis, statistical modeling, and visualization, the project aims to uncover key factors influencing influencer performance and provide actionable insights for brands to make informed marketing decisions.

1.1 Business Question

What factors contribute most to influencer engagement, and how can brands use this knowledge to optimize influencer partnerships?

1.2 Dataset Overview

The dataset includes metrics for top Instagram influencers across various industries. It has 200 entries with attributes such as Influencer Name, Influencer Score Follower Count, Engagement Rate, Average Likes, Number of Posts etc. (Dataset Link)

2 Data Cleaning and Preprocessing

1. Missing Value in 60-Day Engagement Rate

   - One row contained a NaN value in 60_day_eng_rate.
   - Solution: Replaced with the median engagement rate to maintain data consistency.

2. No Duplicate Entries were found, ensuring data integrity.

3. Replaced specified columns and converting them to float. ('total_likes', 'posts', 'followers', 'avg_likes', '60_day_eng_rate', 'new_post_avg_like')

4. Summary statistics were calculated to estimate central tendencies and dispersions.

3 Data Analysis & Visualization

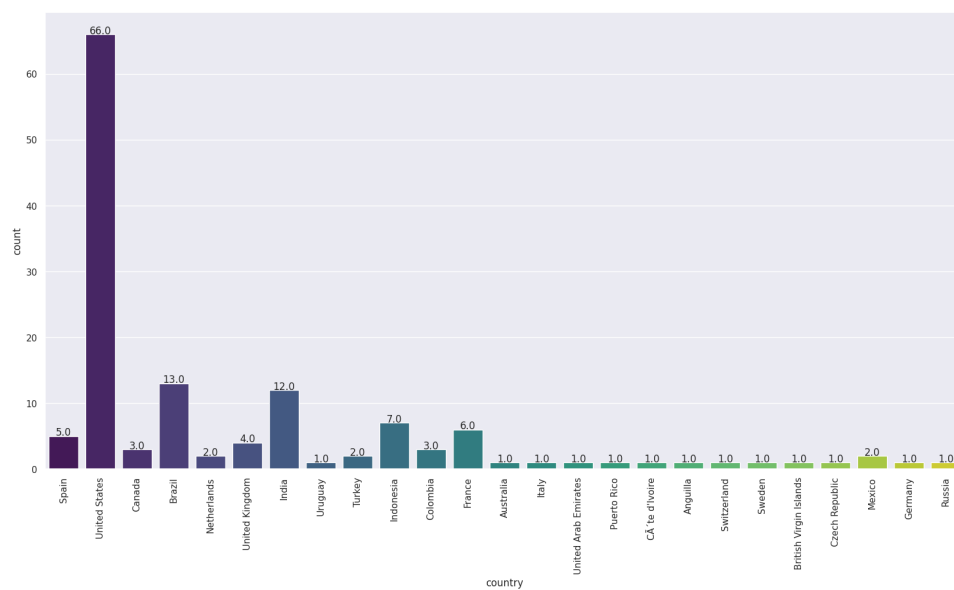## 3.1 Influencer Distribution Across Countries

Figure 1: Influencer Distribution Across Countries

Findings:

USA, UK, India, and Brazil have the most influencers. These nations are significant markets for influencer-marked brand endorsements. The denser population of influencers in these nations implies brands need to act strategically when it comes to the selection of influencers to prevent overcrowding.
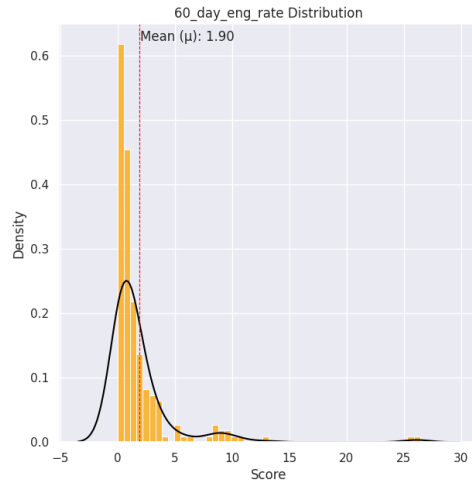
3.2 Distribution of Key Metrics

Figure 2: Distribution of Eng Rate

Figure 3: Box Plots

Findings: Engagement rates are right-skewed → There are few influential people who occupy the space. Most of the influencers have average engagement, but outliers may be extremely profitable. Because the dataset has a small size removing outliers is impractical. However, we can display the outliers as boxplots.
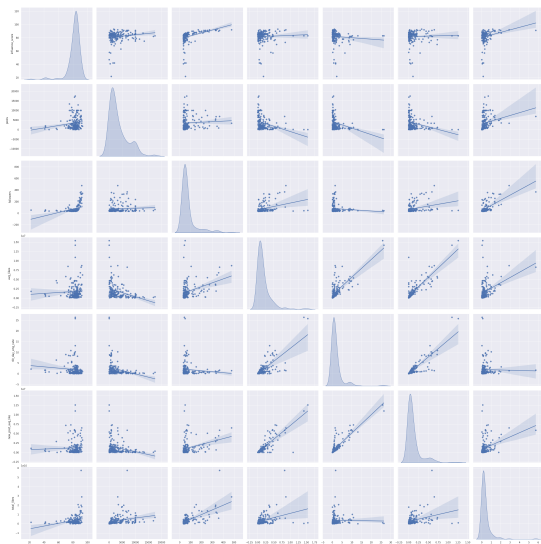
3.3 Correlation Analysis
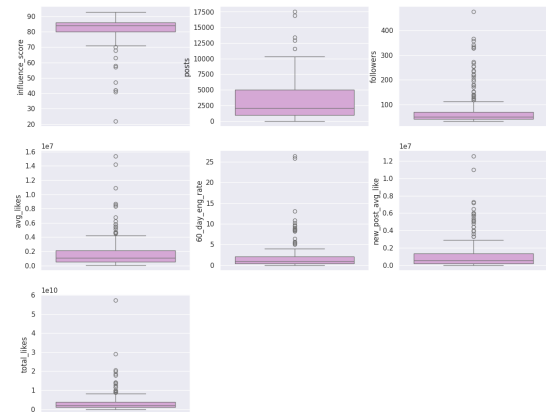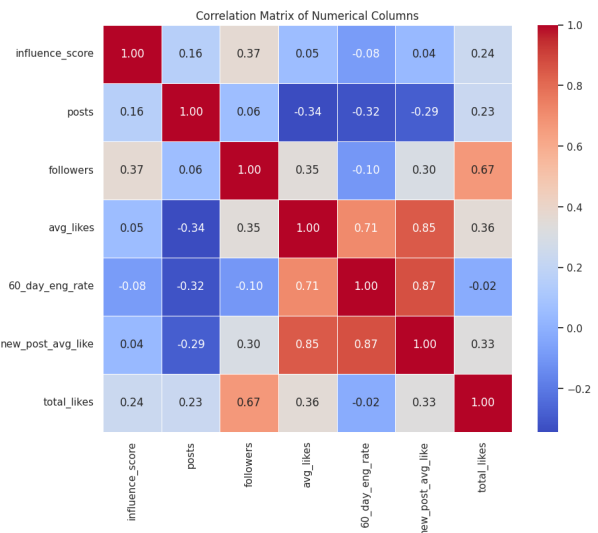

Figure 4: Distribution of Eng Rate

Figure 5: Box Plots

Findings:

On many graphs, there is a positive correlation between the studied parameters. Single outbursts are

also observed.

High correlation of followers with total likes → Followers count increases engagement. Followers don't relate to engagement rate → Micro-influencers work better in terms of engagement.

4   Cluster Analysis
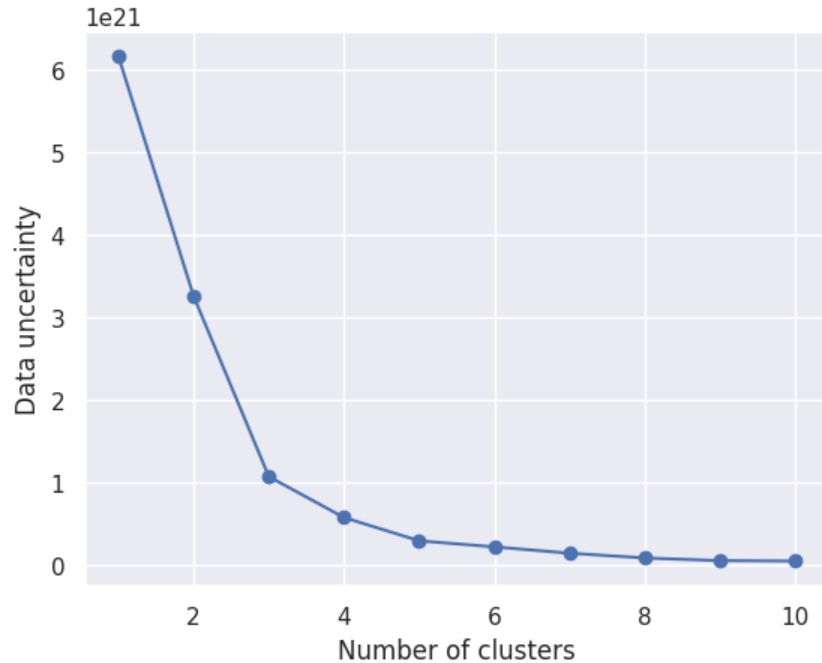

Figure 6: Cluster Analysis

Once K-means clustering is performed and the Elbow method is used it is clear that the optimal number of clusters is 4.

5   Key Takeaways

1. Leverage Micro-Influencers
   Brands would do well to target collaborations with micro-influencers (followers of fewer than 100,000) displaying high engagement levels.

2. Expand Influencer Targeting
   Brands have to broaden their influencer collaborations across various markets and industries to reach larger audiences. The research identified high influencer penetration in the USA, UK, India, and Brazil. By targeting influencers from other markets, brands can enhance their international reach while minimizing the threat of market saturation.

3. Monitor Engagement Metrics
   High-engagement influencers with high engagement metrics will be capable of delivering better campaign results. Brands need to create a process to vet influencers not just based on followers but based on their interaction rate, so they select partners who can better engage with their audience. .

4. Target Outlier Influencers
   Target and activate very high-engagement-rate outlier influencers for specific campaigns or product launchings.These outlier performers can potentially drive much noise and awareness for the campaigns. Engaging with outlier performers has the power to expose new audiences and establish exclusivity about the brand.