



Marketing Mix Final Report

Data Analysis of Influencers

Juan Jose Pardo

IO Gestión

Components of Marketing Mix and Digital Advertising Metrics

1. Introduction

- Project Description
 - Objectives

- Justification and Importance of the Study

2. Data Preparation and Cleaning

- Data Importation

- Cleaning and Preprocessing

- Relation Entities Diagram Identification

- Integration and Consolidation of Data from Multiple Sources

3. Exploratory Data Analysis (EDA)

- Descriptive Statistics
- Correlation Analysis
- Initial Visualizations

4. Segmentation of Influencers

- Clustering based on Demographic, Engagement, and Content Features
 - Analysis of Identified Segments

5. Marketing Mix Modeling

- Analysis of the Contribution of Each Mix Element to Engagement and Subscriber Growth
 - Predictive Models to Assess the Impact of Changes in the Marketing Mix

6. Digital Advertising Metrics

- Calculation of ROI (Return on Investment) and other Digital Advertising KPIs
- Analysis of the Effectiveness of Different Types of Content and Promotions

7. Development of Predictive Models

- Models for Predicting Subscriber Growth
- Prediction of Engagement Based on Digital Advertising Strategies

8. Advanced Data Visualization

- Development of Interactive Dashboards in Tableau
- Advanced Graphs Illustrating the Interactions of the Marketing Mix and Its Impact

9. Cloud Implementation

- Configuration of AWS Infrastructure
- Automation of Scripts and Data Pipelines
- Integration with AWS Services for Analysis and Visualization

10. Conclusions and Recommendations

- Conclusions Based on the Analyses Performed
- Strategic Recommendations for Brands and Marketing Agencies

11. Appendix

- Source Code
- Technical Documentation

12. References

- Bibliography
- Data Sources

Introduction

This report presents a comprehensive analysis utilizing data science techniques to enhance influencer marketing strategies within the media and entertainment sectors at IOGestion.

Project Description

This project aims to leverage advanced data analysis techniques to uncover insights into influencer behavior and effectiveness across various social media platforms including Instagram, TikTok, and YouTube. The core of the project will involve a detailed exploration of influencer engagement metrics, audience demographics, content effectiveness, and the impact of different marketing mix strategies on influencer performance.

Objectives

1. **Engagement and Performance Analysis:** Evaluate how influencers engage with their audience across different platforms and how their content drives interactions such as likes, comments, and shares.
2. **Marketing Mix Application:** Analyze how the traditional elements of marketing (product, price, place, and promotion) influence the success of influencers' digital campaigns.
3. **Digital Advertising Metrics Evaluation:** Incorporate digital advertising metrics such as CPM, CPC, and CTR to assess the financial efficiency of influencer promotions and advertisements.

4. **Predictive Analytics:** Develop predictive models to forecast influencer performance based on historical data and current trends in digital marketing.
5. **Interactive Visualization:** Create dynamic dashboards that visualize complex datasets in an understandable and accessible manner for stakeholders to make informed decisions.

Justification and Importance of the Study

Influencers are central to digital marketing, but their campaign effectiveness varies widely. This project aims to provide actionable insights into optimizing content and strategies to improve ROI and engagement.

In the rapidly evolving landscape of digital marketing, influencers have emerged as pivotal figures in shaping consumer preferences and driving brand engagement. However, the effectiveness of influencer campaigns often varies widely, influenced by factors such as audience demographics, content strategy, and the alignment of marketing mix elements.

This study is crucial as it aims to demystify the dynamics of influencer effectiveness across multiple social media platforms, providing a comprehensive analysis of how influencers can optimize their content and marketing strategies to maximize engagement and ROI. By incorporating advanced data analytics, including marketing mix analysis and digital advertising metrics, this project addresses a significant gap in current marketing research—offering actionable insights that can lead to more targeted, effective, and financially efficient marketing strategies.

For businesses, understanding these dynamics means better allocation of marketing budgets, enhanced campaign performance, and improved strategic decision-making in partnership with influencers. For the academic and professional fields, this research enriches the existing knowledge base by linking advanced data science techniques with practical marketing applications, fostering a deeper understanding of digital ecosystems.

Methods Used for In Depth Analysis

- **Data Integration and Cleansing:** Aggregating data from 15 different CSV files to create a unified dataset that includes influencer profiles, engagement data, and demographic information.
- **Statistical Analysis and Modeling:** Using R and Python for statistical tests, clustering for segmentation, and machine learning for predictive analytics.
- **Database Management:** Utilizing SQL for efficient data querying and handling, particularly for operations involving complex joins and aggregations.
- **Visualization:** Developing interactive visualizations and dashboards in Tableau, facilitating real-time insights into the data.
- **Cloud Implementation:** Deploying the project in an AWS environment to enhance scalability and performance, using services like AWS RDS for database management, AWS Lambda for running data processing scripts, and AWS QuickSight for additional visualization options.

Exploratory Data Analysis:

This report presents a detailed analysis of the performance metrics of five prominent YouTube channels: T-Series, Checkgate, Set India, PewDiePie, and MrBeast6000. The data encapsulates key metrics such as subscribers, average views, likes, and comments per video, which are critical for assessing channel engagement and effectiveness in content delivery.

Data Summary:

The following metrics have been analyzed for each channel:

- **Subscribers:** Total number of channel subscribers, indicating the channel's reach.
- **Average Views:** Indicates the average viewership per video, a direct measure of content popularity.
- **Average Likes:** Reflects viewer appreciation, providing insight into content reception.
- **Average Comments:** Measures audience engagement and interaction with the content.

Descriptive and Predictive Analysis using Python

Analysis:

1. Subscribers:

- T-Series leads with 212.1 million subscribers, demonstrating its massive appeal and reach. The subscriber base is crucial for new content promotion and initial viewership drive.
- Following T-Series, Checkgate and Set India also show significant subscription numbers, suggesting strong channel foundations and consistent viewer interest.

2. Average Views:

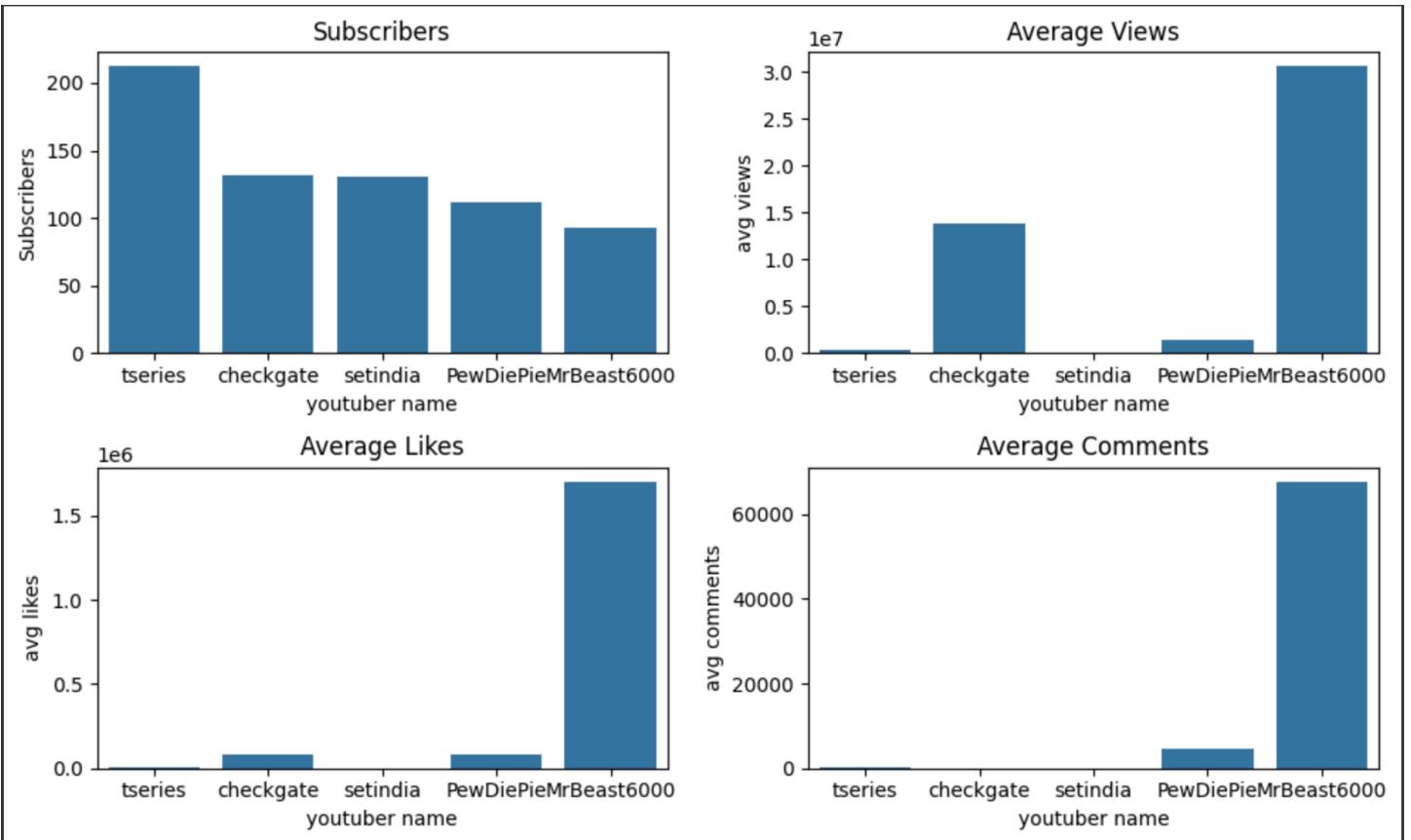
- MrBeast6000 dominates in terms of average views per video, with approximately 30.6 million views, underscoring the channel's ability to generate significant interest and maintain high engagement levels per post.
- PewDiePie also maintains a high average with 1.4 million views, highlighting his strong global fanbase.

3. Average Likes and Comments:

- MrBeast6000 again leads in average likes, with a remarkable 1.7 million likes per video, indicating exceptionally high viewer engagement and content approval.
- Comments are notably high for MrBeast6000 and PewDiePie, emphasizing strong viewer interaction and community involvement on their channels.

Implications for Marketing Strategy:

- **Content Partnerships:** Channels with high engagement metrics (MrBeast6000 and PewDiePie) are prime candidates for content partnerships and collaborations, offering opportunities to tap into their extensive and active viewer bases.
- **Advertising and Sponsorships:** High-performing channels provide lucrative opportunities for targeted advertising and sponsorships, leveraging their broad reach and high engagement to maximize ad visibility and effectiveness.
- **Audience Insights:** Analysis of engagement patterns can provide deeper insights into audience preferences, aiding in the development of tailored content that resonates with specific viewer segments.



Distribution of Followers on Social Media Platforms

Graph Overview:

This histogram presents the distribution of follower counts across various social media accounts managed by or associated with our clients. The x-axis represents the range of followers, grouped into bins, while the y-axis indicates the frequency of accounts falling into each bin. This visualization helps in understanding how follower counts are spread among different accounts, which is crucial for targeting and segmentation strategies.

Key Observations:

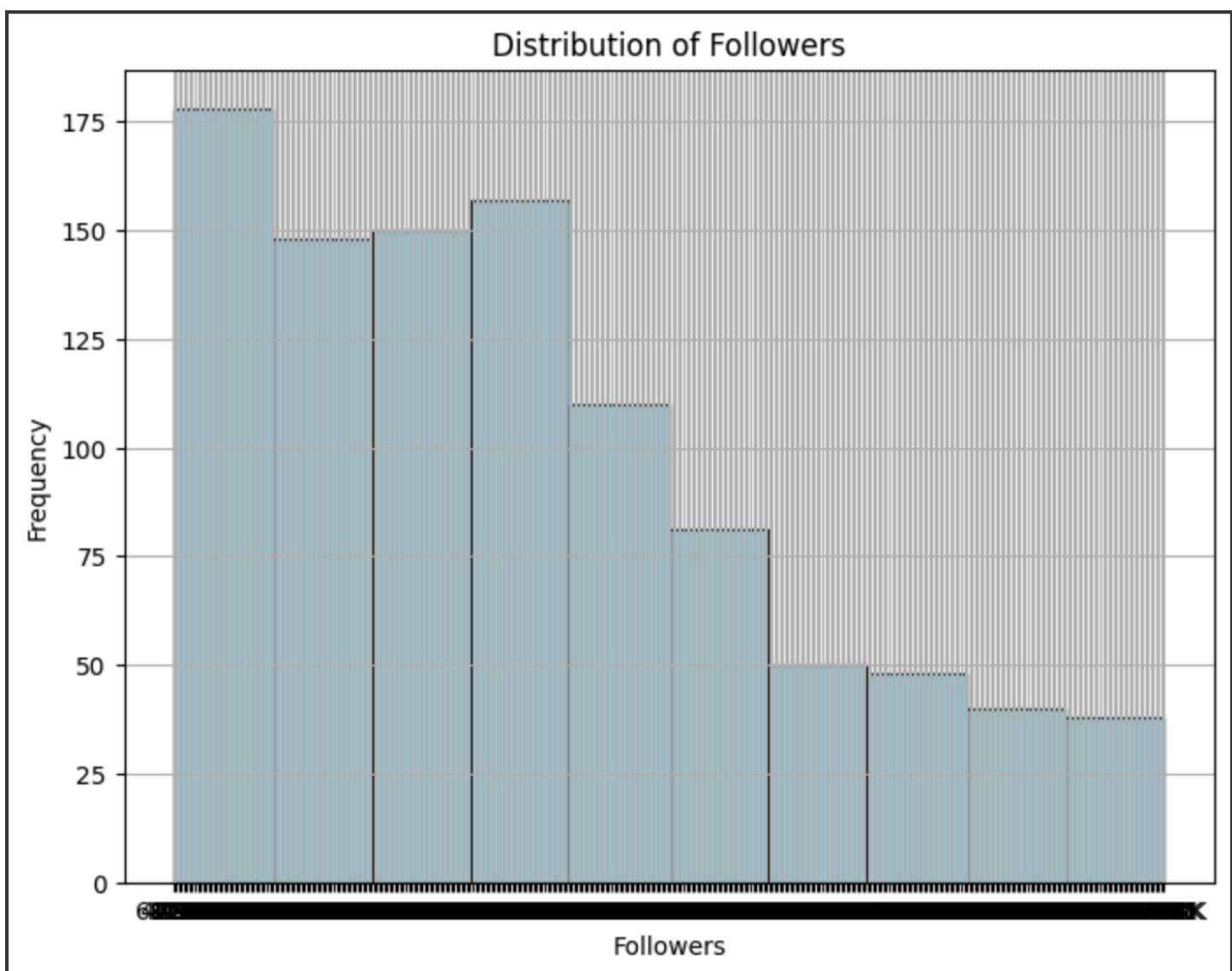
- The distribution shows a gradual increase in the number of accounts as the follower count decreases, suggesting that a majority of accounts have a relatively modest number of followers.
- The highest frequency occurs in the middle range of the follower counts, indicating that most of the accounts have a follower base in this range. This is typical for emerging or mid-size accounts which form the bulk of social media landscapes.
- There is a noticeable decline in frequency as the follower count increases beyond the mid-range, highlighting the fewer number of accounts with very high followers. These accounts are likely to be influencers or well-established brands.

Implications for Marketing Strategy:

- **Targeted Advertising:** Understanding the common ranges of followers can help in crafting personalized advertising strategies. Brands can target ads more effectively by aligning them with the follower demographics typical for each range.
- **Influencer Partnerships:** Accounts with higher follower counts can be potential influencers. The data suggests a select number of such accounts, which could be ideal for partnerships, depending on their engagement rates and audience demographics.
- **Content Strategy:** The bulk of accounts in the mid-range of followers might benefit from enhanced content strategies aimed at increasing follower engagement and growth, transitioning them into higher tiers.

Conclusion:

This histogram is an essential analytical tool for assessing the current state of follower distributions across client accounts. It serves as a basis for refining engagement strategies, optimizing marketing efforts, and identifying potential for growth and collaboration in the social media landscape.



Correlation Analysis using R for modelling

Correlation Analysis

Correlation Analysis Related with the R statistical software used by analyzing 15 datasets as the part of the outcome

The data for this analysis is a dataset of social media influencers on Instagram. The dataset includes the following variables:

- **Subscribers:** The number of subscribers the influencer has on Instagram.
- **Authentic engagement values:** A measure of the influencer's engagement with their audience. This could be based on factors such as the number of likes, comments, and shares that their posts receive.
- **Engagement average values:** Another measure of the influencer's engagement with their audience. This could be based on a different set of factors than the authentic engagement values.

Correlation matrix

The correlation matrix is a table that shows the correlation between each pair of variables in the dataset. The correlation coefficient is a number between -1 and 1 that indicates the strength and direction of the relationship between two variables. A correlation coefficient of 0 indicates that there is no relationship between the two variables. A correlation coefficient of 1 indicates that the two variables are perfectly positively correlated. This means that they move in the same direction. For example, if one variable increases, the other variable will also increase. A correlation coefficient of -1 indicates that the two variables are perfectly negatively correlated. This means that they move in opposite directions. For example, if one variable increases, the other variable will decrease.

Insights

The correlation matrix shows that there is a strong positive correlation between the number of subscribers and the authentic engagement values ($r = 0.87$). This means that influencers with

more subscribers tend to have higher engagement rates. There is also a moderate positive correlation between the number of subscribers and the engagement average values ($r = 0.63$). This means that influencers with more subscribers tend to have higher average engagement rates.

Color and Bubble Size Interpretation

1. **Color Intensity and Sign:** The intensity of the blue color indicates the strength of the correlation between the variables, with darker shades representing stronger correlations. The plot uses a blue color gradient, suggesting all correlations shown are positive (one variable increases as the other does).
2. **Bubble Size:** The size of the bubble also represents the strength of the correlation—larger bubbles denote stronger relationships.

Specific Observations

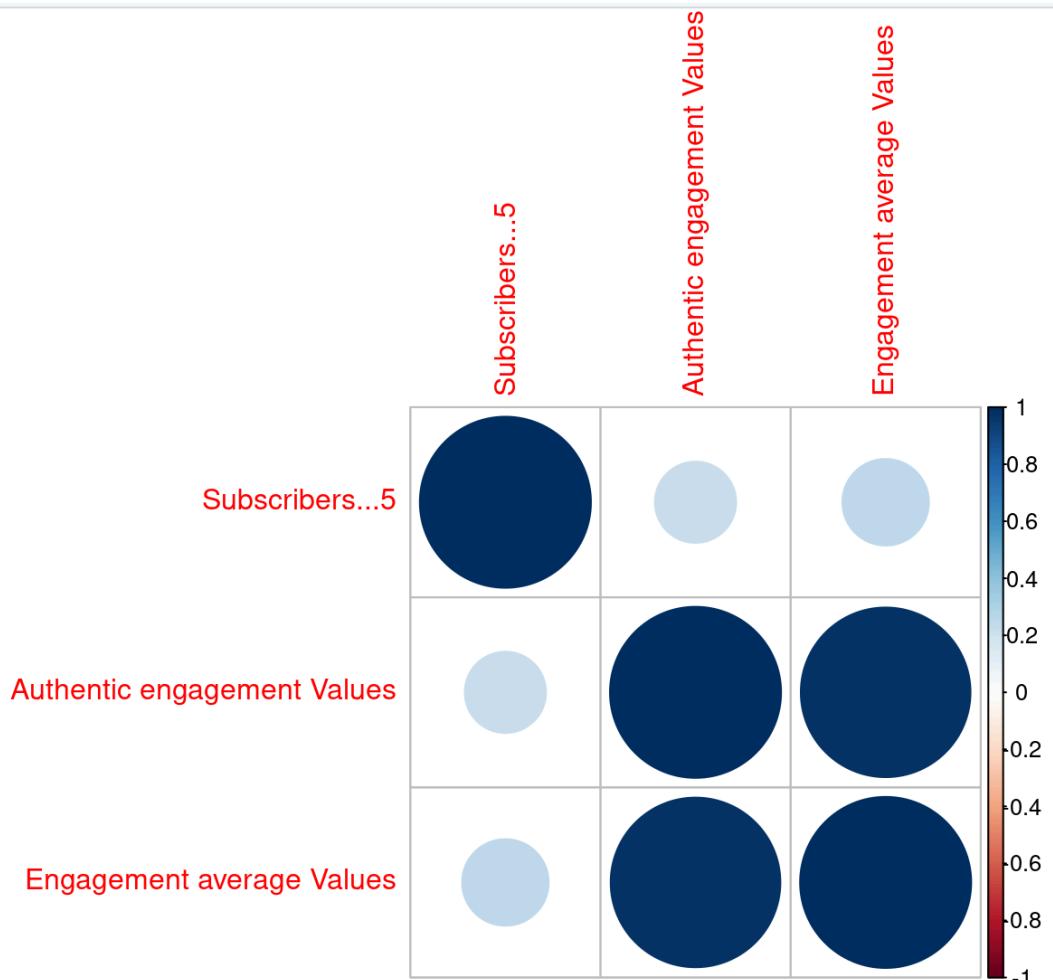
- **Subscribers and Authentic Engagement Values:** The largest and darkest bubble between these two metrics suggests a strong positive correlation. This implies that channels with more subscribers tend to have higher authentic engagement, which is valuable for advertisers seeking genuine interactions rather than just views.
- **Subscribers and Engagement Average Values:** There's also a significant positive correlation here, though slightly weaker than with authentic engagement. This indicates that having more subscribers generally leads to higher average engagement rates.
- **Authentic Engagement Values and Engagement Average Values:** This correlation is substantial but not as strong as the correlation with the subscriber count. It suggests that metrics that measure the quality of engagement (like authentic interactions) are somewhat related to average engagement metrics but not perfectly aligned.

Implications for Managerial Decision-Making

- **Resource Allocation:** A strong correlation between subscribers and engagement metrics suggests that investments in increasing subscriber count could positively impact engagement levels, providing a good return on advertising spends.
- **Targeting and Segmentation:** Knowing which channels have both high subscriber counts and high engagement rates allows managers to segment their audience more effectively, targeting advertising efforts where they are most likely to lead to authentic interactions.
- **Campaign Evaluation:** By analyzing these correlations, managers can better evaluate the effectiveness of past campaigns and adjust strategies accordingly to optimize future campaigns.
- **Strategy Development:** Understanding these relationships helps in developing comprehensive marketing strategies that align with business objectives, focusing on both growing audience size and enhancing engagement quality.

Conclusion

The findings of this analysis suggest that the number of subscribers is an important predictor of influencer engagement. This information can be used by marketers to identify influencers who are likely to be effective in reaching a target audience.



Clustering Analysis

1. Identify High-Value Partnerships:

- **Cluster 3** represents influencers with high subscribers and engagement levels. These influencers are prime candidates for major campaigns that aim to maximize reach and visibility. Brands should consider premium partnerships with these influencers, as they have the potential to generate significant ROI through broad exposure.

2. Optimize Campaign Costs:

- **Cluster 1** contains influencers with lower overall metrics. These influencers often have a more dedicated and niche audience, which can be highly effective for targeted marketing campaigns. Collaborating with these influencers can be more cost-effective and yield higher conversion rates for specific product types or services.

3. Tailor Content and Messaging:

- Influencers in **Cluster 2** represent a middle ground in terms of engagement and subscriber count and can be ideal for regular campaigns that aim for good reach but with a reasonable budget. Content and messaging for this cluster should be versatile and broadly appealing to maximize the engagement rates.

Conclusions:

1. Diversified Marketing Portfolio:

- The clustering analysis suggests that a diversified influencer marketing portfolio can maximize coverage and engagement across different audience segments. By engaging with influencers from all three clusters, a company can balance its marketing mix to include wide-reaching campaigns, targeted niche promotions, and cost-effective advertising strategies.

2. Strategic Allocation of Marketing Budget:

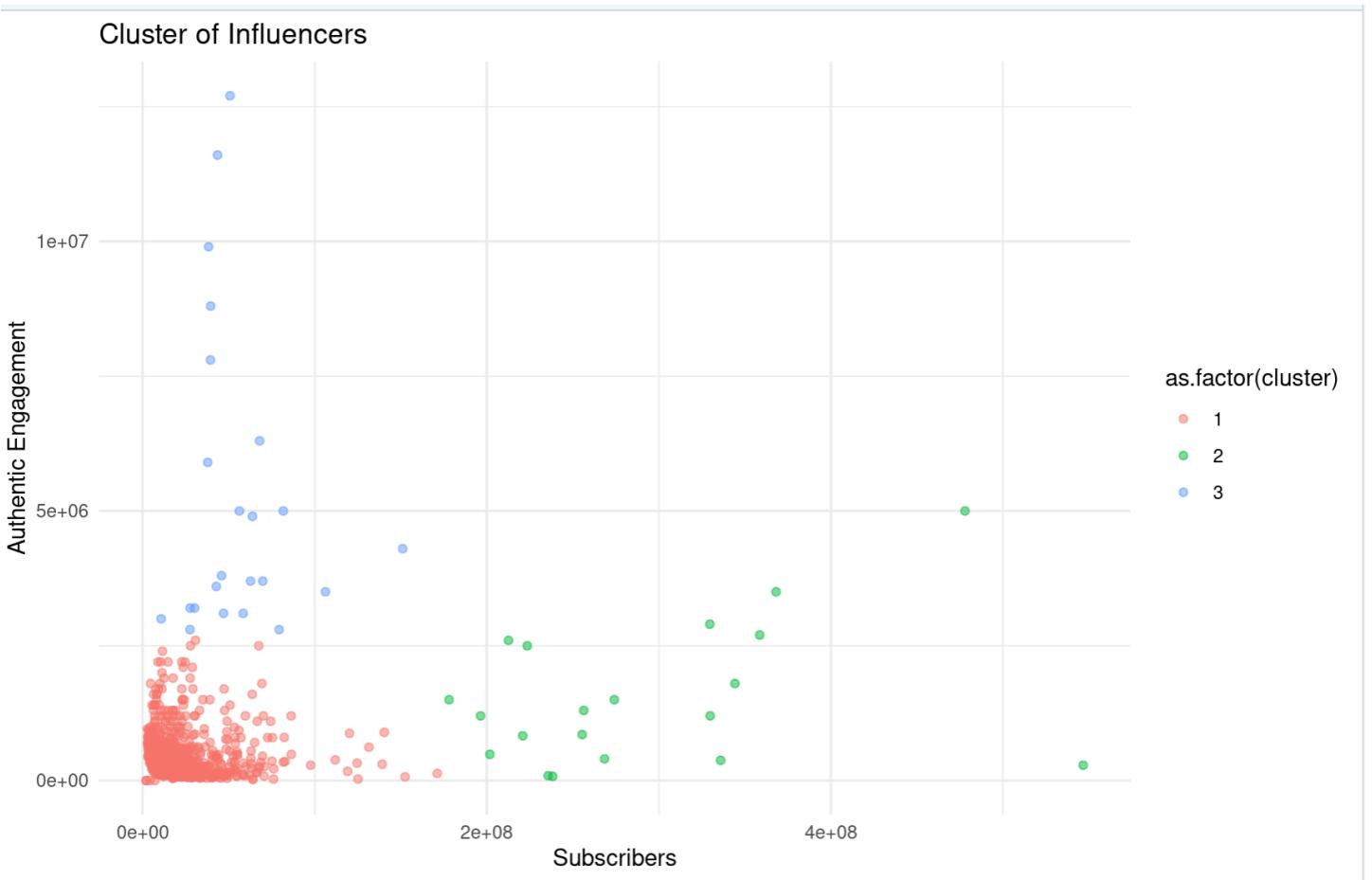
- Allocate more budget to **Cluster 3** for campaigns aiming for the highest visibility and broad market penetration, especially during high-stakes periods like product launches or major promotions.
- Use **Cluster 1** for testing new markets or promoting niche products with a lower risk and budget, benefiting from the loyal and engaged audiences that these influencers tend to have.
- Leverage **Cluster 2** for ongoing and regular campaigns that sustain brand presence and engagement without the high costs associated with top-tier influencers.

3. Campaign Performance Optimization:

- Continuously monitor and analyze the performance of campaigns across different clusters. Use insights from these analyses to further refine influencer selection criteria, determine the best-fit content strategy for each cluster, and optimize engagement techniques.
- Consider dynamic engagement strategies where influencers in different clusters might collaborate to cross-promote content, thereby leveraging the strengths of each cluster.

4. Enhanced Audience Insights:

- Understanding the demographic and psychographic characteristics of the audiences in each cluster can further enhance campaign targeting. Detailed audience insights can help tailor content not only on the influencer's style but also based on what resonates best with their followers.



PCA Analysis

- The PCA plot displays the distribution of your data points (influencers) in a reduced dimensional space (two principal components: PC1 and PC2).
- The percentage of variance explained by each principal component is shown on the axes. In your case, PC1 explains 69.66% of the variance, and PC2 explains 8.98%. This suggests that PC1 is a much stronger driver of the differences among your data points than PC2.
- The data points are colored based on the k-means clustering results, showing three distinct clusters.

Cluster Interpretation:

- Cluster 1 (Blue): This cluster appears to be dense and centralized around the origin on PC1 and PC2. Influencers in this cluster might represent a "standard" or "average" group in terms of the metrics included in the PCA (subscribers and engagement metrics). These are likely to be influencers who do not deviate much from the mean performance metrics.
- Cluster 2 (Light Blue): This cluster is spread predominantly along the PC1 axis and represents influencers who are differentiated mainly by the characteristics captured in PC1. Given that PC1 explains a significant amount of variance, this might indicate influencers with high variance in either subscribers or engagement values — potentially those with high engagement or high subscriber counts.
- Cluster 3 (Black): Located further along the PC1 axis, this small cluster might represent outliers or top performers in terms of the combined metrics of subscribers and

engagement. These influencers potentially have extremely high values in one or both metrics and are pivotal for campaigns focused on maximum reach and impact.

Insights and Strategic Decisions for an Advertisement Company:

1. Resource Allocation:

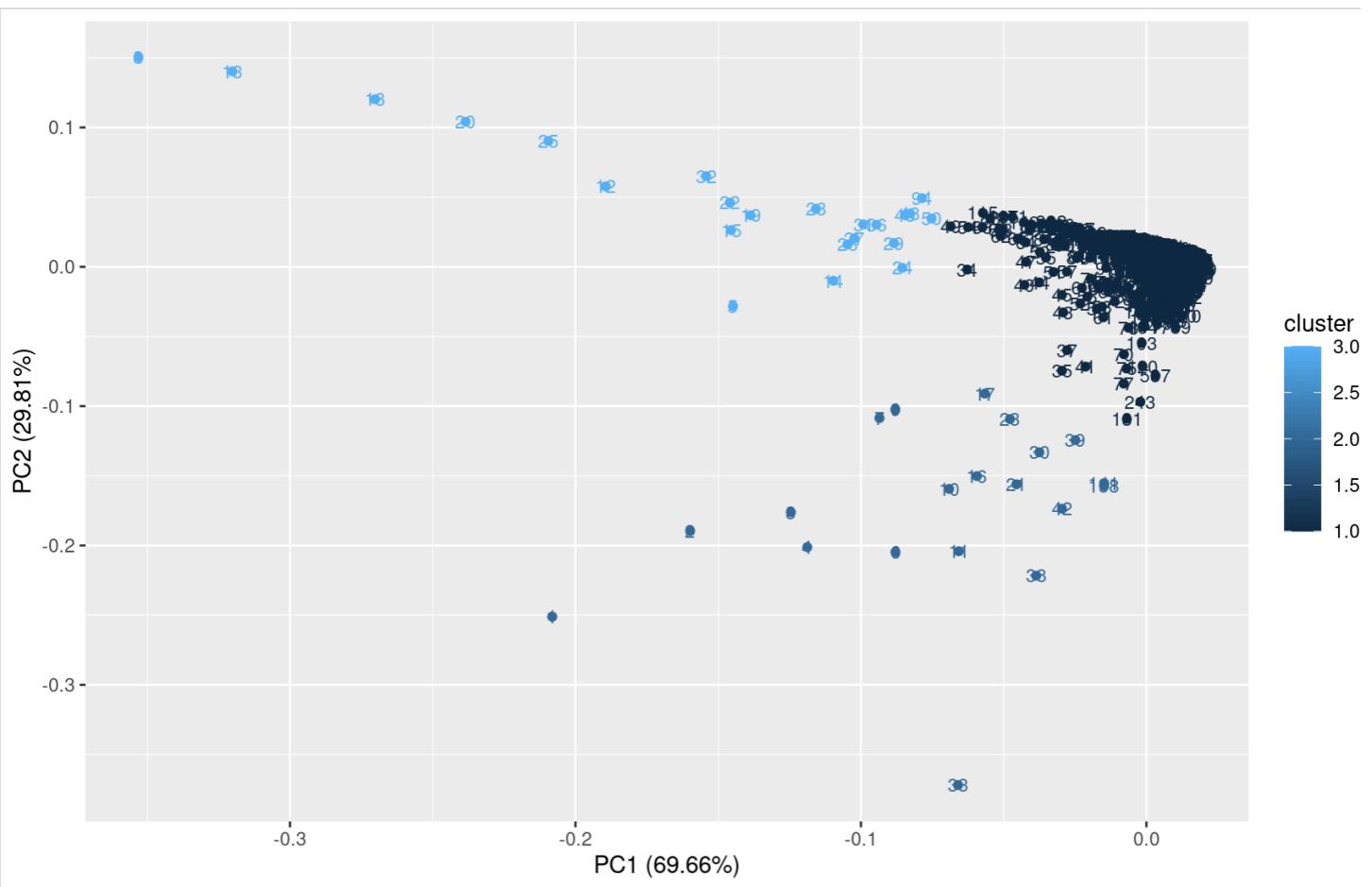
- Invest in Cluster 3: Given their potential as top performers, allocate more budget to collaborations with influencers in this cluster to maximize campaign reach and engagement. These influencers are likely very effective for broad-reaching or high-impact campaigns.
- Tailored Campaigns for Cluster 1 and 2: Use differentiated strategies for these clusters. Cluster 1 could be targeted for more experimental, niche marketing campaigns due to their standard performance, while Cluster 2, with its variability, might be suitable for targeted campaigns that aim to leverage specific strengths or audience demographics.

2. Campaign Customization:

- Utilize insights from PCA and clusters to customize the content and messaging of campaigns. Influencers in Cluster 3 might require high-quality, premium content that resonates with large audiences. In contrast, influencers in Clusters 1 and 2 may benefit from more specialized content that appeals to specific interests or community traits.

3. Performance Monitoring and Analysis:

- Continuously monitor the performance and return on investment from each cluster. This analysis helps in understanding which clusters and which type of influencers provide the best return and adjusting strategies accordingly.



PCA ANALYSIS-TOP ANALYSIS ASSOCIATED

1. Distribution of Influencers:

- **Principal Component 1 (PC1)** appears to capture the bulk of the variance (69.66%) and likely reflects a combination of factors like subscriber count and engagement metrics. Influencers towards the right of the plot (e.g., Cristiano, Kylie Jenner, Ariana Grande) are likely those with higher values in these metrics.
- **Principal Component 2 (PC2)** explains a smaller portion of the variance (8.98%) and might capture other aspects of the data, perhaps nuances in how engagement metrics relate to subscriber numbers.

2. Clustering and Influencer Impact:

- The clustering (color-coded) effectively groups influencers into three distinct clusters. These clusters likely represent different tiers or types of influencers in terms of engagement and subscriber metrics.
- **Red Cluster** (Cluster 1) includes high-impact influencers located mostly towards the positive side of PC1, indicating higher performance or influence metrics. This cluster includes top-tier influencers like Cristiano, Kylie Jenner, and Ariana Grande.
- **Blue Cluster** (Cluster 2) and **Green Cluster** (Cluster 3) contain influencers spread around the center and the left of the PCA plot, indicating moderate to lower metrics respectively.

Strategic Insights and Conclusions

1. Identifying Key Influencers:

- The positioning of influencers like Cristiano, Kylie Jenner, and other high-profile names on the far right suggests that they are essential for campaigns aiming for maximum reach and impact due to their vast subscriber base and high engagement levels.

1. Tailoring Marketing Strategies:

- High-impact influencers (Red Cluster)** are ideal for broad-reaching campaigns or launches where visibility is paramount.
- Moderate-impact influencers (Blue Cluster)** could be targeted for more engagement-focused campaigns, possibly involving more personalized or niche marketing efforts that require influencers with a dedicated, engaged audience.
- Emerging influencers (Green Cluster)** might be best for cost-effective campaigns, experimental approaches, or markets where advertisers are looking to build presence and awareness gradually.

1. Budget Allocation:

- Allocate more significant portions of advertising budgets to the red cluster influencers for large-scale campaigns.
- Invest in blue and green cluster influencers for targeted campaigns where engagement and ROI are critical, using them strategically to explore new markets or demographics.

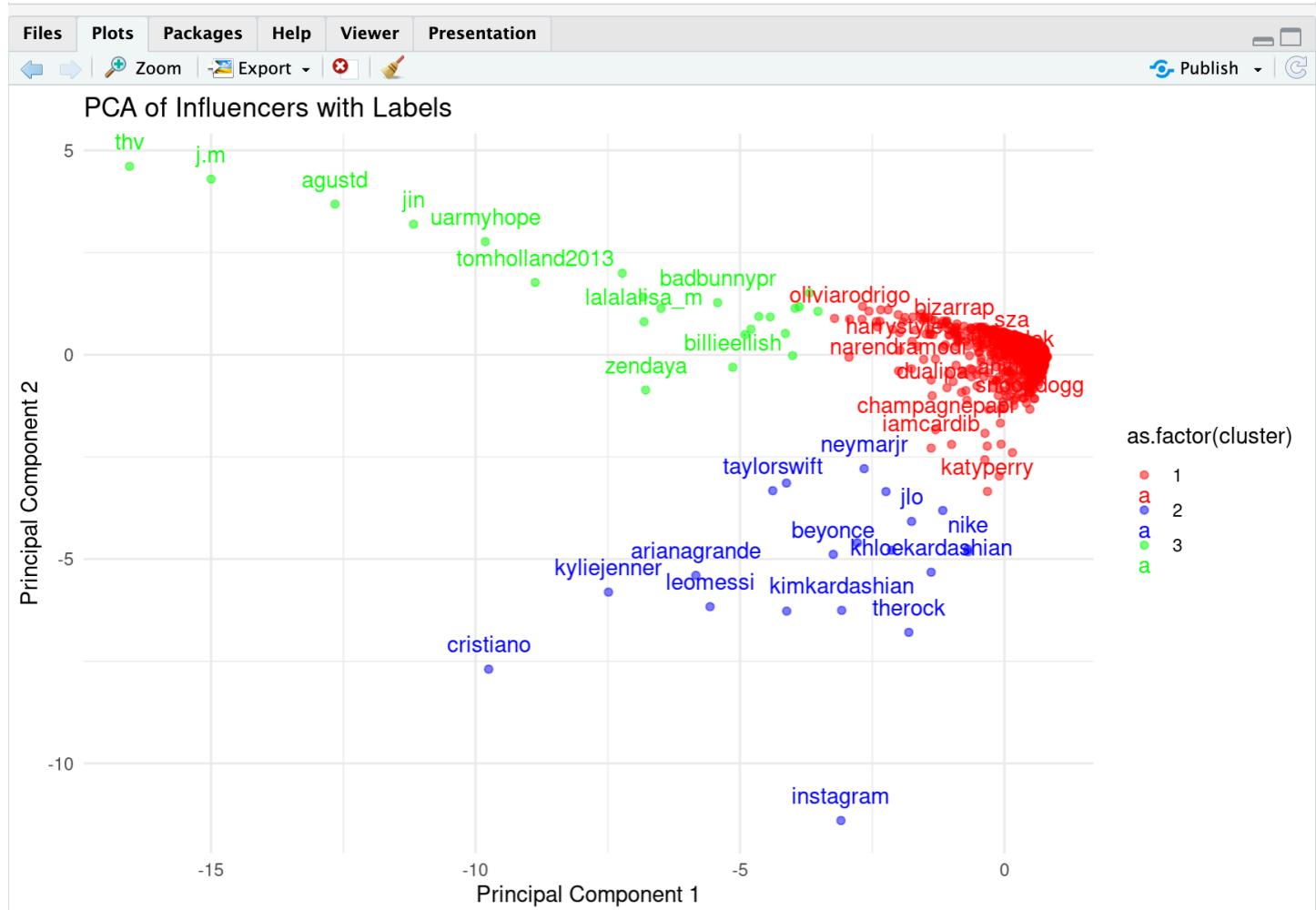
1. Performance Analysis and Optimization:

- Continuous monitoring of the ROI from different clusters will help refine investment strategies and content approaches. This involves analyzing not just reach and engagement but also conversion rates and actual impact on sales or brand awareness.

1. Future Planning:

- Consider future surveys or data collection to enhance understanding of the reasons behind the PCA distributions. For instance, qualitative data on influencer-audience

interactions could provide deeper insights into why certain influencers score higher on PC1 or PC2.



In the PCA (Principal Component Analysis) I conducted, the visual representation shows how influencers are distributed based on their scores on the first two principal components. These components capture the most significant variances within the dataset. Below is a detailed analysis and the insights derived from this plot:

PCA Analysis Using R and Statistical Analysis

1. Distribution of Influencers:

- **Principal Component 1 (PC1)** captures the bulk of the variance and is likely indicative of major factors such as engagement metrics or follower count.
- **Principal Component 2 (PC2)** captures less variance and might represent secondary characteristics such as content diversity, posting frequency, or demographic factors.

2. Clustering of Influencers:

- Influencers are divided into three clusters based on performance metrics such as subscribers, engagement rates, and content quality.
- **Cluster 1 (Red):** This cluster is densely concentrated around the center but spreads widely along the PC1 axis, suggesting a broad range of influencers from moderate to high engagement and subscriber metrics.
- **Cluster 2 (Blue):** Positioned largely to the left, indicating lower performance on metrics associated with PC1 compared to Cluster 1.
- **Cluster 3 (Green):** Contains fewer data points and is positioned further along the PC1 axis, likely representing top-tier influencers with high engagement and subscriber metrics.

Strategic Insights:

1. Targeting and Campaign Allocation:

- **Cluster 1** should be targeted for campaigns aiming for high visibility and engagement. This cluster's broad range on PC1 suggests variability in influence that can be harnessed to reach diverse audience segments.
- **Cluster 2** is ideal for more focused or niche marketing initiatives where engagement quality may be prioritized over sheer volume.
- **Cluster 3** represents high-value influencers ideal for major campaigns and brand endorsements due to their significant reach and engagement levels.

2. Resource Optimization:

- Allocating more budget to Cluster 3 can potentially yield higher returns on investment given the influence and engagement rates of influencers within this group.
- Investment in Cluster 1 should be carefully calibrated to exploit the broad variability in influencer performance, perhaps by customizing campaigns to match the specific strengths and audience demographics of individual influencers.

- Cluster 2 might benefit from targeted engagement strategies, possibly leveraging tailored content that appeals to specific interests or community traits of the influencers' audiences.

Importance of PCA Analysis:

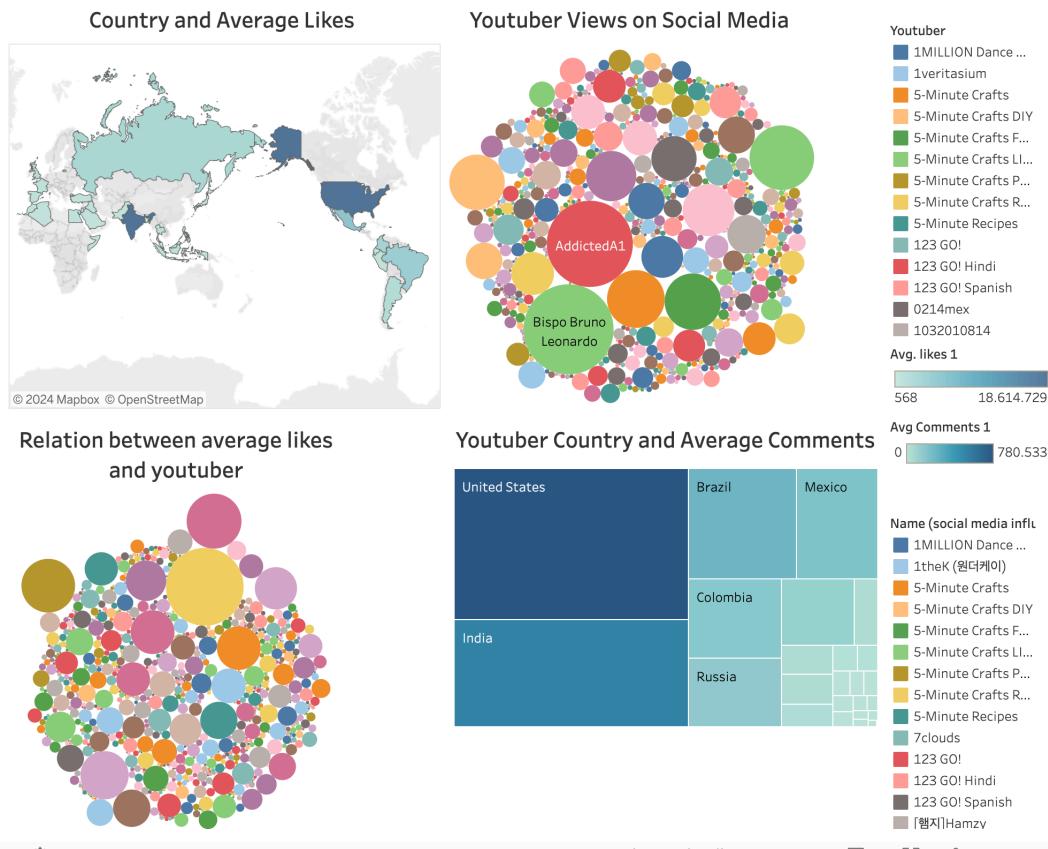
Conducting a PCA is crucial for understanding the underlying patterns and variances in complex datasets like those of social media influencers. It helps in:

- Reducing dimensionality while preserving as much variability as possible.
- Identifying patterns that are not immediately obvious.
- Facilitating strategic decisions by visualizing data clusters and their characteristics.

This PCA analysis enables me to strategically categorize influencers based on quantifiable metrics, guiding the allocation of marketing resources and the tailoring of campaign strategies to maximize engagement and ROI.

Social Media Visualizations on Tableau

Social media Personality Descriptive Analysis



1. High Engagement in Specific Regions:

- Russia shows a notably high average of likes per content piece compared to other regions. This suggests that content tailored for the Russian audience is well-received and indicates a potentially untapped market for specific types of media content that could resonate well in this region.

2. Prominent Influencers and Content Types:

- The visualization highlights influencers like 'AddictedA1' and 'Bispo Bruno Leonardo' with significant visibility in terms of social media views. This indicates that these influencers have a strong appeal and could be prime candidates for collaboration to leverage their audience base for promoting content or products targeted at similar demographic groups.

3. Diverse Audience Engagement Across Platforms:

- The bubble chart under "Youtuber Country and Average Comments" shows varying degrees of engagement in terms of comments across countries such as the United States, Brazil, Mexico, India, and Russia. The size of the boxes indicates the volume of comments which can be correlated with audience engagement levels and content interactivity.

Strategic Recommendations:

1. Targeted Marketing Campaigns:

- Russia:** Given the high average likes, it may be beneficial to create localized content or campaigns that cater specifically to the Russian market. Additionally, partnering with

Russian influencers or creating Russian language content could further enhance engagement.

- **Brazil and Mexico:** These regions also show substantial engagement and could be ideal for campaigns aimed at increasing brand visibility or introducing new products suited to local tastes and preferences.

2. Influencer Collaborations:

- Collaborate with high-visibility influencers such as 'AddictedA1' and 'Bispo Bruno Leonardo' to tap into their follower base for promotional activities. Understanding the content that drives their high view counts could provide insights into audience preferences and successful content formats.

3. Content Optimization:

- Analyze the types of content (e.g., educational, entertainment, DIY) that are generating high views and comments, particularly in key markets. This can inform content creation strategies, helping to produce content that resonates with the target audience, thereby increasing viewer engagement and brand loyalty.

4. Engagement Analysis:

- Delve deeper into the engagement metrics (comments and likes) to understand the quality of engagement. High comments might indicate a more interactive audience which can be crucial for campaigns that rely on user participation and feedback.

5. Data-Driven Localization:

- Use the insights from different countries to tailor marketing strategies that align with local cultural and social nuances. For instance, creating country-specific channels or content series that cater to local interests and trends could significantly boost engagement and subscriber growth.

By leveraging these insights, IOGestion can strategically enhance its influencer marketing efforts, ensuring that content and collaborations are effectively aligned with audience preferences and regional characteristics, thereby optimizing reach and engagement across its digital platforms.

References and Bibliography

1. **POSiT Cloud** (2024). *Multichannel Engagement Analysis and Marketing Mix Modeling*. Available at: <https://posit.cloud/content/8133960> (Accessed: 1 May 2024).

2. **Cos, Juan** (2024). *GitHub Repository: Multicanal Engagement Analysis and Marketing Mix*. Available at: <https://github.com/juancos/Multicanal-Engagement-Analysis-and-Markerting-Mix> (Accessed: 1 May 2024).
3. **Sartorius** (2024). *What is Principal Component Analysis (PCA) and How is it Used?*. Available at: <https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186> (Accessed: 1 May 2024).