Exploring Global YouTube Trends

Akhila Akula

Introduction and Data Description

Introduction & Background

- YouTube: Leading platform for content dissemination
- Billions of daily users consuming diverse videos
- Content spans education, entertainment, corporate brands
- Appeals to various interests and demographics
- Essential to understand channel dynamics, audience engagement, success factors

Objective

GLOBAL LANDSCAPE UNDERSTANDING

SUBSCRIBER DEMGRAPHICS ANALYSIS

CHANNEL CLUSTERING EXPLORATION

TREND TRACKING OVER TIME

Questions

- How do the estimated earnings of the YouTube channels vary across different categories based on the subscriber count?
- What are the top YouTube channel categories by total subscribers?
- In which countries are there the most YouTube channels? Are there any countries with a surprisingly high or low number of channels?
- Is there a steady increase in the creation of YouTube channels globally?
- Are there any specific periods with a surge in new channel creation?
- How does the growth of YouTube channels compare to the growth of internet users worldwide?
- Which metrics have the strongest positive or negative correlations (Uploads, Video Views, Subscribers, and Highest Yearly Earnings)?
- Are there any surprising correlations between seemingly unrelated metrics?
- How does the average gross tertiary education rate compare across different clusters? How does the average urban population density vary between clusters?
- How many channels are grouped into each identified cluster? Are there any particularly large or small clusters?

Impact and Benefits

UNVEIL GLOBAL YOUTUBE LANDSCAPE:

EMPOWER CONTENT CREATORS AND MARKETERS:

ADVANCE DIGITAL MEDIA RESEARCH:

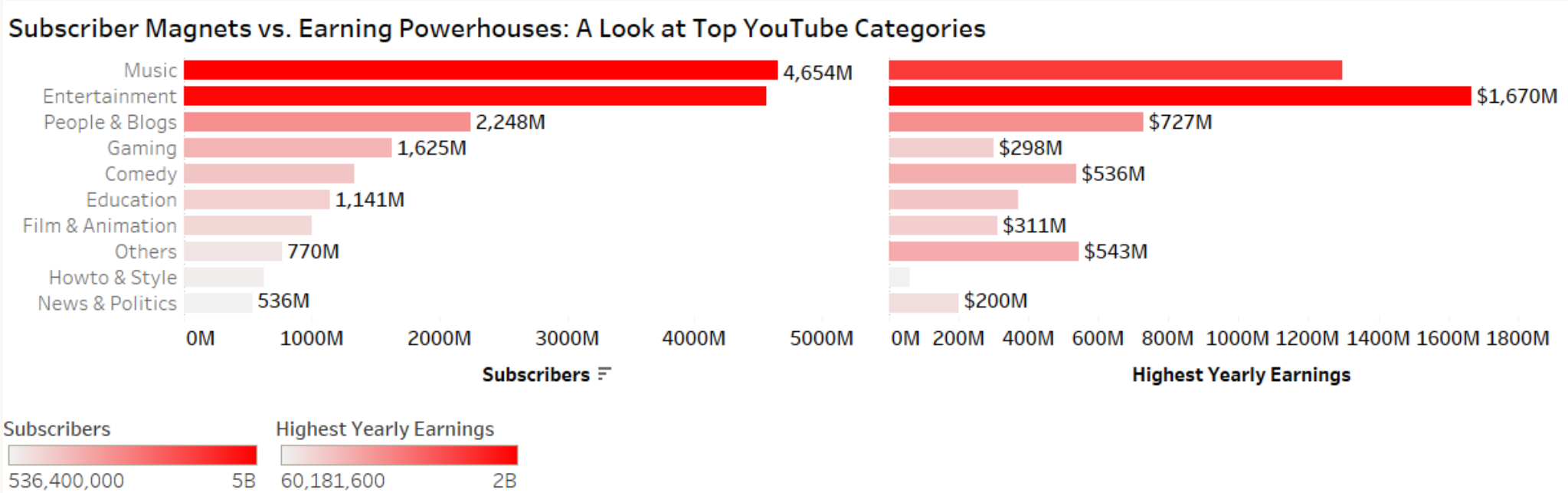
Dataset Description

- Dataset: "Global YouTube Statistics 2023"
  - Sourced from Kaggle, a platform for data science and ML professionals.
  - Hosts diverse datasets for knowledge-sharing in the community.
  - Currently utilized by over 100 data science enthusiasts for YouTube statistics analysis.
  - Contains 24 columns capturing various YouTube channel features.

Dataset – Colum Description

Column	Description
rank	It is the position of the YouTube channel based on the number of its subscribers.
Youtuber	The name of the YouTube channel.
subscribers	The number of subscribers to the channel.
video views	Total views for all videos on the channel.
category	The category of the channel.
Title	The title of the YouTube channel.
uploads	Total number of videos that are uploaded to the channel.
Country	Country where the YouTube channel is from.
Abbreviation	Two letter abbreviation of the country.
video_views_for_the_last_30_days	Total video views in the last 30 days.
lowest_monthly_earnings	Lowest estimated monthly earnings from the channel.
highest_monthly_earnings	Highest estimated monthly earnings from the channel.
lowest_yearly_earnings	Lowest estimated yearly earnings from the channel.
highest_yearly_earnings	Highest estimated yearly earnings from the channel.
subscribers_for_last_30_days	Number of new subscribers gained in the last 30 days.
created_year	Year when the YouTube channel was created.
created_month	Month when the YouTube channel was created.
created_date	Exact date of the YouTube channel's creation.
Gross tertiary education enrollment (%)	Percentage of the population enrolled in tertiary education (college or university) in the country.
Population	Total population of the country
Unemployment rate	Unemployment rate in the country.
Urban_population	Percentage of the population living in urban areas.
Latitude	Latitude coordinate of the country's location.
Longitude	Longitude coordinate of the country's location.

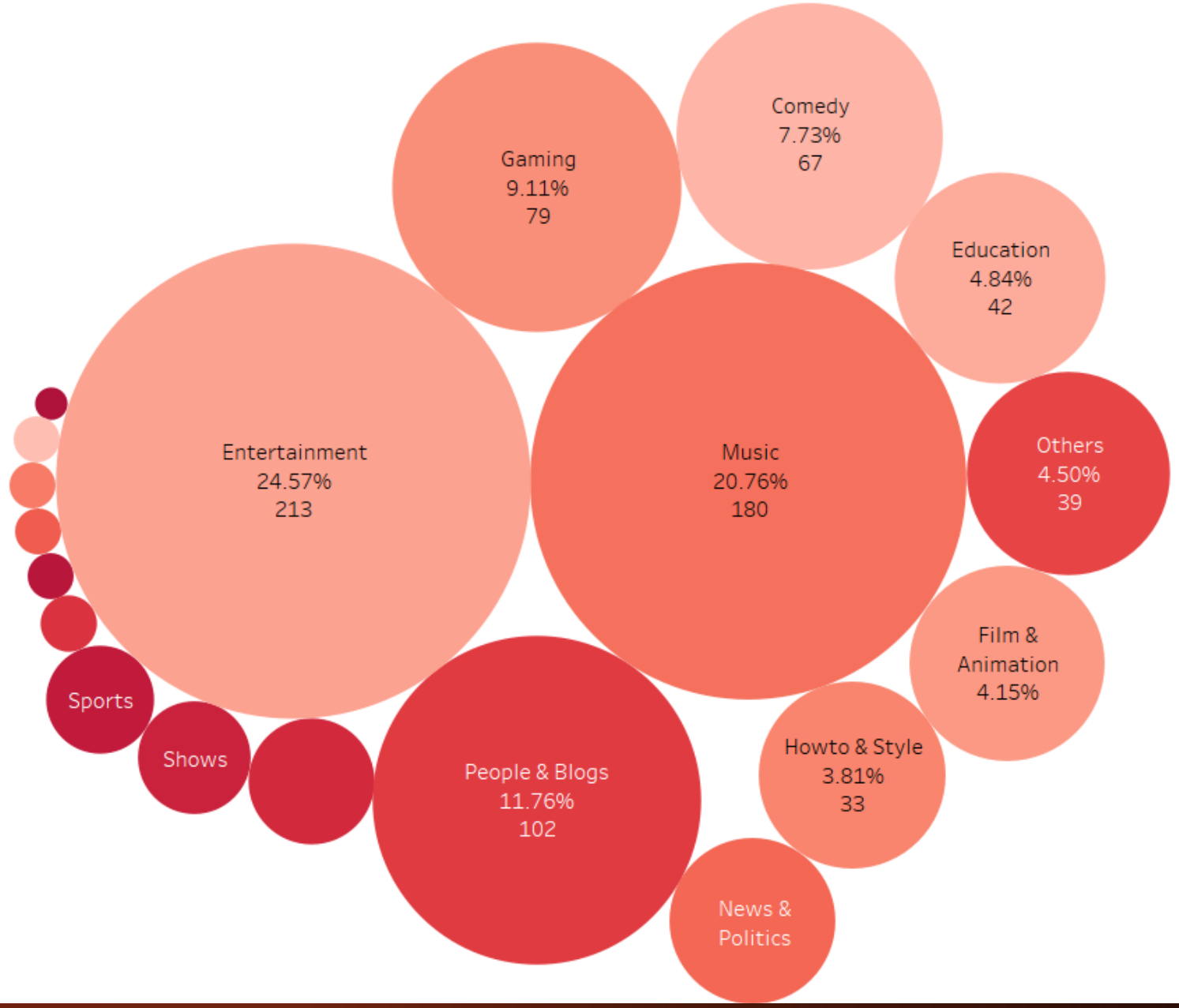
Data Stories



Subscriber Magnets vs. Earning Powerhouses:
A Look at Top YouTube Categories

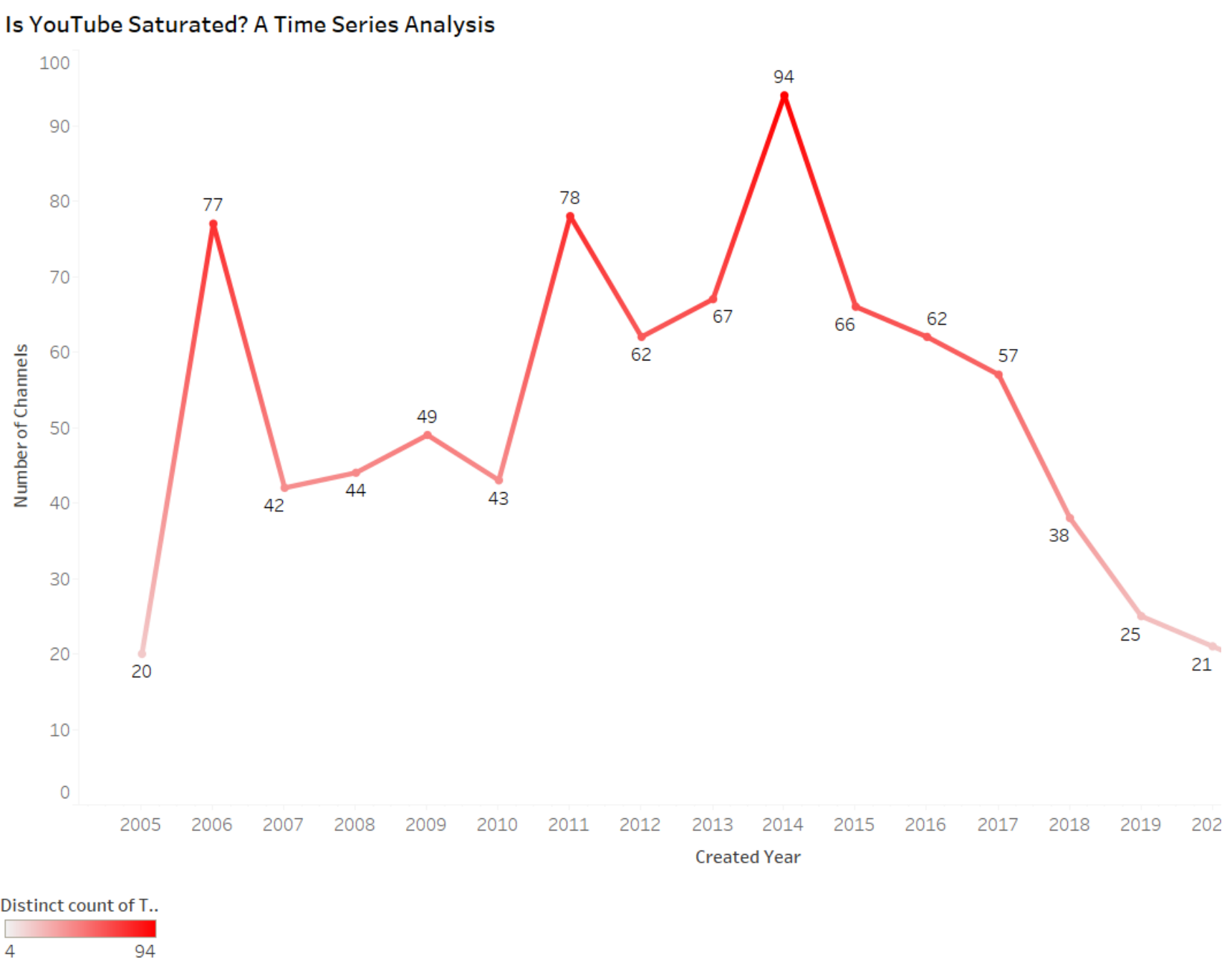
While the "Entertainment" category has the most subscribers, the "Music" category has the highest estimated yearly earnings. The "Comedy" and "People & Blogs" categories also have a large subscriber base. The "How to & Style" category has a relatively lower subscriber base and estimated yearly earnings than other categories.

tent Cosmos



YouTube's Content Cosmos

In this visualization, we depict the distribution of YouTube channel categories based on the quantity of videos within each category. Each bubble's size corresponds to the percentage of total videos attributed to that category, with larger bubbles representing higher quantities and smaller bubbles indicating fewer videos. The colors distinguish between categories, though the legend is omitted from the image.



Time Series Analysis of YouTube Channels Created by Year

- The line chart illustrates the yearly creation of YouTube channels spanning from 2005 to 2022, with columns representing the number of channels and rows denoting creation years. Darker shades indicate higher channel counts. Peak channel creation transpired in 2014 (94 channels), trailed by 2006 (77) and 2011 (78). A consistent decline in creation emerged post-2016. 2022 recorded the fewest channels (4), along with 2021 (17) and 2005 (20). The declining trend post-2016 may stem from YouTube's increasing saturation, intensifying competition among millions of channels vying for audience attention, possibly hindering new creators' ability to garner viewership.

Global Audience

Displayed is a table categorizing YouTube channel subscribers by country. Each row denotes a country, with the value indicating the total subscribers in millions for channels originating from that country (e.g., India boasts 4.3 billion subscribers). Notably, India leads with the highest subscriber count, followed by the United States. Indonesia, Brazil, and Mexico also feature prominently in terms of subscriber numbers.

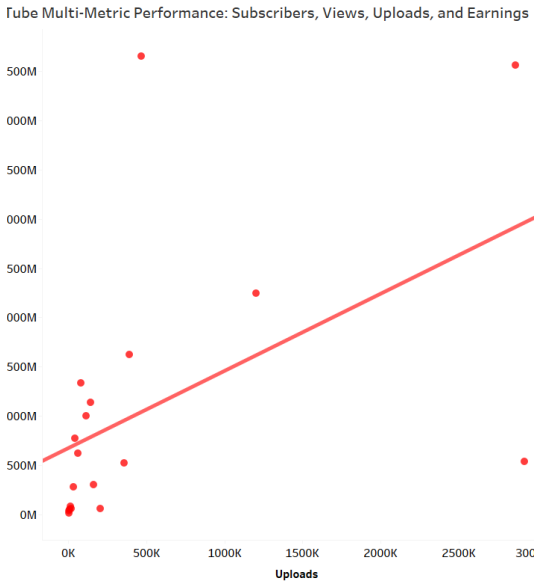
Global Audience

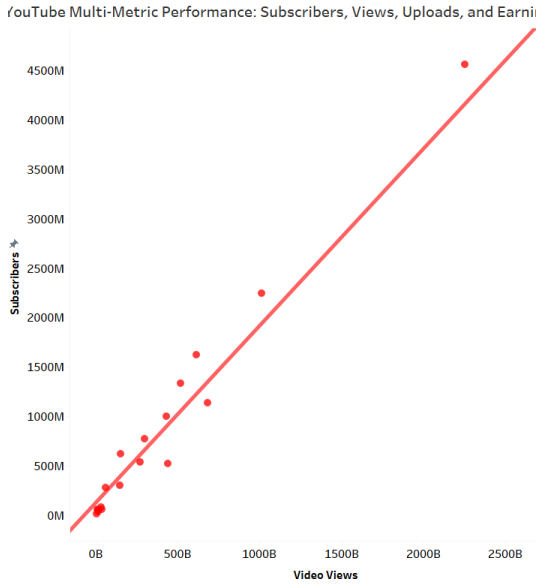
Country	Subscribers
United States	7,315M
India	4,311M
Brazil	1,207M
United Kingdom	917M
Mexico	627M
Indonesia	534M
South Korea	482M
Russia	407M
Thailand	387M
Spain	385M
Canada	364M
Argentina	329M
Colombia	265M
Philippines	240M
Japan	189M
Saudi Arabia	179M
Pakistan	155M
Australia	154M
United Arab Emirates	148M
Ukraine	126M
Germany	116M
Turkey	92M

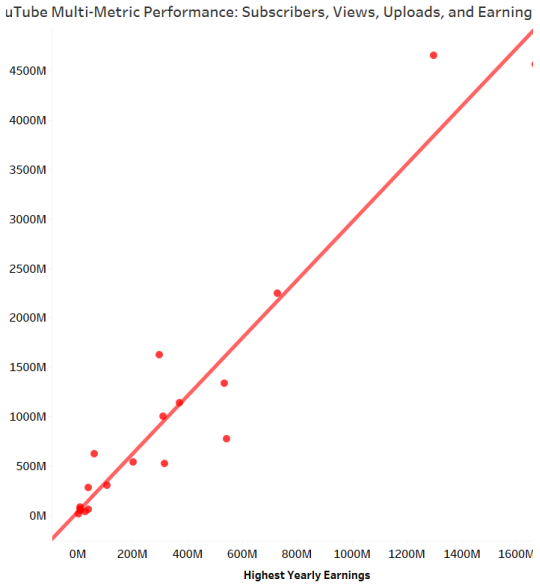
Subscribers
13M 7,315M

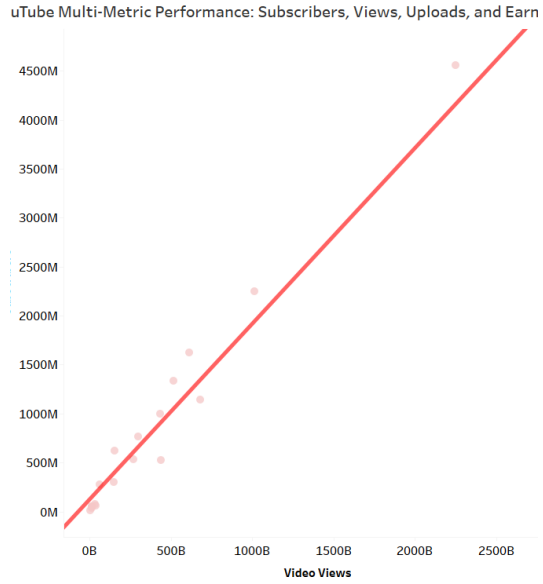
YouTube's Geographical Spread

India and the United States stand out with the highest YouTube channel subscriber counts, with India notably leading by a substantial margin. There appears to be no evident correlation between the quantity of channels (Distinct count of Title) and the total subscriber count. For instance, the United States boasts a significantly larger subscriber base than Brazil, despite having fewer channels









YouTube Channel Landscape

YouTube Channel Analysis Tool

- Interactive scatterplot for analyzing YouTube channel performance metrics across different categories.
- Visualizes data points, each representing a specific YouTube channel.

Subscribers vs. Video Views

- Strong positive correlation observed.
- More subscribers typically lead to more video views.
- Mathematical model provided; intercept may not be statistically significant.

Subscribers vs. Yearly Earnings

- A strong positive correlation was noted between subscribers and the estimated highest yearly earnings.
- Channels with more subscribers have higher earning potential.
- Model includes a mathematical equation, with the intercept (earnings at zero subscribers) potentially insignificant.

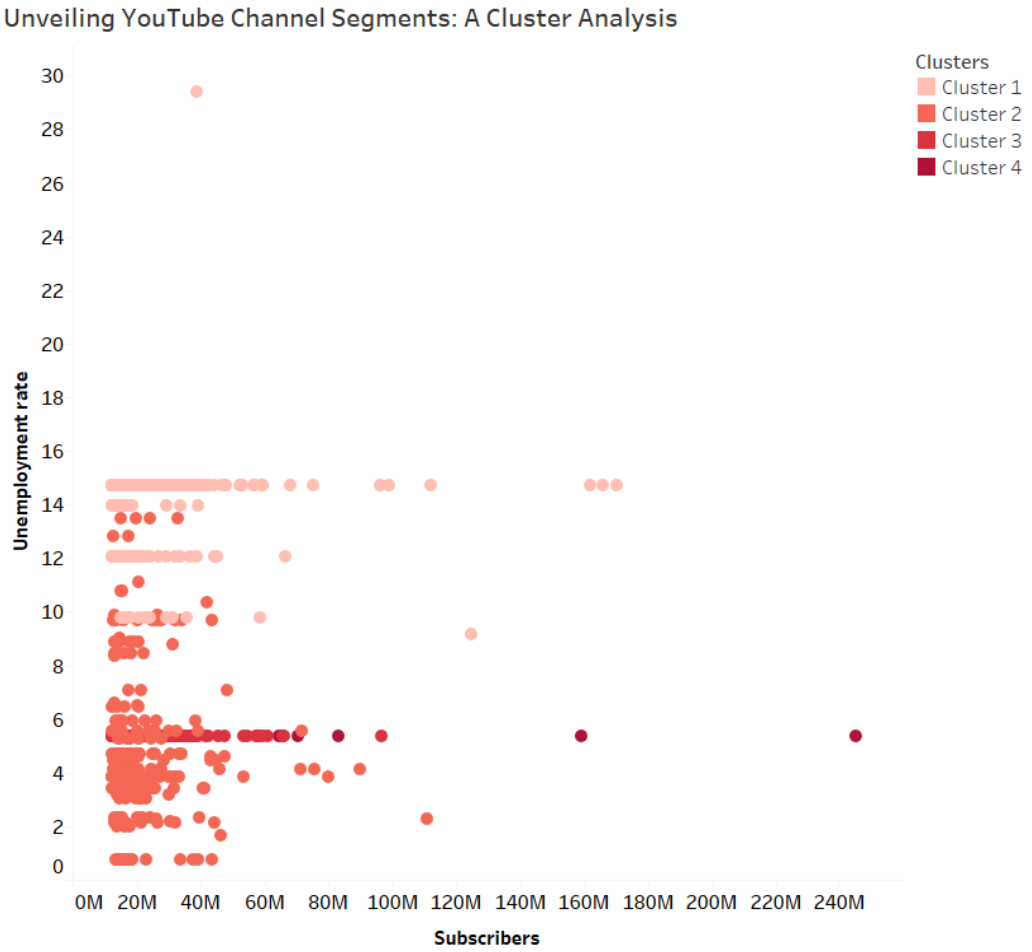
Subscribers vs. Uploads

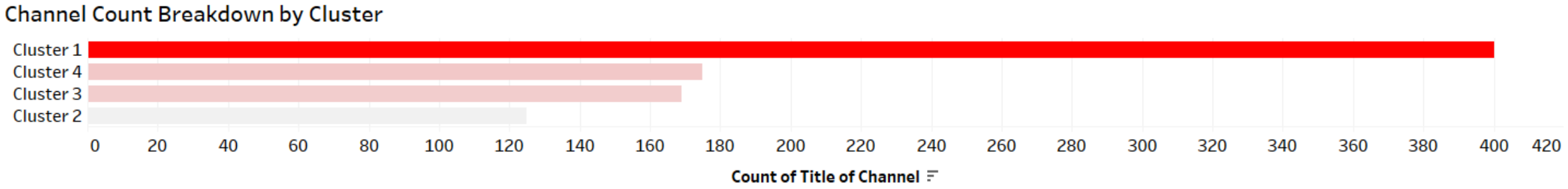
- Positive but weaker correlation compared to views and earnings.
- P-value: 0.0284, indicates a statistically significant but weak correlation.
- R-Squared: 0.2524, suggests 25.24% of subscriber variation is explained by uploads, with other factors influencing the remaining 74.76%.

Model 1 Model 2 Model 3

YouTube Channel Segments : K-Means Clustering

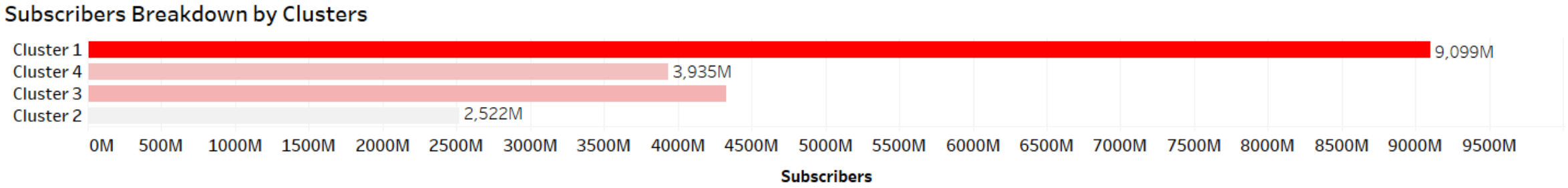
This cluster analysis identifies similar groups of YouTube channels using nine distinct metrics. It provides valuable insights for marketers and content creators, helping them understand various audience segments and adapt their content strategies effectively.





Viewing Clusters by Channels

This bar chart provides a summary of a cluster analysis of YouTube channels, categorizing them into four unique groups (Cluster 1, Cluster 2, Cluster 3, and Cluster 4) based on distinct characteristics. It illustrates the distribution of channels across each cluster, with Cluster 1 containing the highest number of channels at 400, while Clusters 2, 3, and 4 each consist of a smaller, comparable number of channels.



Viewing Clusters by Subscribers

This bar chart explores the distribution of subscribers among YouTube channel segments. It segments channels into four distinct clusters (Cluster 1, Cluster 2, Cluster 3, and Cluster 4) according to their characteristics. Cluster 1 leads with a substantial subscriber base of 9 billion, followed by Clusters 2, 3, and 4 with 2.52 billion, 4.33 billion, and 3.93 billion subscribers, respectively. The chart effectively showcases the varied strengths in subscriber numbers across different channel clusters.

Demographics

Average Gross Tertiary Education

Cluster 1	32,871
Cluster 4	13,003
Cluster 3	4,771
Cluster 2	4,586

Average UnEmployment Rate

Cluster 1	14.21
Cluster 4	5.63
Cluster 3	5.35
Cluster 2	3.89

Average Population

Cluster 3	1,366.60M
Cluster 1	291.83M
Cluster 2	135.01M
Cluster 4	58.10M

Average Urban Population

Cluster 3	473,232,134
Cluster 1	241,755,037
Cluster 2	80,582,308
Cluster 4	47,363,889

Tertiary Education Rate: Reflects the average percentage of tertiary-educated individuals (college or university) in the viewer locations of each channel cluster, suggesting a potentially more educated audience.

Unemployment Rate: Indicates the average unemployment rate in the channel's viewer locations, which may influence viewer's disposable income and leisure time.

Population Size: This shows the average population in the viewer locations, providing insight into the potential reach of channels in each cluster.

Urban Population Percentage: This represents the average proportion of the urban population in each viewer location, relevant to understanding content preferences and internet access trends.

Dashboard Summary

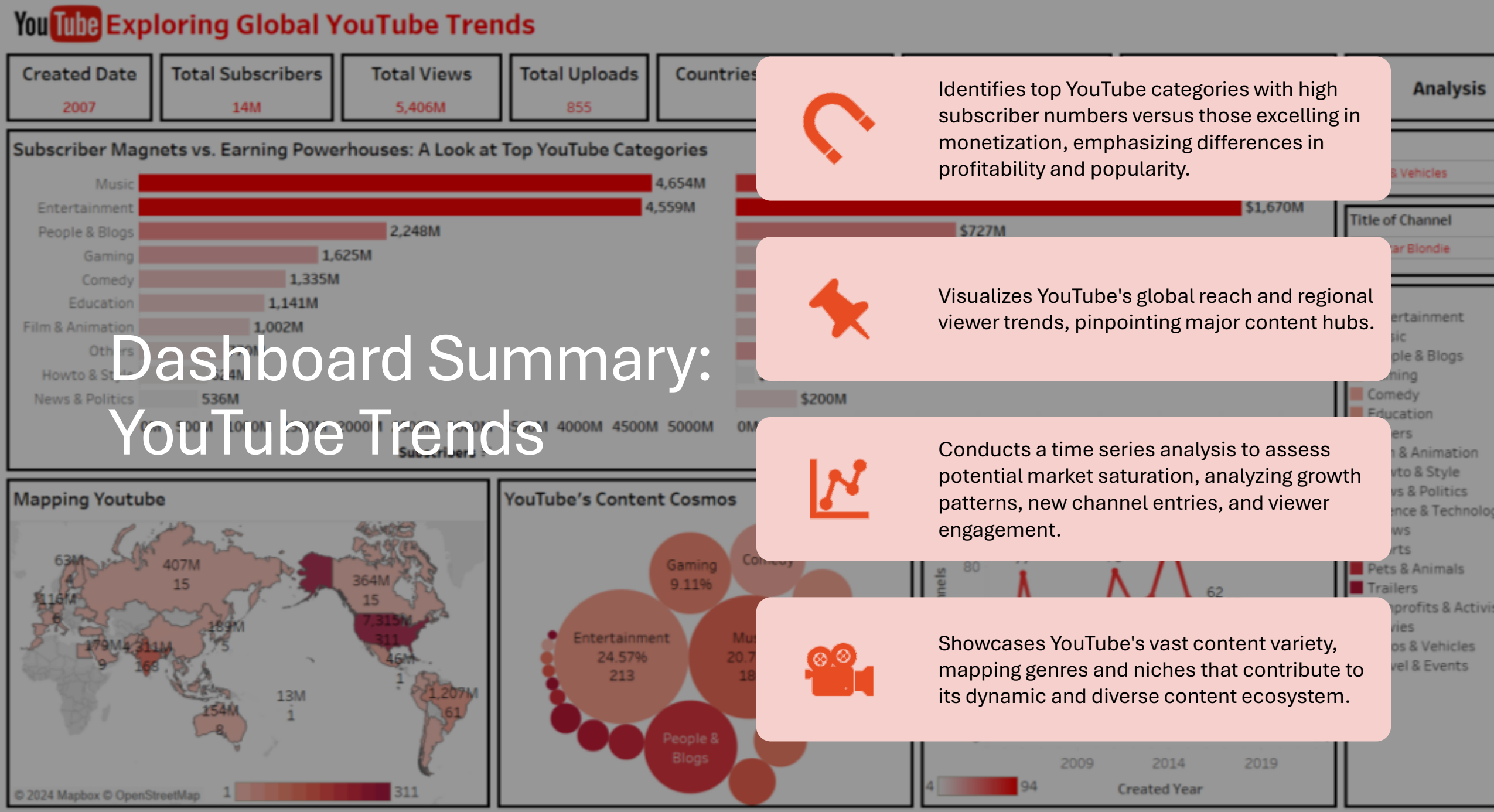
Dashboard Summary: YouTube Trends

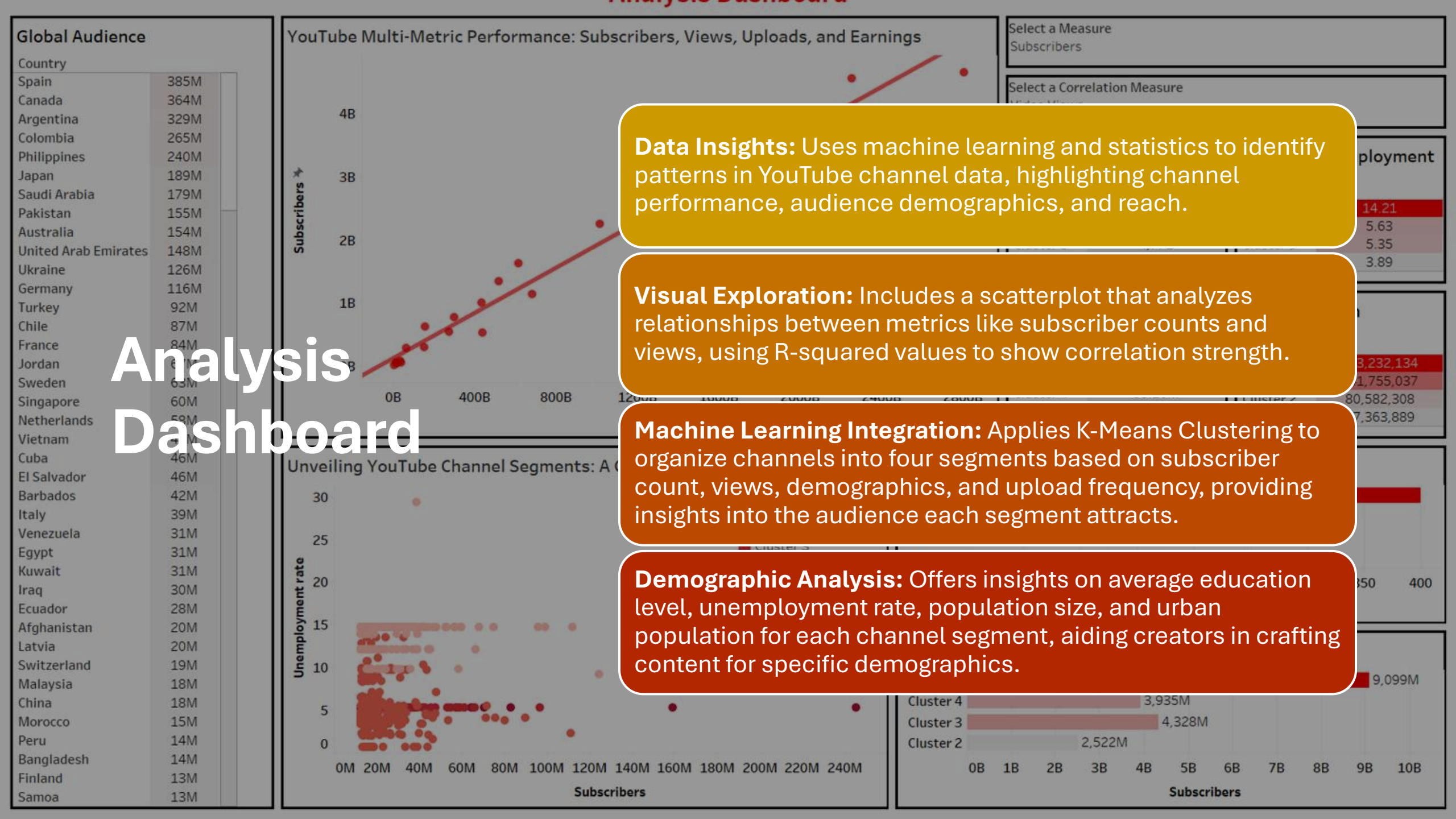
Identifies top YouTube categories with high subscriber numbers versus those excelling in monetization, emphasizing differences in profitability and popularity.

Visualizes YouTube's global reach and regional viewer trends, pinpointing major content hubs.

Conducts a time series analysis to assess potential market saturation, analyzing growth patterns, new channel entries, and viewer engagement.

Showcases YouTube's vast content variety, mapping genres and niches that contribute to its dynamic and diverse content ecosystem.





Analysis Dashboard

Data Insights: Uses machine learning and statistics to identify patterns in YouTube channel data, highlighting channel performance, audience demographics, and reach.

Visual Exploration: Includes a scatterplot that analyzes relationships between metrics like subscriber counts and views, using R-squared values to show correlation strength.

Machine Learning Integration: Applies K-Means Clustering to organize channels into four segments based on subscriber count, views, demographics, and upload frequency, providing insights into the audience each segment attracts.

Demographic Analysis: Offers insights on average education level, unemployment rate, population size, and urban population for each channel segment, aiding creators in crafting content for specific demographics.

Summary And Conclusion

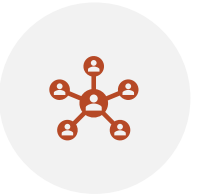
Top Categories: Entertainment channels top the subscriber count, with Music channels leading in estimated yearly earnings.

Content Distribution: Entertainment dominates video counts, with Music and People & Blogs following.

Channel Creation Trends: A steady decline in new channel creations since 2016 suggests possible market saturation.

Global Audience: India and the United States lead in subscriber numbers, with no direct correlation between channel count and subscriber base.

Subscriber Correlations: Strong positive correlations observed between subscribers and video views, highest yearly earnings, and uploads (though uploads show a weaker correlation).



Channel Segmentation: K-means clustering reveals four distinct channel segments based on subscriber count, views, demographics, and upload frequency, each with unique subscriber and audience profiles.

YouTube Trends Project Impact on Different Audiences

Content Creators and Marketers:

Strategic Insights: Gain an understanding of popular content categories and subscriber demographics for targeted content strategies.

Optimization Techniques: Utilize correlations between metrics to enhance video production and boost audience engagement.

Audience Segmentation: Apply cluster analysis for precise content tailoring based on audience segments.

Researchers:

Future Research Directions: Explore channel clustering and trends to study content consumption patterns and the evolving YouTube landscape.

Overall Project Success:

The project met its goals, providing an in-depth look at YouTube channel dynamics through detailed data visualization and analysis.

It offered critical insights into content types, audience demographics, and channel clustering, proving valuable to content creators, marketers, and researchers.