

What Makes a YouTube Channel Popular in Germany?

An Analysis of Subscriber Patterns Across Categories

Bingying Wang

April 30, 2025

Introduction

YouTube platform has evolved from a simple video-sharing platform into a dominant force shaping global media culture around the world. With billions of users worldwide, understanding the factors that make certain YouTube content categories popular is crucial for creators and marketers alike. This report explores a key question: **Which types of YouTube content are most popular in Germany, and are the differences in popularity statistically significant?** By answering this key question can provide valuable insights for content creators to refine their strategies and for businesses looking to effectively engage with target audiences in Germany's digital landscape.

In today's highly competitive content landscape, creators and businesses often highly struggle to capture audience attention effectively. By identifying popular categories or video channels, stakeholders can better allocate resources and strategies, potentially increasing viewership and engagement.

Dataset and Why It Matters

To investigate this question, I utilized the publicly available dataset titled **"Top 1000 Most Subscribed YouTube Channels in Germany"** (<https://www.kaggle.com/datasets/samithsachidanandan/1000-most-subscribed-youtube-channels-in-germany>) which is provided on Kaggle . The dataset includes key attributes necessary for in-depth analysis and explanation, such as:

- Channel name
- Subscriber count
- Video views
- Video count
- Content category
- Channel creation year

These attributes directly align with my research question, and the reason is because they are offering clear quantitative measures to analyze popularity and enabling detailed comparisons across content categories.

Data Collection and Cleaning

The dataset was sourced directly from Kaggle and publicly hosted on my GitHub repository for transparency and reproducibility. You can access it directly via this GitHub link (Or you can copy and paste this link: <https://raw.githubusercontent.com/berylchen3/ds2500-final-project/main/top-1000-most-subscribed-youtube-channels-in-germany.csv>).

Data cleaning was an essential step to ensure validity and accuracy. The cleaning process included:

- Removing improperly labeled categories, such as URLs mistakenly entered as category names.
- Converting subscriber counts and video counts from text strings into numeric values for precise statistical calculations.
- Ensuring consistent labeling of content categories.

Analytical Methods and Visualizations

To systematically examine the data, I conducted a multi-step analysis, including detailed exploratory data analysis (EDA), visualization of critical insights, and rigorous statistical tests.

Visualization 1: Average Subscribers per Category (Bar Chart)

To clearly illustrate the popularity of each content category, a bar chart was utilized (Fig. 1). Categories such as **Shows**, **Comedy**, and **Pets and Animals** showed notably higher average subscriber counts, indicating their broader appeal and consistent popularity among German audiences.

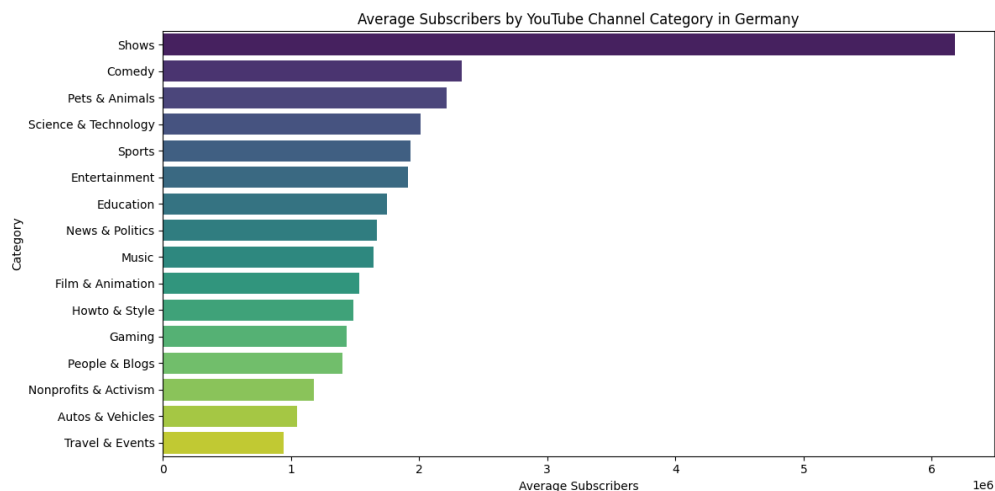


Figure 1: Average Subscribers by YouTube Channel Category in Germany

Visualization 2: Subscriber Count Distribution (Box Plot)

The box plot (Fig. 2) highlights variability within each category. It revealed substantial variance within the **Entertainment** category, indicating the presence of both exceptionally popular channels and moderately performing ones. Conversely, categories like **People & Blogs** demonstrated tighter distributions, suggesting a more consistent subscriber base across channels.

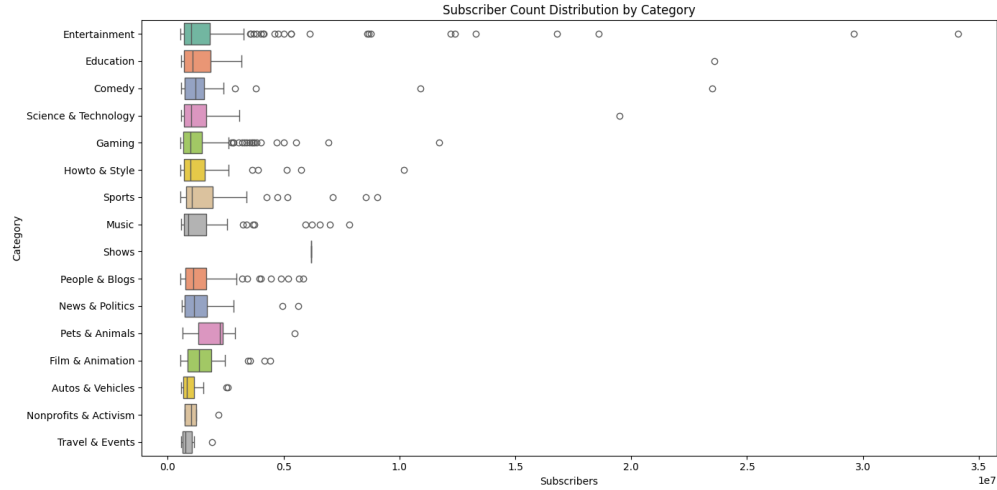


Figure 2: Subscriber Distribution by Category

Visualization 3: Relationship Between Video Count and Subscribers (Scatter Plot)

Exploring the potential correlation between the number of videos uploaded and subscriber count, a scatter plot (Fig. 3) was employed. Interestingly, the visualization suggested a weak relationship, implying that quality or content relevance could be significantly more influential than sheer quantity of videos.

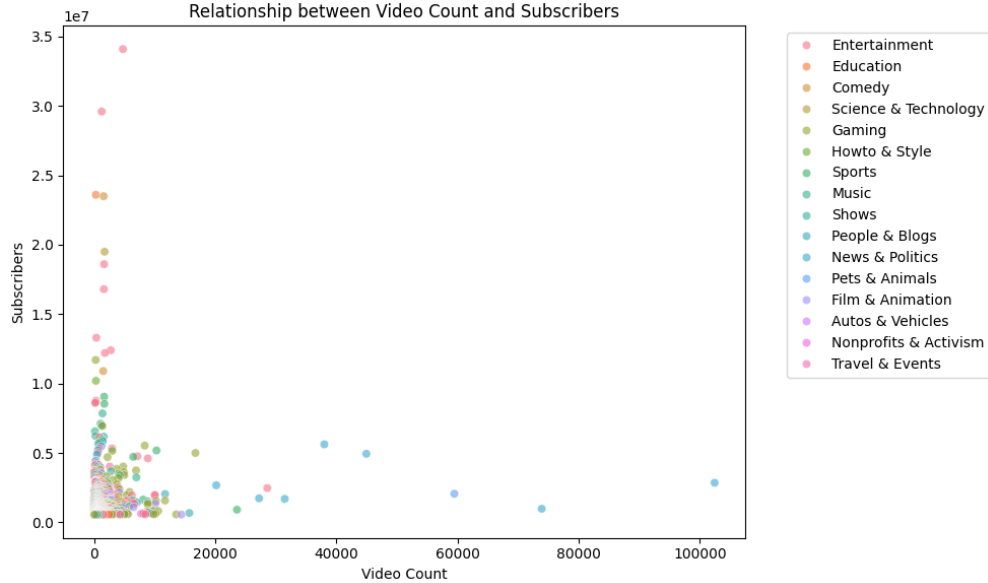


Figure 3: Relationship Between Video Count and Subscribers

Statistical Analysis

Test 1: Advanced Test - Linear Regression

To supplement the categorical analysis, I conducted a linear regression to examine whether the number of videos a channel has posted could predict its subscriber count. Although it's commonly assumed that producing more videos leads to more subscribers, the results of this analysis suggest otherwise.

The regression model yielded the following equation:

$$\text{subscribers} = 21.60 \times \text{video count} + 1,628,189.60$$

However, the R-squared value was only **0.0024**, indicating that the number of videos explains less than 1% of the variation in subscriber counts among top German YouTube channels. This suggests that other factors—such as content quality, topic relevance, engagement strategies, and possibly platform algorithmic boosts—may play a much larger role in channel popularity than video quantity alone.

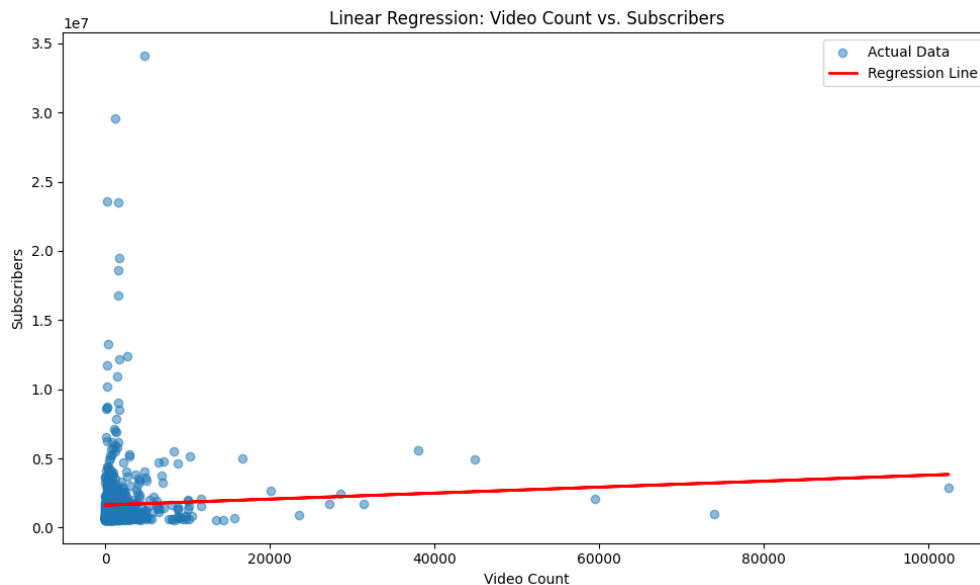


Figure 4: Linear Regression: Video Count vs. Subscribers

In short, while more videos may increase exposure, simply increasing quantity does not guarantee higher subscriber numbers. The relationship is statistically weak, and creators may be better off focusing on targeted and high-quality content rather than sheer volume.

Test 2: ANOVA Test

I conducted a one-way ANOVA test to evaluate whether subscriber counts significantly differed across YouTube content categories. The test returned an F-statistic of 1.0749 and a p-value of 0.375602. Since the p-value is well above the conventional threshold of 0.05, we fail to reject the null hypothesis. In other words, based on this dataset, there is no statistically significant evidence to suggest that average subscriber counts vary across categories.

To examine whether the ANOVA assumptions were reasonably met, I generated a Q-Q plot for one of the larger categories (Education), shown in Figure 5. The distribution shows mild skewness and an outlier,

but given the sample size and robustness of ANOVA, this deviation is not overly problematic. A Q-Q plot (Fig. 5) was used to verify normality assumptions, and while minor deviations were noted, ANOVA’s robustness accommodates this limitation adequately.

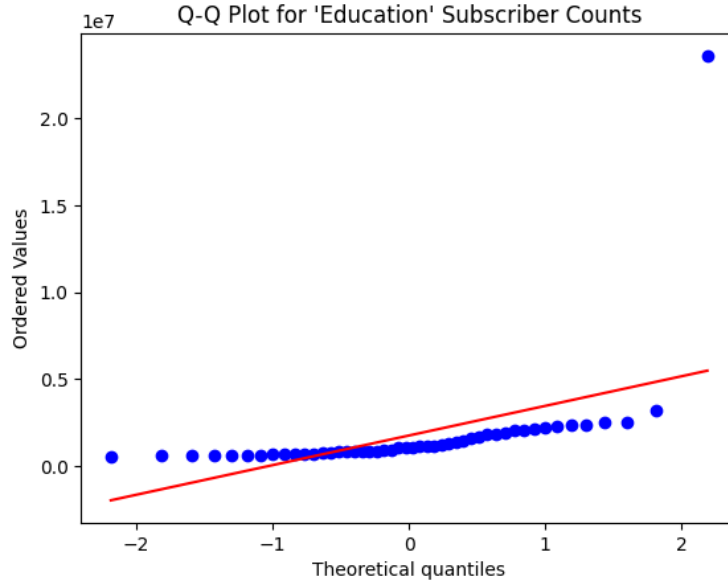


Figure 5: Q-Q Plot for Subscriber Counts in Education Category

Limitations of the Analysis

Despite the thorough analytical process, several limitations exist:

- The dataset captures only the top 1000 channels, excluding smaller creators and potentially biasing results toward already successful channels.
- Subscriber counts alone may not fully represent user engagement or viewer satisfaction.
- Important external factors such as YouTube’s recommendation algorithms, promotional activities, and viewer demographics were not available or considered.

Future analyses could benefit from incorporating these variables, providing a more holistic understanding of channel success factors.

Conclusions and Practical Insights

This comprehensive analysis confirmed significant popularity variations among YouTube categories in Germany. Entertainment-oriented categories dominate subscriber counts, suggesting content creators who focus on entertainment and education are likely to attract larger audiences.

For creators and businesses, this implies strategic advantages in focusing content around these highly appealing categories. Moreover, quality and audience engagement appear more influential than content volume alone, highlighting the importance of targeted, well-produced content.

Resources for Reproducibility

All analyses and visualizations presented here were conducted using Python in Google Colab. The complete and reproducible notebook with all code and detailed explanations can be accessed at the following link: [Colab Notebook](#).