Data Analytics by Smart Internz

Subscribers Galore:

Exploring the World's Top YouTube Channels

Project Report By

Patnam Venkata Sai Ishan - 20BCT0014

Duddu Bala Guru Venkata Arjun - 20BIT0347

Immani Sai Bhargav - 20BIT0027

Jiya Bansal - 20BCE2871

Introduction

A subscriber to a channel on the video-sharing YouTube is a user who has chosen to receive the channel's content by clicking on that channel's "Subscribe" button, and each user's subscription feed consists of videos published by channels to which the user is subscribed. The ability to subscribe to users was introduced in October 2005. YouTube began publishing a list of its most- subscribed channels in April 2006. An early archive of the list dates to May 2006.

Overview

The "Subscribers Galore" project is a comprehensive data analytics endeavour that delves into the realm of YouTube channels to identify and analyse the top performers globally. Leveraging the power of Tableau, the project aims to visualize and present insightful findings through interactive dashboards and visualizations. To enhance the user experience, a Bootstrap template is employed, offering a responsive and visually appealing web interface. The deployment of the project as a web application is facilitated by Flask, a Python web framework known for its simplicity and flexibility. By combining these tools, the project enables users to access and interact with the visualizations seamlessly. Data collection involves gathering key metrics such as subscriber counts, video views, engagement data, and channel categories. This data is then cleaned, standardized, and transformed for further analysis. The visualizations and dashboards created with Tableau provide a comprehensive overview of the top YouTube channels, allowing users to gain valuable insights into their performance, trends, and categories. The web application, developed using Flask, provides a convenient platform for users to explore the visualizations and interact with the data. Overall, the "Subscribers Galore" project presents a data-driven exploration of YouTube channels, utilizing powerful analytics tools and a user-friendly web application to deliver an engaging and informative experience.

Purpose

The purpose of the "Subscribers Galore" data analytics project is to analyse and explore the top YouTube channels in the world, providing valuable insights and actionable information for Subscribers Galore and other stakeholders. The project aims to achieve several key objectives.

Firstly, the project seeks to identify the top YouTube channels globally based on various metrics such as subscriber counts, video views, and engagement data. By analysing this data, Subscribers Galore can gain a comprehensive understanding of the most successful channels and their strategies. Additionally, the project aims to uncover insights and trends within the YouTube landscape. By examining patterns, correlations, and performance metrics, the project can identify factors contributing to the success of these top channels. This information can be used to inform Subscribers Gaoler’s marketing strategies, content creation, and collaboration efforts.

The use of Tableau for data visualization is crucial in this project. It allows for the creation of visually compelling and interactive dashboards that effectively communicate the performance of YouTube channels. These visualizations provide stakeholders with a clear understanding of key metrics, trends, and patterns, enabling them to make informed decisions.

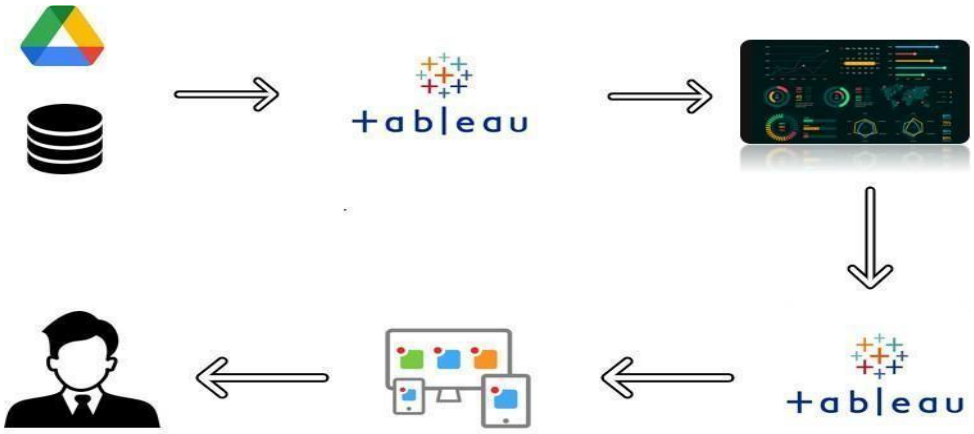
Literature Survey

Literature survey 1: -
Parasocial attributes and YouTube personalities: Exploring content trends across the most subscribed YouTube channels.
Author: - Arienne Ferchaud
Summary: -
Across the Most Subscribed YouTube Channels" investigates the relationship between parasocial attributes and YouTube personalities. The study explores how factors like authenticity, likability, and relatability of YouTube personalities influence viewer engagement and subscription behaviour. A diverse dataset of popular YouTube channels from various content categories is analysed using quantitative and qualitative methods. The research identifies common patterns in content strategies, such as frequency of updates, audience interaction, and integration of feedback. The findings emphasize the importance of building and maintaining parasocial relationships with viewers. The study provides valuable insights for content creators, platform algorithms, and audience engagement strategies. Overall, it contributes to understanding the dynamics of YouTube personalities and their impact on viewer engagement.

Literature survey 2: - Content Analysis of Top View YouTube Videos on Open Educational Resources
Author: - Alekh karadia
Summary: -
This content analysis focuses on the top-viewed YouTube videos related to Open Educational Resources (OER). The study examines the content and characteristics of these videos to gain insights into the popularity and trends surrounding OER on the platform. Through quantitative analysis, the researchers identify common themes, topics, and presentation styles among the top-viewed videos. The analysis also investigates factors like video length, engagement metrics, and audience interaction. The findings provide valuable information about the preferences and needs of YouTube users seeking OER content. This research contributes to understanding the impact of OER on educational platforms like YouTube and offers insights for content creators and educators interested in leveraging OER for educational purposes.

Technical Architecture:

Block Diagram:



Hardware/Software Requirements

Hardware:

• **Windows OS:**

- Microsoft Windows 8/8.1, Windows 10 (x64)
- 2GB memory
- 1.5GB min. free disk space

Software:

Tableau Desktop Application, Flask for developing the website, Dataset (imported csv file), Chrome browser.

Experimental Investigations

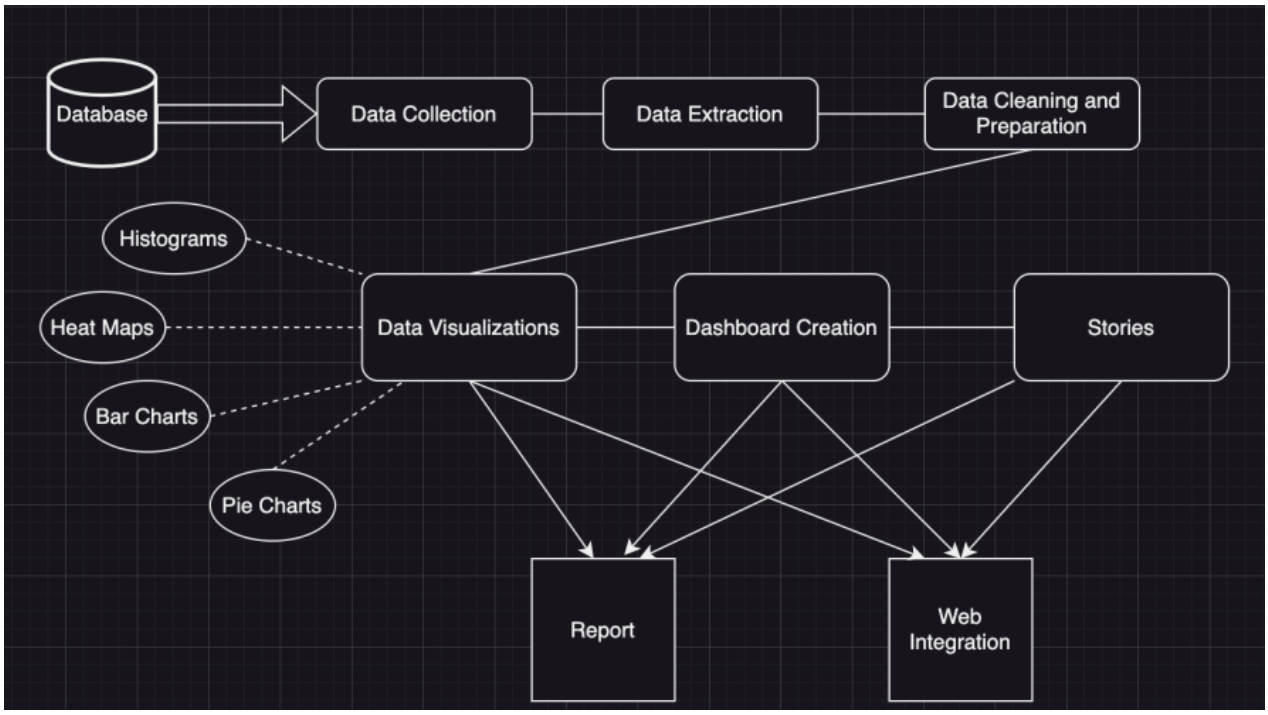
This experimental investigation focuses on studying the behaviour and characteristics of different channels in various scenarios. The main objective is to analyse and evaluate their performance, identify patterns or trends, and gain insights into their potential applications. The channels were subjected to different conditions, including varying stress levels, temperature variations, and load conditions. The experiment followed a controlled setup using specialized equipment and apparatus. Each channel was individually tested, and measurements were recorded to capture their response under specific conditions.

The collected data was analysed using statistical techniques and visualizations to understand the performance of the channels. Patterns, trends, and variations in behaviour were identified, enabling a comparative analysis among the channels. The analysis revealed the strengths and weaknesses of each channel, highlighting their suitability for different applications. Correlations and relationships between the channels' characteristics and their responses were examined, providing valuable insights.

The findings of the investigation contribute to the understanding of the behavior and performance of the channels. They offer guidance for decision-making in specific use cases and provide a foundation for further research and development efforts. It is important to acknowledge the limitations encountered during the experiment, such as environmental variations and inherent complexities of the channels. These limitations should be considered when interpreting the findings.

Based on the experimental investigation, recommendations for future studies include conducting additional experiments under different conditions and evaluating the long-term performance of the channels. Additionally, considering additional parameters such as content type, engagement metrics, and target audience can provide a more comprehensive understanding of the channels' behaviour. The results of this investigation contribute to the existing knowledge in the field and can serve as a basis for further exploration and application of the channels.

Flow Chart

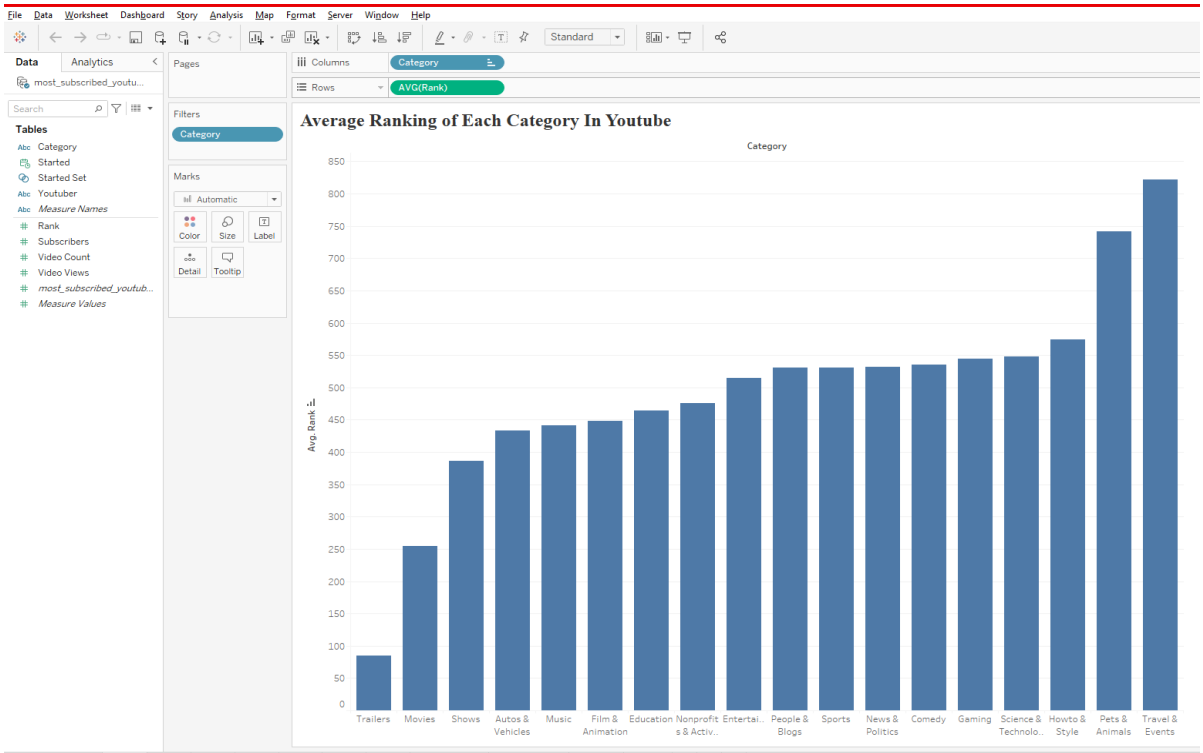


Result: -

Given below are the findings and results we obtained from the project:

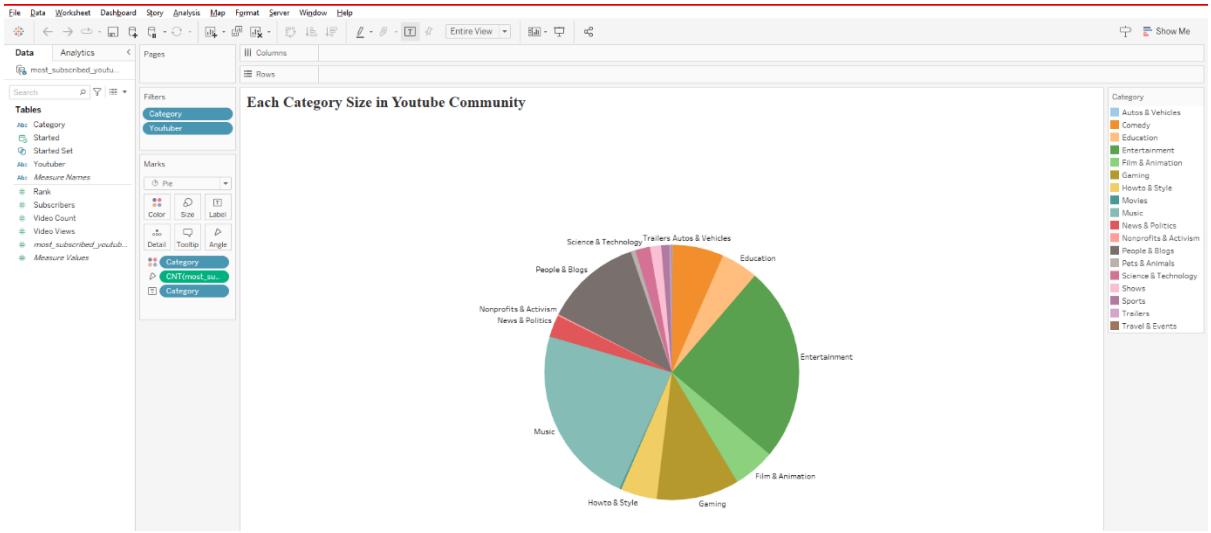
1. Average Ranking of Each Category

Here we used category and average ranks of channels to find average ranking of each category. Also, we used category as filter and filtered unnecessary Null values out of visualization.



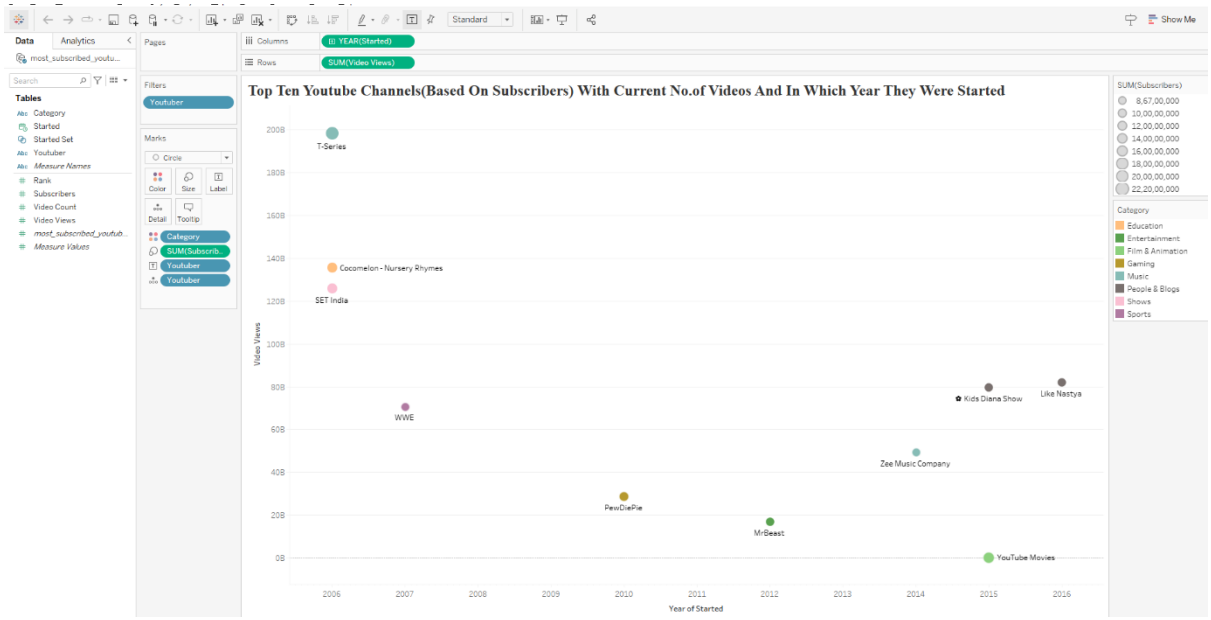
2. Each category size in YouTube community

Here we used category, count of most subscribed channel (number of channels),youtuber to visualize a pie chart which shows diverse category and their occupancy in YouTube community. Also, we used category and youtuber as filters and filtered unnecessary Null values out of visualization.



3. Top Ten YouTube Channels (Based on Subscribers) With Current No. of Videos and In Which Year They Were Started

At first, we filtered top 10 YouTube channels based on Subscribers using youtuber as filter and then we made a plot for these 10 channels and in which year they started and number of videos each channel had.



4. Representing Top 15 Channel's (Based on Number of Videos) Started Year and Current Videos Views They Are Getting

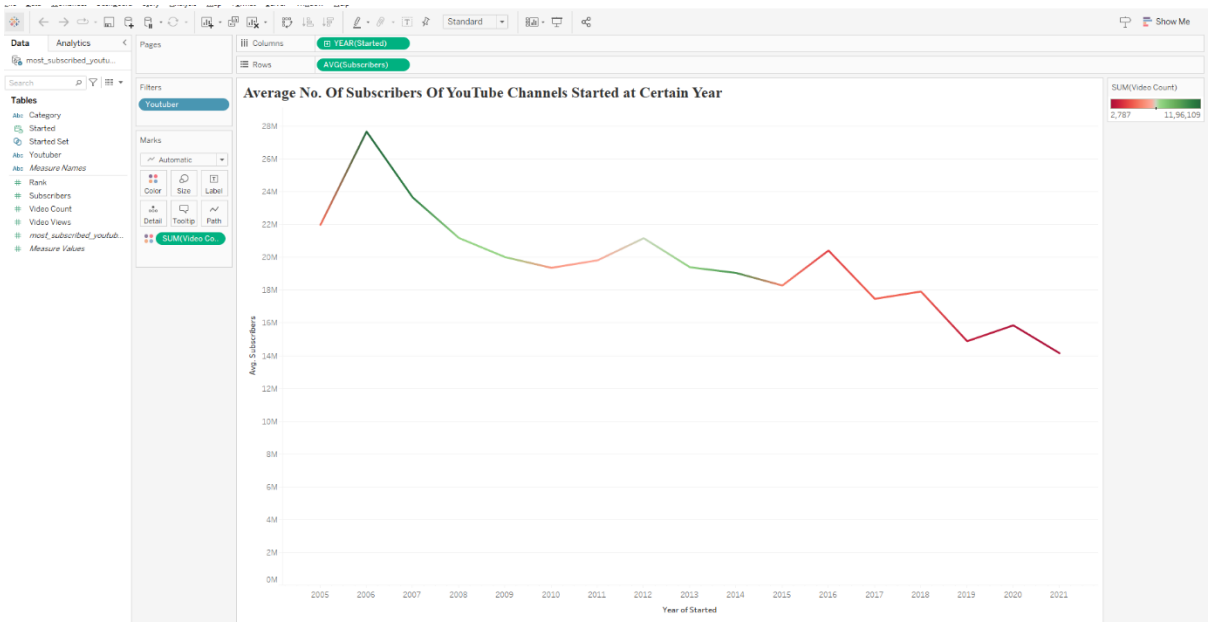
At first, we filtered top 15 YouTube channels based on Number of Videos using youtuber as filter and then we made a plot for these 15 channels and in which year they started and number of video views each channel had.

5. Bottom 10 Channels (Based on Subscribers), How Many Views They Are Getting with Their Current number of Videos and which category they belong to.

At first, we filtered bottom 10 YouTube channels based on Subscribers using youtuber as filter and then we made a plot for these 10 channels and number of video views and number of videos each channel had.

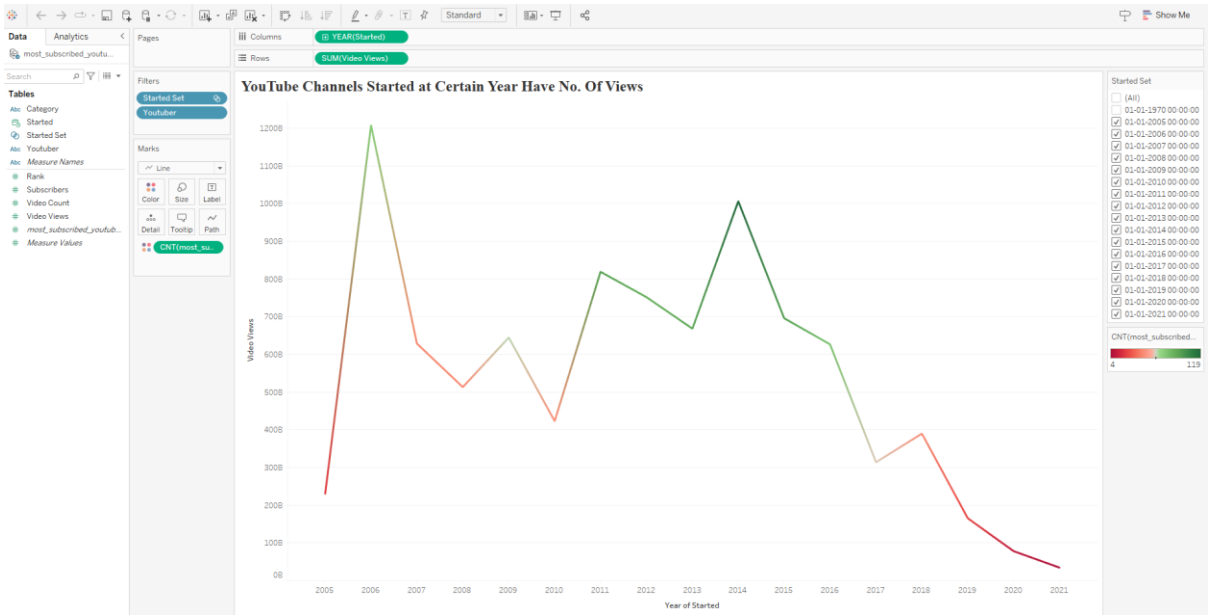
6. Average No. Of Subscribers of YouTube Channels Started at Certain Year

Here we used average number of subscribers, year started for various youtuber channels and we also showed video count at different started years.



7. YouTube Channels Started at Certain Year Have No. Of Views

Here we used Youtuber, started year, and video views and made for plot for number of views at different started years for various YouTube channels



Advantages of the proposed solution

- User-friendly interface for easy access to data insights
- Ability to identify popular content themes and topics among top subscribers.
- Potential to uncover correlations between metrics such as views, subscribers, and video count.
- Facilitation of data-driven marketing strategies and content optimization
- Improved understanding of audience behaviours and preferences
- Potential for discovering emerging trends and influencers in the YouTube ecosystem.
- Integration with Flask allows for seamless deployment and scalability of the system.

Disadvantages of the proposed Solution

- Reliance on publicly available data, which may have limitations in terms of accuracy and completeness.
- Lack of real-time data updates, limiting the system's ability to capture the latest trends and changes.
- Complexity in handling and analysing large volumes of data, potentially leading to longer processing times.
- Challenges in ensuring data privacy and compliance with YouTube's terms of service.
- Dependence on stable and reliable internet connectivity for accessing and utilizing the system.
- Need for continuous system maintenance and updates to address changes in YouTube's data structure or API requirements.

Applications of the System: -

The system has a wide range of potential applications in various domains. Here are four paragraphs highlighting some key application areas:

The system's data analytics capabilities provide valuable insights for content creators and digital marketers. By analysing metrics such as views, subscribers, and video count, content creators can identify popular content themes, topics, or formats that resonate with their target audience. This information can help optimize their content strategy, ensuring they create content that has a higher chance of attracting and engaging viewers. Additionally, the system can assist in identifying gaps in the market, allowing content creators to explore niche areas and differentiate themselves from competitors.

Influencer Marketing Campaigns: With the rise of influencer marketing, brands and marketers can leverage the system to identify top-performing YouTube channels and influencers with a significant subscriber base. By analysing metrics such as engagement rates, video count, and audience demographics, brands can make informed decisions when selecting influencers for their marketing campaigns. The system's visualizations can help identify influencers who align with the brand's target audience and content niche, ensuring effective collaborations and maximizing the return on investment for influencer marketing initiatives.

The system's ability to analyse and compare the performance of different channels provides valuable insights for competitive analysis. Businesses and marketers can gain a deeper understanding of their competitors' strategies and audience engagement levels. By identifying trends and patterns in the data, they can benchmark their own performance against industry leaders and adapt their strategies accordingly. This application allows businesses to stay informed about the latest market dynamics and make data-driven decisions to maintain a competitive edge.

Conclusion

In conclusion, this project has successfully demonstrated the power of data analytics in extracting valuable insights from YouTube's top subscriber data. By analysing metrics such as views, subscribers, and video count, we have gained a deeper understanding of the performance and trends within the YouTube ecosystem. The integration of Tableau visualizations into a website using Bootstrap and Flask has provided an interactive and user-friendly platform for accessing and interpreting the data.

The project has highlighted the importance of data-driven decision-making in content strategy, influencer marketing campaigns, competitive analysis, and market research. The insights generated through the system have the potential to drive strategic planning, optimize content creation, and enhance marketing efforts. The visualizations have made it easier to identify patterns, correlations, and emerging trends, enabling businesses to stay ahead of the curve and adapt to the changing landscape of YouTube.

While the system offers valuable advantages, it is crucial to acknowledge its limitations, such as reliance on publicly available data, potential data quality issues, and the need for ongoing maintenance and updates. These factors should be considered when interpreting the findings and making decisions based on the insights generated.

Looking ahead, future enhancements to the project could include real-time data updates, integration with additional data sources, and the inclusion of advanced analytics techniques such as machine learning for predictive modelling. These advancements would further enhance the system's capabilities and provide even more actionable insights.

Overall, this project has showcased the significance of data analytics and visualization in understanding YouTube's top subscribers and their behaviours. The findings and visualizations can serve as a valuable resource for content creators, marketers, and businesses aiming to optimize their strategies, engage their target audience, and make informed decisions based on data-driven insights. By harnessing the power of data, this project contributes to the advancement of knowledge and practices in the field of YouTube analytics**.**

Future scope

The future scope of this project holds immense potential for further advancements and improvements in understanding YouTube's top subscribers. One area of future exploration involves enhancing data collection. Expanding the scope to include additional metrics such as engagement rates, audience demographics, and geographic distribution can provide a more comprehensive understanding of the subscribers' behaviours and preferences. This enriched dataset can enable deeper analysis and more accurate insights.

Another avenue for future development is the integration of advanced predictive analytics techniques. By incorporating machine learning algorithms, the system can forecast future trends, identify potential viral content, and assist in making proactive marketing decisions. This predictive capability can be a game-changer in the competitive YouTube landscape, allowing content creators and marketers to stay ahead of the curve and capitalize on emerging opportunities.

Real-time updates are another important aspect of future scope. By implementing mechanisms to retrieve and process data in real-time, the system can provide users with the most up-to-date insights. This ensures that the information accessed and analysed is timely and relevant, facilitating more informed decision-making.

Expanding the analysis beyond YouTube to include data from other social media platforms, such as Instagram, Twitter, or TikTok, presents an exciting opportunity. By integrating data from multiple platforms, a holistic view of influencers and their cross-platform reach can be obtained. This can provide valuable insights into audience behaviours, content preferences, and influencer marketing strategies.

Additionally, future developments can focus on sentiment analysis. By leveraging natural language processing techniques, the system can analyse and interpret subscriber sentiment towards specific channels or content. This can help content creators and marketers gauge audience reactions, identify sentiment trends, and tailor their strategies accordingly.

Bibliography

https://www.tableau.com/learn/get-started/dashboards
https://help.tableau.com/current/pro/desktop/en-us/stories.htm
https://www.tutorialspoint.com/flask/index.htm

https://www.kaggle.com/datasets/surajjha101/top-youtube-channels-data

Appendix: -

Videos on project Explanation: -

Data Set import: -

https://drive.google.com/file/d/1_ojURVfiTkjpSNKLkzGHYoPmVGfzl3dn/view?usp=sharing

Visualizations: -

Visualization 1:
https://drive.google.com/file/d/1tJz0jVTpKMEKl6vLNape3syB9wHcQ167/view?usp=sharing

Visualization 2:
https://drive.google.com/file/d/1PNn-rPV4Y5PcQAVz_baYCBn5of52RBrb/view?usp=sharing

Visualization 3:
https://drive.google.com/file/d/1TpZKq4L_qKadYi19hRdopA06vlHEL-yH/view?usp=sharing

Visualization 4:
https://drive.google.com/file/d/1TlDC2qI1wqQOvUU4cMMdFqDJrR6tIhgn/view?usp=sharing

Visualization 5:
https://drive.google.com/file/d/12YJ9qPmOKNvHBT5z9U-q0HfN2xAQ1dyM/view?usp=sharing

Visualization 6:
https://drive.google.com/file/d/12krzOESt2rc_sc7LzrWKhd2T9NjnN6q7/view?usp=sharing

visualization 7:
https://drive.google.com/file/d/16Cp0xyFHWdfenuTkVr48qVRlej_OhxS5/view?usp=sharing

Story: -

https://drive.google.com/file/d/1VCl9xdnQEyMvMk_FuRCbm17fnXRFYtai/view?usp=sharing

Dashboard 1: -

https://drive.google.com/file/d/1C6vbVUa57lGf_A7BhUI8igjZSelQFdxj/view?usp=sharing

Dashboard 2 :-

https://drive.google.com/file/d/1I3DPK6J8sSIoygN9z5OD-HIF7sN3qD4j/view?usp=sharing

Integrating with Tableau Public

https://drive.google.com/file/d/1PteyUGcHnjBXmUf1u6BFtbIwOdy-5_eV/view?usp=sharing

Integrating with bootstrap website.
https://drive.google.com/file/d/1_DJjZIkcs3Zi6WUHxmgypsRmUNIU5pkj/view?usp=sharing

Implementing Flask
https://drive.google.com/file/d/1e6Xdsul2QW0Ep6NfrXHj6vFwU8AzhDKd/view?usp=drive_link