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.

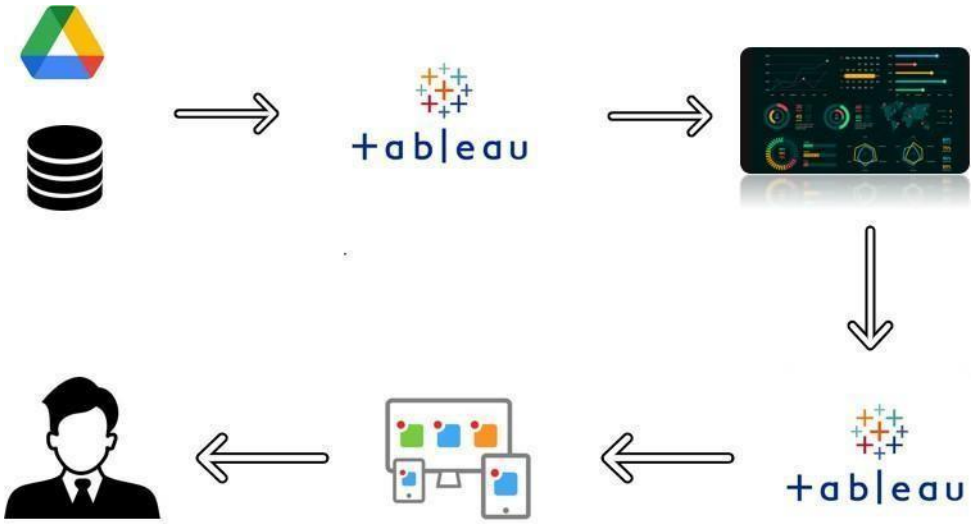
Firstly, the project seeks to identify the top YouTube channels globally based on various metrics such as subscriber counts, video views, and engagement data. By analysing this data, Subscribers Galore can gain a comprehensive understanding of the most successful channels and their strategies.

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.

Furthermore, incorporating a Bootstrap template enhances the user experience by providing a responsive and visually appealing web interface. Users can easily navigate and interact with the visualizations and dashboards, gaining valuable insights into the top YouTube channels.

Technical Architecture:



Business Problem

The business problem addressed by the "Subscribers Galore" data analytics project is the lack of understanding regarding the factors that contribute to the success of top YouTube channels globally. Subscribers' Galore needs insights into the content categories, subscriber counts, video views, and video counts of these channels to make informed decisions. Without comprehensive knowledge of these success factors, they struggle to identify potential channels for partnerships and accurately assess channel performance. The project aims to provide Subscribers Galore with valuable insights into the key metrics and trends of successful YouTube channels, enabling them to optimize their decision-making process, select suitable channels for promotion, and improve their overall competitiveness in the industry.

Business Requirements

Channel Information: Each YouTube channel should have a dedicated page or profile that displays essential information, including the channel name, description, subscriber count, video views, upload frequency, and engagement metrics.

Subscription Management: Users should have the ability to subscribe to their favourite YouTube channels within the platform, allowing them to receive notifications about new uploads and updates from those channels.

Analytics and Insights: The platform should provide analytics and insights to track user engagement, popular channels, and trending content. This data can be used to improve content curation, personalized recommendations, and enhance the overall user experience.

These business requirements aim to create a user-friendly platform that facilitates the exploration and discovery of YouTube channels, while also providing features for personalization, social interaction, and monetization. By meeting these requirements, the platform can enhance user engagement and satisfaction, while also creating opportunities for revenue generation and growth.

Literature Survey

A literature survey conducted by students exploring YouTube channels would typically involve researching existing studies, academic papers, and publications related to the topic.

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.

Social Or Business Impact

Social Impact:
The project could inspire aspiring content creators and YouTubers by showcasing the success stories of the world's top YouTube channels. It may encourage individuals to pursue their creative passions and build their own channels, leading to a more diverse and vibrant YouTube community.

Business Impact:
The project can provide valuable data and insights into the YouTube ecosystem, including trends, demographics, content formats, and strategies employed by successful channels. This information can be useful for industry professionals, marketers, and researchers in understanding the evolving landscape of online video content and optimizing their strategies.

Data Collection & Extraction from Database

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

Collect The Dataset

Activity 1.1: Understand the data

Check out the below link to understand the dataset in detail:
https://www.kaggle.com/datasets/surajjha101/top-youtube-channels-data

The contents of dataset typically include the following information about top YouTube channels.

This dataset contains information about the subscriber count, video count, viewership, and engagement metrics of the most popular YouTube channels. This dataset can be utilized to analyse the trends, patterns, and characteristics of the most popular YouTube channels.

Channel Title: The name of the channel on YouTube.
Category: The type of content that the channel is about.
Subscriber count: the number of people who have joined the channel.
Video Count: This shows how many videos have been uploaded to the channel.

View Count: The number of times each video on the channel has been watched.

Activity 1.2: Understand the data Dataset Import

Importing the dataset into our tableau desktop

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

Data Visualization

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

No Of Unique Visualizations

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyse the performance and efficiency of project include bar charts, line charts, heat maps, scatter plots, pie charts, Maps etc. These visualizations can be used to compare performance, track changes over time, show distribution, and relationships between variables.

Visualizations: -

1)Average Ranking of Each Category link:
Explanation Video link:
https://drive.google.com/file/d/1tJz0jVTpKMEKl6vLNape3syB9wHcQ167/view?usp=sharing

2) Each category size in YouTube community
Explanation Video link:
https://drive.google.com/file/d/1PNn-rPV4Y5PcQAVz_baYCBn5of52RBrb/view?usp=sharing

3) Top Ten YouTube Channels (Based On Subscribers) With Current No. of Videos And In Which Year They Were Started
Explanation Video link:
https://drive.google.com/file/d/1TpZKq4L_qKadYi19hRdopA06vlHEL-yH/view?usp=sharing

4)Representing Top 15 Channel's (Based On Number of Videos) Started Year and Current Videos Views They Are Getting
Explanation Video link:
https://drive.google.com/file/d/1TlDC2qI1wqQOvUU4cMMdFqDJrR6tIhgn/view?usp=sharing

5) Bottom 10 Channels (Based On Subscribers), How Many Views They Are Getting With Their Current number of Videos
Explanation Video link:
https://drive.google.com/file/d/12YJ9qPmOKNvHBT5z9U-q0HfN2xAQ1dyM/view?usp=sharing

6)Average No. Of Subscribers Of YouTube Channels Started at Certain Year
Explanation Video link:
https://drive.google.com/file/d/12krzOESt2rc_sc7LzrWKhd2T9NjnN6q7/view?usp=sharing

7) YouTube Channels Started at Certain Year Have No. Of Views
Explanation Video link:
https://drive.google.com/file/d/16Cp0xyFHWdfenuTkVr48qVRlej_OhxS5/view?usp=sharing

Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

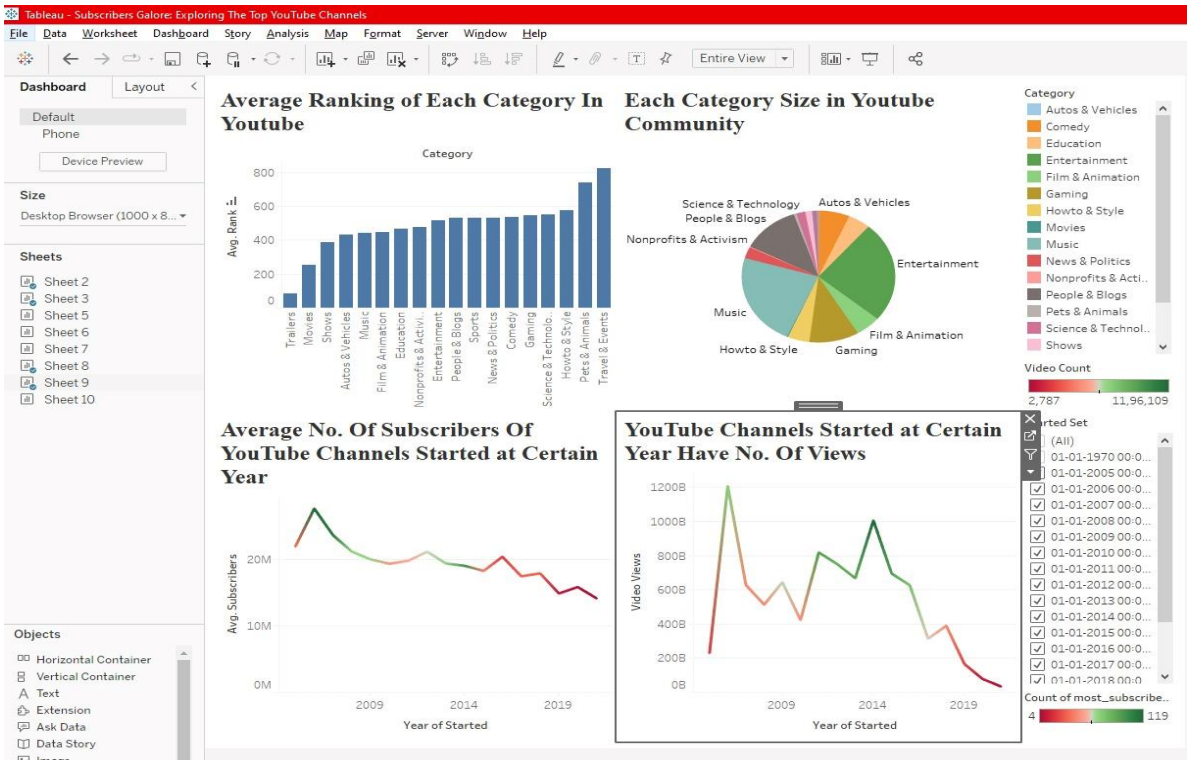
Responsiveness And Design of Dashboard

The responsiveness and design of a dashboard for Data-Driven insights on YouTube channels Analysis is crucial to ensure that the information is easily understandable and actionable. Key considerations for designing a responsive and effective dashboard include user-centred design, clear and concise information, interactivity, data-driven approach, accessibility, customization, and security. The goal is to create a dashboard that is user-friendly, interactive, and data-driven, providing actionable insights.

Dashboard 1

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

Dashboard 2



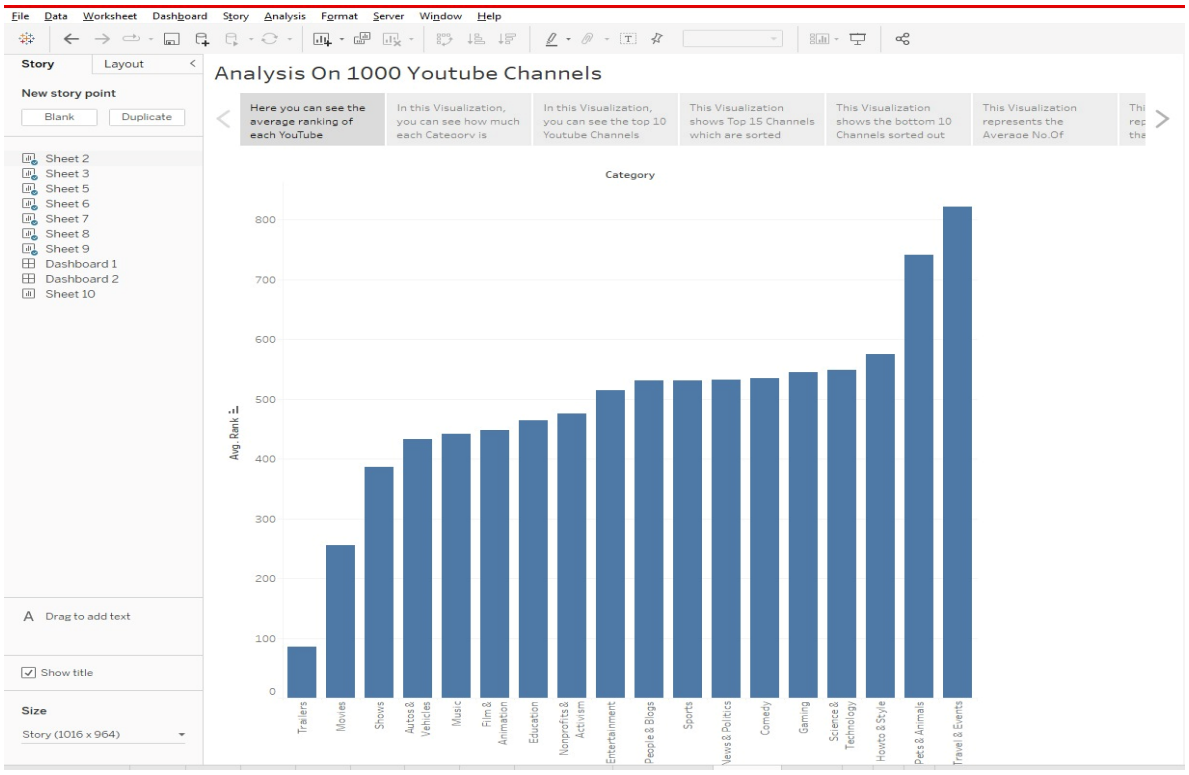
Explanation Video link:
https://drive.google.com/file/d/1I3DPK6J8sSIoygN9z5OD-HIF7sN3qD4j/view?usp=sharing

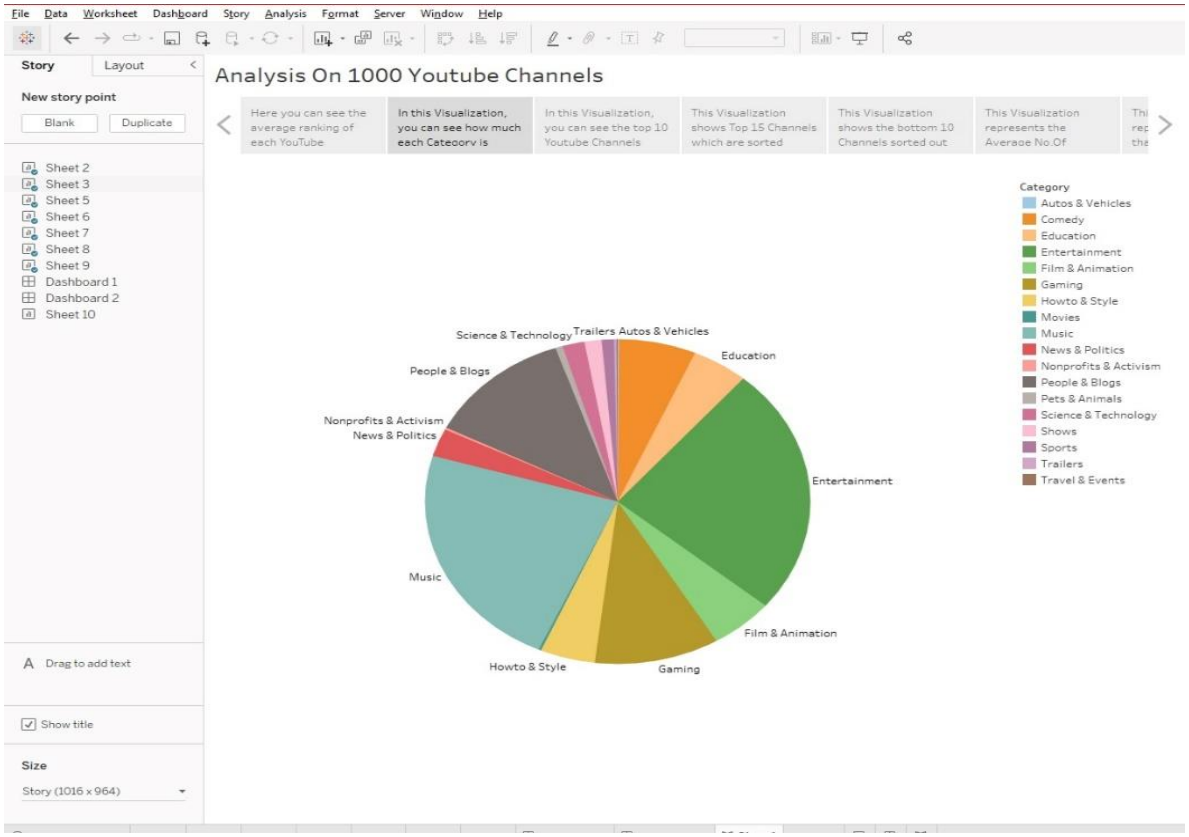
Story

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

No Of Scenes of Story

The number of scenes in a storyboard for Data-Driven insights on YouTube channels Analysis will depend on the complexity of the analysis and the specific insights that are trying to be conveyed. A storyboard is a visual representation of the data analysis process, and it breaks down the analysis into a series of steps or scenes.


File Data Worksheet Dashboard Story Analysis Format Server Window Help
Story Layout
New story point
Blank Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Category is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Category
Avg. Rank
0
100
200
300
400
500
600
700
800
Trailers
Movies
Shows
Autos & Vehicles
Music
Film & Animation
Education
Nonprofits & Activism
Entertainment
People & Blogs
Sports
News & Politics
Comedy
Gaming
Science & Technology
Howto & Style
Pets & Animals
Travel & Events


File Data Worksheet Dashboard Story Analysis Format Server Window Help
Story Layout
New story point
Blank Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Category is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Category
Autos & Vehicles
Comedy
Education
Entertainment
Film & Animation
Gaming
Howto & Style
Movies
Music
News & Politics
Nonprofits & Activism
People & Blogs
Pets & Animals
Science & Technology
Shows
Sports
Trailers
Travel & Events
Science & Technology
Trailers
Autos & Vehicles
Education
People & Blogs
Nonprofits & Activism
News & Politics
Entertainment
Music
Film & Animation
Howto & Style
Gaming

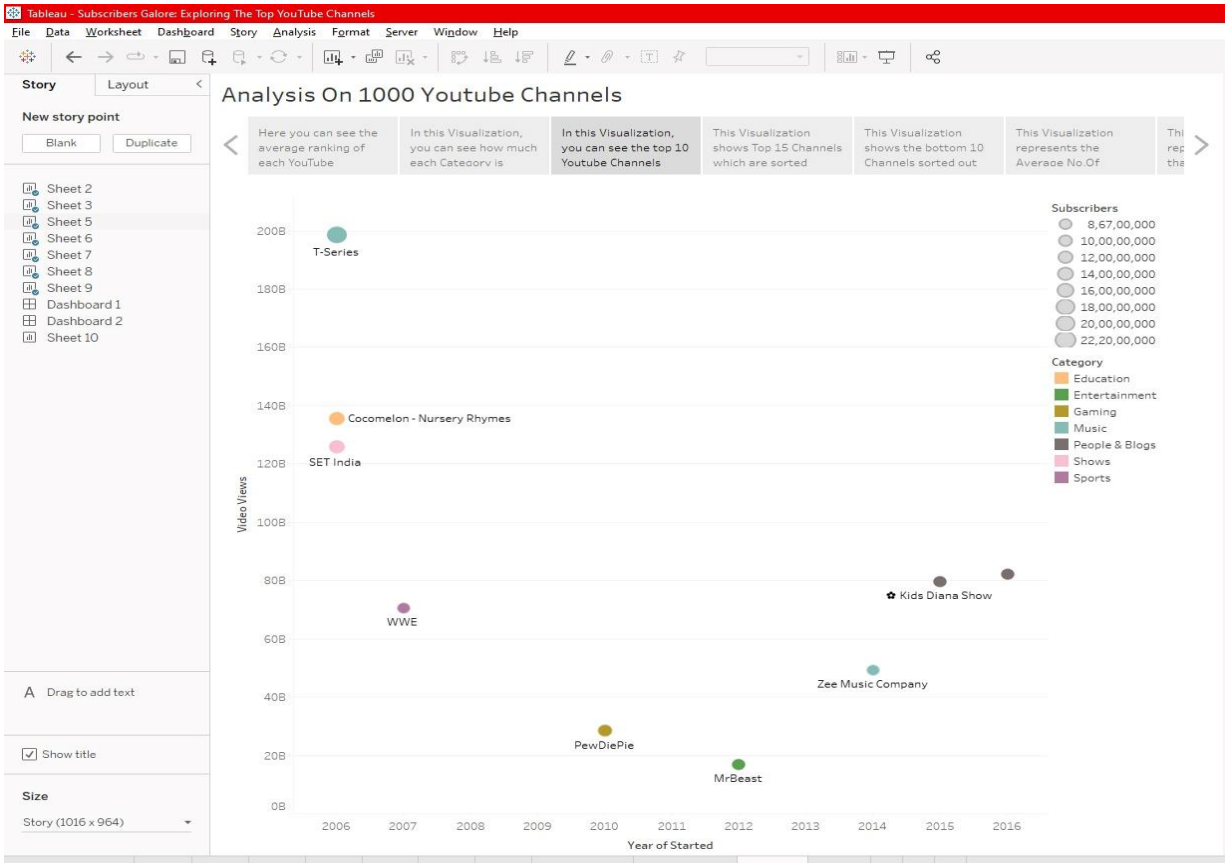


Tableau - Subscribers Galore: Exploring The Top YouTube Channels
File Data Worksheet Dashboard Story Analysis Format Server Window Help
Story Layout
New story point
Blank Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Cateoory is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Subscribers
8,67,00,000
10,00,00,000
12,00,00,000
14,00,00,000
16,00,00,000
18,00,00,000
20,00,00,000
22,20,00,000
Category
Education
Entertainment
Gaming
Music
People & Blogs
Shows
Sports
T-Series
Cocomelon - Nursery Rhymes
SET India
WWE
Kids Diana Show
Zee Music Company
PewDiePie
MrBeast
Video Views
Year of Started

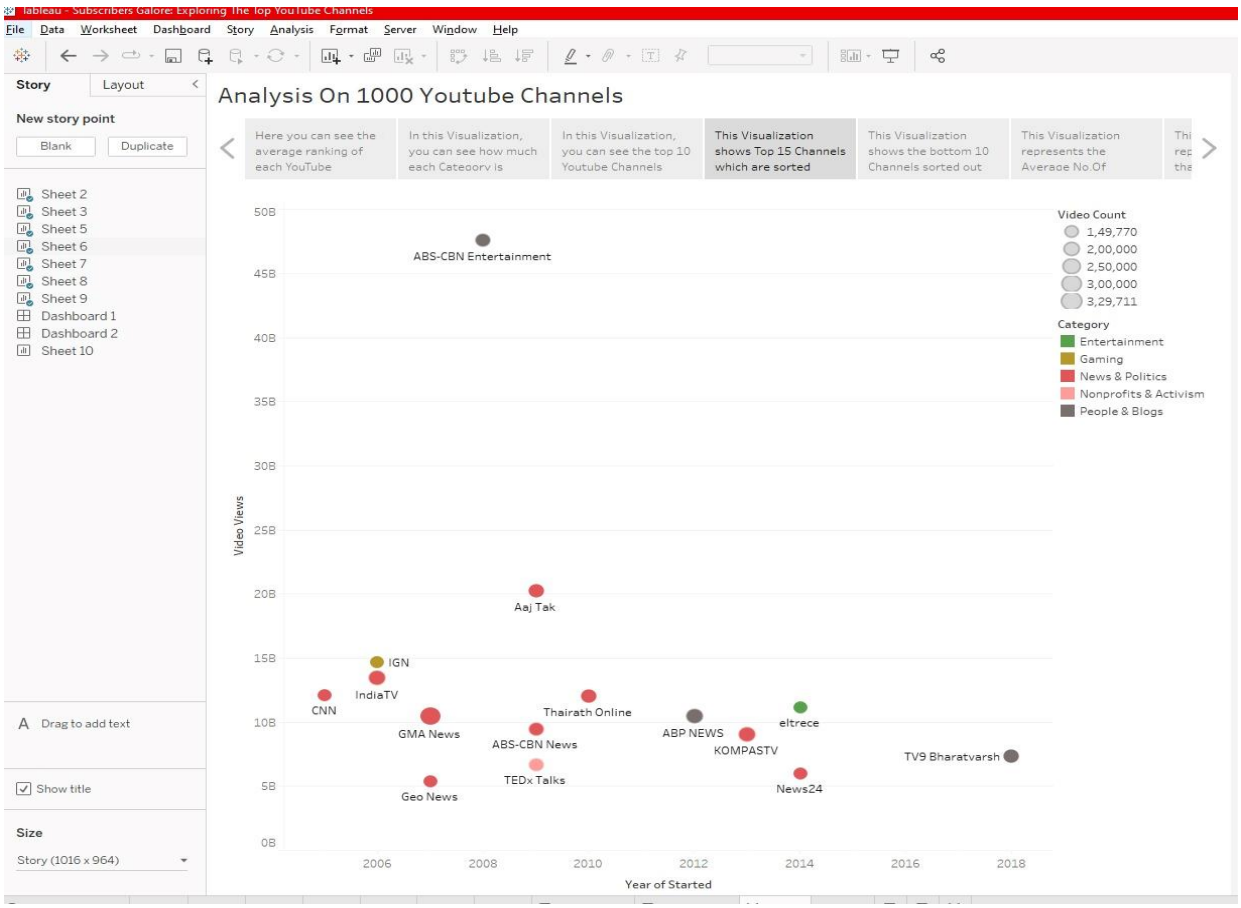
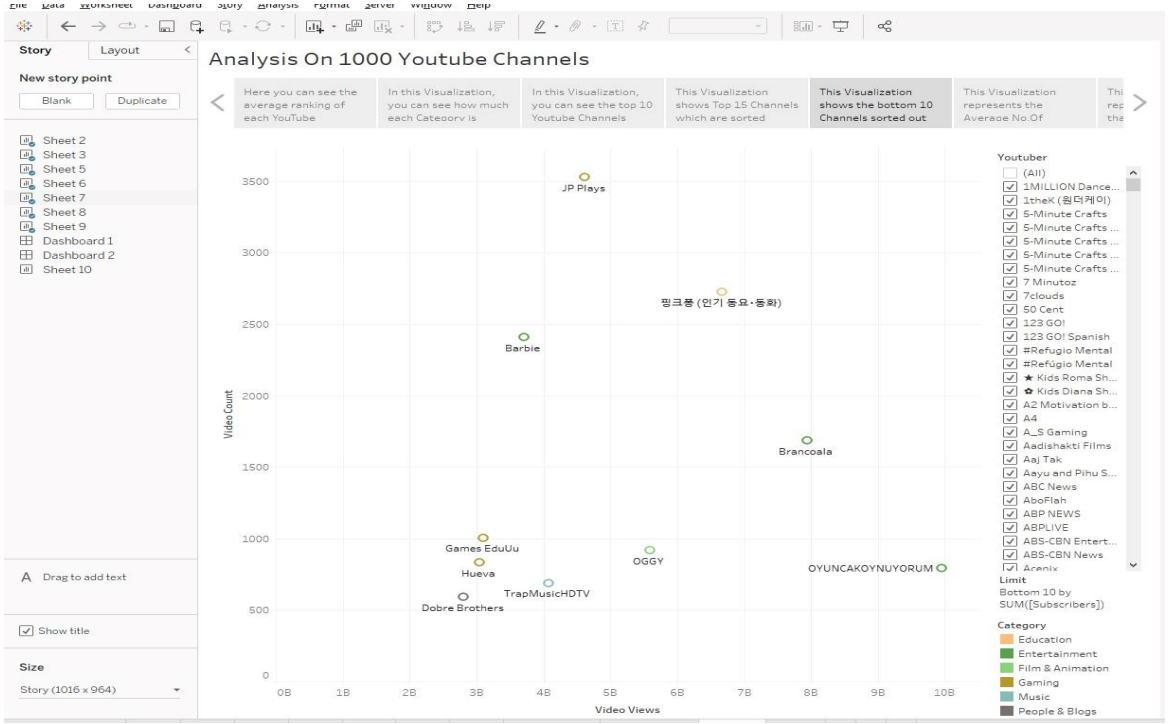
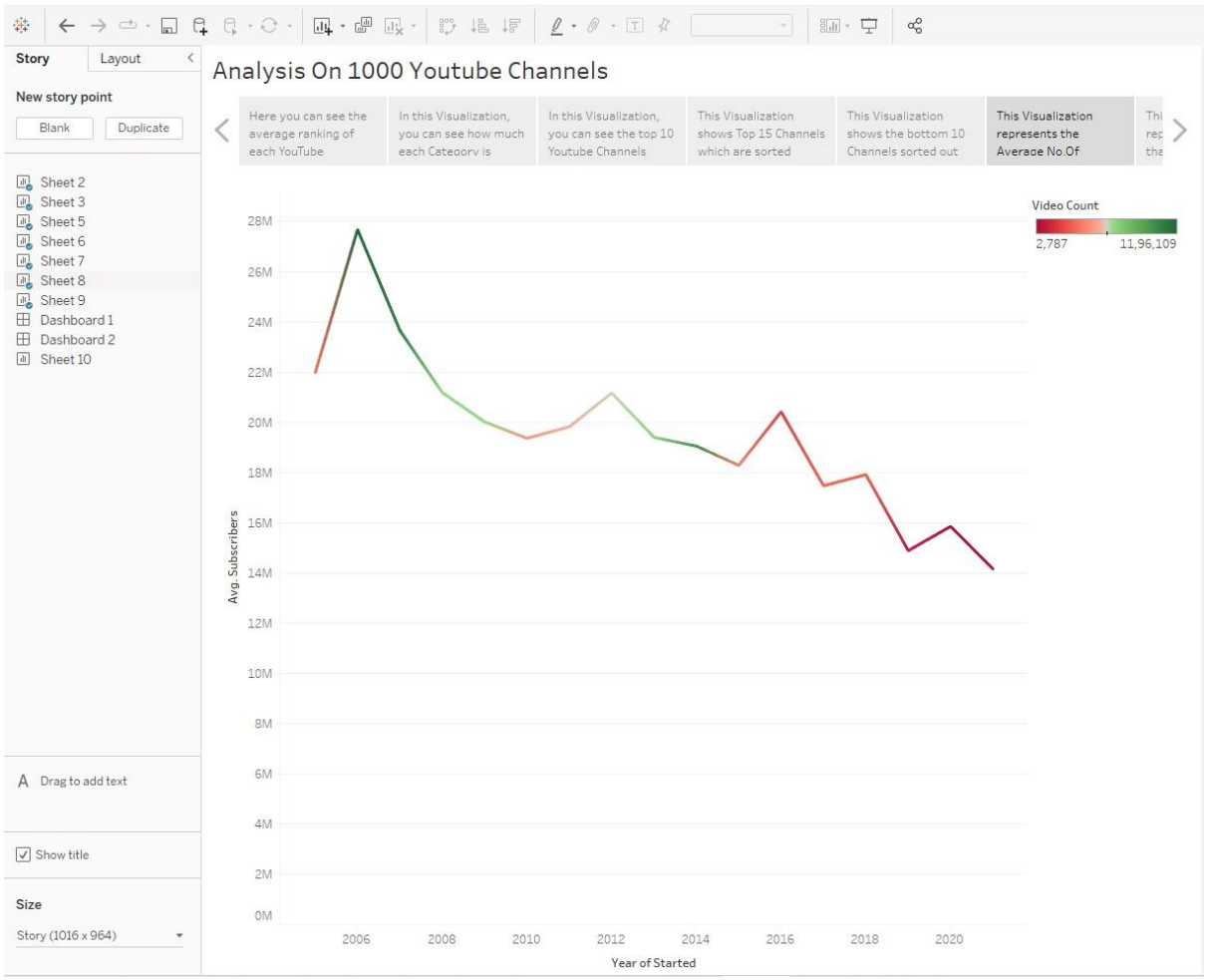
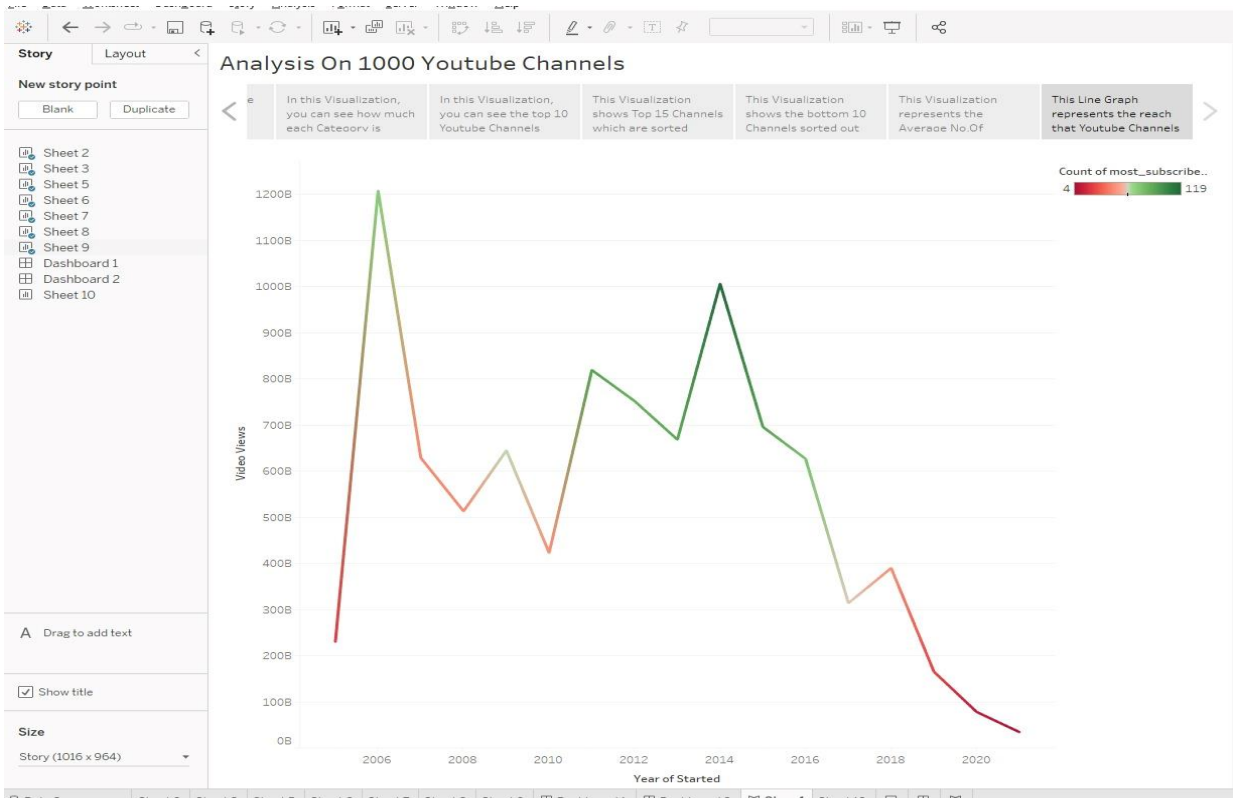


Tableau - Subscribers Galore: Exploring The Top YouTube Channels
File Data Worksheet Dashboard Story Analysis Format Server Window Help
Story Layout
New story point
Blank Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Cateoory is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Video Count
1,49,770
2,00,000
2,50,000
3,00,000
3,29,711
Category
Entertainment
Gaming
News & Politics
Nonprofits & Activism
People & Blogs
ABS-CBN Entertainment
Aaj Tak
IGN
IndiaTV
CNN
GMA News
Thairath Online
ABS-CBN News
TEDx Talks
Geo News
ABP NEWS
KOMPASTV
eltrece
News24
TV9 Bharatvarsh
Video Views
Year of Started


Story
Layout
New story point
Blank
Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Category is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Video Count
Video Views
JP Plays
핑크퐁 (인기 동요·동화)
Barbie
Brancoala
Games EduUu
Hueva
Dobre Brothers
TrapMusicHDTV
OGGY
OYUNCAKOYNUYORUM
Youtuber
Limit
Bottom 10 by SUM([Subscribers])
Category
Education
Entertainment
Film & Animation
Gaming
Music
People & Blogs


Story
Layout
New story point
Blank
Duplicate
Sheet 2
Sheet 3
Sheet 5
Sheet 6
Sheet 7
Sheet 8
Sheet 9
Dashboard 1
Dashboard 2
Sheet 10
Drag to add text
Show title
Size
Story (1016 x 964)
Analysis On 1000 Youtube Channels
Here you can see the average ranking of each YouTube
In this Visualization, you can see how much each Category is
In this Visualization, you can see the top 10 Youtube Channels
This Visualization shows Top 15 Channels which are sorted
This Visualization shows the bottom 10 Channels sorted out
This Visualization represents the Average No.Of
Video Count
2,787
11,96,109
Avg. Subscribers
Year of Started



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

Performance Testing

The goal of performance testing is to identify bottlenecks, optimize performance, and ensure that the system can handle the expected workload effectively.

Amount Of Data Rendered to Database

The amount of data that is rendered to a database depends on the size of the dataset and the capacity of the database to store and retrieve data.

Utilization Of Filters

We used filters in visualizations numbered 3, 4 and 5. In 3 we filtered (limited) the top 10 channels based on highest number of subscribers. In 4 we filtered top 15 channels based on number of videos. In 5 we filtered bottom 10 channels based on least number of subscribers.

No Of Visualizations/ Graphs (7)

- Average Ranking of Each Category
- Each category size in YouTube community
- Top Ten YouTube Channels (Based on Subscribers) With Current No. of Videos and In Which Year They Were Started
- Representing Top 15 Channel's (Based on Number of Videos) Started Year and Current Videos Views They Are Getting
- Bottom 10 Channels (Based on Subscribers), How Many Views They Are Getting with Their Current number of Videos Bar graph shows the Country with its primary language and number of subscribers.
- Average No. Of Subscribers of YouTube Channels Started at Certain Year
- YouTube Channels Started at Certain Year Have No. Of Views

Web Integration

Publishing helps us to track and monitor key performance metrics, to communicate results and progress. help a publisher stay informed, make better decisions, and communicate their performance to others.

Publishing dashboard and reports to tableau public
Step 1: Go to Dashboard/story, click on share button on the top ribbon.

Dashboard, Report and Story Embed With UI With Flask

Activity 1: Integrating with Tableau Public

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

Activity 2: Integrating with bootstrap website.

Explanation Video link:
https://drive.google.com/file/d/1_DJjZIkcs3Zi6WUHxmgypsRmUNIU5pkj/view?usp=sharing

Activity 3: Implementing Flask

Explanation Video link:
https://drive.google.com/file/d/1e6Xdsul2QW0Ep6NfrXHj6vFwU8AzhDKd/view?usp=drive_link

Website Deployment: -