



COURSE: DATA EXPLORATION AND VISUALIZATION-I

**Project Report: Data Scraping and Analysis of the Data by using
Python**

**Submitted By:
Name: Khushi
Roll No: 045027
Batch: BDA-04**

PROJECT OBJECTIVES:

This project involves the collection and analysis of sample data related to **Videos, Subscribers, and Views of a Youtuber**. The purpose of the project was to make us aware of how web scrapping is done, how the data is extracted, and with that how we can clean the data analyze it, and make a managerial report out from the same.

The project objective is to teach practical data skills, including web scraping, data cleaning, in-depth analysis, sorting, and creating a managerial report while maintaining ethical data practices. It focuses on creating clear, visually appealing reports and using predictive models to make informed decisions.

For analysis purposes, I collected data from <https://vidiq.com>, specifically the Top 100 YouTube Channels in India. This data shows the views on their videos, the subscribers that they have, and the number of videos they have posted altogether. My analysis focused on key aspects such as Central Tendency, Variance, Correlation, Time Series Analysis, Statistical Analysis, and Sorting. The goal is to provide a comprehensive report that offers actionable insights while enhancing the understanding of YouTube channel dynamics in India.

The total number of:

VIDEOS- 2876575.0

VIEWS- 1979056720000.0

SUBSCRIBERS- 3649900000.0

CONCLUSION FROM STATISTICAL ANALYSIS OF THE DATA - DESCRIPTIVE ANALYSIS

	Mean	Median	Standard Deviation	Variance	Minimum	Maximum
Subscribers	3.649900e+07	2.840000e+07	3.002954e+07	90177302929292.2	1.800000e+07	2.490000e+08
Views	1.979057e+10	1.178000e+10	2.989978e+10	8.9399670955758e+2	3.517800e+08	2.324100e+11
Videos	28765.75000	3.660000e+03	6.594160e+04	4348294299.987373	67.000000	313720.000000

Analysis of the Table-1

DESCRIPTIVE ANALYSIS:

The descriptive analysis provides insights into the average values and ranges of subscribers, views, and videos on YouTube channels. The descriptive statistics give us a snapshot of the key characteristics of our YouTube channel data:

- **Subscribers:** On average, YouTube channels have approximately 3.649900e+07 subscribers, with a minimum of 1.800000e+07 and a maximum of 2.490000e+08.
- **Views:** The videos on these channels have been viewed 1.979057e+10 times on average, with a range from 3.517800e+08 to 2.324100e+11.
- **Videos:** The channels have uploaded an average of 28765.750000 videos, with the fewest being 67.000000 and the most being 313720.000000.

CONCLUSION FROM STATISTICAL ANALYSIS OF THE DATA–CORRELATION MATRIX

	SUBSCRIBERS	VIEWS	VIDEOS
SUBSCRIBERS	1.000000	0.940893	0.117044
VIEWS	0.940893	1.000000	0.162592
VIDEOS	0.117044	0.162592	1.000000

Analysis of the Table-2

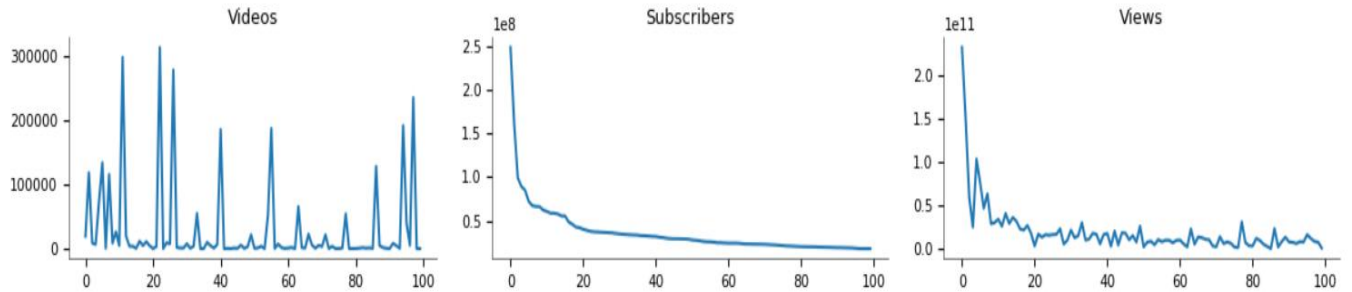
CORRELATION MATRIX:

The correlation matrix shows the degree and direction of relationships between these variables, helping us understand how they influence each other. The correlation matrix reveals how these variables are related to each other:

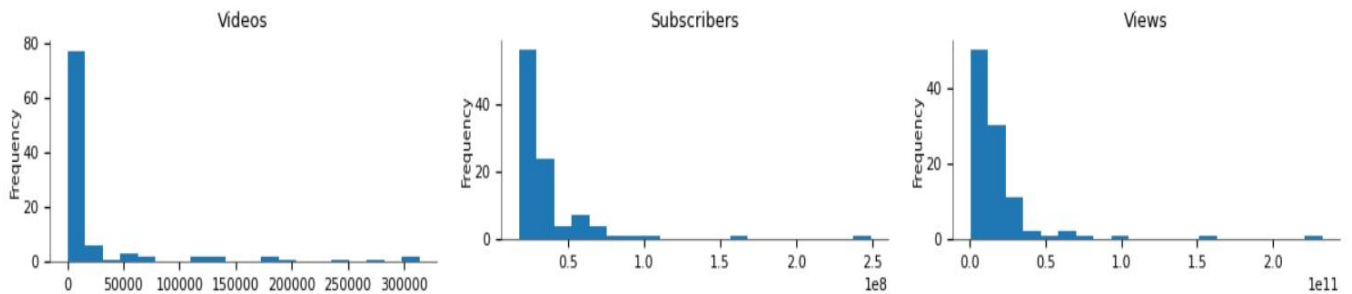
- Subscribers vs. Views: There is a 0.940893 correlation between the number of subscribers and the total views. This suggests that as subscribers increase, views tend to increase as well.
- Subscribers vs. Videos: The correlation between subscribers and the number of videos is 0.117044. This indicates a lack of strong correlation between these two factors.
- Views vs. Videos: Views and the number of videos show a [0.162592] correlation. This means that as the number of videos increases, the total views tend to increase.

GRAPHICAL REPRESENTATION OF THE DATA

Values



Distributions

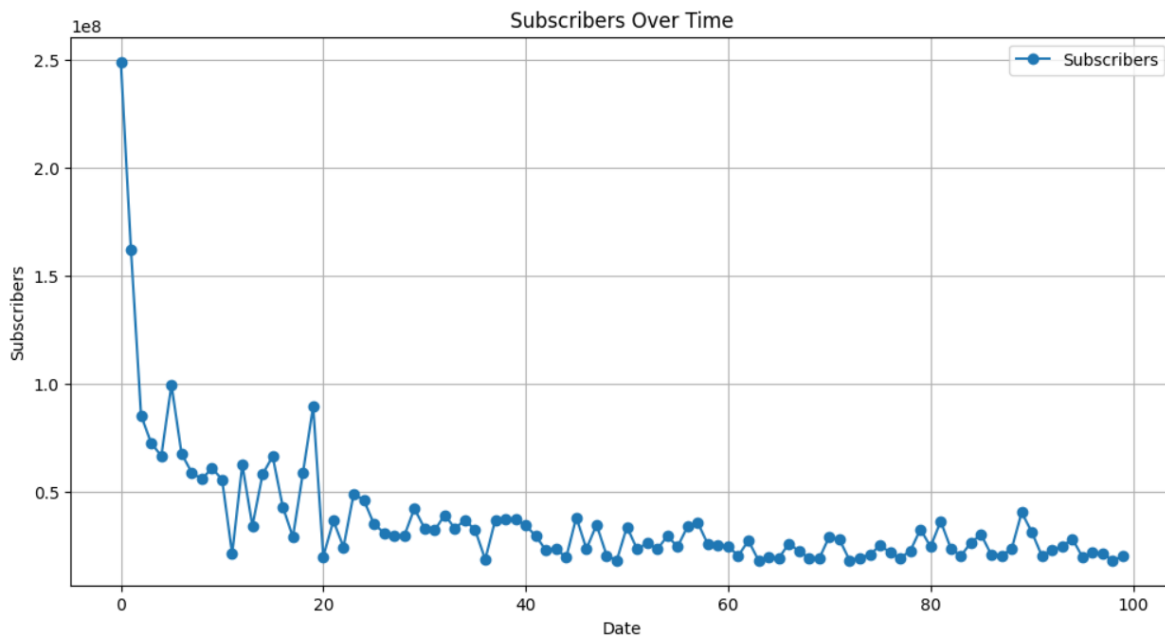


Analysis of the Graph- 1

The graphs show the following:

- The number of subscribers is increasing over time, but the number of views is declining. The increasing number of subscribers suggests that the channel is gaining popularity. However, the declining number of views suggests that the channel's content is not resonating with viewers as much as it used to.
- The distribution of views is also increasing, meaning that there are more videos with a small number of views and fewer videos with a large number of views. The increasing distribution of views suggests that the channel is trying to reach a wider audience by creating a variety of content. However, the declining distribution of subscribers suggests that this strategy is not working as well as it could.
- The distribution of subscribers is declining, meaning that there are fewer channels with a large number of subscribers and more channels with a small number of subscribers. The declining distribution of subscribers suggests that the channel is losing some of its most engaged viewers. This could be due to a number of factors, such as the quality of the content, the frequency of uploads, or the channel's overall direction.

TIME SERIES ANALYSIS OF SUBSCRIBERS OVER TIME

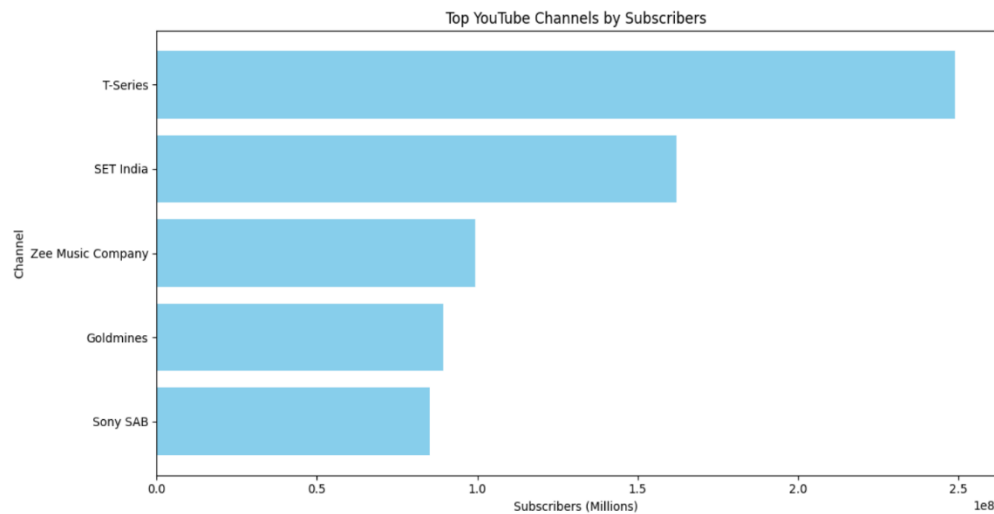


Analysis of the Graph-2

The graph shows the following:

- The overall trend is that the number of subscribers is increasing over time. This is to be expected, as YouTube is a growing platform and new channels are constantly being created.
- However, there are some periods where the number of subscribers decreases. This could be due to a number of factors, such as a decline in the quality of content, changes in the algorithm, or competition from other channels.
- The channels with the most subscribers are typically those that create high-quality, engaging content that is relevant to their audience. They also tend to promote their content effectively through social media and other channels.
- The channels with the fewest subscribers are typically those that create low-quality content, or content that is not relevant to their audience. They may also not promote their content effectively.

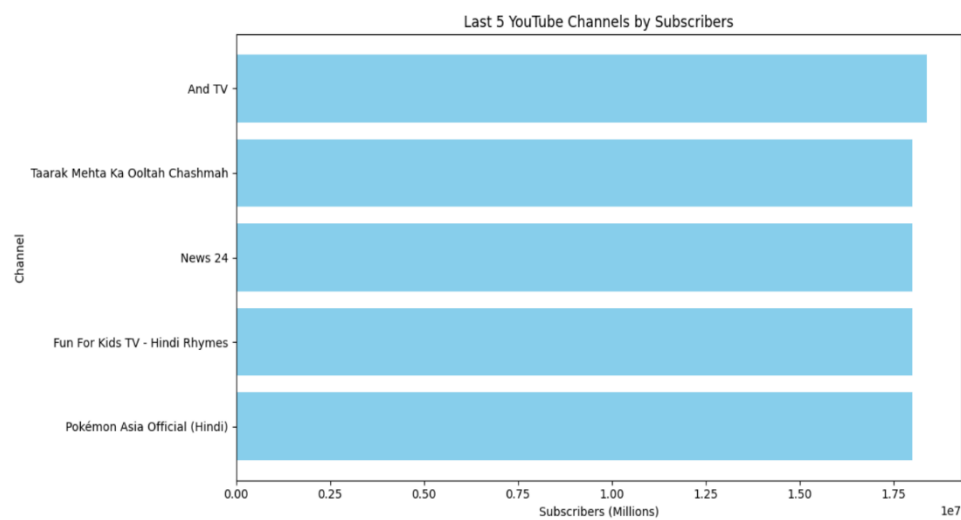
5 TOP SUBSCRIBED AND 5 LEAST SUBSCRIBED YOUTUBE CHANNELS



Analysis of the Graph 3

The graph shows that the Top 5 YouTube Channels by subscribers are:

1. T-SERIES
2. SET INDIA
3. ZEE MUSIC COMPANY
4. GOLDMINES
5. SONY SAB



Analysis of the Graph 4

The graph shows that the Last 5 YouTube Channels by subscribers are:

100. TAARAK MEHTA KA OOLTAH CHASHMAH

99. NEWS 24

98. FUN FOR KIDS TV- HINDI RHYMES

97. POKEMON ASIA OFFICIAL

96. AND TV

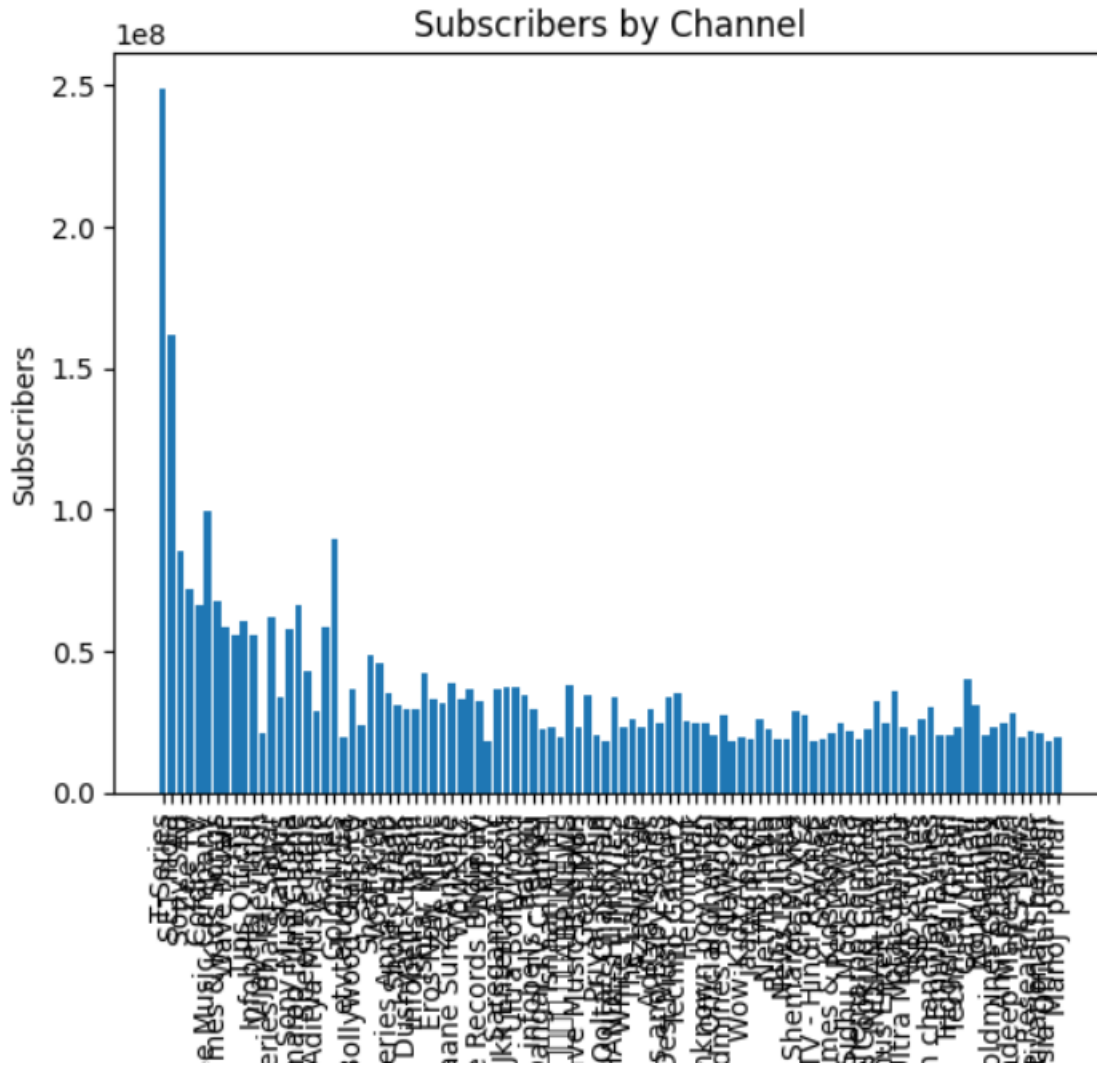
SORTING ON THE BASIS OF VIDEOS, SUBSCRIBERS, AND VIEWS SEPERATELY

On sorting by '*Videos*' I found that ABP NEWS has the highest number of videos (313720.0), whereas ROUND2HELL has the lowest number of videos (67.0) posted on their respective channels.

On sorting by '*Subscribers*' I found that T – SERIES has the highest number of subscribers (313720.0), whereas TAARAK MEHTA KA OOLTAH CHASHMAH has the lowest number of subscribers (18000000.0) subscribed on their respective channels.

On sorting by '*Views*' I found that T- SERIES has the highest number of views (2.324100e+11), whereas MANOJ PARIHAR has the lowest number of views (3.517800e+08) on the videos of their respective channels.

DATA VISUALISATION OF SUBSCRIBERS BY CHANNEL



Analysis of the Graph 5

The graph of data visualization on subscribers by channel shows that:

As the sample size was large, the X-axis of the graph containing the names of the channels got cluttered, but the graph depicts:

T-SERIES has the highest SUBSCRIBERS to VIEWS ratio whereas, MANOJ PARIHAR has the lowest SUBSCRIBERS to VIEWS ratio.

FORECASTING FOR YOUTUBERS

These YouTubers can consider forecasts and general practices for their YouTube channels based on their subscribers, views, and video data.

Subscriber Forecast

1. Interact with your subscribers through comments, live streams, and community posts. Build a strong, loyal audience by responding to feedback and fostering a sense of community.
2. Continue to create high-quality, engaging content that resonates with your target audience. Consistency is key.
3. Collaborate with other YouTubers or influencers in your niche to cross-promote and potentially gain new subscribers.
4. Use time series forecasting models (e.g., ARIMA) to predict future subscriber growth. This can help you set realistic growth targets.

Views Forecast

1. Invest time in creating compelling video titles and eye-catching thumbnails. These elements can significantly impact click-through rates and views.
2. Use relevant keywords and tags in your video descriptions to improve discoverability. Conduct keyword research to identify trending topics in your niche.
3. Diversify your content to appeal to a broader audience. Explore trending topics and formats within your niche.
4. Share your videos on social media platforms and within online communities where your target audience hangs out.

Videos Forecast

1. Develop a content calendar to plan your video releases. Consistency in uploading can keep your audience engaged.
2. Analyze which video lengths perform best for your audience. Tailor your content to meet viewer preferences.
3. Focus on producing high-quality videos rather than increasing the quantity. Quality content is more likely to attract and retain viewers.

4. Forecast future video production based on historical trends and audience preferences.

Some other general recommendations are:

- Regularly review YouTube Analytics to understand which videos perform well and identify areas for improvement.
- Keep up with trends and changes in the YouTube platform and the content industry. Adapt your strategies accordingly, and stay informed.
- Pay attention to comments and feedback. Use constructive criticism to refine your content.
- Network with other content creators and consider collaborations to expand your reach.
- Building a successful YouTube channel takes time. Be persistent, patient, and adaptable in your approach.