# YouTube Channel Analysis Report

## 1. Introduction

This report presents an analysis of YouTube channels in Kenya, using data extracted from YouTube API. The goal is to understand the trends in channel metrics over time and explore relationships between various metrics.

## 2. Data Preparation

Data Columns:

* Channel ID: Unique identifier for each channel.
* Channel Title: Name of the YouTube channel.
* Description: Description of the channel.
* Subscribers: Number of subscribers.
* Views: Total number of views.
* Videos: Total number of videos.

Date Column:

* The 'publishedAt' column was initially set as an index but was reset to a regular column for easier processing. It represents the date when the data was collected.

## 3. Data Cleaning

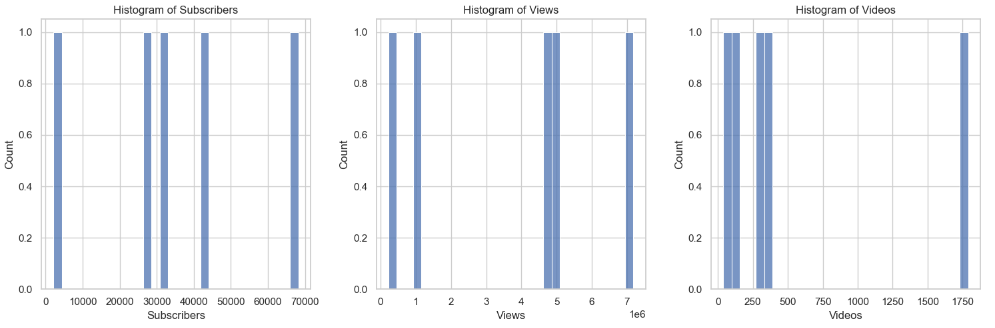
* Missing Values: Rows with missing 'publishedAt' dates were identified and handled.
* Data Conversion: The 'publishedAt' column was converted to datetime format.
* Handling Missing Dates: Missing dates were filled with the mean date of available data.

## 4. Exploratory Data Analysis (EDA)

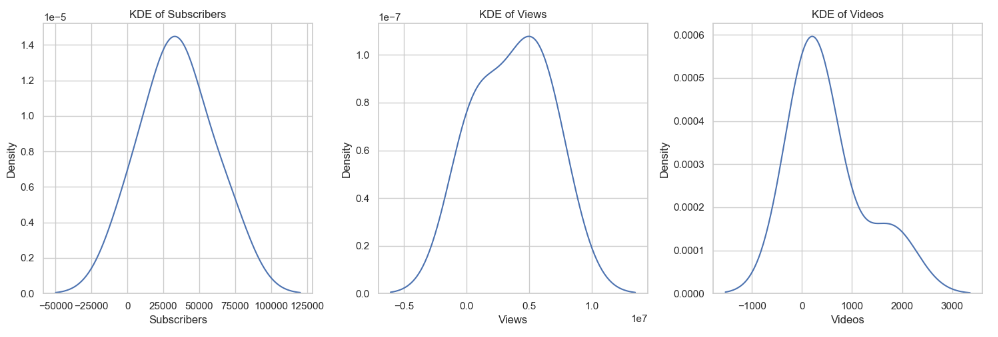
### 4.1 Visualization of Distributions

Histograms, boxplots and KDE plots were used to visualize the distributions of key metrics:

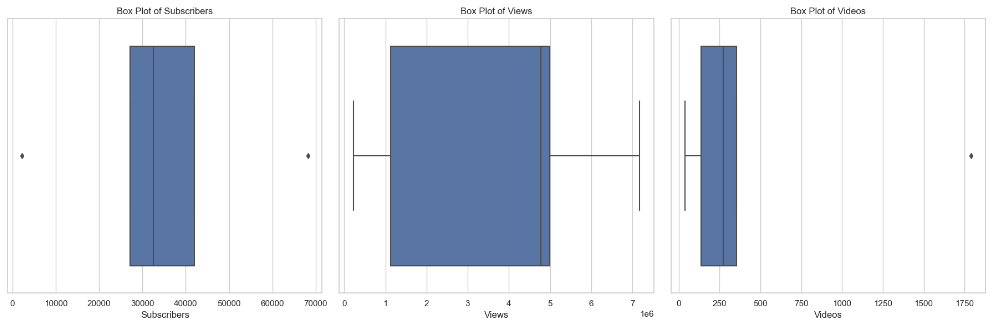
* Subscribers: Distribution of subscriber counts across channels.
* Views: Distribution of view counts.
* Videos: Distribution of the number of videos.



**HISTOGRAMS**



**KDE PLOTS**

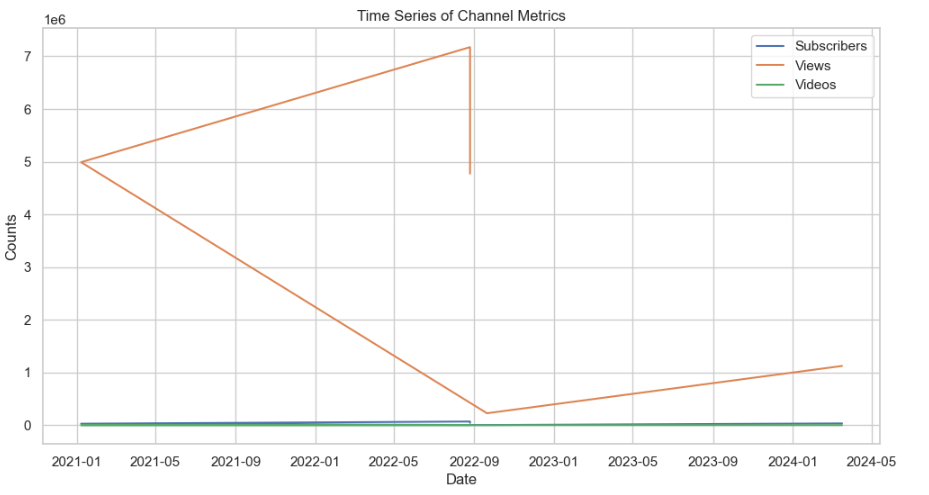


**BOX PLOTS**

### 4.2 Time Series Analysis

A time series plot was created to visualize trends over time:

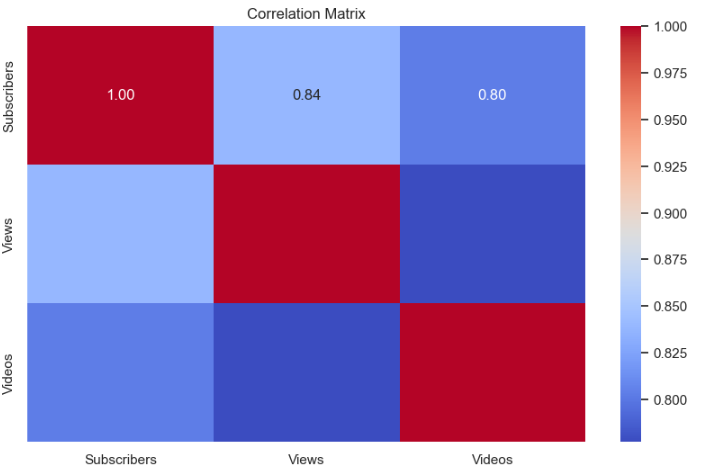
* Subscribers: Trends in subscriber counts.
* Views: Trends in view counts.
* Videos: Trends in the number of videos.



### 4.3 Correlation Analysis

Correlation analysis was conducted to explore relationships between:

* Subscribers and Views.
* Subscribers and Videos.
* Views and Videos.



## 5. Insights and Conclusions

* Distribution Insights:
  + The distribution of subscribers and views shows skewness, indicating that a small number of channels have very high metrics compared to the majority.
  + The distribution of the number of videos shows a more uniform spread.
* Time Series Trends:
  + Trends in subscriber counts, views, and number of videos over time indicate growth or decline patterns which can be useful for understanding channel performance and planning.
* Correlation Analysis:
  + Positive correlations between subscribers and views suggest that channels with more subscribers generally have more views.
  + The relationship between videos and other metrics indicates that more videos can contribute to higher views and subscribers.

6. Recommendations for Further Analysis

* Detailed Seasonal Analysis: Examine seasonal trends and their effects on channel metrics.
* Predictive Modeling: Develop models to forecast future metrics based on historical data.
* Comparative Analysis: Compare metrics across different categories or types of channels.