Software Requirements Specification

for

Youtube Channel Analysis

Prepared by Arpitya Kumar Singh

15-April-2023

Contents

Table of Contents…….....Page No.

1. Introduction

This Software Requirements Specification (SRS) document describes the requirements for the development of a software system that analyzes the top subscribed Youtube channels dataset. The system aims to provide insights into the characteristics of the channels and their categories.

this code is all about analyzing the top subscribed Youtube channels dataset, which has information about the most subscribed Youtube channels in the world. This dataset is used to gain insights into the characteristics of the channels and their categories.

To do this, the code first imports some necessary libraries such as pandas, numpy, and visualization libraries like seaborn, plotly, and matplotlib. Then it reads the dataset from a CSV file and shows the first five rows of the dataset.

After that, the code removes unnecessary columns and checks for missing values. It also checks the unique values of the Category column and removes rows with an unknown category. The final dataset is then described, showing various statistical information.

Next, the code creates two plots to visualize the data. The first plot shows the percentage of Youtube channels by category using a pie chart, where Entertainment, Music, and People & Blogs are the top three categories. The second plot displays a scatter plot of the video views and subscribers by categories using Plotly Express, where Musics, Education, and Shows are the top three categories with the most subscribers and video views.

The code also analyzes the top 10 music channels with the most video count and the top 10 education channels with the most video views. The results are displayed using a bar chart using the seaborn library.

Overall, the code provides insights into the characteristics of the top subscribed Youtube channels, their categories, and their video count and views. It also shows how to perform data processing and visualization using pandas, numpy, seaborn, and plotly libraries.

2. Functional Requirements

2.1 Data Processing

- The system shall be able to remove unnecessary columns from the dataset.
- The system shall be able to check for missing values in the dataset.
- The system shall be able to remove rows with unknown categories in the dataset.

2.2 Data Visualization

- The system shall be able to display a pie chart showing the percentage of Youtube channels by category.
- The system shall be able to display a scatter plot of the video views and subscribers by categories.
- The system shall be able to display a bar chart of the top 10 music channels with the most video count and the top 10 education channels with the most video views.

2.3 Data Analysis

- The system shall be able to provide insights into the characteristics of the top subscribed Youtube channels, their categories, and their video count and views.

3. Non-Functional Requirements

3.1 Performance

- The system shall be able to process the dataset efficiently and in a reasonable time.
- The system shall be able to display the visualizations without significant delay.

3.2 User Interface

- The system shall have an intuitive and user-friendly interface.
- The system shall be able to display the visualizations aesthetically pleasingly.

3.3 Compatibility

- The system shall be compatible with different operating systems and devices.
- The system shall be compatible with different web browsers.

4. Constraints

4.1 Dataset

- The system shall be limited to the analysis of the top subscribed Youtube channels dataset.
- The system shall only process CSV files.

5. References

The development of this system will be based on the following references:

- Pandas library for data processing
- Seaborn, Matplotlib, and Plotly libraries for data visualization
- Python programming language

6. Glossary

- CSV - Comma Separated Values
- SRS - Software Requirements Specification