**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date |  |
| Team ID | 739826 |
| Project Title | View count visionary:a data driven approach to forecasting youtube videos views |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

This template provides a structured approach to planning your data collection process for forecasting YouTube video views. Adjust and expand each section based on your specific project requirements and available resources.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | Project Title: View count visionary: A Data-Driven Approach to forecasting youtube videos views  Objective: To develop a predictive model for estimating future views of YouTube videos based on historical data and video metadata. |
| Data Collection Plan | 1.Data Sources Identification  2.Data types  3.Data collection frequency  4.Data storage and management  5.Data prepfrocessing  6.Tools and technologies  7.Ethical Considerations  8.Timeline |
| Raw Data Sources Identified | YouTube Analytics API:  Access to video-specific metrics such as views, likes, comments, and engagement.  Provides real-time and historical data necessary for forecasting.  Social Media Analytics Platforms:  Social Blade: Offers historical data on YouTube channels and videos, including subscriber growth, view trends, and estimated earnings.  SimilarWeb: Provides insights into traffic sources, referral sites, and audience demographics.  Web Scraping (if applicable):  Extract data from public platforms and forums discussing video trends.  Gather metadata and comments to understand user sentiment and engagement. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Youtube analytics API | Video-specific metrics (views, likes, comments, etc.) provided by YouTube's API. | YouTube Analytics API | JSON, CSV | Varies | API key required |
| social blade | Historical data on YouTube channels and videos including view counts and engagement metrics. | Social Blade | CSV, Excel | Varies | Publicly accessible |
| Web Scraping Tools | Public data from forums, social media, and other platforms discussing video trends and engagement. | Custom scripts or tools like BeautifulSoup for Python | Text, HTML | Varies | Ensure compliance with platform terms |