**CAPSTONE \_PROJECT:YOUTUBE-DATA-HARVESTING-AND-WAREHOUSING USING SQL AND STREAMLIT:**

* Youtube Channel : <https://www.youtube.com/channel/UCZikuVCya6icZj5mWnVwEA>

(Watch full details about this project)

* DemoVideo :

<https://www.loom.com/share/a34ae29b1ab84ae388b2fde582d90dc2>

**INTRODUCTION:**

The YouTube Data Harvesting and Warehousing Project is designed to fetch data from YouTube using the Google API Client, store it in a MySQL database, and provide an interactive user interface for querying and visualizing the data using Streamlit.

**TABLE OF CONTENTS:**

1. Introduction
2. Problem statement
3. Features
4. Technologies Used
5. Installation Version
6. Running of Installation
7. Tools and Libraries Used
8. Workflow

**PROBLEM STATEMENT** :

**>>**Create a Streamlit application that allows users to access and analyze data from multiple YouTube channels.

**>>** Extracting data using Youtube API and then Transforming it to a relational database like MySQL. For getting various info about youtube channels.

**FEATURES:**

**Data Collection**:

* Harvest data from YouTube channels and videos using the Google API Client.

**Data Storage:**

* Store the collected data in a MySQL database.

**Data Analysis:**

* Perform various analytical queries on the data.

**Visualization:**

* Visualize the data using Streamlit.

**User Interface:**

* Interactive Streamlit interface for querying and viewing data.

**TECHNOLOGIES USED:**

* Python(<https://www.python.org/>)
* MySQL(<https://www.mysql.com/>)
* YouTube Data API(<https://developers.google.com/youtube/v3>)
* Streamlit(<https://docs.streamlit.io/library/api-reference>)
* Pandas(<https://pandas.pydata.org/>)

**INSTALLATION VERSION:**

\* Python version 3.12.2

\* Streamlit version 1.35.0

\* Mysql version8.0.36

**RUNNING OF INSTALLATION:**

To run this project, you need to install the following packages:

# pip install google-api-python-client

# pip install mysql.connector python

# pip install pandas

# pip install streamlit

**TOOLS AND LIBRARIES USED**:

**STREAMLIT:**

-Streamlit library was used to create a user-friendly UI.

-It enables users to interact with the programme and carry out data retrieval and analysis operations.

**PYTHON:**

-Python is a high level programming language.

-Python is the primary language employed in this project for the development of the complete application, including data retrieval, processing, analysis, and visualisation.

**GOOGLE API CLIENT:**

-The googleapiclient library communicat with different Google APIs.

- Google API is used to retrieve channel data,video data and comment data from Youtube.

**SQL (MySQL)**:

-A relational database used as a data warehouse for storing migrated YouTube data.

**WORKFLOW:**

**COLLECTION OF DATA:**

Retreiving data from youtube channels using Youtube API key and gaining the information which we required from the channels.

**STORAGE OF DATA:**

Once the information is gained it all the store in MySql database.Migration of data to a SQL database for efficient querying and analysis.

Search and retrieval of data from the SQL database.Support for handling multiple YouTube channels and managing their data.

**STREAMLIT DASHBOARD:**

The above mentioned all informations are to be described in the Streamlit Dashboard and its functionalities to provide a user friendly interactions to visulaize the data.