



# YOUTUBE DATA HARVESTING AND WAREHOUSING

CAPSTONE PROJECT:  
SUJITHA S



# AIM:

- This project aims to develop a user-friendly Streamlit application that utilizes the Google API to extract information on a YouTube channel, stores it in a MongoDB database, migrates it to a SQL data warehouse, and enables users to search for channel details and join tables to view data in the Streamlit app.



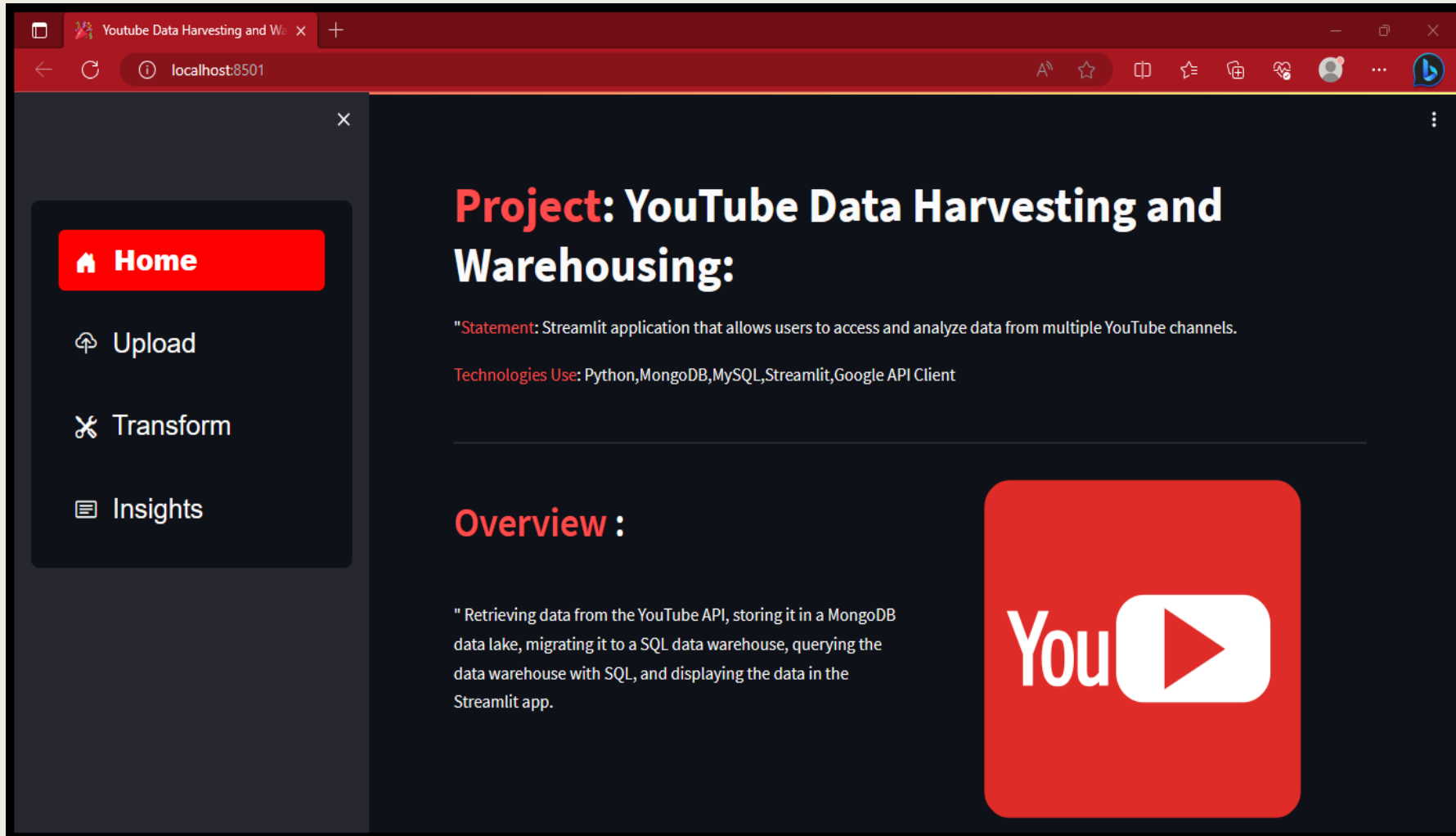
# STEPS TO BE FOLLOWED:

- Set up a Streamlit app
- Connect to the YouTube API
- Store data in a MongoDB data lake
- Migrate data to a SQL data warehouse
- Query the SQL data warehouse
- Display data in the Streamlit app



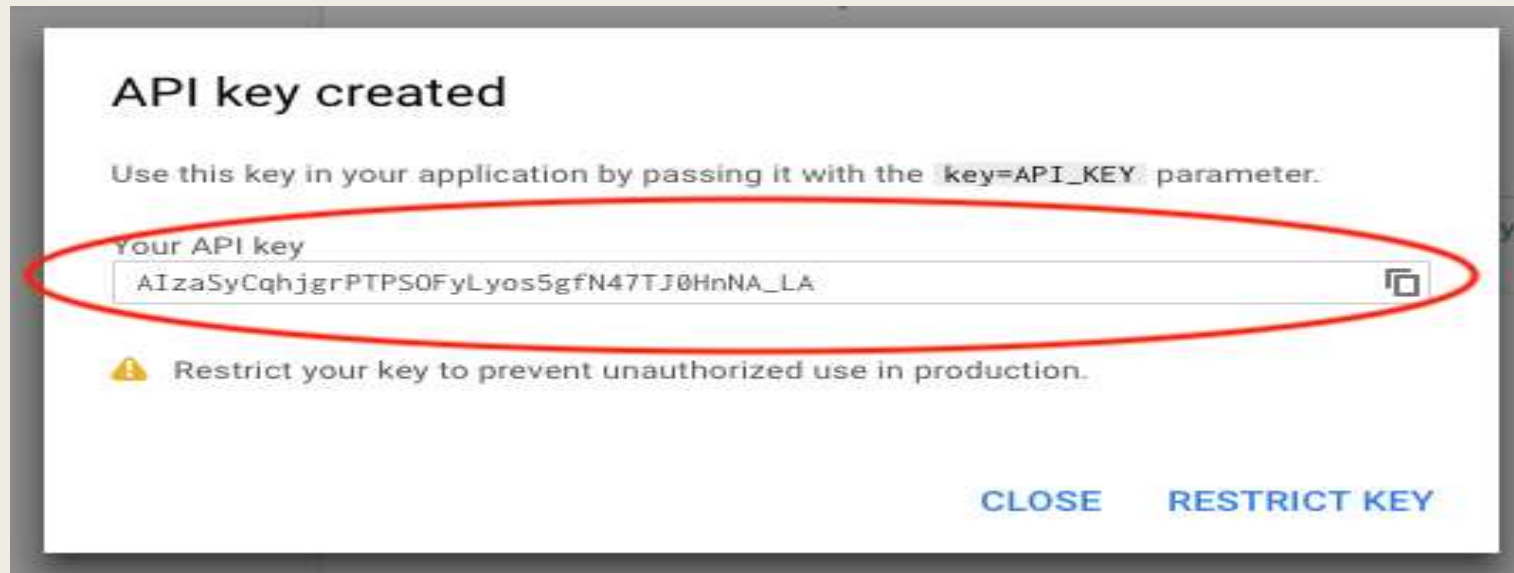
# STREAMLIT PAGE

- This is my **YOUTUBE DATA HARVESTING AND WAREHOUSING** Streamlit app.



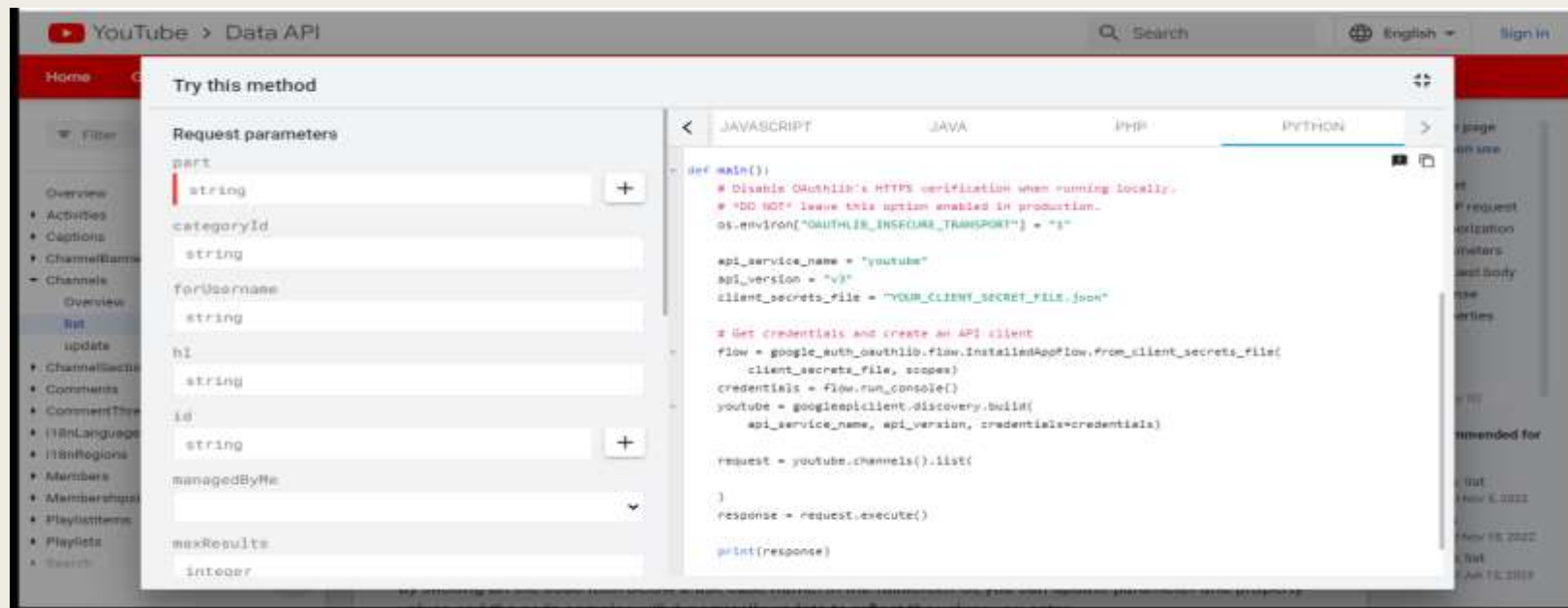
# YOUTUBE API KEY:

- The API provides the ability to retrieve feeds related to videos, users, and playlists.
- We can create our own YOUTUBE API KEY from Google Cloud Platform.
- Image given below is the API KEY for **YOUTUBE Data API V3**.



# RETRIEVE DATA FROM YOUTUBE API

- Select any one YouTube channel and get the channel id
- Refer YouTube API documentation , from the document we can get the codes for data we need.
- In this project , I collected Channel details, Playlist details , Videos details, Comment details.
- Using Functional Blocks I created functions for the details I need , So that I can call it whenever I needed and it can be used for any YouTube channel.



# STORING IN MONGODB

- Store all the data in MongoDB that we collected from the previous step.
- Create a collection , insert them into the database .

The screenshot shows a web application running on localhost:8501. The interface has a dark theme. On the left is a sidebar with a close button (X) and four menu items: 'Home' with a house icon, 'Upload' with a red button and an upload icon, 'Transform' with a scissors icon, and 'Insights' with a list icon. The main content area is titled 'UPLOADING TO MONGODB'. Below the title is a hint: 'Hint : Goto channel's home page > Right click > View page source > Find channel\_id'. There is a button labeled 'migrate to Mongodb'. Below this are two tables. The first table has columns: channel\_name, channel\_id, subscription\_count, channel\_views, channel\_description, and playlist\_id. It contains one row for 'Viu Singapore'. The second table has columns: playlist\_id, channel\_id, and playlist\_name. It contains three rows of playlist data.

UPLOADING TO MONGODB

Hint : Goto channel's home page > Right click > View page source > Find channel\_id

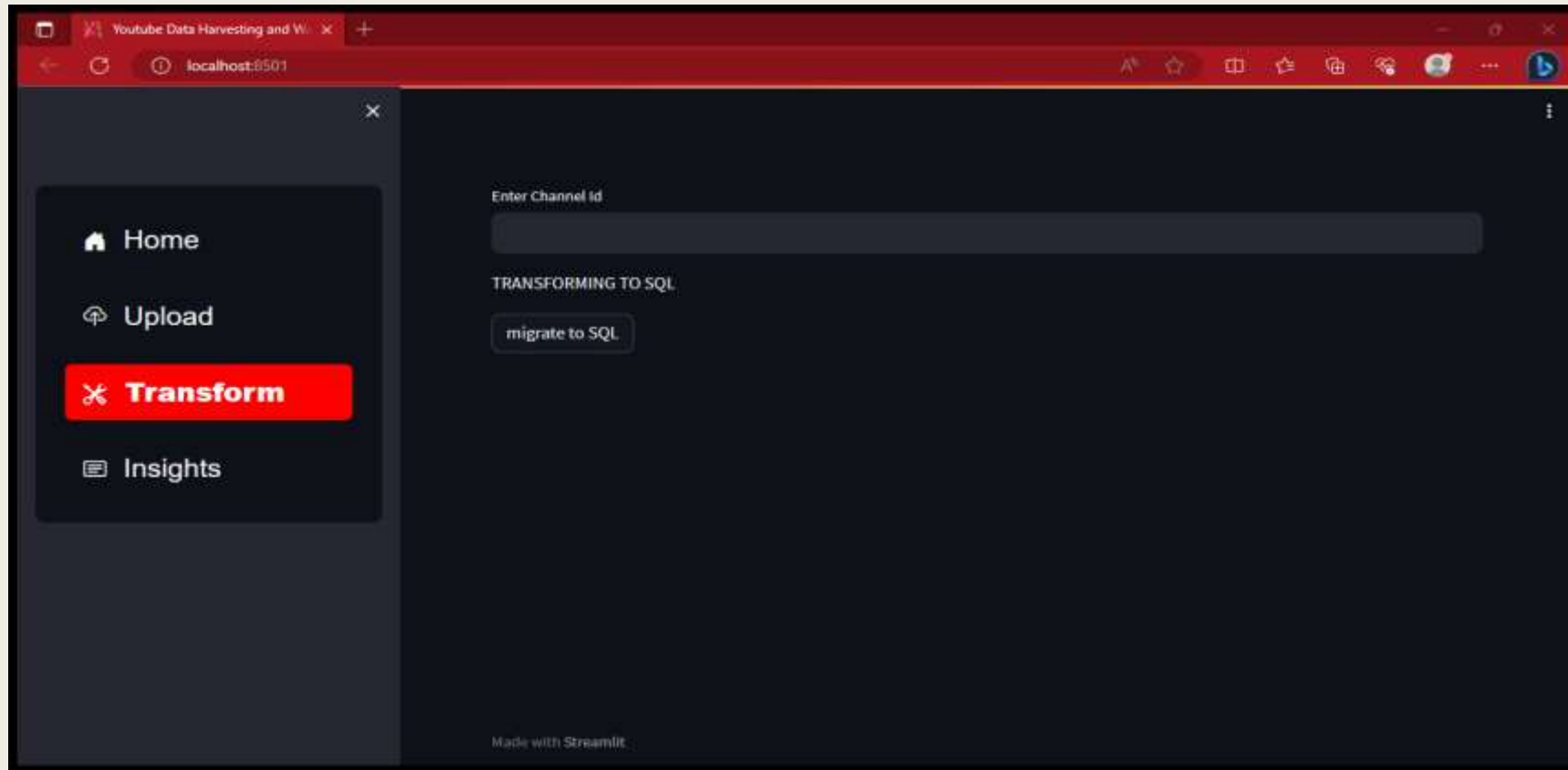
migrate to Mongodb

	channel_name	channel_id	subscription_count	channel_views	channel_description	playlist_id
0	Viu Singapore	UC0RFMDHmTVtsquuswRCU_GA	377000	299111004	The Best of Online Asian Entertainment! Enjoy a wide variety of the latest dramas & variety shows, free downloads with subtitles!	UU0RFMDHr

	playlist_id	channel_id	playlist_name
0	PLg-HQlvgwjvO3sJ3zthPKs1jkZnA87laD	UC0RFMDHmTVtsquuswRCU_GA	What's Hot on Viu - Aug 2023
1	PLg-HQlvgwjvNcUrdOtRJzh5X7AGszj59-	UC0RFMDHmTVtsquuswRCU_GA	Viu Original, My Lovely Liar
2	PLg-HQlvgwjvM8NTcMkwP7yhNHLgJY1Yo2	UC0RFMDHmTVtsquuswRCU_GA	What's Hot on Viu - Jul 2023

# TRANSFER TO SQL

- After you've collected data for multiple channels, you can migrate it to a SQL data warehouse. Here I used MySQL .
- Create Tables for the required details and use the tables for the queries that have to be answered for the questions in INSIGHT PAGE.





# SQL QUERY

The screenshot shows a web browser window with a dark theme. The address bar shows 'localhost:8502'. The application has a sidebar on the left with navigation options: Home, Upload, Transform, and Insights (highlighted in red). The main content area is titled 'Enter Channel Id' and contains a text input field with the value 'UC0RFMDHmTVtsquuswRCU\_GA'. Below this is a dropdown menu labeled 'please select query' with the selected option 'Which channels have the most number of videos, and how many videos do they have?'. A 'Get Answer' button is located below the dropdown. The results are displayed in a table with two columns: 'channel\_names' and 'video-count'. The table contains six rows, with the last row highlighted in yellow.

	channel_names	video-count
0	Simply Coding	255
1	Simply Coding	255
2	SEVENTEEN	1606
3	Coding Is Fun	128
4	Coding Is Fun	128
5	Viu Singapore	2203

THANK YOU