YOUTUBE DATA HARVESTING AND WAREHOUSING

**Introduction:**

YouTube, the online video-sharing platform, has revolutionized the way we consume and interact with media. Launched in 2005, it has grown into a global phenomenon, serving as a hub for entertainment, education, and community engagement. With its vast user base and diverse content library, YouTube has become a powerful tool for individuals, creators, and businesses to share their stories, express themselves, and connect with audiences worldwide.

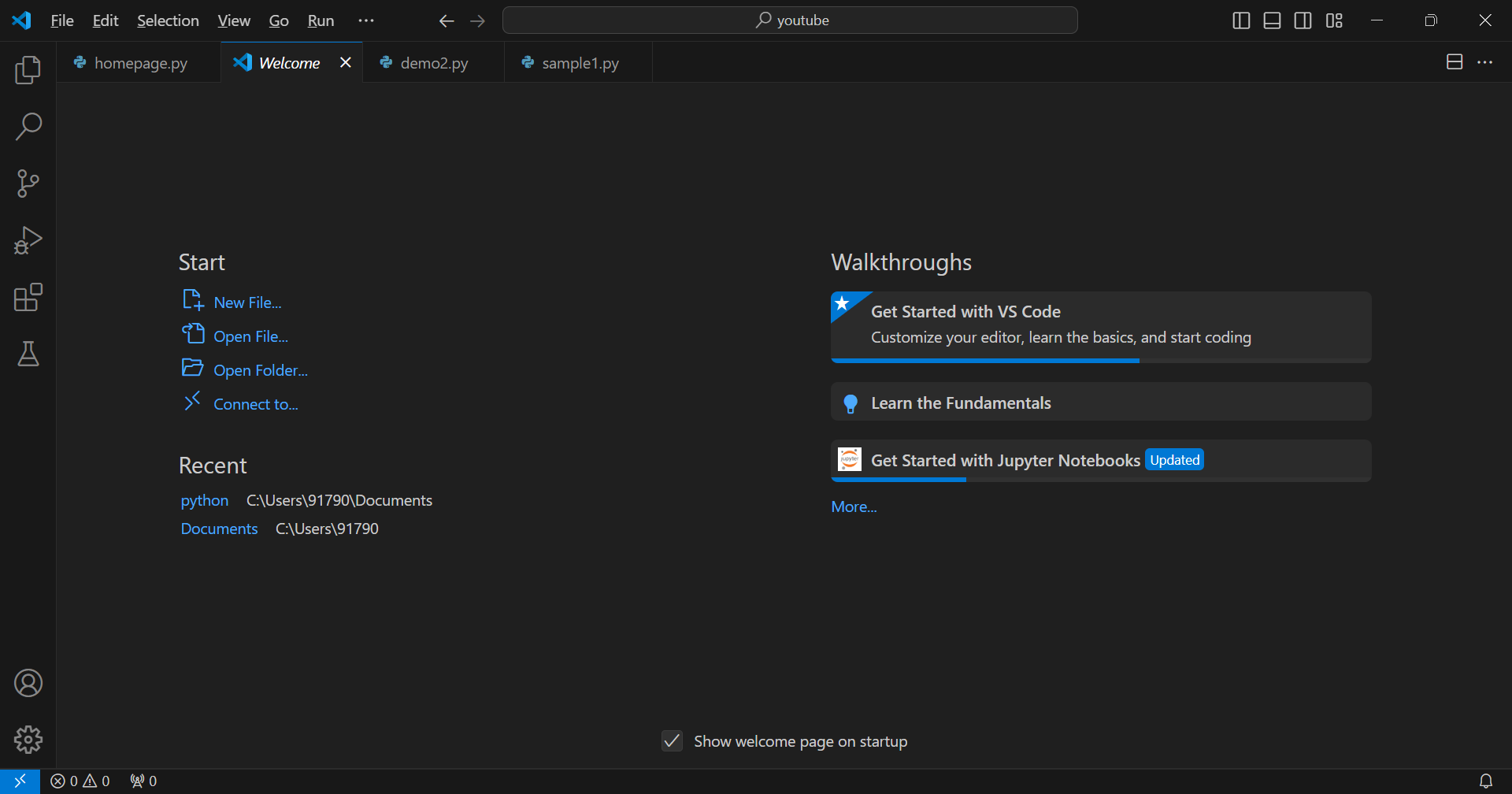
This project extracts the particular youtube channel data by using the youtube channel id, processes the data, and stores it in the MYSQL database. It has analyse the data and give the results depending on the customer questions.

![Intro GUI]( [kanimozhi/Homepage.png at main · kanis11/kanimozhi (github.com)](https://github.com/kanis11/kanimozhi/blob/main/Homepage.png))

**DEVELOPER GUIDE**

**1. Tools Install**

\* Visual Studio code.



Visual Studio Home page

**\* Python 3.12.2 or higher.**

A black screen with white text

Description automatically generated

Python Version

**\* MySQL:**

A screenshot of a computer

Description automatically generated

MYSQL Home page

**\* Youtube API key.**

A screenshot of a computer

Description automatically generated

Create API Key in API Manager

**2. Requirement Libraries to Install:**

\* pip install google-api-python-client, mysql-connector-python, pandas, streamlit.

( pip install google-api-python-client, streamlit-option-menu, PyMySQL , mysql-connector-python, pandas -express streamlit )

**3. Import Libraries:**

# [Youtube API libraries]

#CREATING PARAMETERS FOR THE SYNTAX

#For the ACTIVATION of resource we need Two parameters ( youtube , channel\_ID ) and googlapiclient python library

from googleapiclient.discovery import build

import streamlit as st

# [SQL libraries]

import mysql.connector

# [pandas]

import pandas as pd

#Data visualizaion Libraries

import seaborn as sns

import matplotlib.pyplot as plt

from PIL import Image

#Date and Time Libraries

from datetime import datetime,timedelta

import json

**4. Extract data**

\* Extract the particular youtube channel data by using the youtube channel id, with the help of the youtube API developer console.

a) Process and Transform the data:

After the extraction process, takes the required details from the extraction data.

b) Load data :

After the transformation process, the JSON format data is stored in the SQL database, also It has the option to migrate the data to MySQL database from the MongoDB database.

**5. E D A Process and Framework**

**a) Access MySQL DB :**

Create a connection to the MySQL server and access the specified MySQL DataBase.

**b) Filter the data:**

Filter and process the collected data from the tables depending on the given requirements by using SQL queries and transform the processed data into a DataFrame format.

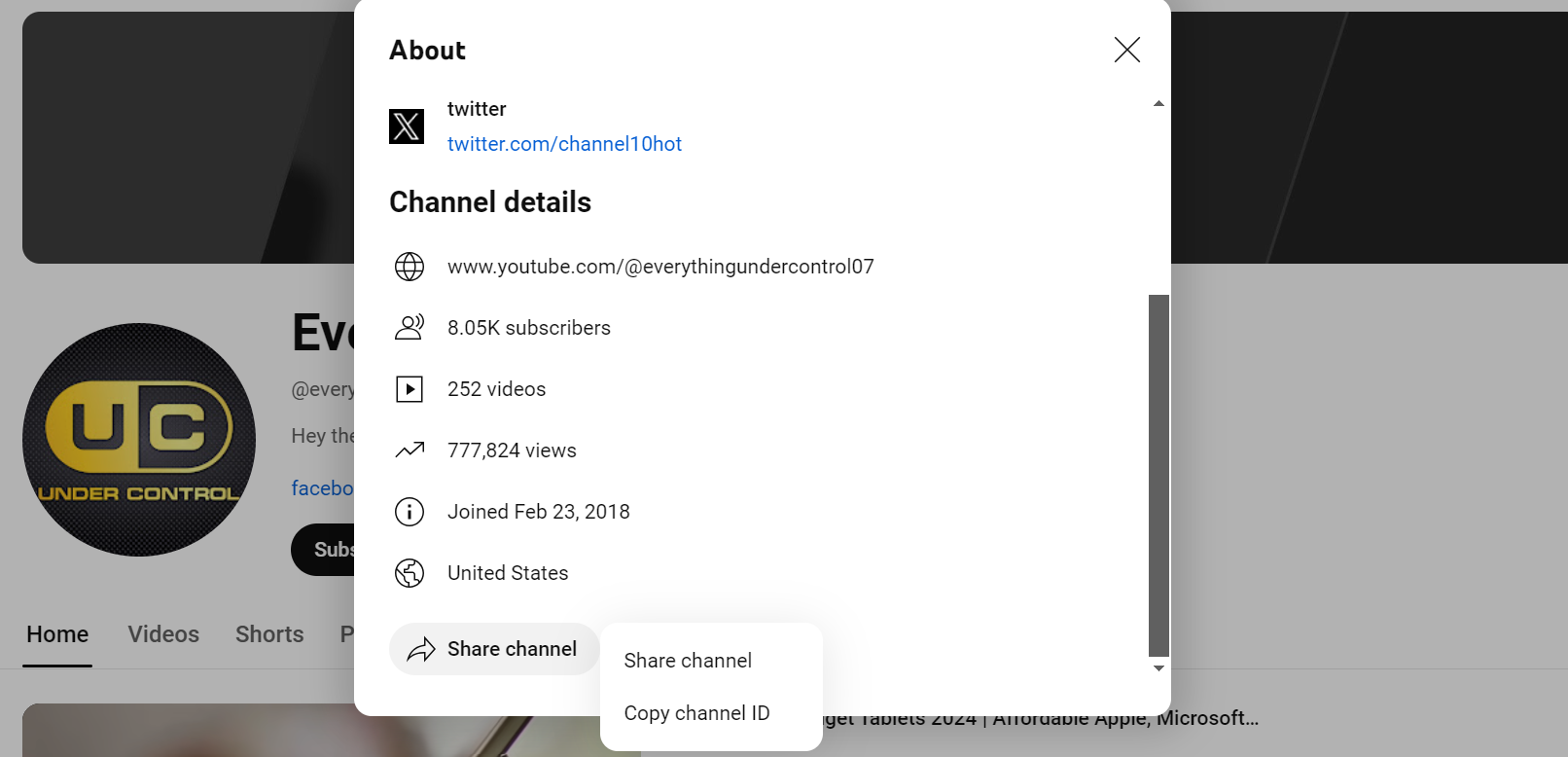
**c) Visualization:**

Finally, create a Dashboard by using Streamlit and give dropdown options on the Dashboard to the user and select a question from that menu to analyse the data and show the output in Dataframe Table and Bar chart.

**USER GUIDE**

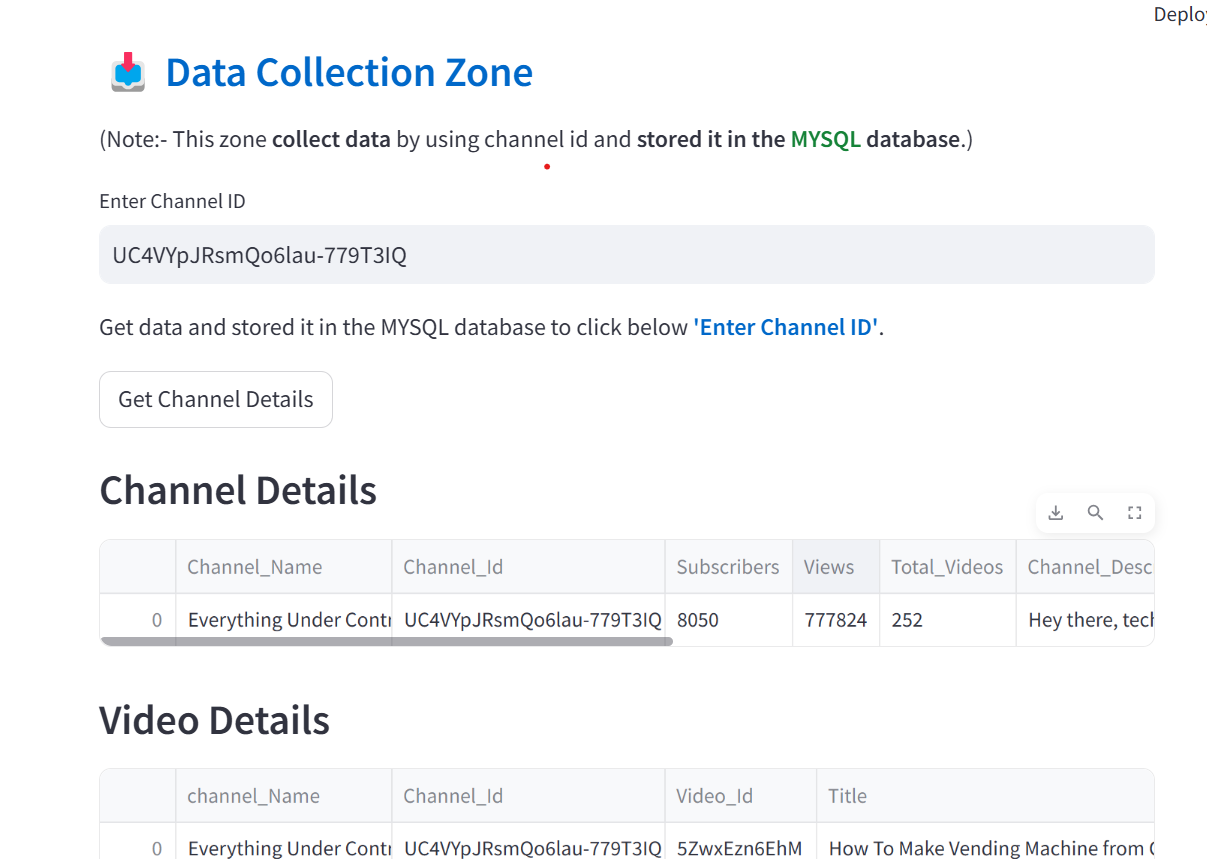
**Data collection zone:**

1.The first step is to collect data from the YouTube using the YouTube Data API. 2.The API and the Channel ID (Extracted from the Channel Page) is used to retrieve channel details, videos details and comment details.



Example of Channel Details from youtube

3.I have used the Google API client library for Python to make requests to the API and the responses are Collected as a Dictionary (Data Collection)

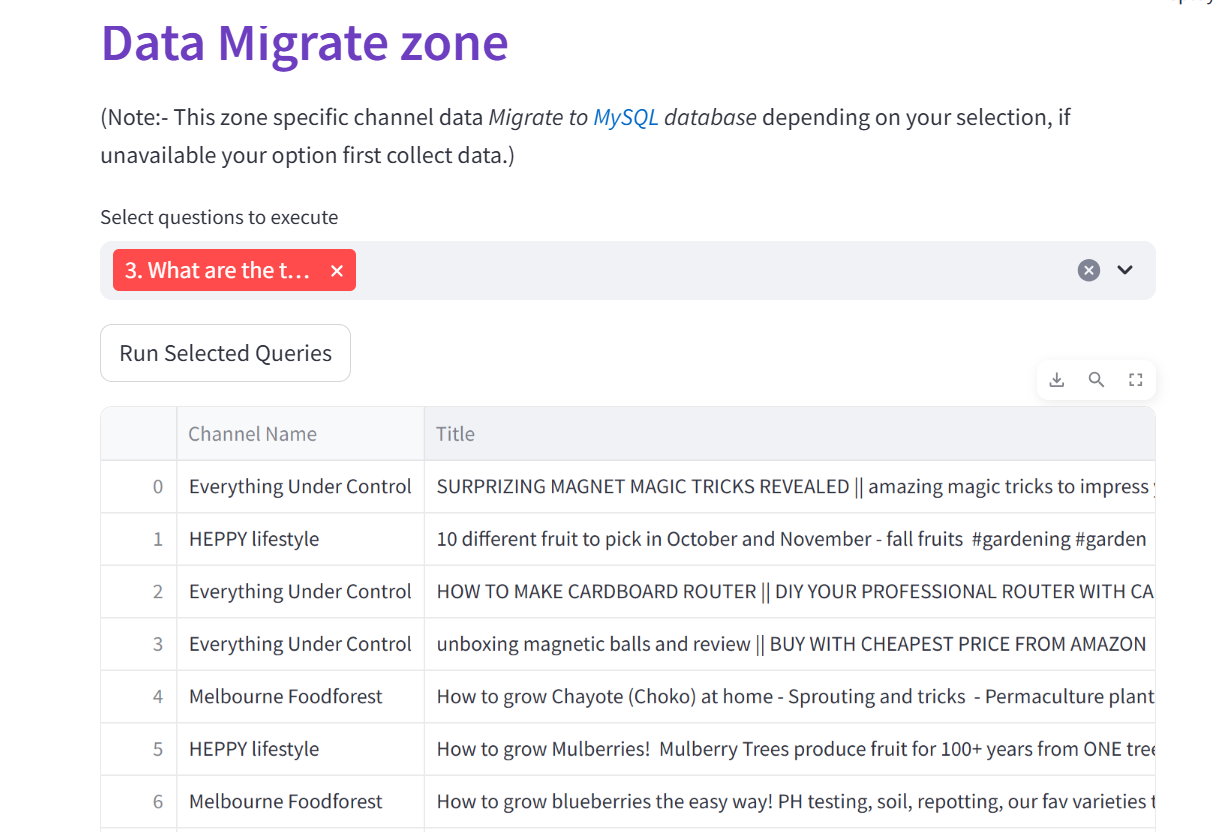


Collect the Channel Information

**Data Analysis and Data Visualization:**

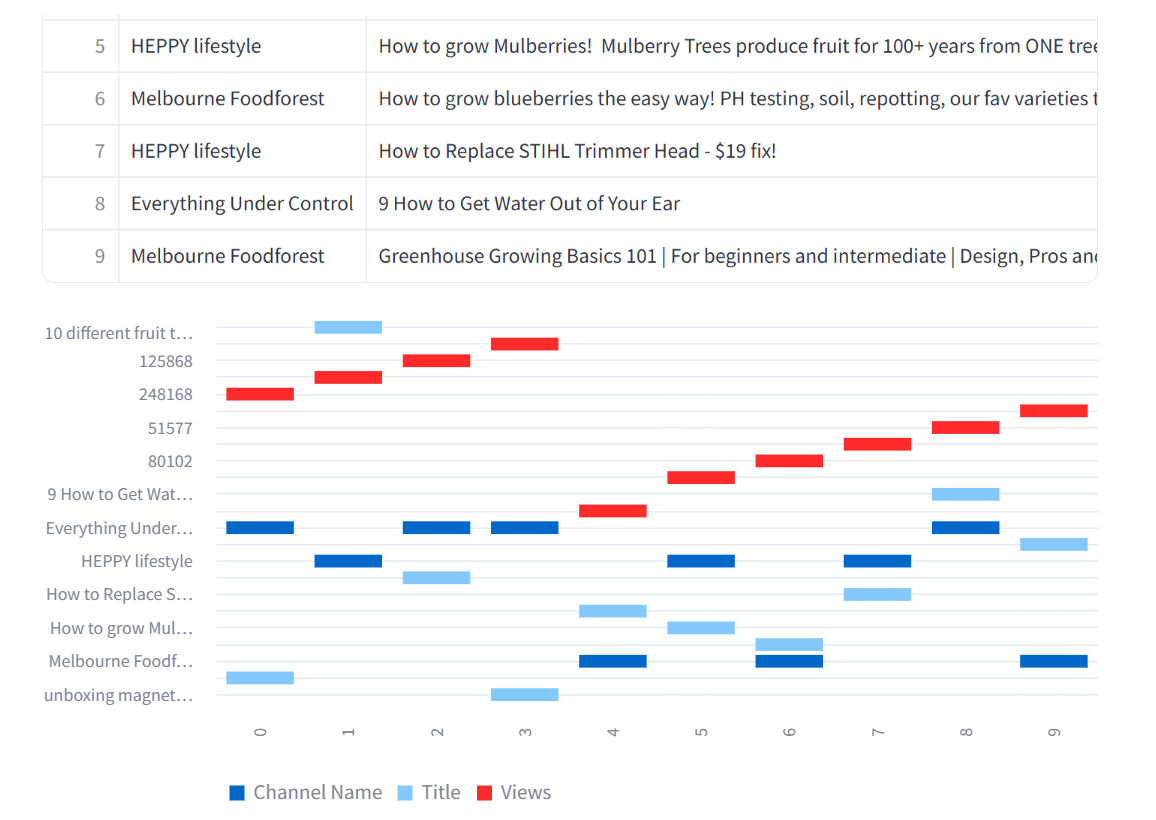
1. Once the Data Collection is done, store it in MYSQL . After Loading all the data. Used SQL queries to join the tables and retrieve data for specific channel

2.With the help of SQL query, I have got many interesting insights about the youtube channels.



Data Retrieved from MYSQL

3.Finally, the data retrieved from the SQL is displayed using the Streamlit Web app. 4.Streamlit is used to create dashboard that allows users to visualize and analyze the data.



Analys the data in Bar-Chart Format

5.Also used Bar\_chart for the Data Visualization to create charts and graphs to analyse the data.

**HOME PAGE** To scrap the YouTube channel details:

**Video link**

\* Click the below Link

[![Intro GUI]( [kanis11/kanimozhi: youtubedataharwesting (github.com)](https://github.com/kanis11/kanimozhi), <https://www.linkedin.com/feed/update/urn:li:ugcPost:7201925977042558976/>