Capstone Project 1

YouTube Data Harvesting and Warehousing

using SQL, MongoDB and Streamlit

GUVI

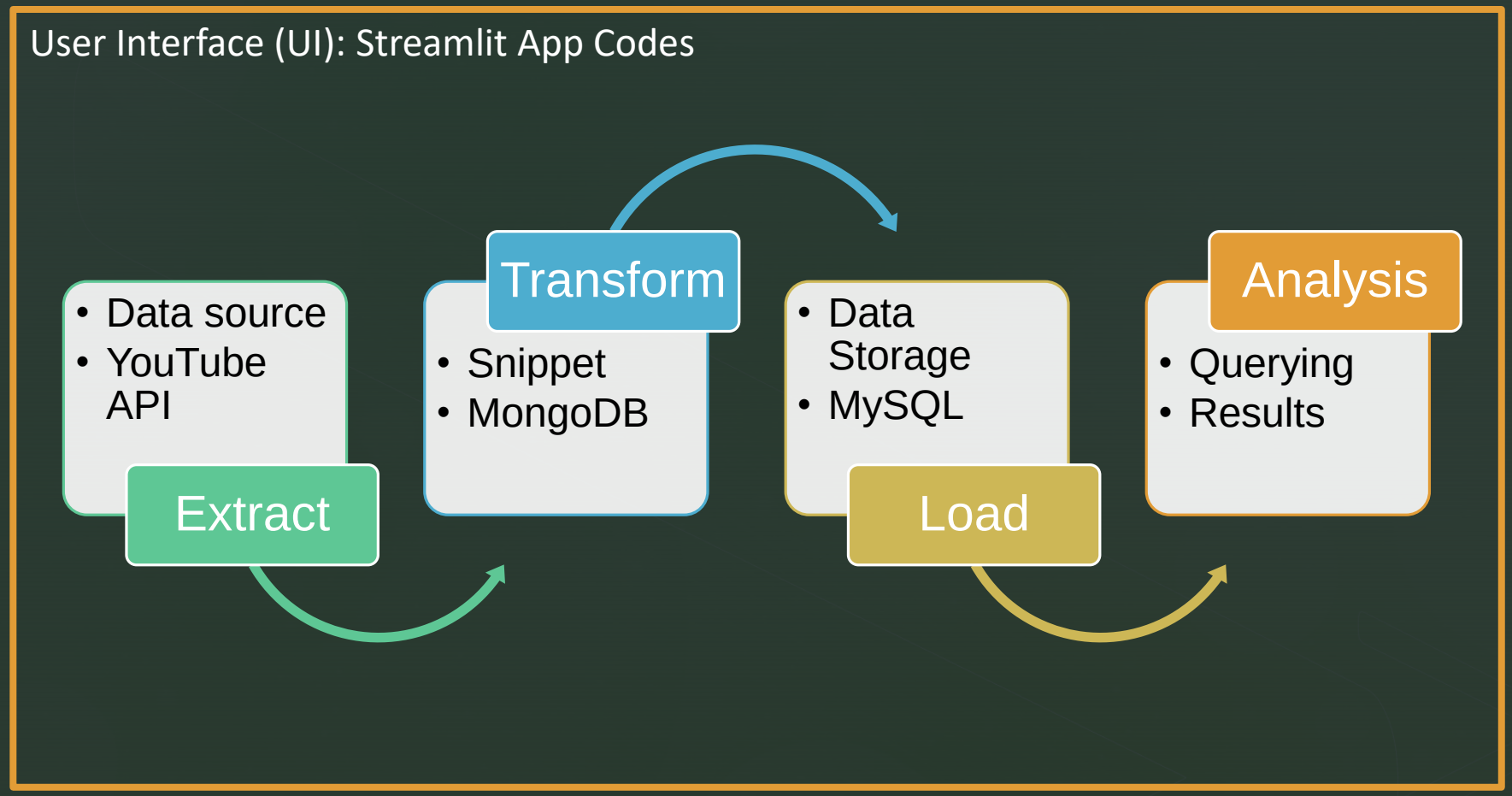
Rajaguru Irusan

Project Overview

Objective

- The project focuses on creating a Streamlit application for accessing and analyzing data from multiple YouTube channels. The application encompasses various features for data extraction, storage in both MongoDB and SQL databases, and querying the SQL database to retrieve data based on questions selected by the user.

Project Workflow



System Requirements

- Visual Studio Code Version: 1.85.1 or later
- MongoDB Compass Version: 1.41.0 or later
- MySQL and MySQL Workbench Version: 8.0.34 or later

Note: Install above tools to execute the project.

PIP Installations

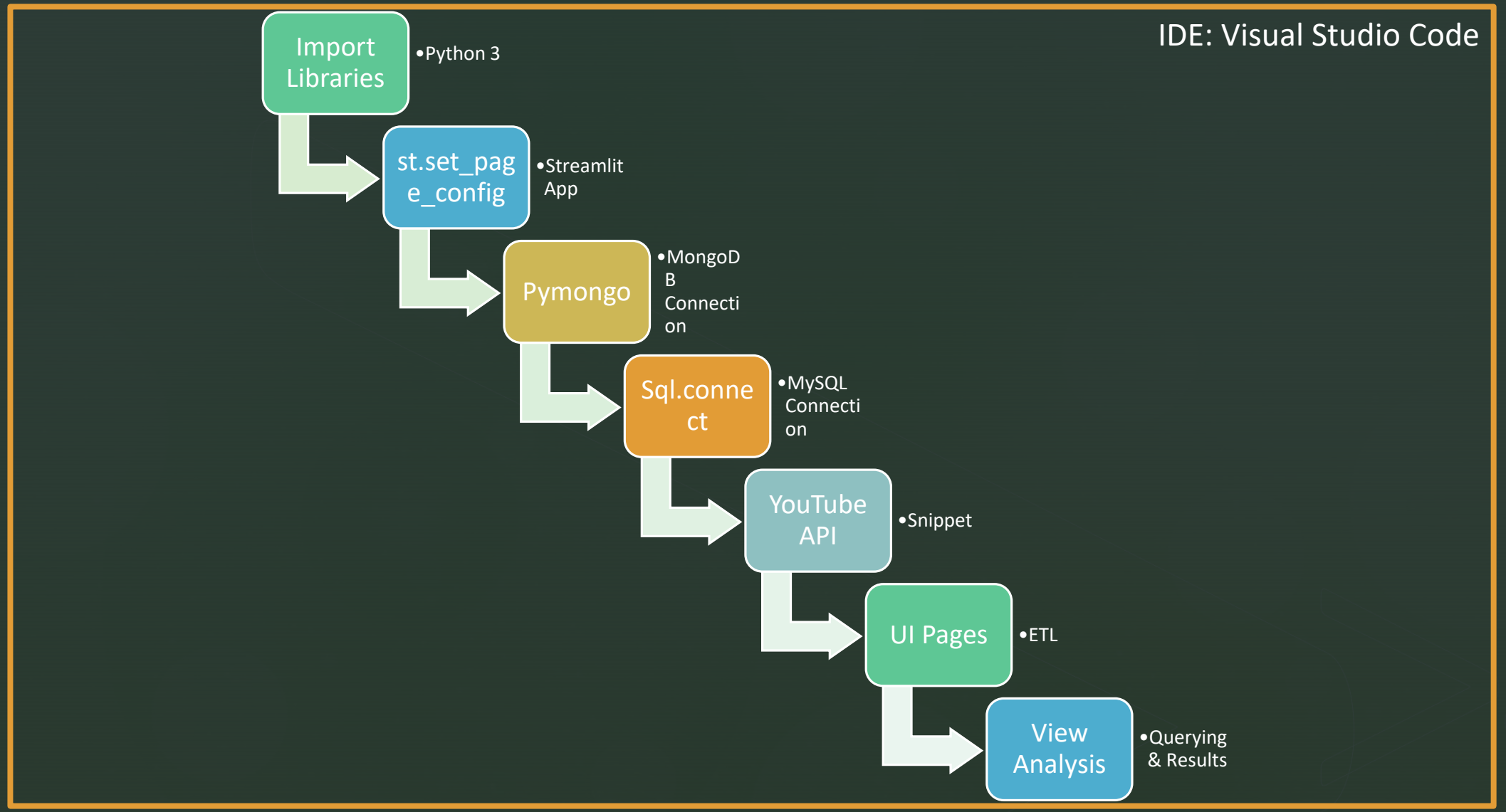
- pip install google-api-python-client
- pip install pandas
- pip install plotly
- pip install plotly.express
- pip install streamlit
- pip install streamlit_option_menu
- pip install mysql
- pip install mysql.connector
- pip install pymongo

Note: Install the above pip in respective IDE terminal (Eg: VS Code terminal or Run in cmd for Windows)

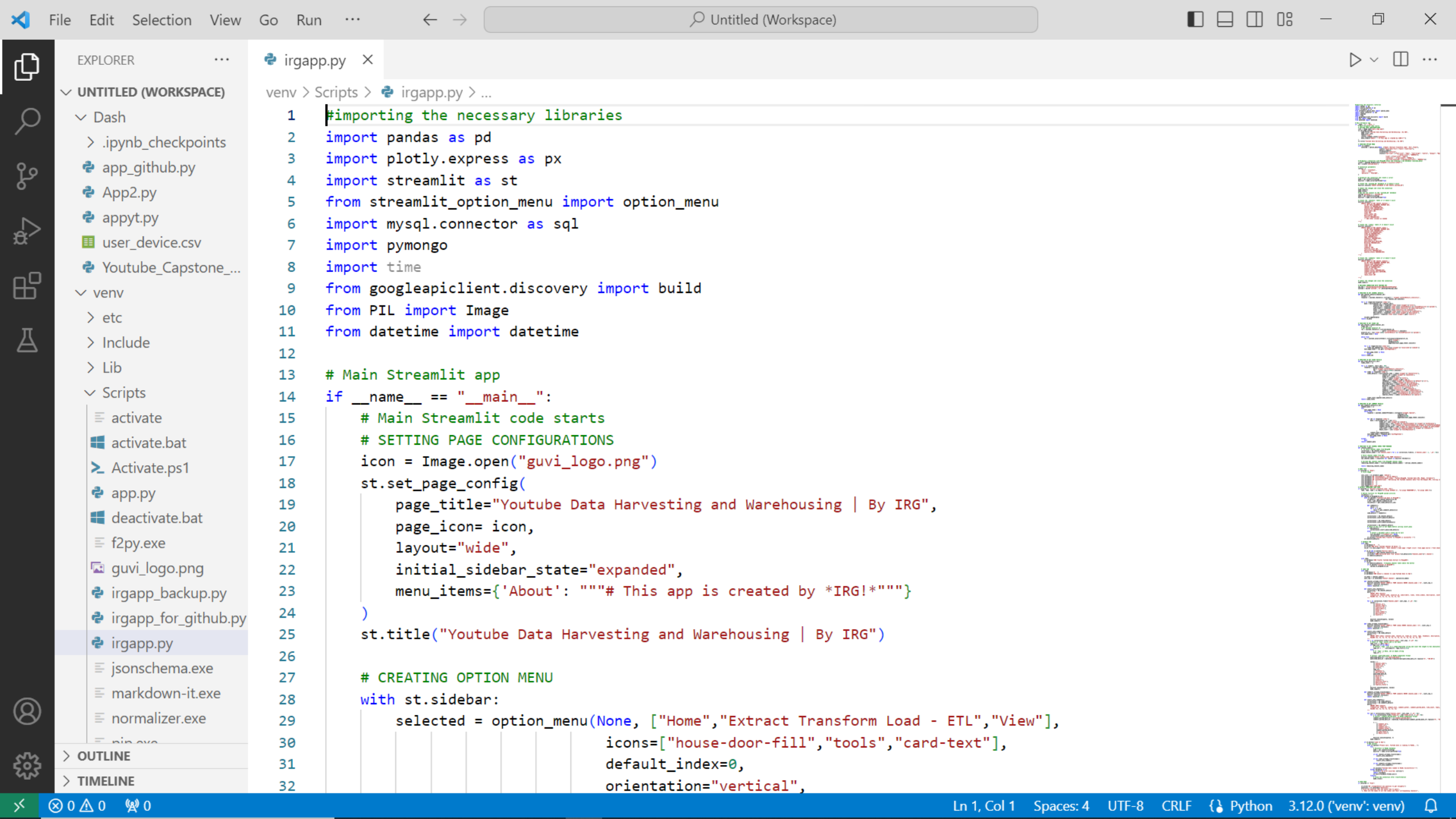
Coding References

- GITHUB Repositories
- YouTube Data API Documentation
- Streamlit Documentation

Coding Flow


IDE: Visual Studio Code
Import Libraries
•Python 3
st.set_page_config
•Streamlit App
Pymongo
•MongoDB Connection
Sql.connect
•MySQL Connection
YouTube API
•Snippet
UI Pages
•ETL
View Analysis
•Querying & Results

Annexures: Screenshots of Results


File Edit Selection View Go Run ···
Untitled (Workspace)
EXPLORER
UNTITLED (WORKSPACE)
Dash
.ipynb_checkpoints
app_github.py
App2.py
appyt.py
user_device.csv
Youtube_Capstone_...
venv
etc
Include
Lib
Scripts
activate
activate.bat
Activate.ps1
app.py
deactivate.bat
f2py.exe
guvi_logo.png
irgapp_backup.py
irgapp_for_github.py
irgapp.py
jsonschema.exe
markdown-it.exe
normalizer.exe
OUTLINE
TIMELINE
irgapp.py
venv > Scripts > irgapp.py > ...
#importing the necessary libraries
import pandas as pd
import plotly.express as px
import streamlit as st
from streamlit_option_menu import option_menu
import mysql.connector as sql
import pymongo
import time
from googleapiclient.discovery import build
from PIL import Image
from datetime import datetime
# Main Streamlit app
if __name__ == "__main__":
# Main Streamlit code starts
# SETTING PAGE CONFIGURATIONS
icon = Image.open("guvi_logo.png")
st.set_page_config(
page_title="Youtube Data Harvesting and Warehousing | By IRG",
page_icon= icon,
layout="wide",
initial_sidebar_state="expanded",
menu_items={'About': """# This app is created by *IRG!*"""}
)
st.title("Youtube Data Harvesting and Warehousing | By IRG")
# CREATING OPTION MENU
with st.sidebar:
selected = option_menu(None, ["Home","Extract Transform Load - ETL","View"],
icons=["house-door-fill","tools","card-text"],
default_index=0,
orientation="vertical",
0 0 0
Ln 1, Col 1 Spaces: 4 UTF-8 CRLF Python 3.12.0 ('venv': venv)

C:\Windows\system32\cmd.exe - streamlit run irgapp.py

```
(venv) C:\Users\Dell\Documents\venv\Scripts>streamlit run irgapp.py

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.1.9:8501
```

Home
Extract Transform Load - ETL
View

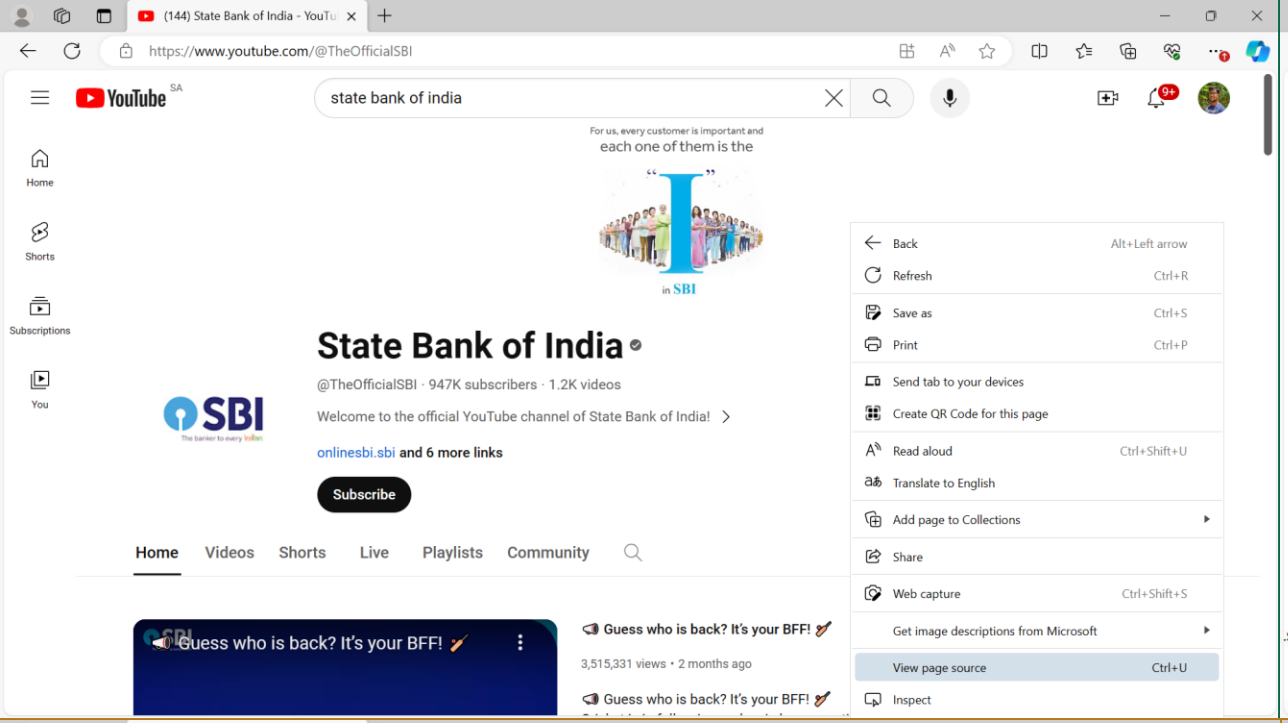
Youtube Data Harvesting and Warehousing | By IRG

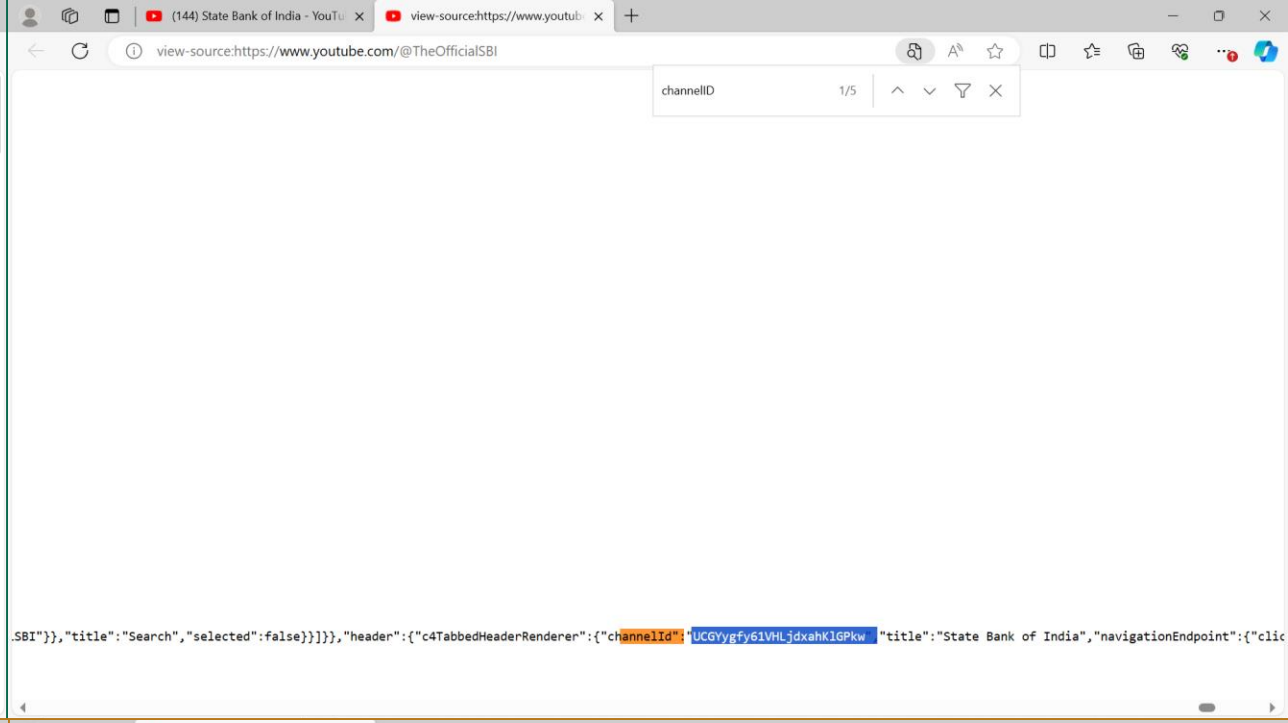
Domain : Social Media

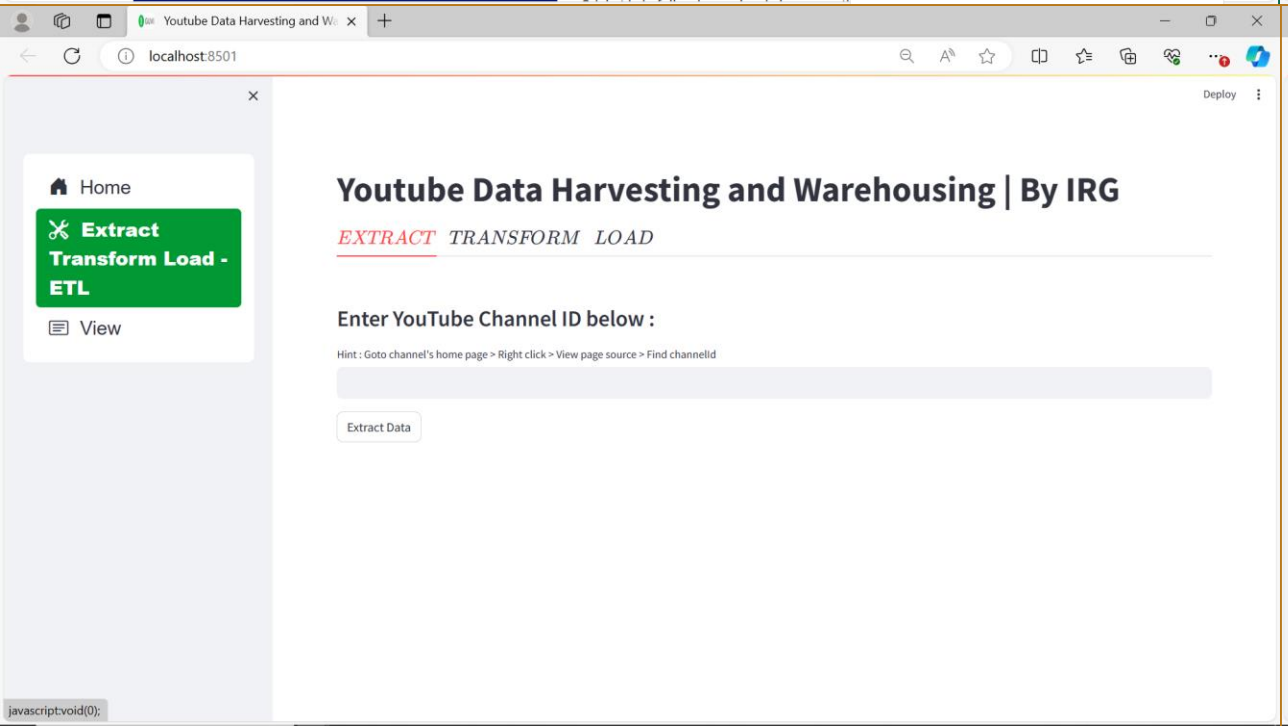
Technologies used :
Python,MongoDB, Youtube Data API, MySql, Streamlit

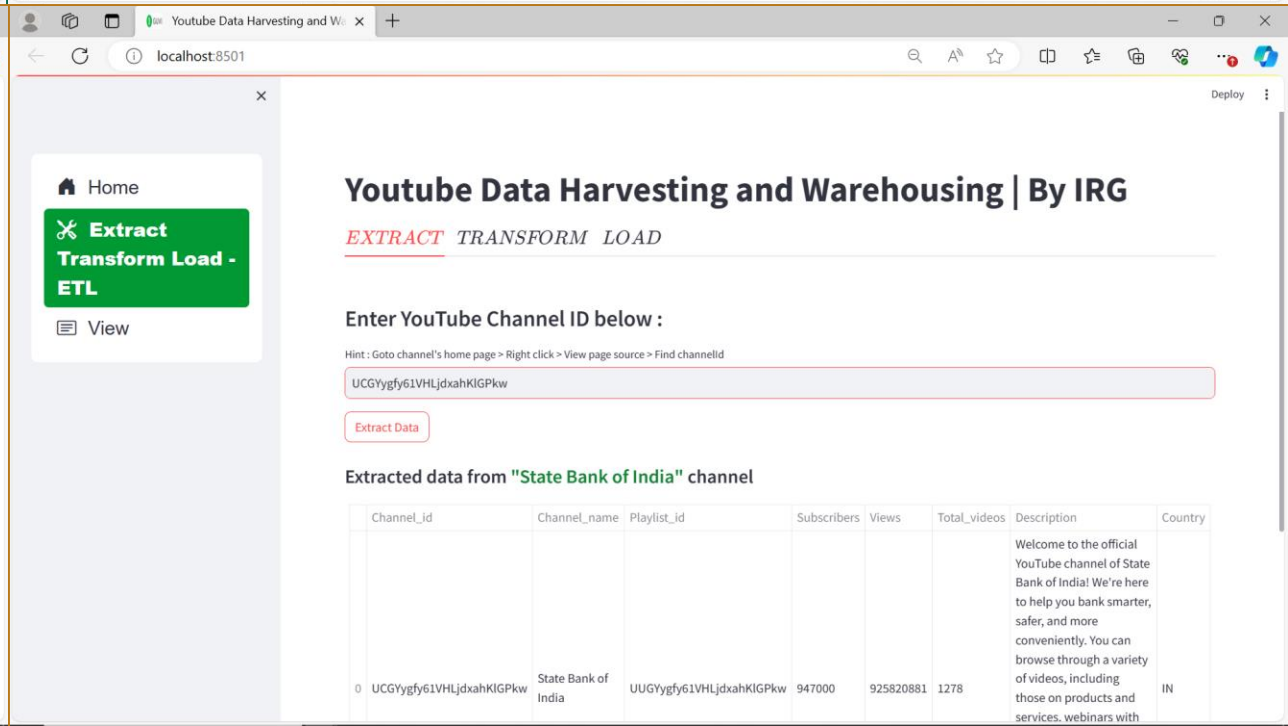
GUVI

Overview : Retrieving the Youtube channels data from the Google API, storing it in a MongoDB as data lake, migrating and transforming data into a SQL database,then querying the data and displaying it in the Streamlit app.


(144) State Bank of India - YouTu
https://www.youtube.com/@TheOfficialSBI
YouTube SA
state bank of india
Home
Shorts
Subscriptions
You
For us, every customer is important and
each one of them is the
in SBI
State Bank of India
@TheOfficialSBI · 947K subscribers · 1.2K videos
Welcome to the official YouTube channel of State Bank of India!
onlinesbi.sbi and 6 more links
Subscribe
Home Videos Shorts Live Playlists Community
Guess who is back? It's your BFF!
Guess who is back? It's your BFF!
3,515,331 views • 2 months ago
Guess who is back? It's your BFF!
Back Alt+Left arrow
Refresh Ctrl+R
Save as Ctrl+S
Print Ctrl+P
Send tab to your devices
Create QR Code for this page
Read aloud Ctrl+Shift+U
Translate to English
Add page to Collections
Share
Web capture Ctrl+Shift+S
Get image descriptions from Microsoft
View page source Ctrl+U
Inspect


(144) State Bank of India - YouTu
view-source:https://www.youtub
view-source:https://www.youtube.com/@TheOfficialSBI
channelID 1/5
.SBI"}},"title":"Search","selected":false}}]}},"header":{"c4TabbedHeaderRenderer":{"channelId":"UCGYygfy61VHLjdxahKlGPkw","title":"State Bank of India","navigationEndpoint":{"clic


Youtube Data Harvesting and W
localhost:8501
Deploy
Home
Extract Transform Load - ETL
View
Youtube Data Harvesting and Warehousing | By IRG
EXTRACT TRANSFORM LOAD
Enter YouTube Channel ID below :
Hint : Goto channel's home page > Right click > View page source > Find channelId
Extract Data
javascript:void(0);


Youtube Data Harvesting and W
localhost:8501
Deploy
Home
Extract Transform Load - ETL
View
Youtube Data Harvesting and Warehousing | By IRG
EXTRACT TRANSFORM LOAD
Enter YouTube Channel ID below :
Hint : Goto channel's home page > Right click > View page source > Find channelId
UCGYygfy61VHLjdxahKlGPkw
Extract Data
Extracted data from "State Bank of India" channel
Channel_id Channel_name Playlist_id Subscribers Views Total_videos Description Country
0 UCGYygfy61VHLjdxahKlGPkw State Bank of India UUGYygfy61VHLjdxahKlGPkw 947000 925820881 1278 Welcome to the official YouTube channel of State Bank of India! We're here to help you bank smarter, safer, and more conveniently. You can browse through a variety of videos, including those on products and services, webinars with IN

Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

EXTRACT *TRANSFORM* *LOAD*

Trasfer YouTube Data Extract to MongoDB

	Channel_id	Channel_name	Playlist_id	Subscribers	Views	Total_videos	Description	Country
0	UCGYygfy61VHLjdxahKlGPkw	State Bank of India	UUGYygfy61VHLjdxahKlGPkw	947000	925820881	1278	Welcome to the official YouTube channel of State Bank of India! We're here to help you bank smarter, safer, and more conveniently. You can browse through a variety of videos, including those on products and services, webinars with industry experts, financial literacy, safe banking and many more. Subscribe and start your journey to a brighter financial future!	IN

Transfer to MongoDB

Youtube Data Harvesting and Wa

localhost:8501

RUNNING... Stop Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

EXTRACT *TRANSFORM* *LOAD*

Trasfer YouTube Data Extract to MongoDB

empty

Transfer to MongoDB

Transferring Data to MongoDB

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

EXTRACT *TRANSFORM* *LOAD*

Trasfer YouTube Data Extract to MongoDB

empty

Transfer to MongoDB

YouTube Data Transfer to MongoDB is successful !!

Channel_id	Channel_name	Playlist_id	Subscribers	Views	Total_videos	Description	Country	_id
						Welcome to the official YouTube channel of State Bank of India! We're here to help you bank smarter, safer, and		

Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

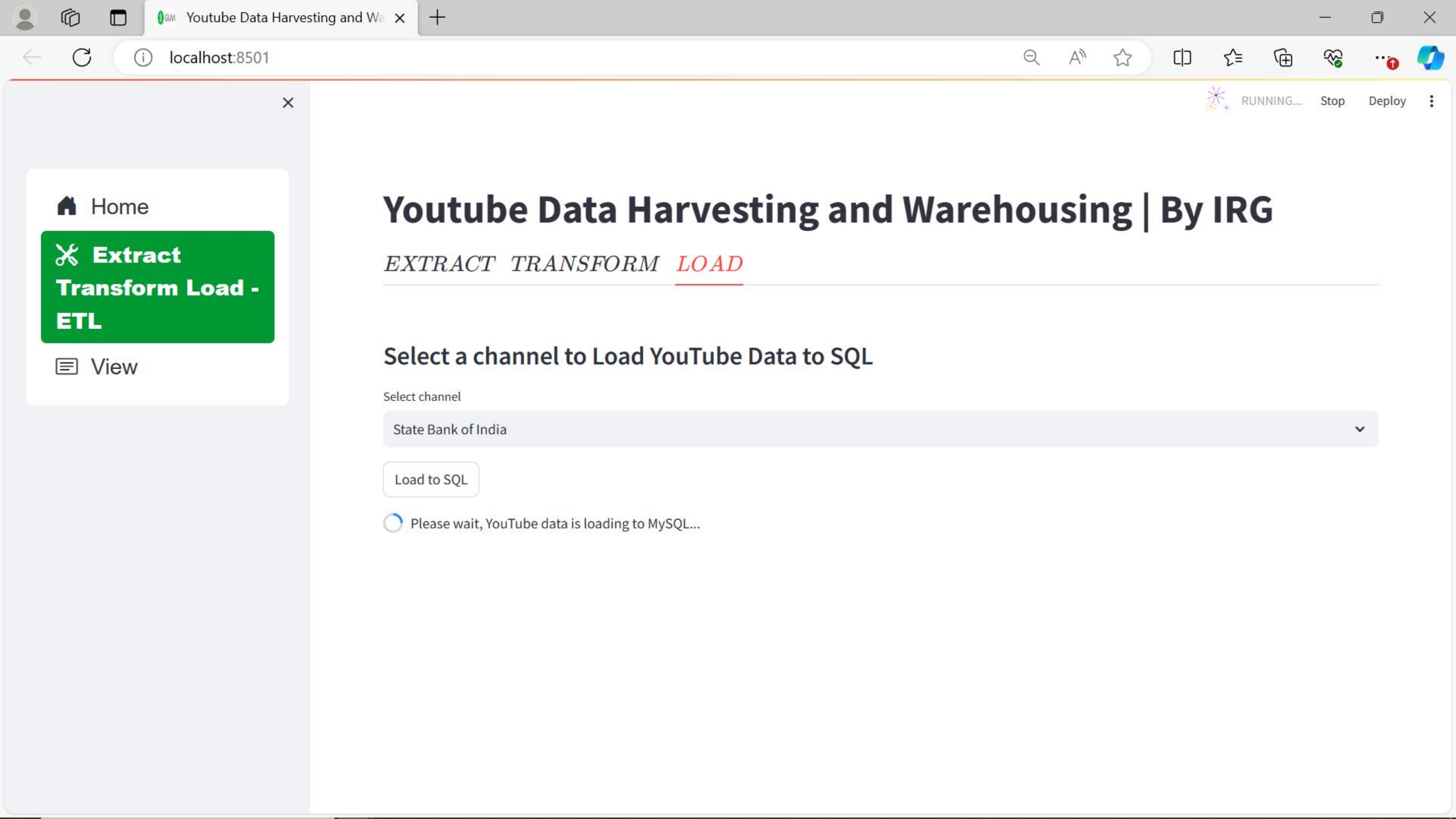
EXTRACT *TRANSFORM* *LOAD*

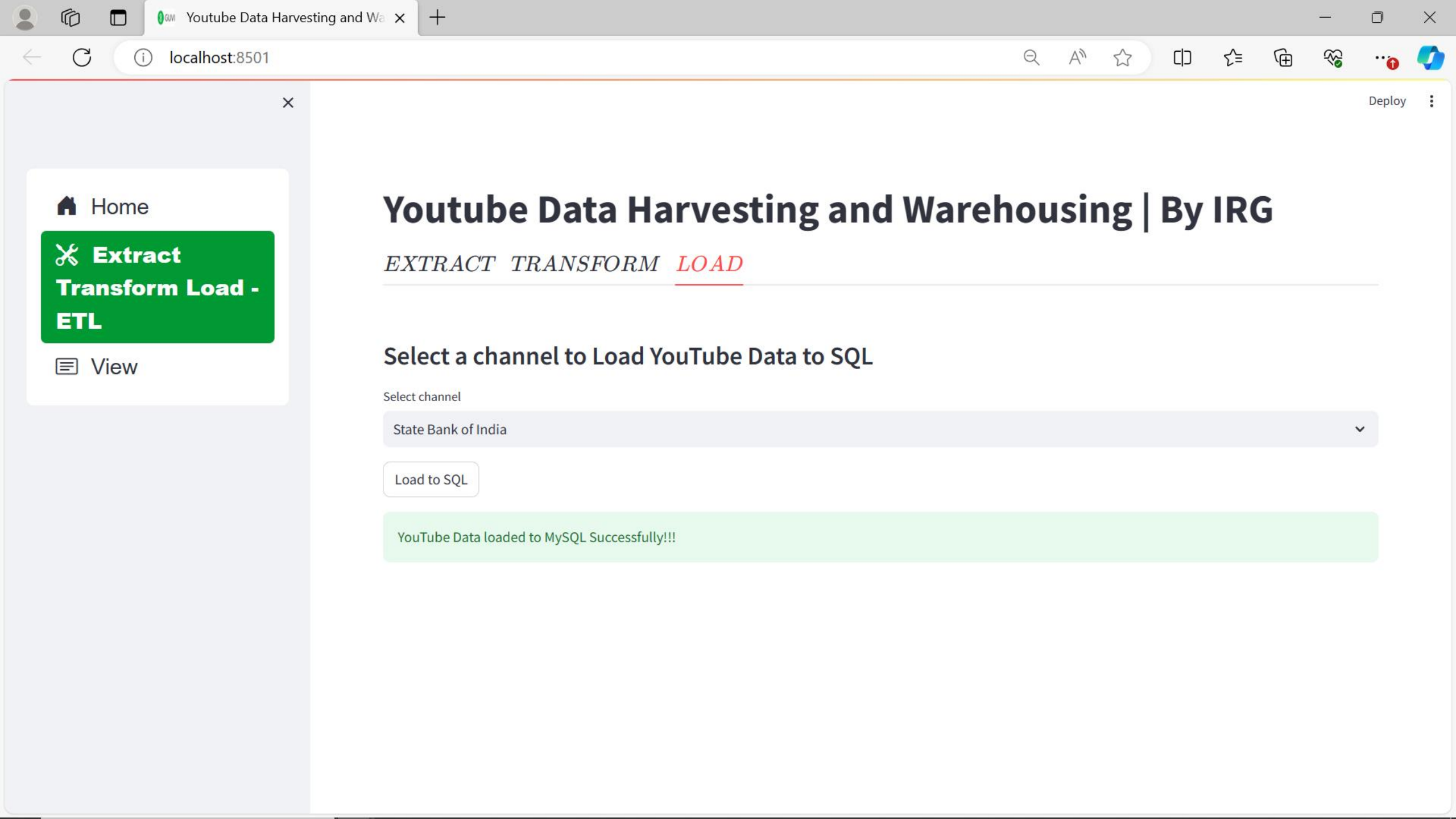
Select a channel to Load YouTube Data to SQL

Select channel

State Bank of India

Load to SQL


Youtube Data Harvesting and Wa
localhost:8501
RUNNING...
Stop
Deploy
Home
Extract Transform Load - ETL
View
Youtube Data Harvesting and Warehousing | By IRG
EXTRACT
TRANSFORM
LOAD
Select a channel to Load YouTube Data to SQL
Select channel
State Bank of India
Load to SQL
Please wait, YouTube data is loading to MySQL...


Youtube Data Harvesting and Wa
localhost:8501
Deploy
Home
Extract Transform Load - ETL
View
Youtube Data Harvesting and Warehousing | By IRG
EXTRACT
TRANSFORM
LOAD
Select a channel to Load YouTube Data to SQL
Select channel
State Bank of India
Load to SQL
YouTube Data loaded to MySQL Successfully!!!

Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

Select any question to get Insights

Questions

Click the question that you would like to query

Click the question that you would like to query

1. What are the names of all the videos and their corresponding channels?
2. Which channels have the most number of videos, and how many videos do they have?
3. What are the top 10 most viewed videos and their respective channels?
4. How many comments were made on each video, and what are their corresponding video names?
5. Which videos have the highest number of likes, and what are their corresponding channel names?
6. What is the total number of likes and dislikes for each video, and what are their corresponding video names?

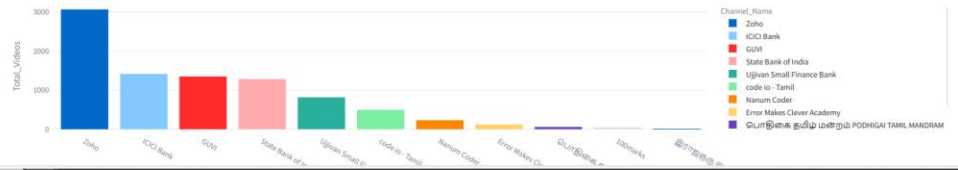
1. What are the names of all the videos and their corresponding channels?

	Video_Title	Channel_Name
1	#Matrix_part_2 - Learn Mathematics in Tamil - Online Class by Ms. Glory Cornelius	100marks
2	#Cramer's rule- Matrix - Learn Mathematics in Tamil - Online Class by Ms. Glory Corne	100marks
3	Structure of C program - Learn Computer in Tamil - Online Class by Cornelius	100marks
4	#Programming Languages - Learn Computer in Tamil - Online Class by Cornelius	100marks
5	#Solved Example Problems on Trigonometry - in Tamil - Online Class by Ms. Glory Cor	100marks
6	#Practice Problems on Trigonometry - in Tamil - Online Class by Ms. Glory Cornelius	100marks
7	#Trigonometry - Learn Mathematics in Tamil - Online Class by Ms. Glory Cornelius	100marks
8	#Transpose Matrix - Learn Mathematics in Tamil - Online Class by Ms. Glory Cornelius	100marks
9	[Episode-14]#Errors and Exceptions in Python(Part - 1)- Learn Python in Tamil - By Dr	100marks
10	#Commutative Property - Learn Mathematics in Tamil - Online Class by Ms. Glory Cor	100marks

2. Which channels have the most number of videos, and how many videos do they have?

	Channel_Name	Total_Videos
1	Zoho	3,051
2	ICICI Bank	1,410
3	GUVI	1,344
4	State Bank of India	1,278
5	Ujjivan Small Finance Bank	814
6	code io - Tamil	496
7	Nanum Coder	230
8	Error Makes Clever Academy	122
9	பொறிகை தமிழ் மன்றம் PODHIGA	61
10	100marks	44
11	இராஜகை இரசன் Rajaguru Irusan	1

Number of videos in each channel :



Youtube Data Harvesting and W

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

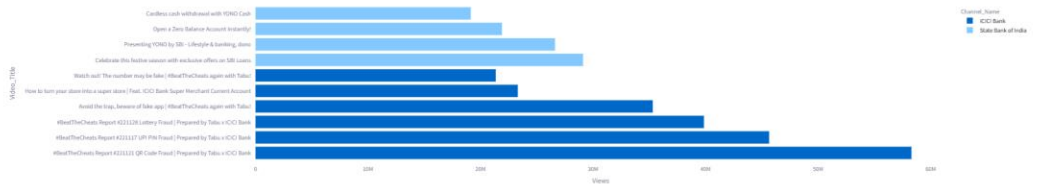
Select any question to get Insights

Questions

3. What are the top 10 most viewed videos and their respective channels?

	Channel_Name	Video_Title	Views
1	ICICI Bank	#BeatTheCheats Report #221121 QR Code Fraud | Prepared by Tabu x ICICI Bank	58,298,666
2	ICICI Bank	#BeatTheCheats Report #221117 UPI PIN Fraud | Prepared by Tabu x ICICI Bank	45,648,300
3	ICICI Bank	#BeatTheCheats Report #221128 Lottery Fraud | Prepared by Tabu x ICICI Bank	39,852,365
4	ICICI Bank	Avoid the trap, beware of fake app | #BeatTheCheats again with Tabu!	35,307,571
5	State Bank of India	Celebrate this festive season with exclusive offers on SBI Loans	29,111,120
6	State Bank of India	Presenting YONO by SBI - Lifestyle & banking, dono	26,616,962
7	ICICI Bank	How to turn your store into a super store | Feat. ICICI Bank Super Merchant Current Ac	23,308,510
8	State Bank of India	Open a Zero Balance Account instantly!	21,905,321
9	ICICI Bank	Watch out! The number may be fake | #BeatTheCheats again with Tabu!	21,351,961
10	State Bank of India	Cardless cash withdrawal with YONO Cash	19,130,521

Top 10 most viewed videos :



Youtube Data Harvesting and W

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

Select any question to get Insights

Questions

4. How many comments were made on each video, and what are their corresponding video names?

	Channel_Name	Video_id	Video_Title	Total_Comments
1	State Bank of India	Vpiq3n32sk8	One Bank Many Solutions	2,119
2	Error Makes Clever Academy	m67-bOpOoPU	Python Tutorial - Python Full Course for Beginners in Tamil	1,392
3	State Bank of India	QqGO7tz7eQ4	We are pleased to onboard MS Dhoni as Brand Ambassador of SBI	1,115
4	Error Makes Clever Academy	FYErehuSuuw	Mastering HTML Basics and Beyond | Your Ultimate Guide to Begin Web Development	992
5	code io - Tamil	fmSnLiAv-zc	C Programming for Beginners in Tamil | Complete Course | code io - Tamil	988
6	State Bank of India	WLecCPJCf_I	Bank Smart With SBI	913
7	State Bank of India	HhqpR-IuP9c	SBI RINB – How to Change Mobile Number Online Without Visiting Branch	850
8	Error Makes Clever Academy	ufRLns-QQh0	Learn Python Programming in a Simplest way 💥	814
9	State Bank of India	J048jBoTrPY	Pay with ease, live with ease.	781
10	Error Makes Clever Academy	LI62YIkEvs8	How to create PERFECT RESUME with zero skills for IT Job Application 🚀 | Resume fi	770

Youtube Data Harvesting and W

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

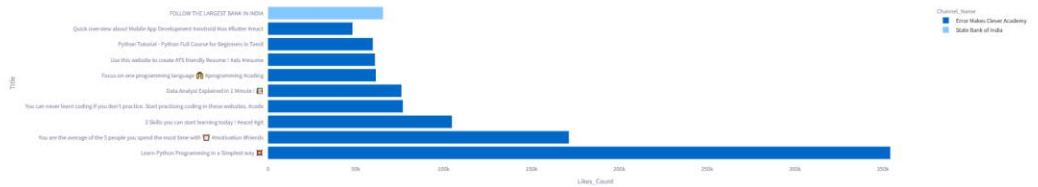
Select any question to get Insights

Questions

5. Which videos have the highest number of likes, and what are their corresponding channel names?

	Channel_Name	Title	Likes_Count
1	Error Makes Clever Academy	Learn Python Programming in a Simplest way 💥	354,208
2	Error Makes Clever Academy	You are the average of the 5 people you spend the most time with ♉ #motivation #fr	171,265
3	Error Makes Clever Academy	3 Skills you can start learning today ! #excel #git	104,679
4	Error Makes Clever Academy	You can never learn coding if you don't practice. Start practising coding in these webs	76,770
5	Error Makes Clever Academy	Data Analyst Explained in 1 Minute ! 🧮	75,964
6	State Bank of India	FOLLOW THE LARGEST BANK IN INDIA	65,466
7	Error Makes Clever Academy	Focus on one programming language 👩‍💻 #programming #coding	61,393
8	Error Makes Clever Academy	Use this website to create ATS friendly Resume ! #ats #resume	60,884
9	Error Makes Clever Academy	Python Tutorial - Python Full Course for Beginners in Tamil	59,673
10	Error Makes Clever Academy	Quick overview about Mobile App Development #android #ios #flutter #react	48,095

Top 10 most liked videos :



Youtube Data Harvesting and W

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

Select any question to get Insights

Questions

6. What is the total number of likes and dislikes for each video, and what are their corresponding video names?

Note: Dislike count was removed by YouTube, Hence counted Likes only

	Channel_Name	Title	Likes_Count
1	Error Makes Clever Academy	Learn Python Programming in a Simplest way 💥	354,208
2	Error Makes Clever Academy	You are the average of the 5 people you spend the most time with ♉ #motivation #fr	171,265
3	Error Makes Clever Academy	3 Skills you can start learning today ! #excel #git	104,679
4	Error Makes Clever Academy	You can never learn coding if you don't practice. Start practising coding in these webs	76,770
5	Error Makes Clever Academy	Data Analyst Explained in 1 Minute ! 🧮	75,964
6	State Bank of India	FOLLOW THE LARGEST BANK IN INDIA	65,466
7	Error Makes Clever Academy	Focus on one programming language 👩‍💻 #programming #coding	61,393
8	Error Makes Clever Academy	Use this website to create ATS friendly Resume ! #ats #resume	60,884
9	Error Makes Clever Academy	Python Tutorial - Python Full Course for Beginners in Tamil	59,673
10	Error Makes Clever Academy	Quick overview about Mobile App Development #android #ios #flutter #react	48,095

Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

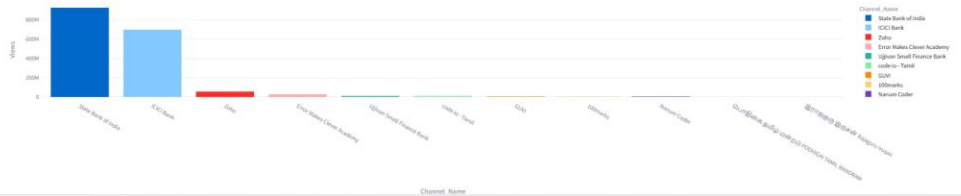
Select any question to get Insights

Questions

7. What is the total number of views for each channel, and what are their corresponding channel names?

	Channel_Name	Views
1	State Bank of India	925,820,881
2	ICICI Bank	698,129,345
3	Zoho	56,253,307
4	Error Makes Clever Academy	27,221,210
5	Ujjivan Small Finance Bank	8,983,731
6	code io - Tamil	5,338,792
7	GUVI	3,614,928
8	100marks	1,409,615
9	Nanum Coder	976,463
10	தொழில்நுட்பத் தமிழ் மன்றம்	3,092
11	இராஜகுரு இருசன் Rajag	228

Channels vs Views :



Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

Select any question to get Insights

Questions

8. What are the names of all the channels that have published videos in the year 2022?

	Channel_Name
1	100marks
2	code io - Tamil
3	GUVI
4	ICICI Bank
5	Nanum Coder
6	State Bank of India
7	Ujjivan Small Finance Bank
8	Zoho

Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

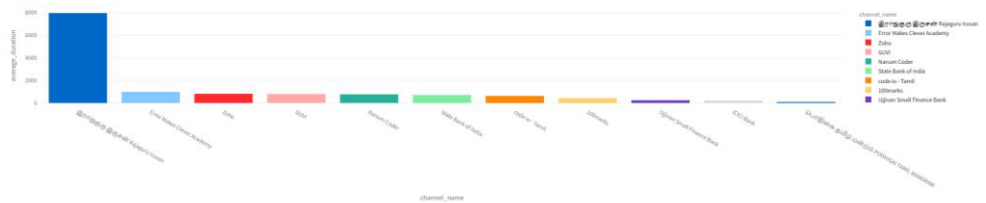
Select any question to get Insights

Questions

9. What is the average duration of all videos in each channel, and what are their corresponding channel names?

	channel_name	average_duration
1	இராஜகுரு இருசன் Rajaguru Irusan	7,952
2	Error Makes Clever Academy	984
3	Zoho	815
4	GUVI	793
5	Nanum Coder	770
6	State Bank of India	709
7	code io - Tamil	623
8	100marks	432
9	Ujjivan Small Finance Bank	246
10	ICICI Bank	216

Average video duration for channels :



Youtube Data Harvesting and Wa

localhost:8501

Deploy

Home

Extract Transform Load - ETL

View

Youtube Data Harvesting and Warehousing | By IRG

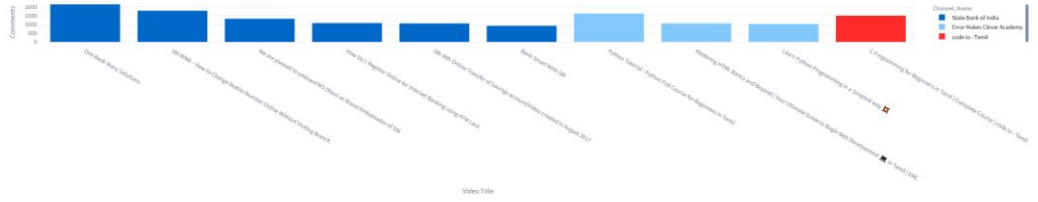
Select any question to get Insights

Questions

10. Which videos have the highest number of comments, and what are their corresponding channel names?

	Channel_Name	Video Title	Video_ID	Comments
1	State Bank of India	One Bank Many Solutions	Vpiq3n32rk8	2,146
2	State Bank of India	SBI RING - How to Change Mobile Number Online Without Visiting Branch	HhqzR-IuP9c	1,784
3	Error Makes Clever Academy	Python Tutorial - Python Full Course for Beginners in Tamil	m67-0OpOoPU	1,622
4	code io - Tamil	C Programming for Beginners in Tamil | Complete Course | code io - Tamil	fmSmLiAv-zc	1,498
5	State Bank of India	We are pleased to onboard MS Dhoni as Brand Ambassador of SBI	QqSOTz7eQ4	1,325
6	State Bank of India	How Do I: Register Online for Internet Banking using ATM card	VAX3QPUd1XM	1,076
7	Error Makes Clever Academy	Mastering HTML Basics and Beyond | Your Ultimate Guide to Begin Web Development	FYEreNuSuuw	1,063
8	State Bank of India	SBI IMB: Online Transfer of Savings account(Video created in August 2017	JQgYELH5NIM	1,055
9	Error Makes Clever Academy	Learn Python Programming in a Simplest way	uRRLms-QQh0	1,022
10	State Bank of India	Bank Smart With SBI	WLecCPJCf_I	920

Videos with most comments :



Thank You!!!