**Youtube\_Data\_Harvesting\_and\_Warehousing**

**Domain : Social Media**

**Problem Statement :**

The problem statement is to create a Streamlit application that allows users to access and analyze data from multiple YouTube channels. Extracting data using Youtube API and storing it on MongoDB then Transforming it to a relational databaselike MySQL. For getting various info about youtube channels.

**Technologies used :**

* [Python](https://www.python.org/)
* [MongoDB](https://www.mongodb.com/atlas/database)
* [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/)
* [Plotly](https://plotly.com/python/)

**Overview :**

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.

**Workflow :**

* Connect to the YouTube API this API is used to retrieve channel, videos, comments data. I have used the Google API client library for Python to make requests to the API.
* The user will able to extract the Youtube channel's data using the Channel ID. Once the channel id is provided the data will be extracted using the API.
* Once the data is retrieved from the YouTube API, I've stored it in a MongoDB as data lake. MongoDB is a great choice for a data lake because it can handle unstructured and semi-structured data easily.
* After collected data for multiple channels,it is then migrated/transformed it to a structured MySQL as data warehouse.
* Then used SQL queries to join the tables in the SQL data warehouse and retrieve data for specific channels based on user input.
* With the help of SQL query I have got many interesting insights about the youtube channels.
* Finally, the retrieved data is displayed in the Streamlit app. Also used Plotly's data visualization features to create charts and graphs to help users analyze the data.
* Overall, this approach involves building a simple UI with Streamlit, retrieving data from the YouTube API, storing it in a MongoDB datalake, migrating it to a SQL data warehouse, querying the data warehouse with SQL, and displaying the data in the Streamlit app.