

# **YOUTUBE DATA HARVESTING**

**BY**

**GOWCIGAN M**



# INTRODUCTION

Welcome to our presentation on harvesting YouTube data. In this presentation, we will explore the various ways in which YouTube data can be collected and analyzed to gain insights into user behavior and preferences. With the explosive growth of video content on the platform, understanding YouTube data has become increasingly important for businesses and content creators alike. Join us as we dive into this fascinating topic.



# YOUTUBE API

The YouTube API is a tool that allows developers to access and interact with YouTube data, including videos, channels, and playlists. This API can be used for data harvesting, which involves collecting and analyzing data from YouTube in order to gain insights and make informed decisions.

Using the YouTube API, developers can retrieve information about videos, such as their titles, descriptions, and view counts, as well as information about channels and playlists. This data can be analyzed to identify trends and patterns, such as which types of videos are most popular or which channels are growing the fastest.

In addition to retrieving data, the YouTube API also allows developers to perform actions such as uploading videos, managing comments, and creating playlists. This can be useful for content creators or businesses looking to automate certain tasks on their YouTube channels.

Overall, the YouTube API provides a powerful tool for data harvesting and analysis, as well as for managing and automating tasks on YouTube channels.

# DATA EXTRACTION TECHNIQUES

YouTube is a goldmine of data for marketers, researchers, and businesses. Here are some of the most effective techniques for extracting data from YouTube:

- **Web Scraping:** This involves using a web scraper to automatically extract data from YouTube. This can include video titles, descriptions, view counts, and more.
- **APIs:** YouTube provides APIs that allow developers to programmatically access YouTube data. This can be a more efficient and reliable way to extract data compared to web scraping.
- **Third-Party Tools:** There are many third-party tools available that can help extract data from YouTube. These tools can range from simple web scrapers to more complex data analytics platforms.

# STORE IN THE MONGODB DATA LAKE

- Once the YouTube data has been harvested, it can be stored in a Mongodb data lake for further analysis and processing. Mongodb is a NoSQL database that is designed to handle unstructured data like the data harvested from YouTube. The data can be stored in JSON format, making it easy to work with and analyze.

# **MIGRATE FROM MONGODB TO SQL DATA WAREHOUSE**

1. Assess your current data model and identify any changes needed to fit a SQL data warehouse
2. Export data from Mongodb using a tool like mongoexport or a custom script
3. Transform the data into a format suitable for a SQL database using ETL tools or custom scripts
4. Load the transformed data into the SQL data warehouse using a tool like SQL Server Integration Services (SSIS) or custom scripts

# QUERY THE SQL DATA BASE

- SQL database definition includes structured data organized into tables, defining their schema, relationships, constraints, and data types. It utilizes SQL (Structured Query Language) for managing, querying, and manipulating data, ensuring efficient storage, retrieval, and management of information.

# DISPLAY THE DATA IN STREAMLIT APPLICATION

- Once the YouTube data has been harvested and preprocessed, it can be displayed in a user-friendly way using the Streamlit app. Streamlit is a Python library that allows for the creation of interactive web apps with minimal code. With Streamlit, we can create a dashboard that displays the YouTube data in a variety of formats, including tables, charts, and visualizations.





THANK YOU