**Project Report: YouTube Data Analysis and Visualization with Power BI**

**Project Overview**

This project involves extracting data from the YouTube API, cleaning and transforming the data using SQL and Python, and then visualizing the insights using Power BI. The goal is to create a comprehensive dashboard that provides valuable insights into the YouTube data.

**Data Extraction**

**Tools Used: YouTube API, PowerBI, SQL, Python**

1. API Setup:

- Registered an application with YouTube and obtained API keys.

- Used Python's requests library to fetch data from the YouTube API.

- Extracted data points such as video titles, view counts, like counts, comment counts, and publication dates.

2. Script Example:

python

import requests

API\_KEY = 'YOUR\_YOUTUBE\_API\_KEY'

CHANNEL\_ID = 'YOUR\_CHANNEL\_ID'

url = f"https://www.googleapis.com/youtube/v3/search?key={API\_KEY}&channelId={CHANNEL\_ID}&part=snippet,id&order=date&maxResults=20"

response = requests.get(url).json()

videos = response['items']

Extract required data

data = []

for video in videos:

video\_data = {

'title': video['snippet']['title'],

'view\_count': video['statistics']['viewCount'],

'like\_count': video['statistics']['likeCount'],

'comment\_count': video['statistics']['commentCount'],

'published\_at': video['snippet']['publishedAt']

}

data.append(video\_data)

Data Cleaning and Transformation

Tools Used: SQL, Python (Pandas)

1. Data Cleaning with Python:

- Loaded the raw data into a Pandas DataFrame.

- Handled missing values and data type conversions.

- Example:

python

import pandas as pd

df = pd.DataFrame(data)

df['published\_at'] = pd.to\_datetime(df['published\_at'])

df['view\_count'] = df['view\_count'].astype(int)

df['like\_count'] = df['like\_count'].astype(int)

df['comment\_count'] = df['comment\_count'].astype(int)

2. Data Transformation with SQL:

- Imported the cleaned data into a SQL database for further transformation.

- Created SQL queries to aggregate data, calculate additional metrics, and prepare the data for visualization.

- Example:

sql

SELECT

title,

view\_count,

like\_count,

comment\_count,

published\_at,

(like\_count / view\_count) 100 AS like\_ratio,

(comment\_count / view\_count) 100 AS comment\_ratio

FROM youtube\_data

Data Visualization

Tool Used: Power BI

1. Importing Data:

- Connected Power BI to the SQL database to import the transformed data.

2. Creating Measures:

- Used DAX to create complex measures for detailed insights.

- Example:

DAX

TotalLikes = SUM(youtube\_data[like\_count])

AverageViewCount = AVERAGE(youtube\_data[view\_count])

LikeRatio =

VAR TotalLikes = SUM(youtube\_data[like\_count])

VAR TotalViews = SUM(youtube\_data[view\_count])

RETURN DIVIDE(TotalLikes, TotalViews) 100

3. Building the Dashboard:

- Created multiple visualizations, including bar charts, line charts, and pie charts to represent data such as:

- Total views, likes, and comments over time.

- Average view count per video.

- Like and comment ratios.

- Popular videos based on views and engagement.

4. Example Visualizations:

- Total Views Over Time: A line chart showing how view counts have changed over time.

- Top Videos: A bar chart highlighting the most viewed and liked videos.

- Engagement Ratios: Pie charts displaying like and comment ratios relative to total views.

Conclusion

This project showcases the end-to-end process of extracting, cleaning, transforming, and visualizing data. By leveraging the YouTube API, SQL, Python, and Power BI, we were able to create a dynamic and interactive dashboard that provides actionable insights into YouTube channel performance.

GitHub Repository

The complete project, including all scripts and the Power BI dashboard file, is committed to the GitHub repository. The repository structure is as follows:

/YouTube-Data-Analysis

├── data\_extraction

│ └── youtube\_api.py

├── data\_cleaning

│ └── data\_cleaning.py

├── data\_transformation

│ └── transformation.sql

├── visualization

│ └── youtube\_dashboard.pbix

├── README.md

The README.md file includes detailed instructions on how to set up the environment, run the scripts, and open the Power BI dashboard.

Repository Link

[GitHub Repository: YouTube Data Analysis](https://github.com/yourusername/YouTube-Data-Analysis)

Feel free to explore and contribute to the project!