

YouTube Data: Scrapping, Analyzing, and Visualizing.

Submitted to
LOVELY PROFESSIONAL UNIVERSITY
PHAGWARA, PUNJAB



LOVELY
PROFESSIONAL
UNIVERSITY

Master of Computer Application

Submission Date: **19th Oct, 2022**

Submitted By:

Suraj Vishwakarma

12115712

Submitted to:

Girish Kumar

To whom so ever it may concern

I, **Suraj Vishwakarma, 12115712**, hereby declare that the work done by me on **“YouTube Data: Scraping, Analyzing and Visualizing”** from **September 2022** to **October 2022**, **Lovely Professional University, Phagwara, Punjab**, is a record of original work for the partial fulfillment of the requirements for the award of the degree, **Master of Computer Application**.

Suraj Vishwakarma (**12115712**)

Dated: 2nd Oct, 2022

ACKNOWLEDGEMENT

Phagwara, Punjab

1st October 2022

I would like to convey my heartfelt gratitude to Mr. Girish Kumar Sir for his tremendous direction and assistance in the completion of my project.

I would also like to thank him for providing me this wonderful opportunity to work on a project with the topic YouTube Data Analysis. This project would not have been accomplished without their help and insights.

Suraj

Vishwakarma

12115712

Table of Contents

Sr. No.	Title	Page No.
1	Introduction:	5
	i. Purpose	5
	ii. Overview	5
	iii. Scope	5
2	Overall Description:	6
	i. Analysis Objective	6
	ii. Operative Environment	6
3	Functional Requirements	6
4	Time Taken	6
5	Non-Functional Requirements:	7
	i. Performance Requirements	7
	ii. Security Requirements	7
	iii. Analysis Quality Attributes	7
6	Design Phase:	7
	i. DFD (Data Flow Diagram)	8-10
	ii. Flow Chart	11
7	Bibliography/ References:	12

Introduction:

1.1 Purpose:

This document provides the requirement specifications for **YouTube Data Analysis for Tech Channel** and Visualizing the YouTube Data. It specifies user interface attributes, functional and nonfunctional requirements, and long-term ideas for the evolution of the system.

1.2 Overview:

It's a data science analysis project in which analysis done in python with the help of YouTube data API v3, pandas and seaborn. In which,

- a) I extracted the data from YouTube by using YouTube API.
- b) Extracting its name.
- c) Total views they have got.
- d) Total Subscribers they have.
- e) Comparing few Tech Channels and comparing their data to measure their growth.
- f) Then, I extracted all the videos details of a particular channel.
- g) Extracting Its videos Views, Title, Total Comments, Likes and Published Date.
- h) Then Analyzing and Visualizing them with the help of seaborn.

Once finished, analysis is exported into the csv file.

1.3 Scope:

-----→ **This is a YouTube data analysis project.** ←-----

YouTube is a website designed for sharing video. Millions of users around the world have created accounts on the site that allow them to upload videos that anyone can watch. Through YouTube people can watch videos, share videos, Upload videos and can even download them.

And with the help of YouTube data analysis, anyone can easily understand the statistics and differences between the channel's growth
And get the whole idea about any channels subscribers, videos count, particular video data, most watched videos and many more.

2. Overall Description

2.1 Analysis Objective:

YouTube is an American Video-sharing website headquarter in San Bruno California. YouTube allows users to upload, view, rate, share, add to favorites, report, and comment on videos, subscribe to other users. It offers a wide variety of user-generated and corporate media videos. Available content includes video clips, TV show clips, music videos, short and documentary films, audio recordings, movie trailers, live streams, and other content such as video blogging, short original videos, and educational videos.

2.2 Operating Environment:

YouTube analysis can be done by other ide too, So, I choose to do it on a documented type format ide named jupyter. Which come separately as well as with the famous package of anaconda.

For Analysis, we require:

- a. API key for scrapping the data from YouTube or you can use dataset for analysis if you like which is available on Kaggle.
- b. Needed an IDE such as Jupyter.
- c. Seaborn for Visualization
- d. A browser for Jupyter, I'm using Microsoft Edge for this.
- e. Pandas for data Analysis.
- f. To Run the Jupyter:
 - i. Ram: 512MB
 - ii. Space: 2GB
 - iii. Core: Dual Core
 - iv. Cache: 512kb

3. Function Requirements

- Required API key to request the data.
- Or Proper formatted csv file dataset.
- A device with the jupyter ide.
- Pandas installed.
- Seaborn installed.

4. Time Taken:

1.	Understanding and Using API for Scrapping the data	2 Days
2.	Design Phase	5 Days
3.	Data Analysis of Few Channels	7 Days
4.	Extracting the insights from a particular channel	6 Days

5. Non-Functional Requirements

5.1. Performance Requirements:

- The user must have a device which is at least have the specification and space to store and run dataset on Jupyter.
- The user must have at least of 500kbps of net connection in order to extract the data from the YouTube without much delay.

5.2 Security Requirements:

- API key should not be shared to anyone.
- Or Dataset should not be shared with anyone.

5.3 Analysis Quality Attributes:

- AVAILABILITY: The API key, a device with working jupyter and internet connection with 512kbps.
- MAINTAINABILITY: Keep updating the data requested from the YouTube.
- USABILITY: The analysis can be done anywhere if the connectivity to the internet and a device.

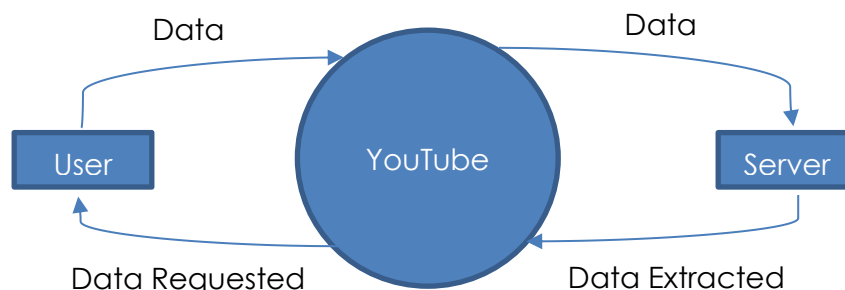
6. Design Phase:

Design phase deals with transforming there requirements, as described in the SRS document into a form that can be used while programming. In design phase of SDLC based on requirement captured in SRS.

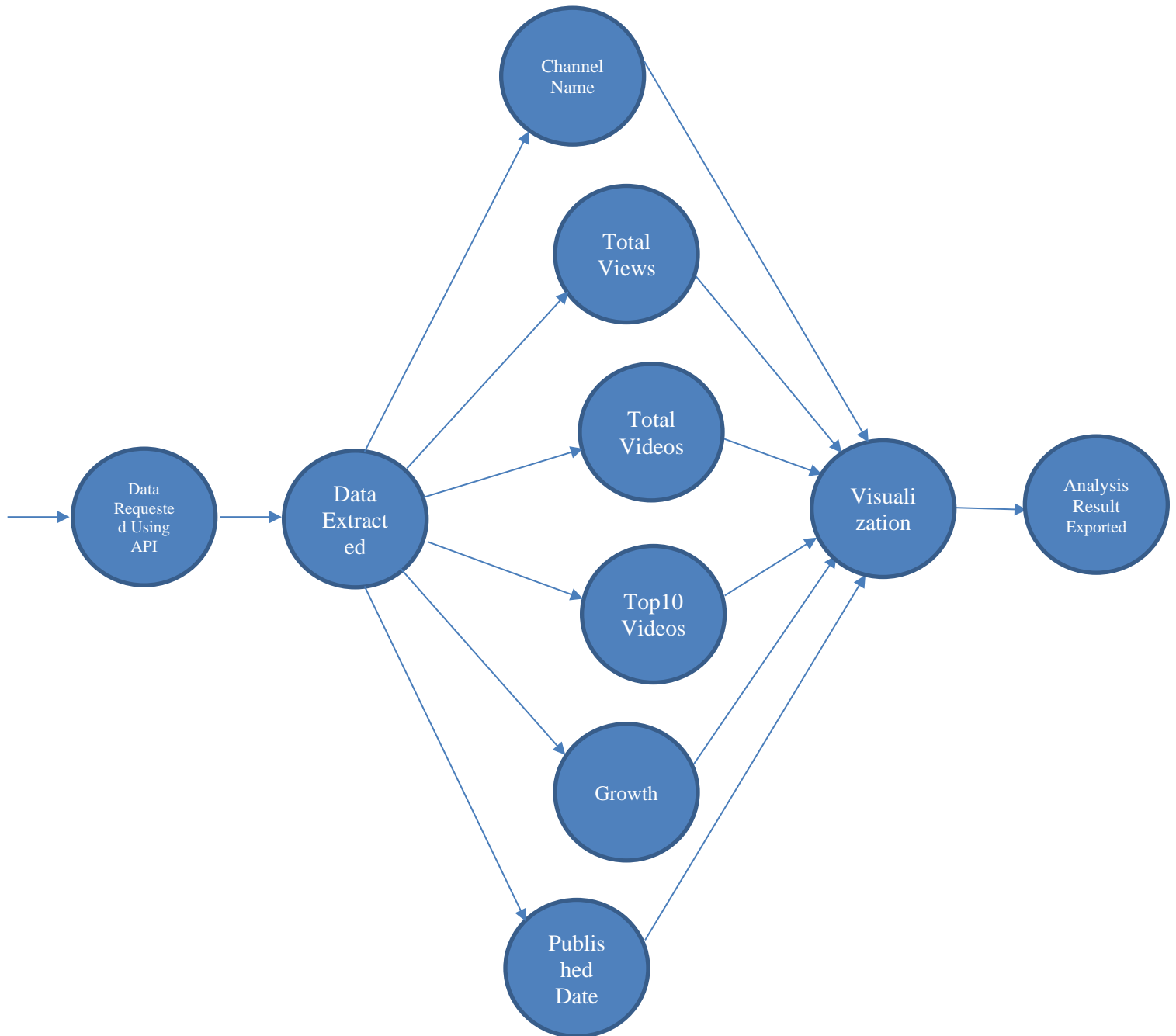
6.1 DFD (DataFlow Diagram)

A data flow diagram (DFD) maps out the flow of information for any processor system.

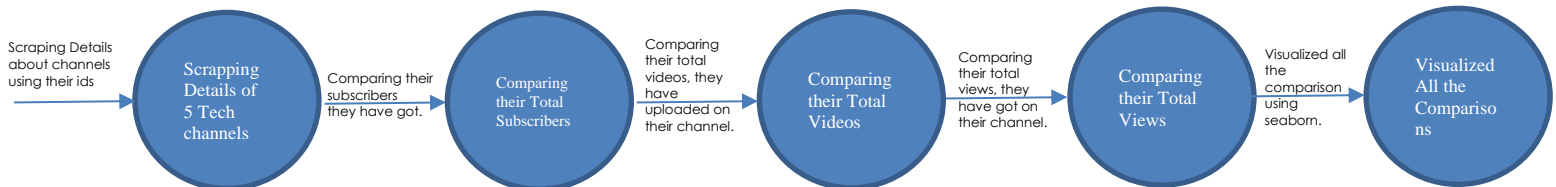
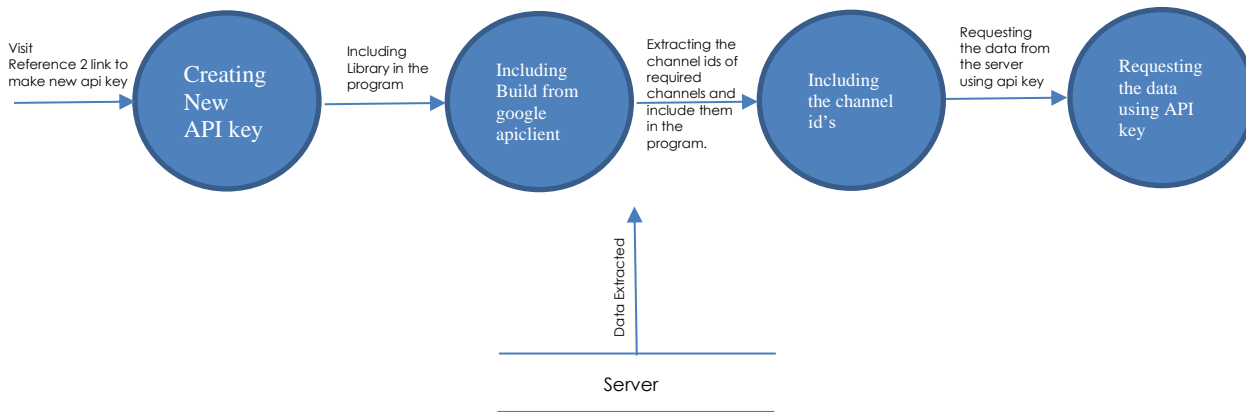
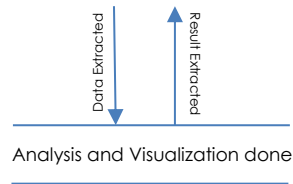
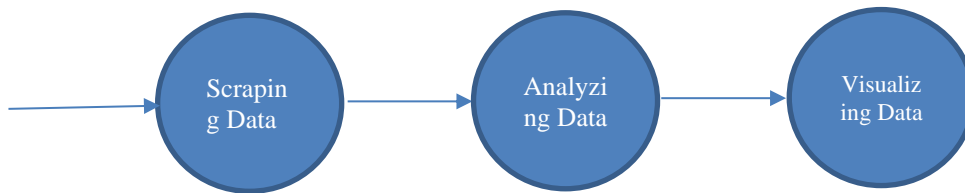
Level 0



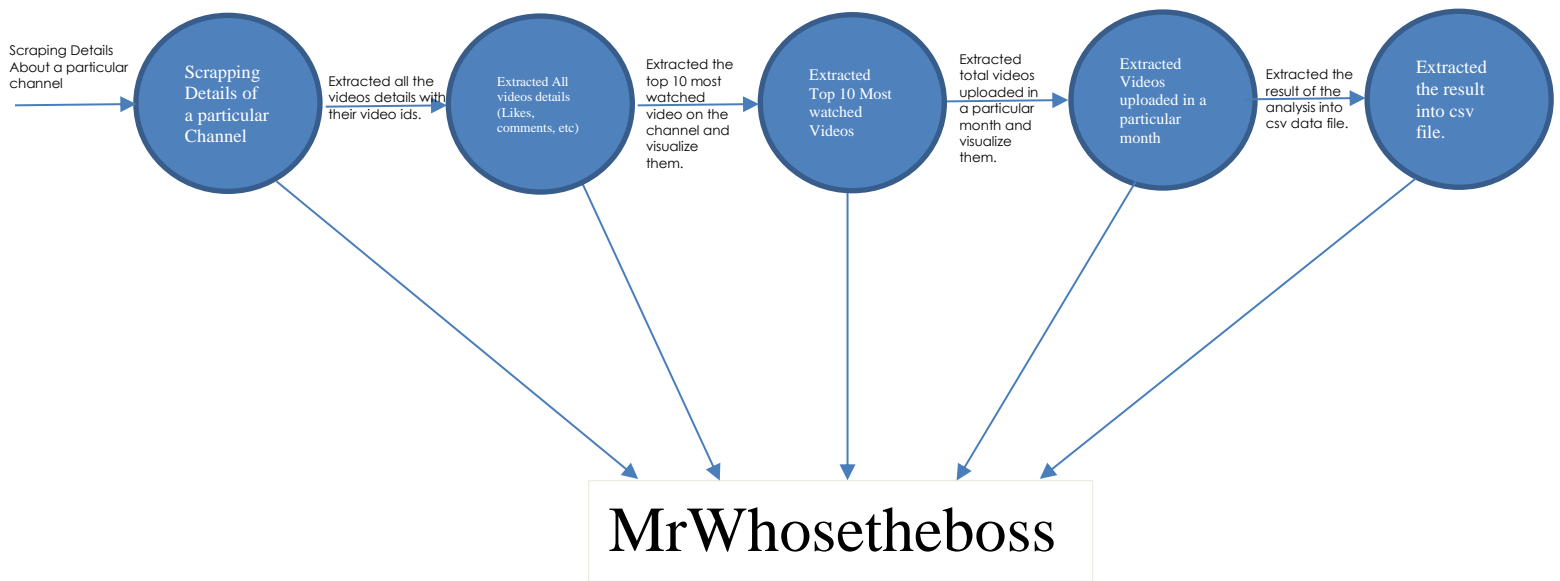
Level 1



Level 2

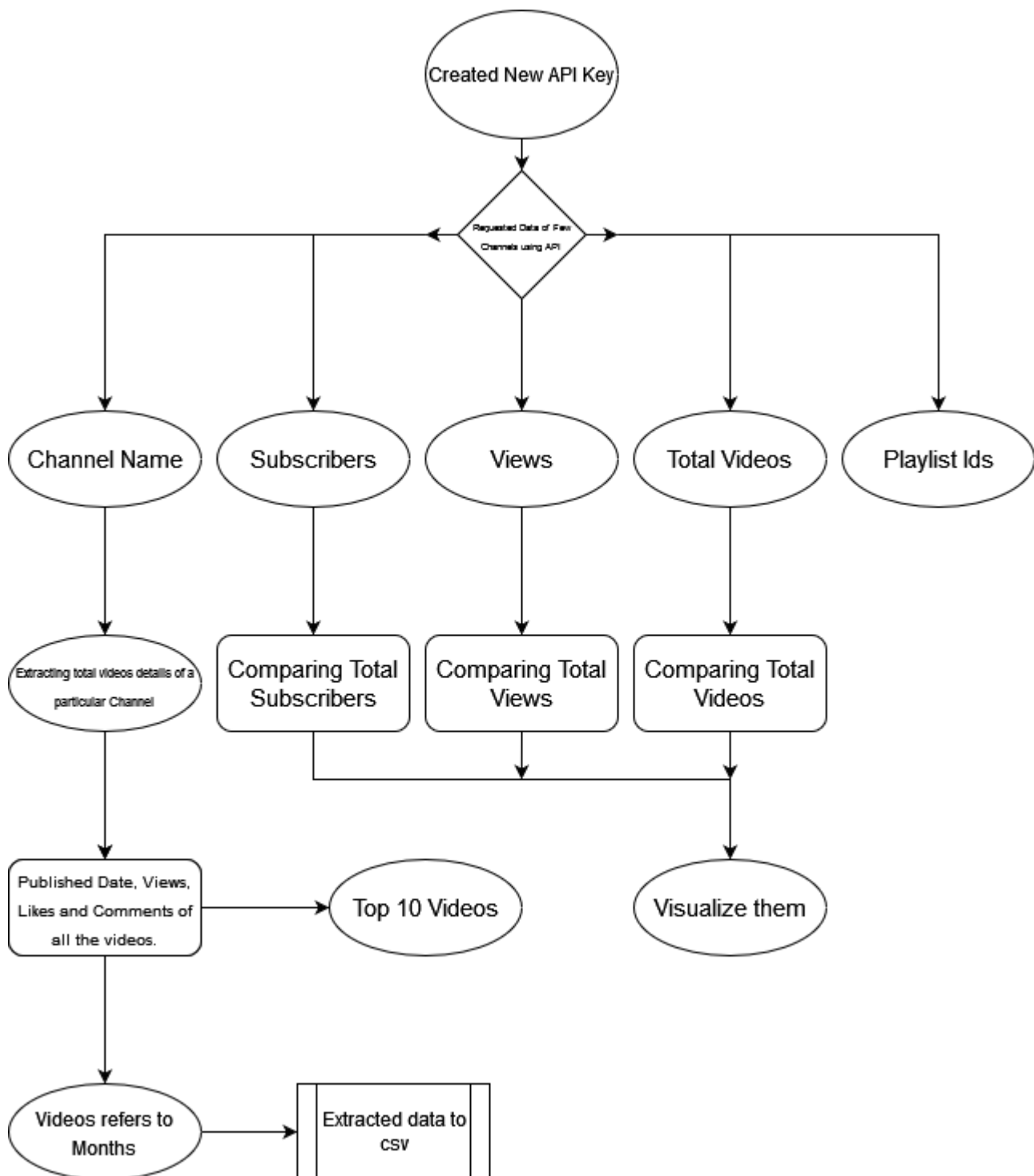


Analysis on a Particular Channel:



6.2 Flow Chart:

A graphical representation of a computer program in relation to **the** sequence of **its** functions (as **opposed to** the data it processes).



7. References:

- <https://developers.google.com/youtube/v3/docs>
- <https://console.cloud.google.com/apis/dashboard?project=yt-analysis-project-364212>
- <https://commentpicker.com/youtube-channel-id.php>
- <https://jsonformatter.curiousconcept.com/#>
- <https://www.kaggle.com/>
- <https://www.google.com/>