



# TrendWave Analytics

"Unlocking the Power of Data, One Video at a Time."

Data Engineering  
Group 4

# Objective:

TrendWave Analytics is Developing a comprehensive data analytics pipeline to analyze and provide insights into Youtube videos, helping business and content creators understand popular video types, key characteristics, and success factors.





# Table of contents

01

Project  
overview

02

Data  
Injection

03

ETL Design

04

Visualization

05

Findings



# Trending YouTube Video Statistics

Data Card

Code (2809)

Discussion (43)

▲ 5225

New Notebook

Download (211 MB)



CA\_category\_id.json (7.91 kB)



```
▼ "root" : { 3 items
  "kind" : string "youtube#videoCategoryListResponse"
  "etag" : string ""ld9biNPKjAjjgV7EZ4EKeEGrhao/1v2mrzYSYG6onNLt2qTj13hkQZk""
  ▼ "items" : [ 31 items
    ▶ 0 : {...} 4 items
    ▶ 1 : {...} 4 items
    ▶ 2 : {...} 4 items
    ▶ 3 : {...} 4 items
    ▶ 4 : {...} 4 items
    ▶ 5 : {...} 4 items
    ▶ 6 : {...} 4 items
    ▶ 7 : {...} 4 items
    ▶ 8 : {...} 4 items
    ▶ 9 : {...} 4 items
    ▶ 10 : {...} 4 items
    ▶ 11 : {...} 4 items
```

Our dataset from  
Youtube

Source: - <https://www.kaggle.com/datasnaek/youtube-new/data>

## Data Explorer

Version 115 (539.22 MB)

{i} CA\_category\_id.json

CAvideos.csv

{i} DE\_category\_id.json

DEvideos.csv

{i} FR\_category\_id.json

FRvideos.csv

{i} GB\_category\_id.json

GBvideos.csv

{i} IN\_category\_id.json

INvideos.csv

{i} JP\_category\_id.json

JPvideos.csv

RU\_category\_id.json

RUvideos.csv

MX\_category\_id.json

MXvideos.csv

{i} RU\_category\_id.json

RUvideos.csv

# CSV's Features



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	thumbnail_url	comments_disabled	ratings_disabled	video_error_reason	description						
2kyS6SVSYSE	17.14.11	WE WANT T	CaseyNeistat	22	2017-11-13T	SHANTell ma	748374	57527	2966	15954	https://i.yt	FALSE	FALSE	FALSE	SHANTELL'S CHANNEL - <a href="https://www.youtube.com/shantellmartin">https://www.youtube.com/shantellmartin</a> \nCANDICE - <a href="https://www.youtube.com/candice">https://www.youtube.com/candice</a>						
1ZAPwfrtAF	17.14.11	The Trump P	LastWeekTo	24	2017-11-13T	last week to	2418783	97185	6146	12703	https://i.yt	FALSE	FALSE	FALSE	One year after the presidential election, John Oliver discusses what we've learned so far.						
5qpjK5DgCt4	17.14.11	Racist Super	Rudy Mancus	23	2017-11-12T	racist supern	3191434	146033	5339	8181	https://i.yt	FALSE	FALSE	FALSE	WATCH MY PREVIOUS VIDEO ,ñð \n\nSUBSCRIBE ,ñf <a href="https://www.youtube.com/channel/UC5qpjK5DgCt4">https://www.youtube.com/channel/UC5qpjK5DgCt4</a>						
puqaWwrc7t	17.14.11	Nickelback L	Good Mythic	24	2017-11-13T	rhett and lin	343168	10172	666	2146	https://i.yt	FALSE	FALSE	FALSE	Today we find out if Link is a Nickelback amateur or a secret Nickelback devotee. GMMB						
d380meDOW	17.14.11	I Dare You: G	nigahiga	24	2017-11-12T	ryan  "higa"	2095731	132235	1989	17518	https://i.yt	FALSE	FALSE	FALSE	I know it's been a while since we did this show, but we're back with what might be the						
gHZ1QzOKIK	17.14.11	2 Weeks wit	iJustine	28	2017-11-13T	ijustine  "we	119180	9763	511	1434	https://i.yt	FALSE	FALSE	FALSE	Using the iPhone for the past two weeks -- here's my thoughts!\nAll my iPhone X Videos						
39idVpFF7N	17.14.11	Roy Moore &	Saturday Nig	24	2017-11-12T	SNL  "Saturd	2103417	15993	2445	1970	https://i.yt	FALSE	FALSE	FALSE	Embattled Alabama Senate candidate Roy Moore (Mikey Day) meets with Vice President						
nc99ccSXSTO	17.14.11	5 Ice Cream	CrazyRussiar	28	2017-11-12T	5 Ice Cream	817732	23663	778	3432	https://i.yt	FALSE	FALSE	FALSE	Ice Cream Pint Combination Lock - <a href="http://amzn.to/2ACipdl">http://amzn.to/2ACipdl</a> \nMini Ice Cream Sandwich						
jr9QtXwC9vc	17.14.11	The Greatest	20th Century	1	2017-11-13T	Trailer  "Hug	826059	3543	119	340	https://i.yt	FALSE	FALSE	FALSE	Inspired by the imagination of P.T. Barnum, The Greatest Showman is an original musical						
TUmyygCMN	17.14.11	Why the rise	Vox	25	2017-11-13T	vox.com  "vo	256426	12654	1363	2368	https://i.yt	FALSE	FALSE	FALSE	For now, at least, we have better things to worry about.\n\n\nSubscribe to our channel						
9wRQJfJFNDV	17.14.11	Dion Lewis'	NFL	17	2017-11-13T	NFL  "Footba	81377	655	25	177	https://i.yt	FALSE	FALSE	FALSE	New England Patriots returner Dion Lewis blasts off for an amazing kickoff return tou						
VifQJit6AO	17.14.11	(SPOILERS) "	amc	24	2017-11-13T	The Walking	104578	1576	303	1279	https://i.yt	FALSE	FALSE	FALSE	Shiva arrives just in time as King Ezekiel attempts to take out an army of walkers.\n\n\n						
5E4ZBSInqU	17.14.11	Marshmello	marshmello	10	2017-11-13T	marshmello	687582	114188	1333	8371	https://i.yt	FALSE	FALSE	FALSE	WATCH SILENCE MUSIC VIDEO ,ñð <a href="https://youtu.be/Tx1sqYc3qas">https://youtu.be/Tx1sqYc3qas</a> \nWATCH YOU & ME						
GgVmn66K	17.14.11	Which Count	NowThis Wo	25	2017-11-12T	nowthis  "no	544770	7848	1171	3981	https://i.yt	FALSE	FALSE	FALSE	The world at large is improving, but some countries and their governments are strugg						
TaTleo4Os8	17.14.11	SHOPPING F	The king of C	15	2017-11-12T	shopping for	207532	7473	246	2120	https://i.yt	FALSE	FALSE	FALSE	Today we go shopping for new fish for some of the aquariums. I wanted to find some						
kgao455yaO	17.14.11	The New Spr	BostonDyna	28	2017-11-13T	Robots  "Bos	75752	9419	52	1230	https://i.yt	FALSE	FALSE	FALSE	For more information . . . stay tuned.						
ZAQs-ctOqX	17.14.11	One Change	Cracked	23	2017-11-12T	pacific rim  "	295639	8011	638	1256	https://i.yt	FALSE	FALSE	FALSE	Pacific Rim was so good, we can,Âot believe they didn,Âot do this one thing to make						
YVfyYrEmzg	17.14.11	How does yo	TED-Ed	27	2017-11-13T	TED  "TED-Ed	78044	5398	53	385	https://i.yt	FALSE	FALSE	FALSE	Check out our Patreon page: <a href="https://www.patreon.com/teded">https://www.patreon.com/teded</a> \n\nView full lesson: <a href="https://www.youtube.com/watch?v=YVfyYrEmzg">https://www.youtube.com/watch?v=YVfyYrEmzg</a>						
NGN5m21k	17.14.11	Home Made	Rate of Fire	28	2017-11-13T	ultimately b	87007	11063	26	2211	https://i.yt	FALSE	FALSE	FALSE	reopened now to fly out of ground effect! The homemade airplane does indeed fly! T						

# Json's Feature

CA\_category\_id.json (7.91 kB)



```
▼ "root" : { 3 items
  "kind" : string "youtube#videoCategoryListResponse"
  "etag" : string ""ld9biNPKjAjjV7EZ4EKeEGrhao/1v2mrzYSYG6onNLt2qTj13hkQZk""
  ▼ "items" : [ 31 items
    ▼ 0 : { 4 items
      "kind" : string "youtube#videoCategory"
      "etag" : string ""ld9biNPKjAjjV7EZ4EKeEGrhao/Xy1mB4_yLrHy_BmKmpBggy2mZQ""
      "id" : string "1"
      ▼ "snippet" : { 3 items
        "channelId" : string "UCBR8-60-B28hp2BmDPdntcQ"
        "title" : string "Film & Animation"
        "assignable" : bool true
      }
    }
    ▼ 1 : { 4 items
      "kind" : string "youtube#videoCategory"
      "etag" : string ""ld9biNPKjAjjV7EZ4EKeEGrhao/UZ1oLIIZ2dxIh045ZTFR3a3NyTA""
      "id" : string "2"
      ▶ "snippet" : { ... } 3 items
    }
    ▼ 2 : { 4 items
```

## Create IAM Role

- Ensure Access for Everyone
- Establish an IAM (Identity and Access Management) role to provide secure and controlled access to AWS resources.
- Regulate who can perform actions on your AWS resources.
- Protect sensitive data and resources.
- Define role permissions based on user requirements.

## Store CSV File to S3 Bucket

- Data Storage for Merging
- Imported the CSV file to an S3 (Simple Storage Service) bucket to facilitate data merging and storage.
- S3 can handle large volumes of data.
- Data is highly available and durable.
- Use AWS Management Console or SDKs to upload the file.

## Store and Clean JSON File

- Properly Format JSON Data
- Store the JSON file and apply data cleaning processes to ensure it's in the correct format for analysis.
- Ensures accurate and reliable data.
- Prepared data is ready for analysis and visualization.
- Use tools like Python and AWS services like Lambda for cleaning and transformation.



## Json error



⊗ Failed

Time in queue: 0.171 sec

Run time: 0.308 sec

Data scanned: -

⊗ Row is not a valid JSON Object - JSONException: A JSONObject text must end with '}' at 2 [character 3 line 1]

This query ran against the "db\_youtube\_cleaned" database, unless qualified by the query. Please post the error message on our [forum](#) or contact [customer support](#) with Query Id: 3d57071e-501c-4b45-ab02-4bc9c5294dba

## CSV error

⊗ Failed

Time in queue: 0.124 sec

Run time: 0.391 sec

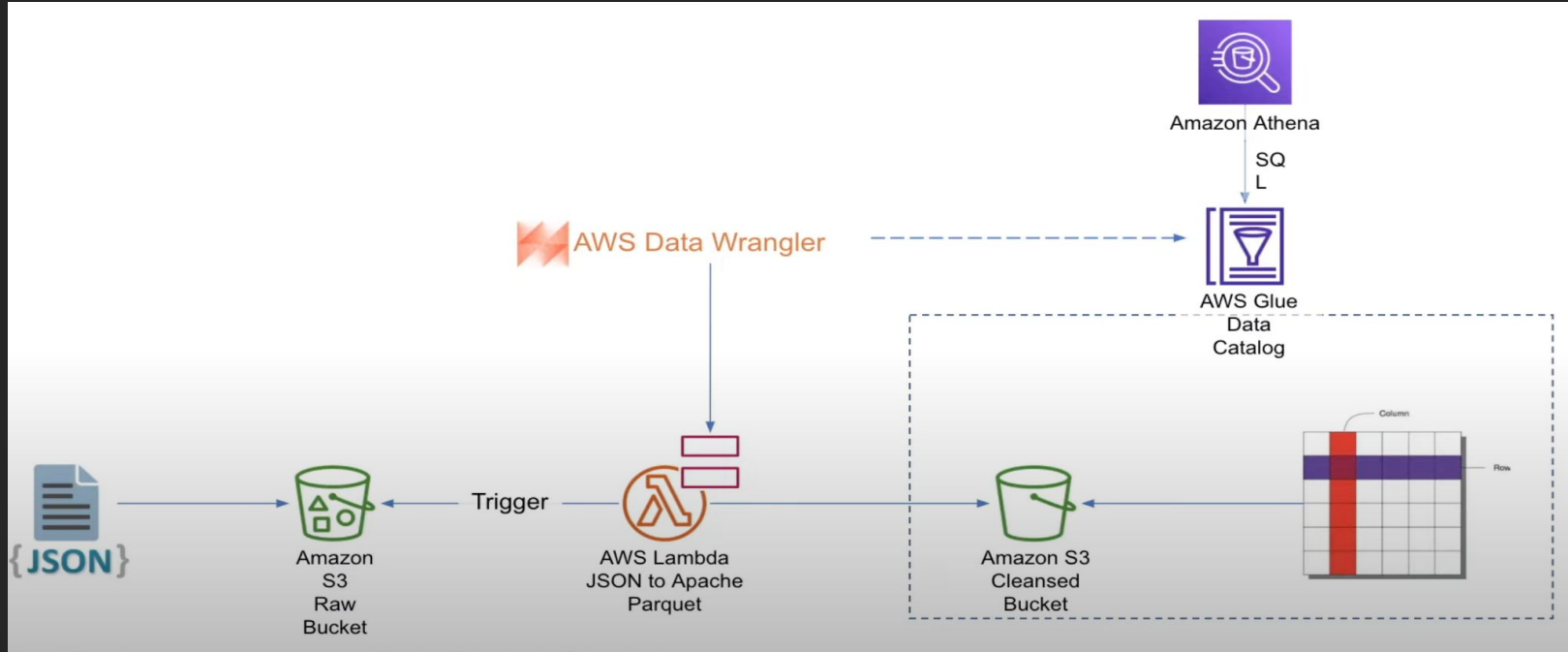
Data scanned: -

⊗ SYNTAX\_ERROR: line 2:87: '=' cannot be applied to bigint, varchar

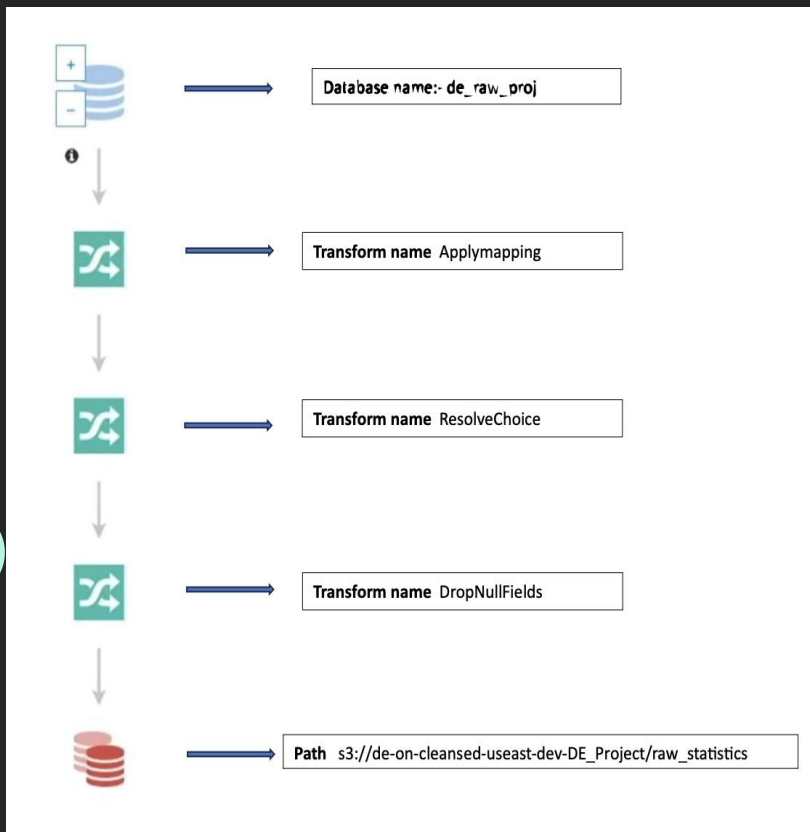
This query ran against the "db\_youtube\_cleaned" database, unless qualified by the query. Please post the error message on our [forum](#) or contact [customer support](#) with Query Id: 56d2bc2f-058f-481d-872a-7de927e5e35b



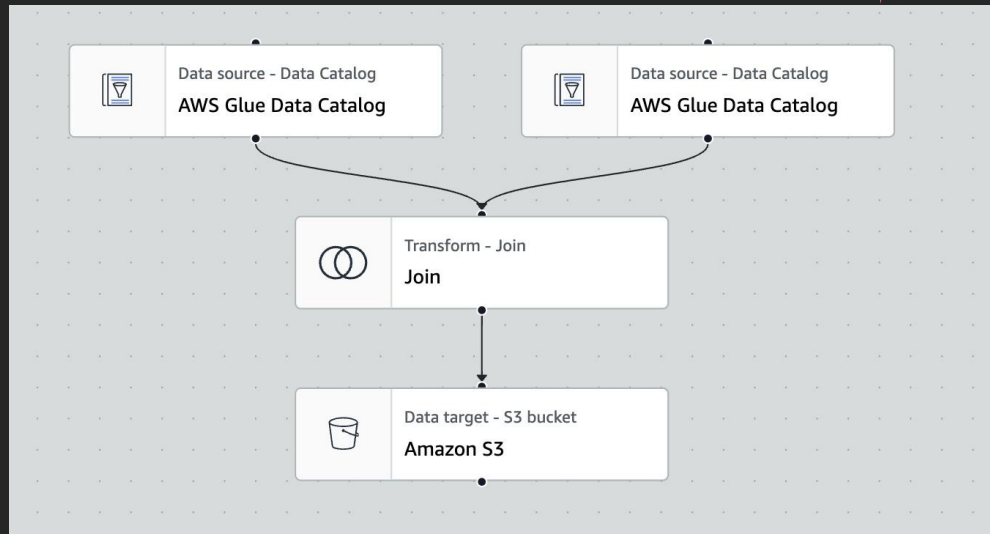
# Cleaning JSon Data



## ETL for CSV cleaning



## ETL for combining csv and Json and final parquet



# Final Parquet file



Amazon Athena > Query editor

Editor | Recent queries | Saved queries | Settings

Workgroup: primary

**Data**

Data Source: AwsDataCatalog

Database: db\_youtube\_analytics

Tables and views: Create

Filter tables and views

Tables (1): final\_analytics

Views (0)

Query 5 × | Query 6 × | Query 7 × | Query 8 × | Query 9 ×

```
1 SELECT * FROM "db_youtube_analytics"."final_analytics" limit 10;
```

SQL Ln 1, Col 1

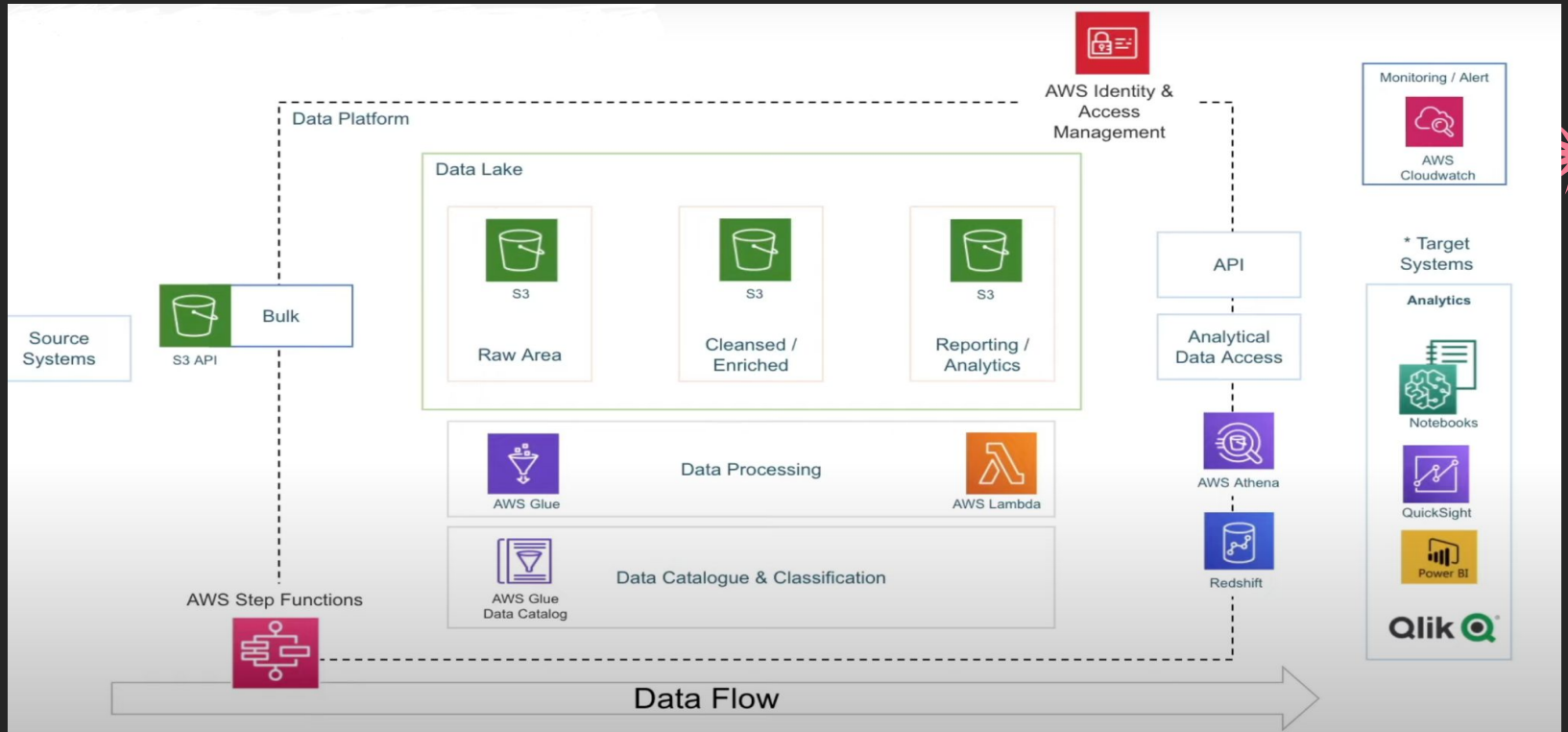
Run again | Cancel | Save | Clear | Create

Completed Time in queue: 0.113 sec Run time: 0.835 sec Data scanned: 248.79 KB

Results (10) Copy Download results

Search rows

#	ratings_disabled	snippet_title	comments_disabled	trending_date	etag	video_id	thumbnail
1	false	Comedy	false	18.14.01	"XI7nbFXulYBipL0ayR_gDh3eu1k/G9LHzQmx44rX25Syaga_Aqtwz8M"	6OqZ6XCEhY	https://i
2	false	Comedy	false	18.14.01	"Id9bINPKjAigIv7EZ4EKeEGrhao/G9LHzQmx44rX25Syaga_Aqtwz8M"	6OqZ6XCEhY	https://i
3	false	Comedy	false	18.14.01	"Id9bINPKjAigIv7EZ4EKeEGrhao/G9LHzQmx44rX25Syaga_Aqtwz8M"	6OqZ6XCEhY	https://i



# Data Visualization



Trends and Patterns  
Dashboard View  
Our Findings

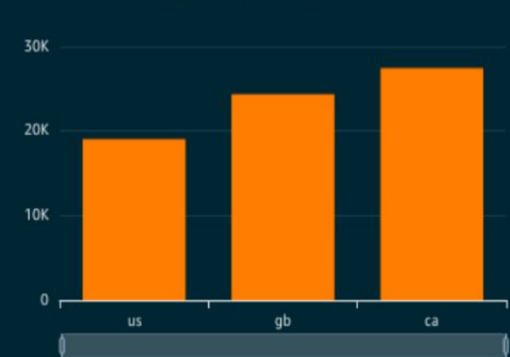


QuickSight

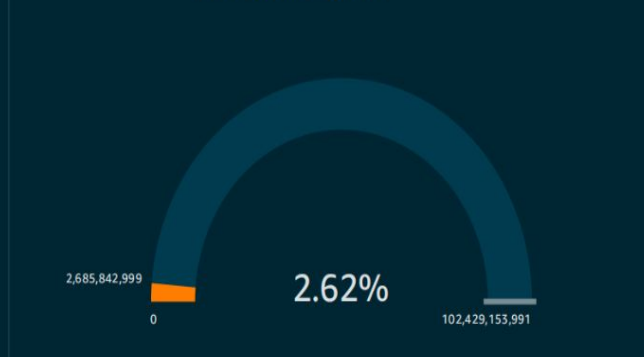


Region	Category	Avg Likes	Avg Dislikes	Avg Views
All	All	55.02K	629.76	1.12M

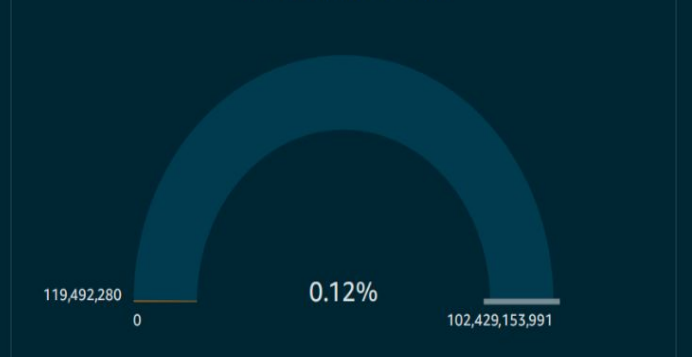
Videos Upload by Country



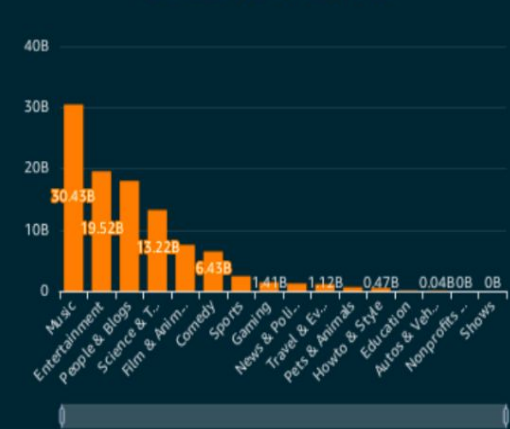
Sum of Likes & Views



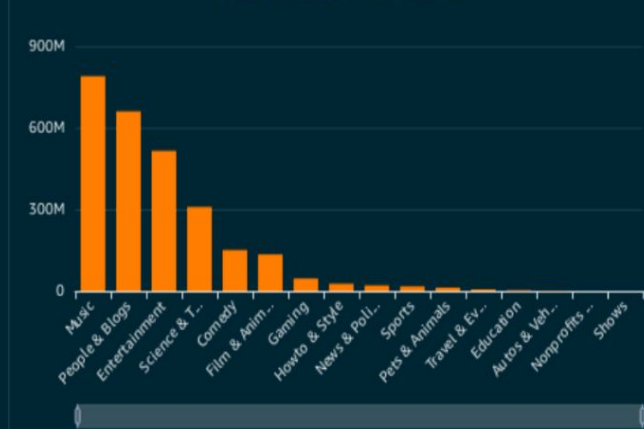
Sum of Dislikes & Views



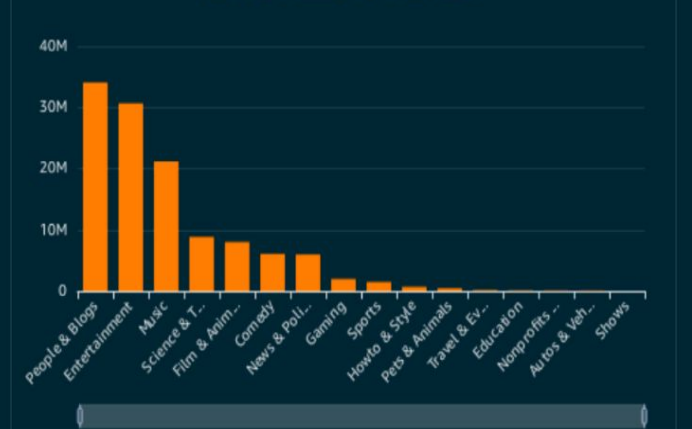
Views Analysis By Categories



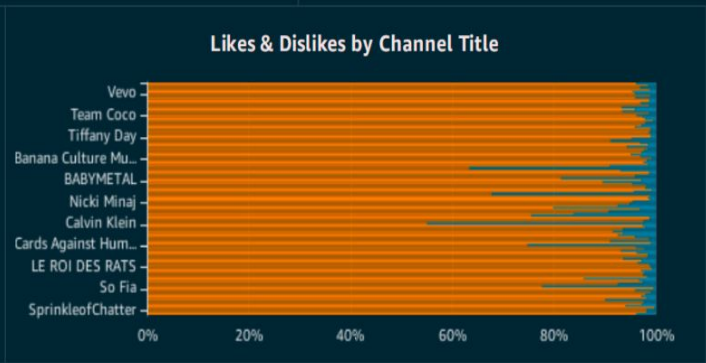
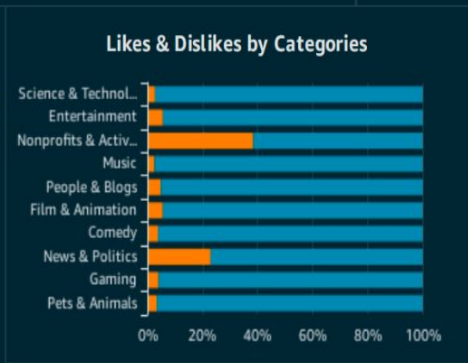
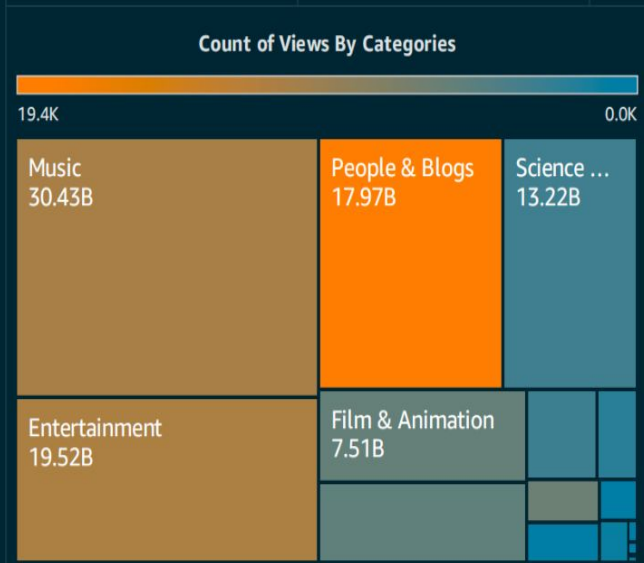
Sum Of Likes By Categories



Sum Of Dislikes By Categories

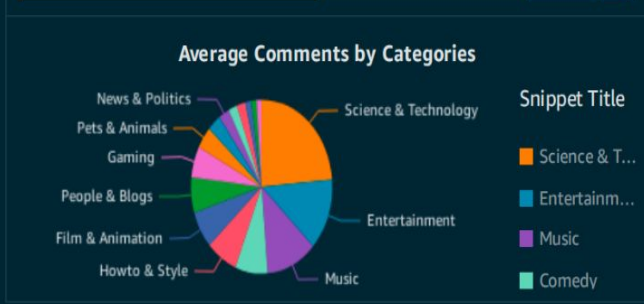


Category	Region	Avg Likes	Avg Dislikes	Avg Views
All	All	55.02K	629.76	1.12M



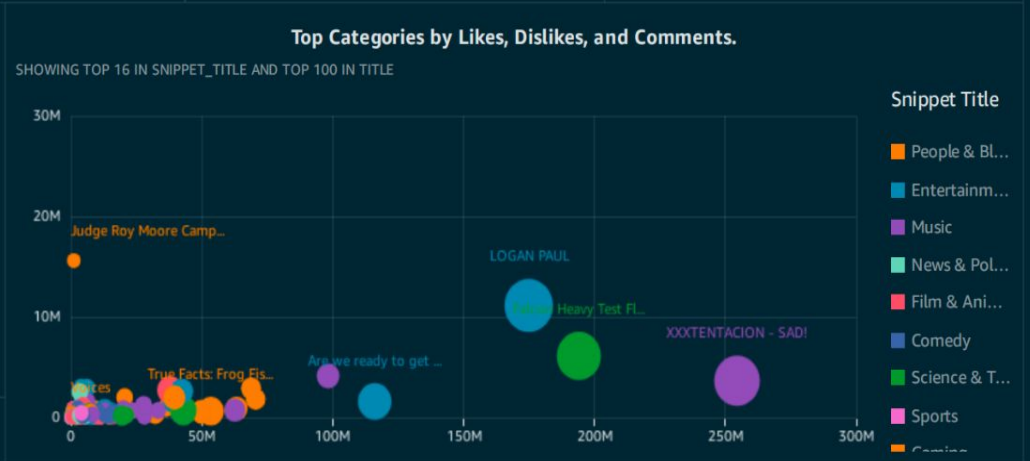
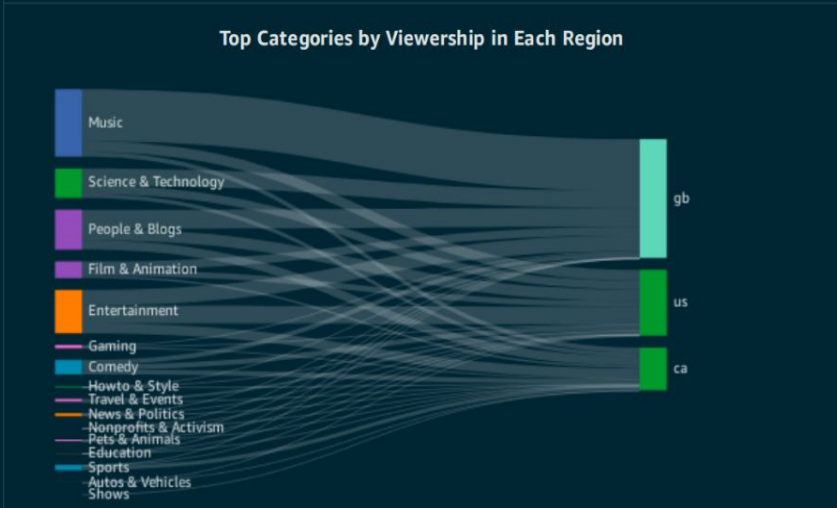
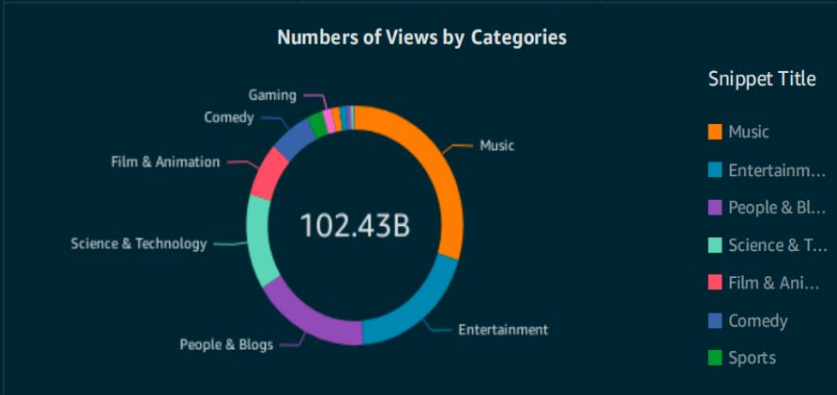
### Trend Report

snippet_title	channel_title	publish_time	trending_date	likes	views	dislikes
Comedy	Best of Simpsons Characters	03-21-2018	Jan 31, 2018	4,261	239,376	1,560
Comedy	Mary Giannone	03-27-2018	Jan 31, 2018	3,977	553,637	2,940
Comedy	PewDiePie	03-27-2018	Jan 31, 2018	373,369	5,114,697	96,470
Comedy	Team Coco	01-23-2018	Jan 31, 2018	11,373	517,174	16,110
Comedy	Team Coco	01-28-2018	Jan 31, 2018	18,320	701,411	11,010
Comedy	Team Coco	03-28-2018	Jan 31, 2018	17,708	1,088,804	2,180
Comedy	Team Coco	05-10-2018	Jan 31, 2018	62,021	2,130,991	9,520
Comedy	blazze999	03-26-2018	Jan 31, 2018	898	214,731	870
				37,923.85	1,446,290.05	119,49...





region equals	snippet_title equals	Avg Views	Avg Dislikes	Avg Likes
All	All	1.12M	629.76	55.02K



# Findings



- ❑ Canada holds the record for the highest number of videos uploaded on YouTube, accounting for approximately 30% of the total videos uploaded.
- ❑ The Music category has the highest number of views on YouTube, totaling 30.43 billion views.
- ❑ “SAD!” is most trending video by XXXTENTACION has the highest like(254.48M), dislikes (3.64M), and comment (20.16M).
- ❑ PewDiePie channel is most popular channel.



# Use Cases

- ❑ **Content Creators' Content Strategy Optimization**
- ❑ **Marketers' Data-Driven Advertising**
- ❑ **Media Companies' Enhanced Audience Engagement**
- ❑ **Data Analysts and Researchers' In-Depth Exploration**

## *Our Team*



Harsh Shah  
Data Engineer



Rutvik G  
Data Scientist



Jay  
Data Analyst



Vrushabh  
Data Engineer



Jenith  
Product Manager  
and Data Analyst

*Thank you!!!!*

