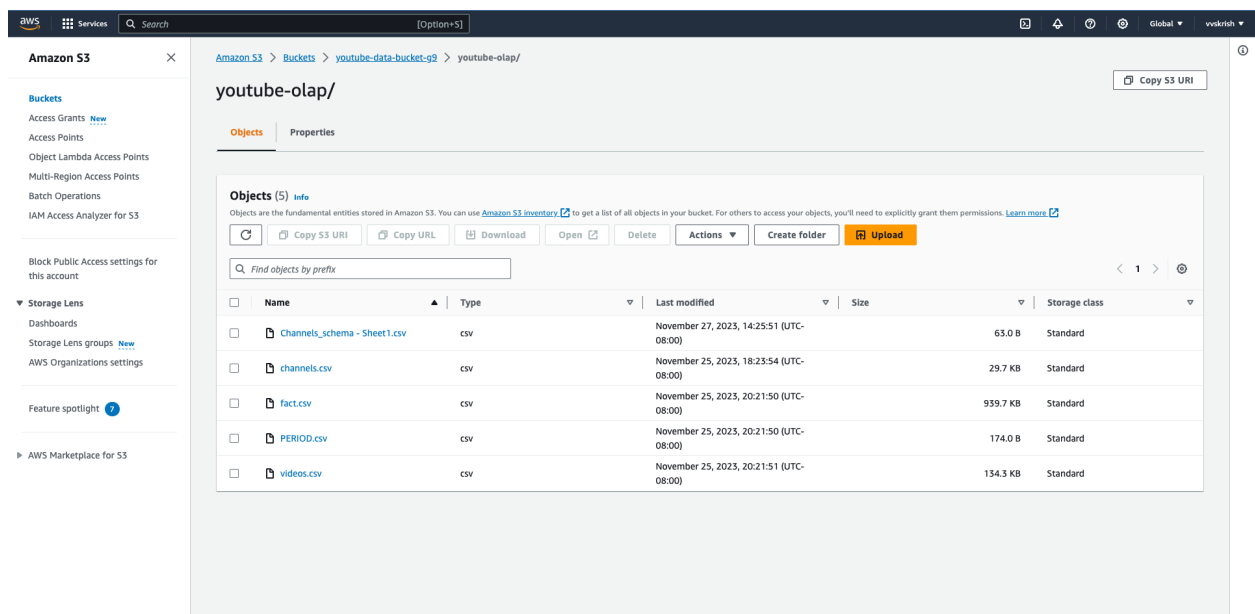


ETL

Amazon S3 Bucket: Link - <https://youtube-data-bucket-g9.s3.amazonaws.com/youtube-olap/>
An S3 bucket in Amazon S3 is a storage container used to hold various types of data, including CSV files. These CSV files are raw data that typically undergo Extract, Transform, and Load (ETL) operations within AWS Glue or similar services.

During ETL, the CSV files are extracted from the S3 bucket, transformed into a structured format suitable for analysis or other purposes, and then loaded back into another location, often another S3 bucket or a database, ready for use. This process helps in organizing, cleaning, and preparing the data for analytics, reporting, or further processing.



IAM Roles:

In Amazon Web Services (AWS), Identity and Access Management (IAM) roles serve as fundamental tools for securely delegating permissions within AWS resources. These roles define a set of permissions that dictate actions authorized for entities, be it an AWS service or a user, on AWS resources. They eliminate the necessity for long-term credentials like usernames and passwords when granting access to resources.

The IAM role in the image below is configured with specific permissions:

- AmazonS3FullAccess: Grants complete access to Amazon S3 (Simple Storage Service), enabling any action on S3 buckets within the AWS account.
- AmazonRedshiftFullAccess: Provides full access to Amazon Redshift, empowering the IAM role to manage Redshift clusters comprehensively. This includes tasks such as creating, modifying, deleting clusters, and handling administrative functions.

- **AWSGlueServiceRole**: Grants the essential permissions for executing actions within AWS Glue, a fully managed ETL service facilitating data preparation and loading for analytics.

By consolidating these permissions within a single IAM role, entities assuming this role gain the combined access and capabilities across Amazon S3, Amazon Redshift, and AWS Glue. This practice aligns with security best practices, adhering to the principle of least privilege by granting only the necessary permissions for specific tasks.

IAM roles offer remarkable flexibility, allowing different entities such as AWS services, applications, or other AWS accounts to assume these roles based on predefined trust policies. Assigning roles to entities enables centralized permission management, upholding security measures, and reducing reliance on persistent credentials, thereby enhancing the overall security posture of AWS resources.

The screenshot displays the AWS IAM console interface for the 'youtube-glue-role'. The left sidebar shows the 'Identity and Access Management (IAM)' navigation menu. The main content area shows the role's summary, including its creation date, ARN, and last activity. Below the summary, the 'Permissions policies' section lists three attached policies: AmazonRedshiftFullAccess, AmazonS3FullAccess, and AWSGlueServiceRole. The 'Permissions boundary' is currently not set.

Policy name	Type	Attached entities
AmazonRedshiftFullAccess	AWS managed	2
AmazonS3FullAccess	AWS managed	1
AWSGlueServiceRole	AWS managed	2

Red Shift Cluster:

Amazon Redshift is a data warehouse service in AWS used for analyzing large datasets. A Redshift cluster is a collection of nodes that work together to handle queries and data storage. After the ETL process, the transformed data from CSV files is often loaded into a Redshift cluster, typically organized into tables.

These tables store structured data and are designed to support efficient querying and analysis. In your case, there are four tables within the Redshift cluster, each containing specific datasets that have undergone transformation. These tables serve as organized repositories of data, allowing users to run complex queries and analytics to derive insights.

Details of Redshift cluster:

Connection details of Redshift Cluster:

The screenshot displays the AWS Management Console interface for an Amazon Redshift cluster. The left sidebar shows navigation options like 'Serverless dashboard', 'Query editor v2', and 'Monitoring'. The main content area is titled 'default-workgroup' and contains two sections: 'General information' and 'Network and security'.

General information

Property	Value
Workgroup	default-workgroup
Namespace	youtube-data-cluster
Workgroup ARN	arn:aws:redshift-serverless:us-east-1:248592503212:workgroup/68d4ac14-e85b-4735-ab9a-584b4e0aa46c
Date created	November 25, 2023, 17:41 (UTC-08:00)
Status	Available
Configuration	Production
Custom domain name	-
Endpoint	default-workgroup.248592503212.us-east-1.redshift-serverless.amazonaws.com:5439/dev
JDBC URL	jdbc:redshift://default-workgroup.248592503212.us-east-1.redshift-serverless.amazonaws.com:5439/dev
ODBC URL	Driver={Amazon Redshift (x64)}; Server=default-workgroup.248592503212.us-east-1.redshift-serverless.amazonaws.com; Database=dev

Network and security

Property	Value
Virtual private cloud (VPC)	vpc-066a2229675bf0f6d
VPC endpoint ID	vpce-041651bd991c0e36e
VPC security group	sg-0879d59bf1cdf8414
Subnet	subnet-074f8843443a4f66b, subnet-017b0bf37bd184111, subnet-0082682e75fa542e7, subnet-08fd993221d829a2, subnet-0761055f8257639b3, subnet-03c0a907210e05f7
Enhanced VPC routing	Off
Publicly accessible	On

Crawlers:

In AWS Glue, crawlers serve as automated processes used for discovering and cataloging metadata from diverse data sources. These processes analyze data structure and schema, facilitating streamlined data processing within the ETL (Extract, Transform, Load) workflow.

The S3 crawler within AWS Glue inspects data stored in Amazon S3 buckets. It identifies file formats and schema, organizing the data for subsequent use in the ETL process. By scanning S3 data, this crawler comprehends its structure, aiding in the preparation of data for transformation.

Similarly, the Redshift crawler in AWS Glue targets Redshift clusters. It assesses the data within Redshift tables, capturing their structure and schema. This acquired information becomes crucial for mapping and efficiently transforming data during ETL operations. It ensures consistency and accuracy in data processing within the Redshift environment.

Both crawlers assume a pivotal role in the ETL process by automatically discovering and organizing metadata. Their function streamlines the transformation and loading of data from its raw state into structured datasets, primed for analysis or other intended uses.

Redshift Crawler:

The screenshot displays the AWS Glue console interface for a crawler named 'yt-redshift-crawler'. The left sidebar shows the navigation menu with 'Data Catalog' and 'Crawlers' highlighted. The main content area shows the crawler's properties and a table of its runs.

Crawler properties

Name	IAM role	Database	State
yt-redshift-crawler	youtube-glue-role	yt-s3-db	READY

Crawler runs (2)

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
November 27, 2023 at 21:58:31	November 27, 2023 at 21:41:05	02 min 33 s	Completed	0.300	4 table changes, 0 partition changes
November 27, 2023 at 21:02:06	November 27, 2023 at 21:05:51	03 min 44 s	Completed	0.301	4 table changes, 0 partition changes

JDBC Connection to Temporary Database:

In AWS Glue, a JDBC connection establishes access to a temporary database, essential for the ETL process. JDBC, as a Java Database Connectivity tool, enables data interaction and transfer between the Glue environment and external databases. This connectivity supports AWS Glue in extracting, transforming, and loading data across diverse databases, ensuring smooth data processing throughout the ETL operations.

The screenshot displays the AWS Glue console interface for a JDBC connection. The left sidebar shows the navigation menu with 'Data Catalog' and 'Connections' highlighted. The main content area shows the connection details and a table of jobs.

Jdbc connection

Connection details

Connector type	Connection URL
JDBC	jdbc:redshift://default-workgroup.248592503212.us-east-1.redshift-serverless.amazonaws.com:5439/youtube

Driver class name

Driver path

Username

Require SSL connection

Subnet

Security groups

Description

Created on

Last modified

Class name

Your jobs (7)

Job name	Type	Last modified	AWS Glue version
yt-eti-chnl-job	Glue ETL	11/28/2023, 7:50:50 PM	4.0
yt-eti-videos-job	Glue ETL	11/28/2023, 7:26:40 PM	4.0
yt-eti-videos-job	Glue ETL	11/28/2023, 7:03:42 PM	4.0
yt-eti-prd-job	Glue ETL	11/28/2023, 6:28:40 PM	4.0
yt-eti-fact-job	Glue ETL	11/28/2023, 4:56:15 PM	4.0
yt-eti-vids-job	Glue ETL	11/28/2023, 4:27:37 PM	4.0
yt-eti-vids-job	Glue ETL	11/28/2023, 3:56:48 PM	4.0

Temporary Database:

AWS

Services

Search

Services

[Option+S]

⏏

⏏

⏏

⏏

N, Virginia

vskrish

AWS Glue

×

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

▼ Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

► Data integration and ETL

► Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

Databases (1)

Last updated (UTC)
November 29, 2023 at 08:20:03

⌵

Edit

Delete

Add database

A database is a set of associated table definitions, organized into a logical group.

🔍 Filter databases

<input type="checkbox"/>	Name	Description	Location URI	Created on (UTC)
<input type="checkbox"/>	yt-s3-db	-	-	November 27, 2023 at 20:32:45

Contents of Temporary Database:

AWS Glue

You can now create Apache Iceberg tables in the AWS Glue Data Catalog. To learn more, visit the documentation.

yt-s3-db

Last updated (UTC) November 29, 2023 at 08:20:00 [Refresh] [Edit] [Delete]

Database properties

Name	Description	Location	Created on (UTC)
yt-s3-db	-	-	November 27, 2023 at 20:32:45

Tables (8)

Last updated (UTC) November 29, 2023 at 08:20:01 [Refresh] [Delete] [Add tables using crawler] [Add table]

View and manage all available tables.

Filter tables

<input type="checkbox"/>	Name ▲	Database ▼	Location ▼	Classification ▼	Deprecated ▼	View data	Data quality
<input type="checkbox"/>	channels_csv	yt-s3-db	s3://youtube-data-bucket-g9/j	CSV	-	Table data	View data quality
<input type="checkbox"/>	fact_csv	yt-s3-db	s3://youtube-data-bucket-g9/j	CSV	-	Table data	View data quality
<input type="checkbox"/>	period_csv	yt-s3-db	s3://youtube-data-bucket-g9/j	CSV	-	Table data	View data quality
<input type="checkbox"/>	videos_csv	yt-s3-db	s3://youtube-data-bucket-g9/j	CSV	-	Table data	View data quality
<input type="checkbox"/>	youtube_public_channel	yt-s3-db	youtube.public.channel	redshift	-	-	View data quality
<input type="checkbox"/>	youtube_public_fact	yt-s3-db	youtube.public.fact	redshift	1701121264315	-	View data quality
<input type="checkbox"/>	youtube_public_period	yt-s3-db	youtube.public.period	redshift	1701121264626	-	View data quality
<input type="checkbox"/>	youtube_public_videos	yt-s3-db	youtube.public.videos	redshift	1701121264914	-	View data quality

☒ Enable compact mode
☒ Enable new navigation

ETL Jobs:

In AWS Glue, an ETL job is created for each of the four tables to manage data flow within the ETL process.

These jobs define Extract, Transform, and Load operations tailored for each table. Here's the breakdown:

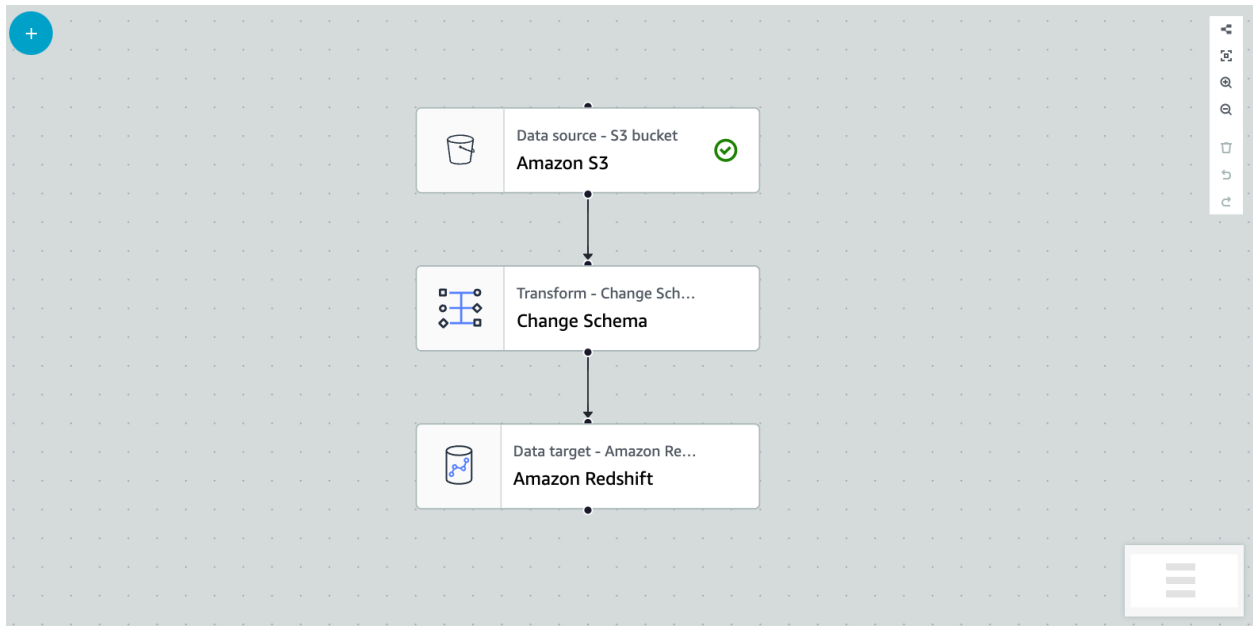
1. Extraction (Extract): Data is extracted from respective sources like Amazon S3 buckets or Redshift databases.
2. Transformation (Transform): Extracted data undergoes predefined logic-based transformations, such as cleaning, restructuring, aggregating, or enriching for analysis.
3. Loading (Load): Transformed data is loaded into the target destination, typically the respective Redshift table. This step ensures structured data storage, ready for efficient querying and analysis.

Each ETL job in AWS Glue addresses unique requirements and transformations specific to its table. Creating separate jobs for each table enables better organization, management, and optimization of the ETL process, aligning with the data characteristics within each table.

The screenshot displays the AWS Glue Studio interface. At the top, a green banner indicates 'Job delete success' with the message 'Your job(s) have been successfully deleted.' Below this, the 'Create job' section offers three options: 'Visual ETL' (Author in a visual interface focused on data flow), 'Notebook' (Author using an interactive code notebook), and 'Script editor' (Author code with a script editor). The 'Example jobs' section includes a 'Create sample job' button. The 'Your jobs (4)' section features a table listing four ETL jobs.

<input type="checkbox"/>	Job name	Type	Last modified	AWS Glue version
<input type="checkbox"/>	yt-eti-chrl-job	Glue ETL	11/28/2023, 7:50:50 PM	4.0
<input type="checkbox"/>	yt-eti-videos-job	Glue ETL	11/28/2023, 7:26:40 PM	4.0
<input type="checkbox"/>	yt-eti-prd-job	Glue ETL	11/28/2023, 6:28:40 PM	4.0
<input type="checkbox"/>	yt-eti-fact-job	Glue ETL	11/28/2023, 4:56:15 PM	4.0

ETL Workflow:



Channel ETL:

Getting started

ETL Jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data Integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

yt-etl-chnl-job

Last modified on 11/28/2023, 7:50:50 PM

Try new UI

Actions

Save

Run

Visual

Script

Job details

Runs

Data quality - updated

Schedules

Version Control

```
graph TD; A[Data source - S3 bucket Amazon S3] --> B[Transform - Change Schema Change Schema]; B --> C[Data target - Amazon Redshift Amazon Redshift];
```

Data target properties - Amazon Redshift

Name: Amazon Redshift

Node parents: Choose one or more parent node

Change Schema

Redshift access type: Direct data connection - recommended

Redshift connection: jdbc connection

Connection: View properties

Database: youtube

Schema: public

Table: channel

Handling of data and target table: APPEND (insert) to target table

Data preview

Output schema

Schema AVAILABLE

Infer schema from session

Key	Data type
channel_id	string
channel_name	string
subscriber_count	string
view_count	string
video_count	string

Videos ETL:

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

yt-etl-videos-job

Last modified on 11/28/2023, 7:26:40 PM

Try new UI

Actions

Save

Run

Visual

Script

Job details

Runs

Data quality - updated

Schedules

Version Control

+

Amazon S3

Transform - Change Sch...

Change Schema

Amazon Redshift

Data preview

Output schema

Data preview

Info

READY

Filter sample dataset

End session

Previewing 0 of 0 fields

Loading data previews. (100% complete)

Data source properties - S3

Name

Amazon S3

S3 source type

info

S3 location

Choose a file or folder in an S3 bucket.

Data Catalog table

S3 URL

s3://youtube-data-bucket-...

View

Browse S3

Recursive

Read files in all subdirectories.

Data format

CSV

Delimiter

Comma (,)

Escape character - optional

Enter a character to use for escaping

Quote character

Double quote (")

First line of source file contains column headers

Records in source files can span multiple lines

Infer schema

Execution of ETL Job:

Getting started

ETL jobs

Visual ETL

Notebooks

Job run monitoring

Data Catalog tables

Data connections

Workflows (orchestration)

Data Catalog

Databases

Tables

Stream schema registries

Schemas

Connections

Crawlers

Classifiers

Catalog settings

Data integration and ETL

Legacy pages

What's New

Documentation

AWS Marketplace

Enable compact mode

Enable new navigation

yt-etl-fact-job

Last modified on 11/28/2023, 4:56:15 PM

Try new UI

Actions

Save

Run

Visual

Script

Job details

Runs

Data quality - updated

Schedules

Version Control

Job runs (1/1)

Info

Last updated (UTC)

November 29, 2023 at 08:23:22

View details

Stop job run

Table View

Card View

Filter job runs by property

Run status

Retries

Start time (UTC)

End time (UTC)

Duration

Capacity (DPUs)

Worker type

Glue version

Succeeded

0

2023/11/29 00:56:19

2023/11/29 00:58:10

1 m 34 s

10 DPUs

G.1X

4.0

Run details

Input arguments (10)

Continuous logs

Run Insights

Metrics

Spark UI

Job name

yt-etl-fact-job

Start time (UTC)

November 28, 2023 12:56:19 AM

Glue version

4.0

Last modified on (UTC)

November 28, 2023 12:58:10 AM

Id

jr_282ec889d932ac56c5d7ba4527f36c1718097fb48a8755c69e3a20b09dc10af2

End time (UTC)

November 28, 2023 12:58:10 AM

Worker type

G.1X

Log group name

/aws-glue/jobs

Run status

Succeeded

Start-up time

16 seconds

Max capacity

10 DPUs

Number of workers

10

Retry attempt number

initial run

Execution time

1 minute 3.4 seconds

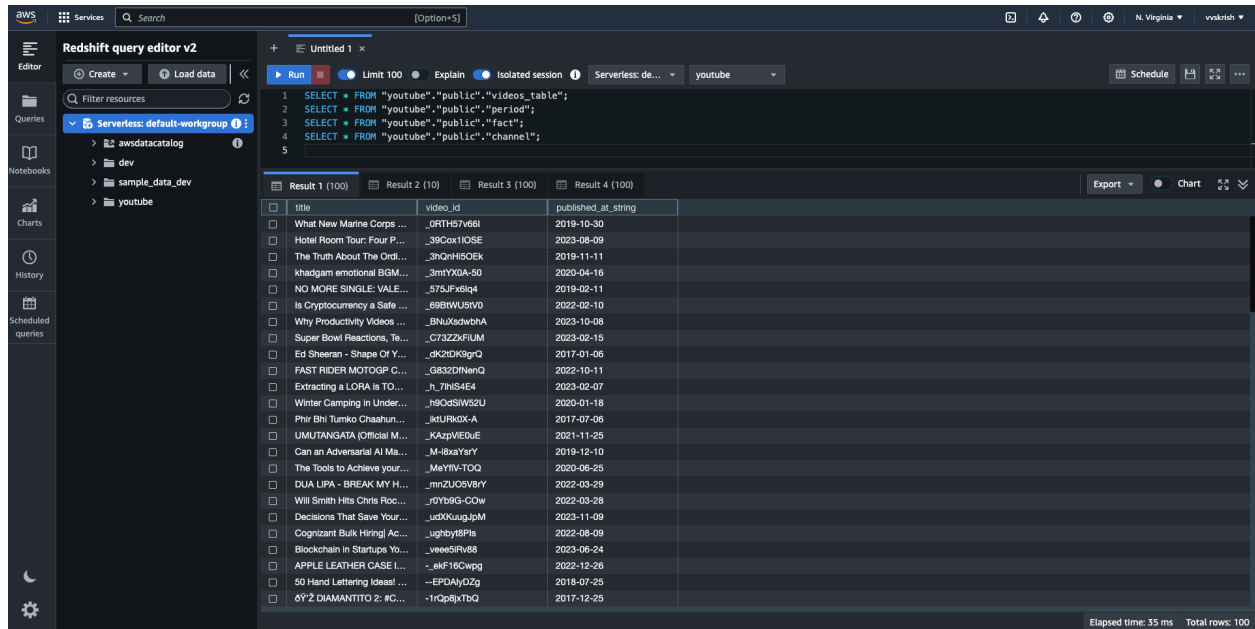
Execution class

Standard

Timeout

7800 minutes

Videos data loaded into the Redshift cluster:



The screenshot shows the AWS Redshift Query Editor v2 interface. The SQL query in the editor is:

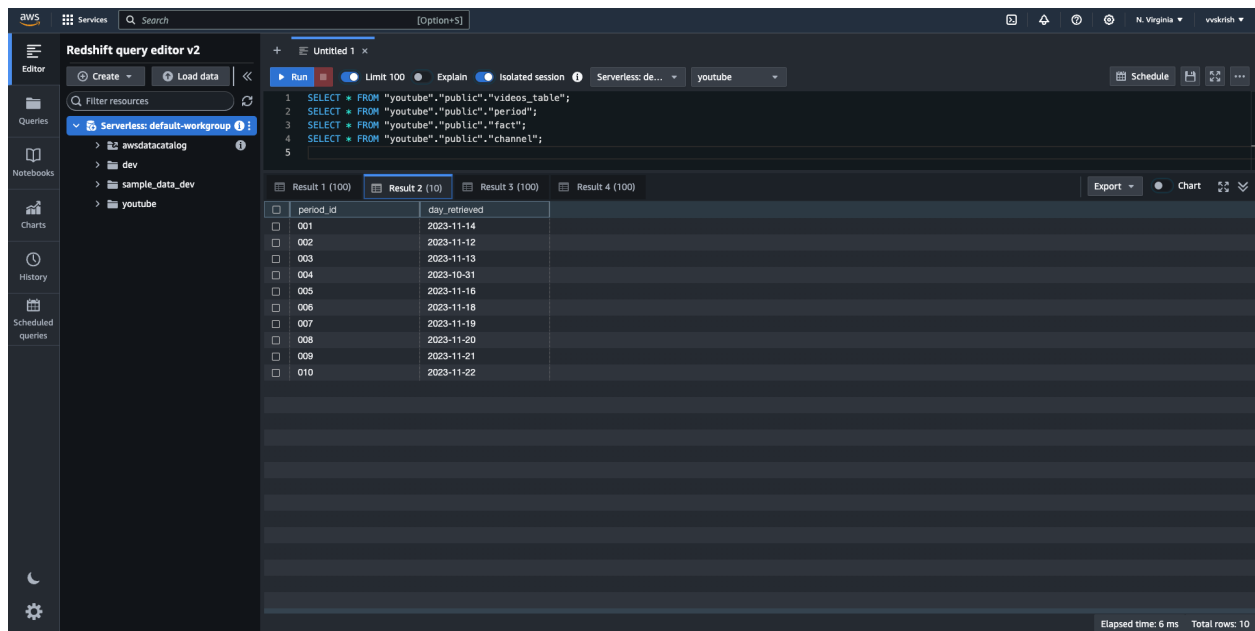
```
1 SELECT * FROM "youtube"."public"."videos_table";
2 SELECT * FROM "youtube"."public"."period";
3 SELECT * FROM "youtube"."public"."fact";
4 SELECT * FROM "youtube"."public"."channel";
5
```

The results are displayed in a table with the following columns: title, video_id, and published_at_string. The table contains 100 rows of data.

title	video_id	published_at_string
What New Marine Corps ...	_0RTH57v66l	2019-10-30
Hotel Room Tour: Four P...	_39Cox1JOSE	2023-08-09
The Truth About The Ord...	_3hQnH5OEK	2019-11-11
khadgam emotional BGM...	_3mtYX0A-50	2020-04-16
NO MORE SINGLE: VALE...	_575JFXeiq4	2019-02-11
Is Cryptocurrency a Safe ...	_69BWUw0V0	2022-02-10
Why Productivity Videos ...	_BNuXsdwbthA	2023-10-08
Super Bowl Reactions, Te...	_C73Z2kFIUM	2023-02-15
Ed Sheeran - Shape Of Y...	_dK2lDK9gQ	2017-01-06
FAST RIDER MOTOGP C...	_G832DNeuQ	2022-10-11
Extracting a LORA is TO...	_h_7hNS4E4	2023-02-07
Winter Camping in Under...	_h8QdSIW52U	2020-01-18
Phir Bhi Tumko Chashun...	_ktURK0X-A	2017-07-06
UMUTANGATA (Official M...	_KApzVIE0UE	2021-11-25
Can an Adversarial AI Ma...	_M-i8kaYarY	2019-12-10
The Tools to Achieve your...	_MeYfIV-TQQ	2020-06-25
DUA LIPA - BREAK MY H...	_mnZU0SV8Y	2022-03-29
Will Smith Hits Chris Roc...	_r0Yb9G-C0w	2022-03-28
Decisions That Save Your...	_udKkuugJpM	2023-11-09
Cognizant Bulk Hiring! Ac...	_ughbytBPis	2022-08-09
Blockchain in Startups Yo...	_veeeSiRv88	2023-06-24
APPLE LEATHER CASE I...	_ekF16Cwpg	2022-12-26
50 Hand Lettering Ideast ...	_EPDAlyQZg	2018-07-25
dyZ DIAMANTITO 2: #C...	_1Rq8jxTbQ	2017-12-25

Elapsed time: 35 ms Total rows: 100

Period data loaded into the Redshift cluster:



The screenshot shows the AWS Redshift Query Editor v2 interface. The SQL query in the editor is:

```
1 SELECT * FROM "youtube"."public"."videos_table";
2 SELECT * FROM "youtube"."public"."period";
3 SELECT * FROM "youtube"."public"."fact";
4 SELECT * FROM "youtube"."public"."channel";
5
```

The results are displayed in a table with the following columns: period_id and day_retrieved. The table contains 10 rows of data.

period_id	day_retrieved
001	2023-11-14
002	2023-11-12
003	2023-11-13
004	2023-10-31
005	2023-11-16
006	2023-11-18
007	2023-11-19
008	2023-11-20
009	2023-11-21
010	2023-11-22

Elapsed time: 6 ms Total rows: 10

Fact Table loaded into the Redshift cluster:

The screenshot shows the AWS Redshift Query Editor v2 interface. The SQL query in the editor is:

```
1 SELECT * FROM "youtube"."public"."videos_table";
2 SELECT * FROM "youtube"."public"."period";
3 SELECT * FROM "youtube"."public"."fact";
4 SELECT * FROM "youtube"."public"."channel";
5
```

The results are displayed in a table with the following columns: period, comment_count, like_count, view_count, channel_id, video_id, and id_string. The table contains 100 rows of data.

period	comment_count	like_count	view_count	channel_id	video_id	id_string
007	109	1081	52541	UCsASYLcaJmhu73vz1...	SCS0M4J-eww	21875
007	110	730	139107	UC1sKieSogV-APEA4Gh...	sYQ3ZV-GY80	21876
007	7	76	2500	UCMxymPG6WLLJ4exb...	Gupq4FpPhM	21877
007	2	5	101	UCIzhonT3_uhN3w1oYtZ...	0vZBNGYH4Gc	21878
007	2	13	1982	UCmnAvVDGSD7EiUEb...	Flr1kaKak1g	21879
007	4	18	6569	UCmnAvVDGSD7EiUEb...	c1oxeYfRkQ	21880
007	6	65	3490	UCLJ00k4HUBRyYf4rt...	3jhdv8MuDg	21881
007	7	22	9796	UCmnAvVDGSD7EiUEb...	mCmn3bb26xc	21882
007	0	8	210	UCBFg4gUnVxQkgGsz...	eyP3u2aPPQ	21883
007	0	1	50	UCF1mnWfCs0_sq_k8L...	7Hak7-NINGC	21884
007	0	2	60	UCF1mnWfCs0_sq_k8L...	ihvGUG0weo	21885
007	1	1	27	UC4GUny57ZJ57IasJv9...	qjFfC-ng2j8	21886
007	5	70	2018	UCyuGoBVJ0oL08FwG...	BBi48HvxE	21887
007	1	2	27	UCLYsP5y9SxQgAlfo3U...	_veee5Rv68	21888
007	0	1839	712785	UCHuiy8bXnmK5n8YHU...	m8baYkVYhw	21889
007	0	4	301	UCwB7c1UaRfU2RnUJK...	3Kgy--ENQdM	21890
007	16	326	18870	UCD78HSHWYqMyeR0...	O_JN32a6x8	21891
007	0	77	8438	UCL9BHSRowqQ1TyBn...	aRhm0Z4bNI	21892
007	0	1	68	UCobq2sQJ-dSIL3vBW...	oHeATDFtT4	21893
007	282	15256	165438	UC2pG-jfnadWgXO_Ify...	GPKNIQRIkFA	21894
007	128	1212	29251	UCHj2_6e0Y2YdzHZDk...	ZSOZw7WavN8	21895
007	2	11	641	UCv-LWwHDTwmS3YgY...	xzra9vJaJqM	21896
007	140	1389	36801	UCHj2_6e0Y2YdzHZDk...	x51L-prDFil	21897
007	0	23	598	UCIK-2xSw_sJBewF7...	T4TMB7MwVz8	21898

Elapsed time: 5 ms Total rows: 100

Channel Table loaded into the Redshift cluster:

The screenshot shows the AWS Redshift Query Editor v2 interface. The SQL query in the editor is:

```
1 SELECT * FROM "youtube"."public"."videos_table";
2 SELECT * FROM "youtube"."public"."period";
3 SELECT * FROM "youtube"."public"."fact";
4 SELECT * FROM "youtube"."public"."channel";
5
```

The results are displayed in a table with the following columns: id, channel_name, video_count, view_count, subscriber_count, and channel_id. The table contains 100 rows of data.

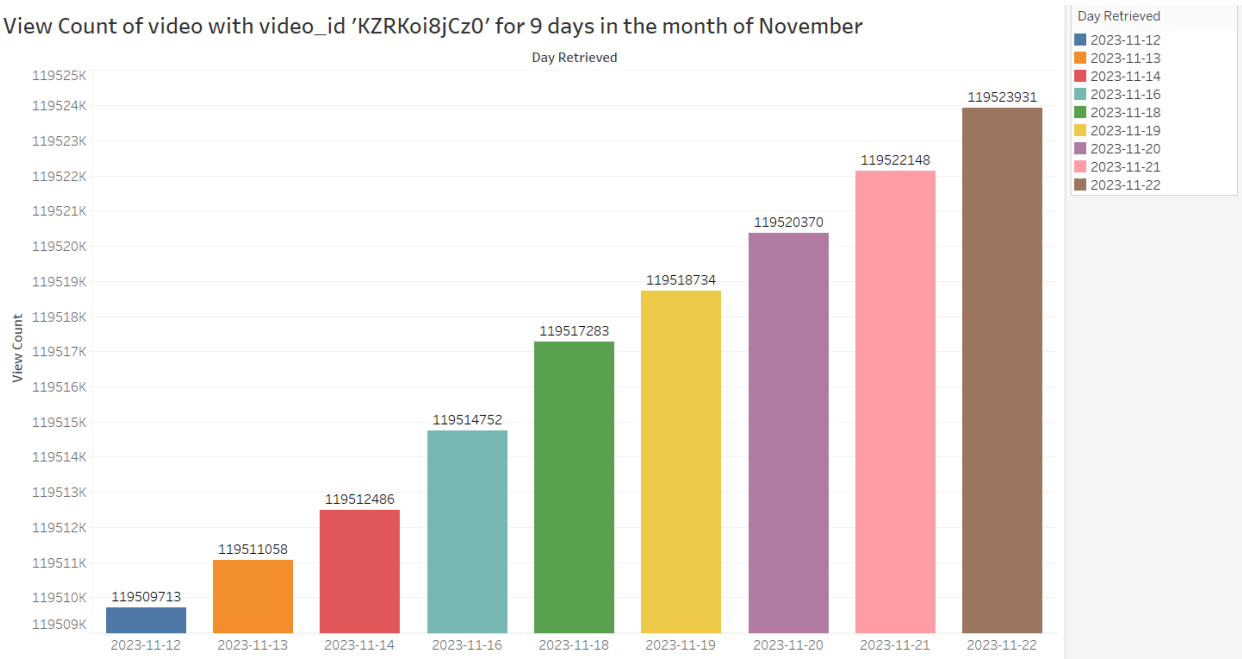
id	channel_name	video_count	view_count	subscriber_count	channel_id
NULL	ABC	13216	2378203254	2860000	UC_JpS8O_-IRA6RNe05...
NULL	Betty Franks Art	209	5397775	77100	UC_Lt0k0ngMzTnGdS...
NULL	Strange Loop Conference	718	8486229	77200	UC_QthvA8uay2CoOisf...
NULL	The Easy Stuff	17	684	11	UC_sriEWjpaP-apQ08...
NULL	Let's Figure Out Adulting	392	831279	4440	UC_uFetXZ0aHdsV4bL...
NULL	Hugo Cruz	2	11281	9	UC_Znd_HJ1tweyQ09K2...
NULL	Koi Molina	9	4370	7	UC_78bDumJubDjvPL3...
NULL	Kuthuty	78	1383520	7190	UC_4_DaYf1K6Cqg5WK...
NULL	Katy Perry	133	25746257988	4450000	UC-8Q-HLdECwQmaWN...
NULL	Stanford	4051	275483027	1840000	UC-EnpmCZ3OXyAaG7...
NULL	PewDiePie	4740	29160377338	11100000	UC-1kUZR3Gqm24_Va...
NULL	DesiMax	33	86223	1600	UC-ZABxgEnBnFOOF0e...
NULL	A Couch Potato Girl	205	8678895	197000	UC-zK_W8TJvZ7OpLe...
NULL	Ed Sheeran	451	31159950083	5380000	UCOC-wOY3pzdXGB8BH...
NULL	Juantika Reynolds-Jackson	4	225	1	UCQDn5cl47patYxrsY...
NULL	Jon Mercado	118	18969	136	UCQfD6dXIBXyK7RTG...
NULL	Miles Minnick	346	6925596	30200	UCQAp1HE2LwGpS4G...
NULL	DIY Pinto	7092	97221279	112000	UCQGHf05mTzjmsRLG...
NULL	Jake Messenger	10	1014	6	UCQHCDUBoqPW4eUBK...
NULL	albert_cancook	627	10602423224	15400000	UCQnKsniC0hoOdykz...
NULL	Sportify	79	98346	461	UCQpWSRkuuQMLrnbtl...
NULL	Ariana Grande	154	24673381554	53100000	UCQVOyTZOCBkdQhF3B...
NULL	Girl Geek X	285	162819	1520	UCQwFWzkhxXJ_g3IsO...
NULL	San Jose Public Library	788	773245	3440	UCQvRhOXDgJBSHLcaB...

Elapsed time: 3551 ms Total rows: 100

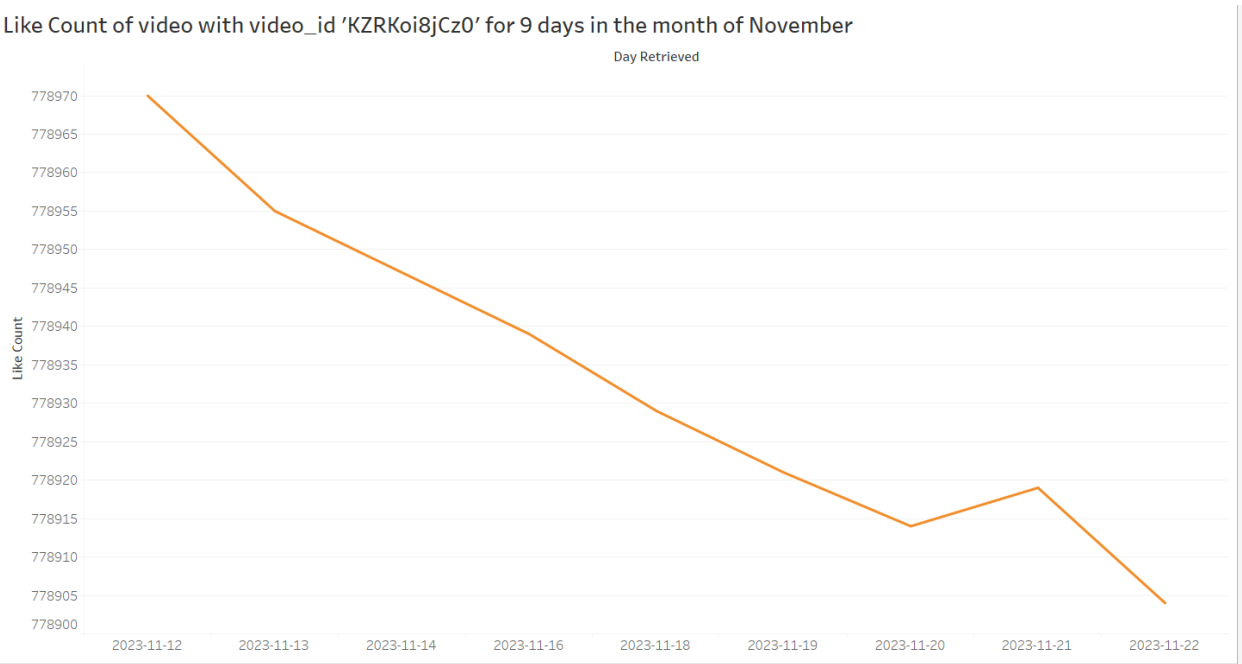
Visualizations (in Tableau):

We connected Amazon Redshift with Tableau and performed visualizations.

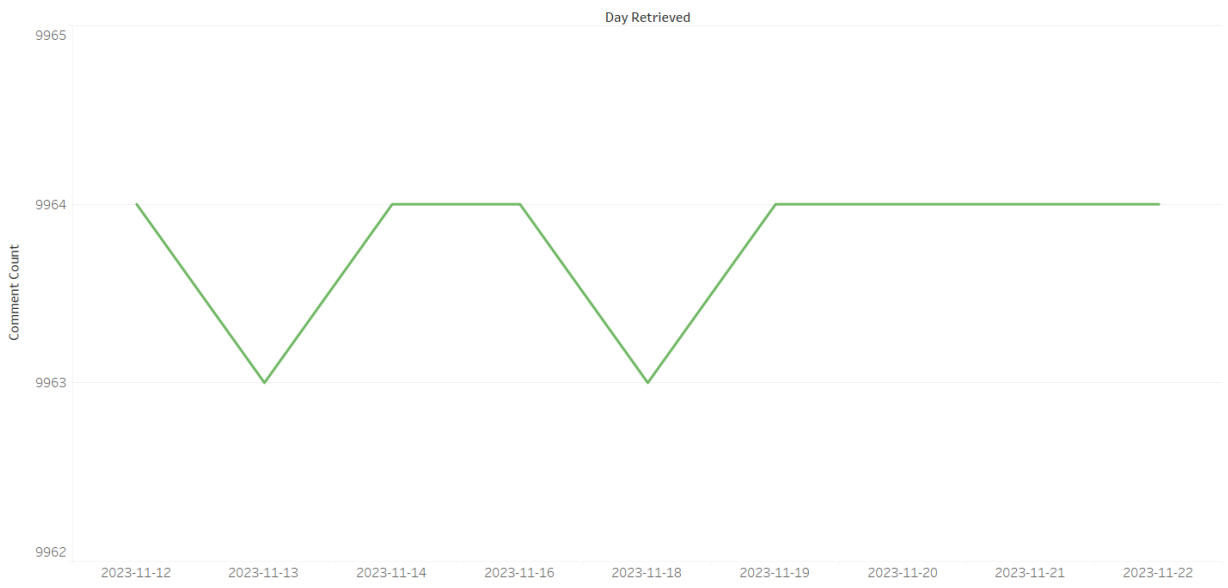
View Count of video with video_id 'KZRKoi8jCz0' for 9 days in the month of November



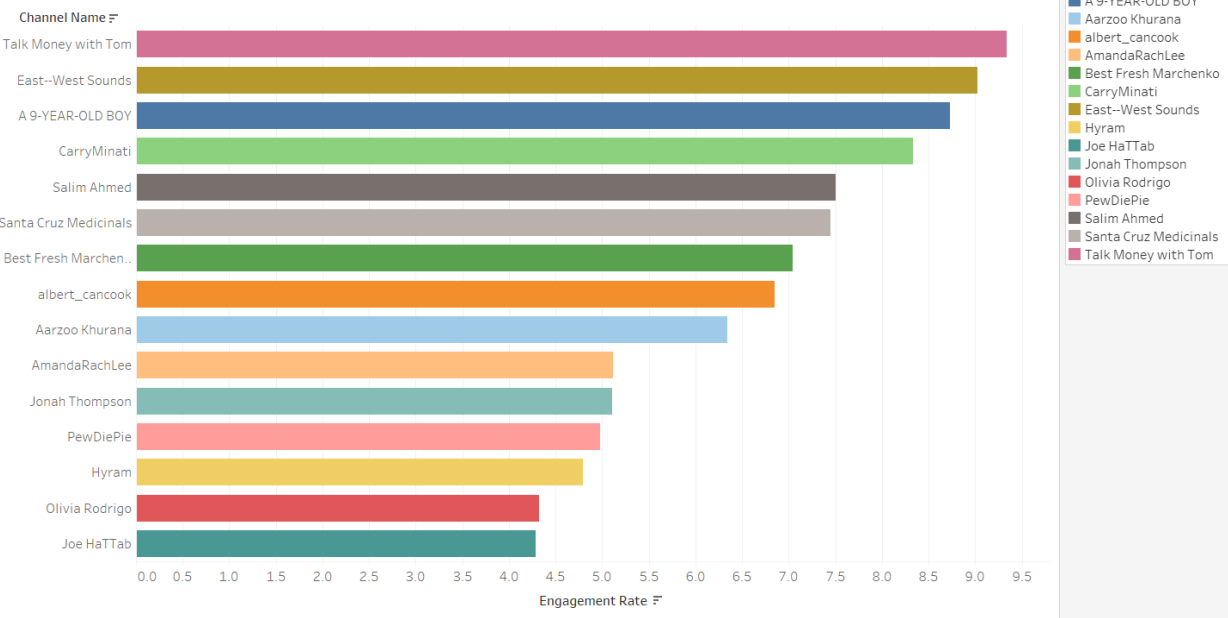
Like Count of video with video_id 'KZRKoi8jCz0' for 9 days in the month of November

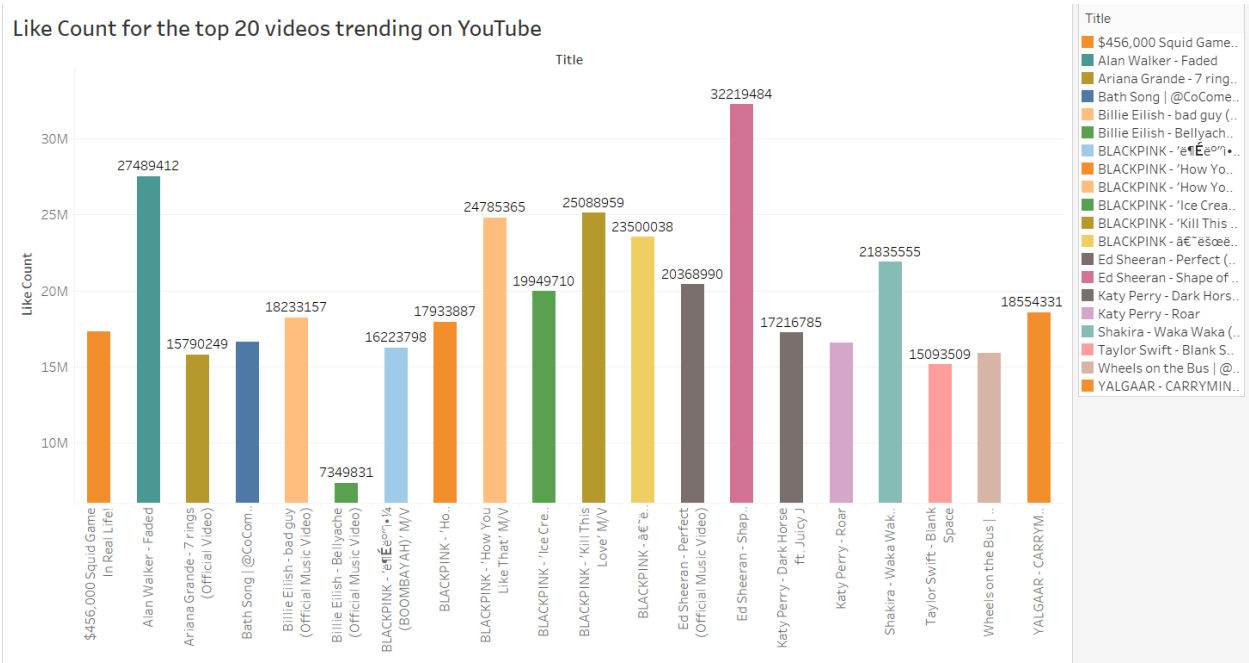
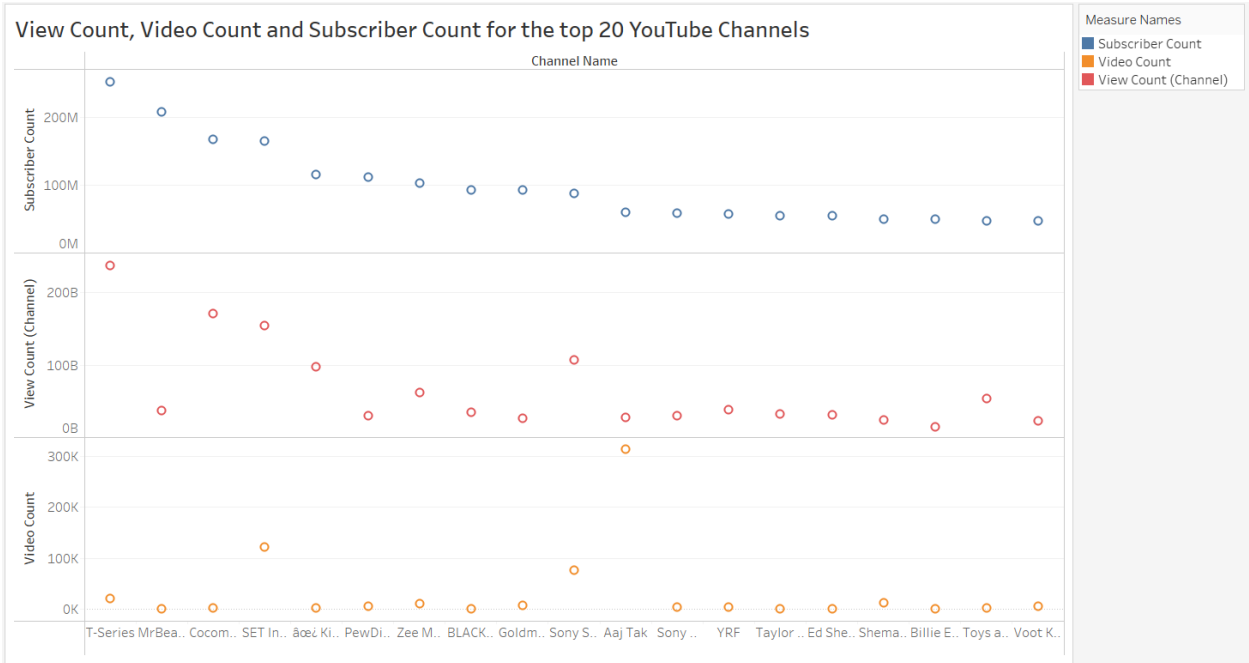


Comment Count of video with video_id 'KZRKoi8jCz0' for 9 days in the month of November



Top 20 Channels with Highest Engagement Rate





View Count, Video Count and Subscriber Count for the top 20 YouTube Channels

