

User Requirements Doc: Top SA Youtubers 2024 Dashboard

This project used the phases of data analysis: **ASK, PREPARE, PROCESS, ANALYSE, and SHARE**

PHASE 1: ASK- PROBLEM DEFINITION

➤ Problems identified

Litu, the regional project coordinator for Project Hope, a non-profit organization that promotes adolescent health among teenagers in South Africa is looking for the top 3 YouTubers in the country to run an outreach campaign. However, she is finding it difficult to identify them. She has performed online research but constantly bumps into overly complicated and conflicting insights. She has also held calls with different third-party providers, but they are all expensive options for underwhelming results. So, she asked her data analyst team to give her insights and make a dashboard for her to make this decision.

➤ Objective

To discover the top performing SA Youtubers to form the outreach adolescent health promoting team for the year,2024.

➤ Stakeholders expectations

The regional coordinator and her team would like to identify the top YouTubers in SA based on the number of subscribers(followers), views accumulated, and the YouTuber name who runs the channel. She can then decide which channels are best to promote adolescent health throughout the campaign. So, they would like a dashboard that included the following:

- List the top YouTube channels by subscribers and views
- Display key metrics (channel name, subscribers, views, views per subscribers)
- Be user-friendly and easy to filter/sort
- Use the most recent data possible

Success criteria

Litu can

- Easily identify the top-performing YouTube channels based on the key metrics mentioned above
- Assess the potential for successful campaigns with top Youtubers based on reach, engagement, and potential revenue
- Make informed decisions on the ideal collaborations to advance based on recommendations

This allows Litu to make data-driven decisions and build relationships with YouTubers for future collaborations, which leads to recognition within the company.

The dashboard to be

PHASE 2: PREPARE THE DATA

DATA SOURCE

The data is extracted from Kaggle(an Excel extract) available @ www.kaggle.com/datasets/bhvyadhingra00020/top-100-social-media-influencers-2024-countrywise?resource=download

Information needed

The data should include:

- Total subscribers
- Total Views
- Channel names and YouTuber names

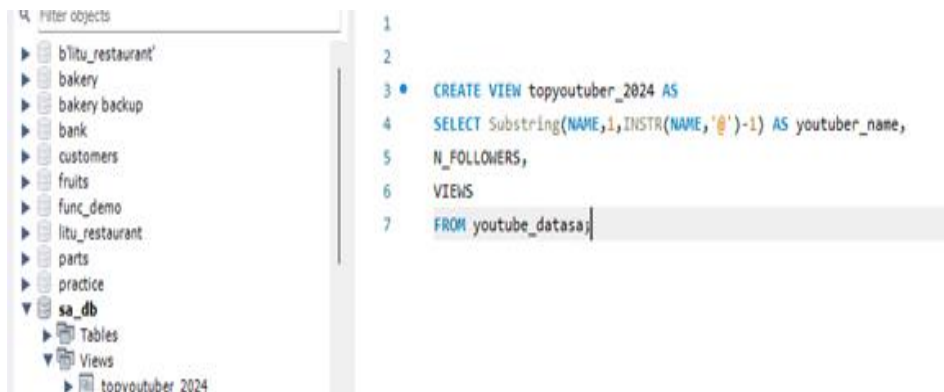
PHASE 3: PROCESS THE DATA

The aim is to clean data and remove errors: both Excel and MYSQL were used

- Data such as total subscribers were in letter formats e.g 2M and were converted to numbers in Excel (e.g.2000000)

MYSQL DATA CLEANING PROCESS

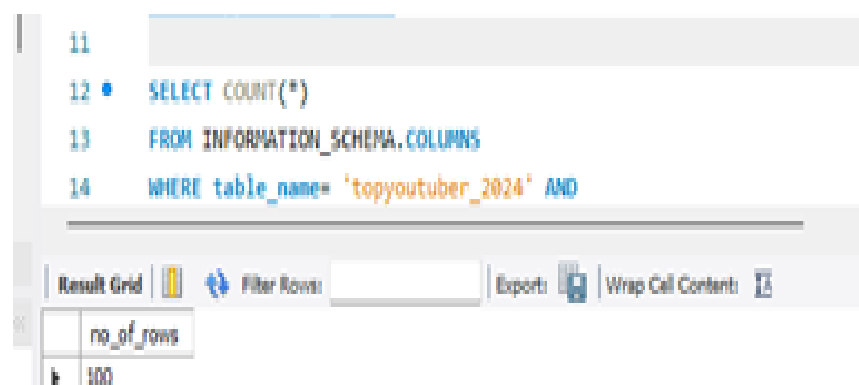
1. Create view



Data quality checks

We need to add measures in place to confirm the dataset contains the data required without any issues – here are some of the data quality check conducted:

- Row count check -PASS



- Column count check-PASS

SCHEMAS

Filter objects

- b'litu_restaurant'
- bakery
- bakery backup
- bank
- customers
- fruits
- func_demo
- litu_restaurant
- parts
- practice
- sa_db
 - Tables
 - Views
 - topyoutuber_2024

Administration Schemas

Information

No object selected

```

6 VIEWS
7 FROM youtube_datasa;
8
9 • SELECT count(*) AS no_of_rows
10 FROM topyoutuber_2024;
11
12 • SELECT COUNT(*)
13 FROM INFORMATION_SCHEMA.COLUMNS
14 WHERE table_name= 'topyoutuber_2024' AND
15 table_schema='sa_db';
16

```

Result Grid

COUNT(*)
3

- Data type check-PASS

Data needed

The dataset to produce the information we need should include the following fields

- Channel name (string, text)
- Total subscribers (integer)
- Total views (integer)

```

17
18 • SELECT *
19 FROM INFORMATION_SCHEMA.COLUMNS
20 WHERE table_name= 'topyoutuber_2024' AND
21 table_schema='sa_db';
22

```

Result Grid

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	CHARACTER
▶	def	sa_db	topyoutuber_2024	N_FOLLOWERS	2	NULL	YES	int		NULL
	def	sa_db	topyoutuber_2024	VIEWS	3	NULL	YES	int		NULL
	def	sa_db	topyoutuber_2024	youtuber_name	1	NULL	YES	mediumtext	16777215	16777215

- Duplicate check-PASS

```

22
23 -- check duplicates---
24 • SELECT youtuber_name,
25 count(*) as duplicate_count
26 from topyoutuber_2024
27 group by youtuber_name
28 having count(*)>1;

```

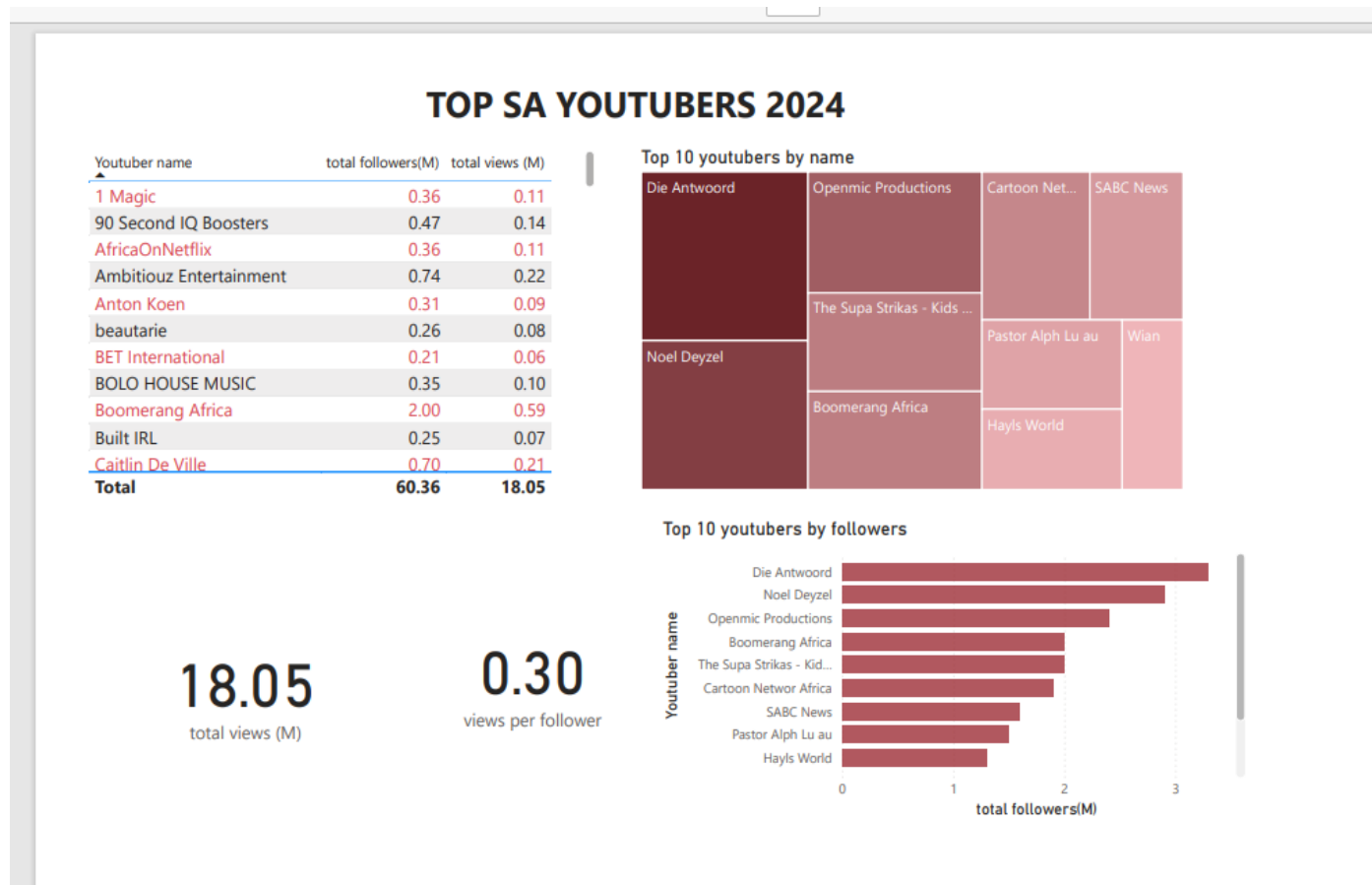
Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

youtuber_name	duplicate_count
---------------	-----------------

PHASE 4: ANALYSE THE DATA: VISUALIZATION

Visualization was done in POWER BI

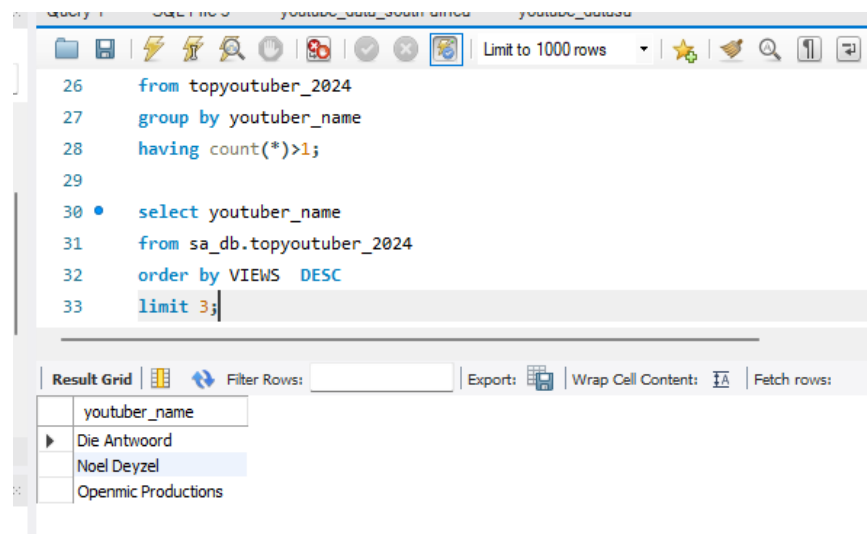
This is the dashboard for the Top SA YouTubers in 2024 so far



PHASE 5: SHARE FINDINGS WITH STAKEHOLDERS

For this analysis, we have answered 2 questions to get the information we need for the regional coordinator to make her decisions.

1. Who are the top 10 YouTubers with the most subscribers?
 - **Die Antwoord, Noel Deyzel and Openmic productions** are the channels with the most subscribers in SA
2. Which 3 channels have the most views?



The screenshot shows a SQL query editor with the following code:

```
26 from topyoutuber_2024
27 group by youtuber_name
28 having count(*)>1;
29
30 • select youtuber_name
31 from sa_db.topyoutuber_2024
32 order by VIEWS DESC
33 limit 3;
```

Below the query editor, the 'Result Grid' is displayed with the following data:

youtuber_name
Die Antwoord
Noel Deyzel
Openmic Productions

Recommendations to the regional coordinator

- I recommend **Die Antwoord, Noel Deyzel and Openmic Productions** as the **top 3 YouTube channels to collaborate with** for the campaign if you want to increase visibility because **these 3 channels have the most YouTube subscribers and attract a lot of views in SA.**