


<div>COMPANY</div>  <a href="#">Click here to access the company site</a>	Project	<a href="#">Click here to access the project</a>
	Date	August - 2022
	Description	It was a data analysis project developed to developed with the aim of obtaining approval for an application to a data analyst vacancy at the Eleflow company.
	Index	<a href="#">1. Challenge proposed by the company</a> <a href="#">2. Report</a> <a href="#">3. Solution overview</a>

### 1. Challenge proposed by the company

In a hypothetical scenario, Eleflow was hired by Google to provide data analysis consultancy for the YouTube website. The candidate should answer the questions:

1. Daily and monthly publications in the selected (category/country/period).
2. Which 5 channels have the highest views in the selected (category/country/period)?
3. Which 5 videos, common to the 4 countries, have the most likes?
4. Which 5 videos have the highest number of dislikes in the selected (category/country/period)?

[^Return to the index.](#)

### 2. Report

Below, Fig. 1, we have the image from the final report. It also may be accessed through the link: <https://app.powerbi.com/view?r=eyJrJoiM2YwZGQ0NzEtNjg2Yi00NGM3LTgxOWYtM2I1ZTNmMjAxNzAxliwidCI6IjYxM2Y2OWM0LTkyOTQtNDMxZS04ZWEyLThtN2Q1NWUxYzZzZCJ9>.

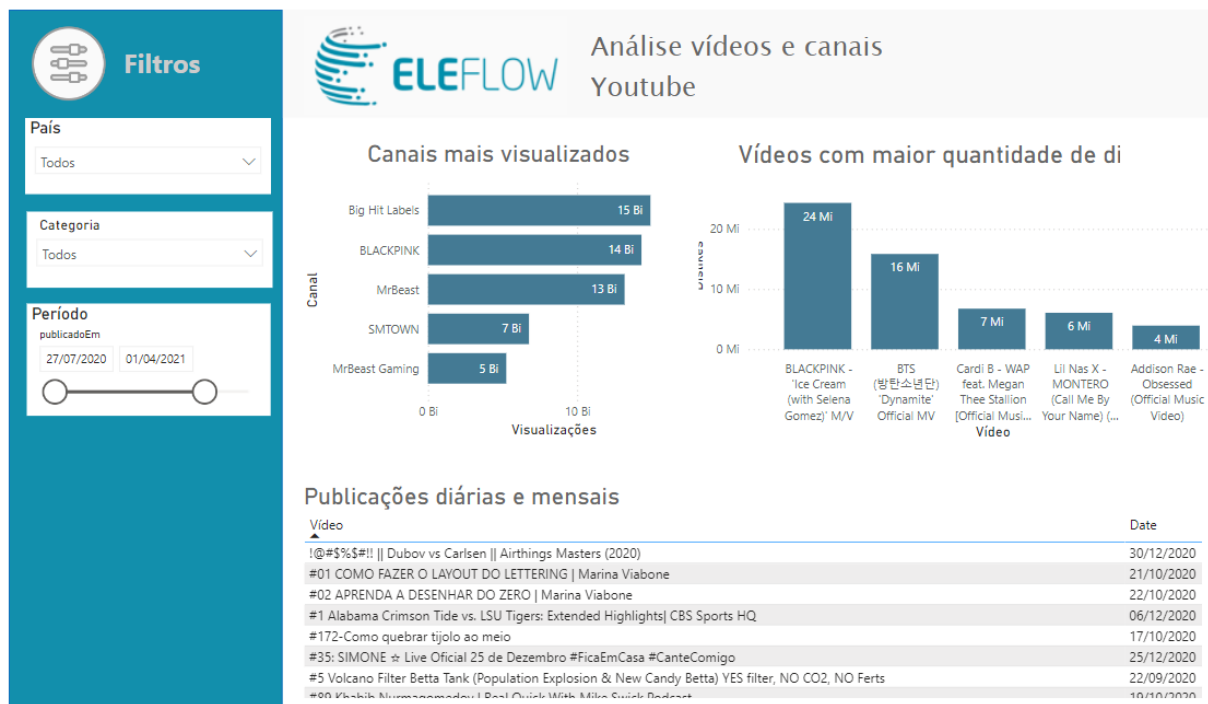


Fig. 1 – Final report.

[^Return to the index.](#)

### 3. Solution overview

To solve the problem was need three steps. To the first step was necessary merge the datasets because the data were sent throw three files, Fig. 2, one to each country.

- CA\_youtube\_trending\_data.csv
- BR\_youtube\_trending\_data.csv
- GB\_youtube\_trending\_data.csv
- US\_youtube\_trending\_data.csv

Fig. 2 – Files containing the data to merge.

The original dataset was available throw a unique table, the Table 1 shows the column's name. Due this condition, at the second step, the data were modeled according to the 'star schema'.

video_id	trending_date	thumbnail_link
----------	---------------	----------------

Title	tags	comments_disabled
publishedAt	view_count	ratings_disabled
channelId	likes	description
channelTitle	dislikes	
categoryId	comment_count	

Table 1 – Columns name from the original dataset.

Below, the star schema developed, Fig. 3. The fact table contains the mainly data about the access to the videos. The dimension table contain specific data or calendar data.

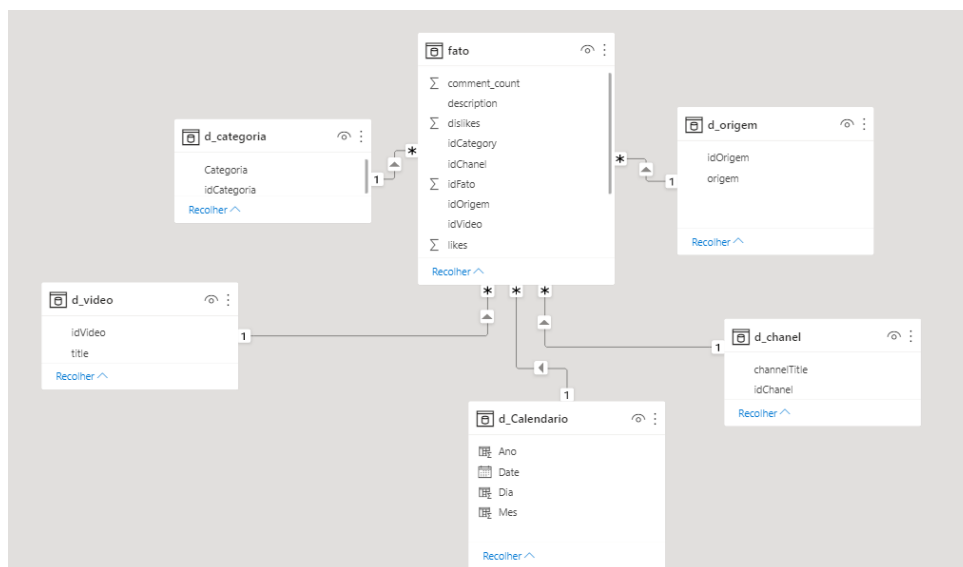


Fig. 3 – ‘Star schema developed to execute the analysis.

Finally, in the third step, the data were cleaned throw DAX and the final report was generated.

[^Return to the index.](#)