Youtube's Statistics

Mariana Rosa 98390, Simão Arrais 85132

Information Visualization, 2022 (Master Degree in Informatic Engineer, University of Aveiro)

Abstract

This report aims to explain the implementation of the platform developed for the project of the Visualization Information course - "conceptualization and implementation of a prototype of a simple application for visual exploration". We will be using D3.js, SVG.js and Google Javascript API to build the visualizations. This project is about Youtube's statistics, where the user must be able to see what videos/categories are trending and what metrics are associated with each one.

Motivation and objectives

YouTube is a popular video sharing platform that allows users to upload, share, and view videos. As a result, it has become a valuable source of information and data for many businesses, organizations, and individuals. In order to make the most of this data, it is important to have effective visualization solutions in place. The motivation for this project is to create an easy-to-use application for Youtube users and marketing managers to better understand and analyze the statistics of Youtube videos. The objectives of the project are to provide answers to questions about the most popular categories of videos, the number of likes and views for videos in a specific category, and which videos are currently trending across the platform or for specific categories. The application uses a dataset that allows users to analyze the popularity and sentiment of specific videos and categories on Youtube.

Users and the Questions

To build an easy-to-use application, the Youtube Statistics platform must know the target users and the questions to be answered.

Characterization of the users and their context

Nowadays, Youtube is not just a hobby. It can be seen as an actual business. Every day Youtubers can make thousands of dollars in revenue and partner with brands to promote them. People are switching from TV ads to Youtube ones.

Therefore, the main users to our website will be Youtubers themselves, marketing managers, or just curious people.

In order to idealize better our users, two personas were invented:

	Persona 1	**Persona 2**
Name	Tom Fox	Sara Sousa
Age	37	27
Country	England	Portugal
Job title	Social Media Manager	Content Creator
Goals and task	Tom works for a Technological brand, TechNow. He is looking for a big Youtuber to do Marketing for his company. Therefore, he needs to analyze some of the candidates.	Check how everyone else is doing (in statistics), to compare to her channel worldwide; See which category is more popular to know which content is trending.

Questions to Answer

The user, using our application, can answer to multiple questions:

1. What are the most seen categories of videos?
2. What are the most liked videos on Youtube?
3. How many likes do videos from the Tech category/keyword have?
4. How many views do videos from Tech category have?
5. Who was the Youtuber with the most views on a video in the Sports category?
6. How many videos from "Tech" category are trending on Youtube?
7. What videos are trending in Sports?

Dataset

The chosen dataset was taken from Kaggle and it contains the following information columns:

- *Title*: Video Title.
- *Video ID*: The Video Identifier.
- *Published At*: The date the video was published in *YYYY-MM-DD*.
- *Keyword*: The keyword associated with the video.
- *Likes*: The number of likes the video received.
- Comments: The number of comments the video has.
- *Views*: The number of views the video got.

This dataset is useful for tasks related to video/keyword popularity and sentiment analysis which is exactly what we need in order to fulfill the personas goals and tasks.

Visualization Solution

To create a minimal website, that is intuitive and user-friendly, that could answer these questions, first, we need to build a Low fidelity prototype and ask users about it.

Low fidelity prototype and user feedback

A Low fidelity prototype was drawn on a digital surface and it was tested with colleagues in the class on 9th November.

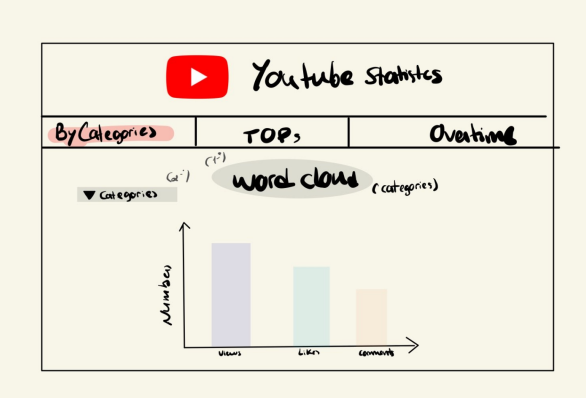


Figure 1: Aspect of the low fidelity prototype page of statistics by category/keyword

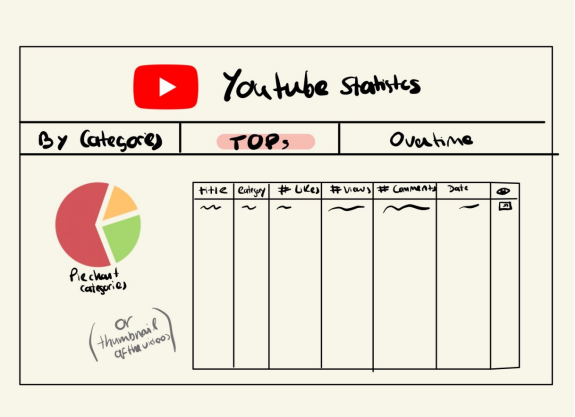


Figure 2: Aspect of the low fidelity prototype page of the hits (top videos)

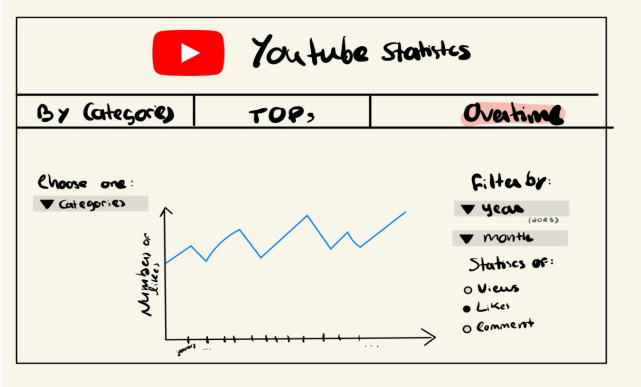


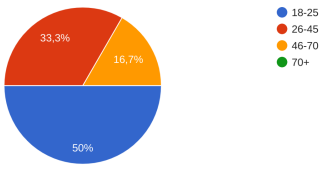
Figure 3: Aspect of the low fidelity prototype page of statistics by time

In this class, we did a user usability test, really important to know if the application we have drawn was natural to use.

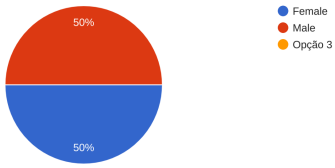
Since we were not satisfied with our testers, because they were all from the same niche (informatic students and they were practically the same age), we invited our family and friends to join the study.

Participants information

Age:



Gender



Ranking

In a 5-10 minutes test, users will perform five tasks and rank them in a scale of 1 to 5, being:

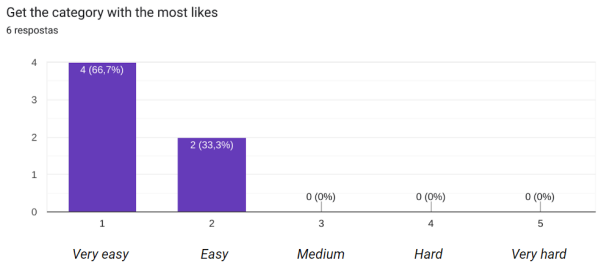
1. Very easy
2. Easy
3. Medium
4. Hard
5. Very hard

Tasks

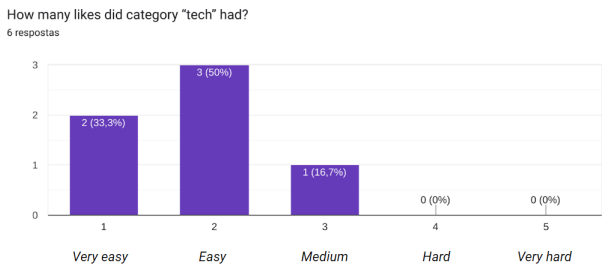
1. Get the category with the most likes;
2. How many likes did category “tech” have?
3. How many views did the category “interview” have?
4. What was the YouTuber with the most viewed video in the category Computer Science?

Results

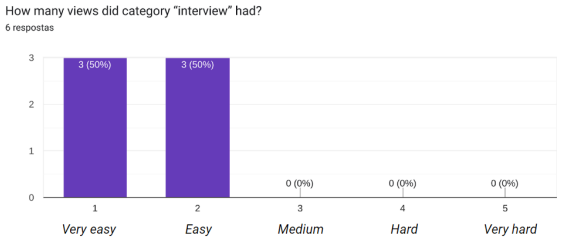
Task 1:



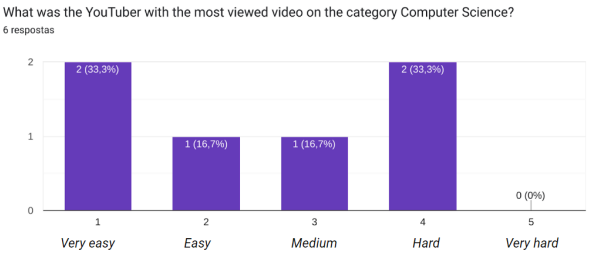
Task 2:



Task 3:



Task 4:



Main feedback and changes

In Task 2 and 3, people sometimes were confused on where to look up the information because it was duplicated. When looking up for a certain category’s statistics, first a WordCloud would show up, the user clicked and then it would appear the bar chart and a dropdown menu to show the categories. Since the Word Cloud would have a lot of information at the same time, it’s better to have two ways of searching. Also, the dropdown menu is much clearer and organized to look up on a word. We also removed the comments bar since the number was irrelevant (Fig. 4).

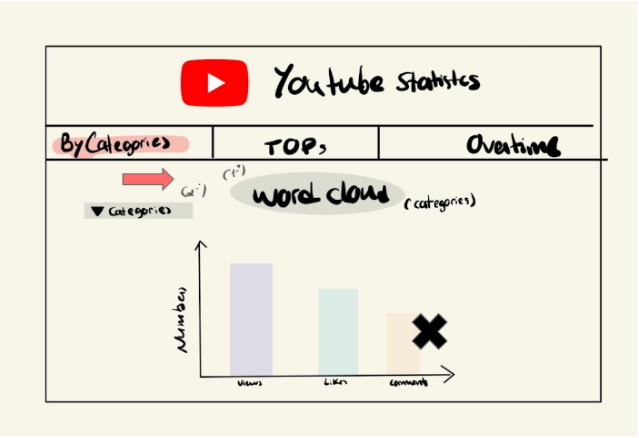


Figure 4: Changes according to the user test to the “By categories” page

Heuristics

After talking with a few participants, we tried to evaluate the system according to Jakob Nielsen's Heuristics. We choose three that can be a small problem:

- Visibility of system status (1, Cosmetic problem)
  - You are able to get the same information from multiple places and sometimes the user get confused on where to pick up the information on the page.

- Error prevention (1, Cosmetic problem)
  - Users may not know that it is possible to choose to see views/likes of a certain category by clicking on the WordCloud graphic.

- Aesthetic and minimalist design (1, Cosmetic problem)
  - The same data it's on both places at the same time, however one gives you a general view and another a more specific one.

Functional prototype

For the functional prototype it was implemented most of the features presented for the low fidelity prototype plus some changes after the user feedback.

A single-page application was the design choice for the web app. All necessary code is loaded to the user's web browser when they first access the app and subsequent interactions with the app are handled dynamically by the browser. This can provide a more seamless and responsive experience. The application makes use of HTML, CSS, Javascript and libraries like D3, SVG and Google.

It starts off with some basic information about the application and some features it offers (Fig 5).

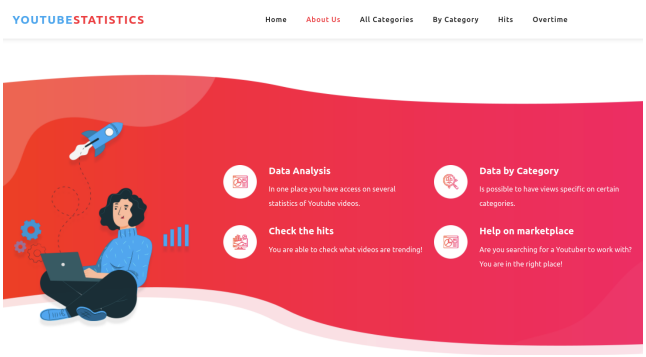


Figure 5: Changes according to students to the "By categories" page

To compare views and likes across all categories, to answer to question 1 and 2, a Bar Chart using D3 and SVG was made. It's possible to toggle between views and likes and even to check on precise numbers for each category represented in the Bar Chart (Fig 6).
tá

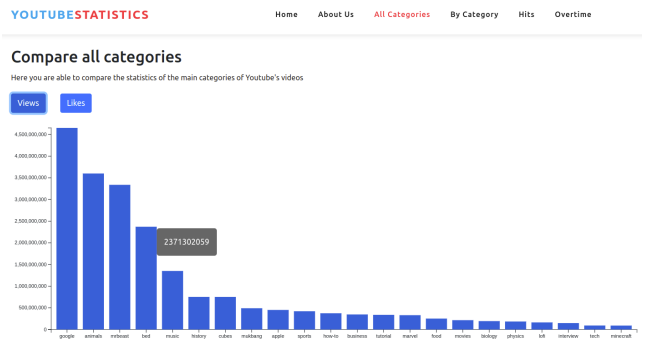


Figure 6: Bar chart of views and likes across all categories

In order to check more detailed information about individual categories we present to the user an interactive WordCloud, a Bar Chart and a dropdown with various options (Fig.7 & 8). We made sure that, when the user hovers a word, the word highlights therefore the user can know that it's clickable.
It's possible to check information on views and likes for specific categories that the user can choose from in the dropdown or WordCloud (this can answer to question 2 and 3). We also allow the user to see specific numbers by hovering the mouse above the bars.

Figure 7: Word Cloud of categories

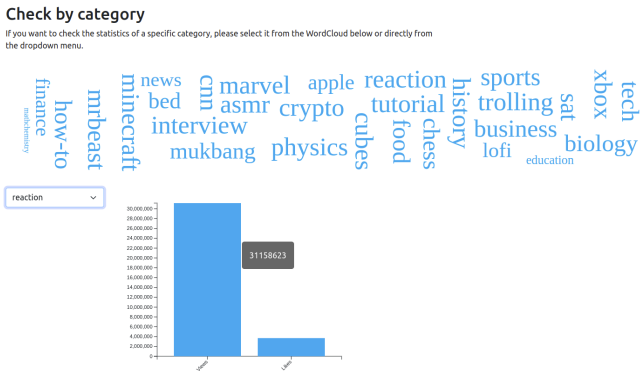
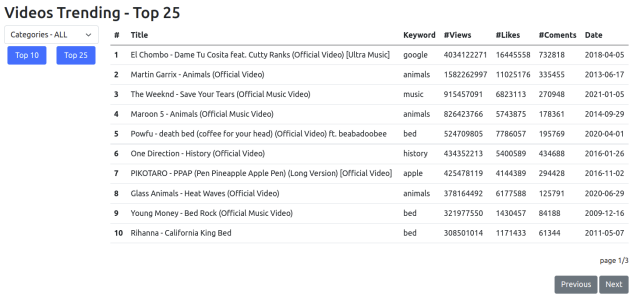


Figure 8: Word Cloud of categories and Bar Chart with views and likes of specific categories

In order to deliver information about trending videos we have developed a top videos table (Fig.9). The user can toggle between top 10 or top 25 and choose to see information about the trending videos across the platform or specific categories. This can answer questions 5 to 7. The data is presented in a paginated table where each video is a clickable link that redirects the user to the Youtube video itself, answering to the question number 5, showing to the user who is the Youtuber who posted that video.



#	Title	Keyword	#Views	#Likes	#Coments	Date
1	El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]	google	4034122271	16445558	732818	2018-04-05
2	Martin Garrix - Animals (Official Video)	animals	1582262997	11025176	335455	2013-06-17
3	The Weeknd - Save Your Tears (Official Music Video)	music	915457091	6823113	270948	2021-01-05
4	Maroon 5 - Animals (Official Music Video)	animals	826423766	5743875	178361	2014-09-29
5	Powfu - death bed (coffee for your head) (Official Video) ft. beabadoobee	bed	524709805	7786057	195769	2020-04-01
6	One Direction - History (Official Video)	history	434352213	5400589	434688	2016-01-26
7	PIKOTARO - PPAP (Pen Pineapple Apple Pen) (Long Version) [Official Video]	apple	425478119	4144389	294428	2016-11-02
8	Glass Animals - Heat Waves (Official Video)	animals	378164492	6177588	125791	2020-06-29
9	Young Money - Bed Rock (Official Music Video)	bed	321977550	1430457	84188	2009-12-16
10	Rihanna - California King Bed	bed	308501014	1171433	61344	2011-05-07

Figure 9: Trending videos Table with top 10 and top 25

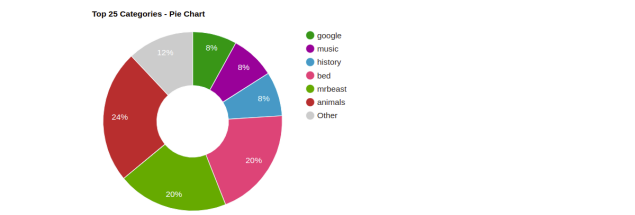


Figure 10:Pie Chart for categories of trending videos

Linked to the top trending videos table there is also a pie chart that, depending on the choice between top 10 or top 25, presents the number of videos that are included in that specific top for each category (Fig 10). This one is meant to answer question number 6.

Implementation challenges

A big implementation challenge was the Overtime feature since we weren't able to showcase the information the way it was intended to be. We managed to get a basic implementation of a Linear Graph using D3.js but the line traced on the graph was not an accurate representation of the data.

Evaluation and changes in the prototype

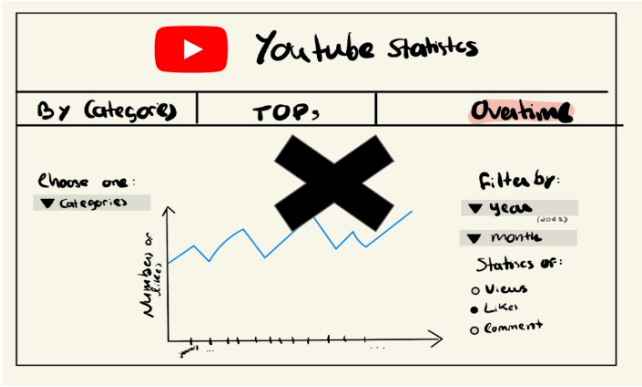


Figure 11:Changes according to our availability.

Due to lack of time and implementation challenges, it was not possible for us to build the “Overtime” page (Fig. 11).

After the functional prototype was made, we asked the same students the same questions before the final presentation, to see if the website was intuitive. Some changes were made:

- Highlight the word cloud and change the mouse pointer, in a way the user can see that it is clickable;
- Highlight the rows in the table to indicate the user that it has a hyperlink.

Conclusion and Future Work

For future work we intend to complete the “Overtime” page since it can offer useful information to the users on sentiment and popularity about categories over a time period.

In conclusion, the Youtube Statistics platform developed in this project provides a useful tool for Youtubers, marketing managers, and curious individuals to better understand and analyze the popularity and metrics of Youtube videos. The platform allows users to view statistics on Youtube by category/keyword and view the top videos on the platform. The use of the D3.js and SVG.js library allows for intuitive and user-friendly visualizations of the data, and the incorporation of user feedback in the design of the platform ensures a high level of usability. Overall, the project successfully achieves its goals and provides a valuable resource for analyzing Youtube statistics.

Links

https://github.com/simaoarrais/VI-Assignment