FIT5147 DATA EXPLORATION AND VISUALISATION

Semester 1, 2019

Assignment

VISUALIZATION PROJECT

YouTubers Analysis – v2.0

How YouTubers make money?

Balaji Mohan

29301092

FIT5147

Table of Contents

Introduction

YouTube is now a new trending business all over the world. YouTube has been converting the social networking users into YouTubers and it even attracts corporate companies to start their own YouTube channels. This is an updated version of my previous "How YouTubers make money" report and here I can call this version as "YouTubers Analysis" since this report details the YouTubers' income and the details of the video. This report could be interesting for the people who watch YouTube very often and they can check whether their favorite channel is listed out in the top ten trending channels.

Design

My initial design was to use a geographical map to select a country for the YouTubers report. Unfortunately, my datasets are only limited for few countries I could not implement the map in my visualization. Hence, I have just updated my design sheets to select a country with a drop-down filter button.

The second design is about the selection of countries to display the available categories in bubble charts. I have used bubble chart here to show the size for the different categories. One can easily figure out the size for the categories of videos in bubble charts.

The third design is about selecting the category to list out the top trending channels in a bar chart. This approach is based on the cross filter and the bar chart keeps on updated when a different country and category is changed.

Displaying the monthly income report for the selected channel is my fourth design. This is just a static outcome which changes corresponding to country, categories, and channels.

The final design is streaming. I have used the streaming concept here to store the datasets for most trended video into an array and display them in a streaming concept. Overall, my design looks like a dashboard where the user can see the YouTubers analysis and how much they are earning with their number of trending videos.

Implementation

Data Wrangling is the first step in my implementation and used python for cleaning the datasets. Wrangling was already discussed well enough in version 1. Next, I have used javascript and D3 to visualize the datasets. My code works well with Bracket and Firefox and I have used D3 v4. Below are the script sources need to be working fine in the HTML page while running my code.

- https://d3js.org/d3.v4.min.js
- https://d3js.org/d3-scale-chromatic.v1.min.js
- http://labratrevenge.com/d3-tip/javascripts/d3.tip.v0.6.3.js

I have used bubble charts, bar charts and a line graph for the overall visualizations.

Why bubble chart?

- A bubble chart is a good way of visualizing the contents in size. I have a demand in my report to showcase the size of categories in size and hence I have implemented bubble size to display the categories in sizes.
- I have added a hover feature for my bubble chart, and it will display the total number of videos uploaded for this category. On mouseover function, I have added text to display the count.
- Also, the bubble charts are acting as a filter for seeing the top ten trending channels for the selected category.
- For filtering, I have added on click function with d3 to link the bar chart.
- The tool I have used here is Scalable Vector Graphic (SVG) for creating the nodes as well as appending it.

Why bar chart?

- A bar chart can be useful while comparing the values for different categories. Here, I have used a bar chart for the same as I want to display the top trending channels and their sizes.
- The bar chart I have implemented here is interactive to operate and this display the income report the selected bars i.e. channel name.
- Just like Bubble chart, I have added hover feature here.
- On mouseover, the bar displays the channel name in full size as the x-axis does not display the full name of the channel due to limited space. Hence displaying the full name on mouseover.
- The on-click feature is also implemented here to link the income report graph for the clicked channel.
- Hence, the user should select the country first, then he has to select the category and then the bar chart displays here for the selected inputs.
- A bar chart is used for streaming as well. This is displayed below the top ten trending channels and income report.
- SVG is used here to create and append the bar chart.

Why line graph?

- I have a line graph here to show the report of the monthly income of a YouTuber channel. This will be easy to analyze the income of a YouTuber.
- Unlike the above charts, this is just a static chart as it is the end of the narration.
- The x-axis displays the timeline of the trending videos found in the datasets.
- Y-axis displays the number of dollars earned by the selected YouTuber.
- SVG is used here to create and display the line graph.

User Guide

Below are the instructions and screenshots for viewing and exploring the **YouTube Analyzer 2.0 – How YouTubers make money?**

The submission folder contains totally 11 files.

  i. 29301092.ipynb
  ii. YT_solution.csv
  iii. 29301092.html
  iv. 29301092.css
  v. 29301092.js
  vi. US_category_id.json
 vii. USvideos.csv
 viii. GB_category_id.json
  ix. GBvideos.csv
  x. IN_category_id.json
  xi. INvideos.csv

First, open jupyter notebook to load all the files except HTML, CSS, js, and YT_solution.csv. The ipynb file will generate the YT_solution.csv after running it.

With the generated YT_solution.csv dataset, create a folder and place all the HTML, CSS, js and YT_solution.csv and run it through Bracket or any available web browser. I prefer Bracket and the code should work with any standard web browser like Firefox.

29301092.html is the index file here and it will display the following the figure.
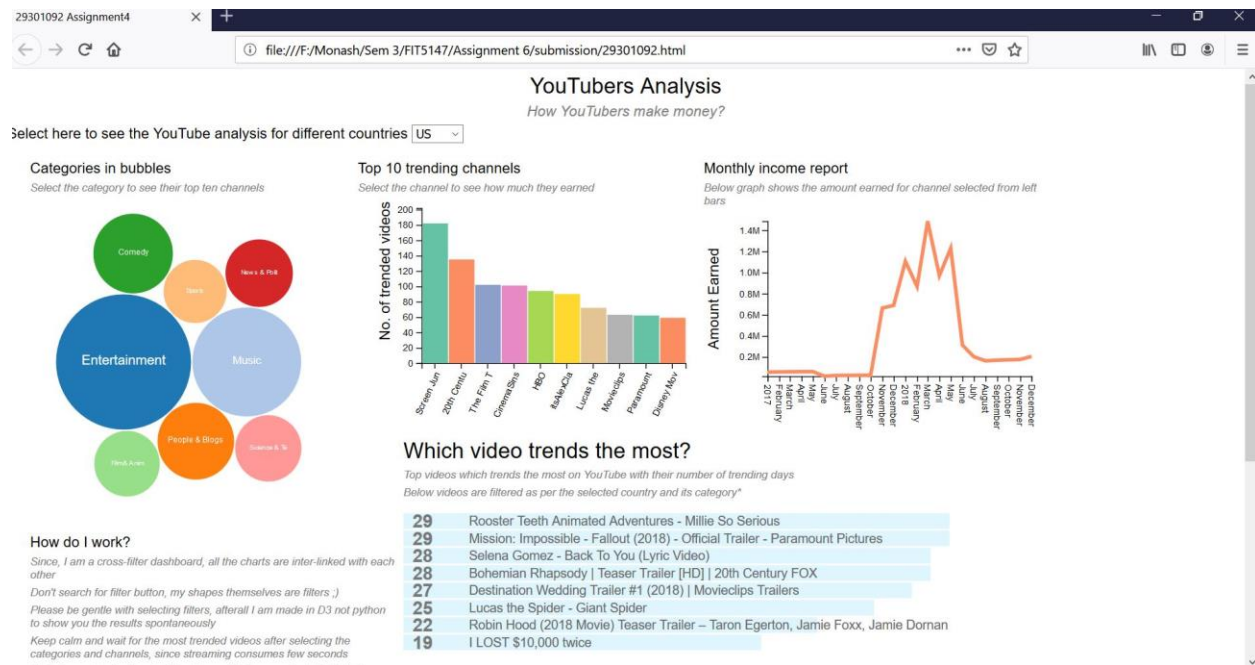
Above is my dashboard design for the datasets which I want to analyze. This design is based on cross filter functions where all the charts are inter-linked with each other.

By default, my design shows the visualization for US country, followed by its film and animation category and the income report for HBO channel. I have provided them as default since the empty figure does not look good.
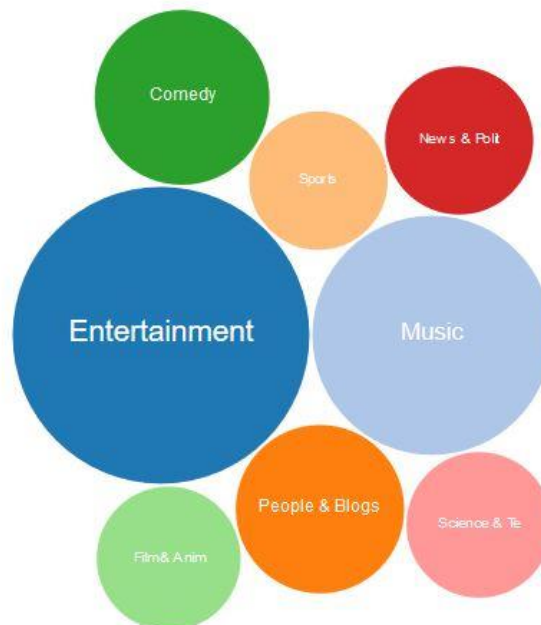
How does it work?

First, the user has to select the country via drop-down button available on the top-left-corner. Remember, the US will be shown by default.

Once the user has selected the country, the bubble chart updates automatically and show the categories in bubble charts like in below figure.
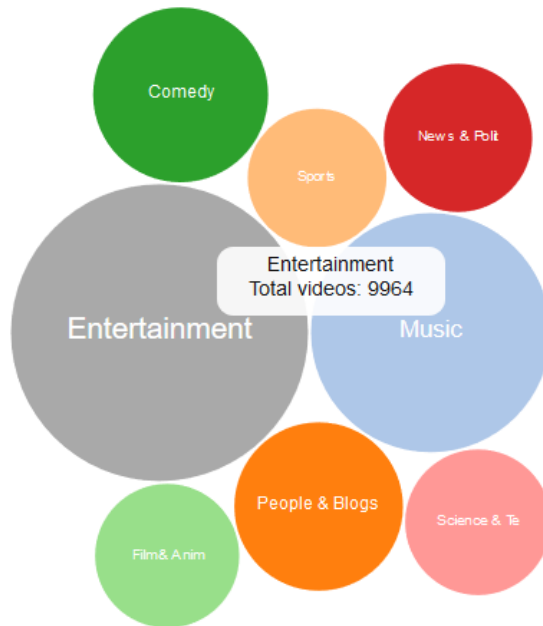

From the above chart, we can see that Entertainment tops the most in trending categories. Yes, the users watch a lot of entertainment videos on YouTube and it induces YouTubers to create many videos in entertainment categories.

Here, I have added hover feature, when the user moves the cursor on the bubble, the value popped up in a box beside to it.
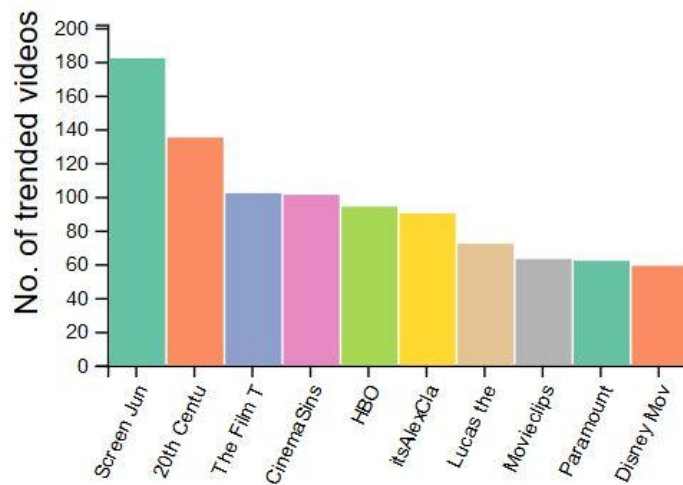
Categories in bubbles

Select the category to see their top ten channels

Comedy

News & Polit

Sports

Entertainment
Total videos: 9964

Entertainment

Music

People & Blogs

Science & Te

Film & Anim

Also, I have created a filter option here. The node itself is a filter when the user clicks the bubble node, the bar chart will be updated accordingly. Say, if the user selects music bubble, the bar chart will be updated to display the top ten trending channels for the selected category. Below figure will be updated in the dashboard when a bubble is clicked.

Top 10 trending channels

Select the channel to see how much they earned

No. of trended videos

200
180
160
140
120
100
80
60
40
20
0

Screen Jun
20th Centu
The Film T
CinemaSins
HBO
itsAlexCla
Lucas the
Movieclips
Paramount
Disney Mov

The bar chart's x-axis and y-axis also will be updated as per the category selection. Since the dataset has so many records and fields, this part was very difficult to achieve in D3. It required a

lot of surfing the blogs on D3 to implement this. As this concept was new to me, I found difficulty in coding with D3 to update the bar chart attributes according to the dataset.

Similar to the above charts, hover feature is added here as well. Below is the figure with the hover option.
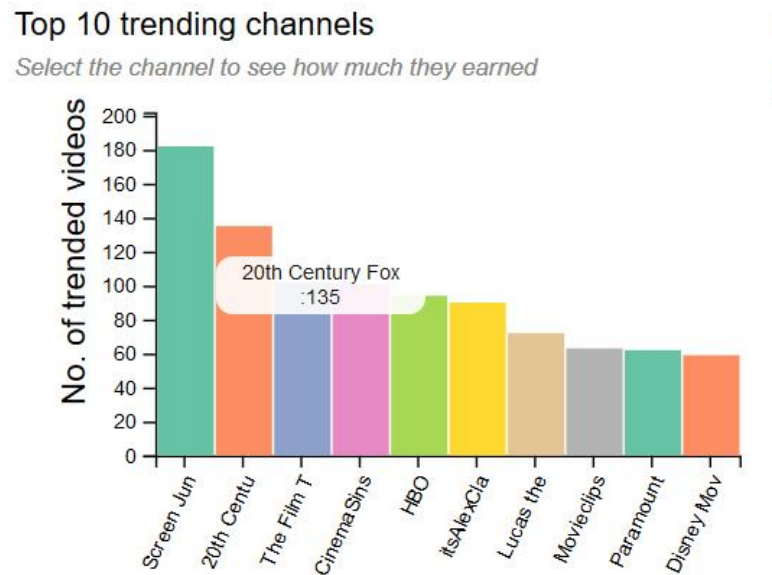
Additionally, I have added one more filter option to this bar chart. When the user clicks the bar, a monthly income report for the clicked channel will be updated right next to it.

Below figure shows the updated line graph for monthly income report for a YouTuber income.
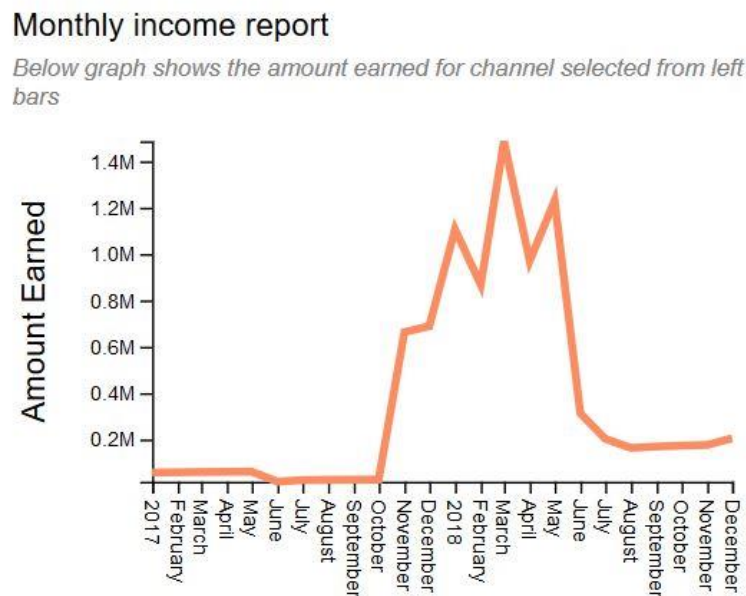
In addition to this, there is one more animated bar graph will be displayed below to these charts. Since it is an animated graph, required a lot of features to visualize it.

I have used streaming pattern here to visualize the streaming data. This bar chart is based on the clicked inputs of country selection and its category.

This stream the data of most trended videos on a particular category like if the user selected music category, it will then display the most trended video in an animated bar chart like below images.

Which video trends the most?

Top videos which trends the most on YouTube with their number of trending days

Below videos are filtered as per the selected country and its category*

29	Mission: Impossible - Fallout (2018) - Official Trailer - Paramount Pictures
29	Rooster Teeth Animated Adventures - Millie So Serious
28	Bohemian Rhapsody | Teaser Trailer [HD] | 20th Century FOX
27	Destination Wedding Trailer #1 (2018) | Movieclips Trailers
25	Lucas the Spider - Giant Spider
22	Robin Hood (2018 Movie) Teaser Trailer – Taron Egerton, Jamie Foxx, Jamie Dornan
18	Christopher Robin Official Trailer
18	Everything Wrong With Thor Ragnarok In 15 Minutes Or Less

Which video trends the most?

Top videos which trends the most on YouTube with their number of trending days

Below videos are filtered as per the selected country and its category*

29	Mission: Impossible - Fallout (2018) - Official Trailer - Paramount Pictures
29	Rooster Teeth Animated Adventures - Millie So Serious
28	Bohemian Rhapsody | Teaser Trailer [HD] | 20th Century FOX
28	Selena Gomez - Back To You (Lyric Video)
27	Destination Wedding Trailer #1 (2018) | Movieclips Trailers
25	Lucas the Spider - Giant Spider
22	Robin Hood (2018 Movie) Teaser Trailer – Taron Egerton, Jamie Foxx, Jamie Dornan

Which video trends the most?

*Top videos which trends the most on YouTube with their number of trending days*
*Below videos are filtered as per the selected country and its category**

29	Rooster Teeth Animated Adventures - Millie So Serious
29	Mission: Impossible - Fallout (2018) - Official Trailer - Paramount Pictures
28	Selena Gomez - Back To You (Lyric Video)
28	Bohemian Rhapsody	Teaser Trailer [HD]	20th Century FOX
27	Destination Wedding Trailer #1 (2018)	Movieclips Trailers
25	Lucas the Spider - Giant Spider
22	Robin Hood (2018 Movie) Teaser Trailer – Taron Egerton, Jamie Foxx, Jamie Dornan
19	I LOST $10,000 twice
18	Everything Wrong With Thor Ragnarok In 15 Minutes Or Less
18	Crow - Simon's Cat	SHORTS
17	Film Theory: Willy Wonka and the Golden Ticket SCAM! (Willy Wonka and the Choco
16	Top breeder 🐈
16	clickbait

As I cannot add a video here, I just captured the bars transitions and attached here. Although it looks like an easy one, it was very hard to implement in D3 since it is inter-linked with multiple category filter options. Most importantly, the bar graph needs to be updated as per the user input selections.

User has to select the filters slowly as it is a cross-filter dashboard design. Each filter is linked with multiple charts and hence the user has to give a few seconds after every filter selection.

Final Output- Each chart is discussed well above. Now the final dashboard looks like this.
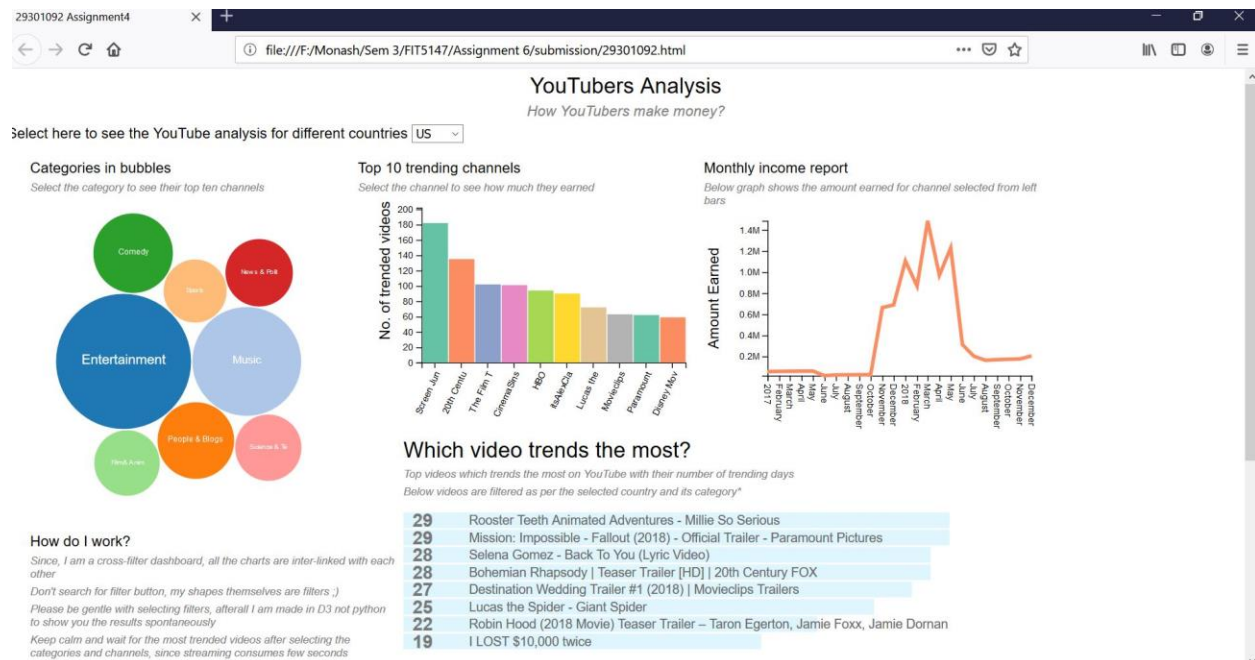
The dashboard is very easy to operate, and anyone can understand this visualization. I have created this report for all types of users even the non-technical person can operate my design and can understand my narration easily.

Most of the factors for "How YouTubers make money?" were discussed previously in my version 1 report and this report is about explaining the things interactively.

I have even added a few instructions in the webpage so that user can understand more about the dashboard and cross-filter design approach.

Thus, the .ipynb file has to run for getting the cleaned dataset and webpage essential files have to be kept in a separate folder to run the HTML file on firefox or bracket.

Testing-This is a very difficult phase in this project. Since the dataset is very huge in size, I had to validate the results with the dataset. Hence, I have smoked test this dashboard and validated that the chart works perfectly for the user input and the result accuracy is obvious.

Difficulty Level- Since it is coded in D3, creating a simple chart itself is a big task and involves a lot of line for simple bubble chart. Moreover, the dataset was also huge in size and it required a lot of conditions to nest the data according to the filter options. Also, the design is based on cross-filter, the function for a chart has to be called from the correct positions. SVG and its positions were really difficult to position in D3 and additionally, CSS skills were also required for this project. Overall, it was a great experience to work with this project and it was really interesting to create this visualization dashboard.

Conclusion

Thus, this project could be really useful for entrepreneurs who think about becoming a YouTuber. This explains how the YouTubers are earning money and can easily identify what the YouTubers are targeting. As seen from the results, YouTubers prefer entertainment category as this category videos are high in numbers comparing to others. More the videos in trending, more the YouTubers get paid, and also other factors were discussed in version 1 report for how YouTubers make money and this version was more concentrated on their income and the details of their video.

Reflection

With this report, I have learned what is D3 and how good it is to visualize the data. I am really familiar with python and R for visualization. I always worked with python or R if I need any visualization for analyzing the datasets. However, after working with D3, I feel D3 is better than python as well as R. Since D3 allows me to implement a lot of features while visualizing the data.

Unfortunately, I could not implement the map feature in this assignment as the dataset is limited to only a few countries. I even tried to add map here, but it was looking unnecessary for this dashboard.

Thus, I would recommend D3 for any detailed and interactive visualization. Although it was very difficult to learn D3 and javascript being a pro python/R user. Nevertheless, I have learned most

9

of the features available in D3 and implemented in my project and explored them. Most importantly, many online blogs have helped me out to learn D3 features and their usages.

References

https://www.kaggle.com/datasnaek/youtube-new/kernels

https://amber.rbind.io/2017/05/02/nesting/

https://www.d3-graph-gallery.com/graph/line_filter.html

https://www.d3-graph-gallery.com/graph/line_several_group.html

http://www.d3noob.org/2014/02/grouping-and-summing-data-using-d3nest.html

https://bl.ocks.org/d3noob/23e42c8f67210ac6c678db2cd07a747e

https://bl.ocks.org/alokkshukla/3d6be4be0ef9f6977ec6718b2916d168

https://www.d3-graph-gallery.com/graph/bubblemap_buttonControl.html

https://www.d3-graph-gallery.com/graph/barplot_animation_start.html
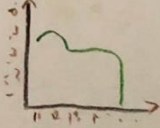
http://bl.ocks.org/charlesdguthrie/11356441

Appendix

Sheet 2 : Initial Design 1

Layout

categories

Hover

total...

↓ clicks

A B C
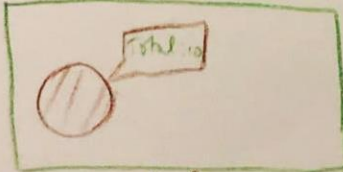
Operation

- When user clicks the bubble, new bar graph will be displayed
- When mouseover, it displays the total number of videos
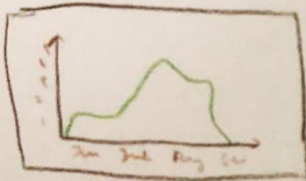
Positives
- Easy to implement
- Easy to understand

Negatives
- Very limited

Sheet 3: Initial Design 2

Layout

→ Line graph for annual income report

↓

It's a static visualization from the clicked channel for the annual report of income

Operation:

- User can view the annual income report with this chart
- this chart changes for every filter selection of country, categories and channels.

Positives.
- Easy to understand the report

Negatives
- Difficult to link with previous charts to show the corresponding values in the selected options

13

Sheet 4 : Initial Design 3

Top :-
10 Trending videos in the
selected category

| I hate you - Cardi B |
| Offset - Cardib |
| More mm |
| How I met you |
| Hamilton |

Animated bar
chart for
streaming

Positives :-

- Rich visualization
- Textual one and easy
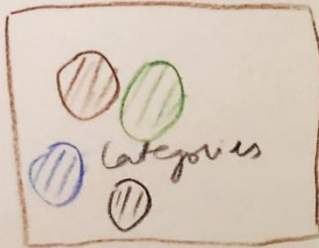 to understand
- Can see the video name
  easily

Negatives :-
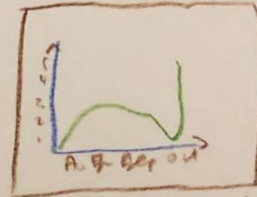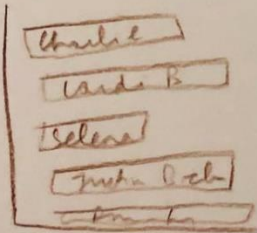
- Very hard to implement

Sheet 5 - Realization