FIT5147 DATA EXPLORATION AND VISUALIZATION

Narrative Visualisation Project

“Trending Youtube Videos Analysis”

Name: Gayatri Aniruddha
Student ID: 30945305
Lab Number : 22
Tutor Name : Pratik Bhumkar

TABLE OF CONTENTS:

Introduction

My dataset was a collection of the statistics of the trending youtube videos. The highlights of my dataset were publishing times, trending dates, likes, dislikes, comments, views and video titles. After performing a lot of data cleaning and wrangling on the dataset, I was able to derive certain attributes such as publishing hour, publishing day, publishing time, video category and video title length.

Intended Audience:
My main intended audience of this narrative visualisation were the aspiring young youtubers who want to establish themselves in the industry and ensure that they reach out to a larger number of people through their work. When it comes to the entertainment industry, it's not only about how much you know, it's also about how well you are able to showcase the stuff you know!

Now, if there is someone out there who has really good content that can make a difference in people's lives, it's really important that he reaches out to people at a time when they are in a position to absorb the knowledge. It is quite vital that his video reaches to a larger audience as well. Furthermore, his video should be attractive to people.

Message my visualisation wants to convey:
Thus, after working on my dataset, I imagined myself as a young, aspiring youtuber. I could narrow down my questions about a video release into four main categories.

- When should I release a video in order to reach a larger audience?
- What time on that day should I release it?
- How long should the Video Title be so that it catches people's attention?
- Which type of video should it be so that I can hold onto people's time

This is exactly what my narrative visualisation is going to convey! There is enough data given about all the contributing factors. The user can choose the different conditions and then can take a call on the day, time and category his video should be in for it to become a trend!

Design

Description of the Visualisation Design Process:
Summary of the Five Design Sheet:
Five Design Sheet 1 : Brainstorming

In order to give the user as much information as possible regarding the different contributing factors that will enable him to release a trending video, I took the following steps :

- First, I pondered and brainstormed. At this stage, I came up with a lot of ideas regarding the things that will be of interest to a user.
    - Which Day?
    - Which Hour?
    - Video Caption Length?
    - Which Category?
    - Which Channel?
    - Video Caption words he is looking for?
    - Country of origin of the video?
- Then, I came up with the different count measures that a user would be interested in, in order to analyze these contributing factors. They are:
    - Video count
    - Likes count
    - Dislikes count
    - Comment count
    - Views count
- Then, I listed out the different possible ways of making this data attractive and informative to the user at the same time. Thus, possible useful visualizations were :
    - Displaying the related Country Map
    - Bar Chart representation
    - Histogram analysis
    - Motion Charts
    - Pie Charts
    - Line Graphs
- I then **filtered** and **categorized** my ideas based on the following:
    - The visualizations can be based on my research questions
    - The division can be based on the country
    - The graphs can be based on the factors the user is interested to display
    - The final analysis should be about the details the user wants to view
- Finally, I finally **combined and refined** my ideas and came up with three main visualisations to support my analysis namely :
    - Bar Charts to explore the relationship between publishing day and publishing hour.
    - Pie Charts to show a relationship between the video caption length and popularity.
    - Bubble/Dot Plots to give a measure of the video category and trending videos.

Five Design Sheet 2 : Layout 1 : Bar Plots

- Here, this deisgn was to show the relationship between the trending videos with their publishing days and hours.
- The user gets to choose the publishing day and hour that interests him in order to get a complete visualization of the videos pertaining to those days and hours.
- Here, the main focus is on the "height" of the Bar Charts.
- This was a simple design, which is easy to read, analyze and interpret.
- However, the only drawback is that this design does not provide other details.

Five Design Sheet 3 : Layout 2 : Pie Charts

- Here, this deisgn was to show the relationship between the trending videos with their video title lengths.
- The user gets to choose the minimum and maximum number of characters in the video title length in order to get a complete visualization of the videos pertaining to those many numbers of characters.
- Here, the main focus is on the "sector" of the Pie Charts. The thickness of each sector is an indication of the popularity of the videos.
- This design provides a good overview of how quickly the video length can capture a person's attention!

Five Design Sheet 4 : Layout 3 : Bubble Plots

- Here, this deisgn was to show the relationship between the trending videos with their corresponding categories.
- The user gets to choose the relevant category in order to get a complete visualization of the videos pertaining to those specific categories.
- Here, the main focus is on the "dots" of the Bubble Plots. The height and positioning of each sector is an indication of the popularity of the videos.
- This design is really easy to interpret which will help the user in analyzing the patterns and trends of the popular videos.
- However, again it does not do that great a job as it fails to cover other details.

Five Design Sheet 5 : Realization Sheet

- This sheet gave a sneak peek of what the final interactive narrative visualisation would look like.
- Here, I have merged the ideas from all the three layouts in order to come with a final image of the layout which will be presented to the user.
- Thus, in order to give the user as much information as possible, we ask the user the day/hour/category/video length he is interested in.
- Then, we plot all 3 visualizations in the same page, giving the user a complete overview of the factors influencing a youtube video!

Slight Modifications to the Realization Sheet :

- While making my final user interactive system, I realised that there was no point asking the user for a particular channel. This was because there are millions of channels out there. More than analyzing the data for a given channel, it makes much more sense to analyze the channel category. Hence, in my final interactive visualisation, I have removed this part of asking the user for a particular channel as this information is repetitive.

- Also, after analysing the dataset further, I have come to the conclusion that the "**number of videos"** is the main count measure that easily distinguishes between the different youtube videos.
- Thus, I have only removed extra repetitive information from my implementation and clearly my final implementation provides the same information and visualisations as my initial proposal.

Justification of final design in terms of the human perceptual system and human communication assumptions :

- My design is completely justified in terms of both the human perception and common human communication assumptions.
- We humans can easily process visual information. We can efficiently identify and differentiate the various visualisations in terms of colors, sizes and shapes etc,
- Also, visualisations which are located in different positions are considered to be different according to our brain. (Knauff, 2013)
- My design has different visualisations :
  - They have different **colors, sizes and shapes.**
  - The different plots and graphs are also located in **different parts of the screen.**
- Hence, the user will be able to clearly identify the factors and analyze its influence and effects clearly in accordance with the common human perceptions and assumptions.

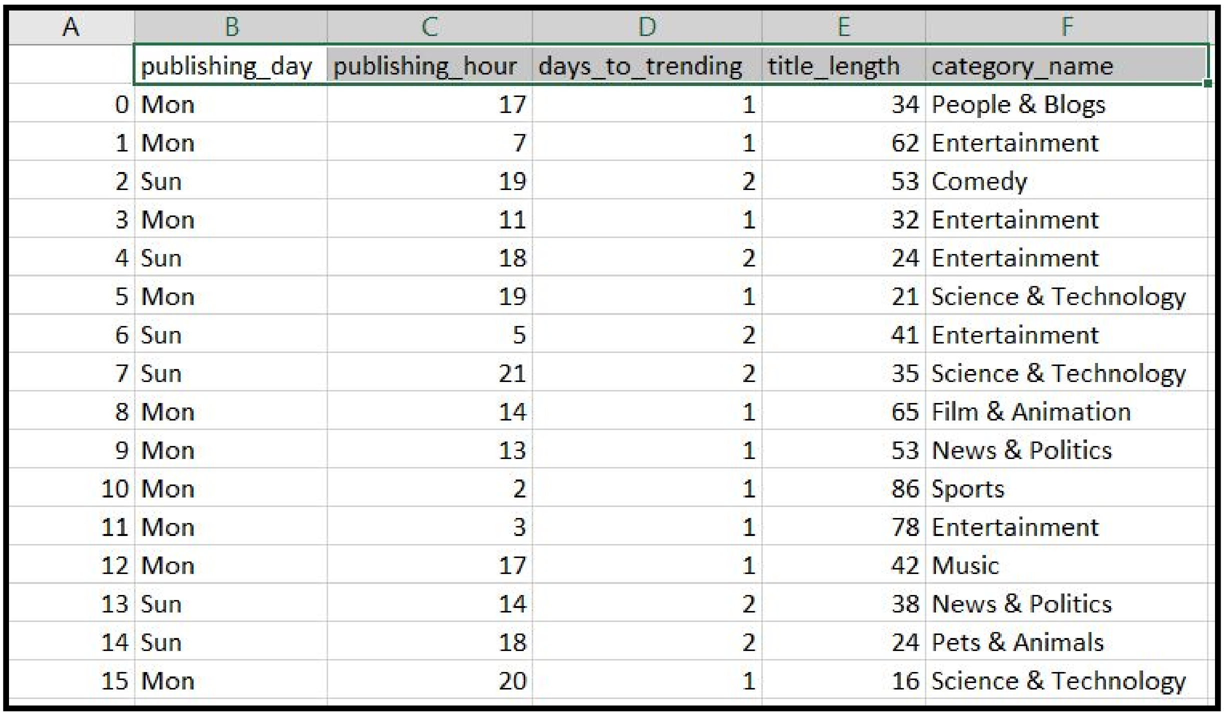
Implementation

Data used for implementation :
The youtube data was taken from : https://www.kaggle.com/datasnaek/youtube

Data Cleaning and Wrangling :

- In my original retrieved data, there were around **22 Columns** after the initial data cleaning.
- Now, in order to make the interactive narrative visualisation, I was using only **5 Relevant Columns.**
- Hence, in my **FINAL FORMATTED DATASET,** I removed all the other columns and have kept only these 5 columns.

Glimpse of how the Final Formatted Dataset for the Narrative Visualization:

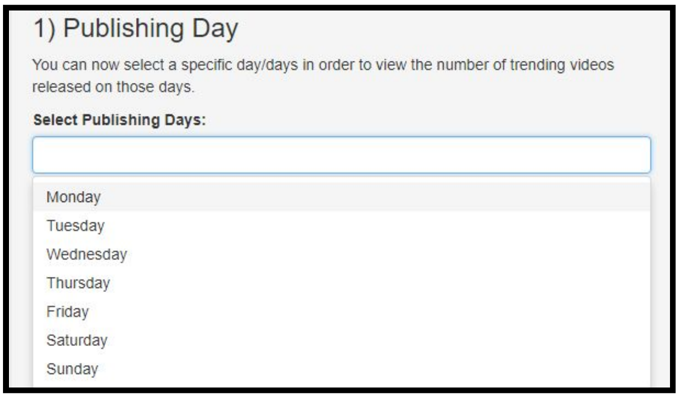


A	B	C	D	E	F
	publishing_day	publishing_hour	days_to_trending	title_length	category_name
0	Mon	17	1	34	People & Blogs
1	Mon	7	1	62	Entertainment
2	Sun	19	2	53	Comedy
3	Mon	11	1	32	Entertainment
4	Sun	18	2	24	Entertainment
5	Mon	19	1	21	Science & Technology
6	Sun	5	2	41	Entertainment
7	Sun	21	2	35	Science & Technology
8	Mon	14	1	65	Film & Animation
9	Mon	13	1	53	News & Politics
10	Mon	2	1	86	Sports
11	Mon	3	1	78	Entertainment
12	Mon	17	1	42	Music
13	Sun	14	2	38	News & Politics
14	Sun	18	2	24	Pets & Animals
15	Mon	20	1	16	Science & Technology

High Level Description of the implementation:

- Here, my interactive visualisation has been completely implemented in R Shiny.
- My Shiny Application is mainly divided into two components :
  - Side Panel
  - Main Panel
- **Side Panel** : ***Taking INPUTS from the USER***
  - This panel is for taking inputs of the factors from the user.
  - The user can play around with these factors in order to view the changing patterns of the data output.

- User has to enter the following details:
  - **Publishing Day :**
    - Users can either choose a single day or multiple days.
    - **Implementation :** Using a Drop Down Box

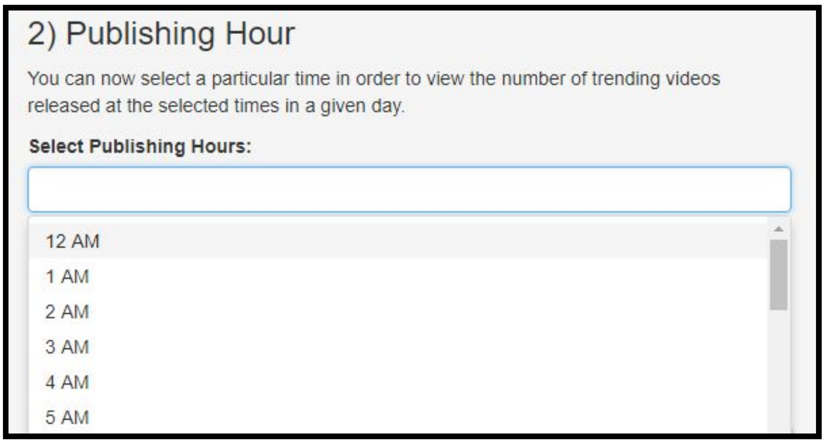


1) Publishing Day

You can now select a specific day/days in order to view the number of trending videos released on those days.

Select Publishing Days:

  - **Publishing Hour :**
    - Users can choose sin single or multiple times of the day for viewing the video statistics.
    - **Implementation :** Using a Drop Down Box

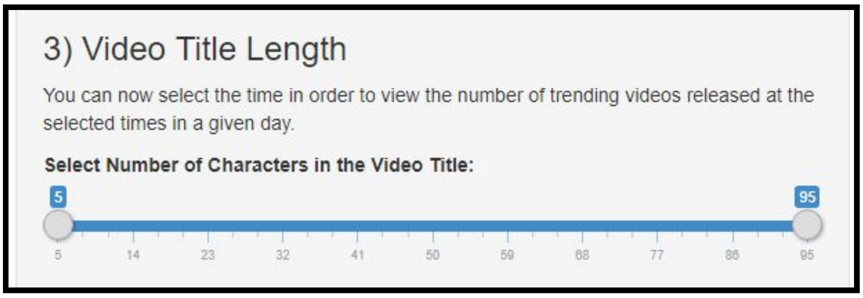


2) Publishing Hour

You can now select a particular time in order to view the number of trending videos released at the selected times in a given day.

Select Publishing Hours:

  - **Video Title Length :**
    - Users can choose the minimum and maximum number of characters they wish to see in the video title length.
    - **Implementation :** Using a Slide Input

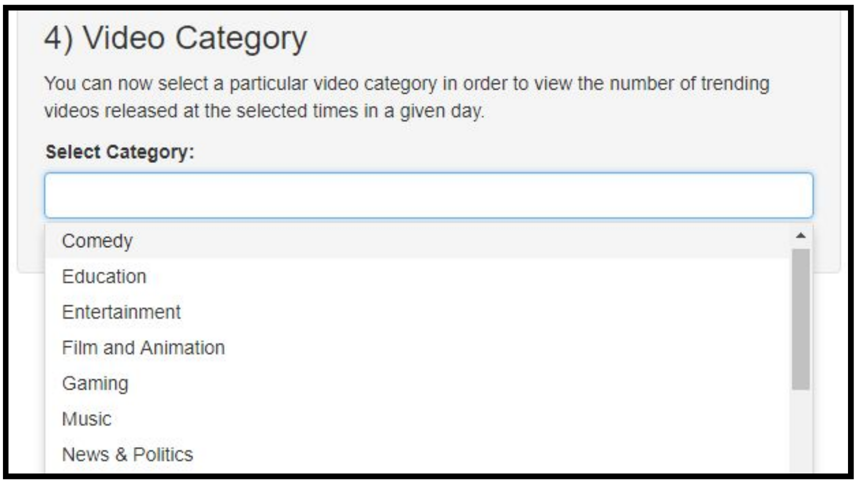


3) Video Title Length

You can now select the time in order to view the number of trending videos released at the selected times in a given day.

Select Number of Characters in the Video Title:

        - **Video Category :**
            - Users get to select the different categories of videos they are interested in to observe the statistics.
            - **Implementation :** Using a Drop Down Box

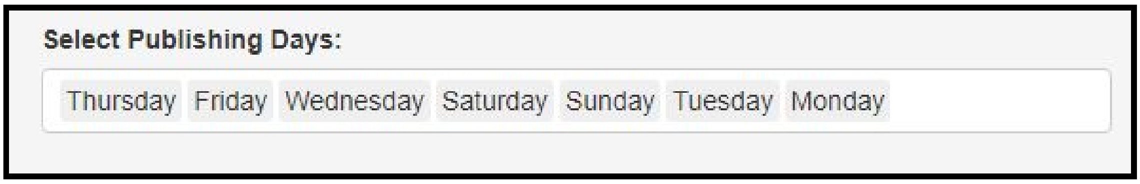


4) Video Category

You can now select a particular video category in order to view the number of trending videos released at the selected times in a given day.

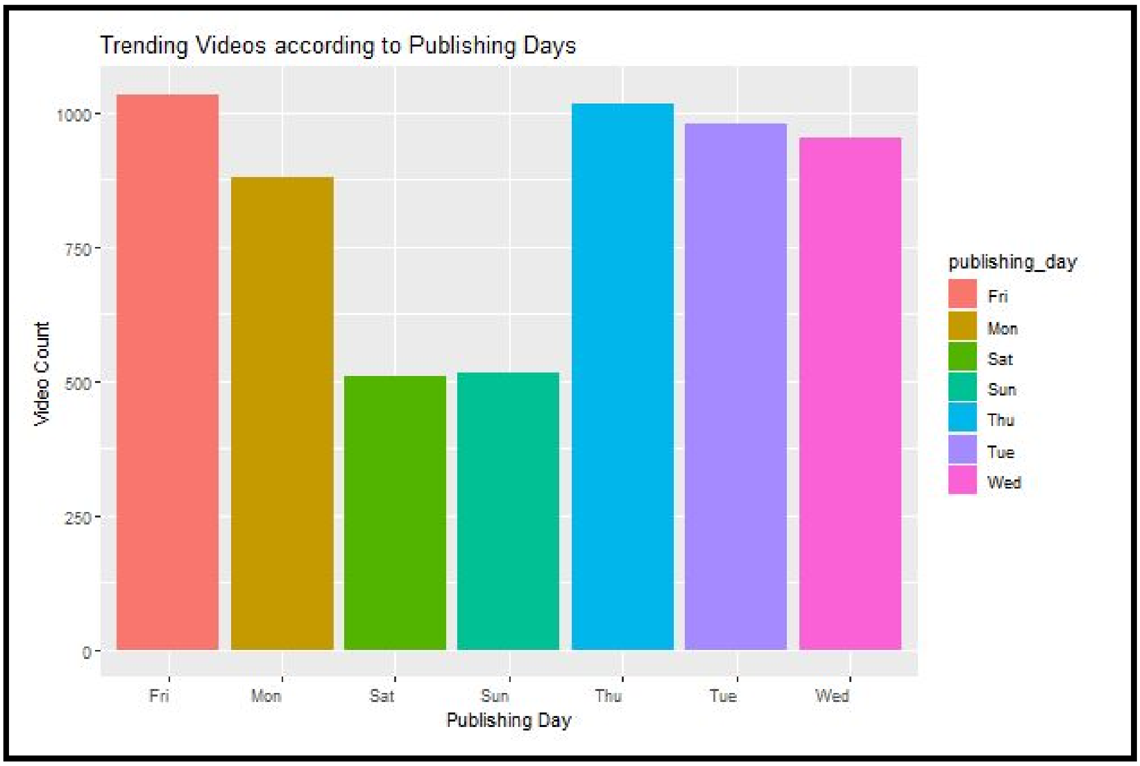
Select Category:

- **Main Panel** : ***Presenting OUTPUTS for the USER***
    - This panel is for displaying all the visualisations.
    - I have divided this dashboard into four equal parts.
        - The **top left part** is for **Bar Plot Visualization.**
            - **Implementation :** Here, the different **publishing days** are marked with different colours.
            - These different colors can be identified and mapped to the publishing days with the help of the legend provided.

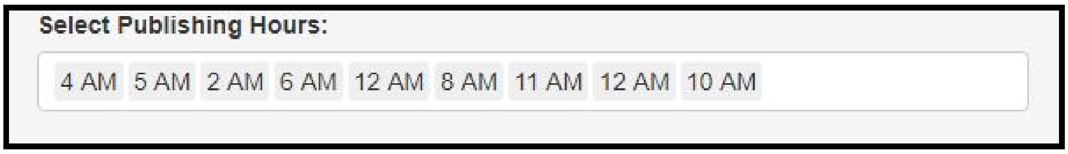


Select Publishing Days:

Thursday Friday Wednesday Saturday Sunday Tuesday Monday

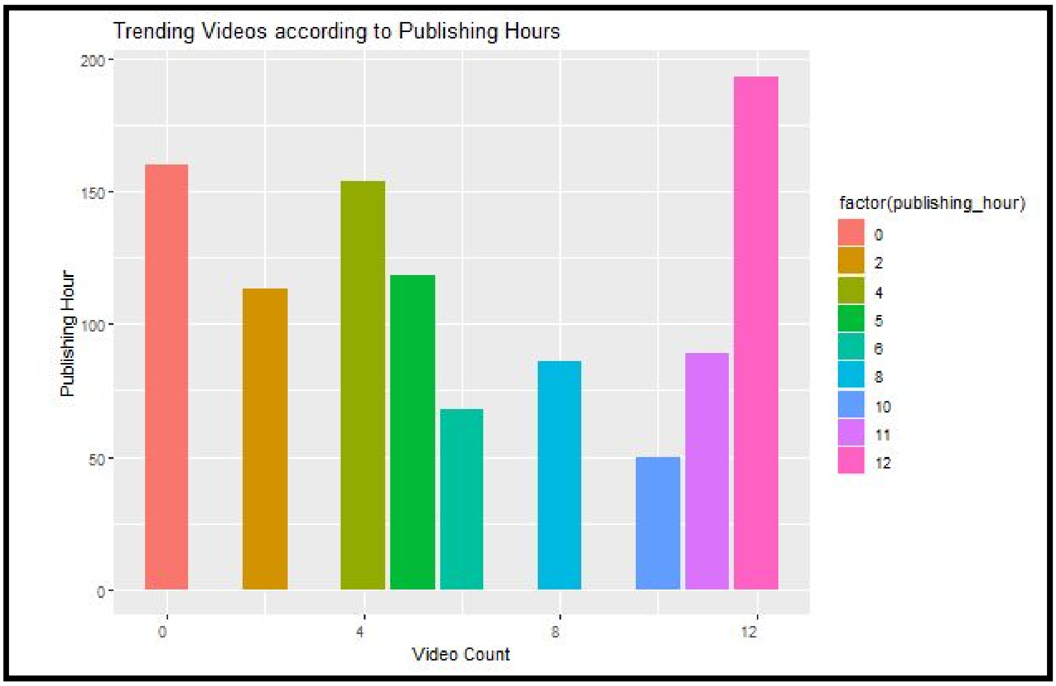


- The **top right part** is for **Bar Plot Visualization.**
    - **Implementation :** Here, the different publishing hours are marked with different colours.
    - These different colors can be identified and mapped to the publishing days with the help of the legend provided.

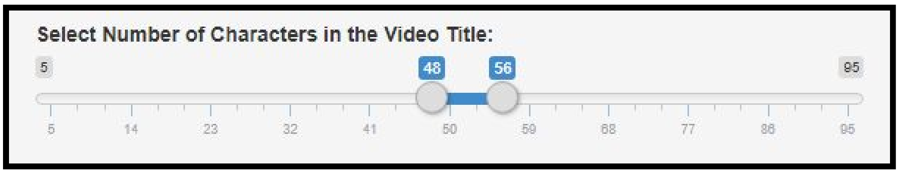


Select Publishing Hours:

4 AM 5 AM 2 AM 6 AM 12 AM 8 AM 11 AM 12 AM 10 AM

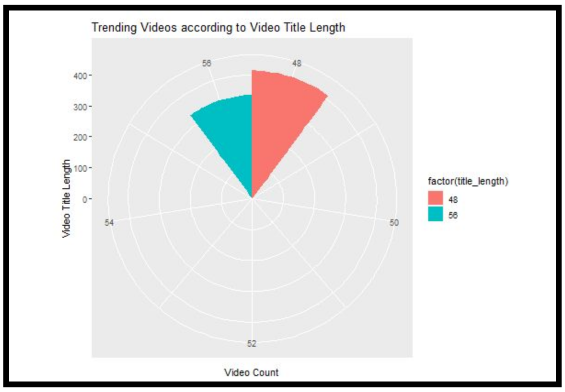


- The **bottom left** one is for the **Pie Chart.**
  - **Implementation :** Here, each sector represents the total number of videos having that many characters in their Video Title.
  - The corresponding exact number of words can be identified through the legend provided.

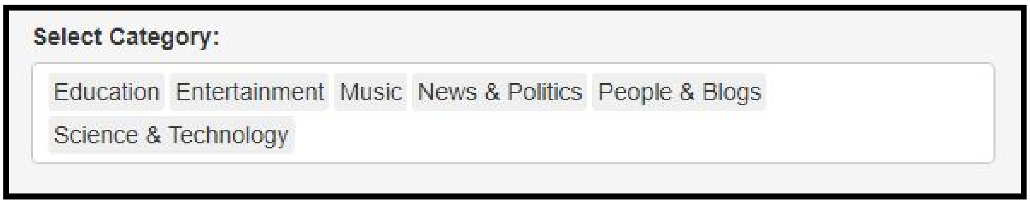


Select Number of Characters in the Video Title:

5 48 56 95

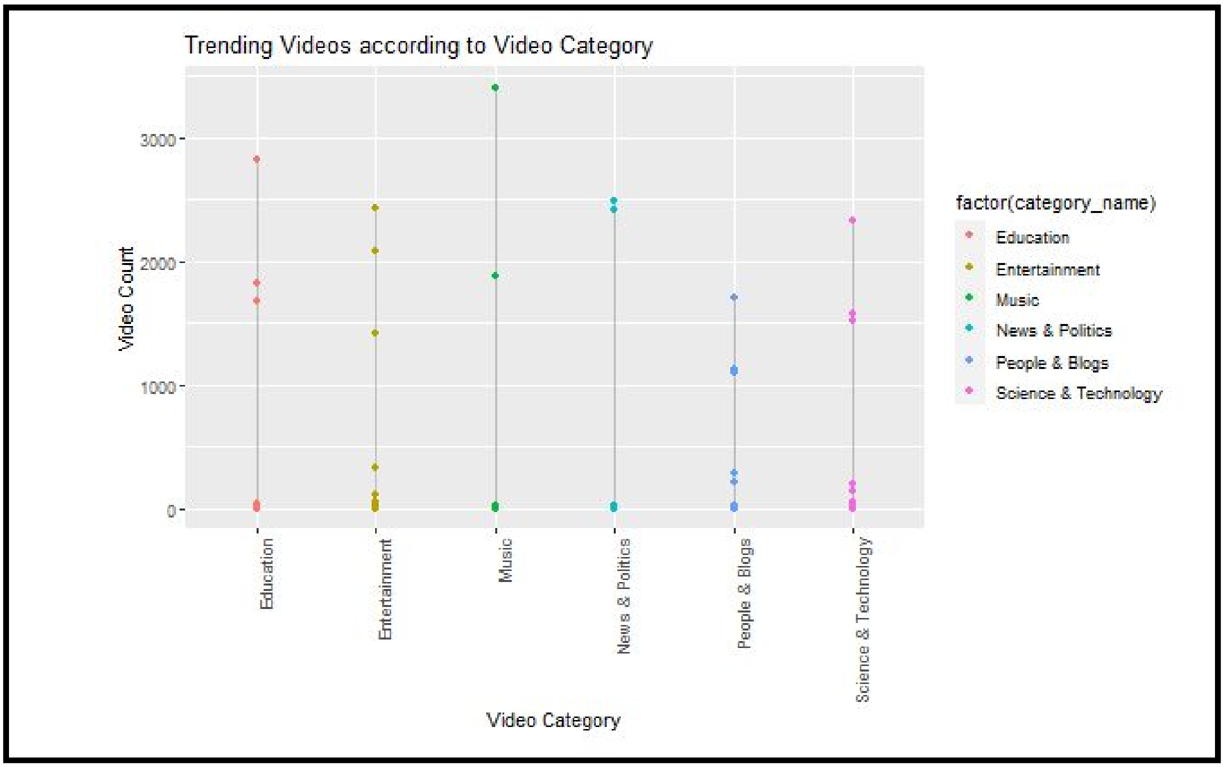


- The **bottom right** one is for the **Bubble Plot.**
  - **Implementation :** Here, each dot represents a given category.
  - We can map the colors to the various categories by observing the legend.



Select Category:

Education Entertainment Music News & Politics People & Blogs Science & Technology



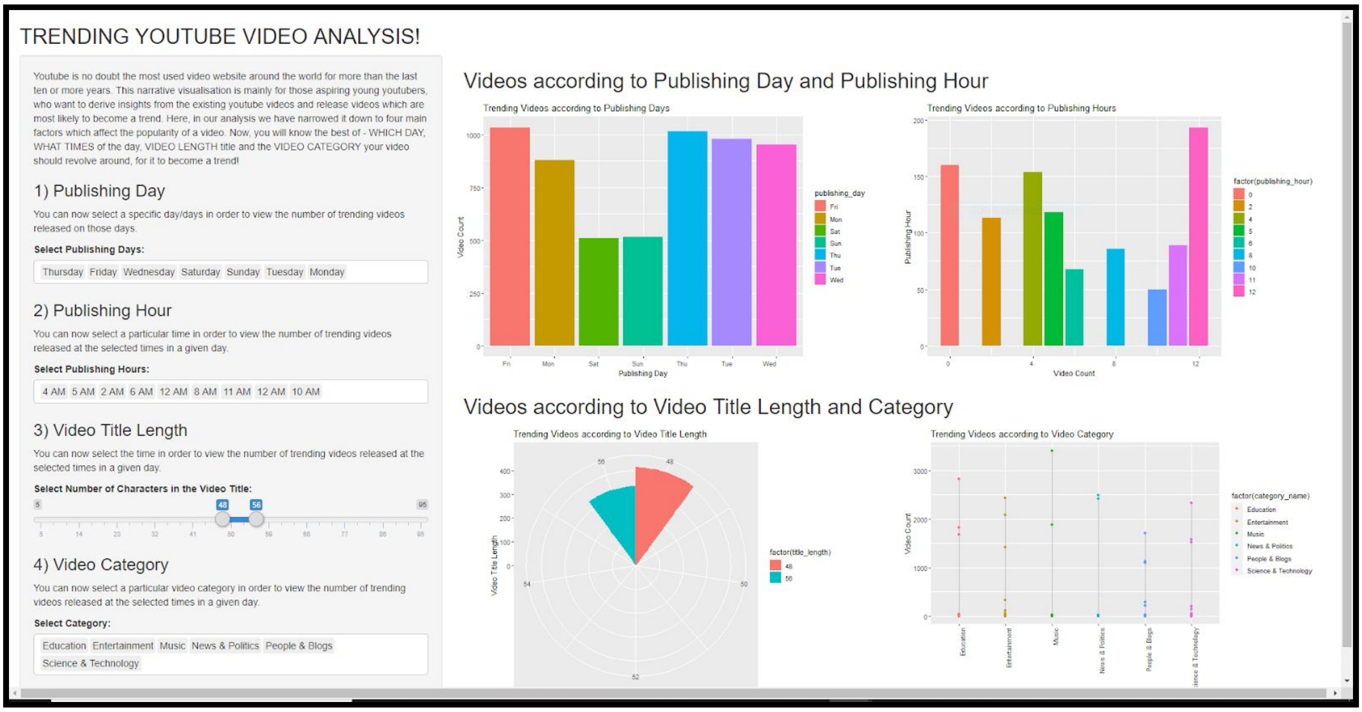
Libraries Used and Why:

- **library(shiny)**
  - Shiny Package was used to create the web application and make it interactive using R.
  - Shiny application helps us to make the interaction “LIVE”.
  - Here, the outputs are directly impacted by the inputs. There is no need for any reloading of the browser as most changes are real time.

- **library(ggplot2)**
  - This package was used to create meaningful and elegant visualizations.
  - The ggplot2 package was used to plot Bar Plots and Bubble Plots.

- **library(gridExtra)**
  - This package was used in order to split the Main Panel into four parts.
  - Here, I have split the dashboard into two rows and every row is split into two boxes based on a factor of 50%.
  - That is how we get working areas to plot our visualisations.

Final Narrative Interactive Visualisation:



TRENDING YOUTUBE VIDEO ANALYSIS!

Youtube is no doubt the most used video website around the world for more than the last ten or more years. This narrative visualisation is mainly for those aspiring young youtubers, who want to derive insights from the existing youtube videos and release videos which are most likely to become a trend. Here, in our analysis we have narrowed it down to four main factors which affect the popularity of a video. Now, you will know the best of - WHICH DAY, WHAT TIMES of the day, VIDEO LENGTH title and the VIDEO CATEGORY your video should revolve around, for it to become a trend!

1) Publishing Day

You can now select a specific day/days in order to view the number of trending videos released on those days.

Select Publishing Days: Thursday Friday Wednesday Saturday Sunday Tuesday Monday

2) Publishing Hour

You can now select a particular time in order to view the number of trending videos released at the selected times in a given day.

Select Publishing Hours: 4 AM 5 AM 2 AM 6 AM 12 AM 8 AM 11 AM 12 AM 10 AM

3) Video Title Length

You can now select the time in order to view the number of trending videos released at the selected times in a given day.

Select Number of Characters in the Video Title: 48 56

4) Video Category

You can now select a particular video category in order to view the number of trending videos released at the selected times in a given day.

Select Category: Education Entertainment Music News & Politics People & Blogs Science & Technology

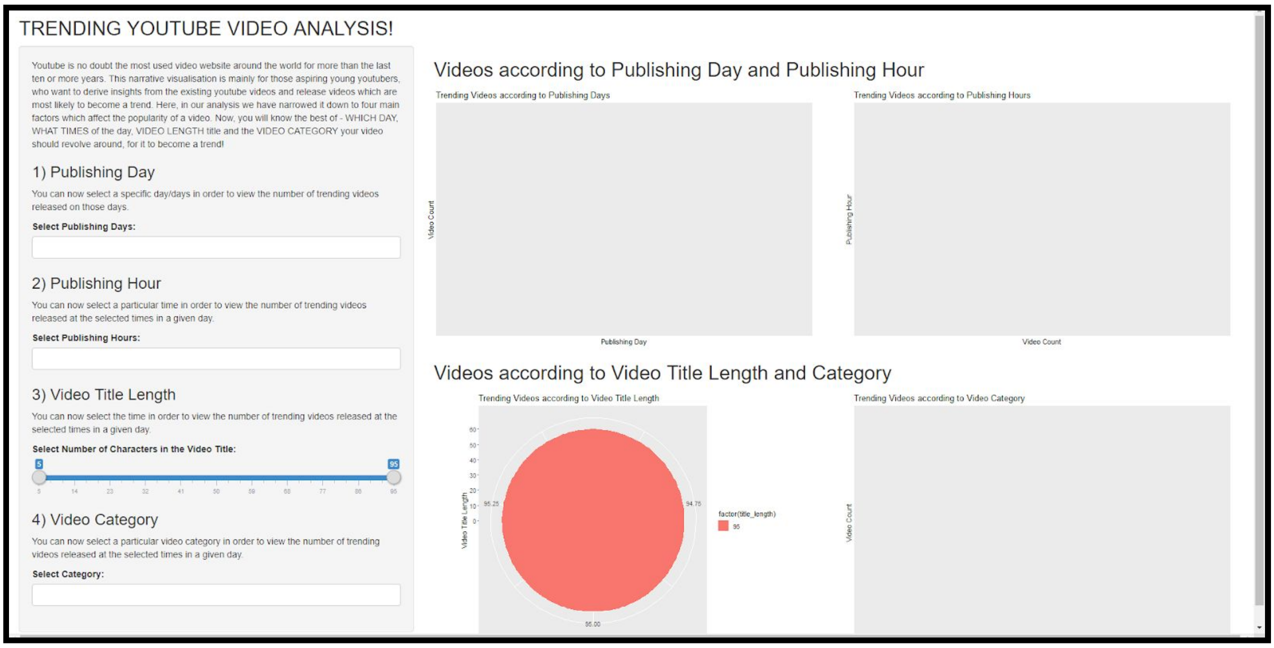
Videos according to Publishing Day and Publishing Hour

Videos according to Video Title Length and Category

User Guide

Instructions for viewing :

- The entire code has been implemented in R.
- Hence, in order to view the application, please open the R file in R Studio.
- Set the working directory as the same folder in which the application has been saved.
- Here, I have used 3 main packages - shiny, ggplot2 and gridExtra. Hence, please install these packages before running the shiny application.
- Here, the data for this interactive visualization is from the csv file : youtube_formatted_data. Please ensure that this csv file is in the same folder as that of the R Shiny Application.
- Now, once you click on "**Run-App"** in he shiny application, the following screen should appear :



TRENDING YOUTUBE VIDEO ANALYSIS!

Youtube is no doubt the most used video website around the world for more than the last ten or more years. This narrative visualisation is mainly for those aspiring young youtubers, who want to derive insights from the existing youtube videos and release videos which are most likely to become a trend. Here, in our analysis we have narrowed it down to four main factors which affect the popularity of a video. Now, you will know the best of - WHICH DAY, WHAT TIMES of the day, VIDEO LENGTH title and the VIDEO CATEGORY your video should revolve around, for it to become a trend!

1) Publishing Day

You can now select a specific day/days in order to view the number of trending videos released on those days.

Select Publishing Days:

2) Publishing Hour

You can now select a particular time in order to view the number of trending videos released at the selected times in a given day.

Select Publishing Hours:

3) Video Title Length

You can now select the time in order to view the number of trending videos released at the selected times in a given day.

Select Number of Characters in the Video Title: 5 – 95

4) Video Category

You can now select a particular video category in order to view the number of trending videos released at the selected times in a given day.

Select Category:

Videos according to Publishing Day and Publishing Hour

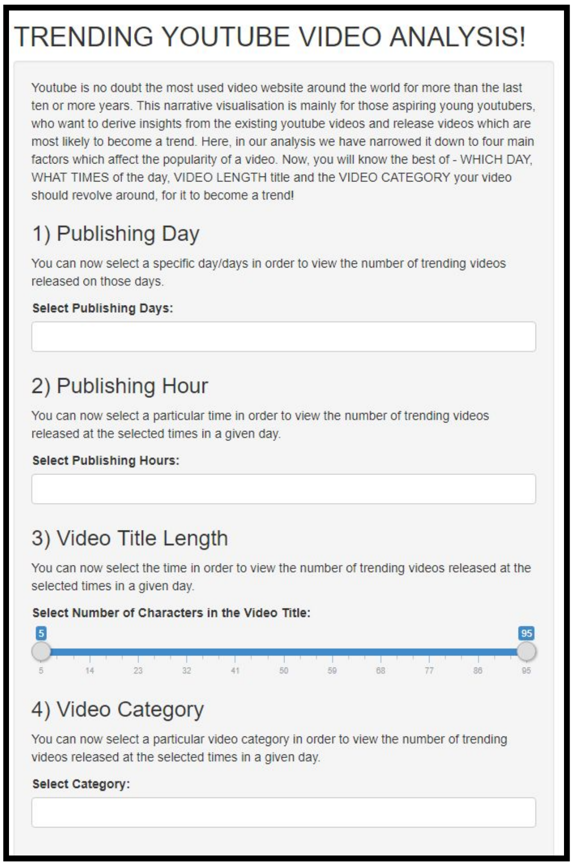
Videos according to Video Title Length and Category

5
14
23
32
41
50
59
68
77
86
95
Trending Videos according to Publishing Days
Video Count
Publishing Day
Trending Videos according to Publishing Hours
Publishing Hour
Video Count
Trending Videos according to Video Title Length
Video Title Length
60
50
40
30
20
10
0
95.25
94.75
95.00
factor(title_length)
🟥 95
Trending Videos according to Video Category
Video Count

- The User can now click on the various options given in the side panel in order to view the various Bar Chart, Pie Chart and Bubble Plot Visualizations.

Instructions for exploring your narrative visualisation :

- This application is mainly for the young, aspiring content creators of youtube!
- If you guys have ever wondered -
    - What should I do to make a difference in youtube?
    - How can I expand my target audience?
    - Is my content attracting sufficient attention?
    - Are my Video Captions too lengthy/too short?
- Then, you have opened the perfect application!
- All that you have to do is, select the options given below and all your doubts would be answered.



TRENDING YOUTUBE VIDEO ANALYSIS!

Youtube is no doubt the most used video website around the world for more than the last ten or more years. This narrative visualisation is mainly for those aspiring young youtubers, who want to derive insights from the existing youtube videos and release videos which are most likely to become a trend. Here, in our analysis we have narrowed it down to four main factors which affect the popularity of a video. Now, you will know the best of - WHICH DAY, WHAT TIMES of the day, VIDEO LENGTH title and the VIDEO CATEGORY your video should revolve around, for it to become a trend!

1) Publishing Day

You can now select a specific day/days in order to view the number of trending videos released on those days.

Select Publishing Days:

2) Publishing Hour

You can now select a particular time in order to view the number of trending videos released at the selected times in a given day.

Select Publishing Hours:

3) Video Title Length

You can now select the time in order to view the number of trending videos released at the selected times in a given day.

Select Number of Characters in the Video Title:

4) Video Category

You can now select a particular video category in order to view the number of trending videos released at the selected times in a given day.

Select Category:

- You can see from the side panel that you have four options to choose from -
  - Publishing Day
  - Publishing Hour
  - Video Title Length
  - Video Category
- Select a day, a time, length and a category and you will get an insight into what is actually happening and making these videos a trend!
- *HINT* : To observe the best variations, you may want to choose the following -
  - Publishing Day - Friday, Saturday and Sunday
  - Publishing Hour - 4 PM, 5 PM, 1 PM, 8 PM
  - Video Title Length - 30 and 60 characters
  - Video Category - Entertainment, Music, News
- Now, these options are for getting the best possible results.
- You can play around with other values, to see the worst possible scenarios and compare the different results.

Conclusion

Summary of my findings and reflection of what I have learnt in this project

In conclusion, after analyzing the results of my dataset and going through my attributes and values multiple times, I narrowed down my findings to four main factors. This entire interactive narrative visualization was intended for those young, aspiring youtubers who want to make a difference out there through their videos and content. I can confidently say that all they have to do in order to make their video into a trend is follow the following -

- **Day of Release** - I have found out after my analysis that videos which release on Fridays have a higher probability of it becoming a trend then videos which are released on other days of the week.
- **Hour of Release** - After the analysis, I have come to the conclusion videos which are published between 4 and 5 pm have a good chance of reaching a wider audience!
- **Video Caption Length** - Now, In Conclusion, it is quite clear that videos having caption length between 30 to 50 characters have the highest chance of them attracting a larger audience.
- **Video Category** : I can confidently say that videos released in the Entertainment and Music Industry will most likely become a trend than videos in the other categories.

What in Hindsight could I have done differently :

- The youtube dataset was very vast and diverse. I had information about different countries available. Hence, I could have done analysis of the trending youtube videos available in different countries.
- Secondly, here I have provided the only four factors by which the user will be able to analyze the trends and patterns of the datasets. I could have even explored the relationship between other factors such as comments, likes/dislikes ratio and their impact on the trending youtube videos.
- I could have given the user more visualisations such as scatter plots, heatmaps and word clouds in order to analyze the commonly found words in the video captions of most common trending youtube videos.

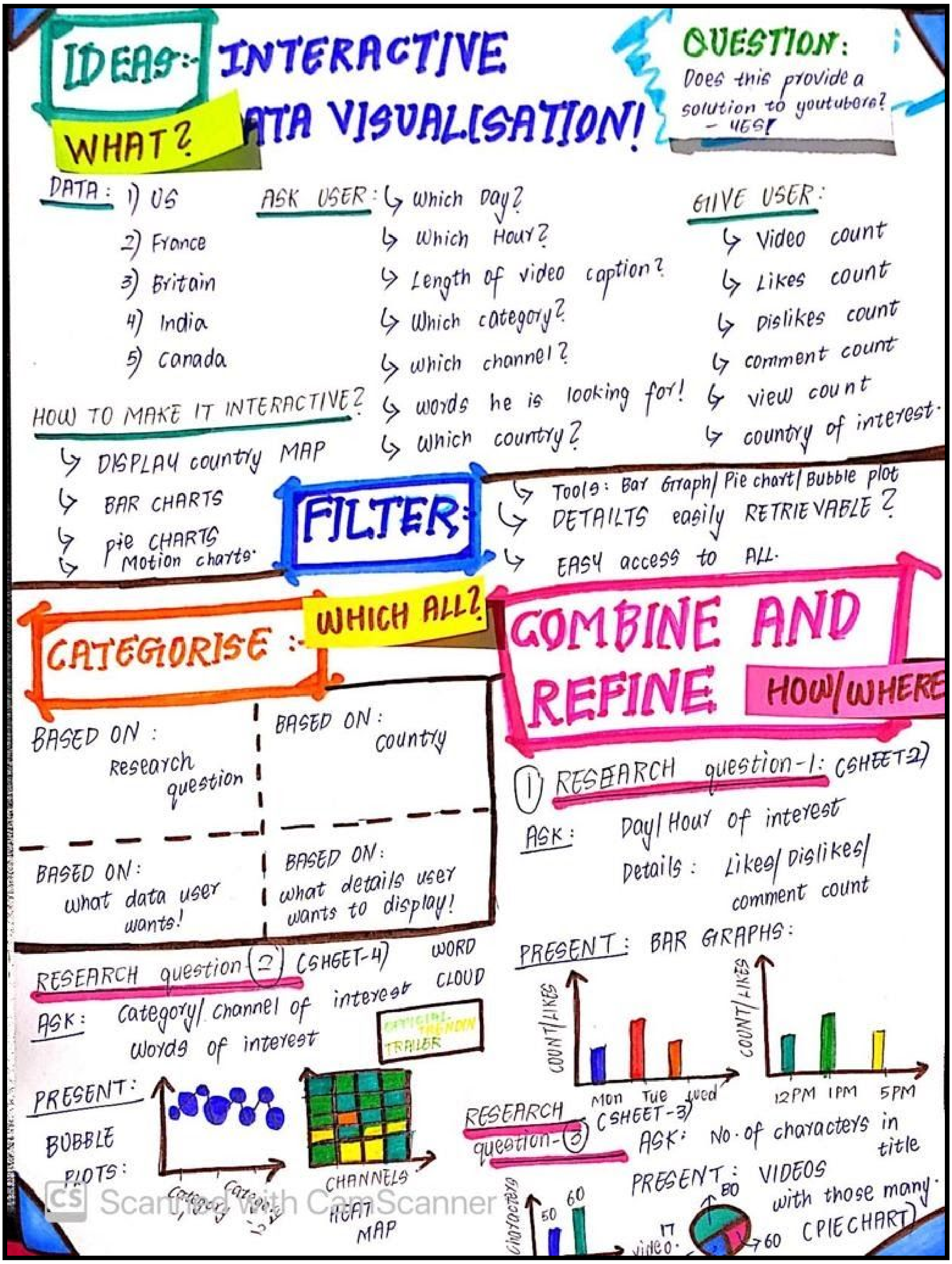
Bibliography

Mitchell, J (2017, October 26). Trending YouTube Video Statistics and Comments. Retrieved from : https://www.kaggle.com/datasnaek/youtube

Markus Knauff (2013, March) . Space to Reason : A Spatial Theory of Human Thought Retrieved from : https://mitpress.mit.edu/books/space-reason

Appendix

- **Kindly find the Initial Five Design Sheets attached to this.**



IDEAS:- INTERACTIVE ATA VISUALISATION!

WHAT?

QUESTION: Does this provide a solution to youtubers? - YES!

DATA: 1) US 2) France 3) Britain 4) India 5) Canada

ASK USER:
- Which Day?
- Which Hour?
- Length of video caption?
- Which category?
- which channel?
- words he is looking for!
- Which country?

GIVE USER:
- Video count
- Likes count
- Dislikes count
- comment count
- view count
- country of interest.

HOW TO MAKE IT INTERACTIVE?
- DISPLAY country MAP
- BAR CHARTS
- pie CHARTS
- Motion charts.

FILTER:
- Tools: Bar Graph/ Pie chart/ Bubble plot
- DETAILS easily RETRIEVABLE?
- EASY access to ALL.

CATEGORISE :- WHICH ALL?

BASED ON: Research question

BASED ON: country

BASED ON: what data user wants!

BASED ON: what details user wants to display!

COMBINE AND REFINE HOW/WHERE

① RESEARCH question-1: (SHEET-2)

ASK: Day/ Hour of interest

Details: Likes/ Dislikes/ comment count

PRESENT: BAR GRAPHS:

RESEARCH question ② (SHEET-4)

ASK: Category/ channel of interest

Words of interest

WORD CLOUD

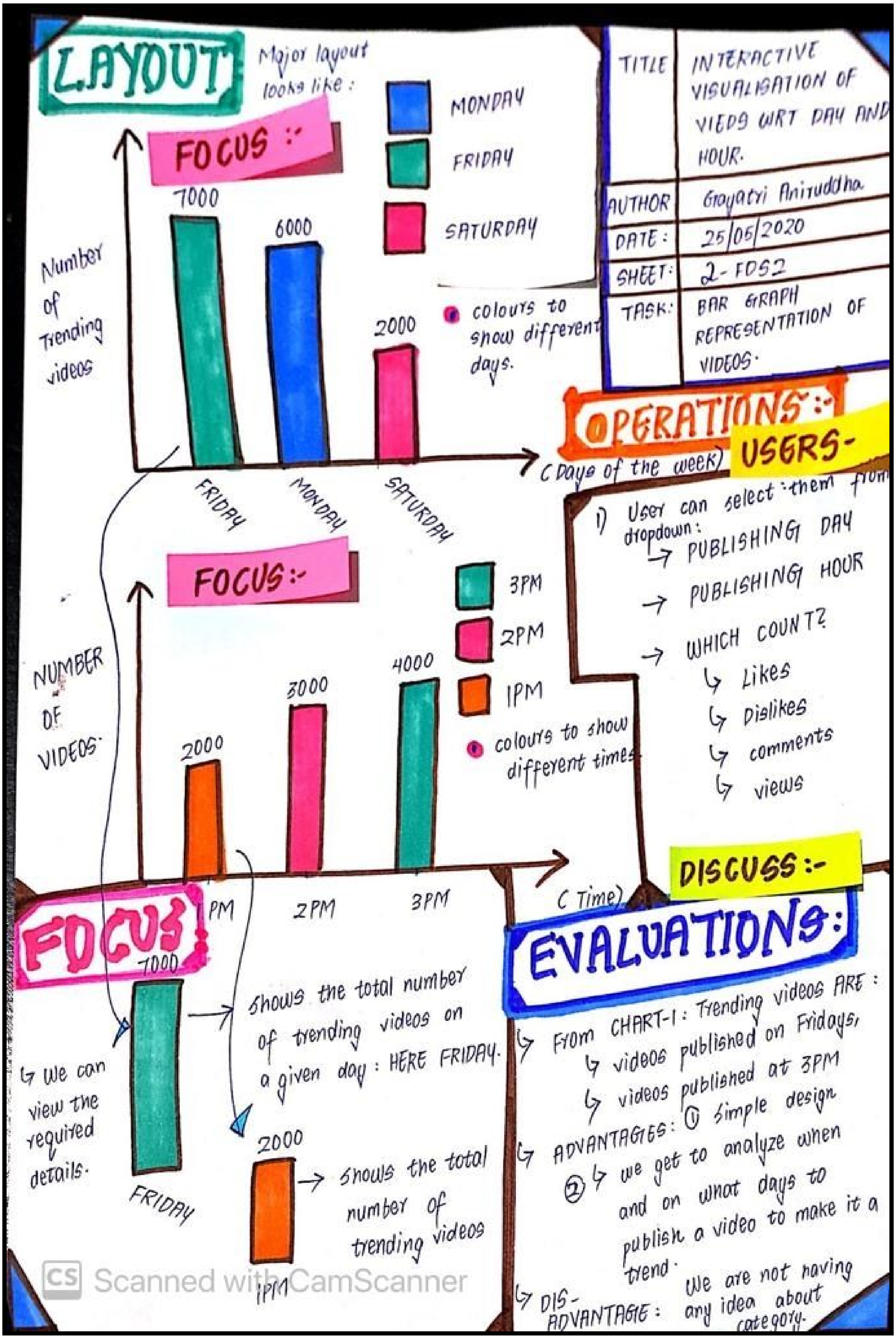
PRESENT: BUBBLE PLOTS:

CHANNELS HEAT MAP

RESEARCH question-③ (SHEET-3)

ASK: No. of characters in title

PRESENT: VIDEOS with those many. (PIECHART)



LAYOUT

Major layout looks like :

● colours to show different days.

● colours to show different times.

TITLE	INTERACTIVE VISUALISATION OF VIEDS WRT DAY AND HOUR.
AUTHOR	Gayatri Aniruddha
DATE :	25/05/2020
SHEET :	2- FDS2
TASK :	BAR GRAPH REPRESENTATION OF VIDEOS.

OPERATIONS :-

USERS-

1) User can select them from dropdown :
→ PUBLISHING DAY
→ PUBLISHING HOUR
→ WHICH COUNT?
↳ Likes
↳ Dislikes
↳ comments
↳ views

FOCUS

↳ We can view the required details.

→ Shows the total number of trending videos on a given day : HERE FRIDAY.

→ Shows the total number of trending videos

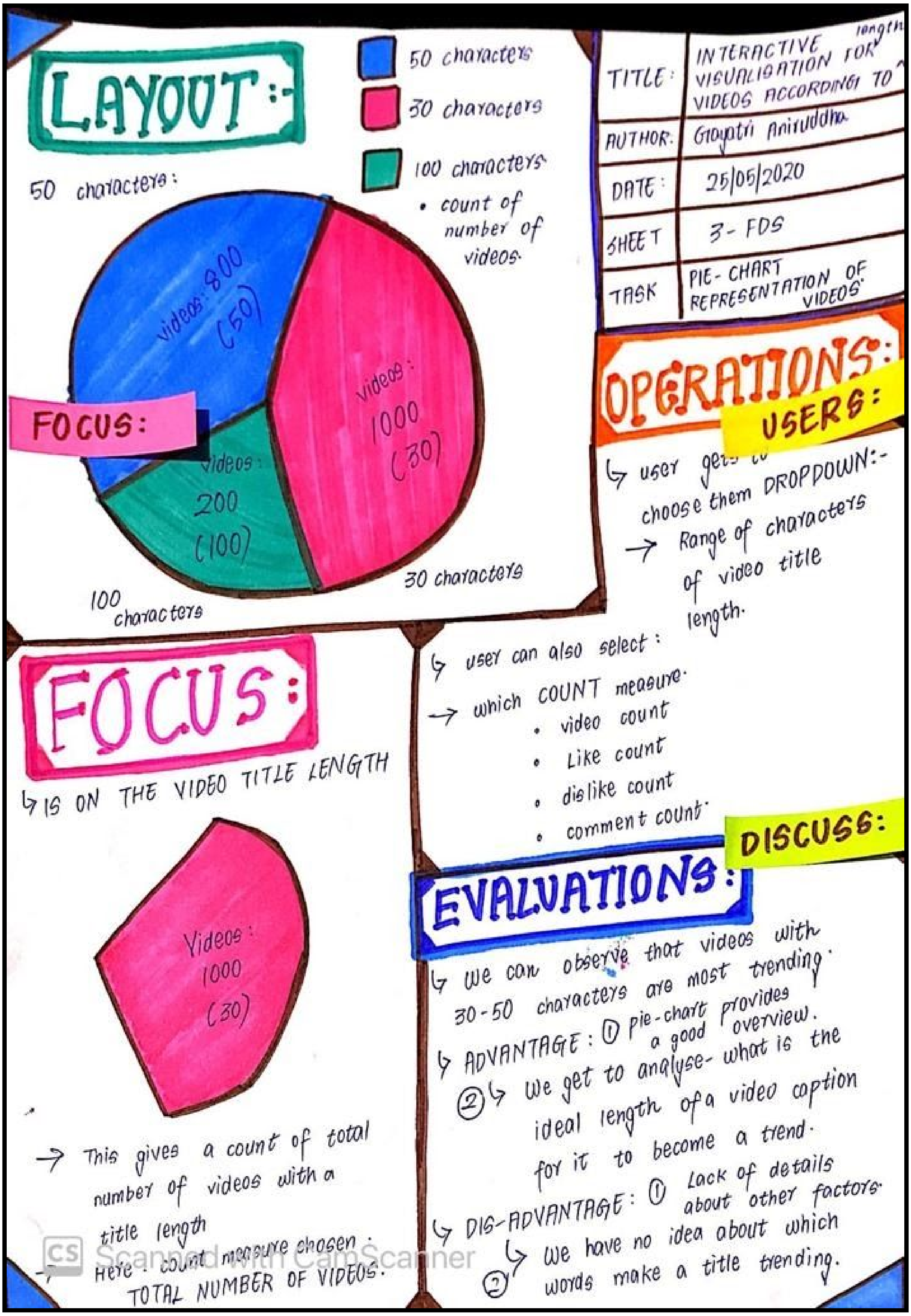
DISCUSS :-

EVALUATIONS:

↳ From CHART-1 : Trending videos ARE :
↳ videos published on Fridays,
↳ videos published at 3PM

↳ ADVANTAGES: ① simple design
② ↳ we get to analyze when and on what days to publish a video to make it a trend.

↳ DIS-ADVANTAGE : We are not having any idea about category.



TITLE:	INTERACTIVE VISUALISATION FOR VIDEOS ACCORDING TO length
AUTHOR.	Gayatri Aniruddha
DATE:	25/05/2020
SHEET	3- FDS
TASK	PIE- CHART REPRESENTATION OF VIDEOS'

LAYOUT :-

🟦 50 characters

🟥 30 characters

🟩 100 characters

- count of number of videos.

50 characters :
videos: 800 (50)
videos : 1000 (30)
videos : 200 (100)
FOCUS :
30 characters
100 characters

OPERATIONS:

USERS:

↳ user gets to choose them DROPDOWN:-

→ Range of characters of video title length.

↳ user can also select :

→ which COUNT measure.

- video count
- Like count
- dislike count
- comment count.

FOCUS:

↳ IS ON THE VIDEO TITLE LENGTH

→ This gives a count of total number of videos with a title length

→ Here : count measure chosen : TOTAL NUMBER OF VIDEOS.

DISCUSS:

EVALUATIONS :

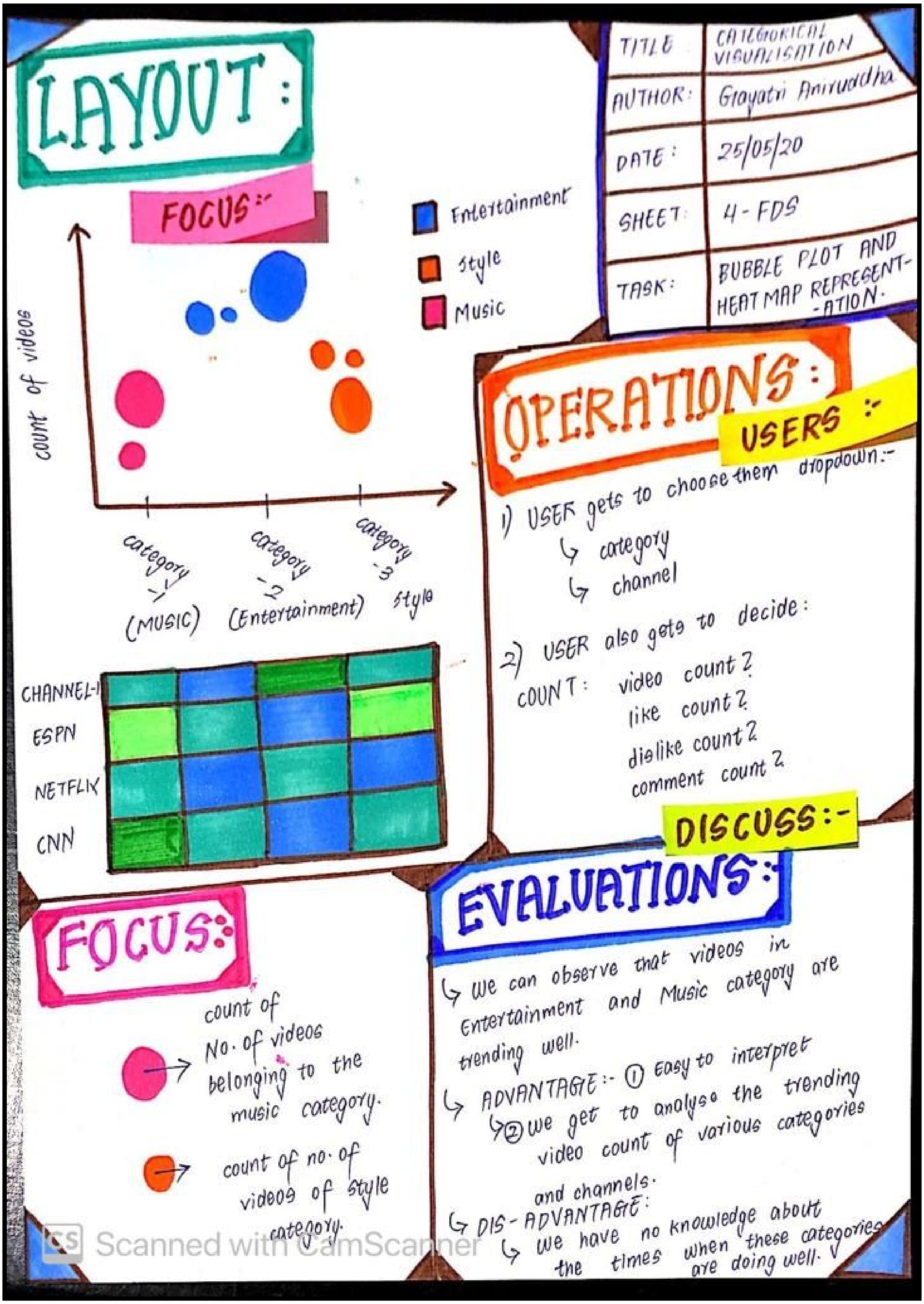
↳ We can observe that videos with 30-50 characters are most trending.

↳ ADVANTAGE : ① pie-chart provides a good overview.

② ↳ We get to analyse- what is the ideal length of a video caption for it to become a trend.

↳ DIS-ADVANTAGE : ① Lack of details about other factors.

② ↳ We have no idea about which words make a title trending.



TITLE	CATEGORICAL VISUALISATION
AUTHOR:	Gayatri Anirudha
DATE:	25/05/20
SHEET:	4-FDS
TASK:	BUBBLE PLOT AND HEAT MAP REPRESENT-ATION.

LAYOUT:

OPERATIONS:

USERS :-

1) USER gets to choose them dropdown:-
   - category
   - channel

2) USER also gets to decide:
   COUNT: video count?
   like count?
   dislike count?
   comment count?

FOCUS:

count of No. of videos belonging to the music category.

count of no. of videos of style category.

DISCUSS:-

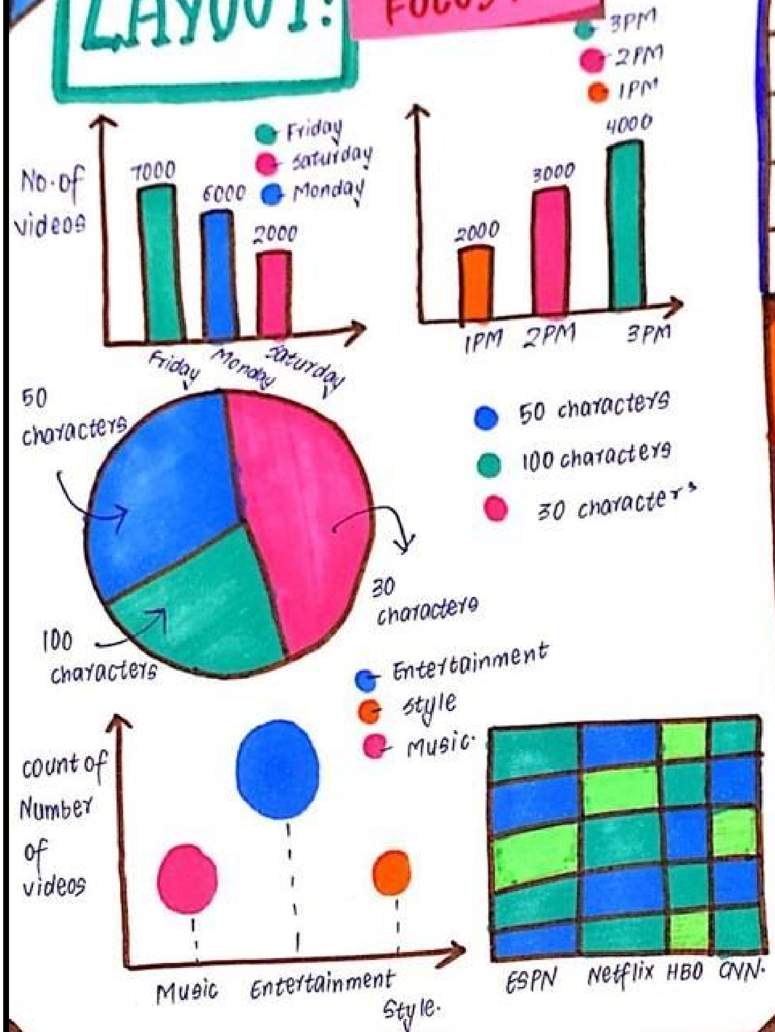
EVALUATIONS:

- We can observe that videos in Entertainment and Music category are trending well.
- ADVANTAGE:- ① Easy to interpret
  - ② we get to analyse the trending video count of various categories and channels.
- DIS-ADVANTAGE:
  - we have no knowledge about the times when these categories are doing well.

TITLE:	INTERACTIVE DATA VISUALISATION
AUTHOR:	Gayatri Aniruddha
DATE:	25/05/20
SHEET:	5-FDS
TASK:	COMPLETE : Realization VISUALISATION :-

LAYOUT:

FOCUS :-

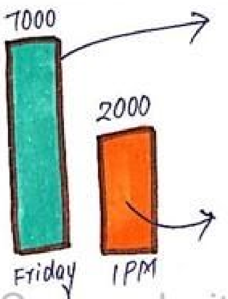


FOCUS:

Most Important Factor :

① PUBLISHING HOUR

② PUBLISHING DAY.



Displays the total number of videos which became a trend on Friday.

Displays the total number of videos that became a trend at 1PM.

OPERATIONS:

USERS:

→ User chooses the following from dropdown

1. publishing day
2. publishing hour
3. title character length
4. category type
5. channels

→ count of : (measure)

- comment count
- view count
- like/ dislike count

DETAIL :

DISCUSS:

Technology:

↳ Displayed in R-shiny application.

↳ Estimated time :

↳ 100 Hours for Designing visualisation.

Estimated schedule :

↳ 10 days at 10 hours per day.

Hardware/ software needed :

R studio.