STREAMING PLATFORM COMPARISON

(NETFLIX – PRIME VIDEO – HULU – DISNEY+)

Visualization Concept Report

Group Members:

- Dhruv Arora (Programmer, Designer, Usability Evaluator)
- Divya Pandey (Programmer, Designer, DevOps)
- Lakshmi Venkatasubramanian (Programmer, Designer, DevOps)
- Priyanka Bijlani (Programmer, Designer, DevOps)
- Sharmeelee Bijlani (Project Manager, Designer, Usability Evaluator)

Contents

EXECUTIVE SUMMARY

Streaming services started as an add-on to DVD and digital download offerings with a trickle of second-run movies and TV shows. They were supplements to the programs you watched on their first (and second) runs on cable TV. But speedier internet connections and an abundance of video streaming devices have accelerated the decline of traditional cable. More and more viewers are cutting the cord entirely in favor of dedicated streaming alternatives.

With so many video streaming platforms in the market, monetizing their platforms through monthly subscription payment model, we wanted to enable the data driven decisions for choosing the video streaming platform through information visualization for the consumers. The core product values of these platforms are the availability of quality movies proxied by user ratings on IMDB or rotten tomatoes, subscription cost, number of movies to watch, different genres of movies, availability of different languages etc. We used these rich factors to enable data driven decisions via representing these factors in an information visualization dashboard.

In the time of COVID-19, the streaming platform has gone far enough to enable remote watching together i.e. you may not be in the same physical place as your friends and family, you can still watch the same shows together. Disney+ supports Scener, a remote co-watching tool, and you can use the Netflix Party Chrome extension to sync that service's video playback across devices. Hulu ad-free subscribers, Amazon Prime Video members, and Plex users also get co-watching features.

The visualization dashboard can be accessed using the following link:

https://public.tableau.com/profile/priyanka5190#!/vizhome/StreamingData_15978090421400/StreamingPlatformDB?publish=yes

A CONCEPT BACKGROUND

The topic was motivated by common interest of our team members which is watching movies. We strongly believe that users will get benefited out of this dashboard as the results are user driven. Based on the subscription cost and other metrics, they will be able to select a platform that most suits their needs. For example, a family might be inclined to subscribe to a platform that has a lot of kid friendly movies. But when there are so many platforms offering the same feature, how would they know which one is the best for them?

The goal of our project is to create an interactive visualization that can be perceived as a go-to marketplace for comparing popular streaming platforms on various metrics. It enables the users to select a platform they want to subscribe to base on their needs. Users can compare streaming platforms based on Genre, content type, average IMDb rating, languages supported, runtime, number of movies and choose the one based on the metric of their interest.

The typical users of the dashboard are existing subscribers of streaming platforms, and movie enthusiasts and viewers.

PROTOTYPE DEVELOPMENT AND USABILITY TEST RESULTS:

We evaluated the effectiveness of the Dashboard by testing out its visual design by directly assessing the participants. We devised the below three questions that would require the users to go through all our designed workflows.

- Which platform has the maximum number of movies in the language 'Hindi'?
- Which platform is the most kid friendly?
- Which genre has the most movies for the decade '1950s'?

The participants in our survey are a few of our classmates as well as professionals from the software and data science community. They make up a diverse target audience of our dashboard which is aimed at anyone interested in streaming movies. All the users come with different preferences and values for what makes a streaming platform more desirable. Their feedback was valuable for improving the usability and optimizing the effectiveness of the dashboard.

Overall, our dashboard was well-received by all the participants. Each one indicated that it was easy to follow. We also got positive feedback for our concept. The purpose of our dashboard is clear, interesting, and relevant to users.

On the other hand, we also received constructive feedback and identified visuals that required improvement. The issues identified are ineffectiveness of the bubble chart in representing the metrics due to limitations of the chart, more emphasis on quantitative metrics rather than qualitative metrics that were not useful to users in genre-decade chart, misleading color encodings between different visuals and ambiguity in the placement of filters which were confusing to the users.

All these findings directly impact on how we implement the design changes, and how the users use the interface.

PROCESS DESCRIPTION:

Based on the feedback from user research and usability testing results, we identified the issues, prioritized them based on multiple criteria, tried out potential solution ideas and refined them further.

The major changes include replacing hover cards in the bubble chart by table of KPIs to enable the users to see the dynamic changes to the metrics, replacing genre-decade chart with genre-content type visual and adding distribution of IMDB ratings by platform to elucidate the quality metrics that the users are interested in. We also paid attention to the

choice of colors between visuals and placement of filters to correctly associate with the relevant visuals. Below is the breakdown of the process steps on how we arrived at the final version of the dashboard.

DATA HANDLING (ACQUISITION, PROFILING, PREPARATION):

We acquired the dataset from Kaggle.com which is an amalgamation of data scraped from Reelgood.com and IMDb dataset. This dataset consists of Movies data across four different platforms 'Hulu' , Netflix', 'Prime videos' and 'Disney+'.

https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney

The motivation behind choosing this data set is it is rich and informative. Multiple variables such as Title, Year, Age, IMDb, Rotten Tomatoes, Netflix, Hulu, Prime Video, Disney+, Type, Directors, Genres, Country, Language, Runtime provide useful information for visual analysis and help uncover meaningful insights. The dataset is also historical in nature and provides information on streaming platforms over time dating back to 1932. has 16744 rows and 16 different columns and contains nominal, ordinal, interval, and ratio data types.

DATA PROCESSING:

In this section, we will walk through a variety of issues in handling and analyzing the data set. We reduced the data for analysis to 11 variables as seen in Table 1: Variables of Interest

Table 1: Variables of interest

Column Name	Data Type
Title	Nominal
Year	Interval
Age	Ordinal
IMDb	Ratio
Rotten Tomatoes	Ratio
Directors	Nominal
Genres	Nominal
Country	Nominal
Language	Nominal
Runtime	Ratio
Platform	Nominal

PROBLEMS ENCOUNTERED AND MODIFICATIONS TO THE PLAN:

- ***DATA CLEANING ISSUES:***

Issue #1: Comma delimited columns

In the original dataset, columns Language, Genre are comma delimited. In order to visualize the distribution of these variables, we split these variables into a list and pivoted the resulting list values into rows.

Issue #2: Binary values for platforms

In order to make the dataset more useful for visualization, 4 columns representing each of the 4 platforms have to be converted from binary values into string and pivoted into rows.

Issue #3: Missing values

After addressing issues 1 and 2, the total number of rows ended up being 70950. A handful of variables reported missing values in rows, as seen in Table 2: Missing Values.

Table 2: Missing Values

Variable Name	# Rows Missing	% Missing values
Rotten Tomatoes	39045	55.03%
Age	32703	46.09%
IMDb	1237	1.74%
Directors	1622	2.29%
Genre	286	0.40%
Language	1005	1.42%
Country	616	0.87%
Runtime	1220	1.72%

We can see from table 2 that Rotten Tomatoes has nearly 55% of values missing and Age has nearly 46% of values missing. As we are interested in comparing Rotten Tomatoes and Age across platforms as they provide important insights, instead of dropping the missing values, we imputed the missing values in such a way the overall average of Rotten Tomatoes for each platform wouldn't change. For Age, we imputed the values such that the missing values were assigned the highest age content type for each platform. Similarly, imputed the missing values for Runtime and IMDb rating with overall mean values for each platform. As the remaining variables are hard to impute due to their nominal nature, we considered removing the missing values. These variables include Language, Directors, Genres and Country. After all the cleaning, we are left with 68662 rows and 11 columns.

- ***IDENTIFYING A VISUALIZATION TECHNIQUE APPROPRIATE FOR THE DATASET AND ITS INTEGRATION WITH OUR WORK:***

  Working through the five design sheets helped our group define our final concept with clarity. The individual brainstorm activity through the first design sheet gave each of our team member's the opportunity to explore different user stories and visualization ideas. This led to a rich discussion, which started with a concept of recommending movies to users and culminated with a concept of finding the most suitable streaming platform for users.

  The subsequent design sheets helped our group design the implementation or deliverable of our concept in terms of visuals. We were able to focus on at least three different visuals including a top view of metrics, a filtering capability on movie characteristics such as Age, Language and Genre and a drill down to a detailed table of movies. Finally, we used the five-design sheet process to design a prototype or draft of our concept dashboard which gave us clarity on which ideas will work as well as which ideas need additional exploration and refinement.

- ***METRIC/KPI DEVELOPMENT:***

  From our brainstorm, we came up with a set of metrics that are important to evaluating the value of a streaming platform to a subscriber, which is our target audience. These platform-level metrics included count of movies, average rating, and languages supported which we could express by aggregating our data by platform and dimension filters selected. Additionally, we added static KPIs such as monthly subscription price and subscriber base to our set of metrics since these are key considerations of subscribers when selecting streaming platforms. These assumptions were validated in our user interviews.

  In the initial prototype, we displayed all of these metrics through a hover card with the count of movies illustrated through a bubble chart. In this visualization style, count of movies was the primary KPI. However, we modified our approach to visualizing all the metrics as primary by expressing them in a dynamic table which enables comparison across platforms as well as deep dive into a single platform based on selection filters.

- ***PUBLISHING THE FINAL VISUALIZATION TO TABLEAU PUBLIC:***

  The medium we chose to publish our final dashboard is Tableau public server. Using this option, we can enable sharing and discovery of our dashboard to any interested user by sharing a URL. The URL below links to our final dashboard which can be viewed and downloaded for further exploration.

https://public.tableau.com/profile/priyanka5190#!/vizhome/StreamingData_15978090421400/StreamingPlatformDB?publish=yes

INSIGHTS FROM VISUALIZATION:

In this section, we will cover 3 major workflows that touch upon all the visuals and their interactions to provide valuable insights into Streaming data.

- ***PLATFORMS COMPARISON :***

  The table of KPIs includes qualitative and quantitative metrics that enable comparison between platforms that users care about such as price, IMDb rating, number of movies, Rotten Tomatoes, number of languages supported and number of subscribers.

  For example, in **Figure 1**, Rotten Tomatoes > 63.1% filter has been applied using the slider and the KPI table shows how different platforms compare on various metrics based on this filter selection. If the user is interested in choosing a platform based on IMDb rating, then Disney+ is the platform that will be selected.

  In another example, in **Figure 2**, Genre = 'Documentary' filter has been applied and the KPI table is updated based on this filter selection. If the user is again interested in choosing a platform based on IMDb rating, then Hulu is the platform that will be selected.

  Figure 1: Comparison of platforms with Rotten Tomatoes filter applied.

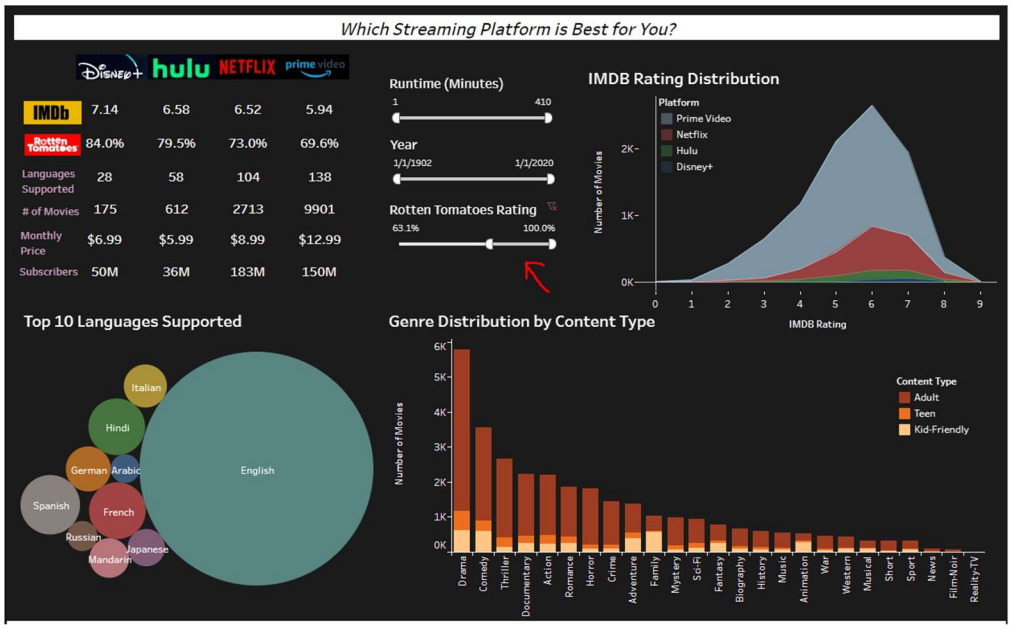
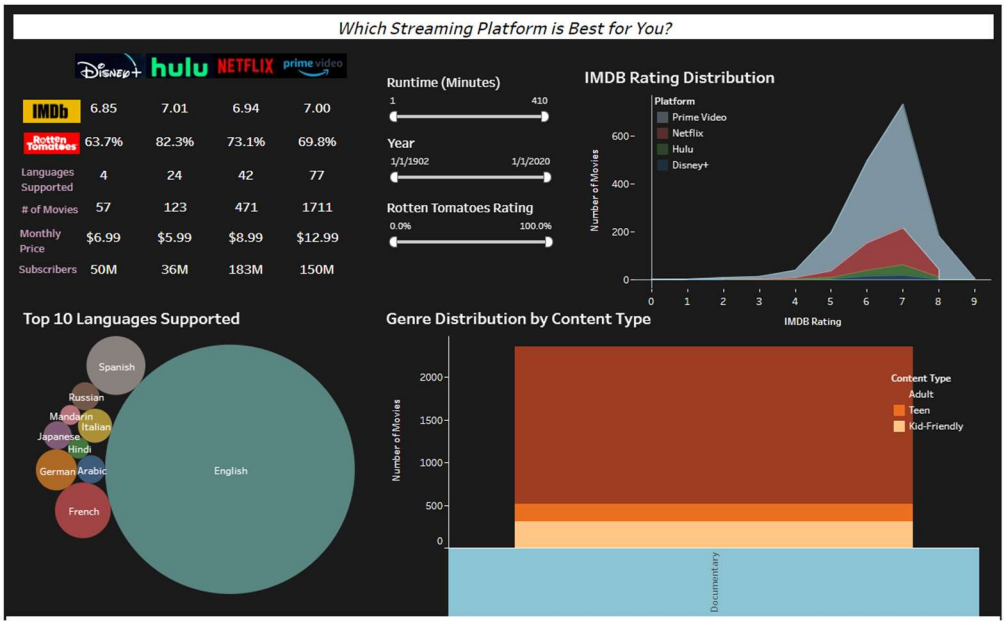


Figure 2: Comparison of platforms on Genre filter applied.

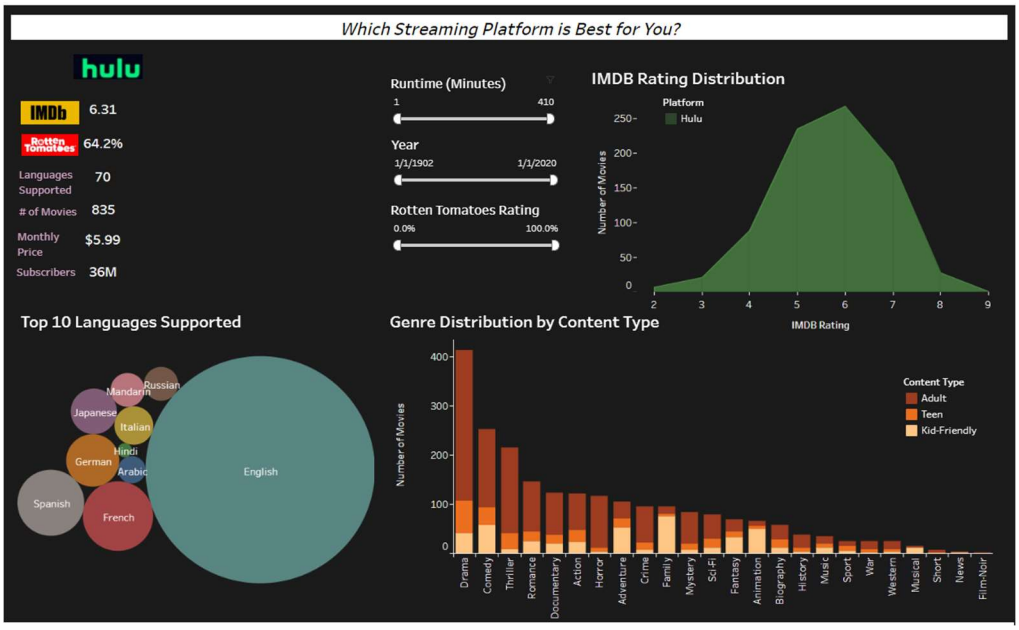


- ***PLATFORM DEEP DIVE :***

  The headers of the KPI table are also buttons and clicking a particular platform button would select that platform, provide the KPIs for that platform and all the other visuals would also refresh to reflect the platform selection. For example, clicking the Hulu button **(Figure 3)** provides the KPIs for the Hulu platform. IMDb distribution visual shows the breakdown of the count of Hulu movies by IMDb rating. Out of the total 835 movies for Hulu, 7 movies have IMDb rating 2, 21 movies have rating 3, 88 movies have rating 4, 235 have rating 5, 268 movies have rating 6, 187 movies have rating 7, 28 movies have rating 8 and 1 has rating 9. Similarly, the 'Top 10 Languages Supported' and 'Genre distribution by Content Type' visuals represent Hulu's data appropriately.

  We can also make use of the sliders and other filters to get further insights into Hulu platform.

Figure 3: Clicking Hulu button provides Hulu data in all the visuals.



- ***EXTRACT DETAILED MOVIES LIST***

  We can get a detailed list of movies from 'Top 10 Languages Supported' and 'Genre Distribution by Content type' visuals by clicking anywhere on the visuals. For example, in **Figure 4**, Rotten Tomatoes and Runtime filters have been applied using the sliders and the visuals change dynamically based on this filter selection.

  To get the list of movies, click either on the Genre chart bars or on the language chart circle to see the link to movies. Then further click on the movies list link to get the list of movies on a new page.

  In this particular example, the detailed movies list table **(Figure 5)** has one movie 'Tom Petty and the Heartbreakers: Running' Down a Dream' that is supported in English language, falls under 2 genre - categories namely Documentary and Music, has an IMDb rating of 8.60, Rotten Tomatoes rating of 100% and Runtime of over 227 minutes and streamed on Prime Video.

Figure 4: Extract movies list

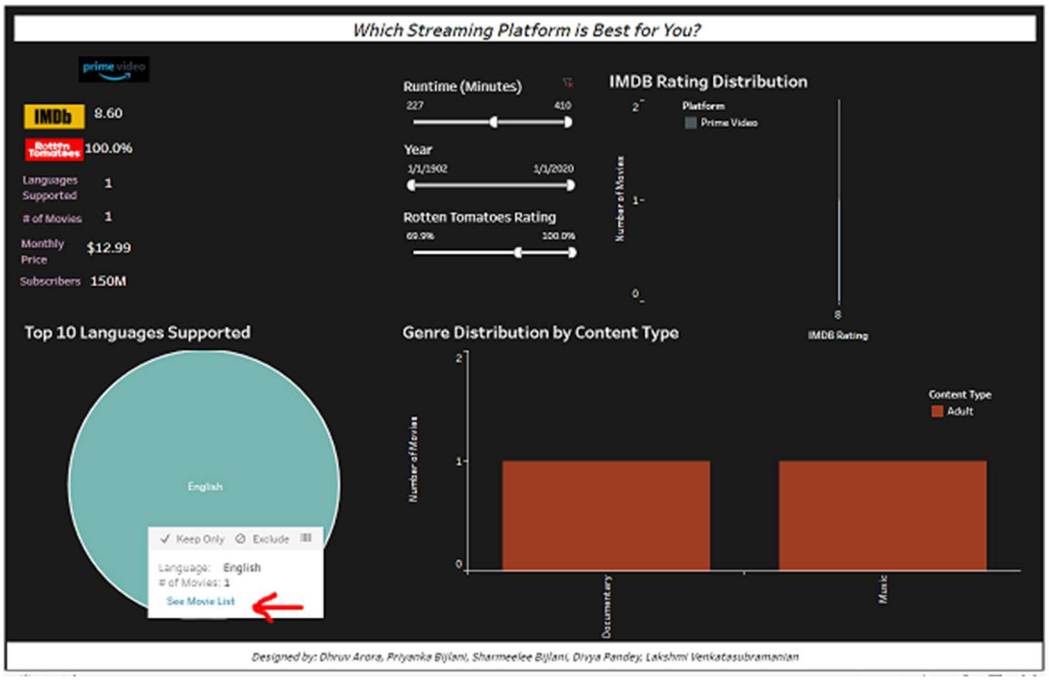
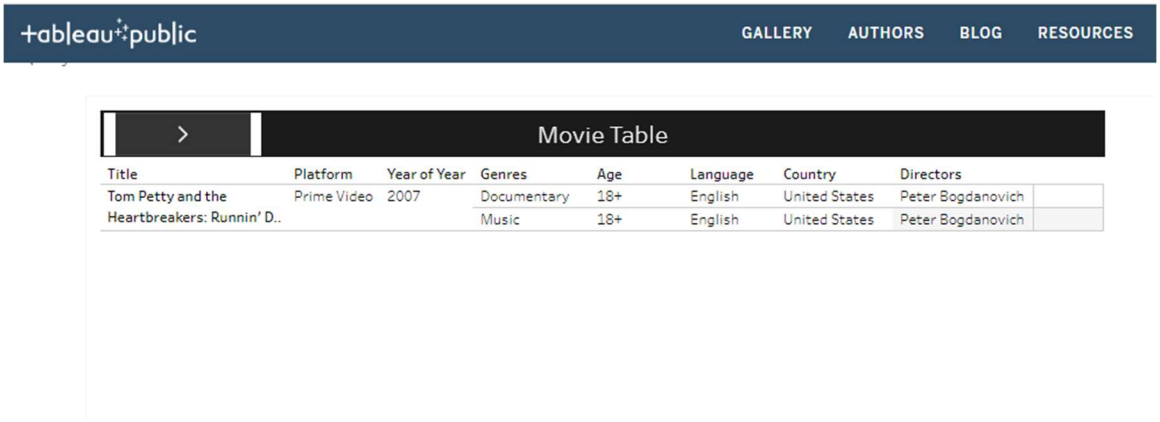


Figure 5: Detailed list of movies



CRITICAL EVALUATION OF STREAMING PLATFORM COMPARISON DASHBOARD:

The four-visualization part of the dashboard has been developed after a rigorous process of understanding the scope of the data and then focusing on the key components required by User to land on the answer to the key question - 'What is the best platform according to my needs?' All four visualizations together are designed to work cohesively and meticulously following design rules such 'more data to ink ratio' and 'Balance and alignment.'

First visualization is a simple yet very effective way to understand the comparison of various parameters such as Count of movies, Language supported, IMDb ratings, rotten tomatoes ratings, monthly price and number of subscribers for all four platforms. This visualization being presented in a tabular format proves to be very effective to capture the initial question 'Where does every platform stand for these parameters'? Once a user has chosen a platform of their interest, they can further deep dive into the details of various parameters with the help of remaining all visualizations. Some of the key observations enabled are platforms with the highest number of movies are Prime videos, however it has the lowest average IMDb rating and Prime videos has the highest subscription price.

Second visualization 'IMDb Rating Distribution' uses line as a mark and color and area encodings to present the distribution of IMDb ratings across four platforms for comparisons and enable the view of 'Quantity vs Quality'. Third visualization 'Top 10 languages supported' is created using 'Packed bubble' chart by using 'size' marks and color encodings to display top 10 languages by movie count, which is often an interest area for people who are interested in content limited to few languages. Packed graph visualization serves as an appropriate visualization because the size of the bubbles for each language provides an overall comparison dimension, however hovering over the bubble can provide the definitive count of movies.

Final visualization 'Genre Distribution by Content type' presents a distribution of movies split by Genre and divided into content type buckets per age groups – Adult, Teen and Kid friendly. This is achieved by stacked bar graph visualization where each stack represents age group. There are a couple of great visual techniques displayed, first length encoding symbolizes the number of movies in each genre and since the axis is sorted, the top genre with the most number of movies can be immediately seen. Further each stack when divided across age buckets using 'Hue' encoding, we can clearly capture if a genre is specific age group friendly or not. For example – Horror being mostly biased toward adult age group movies whereas animation serves the majority of movies to kids. A user can further view the list of all movies under certain areas of interest such as Genre - Comedy and Age group – Teen.

Moreover, dashboard also provides key operations such as 'Filter' and 'Navigation' for users to follow their own user journey and make data driven decisions. Platform level filtration is applied by logo buttons on the first visualization where users can opt to deep dive on areas of interest and all other visualizations will reflect the values limited to the chosen platform. Furthermore, users can also filter on Runtime (length of movie), Year of release and Rotten Tomatoes Rating to narrow the scope of their exploration and find the relevant information.

Overall the dashboard is great example of product produced by following the Fry's model of 'Information Visualization design process where every step followed thoroughly and meticulously has produced a visualization that captures the key elements of good visualization i.e. being user centric, answering the key question, following the encoding

practices according to the nature of data (quantitative, qualitative, ordinal etc.) and being a self-service tool.

'Information Visualization design process'

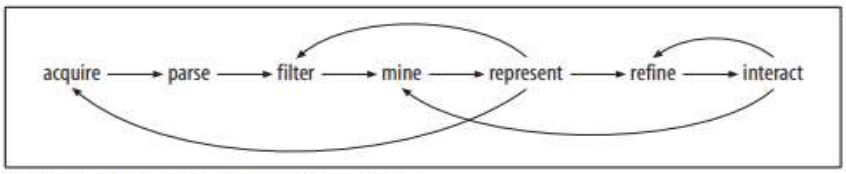


Figure 1-12. Interactions between the seven stages

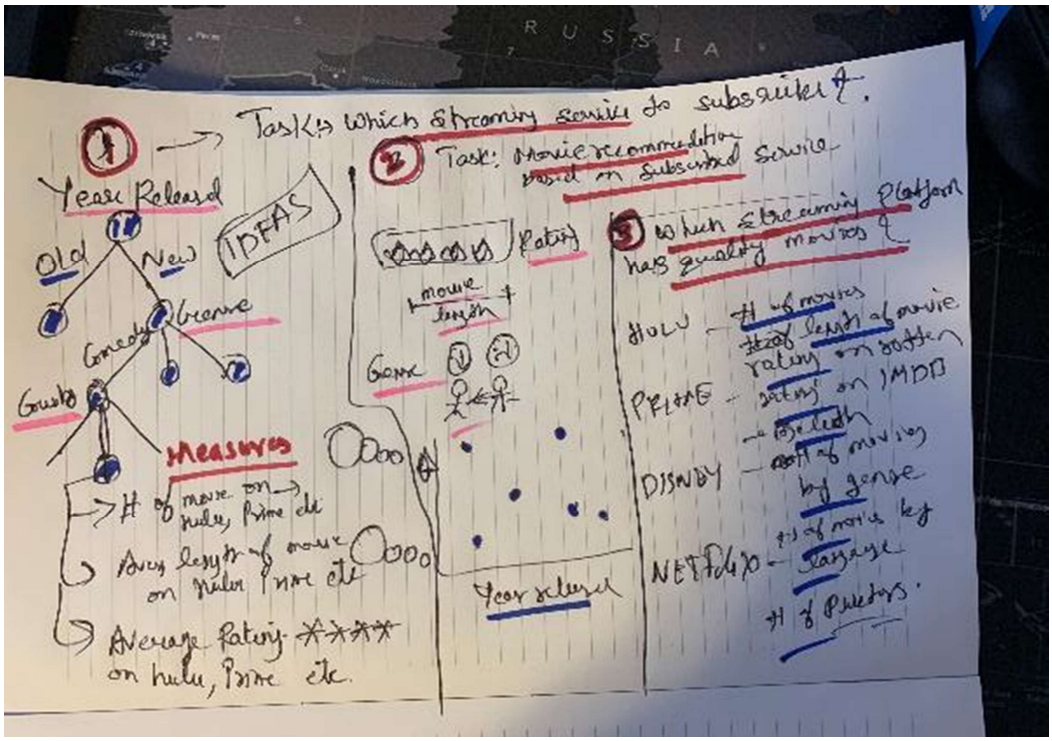
REFERENCES:

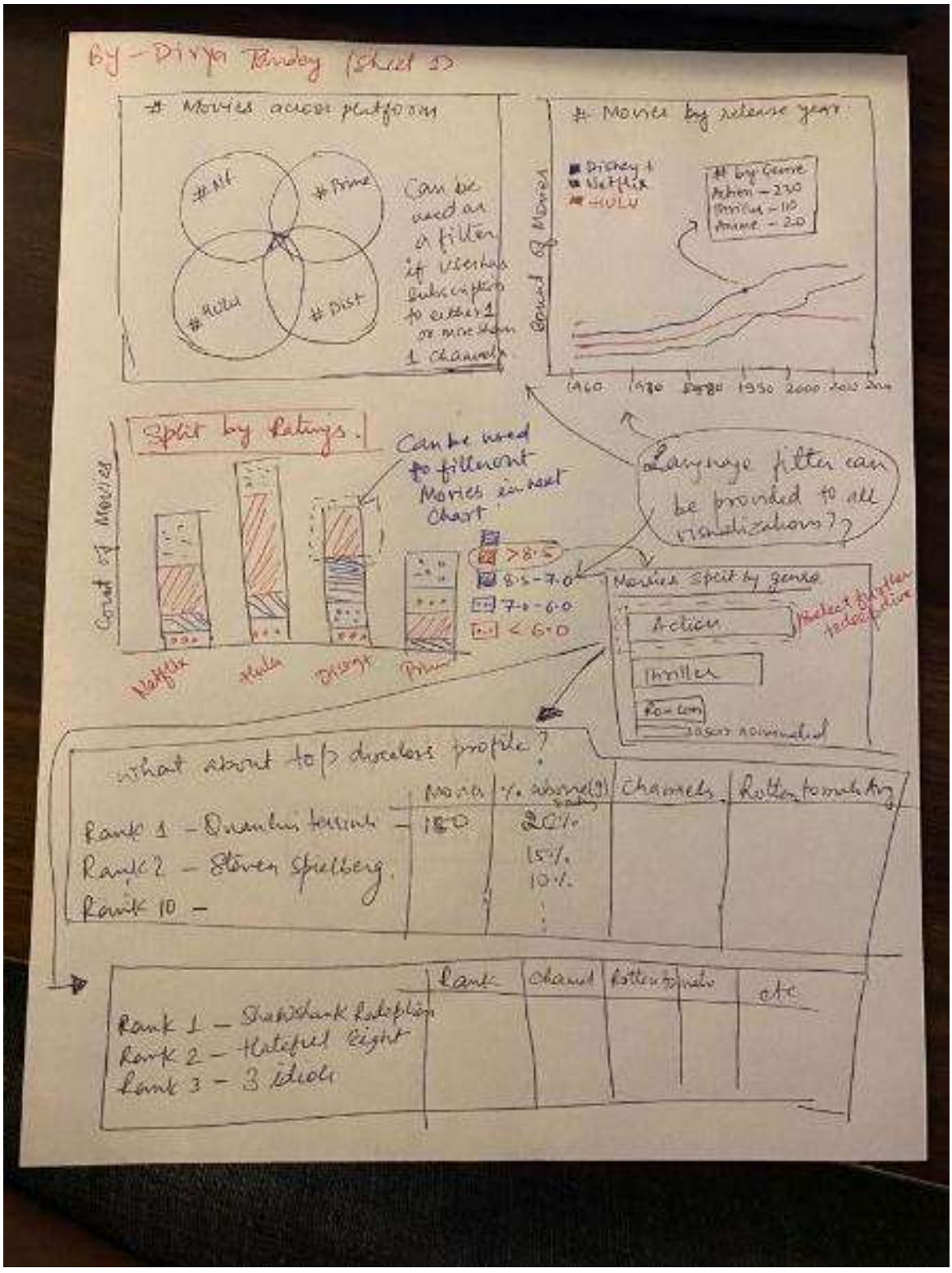
1. Data 511 Reading – Heer Schneiderman, Interactive Dynamics for Visual Analysis
2. Data 511 Lecture – Visual Encoding of Data, Graphical Excellence, and Integrity
3. Data 511 Lecture – Pre-attentive vs attentive visual processing

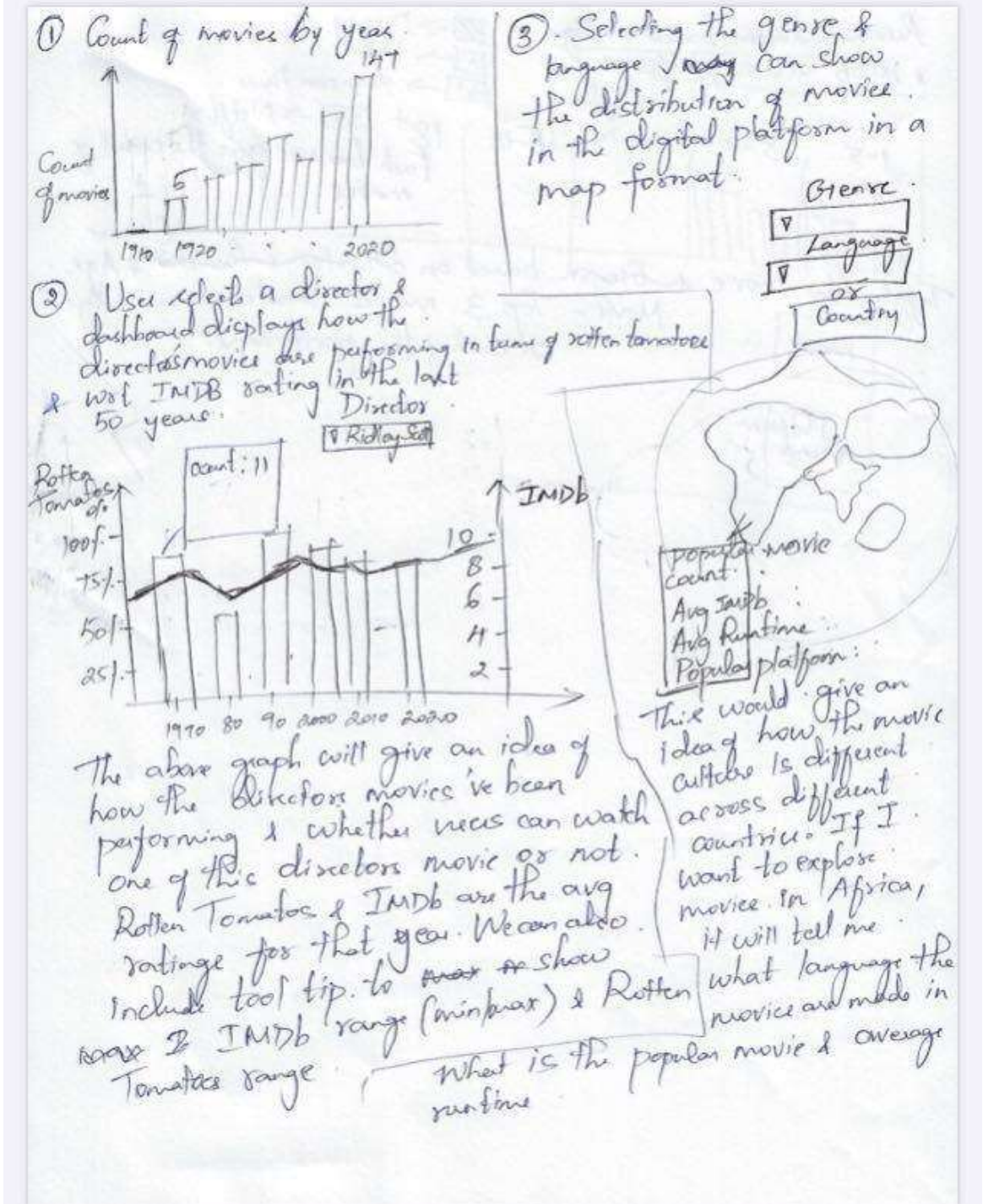
APPENDIX: DESIGN SKETCHES AND FINAL DESIGNS

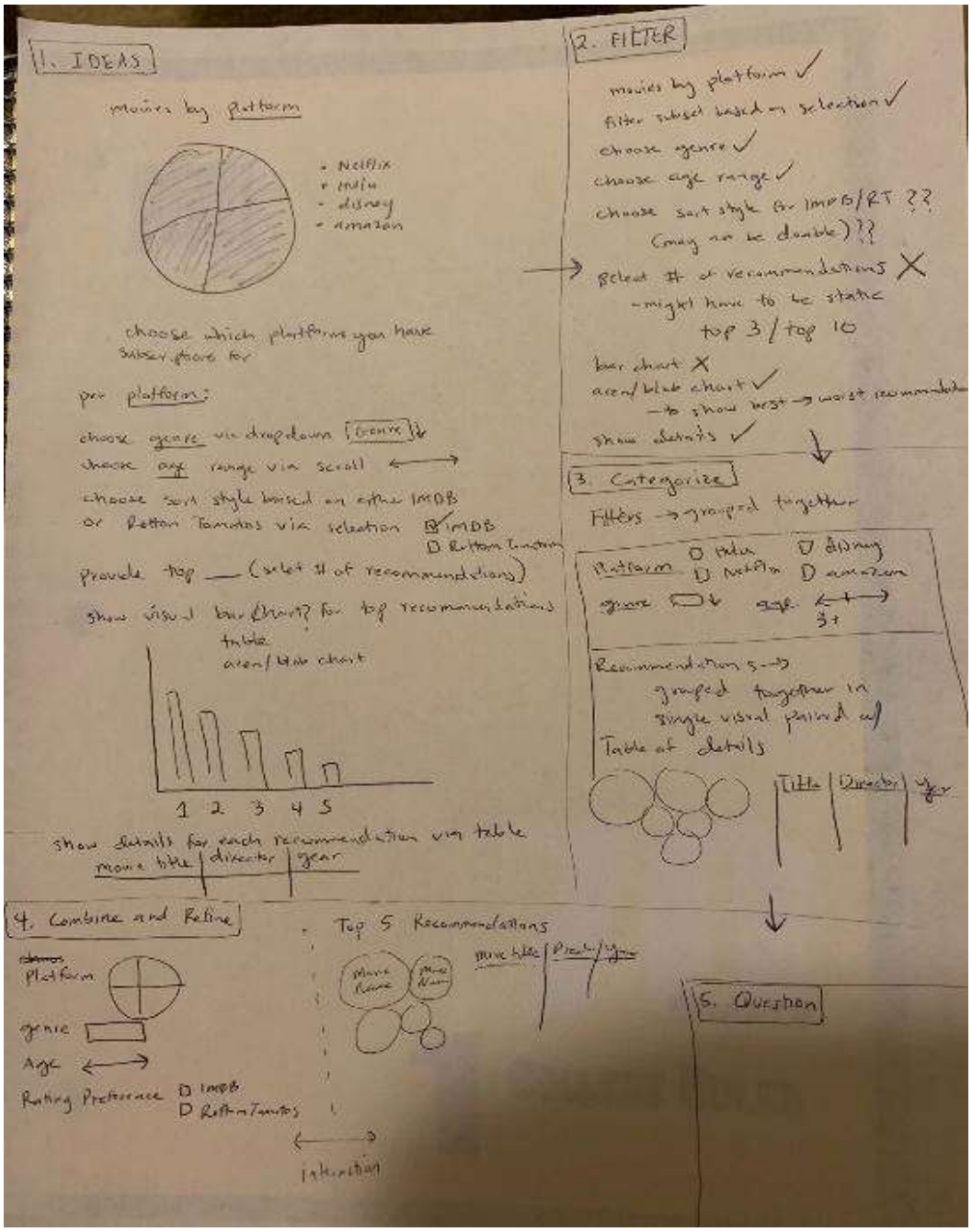
FDS Design sheets:

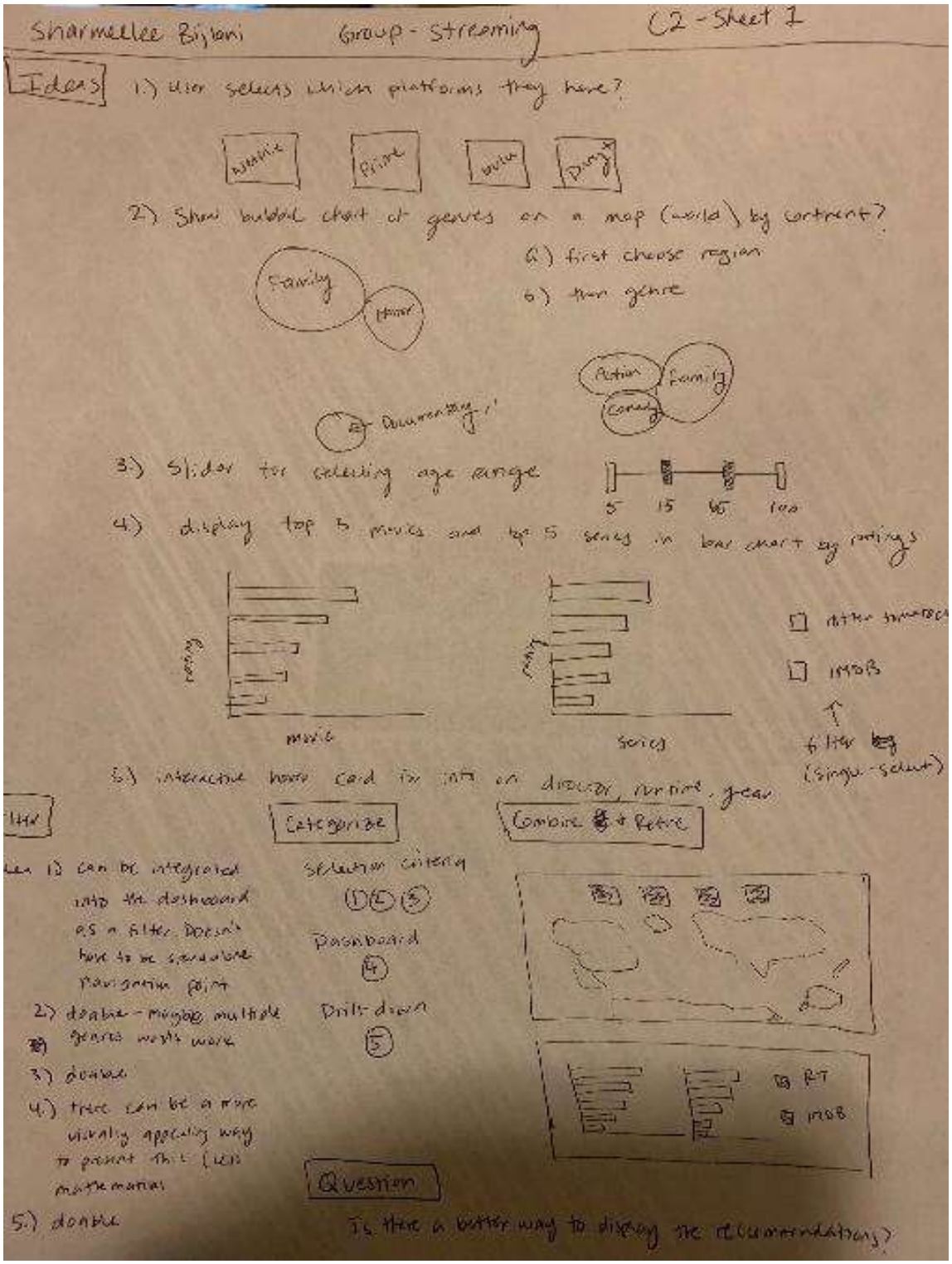
Design sheet 1:



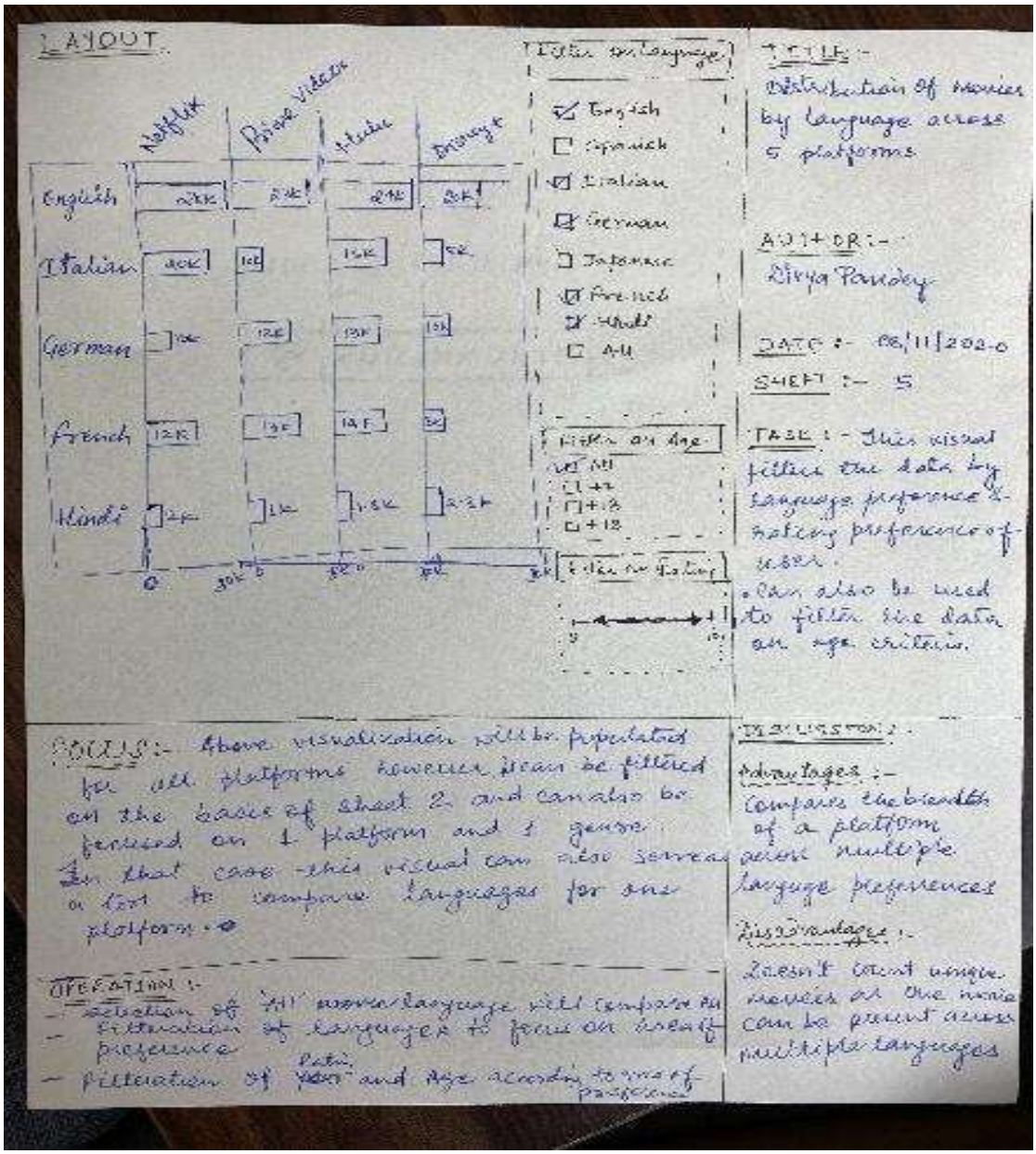

By - Divya Tendey (Sheet 1)
Movies across platforms
#Hit
#Prime
#Hulu
#Disc
Can be used as a filter if viewer has subscription to either 1 or more than 1 channel
Movies by release year
Disney +
Netflix
HULU
by Genre
Action - 230
Thriller - 110
Anime - 20
Count of Movies
1960
1980
1990
1950
2000
2010
2020
Split by Ratings.
Can be used to fillerout Movies in next Chart
Count of Movies
>8.5
8.5-7.0
7.0-6.0
< 6.0
Netflix
Hulu
Disney
Prime
Language filter can be provided to all visualizations??
Movies split by genre
Action
Thriller
Rom-Com
Select filter to delete
what about top 5 directors profile?
Rank 1 - Quentin Tarantino - 180 - 20%
Rank 2 - Steven Spielberg - 15%
Rank 10 - 10%
Rank 1 - Shawshank Redemption
Rank 2 - Hateful Eight
Rank 3 - 3 Idiots

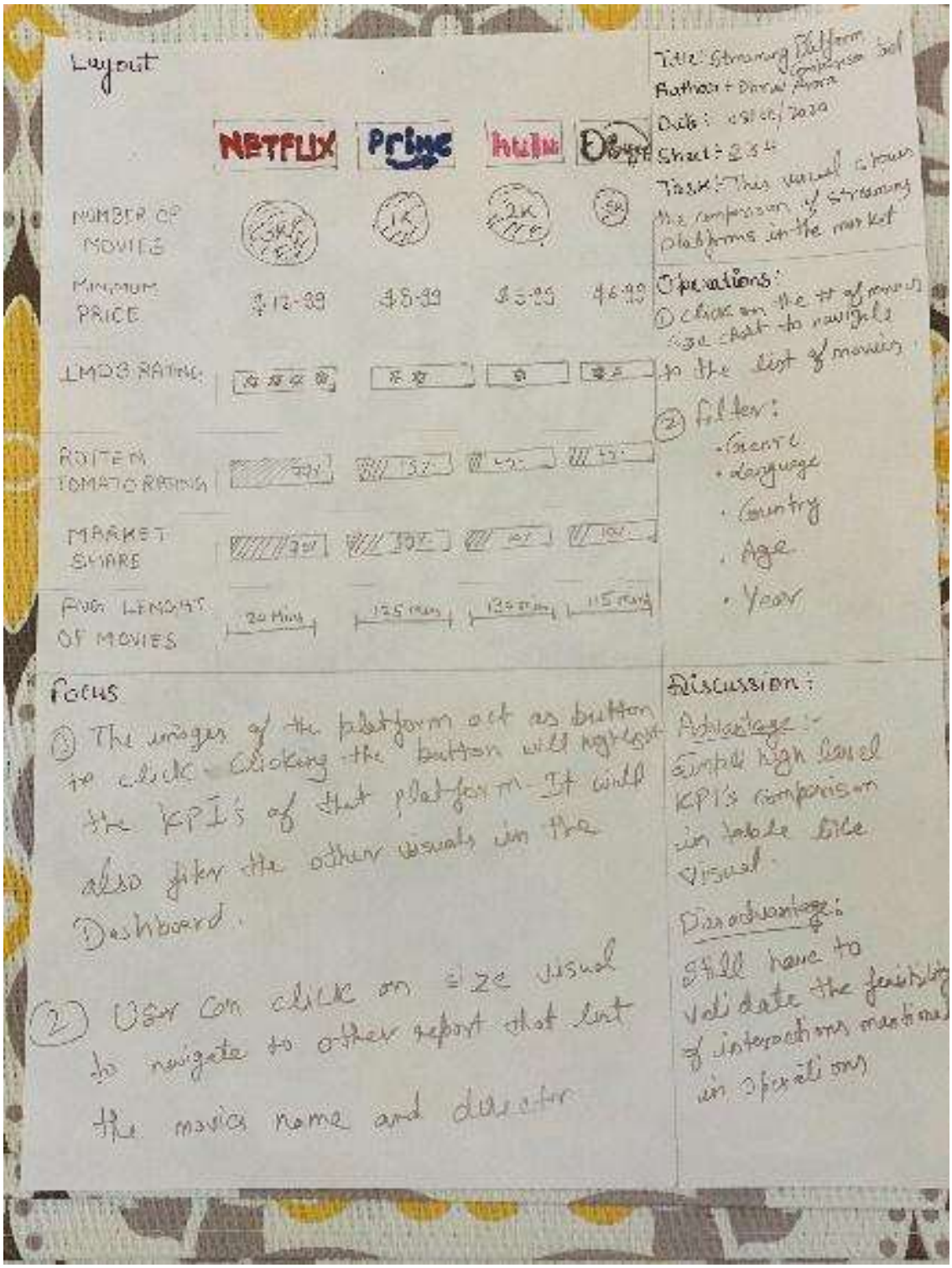

① Count of movies by year
147
5
Count of movies
1910 1920 2020
② User selects a director & dashboard displays how the director's movies are performing in terms of rotten tomatoes & wrt IMDB rating in the last 50 years.
Director
Ridley Scott
count: 11
Rotten Tomatoes %
100%
75%
50%
25%
IMDb
10
8
6
4
2
1970 80 90 2000 2010 2020
The above graph will give an idea of how the Directors movies've been performing & whether users can watch one of the directors movie or not.
Rotten Tomatoes & IMDb are the avg ratings for that year. We can also include tool tip. to show IMDb range (min/max) & Rotten Tomatoes range
③. Selecting the genre & language can show the distribution of movies in the digital platform in a map format.
Genre
Language
or
Country
popular movie
count:
Avg IMDb
Avg Runtime
Popular platform:
What is the popular movie & average runtime
This would give an idea of how the movie culture is different across different countries. If I want to explore movies in Africa, it will tell me what language the movies are made in

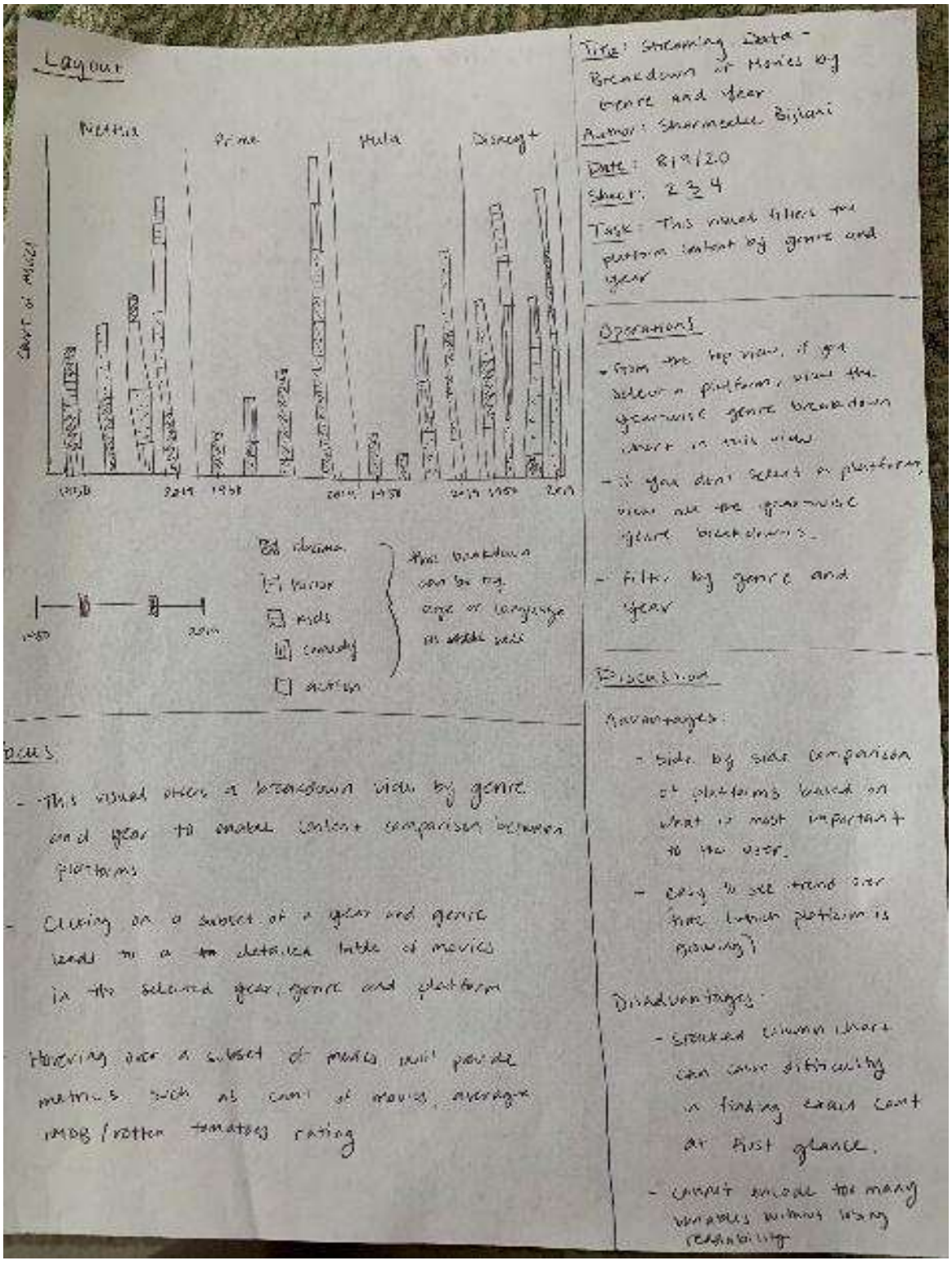

1. IDEAS
movies by platform
Netflix
Hulu
disney
amazon
choose which platforms you have subscriptions for
per platform:
choose genre via dropdown
choose age range via scroll
choose sort style based on either IMDB or Rotten Tomatoes via selection
IMDB
Rotten Tomatoes
Provide top ___ (select # of recommendations)
show visual bar chart? for top recommendations
table
area/line chart
1 2 3 4 5
show details for each recommendation via table
movie title
director
year
2. FILTER
movies by platform
filter subset based on selection
choose genre
choose age range
choose sort style (ie IMDB/RT ??
(may not be doable)??
select # of recommendations
- might have to be static
top 3 / top 10
bar chart
area/line chart
- to show best → worst recommendations
show details
3. Categorize
Filters → grouped together
Platform
Hulu
Disney
Netflix
Amazon
genre
age
3+
Recommendations →
grouped together in
single visual paired w/
Table of details
Title
Director
Year
4. Combine and Refine
Platform
genre
Age
Rating Preference
IMDB
Rotten Tomatoes
interaction
Top 5 Recommendations
movie title
Director
year
5. Question

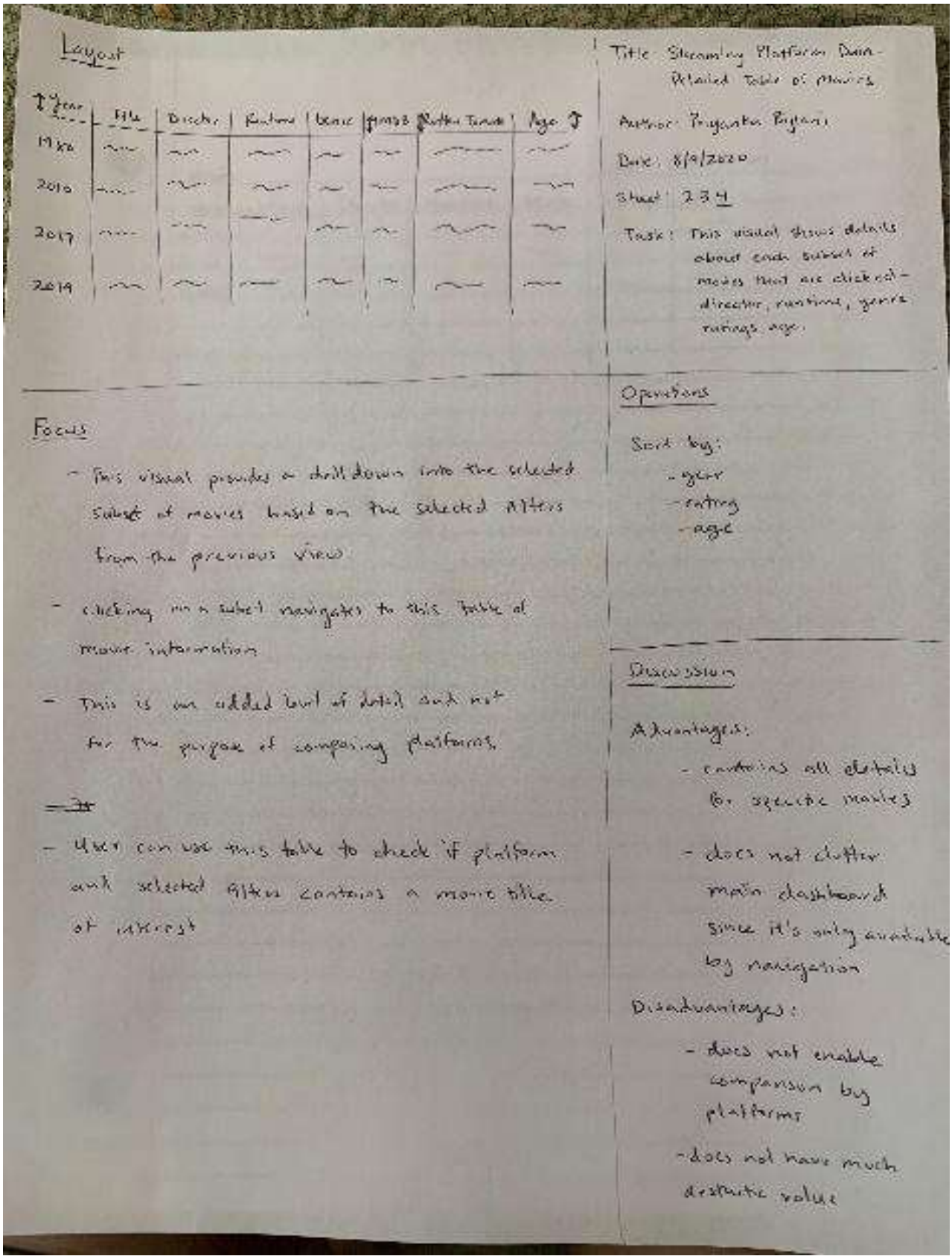


Design sheets 2, 3 and 4:


LAYOUT
Netflix
Prime Video
Hotstar
English
Italian
German
French
Hindi
Filter on Language
English
Spanish
Italian
German
Japanese
French
Hindi
All
Filter on Age
All
+7
+13
+18
Filter on Rating
TITLE :- Distribution of movies by language across 5 platforms
AUTHOR :- Divya Pandey
DATE :- 08/11/2020
SHEET :- 5
TASK :- This visual fillers the data by language preference & rating preference of user.
• Can also be used to filter the data on age criteria.
FOCUS :- Above visualization will be populated for all platforms however it can be filtered on the basis of sheet 2 and can also be focused on 1 platform and 1 genre. In that case this visual can also be used to compare languages for one platform.
OPERATION :-
- selection of 'All' movie/language will compare all
- Filtration of languages to focus on user's preference
- Filtration of Rating and Age according to user's preference
DISCUSSION :-
Advantages :- Compares the breadth of a platform across multiple language preferences
Disadvantages :- Doesn't count unique movies as one movie can be present across multiple languages


Layout
NETFLIX
Prime
hulu
Disney
NUMBER OF MOVIES
MINIMUM PRICE
$12.99
$6.99
IMDB RATING
ROTTEN TOMATO RATING
MARKET SHARE
AVG LENGHT OF MOVIES
20 Mins
Title: Streaming Platform Comparison Tool
Author + Data Source
Date: 05/06/2020
Sheet: 2 of 2
Task: This visual shows the comparison of streaming platforms in the market
Operations:
① click on the # of movies to navigate to the list of movies
② filter:
· Genre
· Language
· Country
· Age
· Year
Focus
① The images of the platform act as button to click. Clicking the button will highlight the KPI's of that platform. It will also filter the other visuals in the Dashboard.
② User can click on the visual to navigate to other report that list the movie name and director
Discussion:
Advantage:
Simple high level KPI's comparison in table like visual.
Disadvantage:
Still have to validate the feasibility of interactions mentioned in operations


Layout
Netflix
Prime
Hulu
Disney+
Count of Movies
1950
2019
Bar chart
Pie chart
Kids
Comedy
Action
Title: Streaming Data - Breakdown of Movies by Genre and Year
Author: Sharmeelee Bijlani
Date: 8/9/20
Sheet: 2 & 4
Task: This visual filters by platform content by genre and year
Operations
- From the top view, if you select a platform, view the year-wise genre breakdown chart in this view
- If you don't select a platform, view all the year-wise genre breakdowns.
- Filter by genre and year
Decisions
Advantages:
- Side by side comparison of platforms based on what is most important to the user.
- Easy to see trend over time within platform is growing?
Disadvantages:
- Stacked column charts can cause difficulty in finding exact count at first glance.
- Cannot include too many variables without losing readability
Specs
- This visual offers a breakdown view by genre and year to enable content comparison between platforms
- Clicking on a subset of a year and genre leads to a detailed table of movies in the selected year, genre and platform
- Hovering over a subset of movies will provide metrics such as count of movies, average IMDB / Rotten Tomatoes rating



Layout

Year	Title	Director	Runtime	Genre	Genres	Rotten Tomatoes	Age
Mix							
2016							
2017							
2019							

Title: Streaming Platform Data - Detailed Table of Movies

Author: Priyanka Pujari

Date: 8/9/2020

Sheet: 2 3 4

Task: This visual shows details about each subset of movies that are clicked - director, runtime, genre, ratings, age.

Focus

- This visual provides a drill down into the selected subset of movies based on the selected filters from the previous view
- clicking on a subset navigates to this table of movie information
- This is an added level of detail and not for the purpose of comparing platforms.

~~-~~

- User can use this table to check if platform and selected filters contains a movie title of interest

Operations

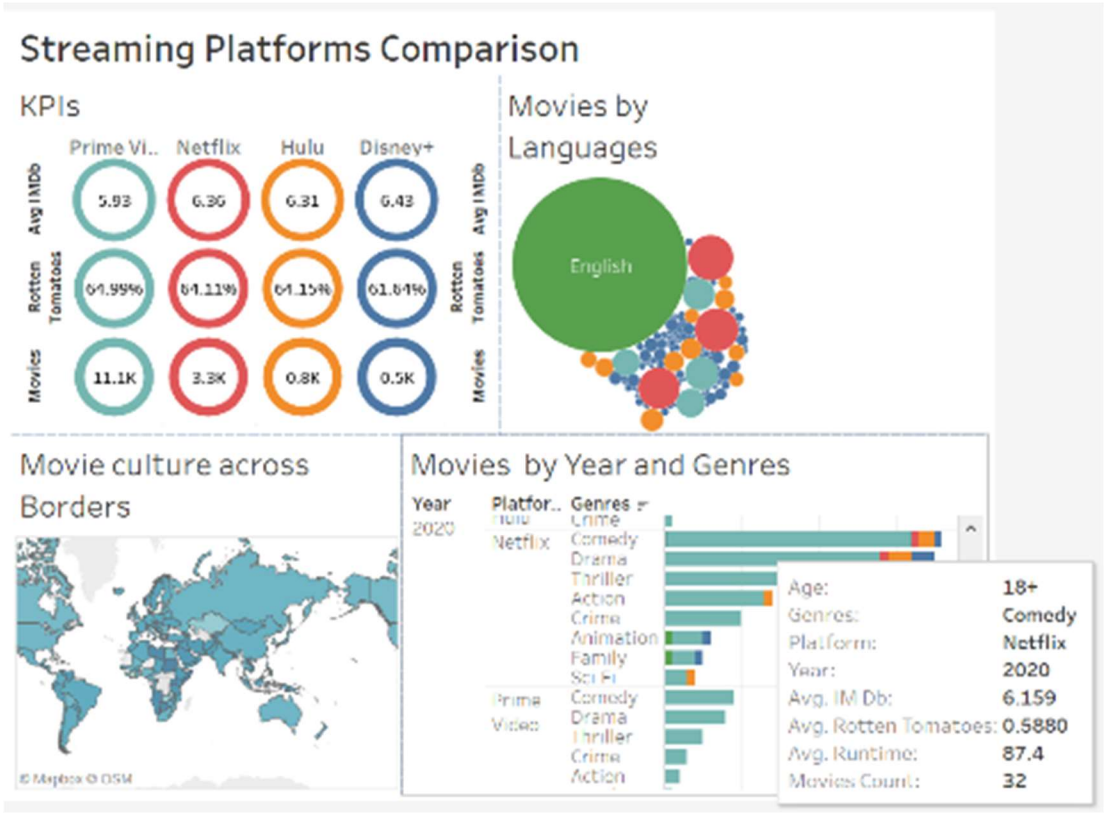
Sort by:
- year
- rating
- age

Discussion

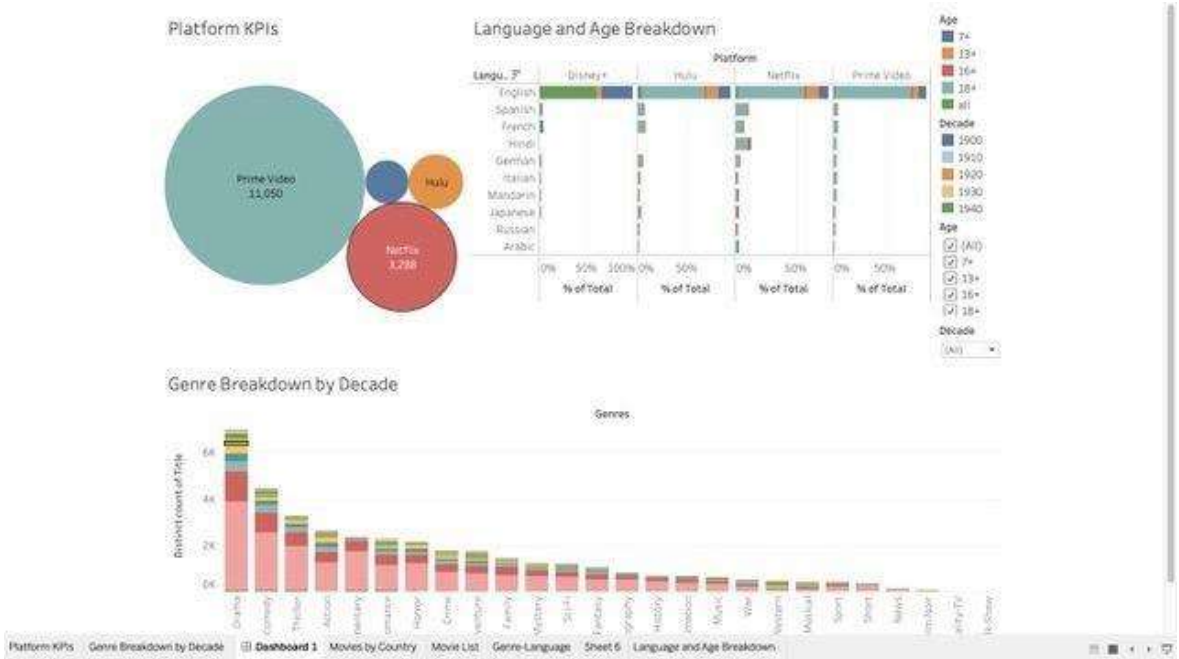
Advantages:
- contains all details for specific movies
- does not clutter main dashboard since it's only available by navigation

Disadvantages:
- does not enable comparison by platforms
- does not have much aesthetic value

Design sheet 5:



Prototype:



Final Dashboard: