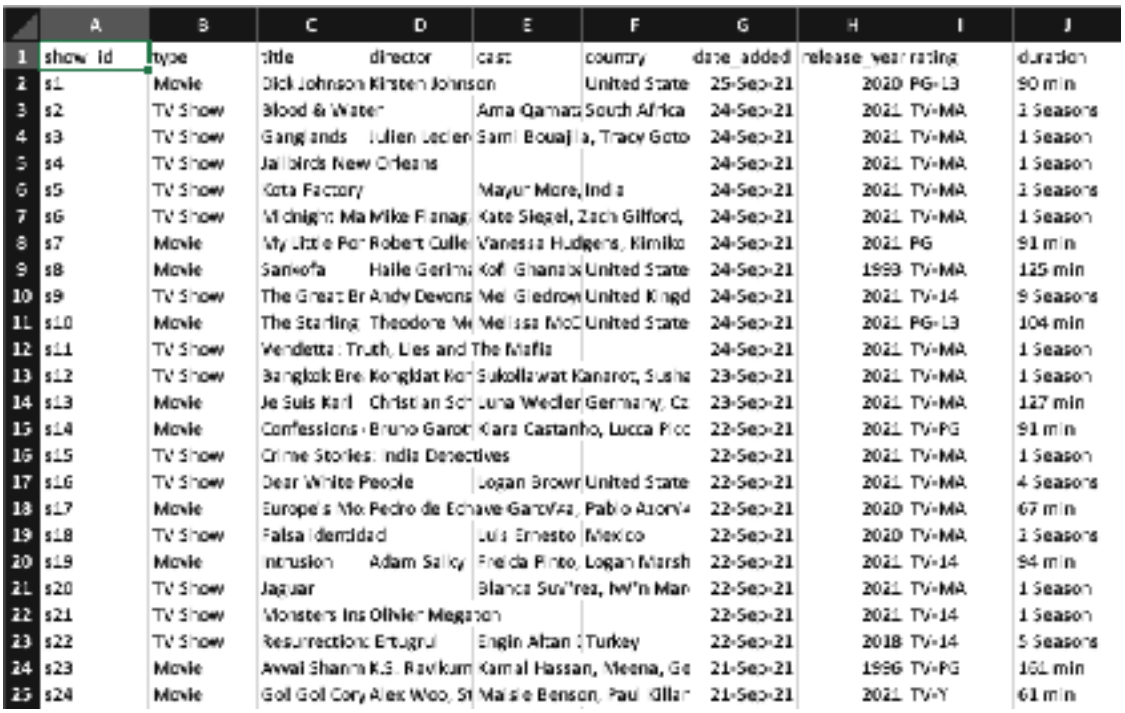
FINAL PROJECT 5304

A brief recap of your data, goals, and tasks, focusing on those that most directly influence your design.

My data was downloaded from kaggle dataset. Dataset : Netflix Titles.



	A	B	C	D	E	F	G	H	I	J
1	show_id	type	title	director	cast	country	date_added	release_year	rating	duration
2	s1	Movie	Dick Johnson	Kirsten Johnson		United State	25-Sep-21	2020	PG-13	90 min
3	s2	TV Show	Blood & Water		Ama Qamata	South Africa	24-Sep-21	2021	TV-MA	2 Seasons
4	s3	TV Show	Ganglands	Julien Leclen	Sami Bouaji	a, Tracy Goto	24-Sep-21	2021	TV-MA	1 Season
5	s4	TV Show	Jailbirds New	Orleans			24-Sep-21	2021	TV-MA	1 Season
6	s5	TV Show	Kota Factory		Mayur More,	Indi a	24-Sep-21	2021	TV-MA	2 Seasons
7	s6	TV Show	Midnight Ma	Mike Flanag	Kate Siegel,	Zach Gilford,	24-Sep-21	2021	TV-MA	1 Season
8	s7	Movie	My Little Pon	Robert Culle	Vanessa Hudgens,	Kimiko	24-Sep-21	2021	PG	91 min
9	s8	Movie	Sankofa	Haile Gerim	Kofi Ghanaba	United State	24-Sep-21	1993	TV-MA	125 min
10	s9	TV Show	The Great Br	Andy Devons	Mel Giedrow	United Kingd	24-Sep-21	2021	TV-14	9 Seasons
11	s10	Movie	The Starling	Theodore Me	Melissa McC	United State	24-Sep-21	2021	PG-13	104 min
12	s11	TV Show	Vendetta: Truth, Lies and	The Mafia			24-Sep-21	2021	TV-MA	1 Season
13	s12	TV Show	Bangkok Bre	Kongdat Kor	Sukollawat	Kanarot, Susha	23-Sep-21	2021	TV-MA	1 Season
14	s13	Movie	Je Suis Karl	Christian Sch	Luna Wedler	Germany, Cz	23-Sep-21	2021	TV-MA	127 min
15	s14	Movie	Confessions	Bruno Garot	Kara Castanho,	Lucca Picc	22-Sep-21	2021	TV-PG	91 min
16	s15	TV Show	Crime Stories:	India Detectives			22-Sep-21	2021	TV-MA	1 Season
17	s16	TV Show	Dear White People		Logan Brown	United State	22-Sep-21	2021	TV-MA	4 Seasons
18	s17	Movie	Europe's Mo	Pedro de Echave Garcí	Pablo Azorín		22-Sep-21	2020	TV-MA	67 min
19	s18	TV Show	Falsa Identidad		Luis Ernesto	Mexico	22-Sep-21	2020	TV-MA	1 Seasons
20	s19	Movie	Intrusion	Adam Salky	Freida Pinto,	Logan Marsh	22-Sep-21	2021	TV-14	94 min
21	s20	TV Show	Jaguar		Blanca Suárez,	Iv'n Mar	22-Sep-21	2021	TV-MA	1 Season
22	s21	TV Show	Monsters Ins	Olivier Megaton			22-Sep-21	2021	TV-14	1 Season
23	s22	TV Show	Resurrection: Ertuğrul		Engin Altan [	Turkey	22-Sep-21	2018	TV-14	5 Seasons
24	s23	Movie	Avval Shanm	K.S. Ravikum	Kamal Hassan,	Meena, Ge	21-Sep-21	1996	TV-PG	161 min
25	s24	Movie	Gol Gol Cor	Alex Woo, St	Maisie Benson,	Paul Gilla	21-Sep-21	2021	TV-Y	61 min

Netflix is good and favorite streaming platform of the increasingly large bunch.It has a huge selection of movies and TV shows old and newtons of high-quality original programs, and easy to navigate interface.

Netflix, Inc. is an American streaming serviced production company based in Los gatos, California. Founded on August 29, 1997 by Reed Hastings and Marc Hastings, Scotts valley, California it offers a film and television series library through distribution deals as well as its own productions, known as Netflix Originals.

As of June 30, 2022, Netflix had 220.7 million subscribers worldwide, including 73.3 million in the United States and Canada, 73.0 million in Europe, the Middle east and Africa, 39.6 million in Latin America and 34.8 million in the Asia Pacific region. It is available worldwide aside from Mainland China, Syria, North Korea, and Russia. Netflix

has played a prominent role in independent film distribution, and it is a member of the Motion Picture Association (MPA).

The key attributes i would like to explore are Types, TV shows, Director, title, country, data_added. The key questions are

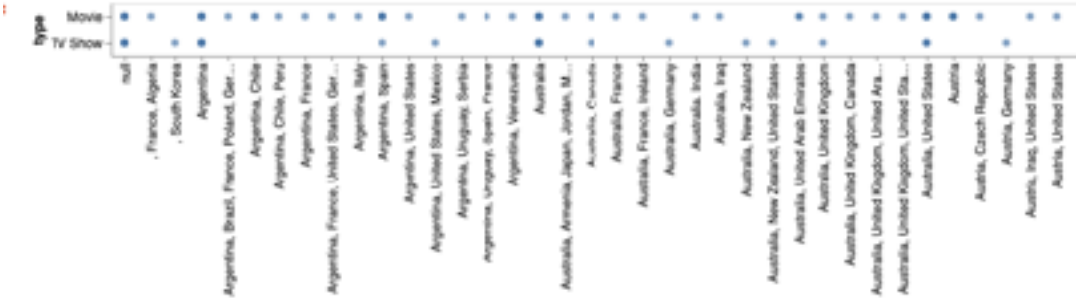
1. How many titles of Movies and TV shows in a year 2021?
2. Duration within 1 hour movies and TV shows?
3. What type of listed_in in the year of 2021?

Screenshots of and/or a link to your visualization implementation.

Let's start the visualization about Netflix dataset.

Number of shows based on types like (TV Shows and Movies) listed below as bar plot.

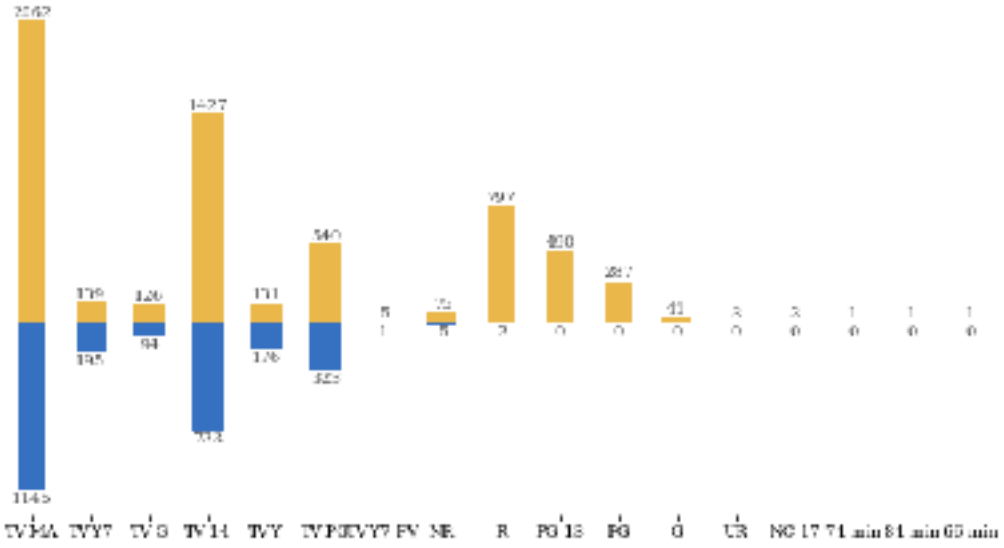
Top 10 countries to be watched in Netflix

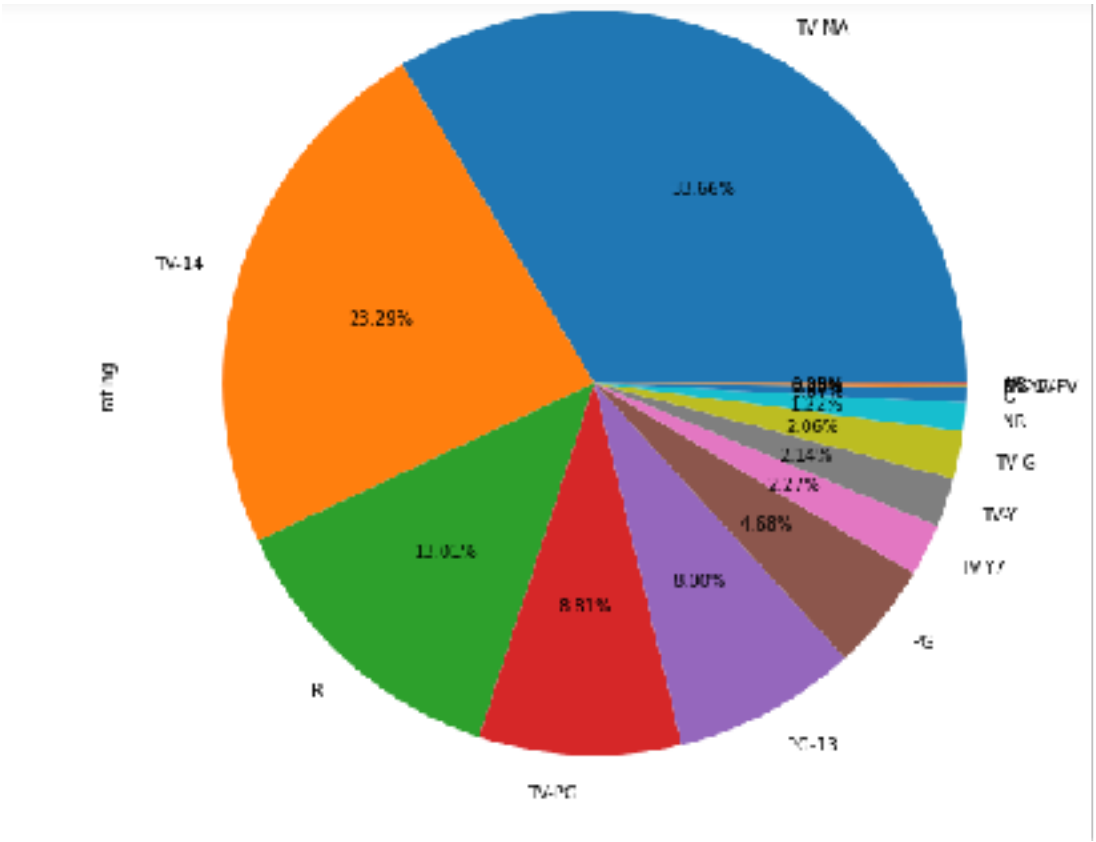


Rating distribution by Film & TV Show

We observe that some ratings are only applicable to Movies.
The most common for both Movies & TV Shows are TV-MA and TV-14.

Movie | TV Show





Most watched type as Documentaries in Netflix.

Provided link to Data visualizations of Netflix dataset
http://localhost:8888/notebooks/Final_Project_5304.ipynb#

A summary of the key elements of your design and accompanying justification

1.Task

- **Why is a task pursued? (goal)**

Analyzing type and country listed in datasets.

- **How is a task conducted? (means)**

The task conducted using bar plot Altair function. Organize the data by grouping movies and Tv shows. End user can see how much differentiated between countries and regions of subscribed lists.

- **What does a task seek to learn about the data? (characteristics)**

We observe that some ratings are only applicable to Movies.
The most common for both movies and TV shows are TV-Ma and TV-14.
United states is most popular on Netflix Movies and TV shows being taking unto 69% and 31% of the overall top 10 slots on average since 2020.

- **Where does the task operate? (target data)**

The most prolific producers of content for Netflix are, primarily, the USA, with India and UK a significant distance behind. It makes sense that the USA produces the most content as, after-all, Netflix is a Us company.
Some of the Insights regarding types listed such as Movies and TV shows vary accordingly with regions, it is really interesting to see how the split of Movies and TV South Korea dominated by TV shows and India dominated by Movies.shows varies by country.

- **When is the task performed? (workflow)**

Number of people watched documentaries most and movies. Information which we give as input of country and type listed in datasets used a s bar chart using Altair. Information as output as list of countries watched Movies and TV shows.

- **Who is executing the task? (roles)**

Data scientists as principle. To have great perspective of data visualizations model.

2. Task

- **Why is a task pursued? (goal)**

Rating distribution by Movies and TV shows.

- **How is a task conducted? (means)**

The task conducted using bar plot Altair function. Organize the data by grouping ratings and type. sorted index with Movies by colts and TV shows by label, colors. End user can see how much differentiated between countries and
regions of ratings.

- **What does a task seek to learn about the data? (characteristics)**

We observe that some ratings are only applicable to Movies. The most common for both Movies & TV shows are TV-MA and TV-14.

- **Where does the task operate? (target data)**

The most prolific producers of content for Netflix are, primarily, the USA, with India and UK a significant distance behind. It makes sense that the USA produces the most content as, after-all, Netflix is a Us company.
Netflix over several years, Things begin to pick up in 2015 and then there is a rapid increase from 2016.It looks like content additions slowed down in 2020, likely due to the COVID-19 pandemic.

- **When is the task performed? (workflow)**

Number of people watched documentaries most and movies. Information which we give as input of country and type listed in datasets used a s bar chart using Altair. Information as output as list of countries watched Movies and TV shows.

- **Who is executing the task? (roles)**

Data scientists as principle. To have great perspective of data visualizations model.

3.Task

- **Why is a task pursued? (goal)**

There are 20 types in Netflix dataset genre for Movies. Movie ratings.

- **How is a task conducted? (means)**

The task conducted using mat plot library plot Altair function. Organize the data by grouping release_year and type.Sorted duration of the year regions of ratings.

- **What does a task seek to learn about the data? (characteristics)**

We observe that some movie ratings have increased duration time from 90 min to 124

min as year increases. From 1940 till 2020 duration increased gradually.

- **Where does the task operate? (target data)**

There are high number of shows under
TV-MA(Mature audience only) followed by
TV-14(Parents strongly cautioned),
TV-PG(Parental guidance suggested) and
R(Restricted,Children).

- **When is the task performed? (workflow)**

We can see most of the shows in the list belong to restricted and matured audience , thus we can say that Netflix has more of matured content and should be restricted for children.

- **Who is executing the task? (roles)**

Data scientists as principle. To have great perspective of data visualizations model.

A discussion of your final evaluation approach, including the procedure, people recruited, and results. Note that, due to the difficulty of recruiting experts, you can use colleagues, friends, classmates, or family to evaluate your designs if experts or others from your target population are unavailable.

- **The target question you want to answer**

Does the visualization provide insight into Netflix subscribers?

- **The people you would recruit to answer that question**

Data scientists.
Due to online course, got help from my family members.

- **The kinds of measures you would use to answer your data (e.g., insight depth, use cases, accuracy) and what these measures would tell you about the core question**

Used many python packages as core coding measures and visualized many plots with some research on online datasets and feedbacks about Netflix.
These measures have the ability to predict future viewers accordingly.

- **The approach you will use to answer that question (e.g., a journaling study, a formal experiment, etc.)**

Approach which i used was journaling study for qualitative feedback and experimental design for quantitative feedback.

- **How you would instantiate those methods (i.e., what would your participants**

do?)

I would allow my participants to explore the data and visualization tools , then get the structured feedback from them.

I follow them through the usage of tools and easy to access the information provided visually.

- **What criteria would you use to indicate that your visualization was successful**

Qualitative feedback from users about the system adoption.

Able to quickly grasp the information needed for future analyses.

A synthesis of your findings, including what elements of your approach worked well and what elements you would refine in future iterations.

Data visualization with Altair approach was useful and visual treat to others.

We can see real-time datasets as visualizations using Altair python packages.

Showing the viewers how the tool can be used to visualize and what to look for it through. Without proper introduction or User manual is easy to understand the tool kit provided.

For future prediction , we can use the measure to explore the interactions further.

we can explore the data using thinkalouds and quantitative experimental designs.