

**VIETNAM NATIONAL UNIVERSITY – HO CHI MINH
CITY INTERNATIONAL UNIVERSITY**

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING



DATA SCIENCE & DATA VISUALIZATION

PROJECT PROPOSAL

Course by Dr. Tran Thanh Tung

TOPIC: NETFLIX MOVIES & TV SHOWS

TEAM MEMBERS

NGUYEN QUANG SANG	ITDSIU21113
DO THANH DAT	ITDSIU21079
PHAN MANH SON	ITDSIU21116
VAN PHU MINH SANG	ITDSIU21112



TABLE OF CONTENTS

I. INTRODUCTION.....	3
II. DATA GATHERING.....	6
III. DATA PROCESSING.....	7
IV. OBJECTIVES AN FEATURES.....	8
V. ENVIRONMENT/TOOLS.....	10
VI. PROJECT TIMELINE.....	11
VII. PROTOTYPE DESIGN.....	13
VIII. REFERENCES.....	15

I. INTRODUCTION

1. Abstract

Netflix, a prominent global streaming platform, has transformed the way we engage with entertainment. Its extensive collection of films and television series provides a multitude of options for audiences worldwide.

Netflix has achieved significant expansion, establishing itself as a dominant player in the streaming industry due to its remarkable growth.

By the beginning of the second quarter of 2022, Netflix had garnered approximately 222 million international subscribers across more than 190 countries (excluding China, Crimea, North Korea, Russia, and Syria), underscoring its widespread global popularity. The platform's remarkable success can be attributed to its extensive international expansion efforts, which include localizing content through subtitles and dubbing in multiple languages, ensuring accessibility to a diverse and global audience.

2. Background and Motivation

This is a mini project of a group of students on Data Science and Data Visualization subjects.

This project aims to create **an interactive data visualization** to explore the intriguing **trends, patterns** and **insights** hidden within Netflix's content landscape.

Using the capabilities of **Python** and its data analysis libraries, we delve into Netflix's extensive content library to unearth valuable **insights** regarding content additions, duration distributions, genre correlations, and even the most frequently used words in titles and descriptions.



Fig 1. Netflix Logo (taken from brand.netflix.com)

With the aid of in-depth code examples and visual representations, we dissect **Netflix's content ecosystem**, offering a new perspective on its evolution. Our analysis delves into release patterns, seasonal fluctuations, and viewer preferences, all contributing to a deeper understanding of the content dynamics within Netflix's extensive universe.

3. Goal

- ❖ Initial **data preparation and analysis** to identify the essential attributes for **a clear visualization**.
- ❖ Providing a **variety chart** options and **dashboard layout** designs to assist users uncover **trends, patterns** and **insights** hidden within Netflix's content landscape.

- ❖ Providing **essential insights** to address **questions derived from the dataset** and using them to resolve issues aimed at **boosting Netflix's growth**.
- ❖ Continuously managing and adjusting **charts using D3.js libraries** to align with potential future modifications.

II. DATA GATHERING

1.About

- ❖ About this Dataset: Netflix is one of the most popular media and video streaming platforms. They have over **8000 movies or tv shows** available on their platform, as of mid-2021, they have **over 200M Subscribers** globally. This tabular dataset consists of listings of all the movies and tv shows available on Netflix, along with details such as - cast, directors, ratings, release year, duration, etc.
- ❖ This dataset is collected from:
Flixable(<https://flixable.com/>) which is a third-party Netflix search engine, and available on **Kaggle**(<https://www.kaggle.com/https://www.kaggle.com/datasets/shivamb/netflix-shows>) website for free.

2.Data files

- ❖ The dataset file is in CSV(Excel) format with null-values waiting to be deadling

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description									
1	s1	Movie	Dick Johnson	Kirsten Johnson	United States	#####	2020	PG-13	90 min	Documentary	As her father nears the end of his life, filmmaker Kirsten Johnson stages his death in inventive and										
3	s2	TV Show	Blood & Water	Ama Qam	South Africa	#####	2021	TV-MA	2 Seasons	International	After crossing paths at a party, a Cape Town teen sets out to prove whether a private-school swimmer										
4	s3	TV Show	Ganglands	Julien Lec	Sami Bouajila, Tracy	#####	2021	TV-MA	1 Season	Crime TV	To protect his family from a powerful drug lord, skilled thief Mehdi and his expert team of robbers										
5	s4	TV Show	Jailbirds	New Orleans		#####	2021	TV-MA	1 Season	Docuseries	Feuds, flirtations and toilet talk go down among the incarcerated women at the Orleans Justice Cen										
6	s5	TV Show	Kota Factory		Mayur Mo	India	#####	2021	TV-MA	2 Seasons	International	In a city of coaching centers known to train India's finest collegiate minds, an earnest but unexc									
7	s6	TV Show	Midnight	Mike Flan	Kate Siegel, Zach Gil	#####	2021	TV-MA	1 Season	TV Drama	The arrival of a charismatic young priest brings glorious miracles, ominous mysteries and renewed r										
8	s7	Movie	My Little	Robert Cu	Vanessa Hudgens, Ki	#####	2021	PG	91 min	Children &	Equestria's divided. But a bright-eyed hero believes Earth Ponies, Pegasi and Unicorns should be pa										
9	s8	Movie	Sankofa	Haile Geri	Kofi Ghan	United States	#####	1993	TV-MA	125 min	Dramas, Ir	On a photo shoot in Ghana, an American model slips back in time, becomes enslaved on a plantation									
10	s9	TV Show	The Great	Andy Devi	Mel Giedr	United Kingdom	#####	2021	TV-14	9 Seasons	British TV	A talented batch of amateur bakers face off in a 10-week competition, whipping up their best dishe									
11	s10	Movie	The Starlin	Theodore	Melissa M	United States	#####	2021	PG-13	104 min	Comedies	A woman adjusting to life after a loss contends with a feisty bird that's taken over her garden and an									
12	s11	TV Show	Vendetta: Truth, Lies and The Mafia			#####	2021	TV-MA	1 Season	Crime TV	Sicily boasts a bold "Anti-Mafia" coalition. But what happens when those trying to bring down organ										
13	s12	TV Show	Bangkok	E Kongkiat	Sukollawat Kanarot,	#####	2021	TV-MA	1 Season	Crime TV	Struggling to earn a living in Bangkok, a man joins an emergency rescue service and realizes he must										
14	s13	Movie	Je Suis Kai	Christian	Luna Wed	Germany,	#####	2021	TV-MA	127 min	Dramas, Ir	After most of her family is murdered in a terrorist bombing, a young woman is unknowingly lured in									
15	s14	Movie	Confessio	Bruno Gar	Klara Castanho, Lucc	#####	2021	TV-PG	91 min	Children &	When the clever but socially-awkward Tetã joins a new school, she'll do anything to fit in. But the										
16	s15	TV Show	Crime Stories: India	Detectives		#####	2021	TV-MA	1 Season	British TV	Cameras following Bengaluru police on the job offer a rare glimpse into the complex and challengin										
17	s16	TV Show	Dear White People	Logan Bro	United States	#####	2021	TV-MA	4 Seasons	TV Comedy	Students of color navigate the daily slights and slippery politics of life at an Ivy League college that's										
18	s17	Movie	Europe's	Pedro de	Echave Garc	Ã-a, Pabl	#####	2020	TV-MA	67 min	Documentary	Declassified documents reveal the post-WWII life of Otto Skorzeny, a close Hitler ally who escaped									
19	s18	TV Show	Falsa	Identidad	Luis Ernes	Mexico	#####	2020	TV-MA	2 Seasons	Crime TV	Strangers Diego and Isabel flee their home in Mexico and pretend to be a married couple to escape									
20	s19	Movie	Intrusion	Adam Sall	Freida Pinto, Logan	#####	2021	TV-14	94 min	Thrillers	After a deadly home invasion at a couple's new dream house, the traumatized wife searches for										
21	s20	TV Show	Jaguar		Blanca Su	Ãirez, IvÃ	#####	2021	TV-MA	1 Season	International	In the 1960s, a Holocaust survivor joins a group of self-trained spies who seek justice against Nazis fi									
22	s21	TV Show	Monsters	Olivier	Megaton		#####	2021	TV-14	1 Season	Crime TV	In the late 1970s, an accused serial rapist claims multiple personalities control his behavior, setting									
23	s22	TV Show	Resurrection: Ertugr	Engin Alta	Turkey	#####	2018	TV-14	5 Seasons	International	When a good deed unwittingly endangers his clan, a 13th-century Turkish warrior agrees to fight a s										
netflix titles																					

Fig 2. Quick view of dataset

III. DATA PROCESSING

The data will undergo a process of cleaning, standardization, and conversion into a format that is appropriate for visualization. We will address any missing data points and outliers, as well as carry out any required calculations for rates and percentages at this stage.

- ❖ Using some basic calculations about the statistic in Python/R to deal with the missing values.

- ❖ Providing some basic plot (pie chart, bar chart, heatmap,...) and support library to visualize the missing values.
- ❖ Remove unnecessary columns, fixing the data type. Finding and filling some important necessary missing values

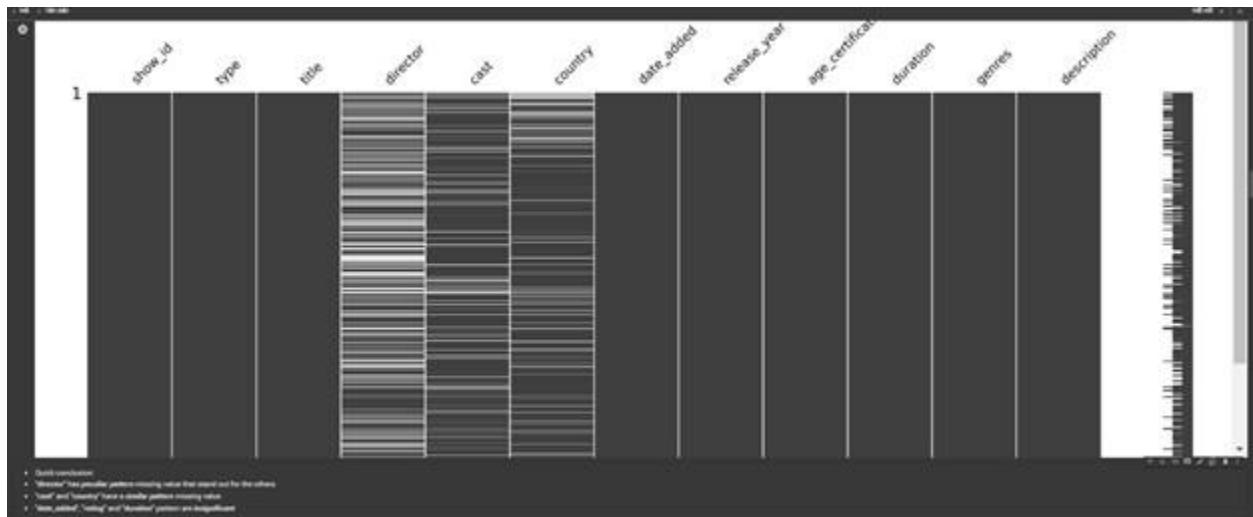


Fig 3. (Quick visualize of missing values by **missingno** library)

IV. OBJECTIVES AND FEATURES

1.Objectives: The Primary Questions to Answer

- Number of Movies vs Number of TV shows?
- Distribution of Content Types (Percentage of Movies and TV Shows)?
- Top 10 Countries Where Netflix is Popular?
- Top 10 Categories by Movie/TV Shows Count?
- Movies & TV Shows Added Over Time?
- Content Added by Month?
- Distribution of Ratings?
- Distribution of Movie Lengths and TV Show Episode Counts?

- The Trend of Movie/TV Show Lengths Over the Years?
- Most Common Words in Titles and Descriptions?
- Duration Distribution for Movies and TV Shows?
- If a content producer aims to release their content at a time when the lowest volume of new content is being added, which month should they choose?
- TV shows with the largest number of seasons?
- Lowest number of seasons?
- How many Netflix Shows/ Movies are made from books as their storylines?

2.Features

- An interface that encourages interaction and is easy for users to navigate.
- Visualization of geospatial data with the ability to zoom and pan.
- Analytical tools for comparing data across various regions.
- Interactive data tooltips to provide detailed information.
- Forecasting models for anticipating future trends.
- Seamless integration with vaccination appointment scheduling.
- A mobile application version for convenient on-the-go access.
- Support for multiple languages.
- Options for sharing on social media platforms.

V. ENVIRONMENT/ TOOLS

- ❖ Google Colab/Jupyter Notebook (using Python) for data processing and analysis.
- ❖ Visual Studio Code for CSS/JS for interactive visualization website building.
- ❖ D3js support to the website building.
- ❖ Git and Github for Version Control for managing projects.

➤ GITHUP LINK ([HERE](#))

VI. PROJECT TIMELINE

Our initial intention was to divide the project into **five phases**, each aligned with a distinct phase, task, and project manager.

STAGE	TASK	MEMBER	WEEK
PLANNING	Commence the project, assemble the team, and establish the GitHub repository	Dat, Quang Sang	1 - 2
	Individual topics research	Dat	
	Topic confirmation	All	
	Determine the goal and the purpose of the analysis	All	
	Research for Analysis	Quang Sang, Minh Sang	

	Research for Website Building	Dat, Son	
DATA GATHERING	Collection the dataset	Quang Sang, Dat	2 - 3
	Research for the references and documentation	All	
	Identify the appropriate analysis method for the project.	All	
	Group discussion before the process	All	
DATA PROCESSING	Identify the missing values and outliers	Quang Sang	3 - 4
	Remove unnecessary missing value	Minh Sang, Quang Sang	
	Analyze the factors that influenced the accuracy and precision of the findings.	Minh Sang, Quang Sang	
	Fill in and fixing missing value	Minh Sang, Quang Sang	
	Clean the data	Minh Sang. Quang Sang	
CONCEPTUAL DESIGN	Design UI/UX	Dat, Son	4 - 8
	Select a theme and chart type for data visualization.	Dat, Son	
	Sketch the diagrams and user flow	Dat, Son	

	Prototype design and development	Dat, Son	
	Finishing the design	Dat, Son	
IMPLEMENTING DESIGN	Structure the web layout	Dat, Sang	8 - 13
	Coding the web function and generate charts	All	
	Deployment of the visualization tool	All	
	Final testing and bug fixes	All	
PRESENTATION	Final report	Quang Sang, Dat	14
	Presentation slides	Minh Sang, Son	

VII. PROTOTYPE DESIGN

Prototype Design 1: Comparative Dashboard

The prototype will offer a dashboard interface enabling users to compare Netflix Movies and TV shows data among various countries, genres, or target ages. It will incorporate charts, graphs, and tables to facilitate a thorough analysis.

Prototype Design 2: Data Visualization on website

This prototype focuses on creating a data visualization web page using HTML, CSS, JavaScript (JS), and the D3.js library. D3.js is a powerful library for creating dynamic and interactive data visualization directly within a web page.

Prototype Design 3: Get Insights and Conclusion

The "Get Insight and Conclusion" prototype serves as the final stage of a data visualization project, where you aim to provide users with actionable insights and conclusions drawn from the data. This prototype is not just about presenting data; it's about guiding users through the data analysis process and helping them make informed decisions or draw meaningful conclusions.

VIII. REFERENCES

- [1] Flixable. " Full List of Movies and TV Shows on Netflix". Flixable website, n.d. <https://flixable.com/>
- [2] SHIVAM BANSAL. "Netflix Movies and TV Shows". Kaggle website, 2021. <https://www.kaggle.com/datasets/shivamb/netflix-shows>

THE END