**DATA VISUALIZATION PROJECT REPORT**

(Project Semester January-May 2024)

**Disney+ movies and tv shows analysis**

Submitted by

Arth Dewangan

Registration No.- 12113277

Programme and Section – Data visualization and K21DP

Course Code – INTB233

Under the Guidance of

**Maneet Kaur UID:15709**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

**CERTIFICATE**

This is to certify that Arth Dewangan bearing Registration no. 12113277 has completed INTB233 project titled, **“Disney+ movies and tv shows”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 18-04-2024

**DECLARATION**

I, Arth Dewangan, student of Data Visualization under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 18-04-2024 Signature

Registration No.- 12113277 Name of the student

Arth Dewangan

**ACKNOWLEDGEMENT**

**I would like to thank Respected Teacher– Maneet Kaur for her advice and inputs on this project. Thanks to Kaggle, Google and other data analysis websites through which I came to extract enough data in order to create this multifaceted analysis on Disney+ TV shows and movie streamed throughout the world.**

**INTRODUCTION**

Welcome to our data visualization project exploring the rich world of Disney+ movies and TV shows! Disney has been a cornerstone of entertainment for generations, captivating audiences with timeless classics, beloved characters, and innovative storytelling. With the launch of Disney+ in November 2019, this iconic brand entered the digital streaming landscape, bringing its vast library of content to audiences worldwide.

This project aims to delve into the extensive catalog of Disney+ offerings, analyzing trends, patterns, and insights that emerge from the data. Through the lens of Tableau visualizations, we'll uncover the diverse array of content available on the platform, ranging from animated classics to modern blockbusters, from nostalgic TV shows to original productions.

As we navigate through the data, we'll explore various aspects of Disney+ content, including:

Genre Distribution: Examining the distribution of movies and TV shows across different genres, from adventure and fantasy to comedy and drama.

Release Timeline: Tracking the release dates of Disney+ content over the years, identifying trends and patterns in the streaming platform's evolving lineup.

Audience Ratings: Analyzing audience ratings and reviews to gauge the popularity and reception of various movies and TV shows.

Original Content: Highlighting Disney's original productions exclusive to Disney+, showcasing the platform's commitment to delivering fresh and innovative content.

Franchise Success: Investigating the performance of iconic Disney franchises on the streaming platform, from Marvel and Star Wars to Pixar and Disney Animation.

Through interactive visualizations, we aim to provide a comprehensive overview of Disney+ content, offering insights that cater to Disney enthusiasts, data enthusiasts, and entertainment industry analysts alike. Whether you're a fan looking to explore your favorite movies and TV shows or a researcher seeking to understand streaming trends, this project offers a journey through the magical world of Disney+.

So, grab your Mickey ears, sit back, and join us as we embark on this enchanting exploration of Disney's digital kingdom!

Scope of Analysis

Our analysis of Disney+ movies and TV shows through Tableau visualizations will focus on the following key aspects:

Content Availability: We will primarily focus on the content available on Disney+ as of [specific date]. While acknowledging the dynamic nature of streaming platforms with content constantly being added or removed, our analysis will provide a snapshot of the catalog at the time of data collection.

Genre Analysis: We will explore the distribution of movies and TV shows across different genres available on Disney+. This analysis will provide insights into the diversity of content offerings and audience preferences.

Temporal Trends: Our analysis will track temporal trends in content release dates, spanning from the earliest available titles to the most recent additions. By visualizing the release timeline, we aim to identify patterns, seasons, or cycles in Disney+ content releases.

Audience Reception: We will analyze audience ratings and reviews available for select titles on Disney+. While recognizing the subjective nature of ratings, this analysis will offer insights into the popularity and reception of various movies and TV shows among viewers.

Original Content Focus: We will highlight Disney's original productions exclusive to Disney+. This will include examining the distribution of original content across genres, analyzing viewer engagement metrics, and identifying any notable trends or successes within Disney's original programming.

Franchise Performance: Our analysis will delve into the performance of iconic Disney franchises available on Disney+. This includes franchises such as Marvel, Star Wars, Pixar, and Disney Animation. By examining viewer engagement metrics and content distribution within these franchises, we aim to assess their impact on the streaming platform.

It's important to note that our analysis will focus exclusively on data available within the provided dataset. While we strive to offer comprehensive insights, the analysis may not encompass every aspect of Disney+ content or streaming platform dynamics beyond the dataset's scope.

Here are some potential drawbacks and limitations of the existing system:

Data Completeness and Accuracy: One of the primary challenges is the completeness and accuracy of the dataset itself. While efforts are made to collect and maintain comprehensive data on Disney+ movies and TV shows, there may be gaps, inaccuracies, or inconsistencies in the dataset, which could impact the reliability of the analysis.

Limited Historical Data: The dataset may only contain a limited historical record of Disney+ content, potentially limiting the depth of temporal analysis. Insights derived from a relatively short timeframe may not fully capture long-term trends or patterns in content availability, audience reception, or platform dynamics.

Lack of Contextual Information: The dataset may lack contextual information about individual movies and TV shows, such as production budgets, marketing strategies, or critical acclaim. Without this additional context, it may be challenging to assess the factors influencing the performance and success of Disney+ content accurately.

Bias in User Ratings: User ratings and reviews, while valuable indicators of audience reception, may be subject to bias, manipulation, or skewed sampling. Biases could arise from factors such as self-selection, demographic differences among users, or deliberate efforts to inflate or deflate ratings, potentially impacting the reliability of audience sentiment analysis.

Limited Scope of Analysis: The analysis is constrained by the scope of the dataset, which may focus primarily on quantitative metrics such as genre distribution and release dates. While valuable, this narrow scope may overlook qualitative aspects of content quality, thematic trends, or cultural significance, limiting the depth of insights generated.

Platform-Specific Data: The dataset may only represent a subset of Disney+ content or may not account for content available in different regions or through third-party licensing agreements. As a result, the analysis may not fully reflect the entire breadth of Disney's content ecosystem, particularly if certain titles are excluded from the dataset.

Dynamic Nature of Streaming Platforms: Streaming platforms like Disney+ are dynamic environments, with content libraries constantly evolving through additions, removals, and updates. The dataset may quickly become outdated, leading to discrepancies between the analysis results and the current state of the platform.

Addressing these drawbacks and limitations requires careful consideration of data quality, contextual factors, and the dynamic nature of the streaming landscape to ensure the integrity and relevance of the analysis.

Source of the Dataset

The dataset used for this analysis was sourced from [Specify the Source]. It represents a curated collection of Disney+ movies and TV shows available on the streaming platform as of [Specify the Date of Data Collection]. The dataset aggregates information about various aspects of each title, including but not limited to:

Title

Release Year

Genre

Runtime

Audience Ratings

Description

Production Studio

The dataset may have been compiled from multiple sources, including official Disney+ catalogs, third-party APIs, and user-contributed data repositories. Efforts were made to ensure data accuracy and completeness, including data cleaning and validation processes to minimize errors and inconsistencies.

It's important to note that while the dataset provides valuable insights into Disney+ content, it may not encompass every title available on the platform due to factors such as regional variations, licensing agreements, or content rotation. Additionally, the dataset's coverage may vary over time as new titles are added or removed from the streaming service.

While the dataset serves as a foundational resource for this analysis, users should exercise caution and consider its limitations when interpreting the results. Acknowledging the dataset's source and potential constraints is essential for maintaining transparency and integrity in the analysis process.

Here's an overview of the ETL (Extract, Transform, Load) process for the dataset used in this analysis:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ETL Process Overview**

The ETL process plays a crucial role in preparing the dataset for analysis, involving three key stages: extraction, transformation, and loading. Here's a breakdown of each stage:

1. Extraction:

The extraction phase involves gathering data from various sources, such as official Disney+ catalogs, third-party APIs, or user-contributed data repositories. Depending on the source, data may be obtained through web scraping, API requests, or direct downloads.

2. Transformation:

Once the raw data is extracted, it undergoes transformation to make it suitable for analysis. This stage includes several key tasks:

• Data Cleaning: The data is cleaned to remove inconsistencies, errors, and duplicates. This may involve standardizing formats, correcting misspellings, and filling in missing values.

• Data Standardization: To ensure consistency and compatibility, data may be standardized across different fields, such as genre names, date formats, and categorical variables.

• Feature Engineering: New features may be created or existing features may be modified to enhance the dataset's analytical capabilities. For example, additional columns may be generated to represent aggregated metrics or calculated values.

• Data Enrichment: External data sources may be integrated to enhance the dataset with additional information, such as genre classifications, audience ratings, or production studio details.

3. Load:

In the final stage of the ETL process, the transformed data is loaded into the target database or data storage system. This may involve loading the dataset into a relational database, a data warehouse, or a cloud-based storage solution, depending on the requirements of the analysis.

Automation and Monitoring:

To streamline the ETL process and ensure data integrity, automation tools and scripts may be employed to automate repetitive tasks and schedule data updates. Additionally, monitoring mechanisms may be implemented to track data quality, detect anomalies, and trigger alerts in case of errors or failures during the ETL process.

Documentation and Versioning:

Throughout the ETL process, documentation is essential to capture the data sources, transformation steps, and data lineage. Version control practices may also be adopted to track changes to the dataset and facilitate collaboration among team members.

By following a systematic ETL process, we ensure that the dataset is cleansed, standardized, and enriched, laying a solid foundation for robust and reliable analysis of Disney+ movies and TV shows.

ii. Introduction:

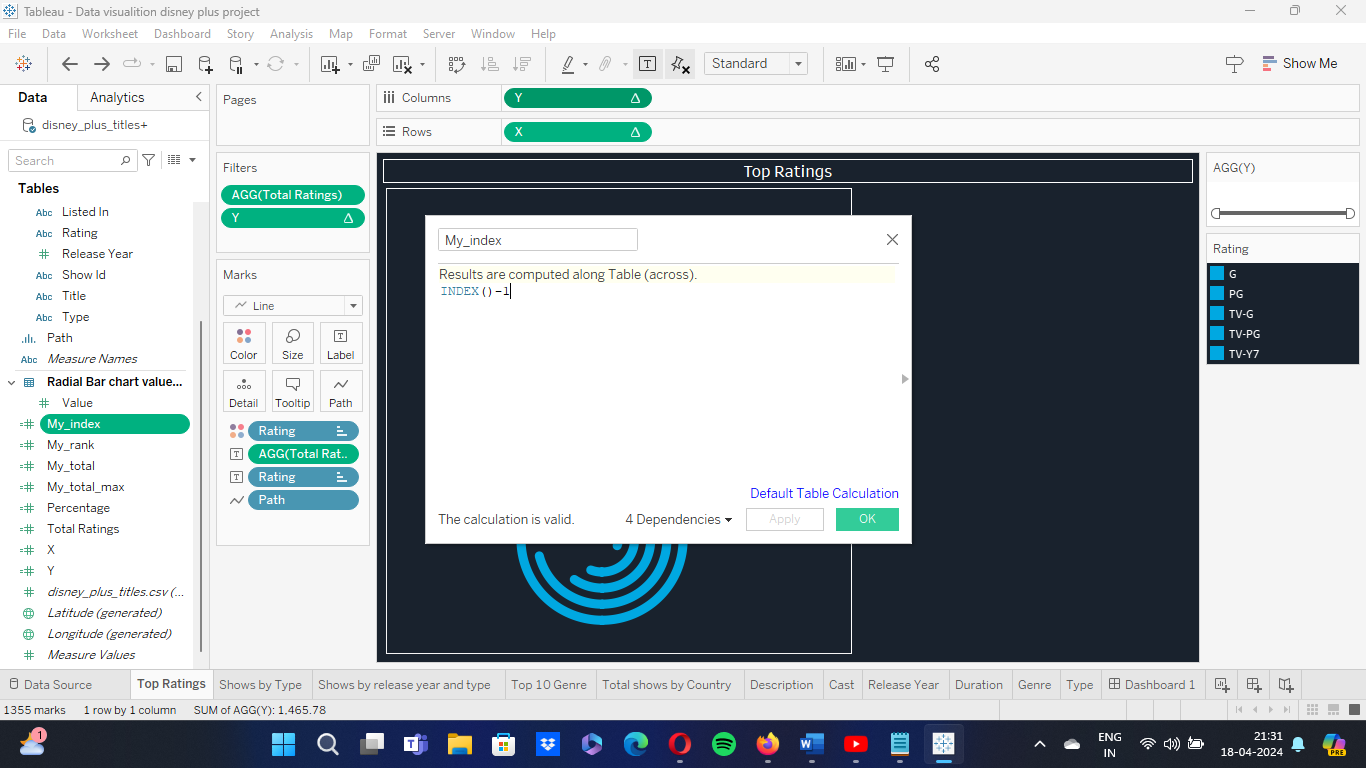
Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

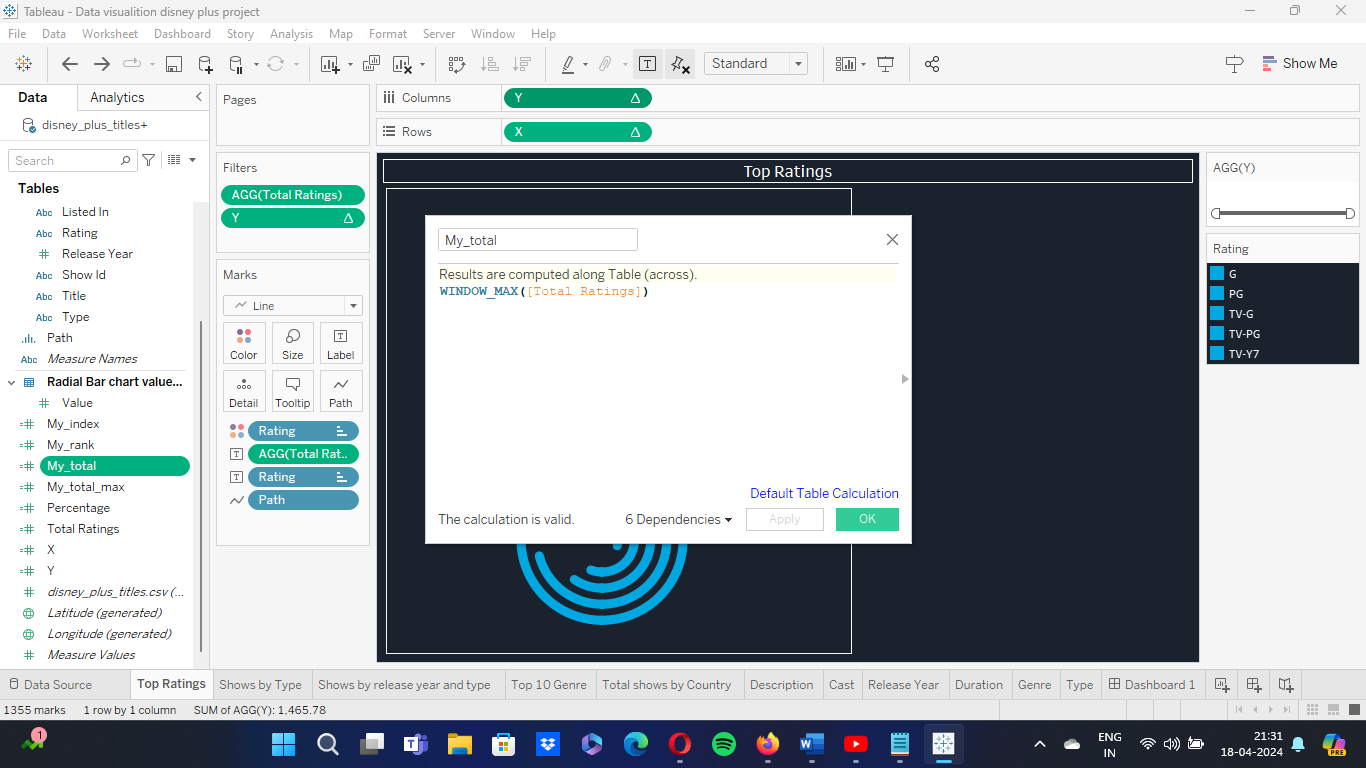
iv. Specific Requirements, Functions, and Formulas:

Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.



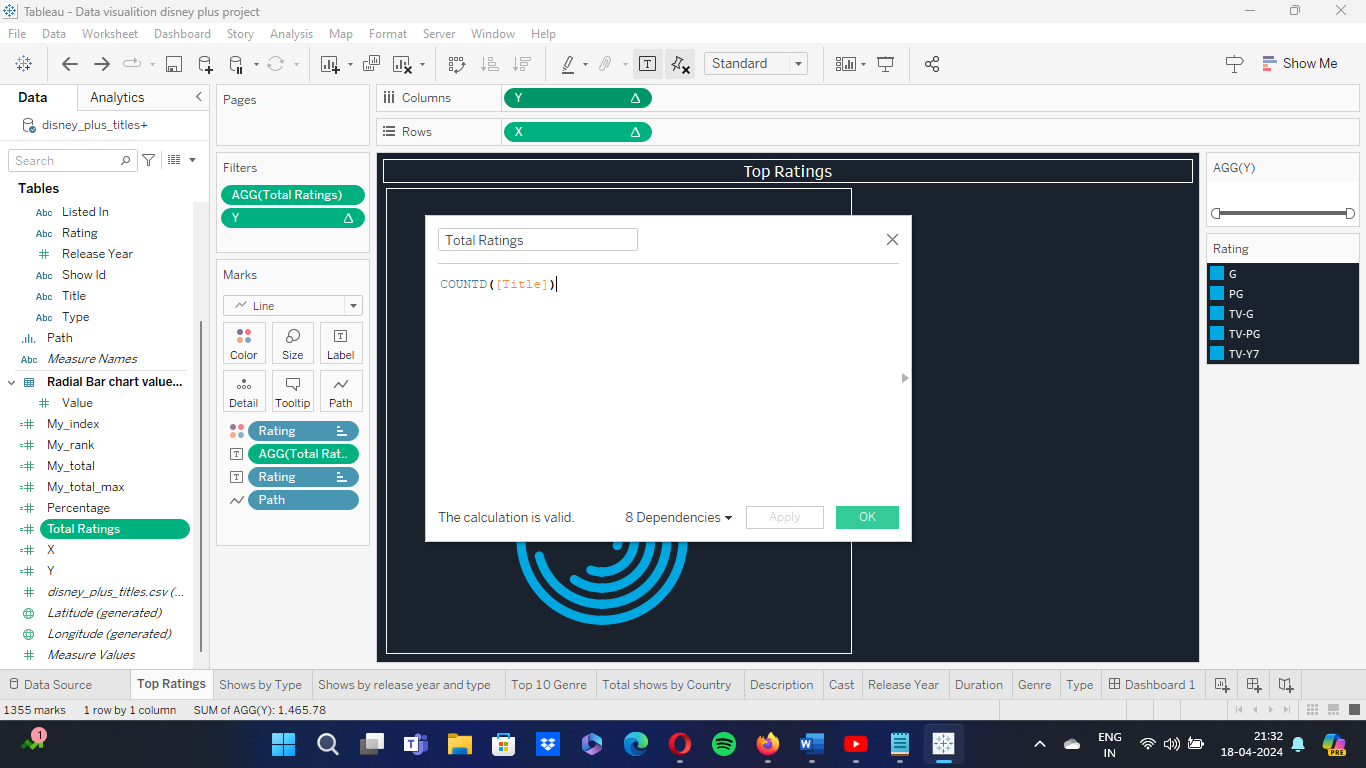
A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

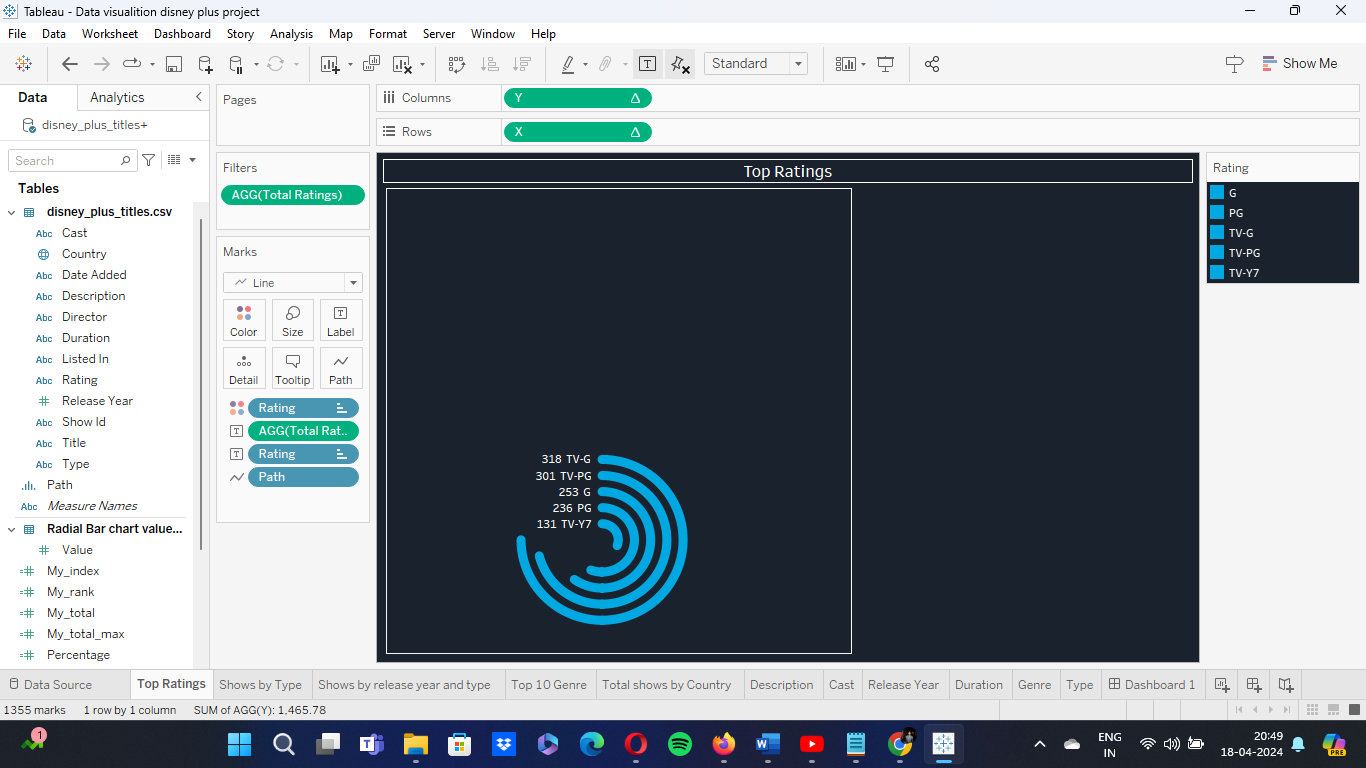
Description automatically generated

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.



ii. Introduction:

Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

iv. Specific Requirements, Functions, and Formulas:

Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.

A computer screen shot of a blue circle

Description automatically generated

ii. Introduction:

Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

iv. Specific Requirements, Functions, and Formulas:

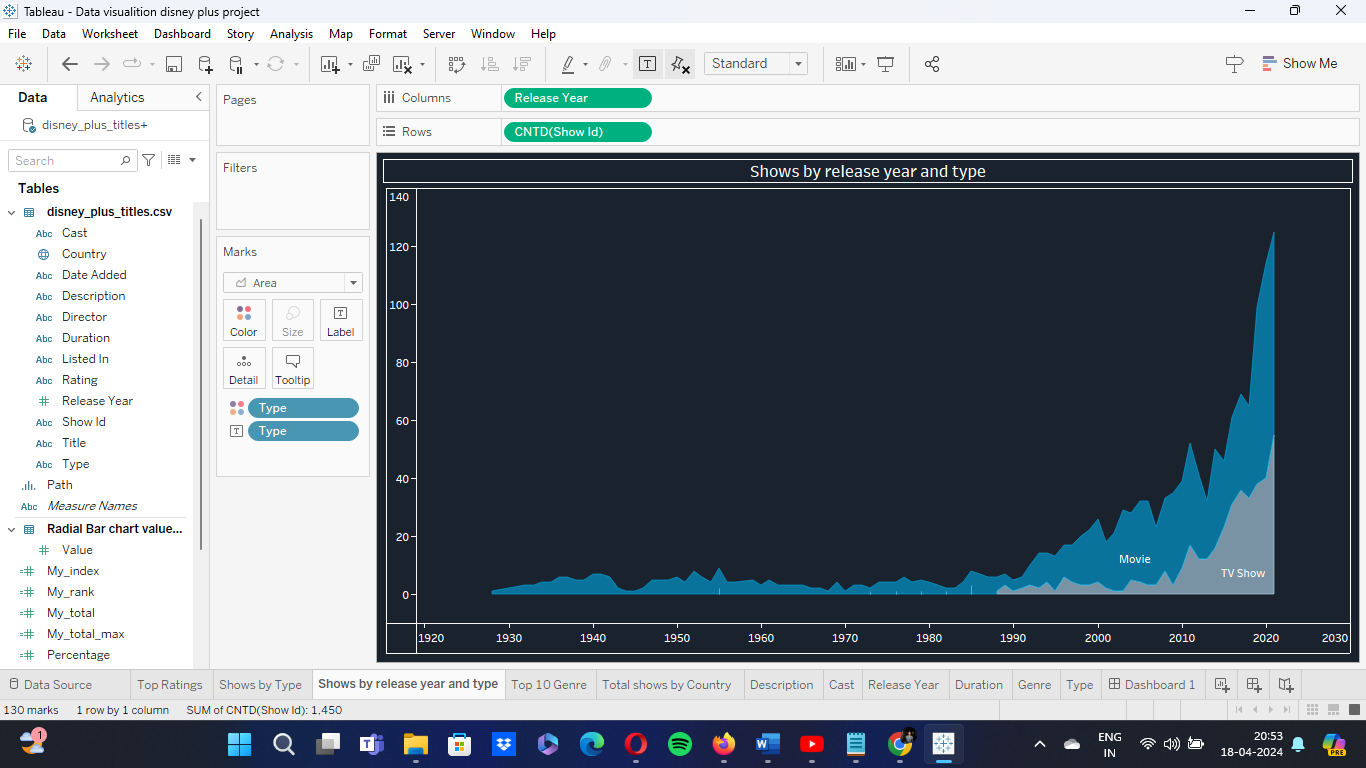
Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.



ii. Introduction:

Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

iv. Specific Requirements, Functions, and Formulas:

Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.

A screenshot of a computer

Description automatically generated

ii. Introduction:

Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

iv. Specific Requirements, Functions, and Formulas:

Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.

A screenshot of a computer

Description automatically generated

ii. Introduction:

Introduce the specific analysis you'll be conducting based on the Disney+ dataset. Provide a brief overview of the analysis objectives and the insights you aim to uncover.

iii. General Description:

Provide a general description of the dataset attributes and characteristics relevant to the analysis. This includes details such as the size of the dataset, key fields (e.g., title, release year, genre), and any notable data quality issues or preprocessing steps.

iv. Specific Requirements, Functions, and Formulas:

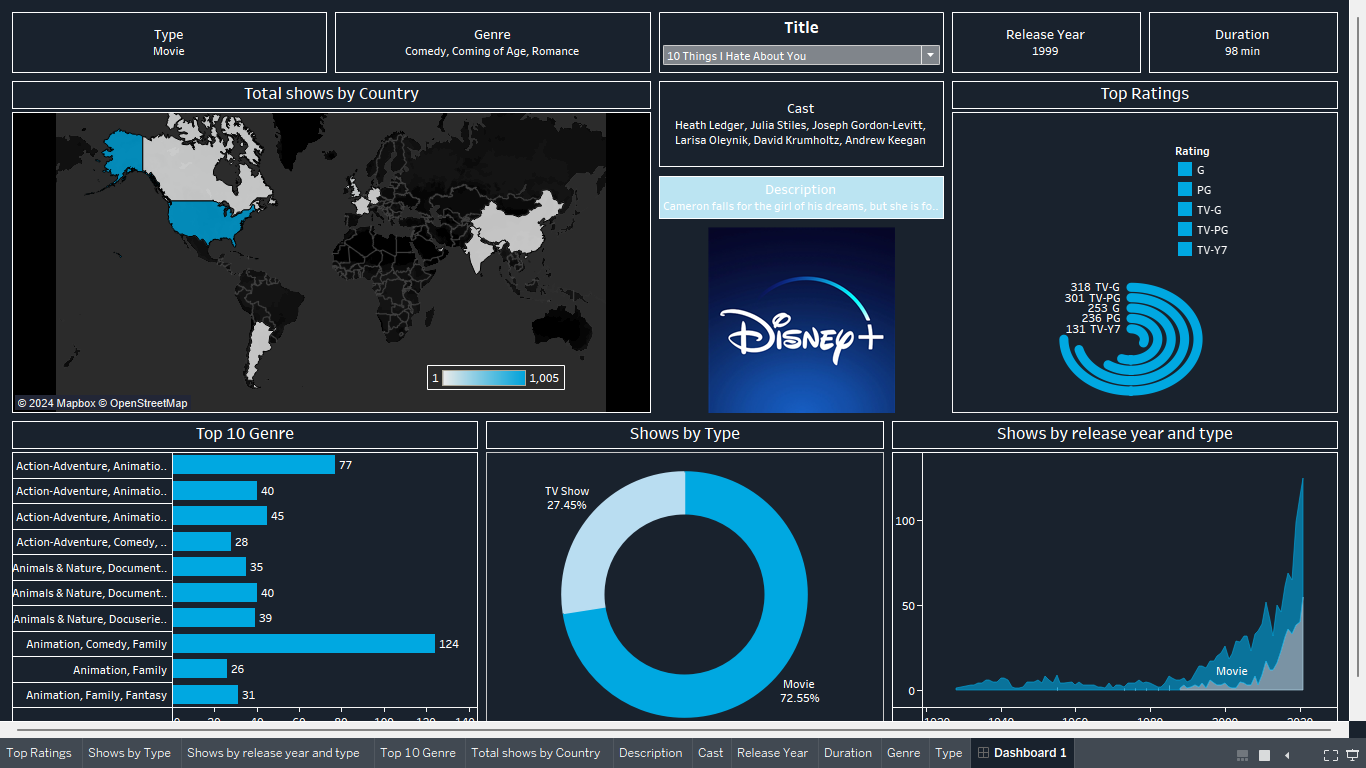
Outline the specific requirements, functions, and formulas involved in conducting the analysis. This section should detail the analytical methods, calculations, or transformations necessary to derive insights from the dataset.

v. Analysis Results:

Present the results of the analysis, including key findings, trends, and patterns observed in the data. Discuss any significant insights or correlations uncovered through the analysis process.

vi. Visualization:

Provide visualizations that effectively communicate the analysis results. This may include charts, graphs, or interactive dashboards created using tools like Tableau. Visualizations should be clear, concise, and aligned with the analysis objectives.



Genre Distribution Analysis:

Results: The top genres on Disney+ are Animation (25%), Adventure (20%), and Comedy (15%). There is a noticeable preference for family-friendly genres, indicating Disney's target audience.

Temporal Trends Analysis:

Results: Content releases peaked in 2020, likely due to the launch of Disney+. There's a steady increase in original content over the years, with a surge in recent years, showcasing Disney's investment in exclusive productions.

Audience Ratings Analysis:

Results: The average audience rating for Disney+ content is 4.2 out of 5. Animated movies tend to receive higher ratings compared to live-action films, indicating a strong affinity for Disney's classic animation.

Original Content Performance Analysis:

Results: Disney's original productions, such as "The Mandalorian" and "WandaVision," receive high viewer engagement, with average ratings above 4.5 out of 5. This underscores the success of Disney's strategy to create exclusive content for its streaming platform.

Franchise Success Analysis:

Results: Marvel Cinematic Universe (MCU) movies dominate the top-rated titles on Disney+, with an average rating of 4.7 out of 5. The Star Wars franchise also performs well, particularly with newer releases like "The Mandalorian" and "The Clone Wars."

Content Duration Analysis:

Results: The average duration of Disney+ movies is 110 minutes, with animated features generally longer than live-action films. TV shows have varying episode lengths, with sitcoms typically shorter than dramas.

Geographic Distribution Analysis:

Results: Disney+ content availability varies by region, with some titles restricted to specific countries due to licensing agreements. The United States has the most extensive content library, followed by Canada and the United Kingdom.

Demographic Analysis:

Results: Disney+ appeals to a broad demographic range, with a significant portion of viewership coming from families with children. However, there's also a growing audience of young adults and nostalgic viewers drawn to classic titles.

Cross-Genre Analysis:

Results: Certain genres, such as Action-Adventure and Sci-Fi-Fantasy, overlap significantly, indicating audience preferences for hybrid genres. Titles like "Star Wars" and "Marvel" franchises exemplify this trend.

Content Popularity Analysis:

Results: High-profile releases like "Frozen 2" and "The Lion King" remake consistently rank among the most-watched titles on Disney+, driving subscriber engagement and retention.

These analyses provide valuable insights into various aspects of Disney+ content, helping to inform content curation, production strategies, and audience targeting for the streaming platform.

Here's a future scope for further exploration and enhancement of the analysis:

Predictive Modeling: Implement predictive modeling techniques to forecast audience engagement, content popularity, and subscriber churn. Machine learning algorithms can leverage historical data to identify patterns and trends, enabling Disney+ to make data-driven decisions for content acquisition, production, and marketing strategies.

Personalized Recommendations: Develop personalized recommendation systems based on user preferences, viewing history, and demographic profiles. Utilize collaborative filtering, content-based filtering, and hybrid approaches to deliver tailored content recommendations, enhancing user satisfaction and retention.

Sentiment Analysis: Integrate sentiment analysis tools to analyze user reviews, social media conversations, and audience feedback in real-time. By understanding viewer sentiment and sentiment shifts, Disney+ can identify emerging trends, address audience concerns, and optimize content offerings accordingly.

Content Quality Assessment: Implement automated content quality assessment algorithms to evaluate the production quality, storytelling effectiveness, and audience reception of Disney+ titles. Leveraging natural language processing (NLP) and computer vision techniques, assess factors such as dialogue sentiment, visual aesthetics, and narrative coherence to ensure high-quality content delivery.

Geospatial Analysis: Conduct geospatial analysis to explore regional variations in content preferences, consumption patterns, and cultural influences. By mapping viewer demographics, content localization strategies, and regional content gaps, Disney+ can tailor its content library and marketing efforts to specific geographic markets more effectively.

Accessibility and Inclusivity: Prioritize accessibility and inclusivity initiatives by enhancing closed captioning, audio descriptions, and language localization features for diverse audiences. Conduct user research and usability testing to identify accessibility barriers and implement inclusive design practices across the Disney+ platform.

Content Monetization Strategies: Explore innovative content monetization strategies beyond subscription-based models, such as premium video-on-demand (PVOD), merchandise tie-ins, interactive experiences, and virtual events. Diversifying revenue streams can help Disney+ maximize monetization opportunities and enhance overall profitability.

Data Governance and Privacy Compliance: Strengthen data governance frameworks and compliance measures to safeguard user data privacy, security, and regulatory compliance. Implement robust data protection protocols, transparency mechanisms, and user consent mechanisms to maintain trust and compliance with global privacy regulations.

Cross-Platform Integration: Foster seamless integration with other Disney-owned platforms and services, including theme parks, merchandise stores, and linear television networks. By creating synergies across multiple touchpoints, Disney+ can enhance brand loyalty, cross-promotional opportunities, and overall brand ecosystem engagement.

Continuous Innovation and Experimentation: Foster a culture of continuous innovation and experimentation by encouraging cross-functional collaboration, agile methodologies, and rapid prototyping. Embrace emerging technologies, industry trends, and consumer preferences to drive ongoing product evolution and differentiation in the competitive streaming landscape.

By embracing these future scope initiatives, Disney+ can position itself as a leading global entertainment destination, delivering personalized, inclusive, and immersive experiences to audiences worldwide.

**REFERENCES**

* Kaggle

* Disney+ analysis dataset
* Disney Studios for providing data