|  | **Marvel-ous Insights on Marvel Cinematic Universe**  **Data 230 - Data Visualization**  **Team 7**  Sowmya Kuruba  Madhura Bhatsoori  Pooja Manjunatha  Anitha Balachandran  Nikhil Thota |
| --- | --- |

**TABLE OF CONTENTS**

[**1. INTRODUCTION 5**](#_ie7687l4cwh2)

[1.1 PROJECT BACKGROUND 5](#_9c329yl35m3e)

[1.2 PROBLEM DEFINITION 5](#_5f6w31wpf3ie)

[1.3 OBJECTIVE 6](#_vmc7hbsu213v)

[1.4 LITERATURE REVIEW 6](#_l61xltasqqf)

[**2. CRISP-DM METHODOLOGY 9**](#_xbqsmpmr0ks)

[**3. DATA FLOW DIAGRAM 11**](#_gwh5fkulnn5)

[**4. DATA COLLECTION 12**](#_34idwf5t8nt1)

[**5. DATA PRE-PROCESSING 14**](#_8su4e0vcnkku)

[5.1 MARVEL Historical Facts 14](#_v8rsejz0j828)

[5.1.1 Data Cleaning 14](#_lp7jkkhiis0z)

[5.1.2 EDA 19](#_3idkpestgvmo)

[5.2 LIVE SURVEY DATA 41](#_hr0w6xf91k5z)

[5.2.1 EDA 41](#_oag9tv60i60l)

[5.2.2 Data Cleaning 43](#_l6e1w1lv3w4w)

[**6. VISUALIZATIONS 46**](#_fm75tar075xj)

[6.1 Marvel Movie Audience Dashboard: Insights on Popularity and Reviews 47](#_kmtfyhimh6ud)

[6.2 Marvel Audience Survey Insights: A Guide for Marvel Writers 48](#_efczwubv6zf8)

[6.3 Marvel Audience Survey Insights: A Guide for Production and Marketing 52](#_gfxecdmfudvo)

[**7. DISCUSSIONS 57**](#_8g2lqp2rjubi)

[**8. CONCLUSION 60**](#_1mzuhlacu53k)

[**9. FUTURE SCOPE 61**](#_2uzrwsoucowh)

[**10. TABLE OF FIGURE 62**](#_ikjluq8xgd6a)

[**11. REFERENCES 64**](#_oijichyhdqrx)

**Abstract**

Marvel's vast and complex mythology, along with its global impact on popular culture, makes it a fascinating subject for analysis. The proposed project aims to analyze Marvel data to gain insights into its impact on popular culture and the evolution of the entertainment industry. To accomplish this, the proposed project will gather a historical dataset from various sources, including Marvel APIs and web scraping, and conduct data exploration to uncover insights into its characteristics. Additionally, the project team will create a survey to gather data from Marvel fans on character popularity and expectations, which will be analyzed for trends. The collected data will be visualized to identify patterns and trends related to character popularity, box office revenue, and critical reception. By analyzing these trends, the project seeks to extract valuable insights into the factors that contribute to a successful franchise and how it can remain relevant over time and it aims to provide an in-depth analysis of the information gathered, offering valuable insights into the entertainment industry, consumer behavior, and cultural trends.

## **INTRODUCTION**



The Marvel Cinematic Universe (MCU) has evolved into a vast and interconnected universe, spanning comics, movies, TV shows, and more. With the proliferation of data available over the internet, there is a tremendous opportunity to leverage data visualization techniques to explore and understand the Marvel dataverse comprehensively. This project aims to embark on a data visualization journey to uncover insights, patterns, and relationships within the vast expanse of Marvel's data, enabling stakeholders and enthusiasts to gain a deeper understanding of this captivating universe.

### 1.1 PROJECT BACKGROUND

The Marvel Cinematic Universe (MCU) has become a global phenomenon, captivating audiences with its rich characters, thrilling storylines,stunning visual effects and a lot of comics. As the franchise continues to expand, it is crucial for stakeholders, including Marvel's story writers, production teams, and enthusiasts, to gain valuable insights into the factors driving the success of these franchises. This project aims to utilize data visualization techniques, data models, methodology, and analytics to uncover the key success factors of Marvel franchises, including character popularity, box office revenue, and critical reception.

### 1.2 PROBLEM DEFINITION

Marvel enthusiasts often find themselves faced with unanswered questions when it comes to the Marvel Cinematic Universe (MCU) and its expansive world of comics and movies. Despite their passion for the franchise, they lack clear answers to key inquiries such as identifying the best character, determining the most popular comic, and understanding various aspects related to the MCU. This lack of clarity creates a void in their knowledge and hinders their ability to engage in meaningful discussions and analysis within the Marvel community.

Furthermore, stakeholders such as Marvel's story writers, production teams, marketers, and franchise developers also face challenges in making informed decisions without comprehensive insights into the MCU universe. Moreover, understanding the preferences and interests of fans becomes crucial for the success of Marvel's merchandise and immersive experiences, including theme parks.

Therefore, there is a pressing need to conduct comprehensive research and analysis that can provide definitive insights into these MCU-related questions, enabling major Marvel enthusiasts and stakeholders to gain a deeper understanding of each other and make informed decisions based on different aspects of the MCU universe.

### 1.3 OBJECTIVE

Our project seeks to leverage data visualization techniques to uncover the key success factors of Marvel franchises such as character popularity, box office revenue, and critical reception, with the goal of benefiting various stakeholders. We aim to provide valuable information to Marvel story writers and production teams to guide content creation, marketing strategies, and overall franchise development. Furthermore, we aim to assist Marvel enthusiasts in developing a more comprehensive understanding of the most popular Marvel movies by analyzing ratings from diverse sources.

### 1.4 LITERATURE REVIEW

The paper by Inseok Ko, and Hyejung Chang, titled **"Interactive Visualization and Analysis of Healthcare Data Using Tableau"** presents a procedure for using Tableau as a business intelligence tool for the interactive visualization and analysis of healthcare data. The authors emphasize the importance of visualization in presenting the results of data analysis and the overall process of data handling. The paper highlights Tableau as software that facilitates the exploration and understanding of data through interactive visualizations. We have drawn inspiration from the concepts and techniques presented in the paper, such as interactive visualizations, customizations, and different chart types, and adapted them to suit the requirements of your Marvel dashboard.

The paper by María Teresa Rodríguez, Sérgio Nunes, and Tiago Devezas titled **“Telling Stories with Data Visualization”** surveys the historical development of data visualization and its intersection with storytelling in data journalism. It highlights the usefulness of visualization in helping readers understand complex information and explores how narrative techniques enhance the power of visual expression. The paper aims to inspire further research and emphasizes the value of incorporating storytelling with data in creating engaging and informative experiences. It can guide the design of effective visualizations for the Marvel project, inspire the incorporation of narrative elements, and encourage further research in the intersection of data visualization and storytelling.

The review paper by Jonas Van Der Donckt, Jeroen Van der Donckt, Emiel Deprost, Sofie Van Hoecke titled ” **An overview and comparison of free Python libraries for data mining and big data analysis”** explores the growing popularity of Python in the field of data science and its associated free libraries. The paper aims to provide a comprehensive description and comparison of various data mining and big data analysis libraries in Python, which has not been covered in previous research. The paper recommends: pandas for data preparation; Matplotlib, seaborn or Plotly for data visualization. This paper offers valuable insights for our Marvel data visualization project by comparing data visualization libraries in Python, and providing recommended libraries for tasks like data preparation and analysis

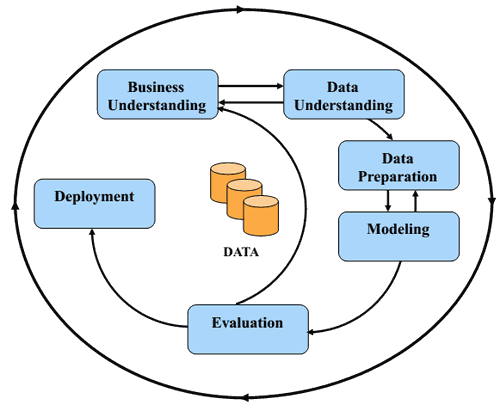
The paper by Jaehoon Lee,Giseop Noh, Chong Kwon Kim titled **“Analysis & Visualization on movie’s popularity and reviews”** presents visualizations using tools to map the relationship between the audience and reviews.This paper focuses on visualization tools that utilize normalized graphs to uncover the underlying moods of movies and illustrate their intricate relationships. Through extensive experimentation using real datasets obtained from naver-movie.com and the Korean Film Council, which encompass the number of customers for each movie, they uncover distinctive relations and gain insights into the profound impact of word-of-mouth effects. This paper on movie analysis and visualization can be useful for your Marvel project by providing insights into the relationship between audience reviews and popularity through visualizations and uncovering the impact.

The article by Louis T. Becker and Elyssa M. Gould titled “**Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data”** introduces Microsoft's Power BI software and its integration with newer versions of Excel, highlighting its potential benefits for librarians in analyzing and visualizing data from diverse sources encountered in library operations. As data visualization gains importance in managing "big data," Power BI provides a user-friendly tool for librarians to enhance their understanding and utilization of data, making it relevant for our Marvel project in terms of analyzing and visualizing data related to writer and production survey aspects of the Marvel universe.

## **CRISP-DM METHODOLOGY**



We are using crisp DM (Cross Industry Standard Process for Data Mining) for our project solution. It uses a structured approach to planning a project. It consists of six stages:



##### *Figure 1: CRISP-DM Methodology*

1) **Business Understanding:** This is the initial step in the project which involves understanding the business problem that we are trying to solve. Here we did a literature survey on the topic finalized. In our project, analyzing and visualizing the MCU movies is the problem statement. Here we then defined goals of the project: analyzing box office performance, reviews, characteristics, and occurrence of marvel superheroes.

2) **Data Understanding**: In this step, we collected the datasets from various sources. We used marvel API, web scraping and Kaggle datasets to do analysis and visualization. We also decided on conducting surveys to plan future activities for writers and the production team. We got 86 and 66 responses for the survey. Then we analyzed these both dataset for its quality, completeness, and relevance to the business problem.

3) **Data preparation:** In this step, the focus is on cleaning and transforming the data to ensure its quality and suitability for analysis. Here we removed missing values, duplicates, unwanted columns, and transforming columns as needed. This helped to remove inconsistencies, errors, or outliers that may affect the accuracy of the analysis. Then we merged the data as required by analysis. This created a reliable and structured dataset that is ready for further analysis and modeling. For survey data, we renamed the columns to shorter names, removed the timestamp column as it was not necessary for analysis.

4) **Data modeling:**  In data modeling initially we performed exploratory data analysis in tableau, created various plots and for complex data analysis we used python to create plots to understand hidden patterns in data. Then we created an interactive dashboard which gives insightful information about the datasets we collected. For survey data, we created visualization charts in Power BI and combined them into two static dashboards which give a comprehensive overview of general audience opinions and expectations from marvel.

5) **Data Evaluation:** In this step, we evaluate the performance of the model and check its effectiveness in solving the business problem. This also involves testing the model on a new dataset and comparing its results to the expected outcomes. This we are doing by comparing marvel fact data with survey data.

6) **Deployment :** As the final phase of our project, we focused on developing a personalized front-end using HTML 5 and Bootstrap. To ensure an intuitive user experience, we created three components for each dataset. We integrated an interactive dashboard from Tableau and two other static dashboards created using Power BI. To pull our visualizations from Tableau to our HTML dashboard, we used the complete Bootstrap and CSS libraries. This allowed us to design our own panels and display the plots in any style we desired. Additionally, any changes we make to our Tableau Public account will automatically get updated in our application. The end result is a comprehensive and interactive data visualization tool that allows users to explore the world of Marvel in a unique and informative way.

## **DATA FLOW DIAGRAM**



The Data Flow Diagram shows the complete high-level process of our proposed solution.As shown in Figure 2, We have collected data of marvel historical facts and general audience survey. Marvel Historical facts data is then cleaned and prepared using python by removing missing and unwanted values. Then we merge all data to create a clean dataset. Next we did exploratory data analysis to understand data and examine relationships between different variables. Based on this we created different visualizations. We then combined these visuals to an interactive dashboard.

Similarly survey data collected for writers and production were cleaned by removing unwanted columns and renaming columns suitable for visualization. Then we did exploratory data analysis. Then created visualizations. We then combined these visualizations into two static dashboards for marvel writers and production. FInally we integrated these three dashboards into a user-friendly web page, allowing users to easily navigate through the visualizations and gain insights into the analyzed data which helps in decision-making.

##### *Figure 2: Data Flow Diagram*

## **DATA COLLECTION**



We have collected data from various sources to help in identifying trends and patterns in the data that may not be apparent from a single source. We have collected from four different sources, as given in Table 1 with detailed information regarding the collection method and description.

Table 1:Data Collection Details

| Name | Collection Method | Description |
| --- | --- | --- |
| Marvel character dataset | Marvel API | We used the Marvel API to collect data on various Marvel characters, using Postman. This is used to analyze characteristics of different superheroes. |
| Marvel box office revenue dataset and marvel movie review dataset | Data downloaded from kaggle | We collected Marvel movies revenue and ratings dataset available on Kaggle. This is used to understand various patterns and trends in box office and review records. |
| Marvel comics sales data | Web scraping various online websites | We scraped data from various websites that track comic book sales from Kaggle and other online resources. We used this to understand the origin of marvel character popularity. |
| General Audience Survey (Guide for marvel writers and production) | It was collected using Google Forms. The survey was shared on SAMMY, Whatsapp to gather information from movie audiences. | We distributed Marvel surveys to personal networks and SJSU Sammy App to collect data that helps the Marvel writers and production team. |

## **DATA PRE-PROCESSING**



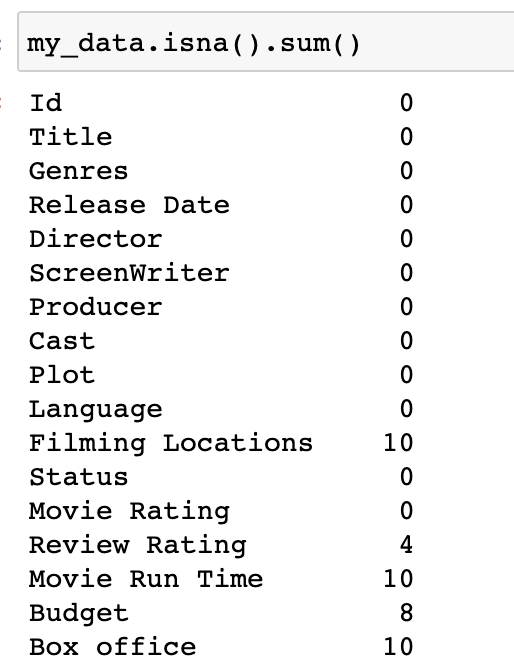
### 

### 5.1 MARVEL Historical Facts

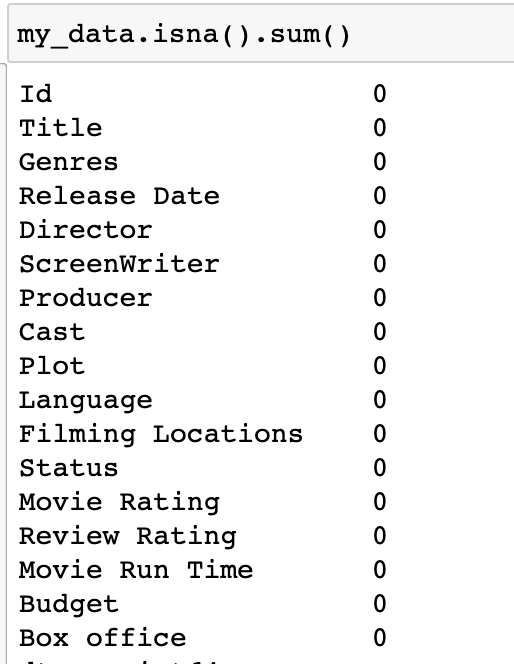
#### **5.1.1 Data Cleaning**

In our analysis of historical facts data, we encountered null values within the dataset. Recognizing the importance of preserving the integrity of the data and ensuring accurate analysis, we manually replaced these null values with appropriate values. By doing so, we aimed to maintain the completeness of the dataset and avoid any loss of valuable information.

This meticulous approach allowed us to address the null values on a case-by-case basis, considering the context and significance of each data point. Through careful examination and research, we assigned appropriate values to the previously null entries, ensuring that no crucial information was overlooked or disregarded.

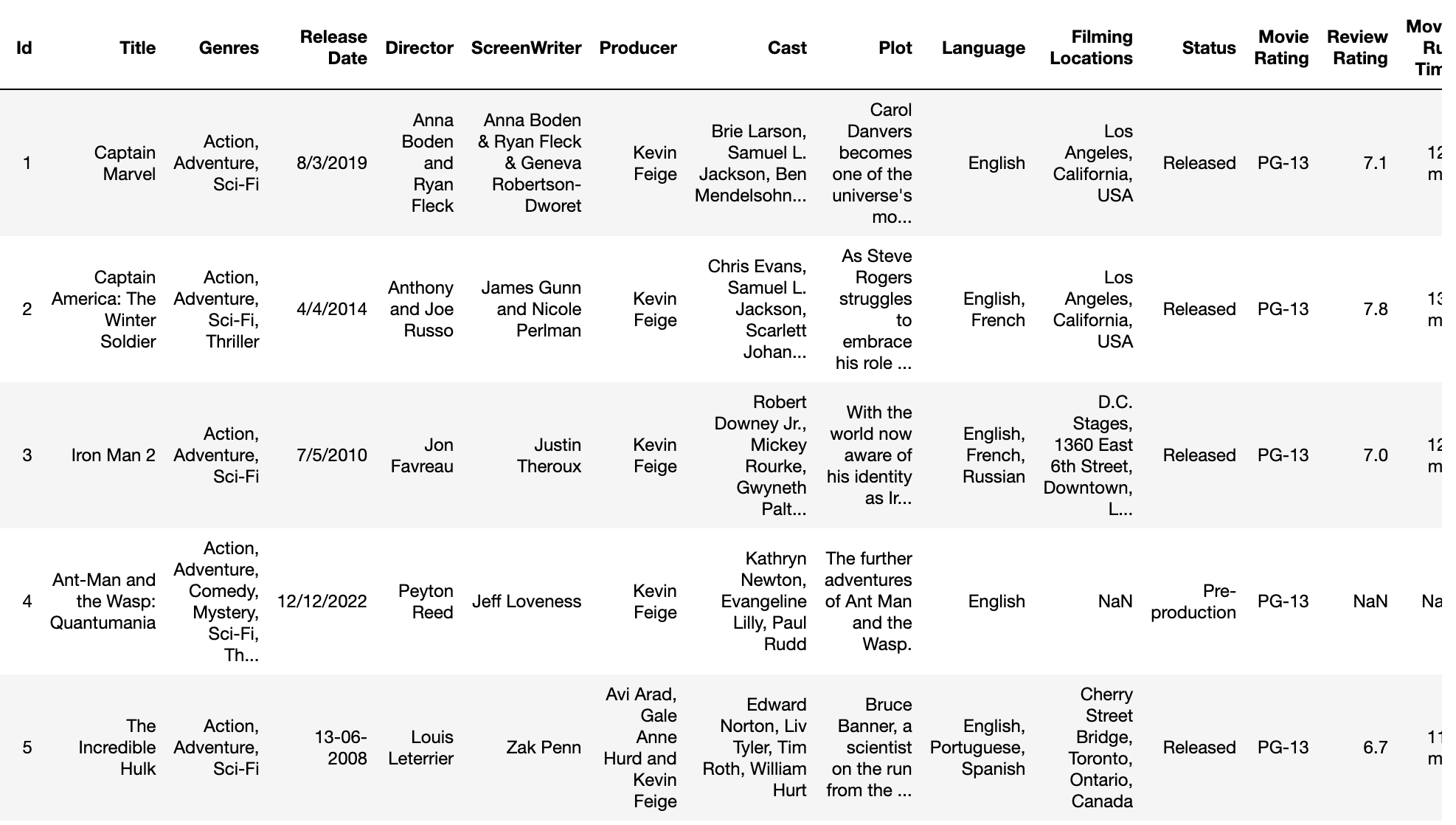
****

##### *Figure 3: Before Handling Null Values*

****

##### *Figure 4: After Handling Null Values*

Before the cleaning process, the dataset contained unnecessary columns and long column names that were not relevant to our analysis.

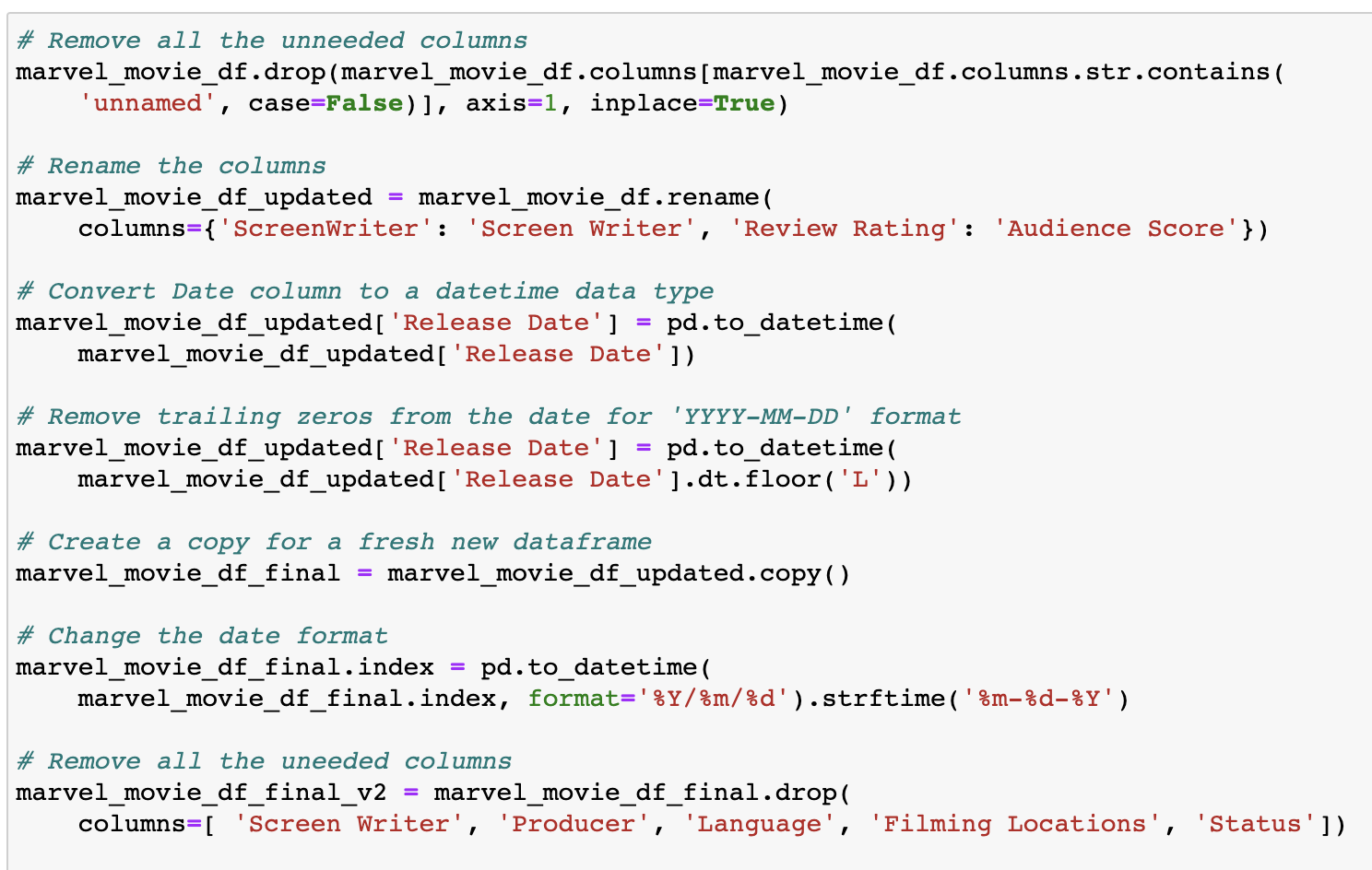
****

##### *Figure 5: Before Data Cleaning*

To streamline the dataset and improve clarity, we made the decision to remove these unwanted columns. Additionally, we identified certain columns with improper names and took the initiative to rename them appropriately, ensuring clear and concise representation of the data.

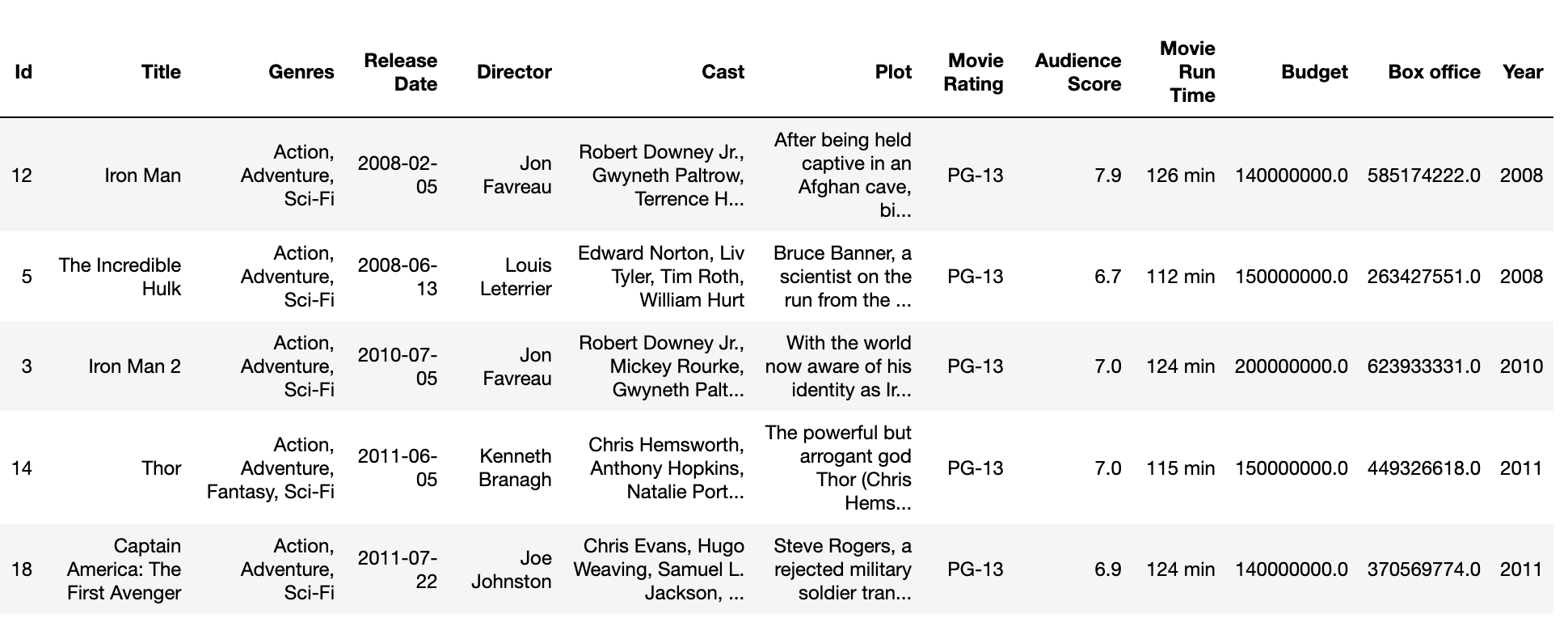
Furthermore, to maintain uniformity and consistency within the dataset, we converted the date format to datetime. This conversion allowed for standardized handling of date values, facilitating accurate analysis and comparisons across the dataset.

Through these data cleaning steps, we aimed to enhance the overall quality and usability of the dataset, making it more suitable for our analysis and improving the efficiency of data processing.

****

##### *Figure 6: Data Cleaning*

After performing the necessary data cleaning steps, the dataset has been transformed into a more organized and structured format. The unwanted columns have been removed, and the remaining columns have been properly renamed for clarity. The date format has been standardized to ensure uniformity and ease of analysis. By cleaning the data, we have eliminated null values and inconsistencies, resulting in a cleaner and more reliable dataset that can be used for further analysis and visualization.

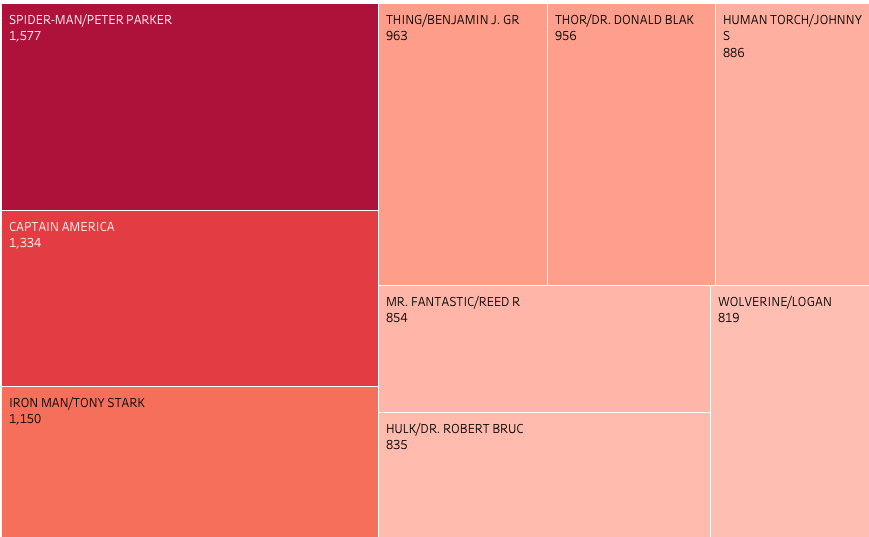
****

##### *Figure 7: After Data Cleaning*

After cleaning the dataset, we sorted the data in chronological order to ensure that the movies are presented in the correct sequence. Sorting the data based on the movie release dates allows for easy comprehension and analysis of the dataset, enabling us to follow the progression of events and trends over time. By organizing the data chronologically, we create a clear and coherent timeline that facilitates a comprehensive understanding of the movie-related information and aids in drawing meaningful insights from the dataset.

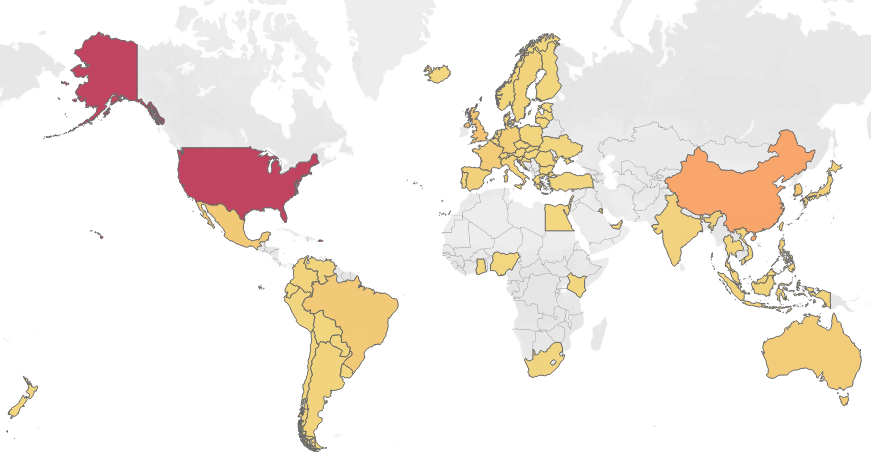
##### *Figure 8: Cleaned Sorted Data*

#### **5.1.2 EDA**

****

##### *Figure 9: Heat-map for popular marvel character*

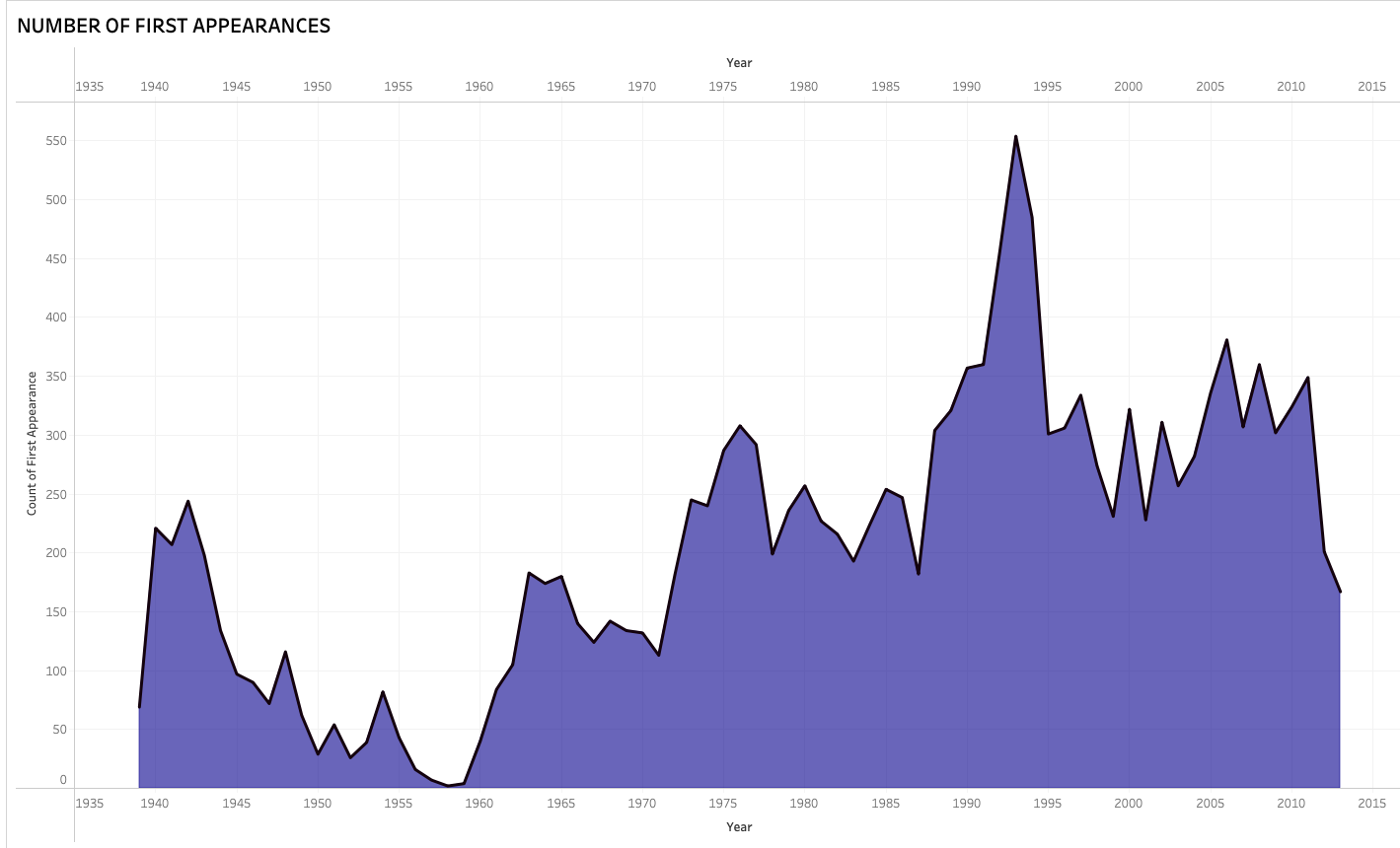
Spider-Man (Peter Parker) emerges as the most popular Marvel character, followed closely by Captain America and Iron Man. This popularity can be attributed to several factors, including their iconic status within the Marvel Universe, their involvement in major storylines and events, as well as their appearances in successful film adaptations.Spider-Man, in particular, has captured the hearts of fans worldwide with his relatable persona as a young, witty, and agile superhero. His enduring popularity can be attributed to his compelling backstory, memorable villains, and his ability to resonate with audiences of all ages. Captain America and Iron Man, on the other hand, have garnered immense popularity due to their central roles in the Marvel Cinematic Universe. Captain America embodies the values of honor, courage, and patriotism, making him a symbol of heroism and leadership. Iron Man, with his technological genius and charismatic personality, has become a fan-favorite for his wit, charisma, and complex character development.It's worth noting that popularity can vary among different demographics, regions, and generations. The Marvel Universe is vast, with a rich roster of beloved characters, each with their own dedicated fan base. However, based on our analysis, Spider-Man, Captain America, and Iron Man stand out as some of the most beloved and widely recognized characters within the Marvel Universe.

****

##### *Figure 10: Box Office Collection Worldwide*

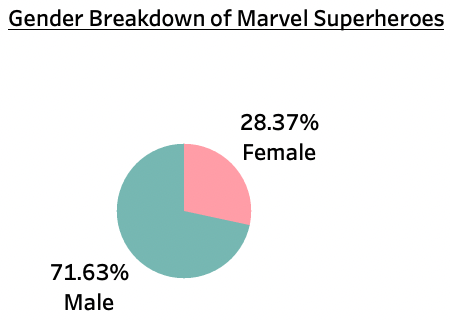
​​

The United States has the highest number of Marvel fans in terms of box office records followed by China. This finding aligns with the global appeal and widespread popularity of Marvel movies, which have garnered massive audiences and generated significant box office revenues in these two countries.The United States, being the home country of Marvel Comics and the birthplace of the Marvel Cinematic Universe, has a longstanding relationship with Marvel characters and stories. The release of Marvel movies has created a strong fan base, with audiences eagerly anticipating each new installment and contributing to the box office success.China, on the other hand, has emerged as a major market for Hollywood films, including Marvel movies. The growing popularity of Marvel superheroes in China can be attributed to the expansion of the Chinese film industry, the increasing accessibility of international films, and the strong appeal of superhero stories and action-packed blockbusters among Chinese audiences.The success of Marvel movies in both the United States and China demonstrates the global impact of the Marvel brand and its ability to captivate audiences across different cultures and markets. As Marvel continues to expand its cinematic universe and introduce new characters and storylines, it is likely that the fan base will continue to grow in these and other countries around the world.

****

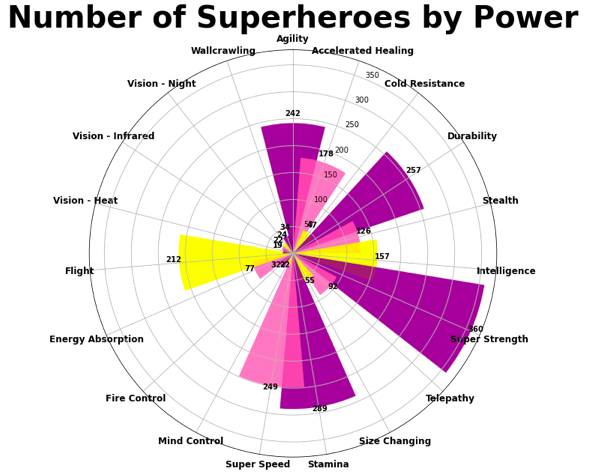
##### *Figure 11: Area plot for number of appearances*

The figure presented illustrates the frequency of first appearances in Marvel over time. From the analysis, it is evident that the period between 1990 and 1995 witnessed the highest number of character first appearances, indicating a significant influx of new characters during that time. On the other hand, the period between 1955 and 1960 exhibited the lowest number of first appearances, suggesting a relatively slower introduction of new characters during that particular time frame.

****

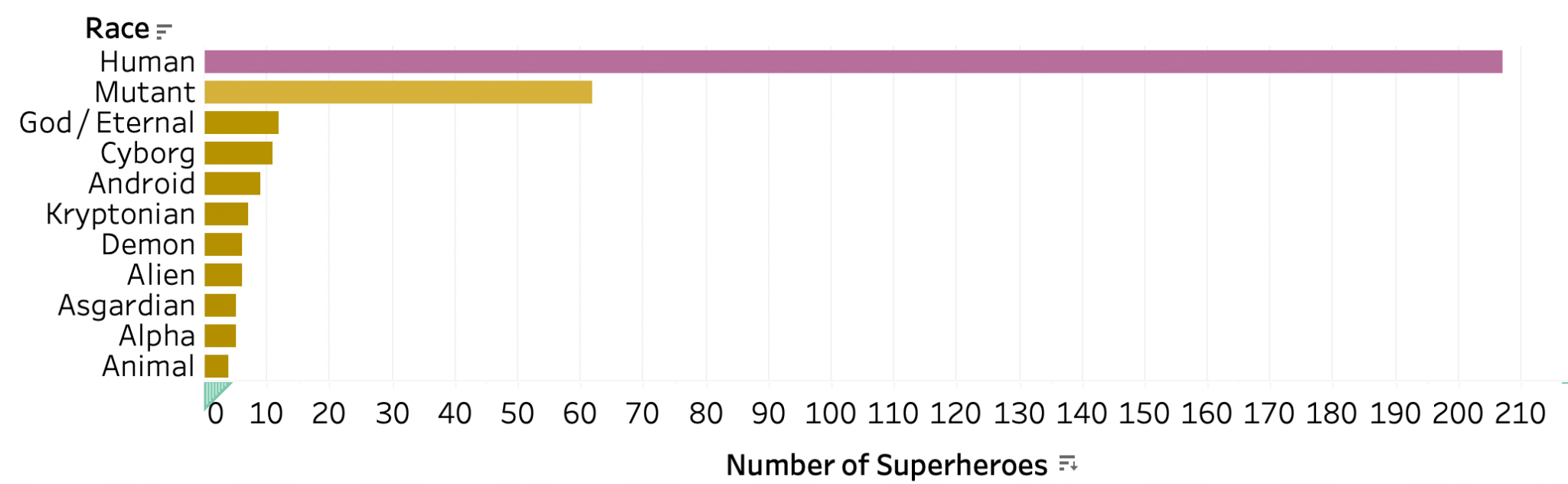
##### *Figure 12:Pie plot for Gender*

Gender imbalance within the Marvel Universe, with male superheroes being more prevalent than female superheroes. This finding aligns with the broader trend observed in the comic book industry, where male characters have historically dominated the superhero genre.There could be several factors contributing to the higher representation of male superheroes. One possible explanation is the historical context in which many of these characters were created. Traditional comic books and superhero stories were often developed during a time when societal norms and gender stereotypes played a significant role in shaping the narratives and characterizations.

****

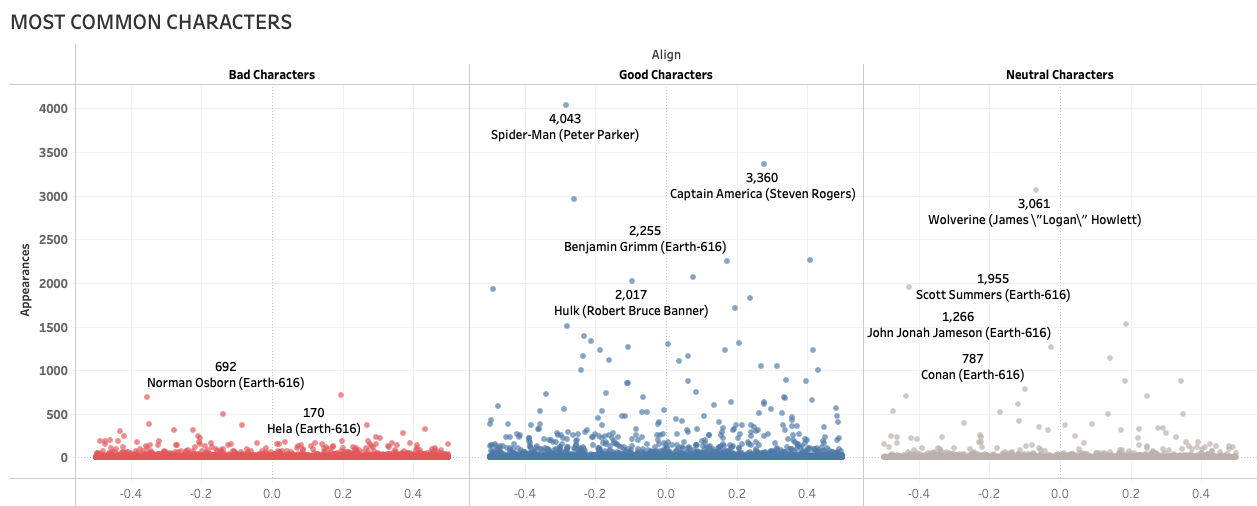
##### *Figure 13: Spectrum of Superhero abilities*

​​The analysis of the number of superheroes by power revealed interesting patterns within the Marvel Universe. The figure illustrates the distribution of superheroes across different powers, providing insights into the prevalence and popularity of specific abilities.The most common power among superheroes is super strength, with a significantly higher number of characters possessing this ability compared to others. Super strength represents the embodiment of physical power, enabling characters to perform extraordinary feats of strength and overpower their opponents. This power resonates with audiences due to its visual appeal and the idea of characters being able to overcome challenges through sheer force.Following super strength, the power of stamina ranks second in terms of the number of superheroes possessing it. Stamina represents a character's endurance and ability to sustain physical exertion over an extended period. This power is often associated with characters who engage in prolonged battles or physically demanding activities, showcasing their resilience and tenacity.The figure reveals that Vision-heat is the least common power among superheroes. Vision-heat refers to the ability to emit or manipulate heat energy, allowing characters t o generate intense heat or control thermal energy. While this power may be less prevalent, it still adds diversity to the range of abilities within the Marvel Universe.

****

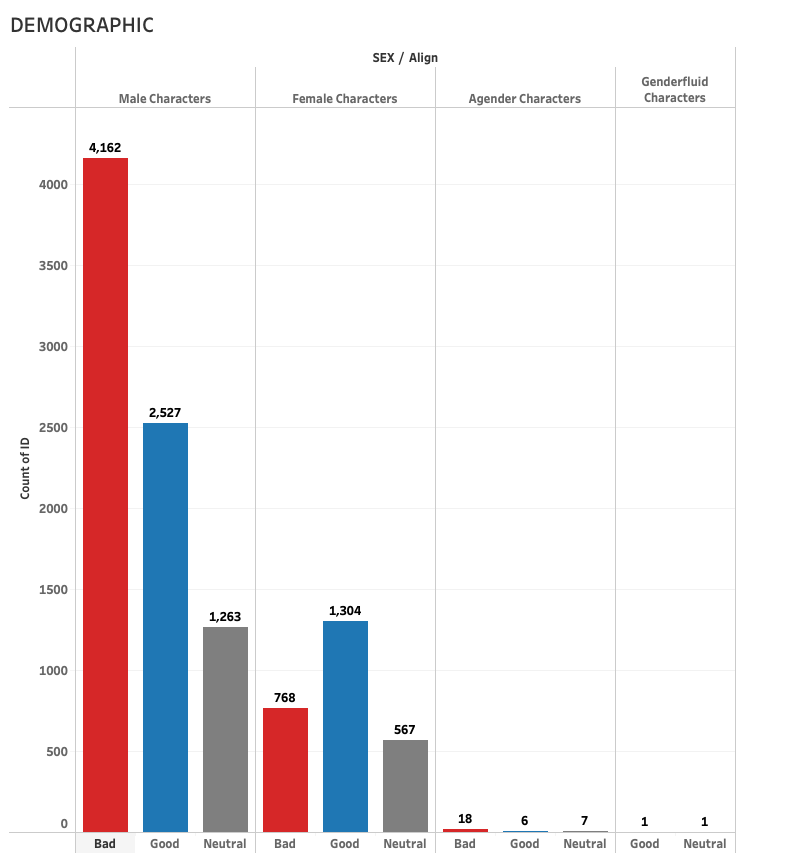
##### *Figure 14: Horizontal bar plot*

Human superheroes are the most common type of characters in the Marvel Universe. While Marvel comics feature a diverse array of characters with various powers and origins, human superheroes have always held a significant presence and appeal to readers.One of the reasons for the prevalence of human superheroes is the relatability factor. Human characters often serve as a relatable point of identification for readers, as they possess human qualities, emotions, and vulnerabilities. They face everyday challenges, allowing readers to connect with their struggles and triumphs on a personal level.Human superheroes often represent the ideals and aspirations of humanity. They demonstrate courage, resilience, and the ability to overcome obstacles, which can inspire readers and reflect the human spirit. Human characters with extraordinary abilities showcase the potential within each individual to rise above their limitations and make a positive impact.While there are characters with extraterrestrial origins, mutants, and cosmic beings, human characters ground the stories in a recognizable reality, allowing for more relatable storytelling and exploring social, political, and ethical issues that resonate with readers.It is important to note that Marvel comics also feature a diverse range of non-human superheroes, including mutants, aliens, gods, and cosmic entities. These characters bring unique perspectives, powers, and storylines to the Marvel Universe, enriching its overall diversity and expanding the possibilities for storytelling.

****

##### *Figure 15: Jitter plot for most common character*

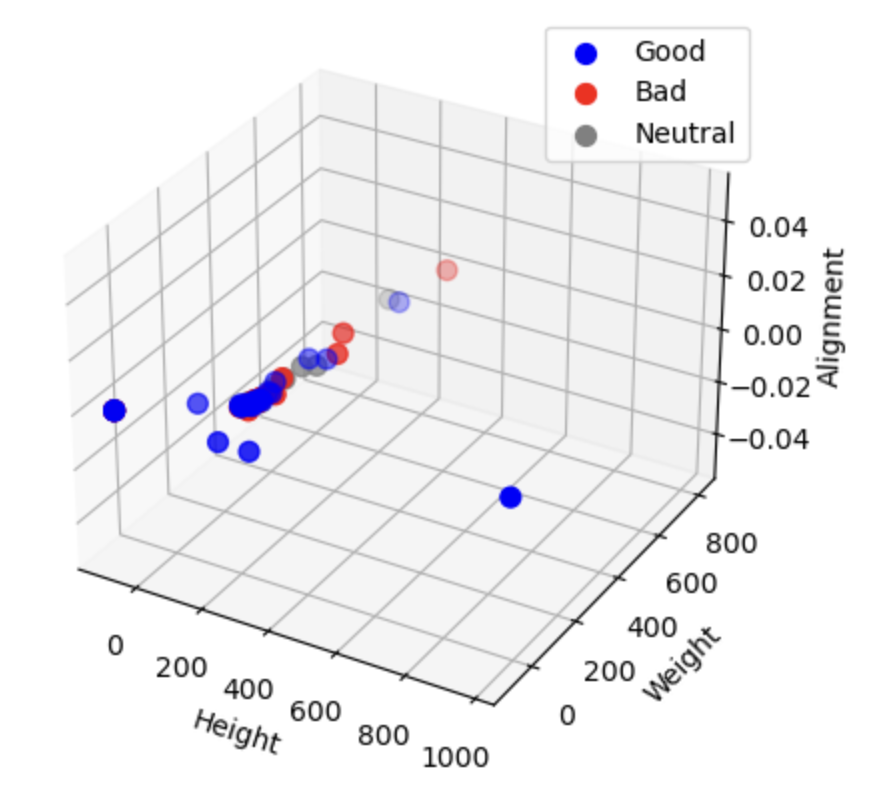
The most common bad character in the Marvel Universe is Norman Osborn. Norman Osborn is a complex and formidable villain who has appeared in various Marvel comics and storylines. Also known as the Green Goblin, he is recognized for his cunning intelligence, advanced technology, and intense rivalry with Spider-Man. Norman Osborn's character has been a prominent antagonist in the Spider-Man series and has had significant impacts on the Marvel Universe as a whole.The most common good character is Spider-Man, one of Marvel's most iconic and beloved superheroes. Spider-Man, also known as Peter Parker, has captured the hearts of fans worldwide with his relatable nature, incredible agility, and web-slinging abilities. Spider-Man embodies the ideals of heroism, justice, and responsibility, making him a fan-favorite character and a symbol of hope within the Marvel Universe.The most common neutral character is Scott Summers, also known as Cyclops. Cyclops is a member of the mutant superhero team known as the X-Men and possesses the ability to emit powerful optic blasts. As a leader and strategist, Cyclops plays a vital role in the X-Men's missions and their fight for mutant rights. His neutral classification indicates that he doesn't align strictly with the traditional notions of good or bad but operates within his own moral compass.Analyzing the prevalence of these characters provides insights into the narrative dynamics of the Marvel Universe. It highlights the significance of complex villains like Norman Osborn, who pose substantial challenges to the heroes, and the enduring popularity of iconic heroes like Spider-Man. The inclusion of neutral characters like Scott Summers adds depth and variety to the Marvel Universe, showcasing characters who navigate the shades of gray between traditional notions of good and bad.

****

##### *Figure 16: Grouped bar plot for Demography*

It is evident that there is a gender imbalance among the characters, with a higher proportion of male characters categorized as good, bad, and neutral compared to female characters.

This gender disparity reflects a historical trend in the comic book industry, where male characters have traditionally dominated the superhero narrative. It is worth noting that in recent years, efforts have been made to address this imbalance and introduce more diverse and well-rounded female characters into the Marvel Universe.The dataset highlights the limited representation of agender characters, who do not identify with any specific gender, and the presence of only one genderfluid character. This emphasizes the need for greater inclusion and representation of gender diverse characters within the Marvel Universe and the broader entertainment industry.By acknowledging the gender imbalance in the dataset, we recognize the importance of promoting gender equality and inclusivity in storytelling. It is crucial for future developments in the Marvel Universe and the comic book industry as a whole to strive for more diverse and inclusive character representation, allowing for a wider range of gender identities to be celebrated and reflected in these narratives.

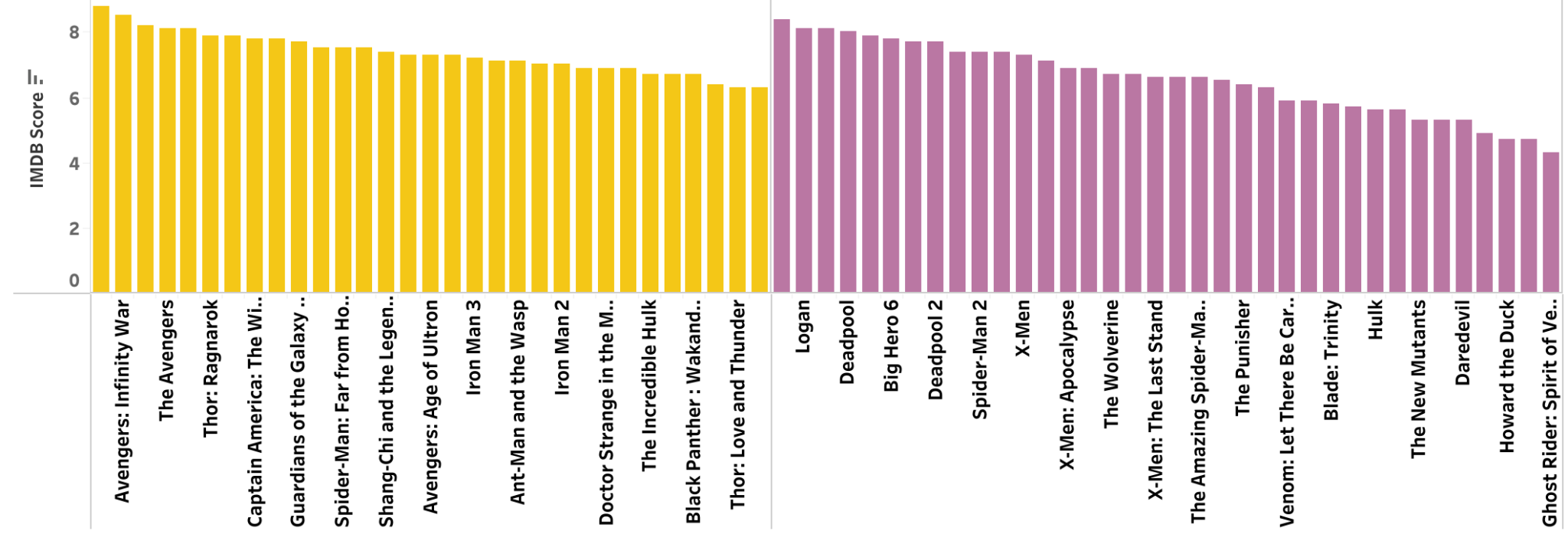
****

##### *Figure 17: 3D Scatter plot*

We have observed a correlation between the physical attributes of characters and their alignment as good, bad, or neutral within the Marvel Universe. It appears that there is a tendency for good characters to exhibit taller stature and lighter weight, while bad characters tend to be larger in size. Neutral characters show moderate height and weight. This observation may reflect certain stereotypes and archetypes prevalent in storytelling, where heroes are often portrayed as leaner and agile, while villains are depicted as imposing and physically intimidating. These physical attributes can help convey the characteristics and traits associated with each character's alignment, adding to their visual representation and the overall narrative.However, these observations are based on the dataset and the existing depictions within the Marvel Universe. It does not imply that all characters adhere to these physical stereotypes or that height and weight alone determine a character's alignment. Marvel, like any storytelling medium, is dynamic and evolving, with diverse character representations and interpretations.

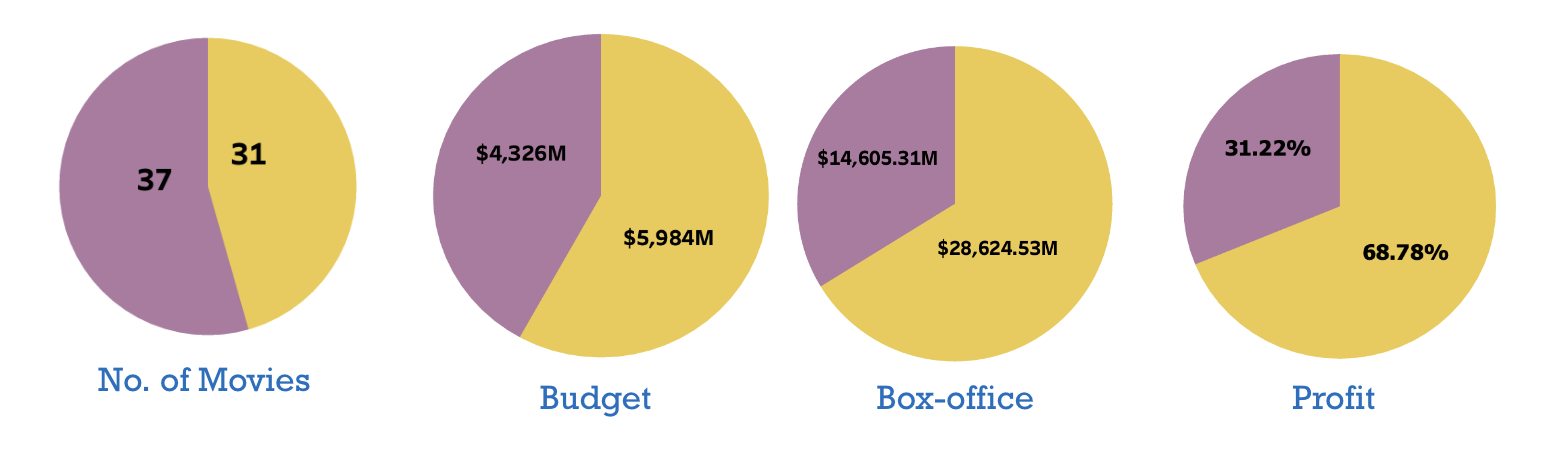
**A picture containing text

Description automatically generated**

****

##### *Figure 18: Grouped bar plot for MCU and Non-MCU*

We have conducted a comprehensive analysis of both MCU (Marvel Cinematic Universe) and non-MCU movies within the Marvel franchise. Through our research, we have uncovered several significant findings that shed light on the differences between these two categories.

****

##### *Figure 19: Pie plots for No. of movies, Budget, Box-office, Profit for MCU nad Non-MCU*

Firstly, we discovered that the number of non-MCU movies exceeds the number of MCU movies. This indicates that the Marvel franchise extends beyond the boundaries of the Marvel Cinematic Universe, encompassing a wider range of movies that may not directly connect to the overarching MCU storyline. This finding highlights the diversity and depth of the Marvel brand and its ability to captivate audiences through various cinematic experiences.

When comparing the financial aspects of MCU and non-MCU movies, we observed that the total budget and box-office collections of MCU movies are significantly higher than those of non-MCU movies. This suggests that the MCU movies have been allocated larger budgets and have enjoyed greater commercial success at the box office. The higher investment in the production of MCU movies may be attributed to the grand scale, special effects, and interconnected storytelling that characterizes the Marvel Cinematic Universe.

Furthermore, we examined the profitability of both categories and found that the profit margins of MCU movies are nearly double that of non-MCU movies. This indicates that the MCU movies not only generate higher revenue but also have better cost-efficiency and return on investment. The financial success of the MCU movies can be attributed to their strong brand recognition, dedicated fan base, and effective marketing strategies employed by Marvel Studios.

**Chart, box and whisker chart

Description automatically generated**

##### *Figure 20: Box plot for Box-office*

**Chart, shape, bubble chart

Description automatically generated**

##### *Figure 21: Horizontal multiple distribution of Box-office*

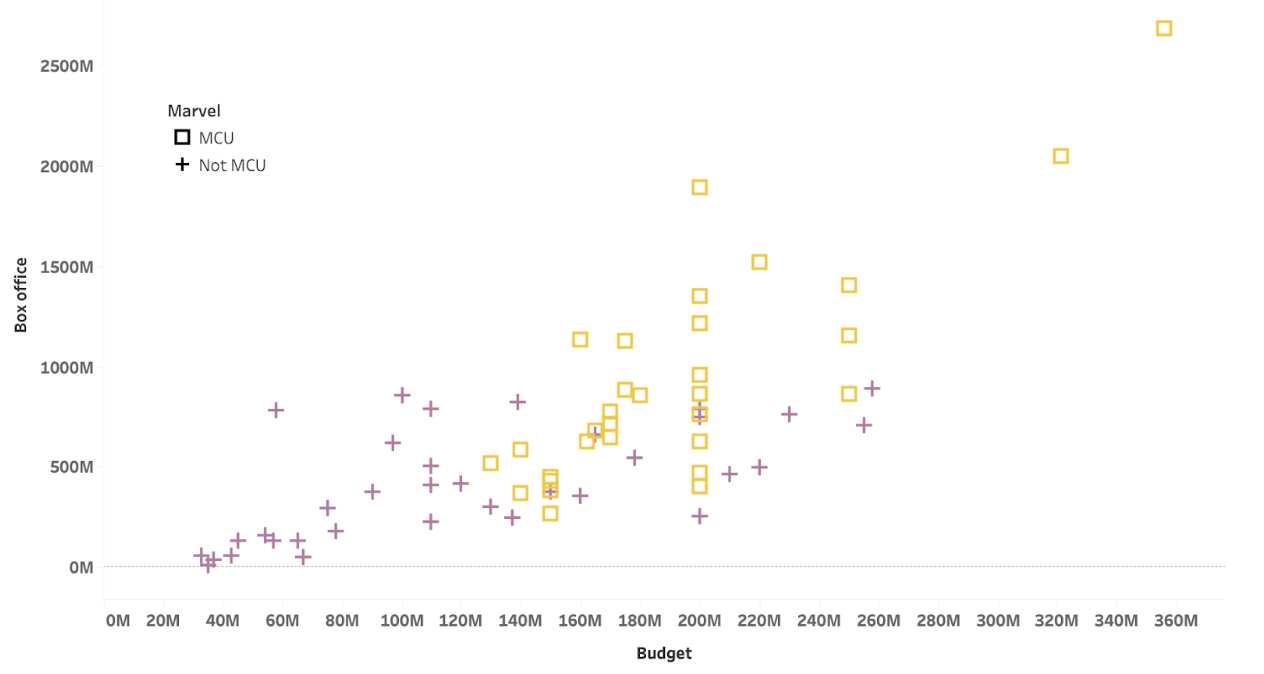
We have analyzed the box-office collections of both MCU (Marvel Cinematic Universe) and non-MCU movies within the Marvel franchise. Our findings reveal intriguing patterns and significant differences between these two categories, shedding light on their respective performances in terms of box-office success and audience reception.

One notable observation is that almost all MCU movies surpass the mean box-office collection, indicating a consistently strong performance in terms of revenue generation. On the other hand, non-MCU movies tend to fall below the mean box-office collection, suggesting a comparatively lower level of commercial success. This trend of box-office collections distribution holds true not only for the overall box-office collections but also for domestic box-office collections, collections from other territories, and opening weekend collections. This indicates a consistent pattern in the box-office performance of MCU and non-MCU movies.

Moreover, it is worth noting that the MCU movies hold the record for the highest box-office collection within the Marvel franchise. This reflects the immense popularity and widespread appeal of the interconnected MCU storytelling, which has captivated audiences worldwide. In contrast, non-MCU movies have the lowest box-office collection record, indicating a lower level of financial success compared to their MCU counterparts.

Within this context, it is remarkable to highlight the exceptional box-office performance of the MCU movie "Avengers: Endgame." This film stands out as the highest-grossing movie in the Marvel franchise, demonstrating its extraordinary commercial success. Additionally, "Avengers: Endgame" boasts an impressive IMDB rating of 8.8, further attesting to its critical acclaim and positive reception among audiences.

Conversely, the non-MCU movie "Inhumans" holds the lowest box-office collection record. This suggests that the movie did not resonate well with audiences and experienced limited commercial success. It is noteworthy that "Inhumans" received a relatively low IMDB rating of 4.9, indicating a less favorable reception compared to other Marvel movies

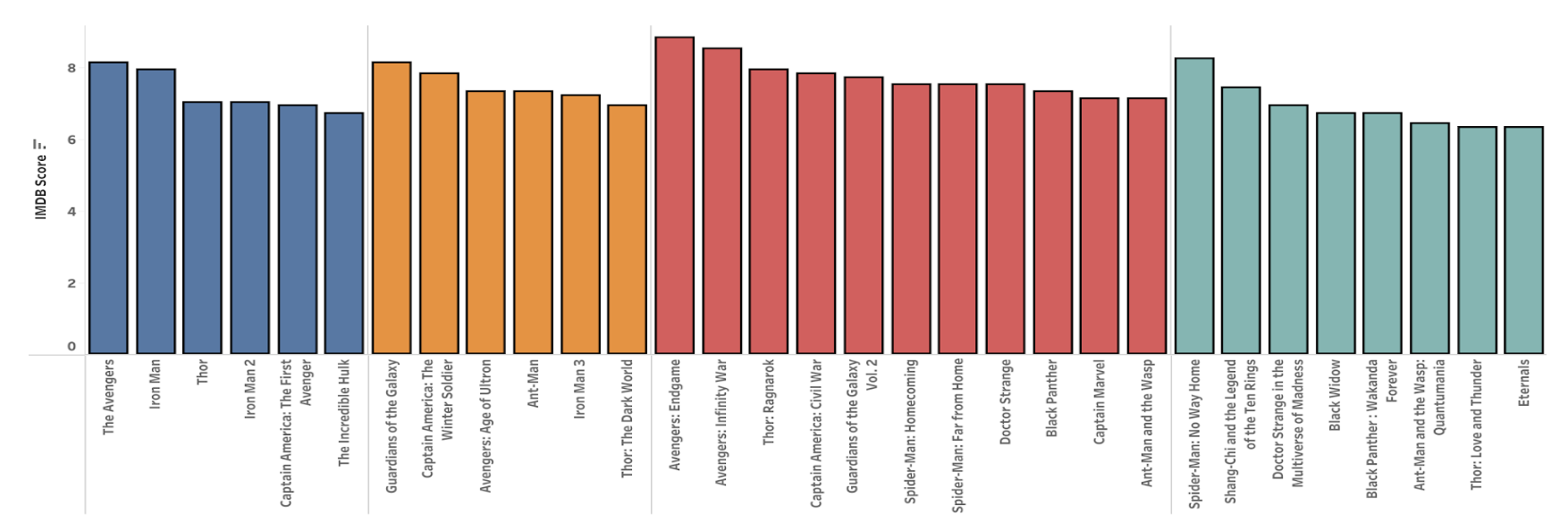
****

##### Figure 22: Scatter plot for Box-office vs Budget

Additionally, our analysis revealed a linear correlation between the budget and box-office performance of both MCU and non-MCU movies. This finding suggests that a higher budget investment tends to positively influence the box-office collections of Marvel movies, regardless of whether they are part of the MCU or not. This correlation emphasizes the importance of allocating sufficient resources to create visually stunning and captivating cinematic experiences that resonate with audiences.

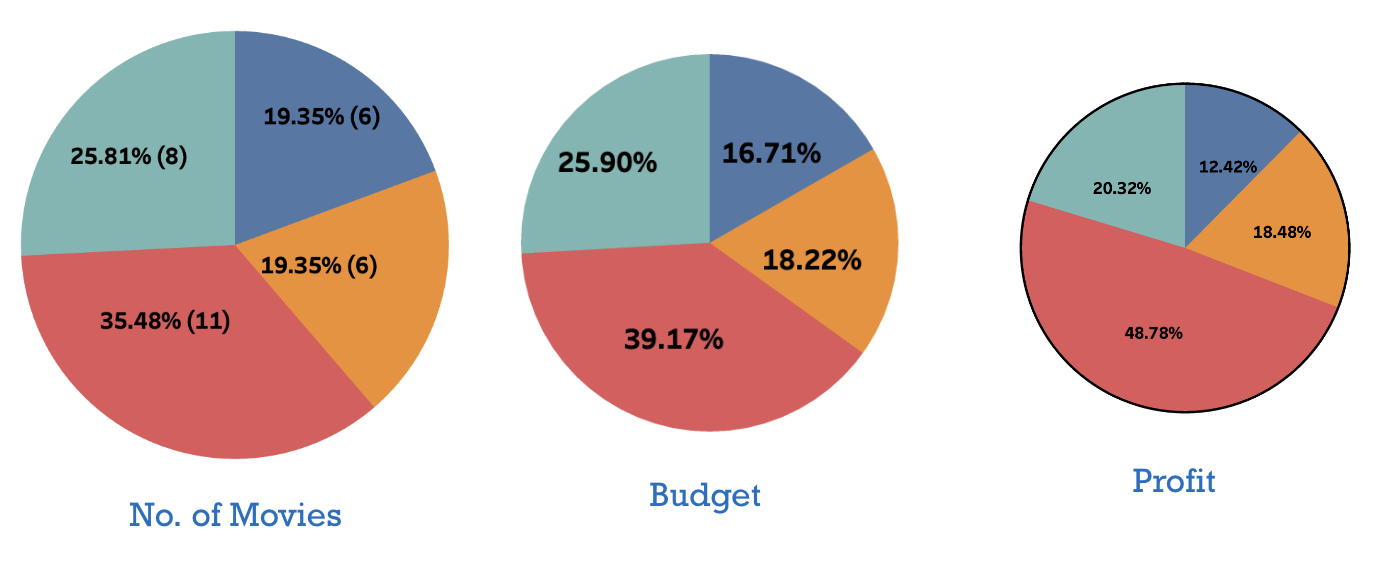
**A picture containing chart

Description automatically generated**

****

##### *Figure 23: Group bar plots for movies w.r t phase*

We have examined the different phases of Marvel movies and analyzed various aspects related to their releases, budgets, profits, and audience ratings. Our findings reveal interesting trends and comparisons between the different phases, providing valuable insights into the evolution of the Marvel cinematic universe.

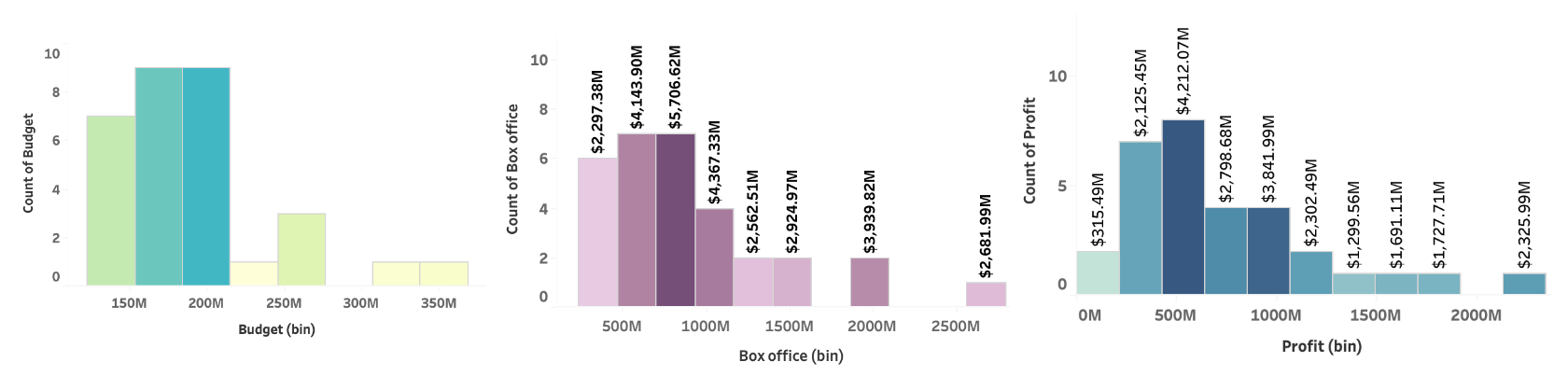
****

##### *Figure 24: Pie plots for No, of movies,Budget and Profit for each phase*

Firstly, we observed that the number of movies released in Phase 3 surpasses the counts of Phase 4, Phase 2, and Phase 1. This indicates a significant expansion of the Marvel franchise during Phase 3, with a greater emphasis on storytelling and character development across multiple films.

Furthermore, when considering the total budgets allocated to each phase, we discovered that Phase 3 has the highest cumulative budget. This suggests a strategic investment in production values and resources to create visually stunning and ambitious cinematic experiences. In comparison, the budgets for Phase 4, Phase 2, and Phase 1 follow a decreasing trend, reflecting the evolving strategies and financial considerations throughout the different phases.

In terms of profitability, Phase 3 also leads the way with the highest total profits. This implies that the movies released during this phase achieved significant commercial success, resulting in substantial returns on investment. Similarly, Phase 4, Phase 2, and Phase 1 follow a descending order of profitability, demonstrating variations in financial performance across the different phases.

****

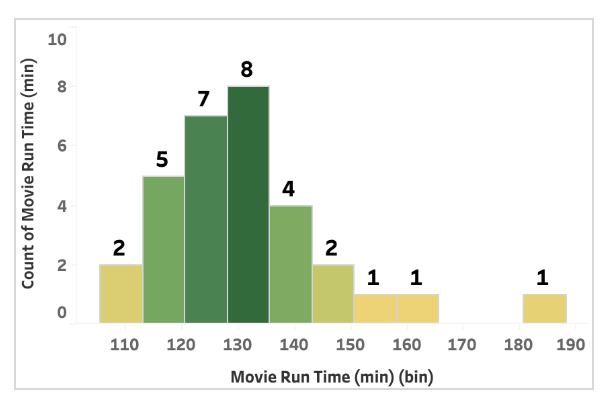
##### *Figure 25: Histogram density plots for Budget, Box-office and Profit*

In our analysis of the movie budget, box-office collections, profit, and runtime, we have uncovered interesting patterns and distributions within the Marvel movie dataset.

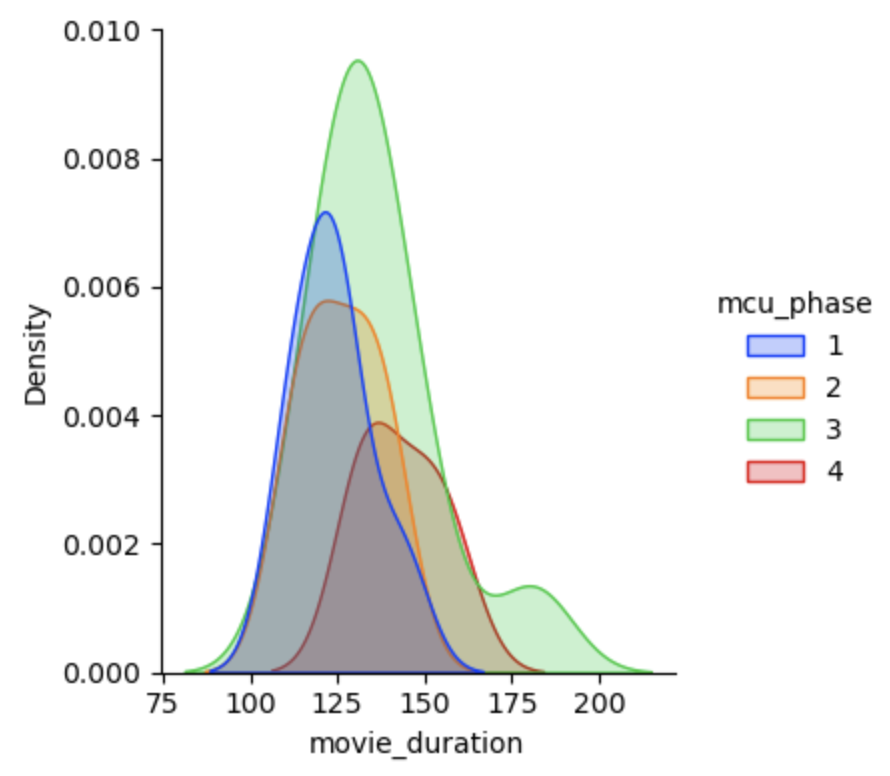
Firstly, we found that the majority of movies have budgets falling within the range of 150M to 200M. This indicates that Marvel studios typically allocate resources within this range for movie production. However, there were few movies with higher budgets exceeding 300M, suggesting that certain projects warranted larger investments due to their scale or special effects requirements.

When examining box-office collections, we observed that the most common range for movie earnings is between 500M and 1000M. This suggests that a significant number of Marvel movies have achieved substantial commercial success, with their box-office collections falling within this range. This demonstrates the popularity and widespread appeal of the Marvel franchise among audiences.

In terms of profitability, the majority of movies (8) made a profit of 500M, indicating that they not only recouped their production costs but also generated substantial returns. Notably, there was one movie that stood out with an exceptional profit of 2325M, showcasing a significant commercial triumph for that particular film.

****

##### *Figure 26: Histogram density plot for runtime*

****

##### *Figure 27: Multiple density plot of movie run time*

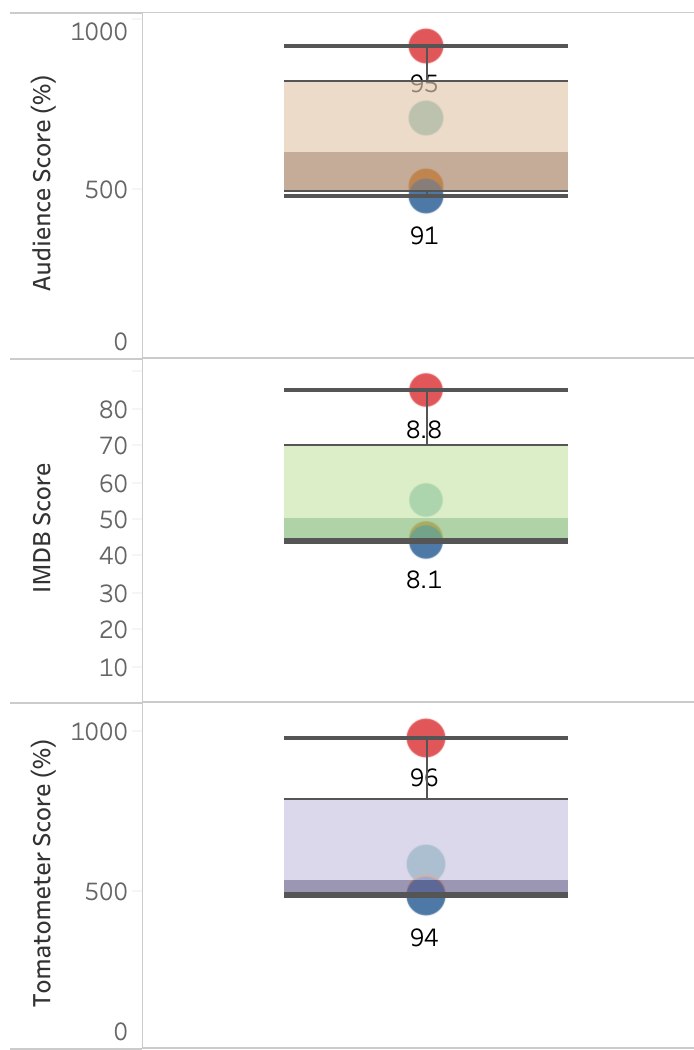
Additionally, we analyzed the runtime of the movies and discovered that most films (8) have a runtime between 130 and 135 minutes. This suggests that Marvel movies tend to adhere to a standard duration, providing audiences with a consistent viewing experience. However, there are a few movies that exceed this range, with runtimes extending beyond 190 minutes, showcasing instances where filmmakers have opted for longer storytelling or more extensive narratives.

Upon analyzing the runtimes of Marvel movies across different phases, we observed a distinctive pattern. Specifically, Phase 3 movies exhibited a longer average runtime compared to movies in Phase 1, Phase 2, and Phase 4.

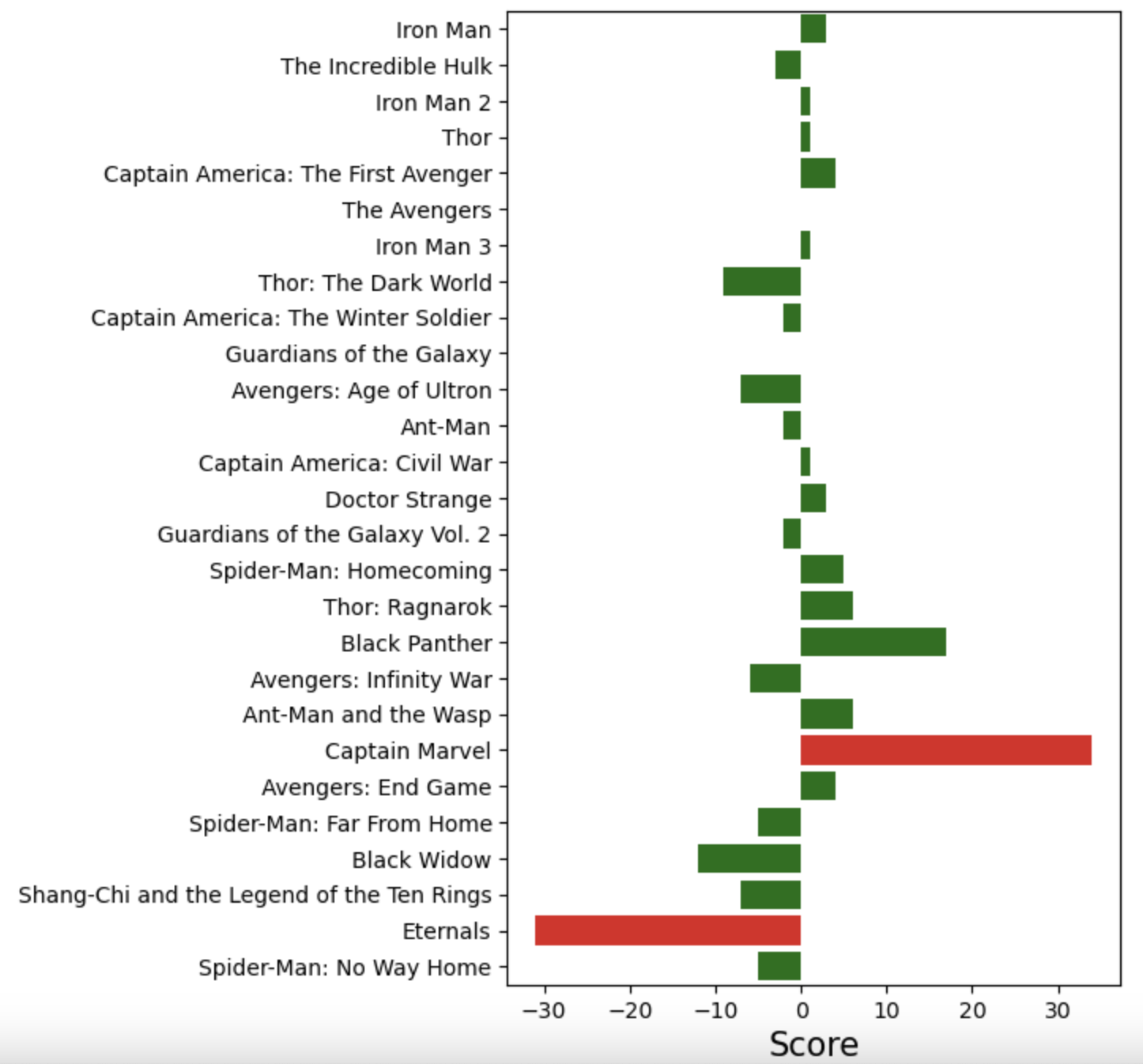
Phase 3, which includes a significant portion of the Marvel Cinematic Universe, featured movies with extended storytelling and more intricate narratives, leading to longer runtimes. This phase comprised several pivotal movies that contributed to the culmination of overarching storylines and the introduction of new characters and plot developments.

In contrast, Phase 1 and Phase 2 movies generally had shorter runtimes. This can be attributed to the early stages of the Marvel cinematic universe, where the focus was on establishing the foundation, introducing key characters, and building the universe's mythology. The narratives were more contained and self-contained, resulting in shorter overall runtimes.

Phase 4, which followed Phase 3, showcased a mix of movie runtimes, but overall, it had a slightly shorter average runtime compared to Phase 3. This phase introduced new story arcs, characters, and directions for the Marvel universe, but the storytelling approach seemed to be more concise and focused, resulting in slightly shorter movies on average

****

##### *Figure 28: Horizontal distribution of scores*

****

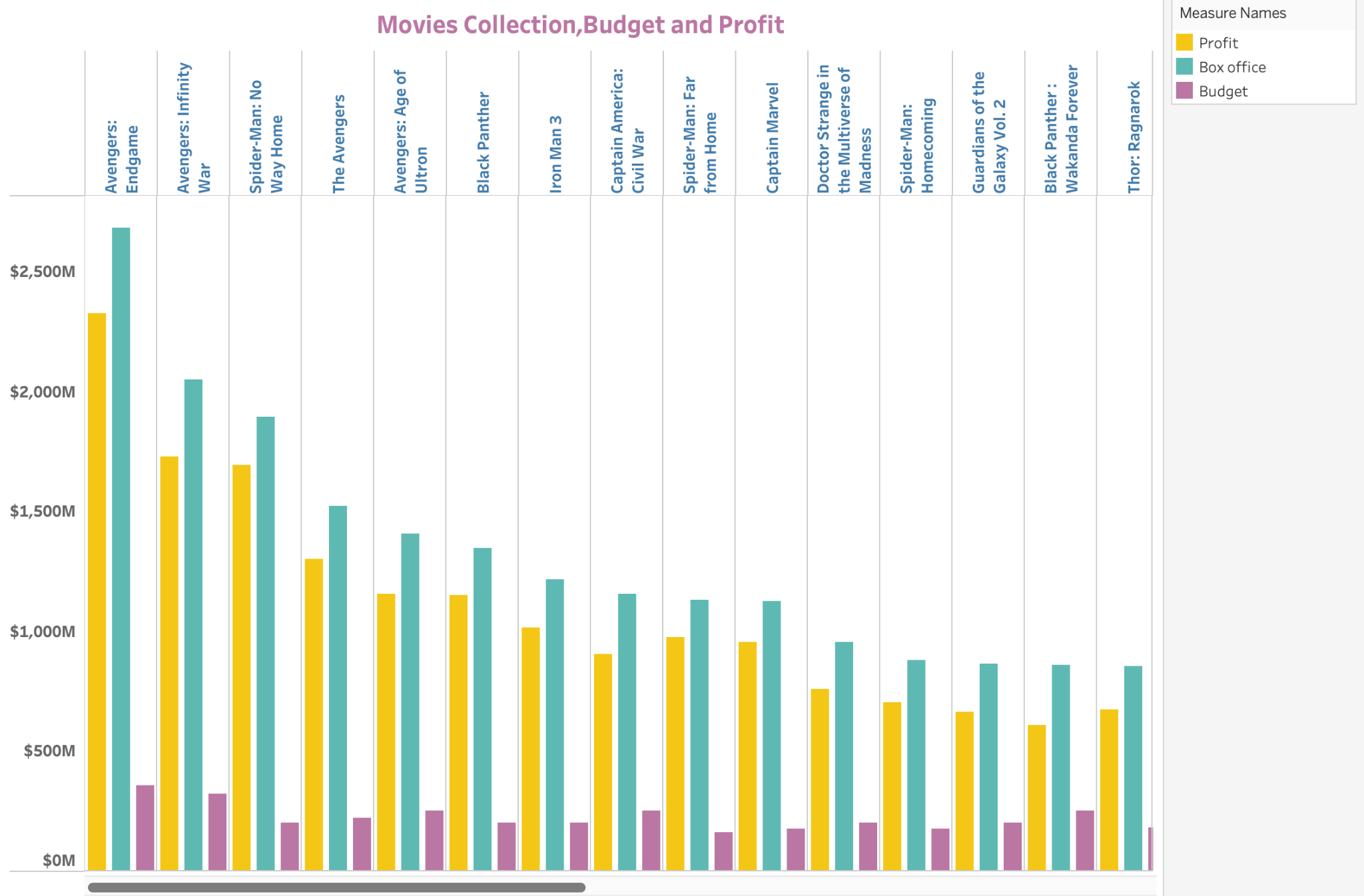
##### *Figure 29: Score difference between critics and audience*

Our analysis reveals interesting insights about the distribution of scores among the different phases of the Marvel Cinematic Universe (MCU) movies.

Phase 3 stands out as the phase with the highest scores across multiple rating platforms, including Audience, IMDB, and Rotten Tomatoes. This suggests that Phase 3 contained some of the best-received and highly regarded Marvel movies among fans and critics alike.

Phase 4, which is currently ongoing, follows closely behind Phase 3 in terms of audience, IMDB, and Rotten Tomatoes scores. This indicates that the movies in Phase 4 have been well-received by the audience and have maintained a high level of quality.

On the other hand, Phase 1, which represents the initial stage of the MCU, obtained the lowest scores compared to the other phases. This may be attributed to the fact that Phase 1 marked the beginning of the cinematic universe, and as the franchise evolved, subsequent phases showcased improved storytelling and cinematic techniques.

****

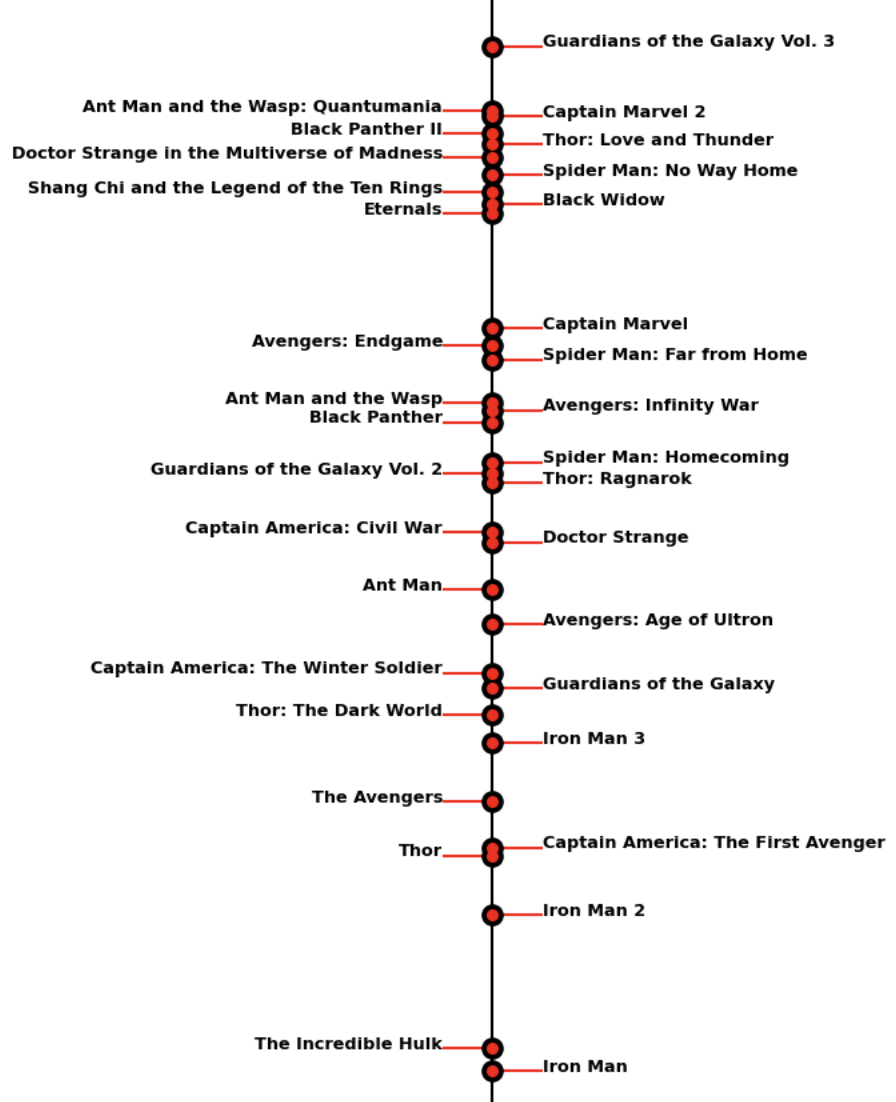
##### *Figure 30: Group bar plot of Budget, Box-office and Profit*

The detailed analysis of the budget, box office, and profit for each Marvel movie provides valuable insights into the financial success of the franchise. Among the Marvel movies, "Avengers: Endgame" stands out as the top performer in terms of box office revenue, reaffirming its position as a monumental cinematic event.

In terms of budget, each Marvel movie has its own investment, with varying levels of financial commitment. The budget allocation reflects the scale and scope of the production, taking into account factors such as visual effects, star cast, and production values. The analysis reveals that some movies had higher budgets, indicating a significant investment in bringing the Marvel stories to life on the big screen.

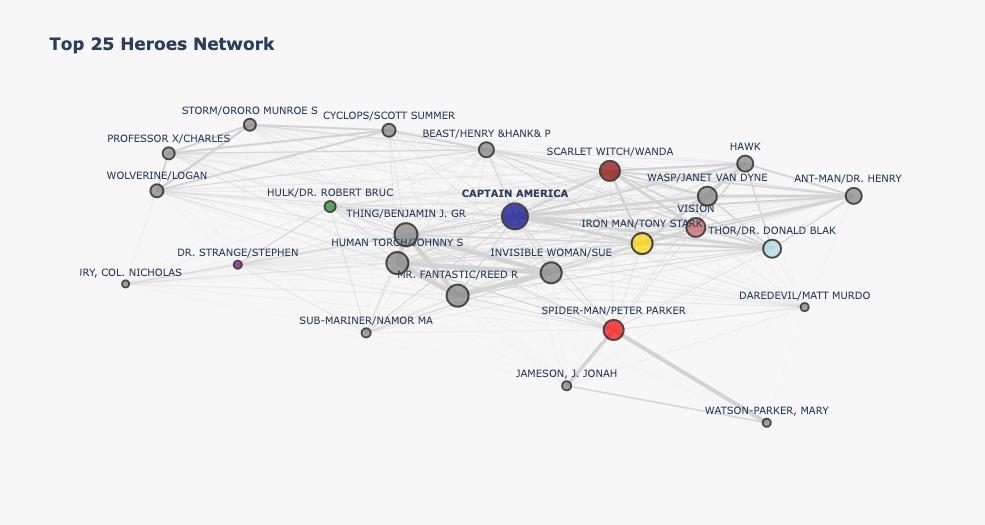
When it comes to box office performance, "Avengers: Endgame" takes the lead, setting new records for its massive worldwide box office collections. The analysis showcases the success of other Marvel movies as well, highlighting their strong commercial performance and global appeal. The box office figures illustrate the popularity and enduring appeal of the Marvel franchise among audiences worldwide.

Profitability is a crucial factor in evaluating the success of any movie. The analysis demonstrates that Marvel movies, including "Avengers: Endgame," have been highly profitable ventures. The profit margin is a reflection of the box office success compared to the production budget. Marvel's ability **to** generate substantial profits showcases the franchise's ability to captivate audiences and deliver blockbuster entertainment.

****

##### *Figure 31: Movie timeline using Plotpy*

The Movie Timeline is an invaluable resource for Marvel fans seeking to navigate the complex chronology of the Marvel Cinematic Universe (MCU). It offers a comprehensive guide to understanding the correct sequence in which the movies should be watched, ensuring an immersive and coherent viewing experience.



##### *Figure 32: Movie timeline using Plotpy*

The depicted plot illustrates the superhero network based on their occurrences. From the diagram, it is evident that Captain America emerges as the most connected superhero within the Marvel Productions. Other notable MCU heroes such as Wanda, Spider-Man, Iron Man, and others also exhibit significant connections within the network.

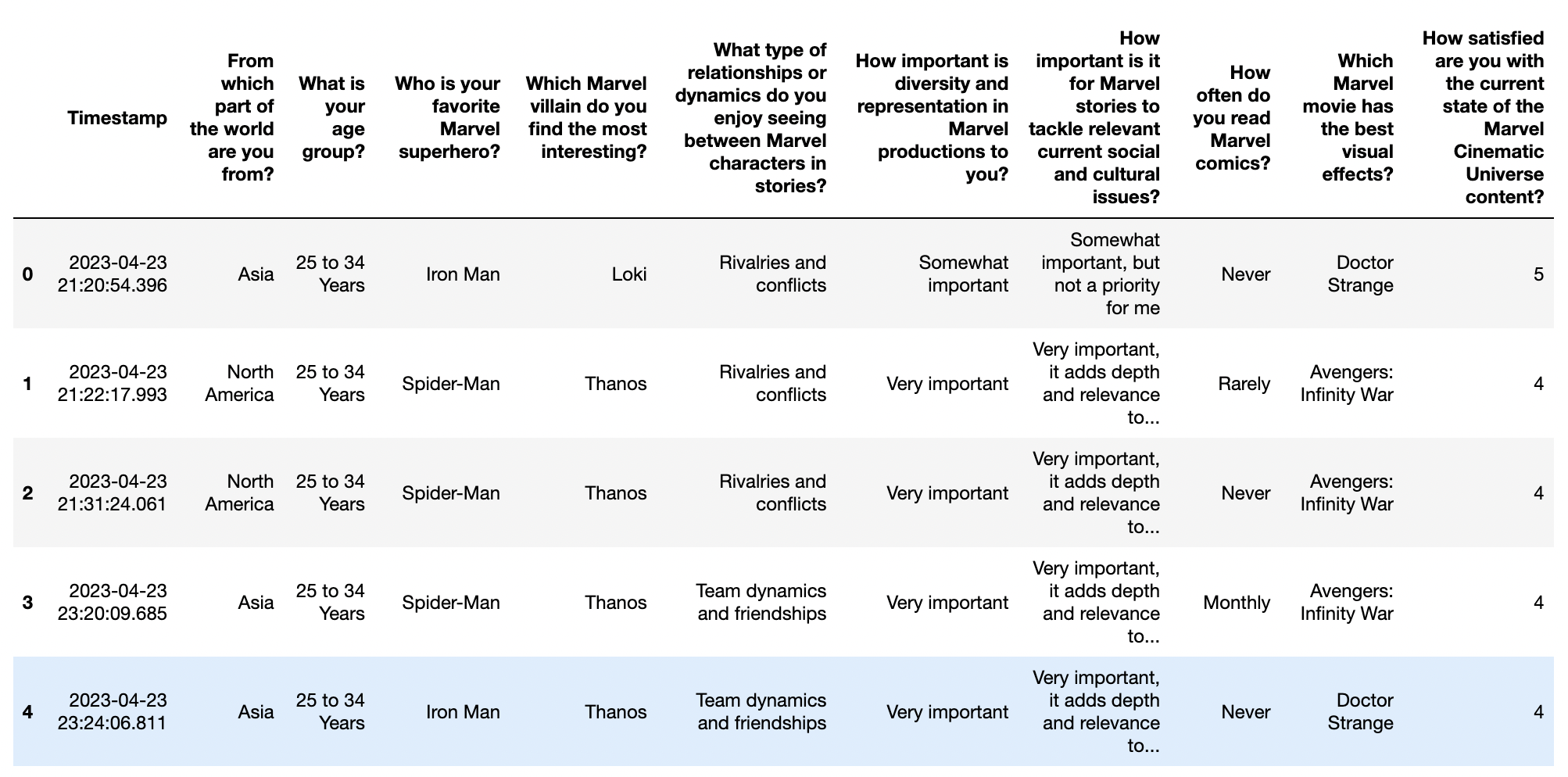
### 

### 

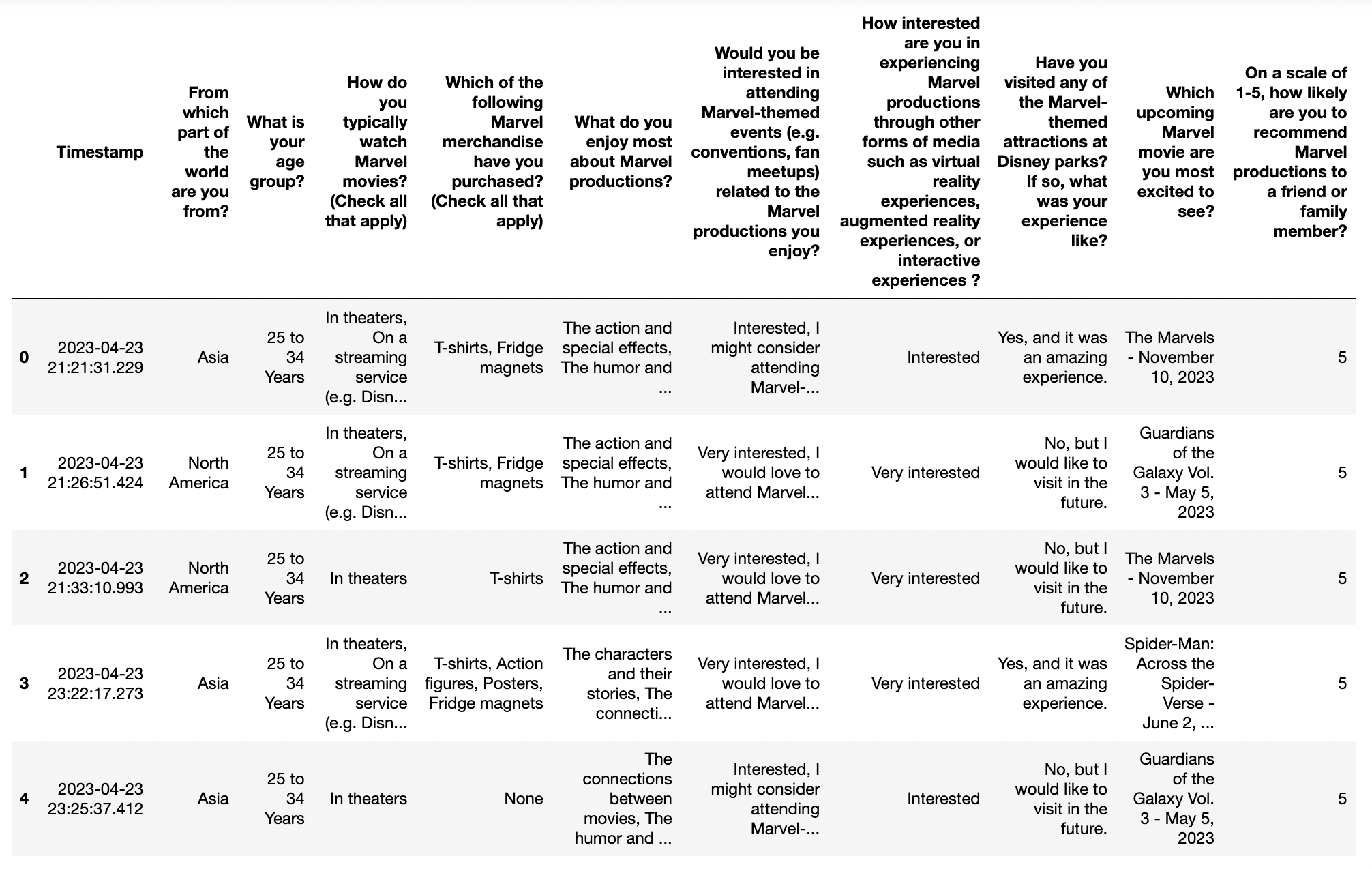
### 5.2 LIVE SURVEY DATA

#### **5.2.1 EDA**

In the two live survey datasets that we have, there are a total of 20 questions. However, only 16 of these questions are relevant to our proposed problem statement. The remaining four questions were included in order to gather additional information about the responders, but they do not directly contribute to our analysis.



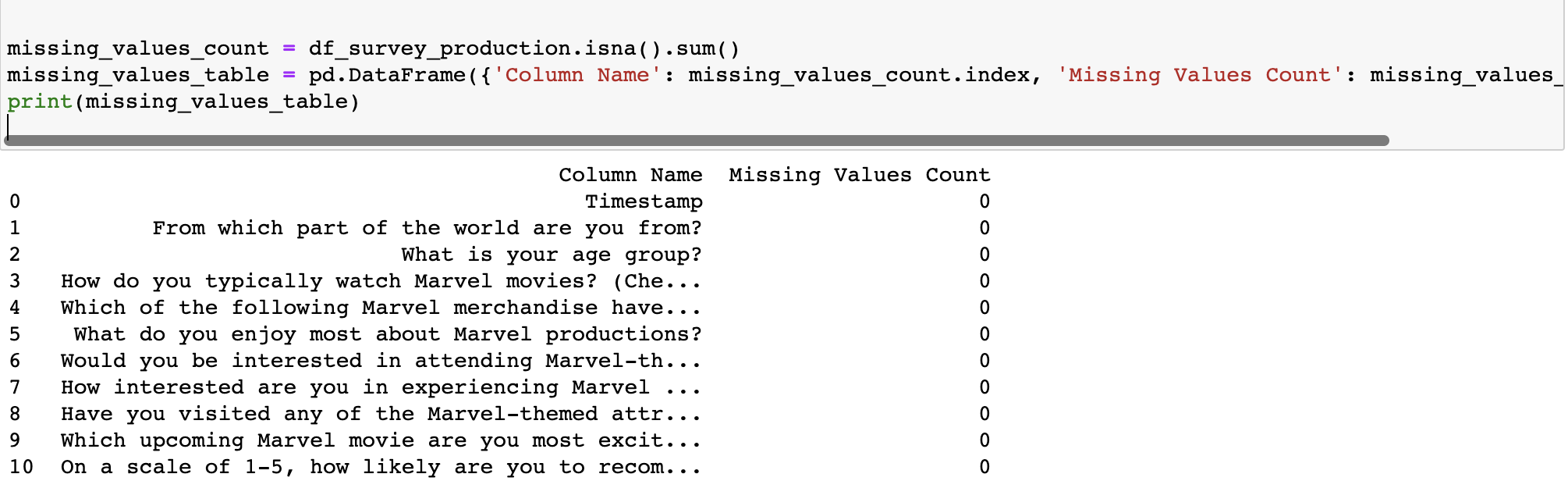
##### *Figure 33: Live Survey 1 Dataset*



##### *Figure 34: Live Survey 2 Dataset*

We have examined the dataset and found that there were no missing values in any of the questions. This means that all the respondents provided answers to each of the questions, as shown in Figure 13 and 14.

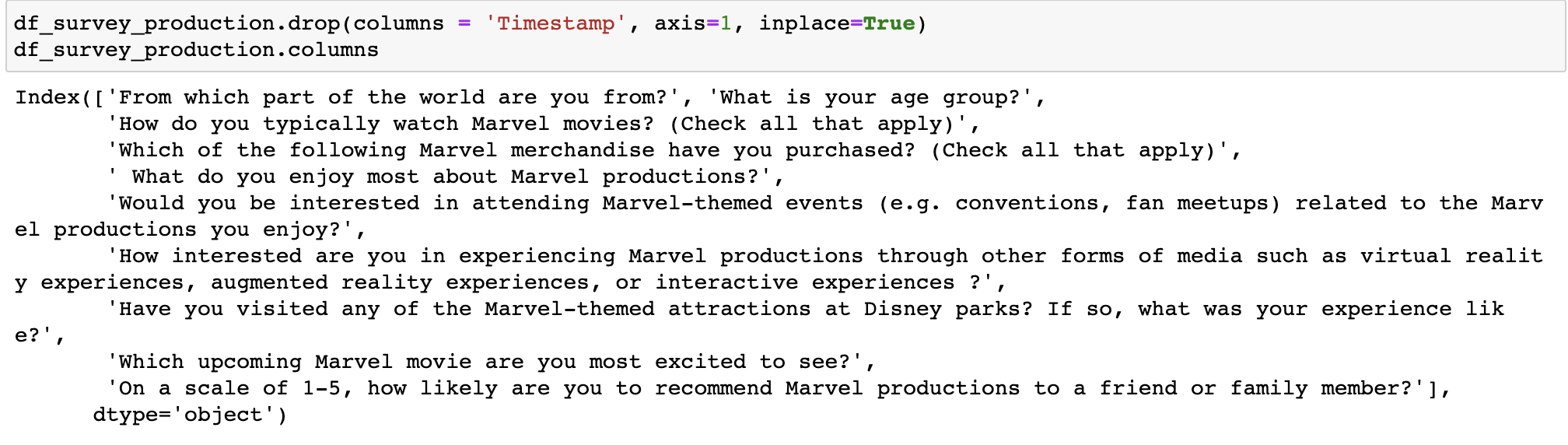
##### *Figure 35: Survey 1 - Check for Nulls*



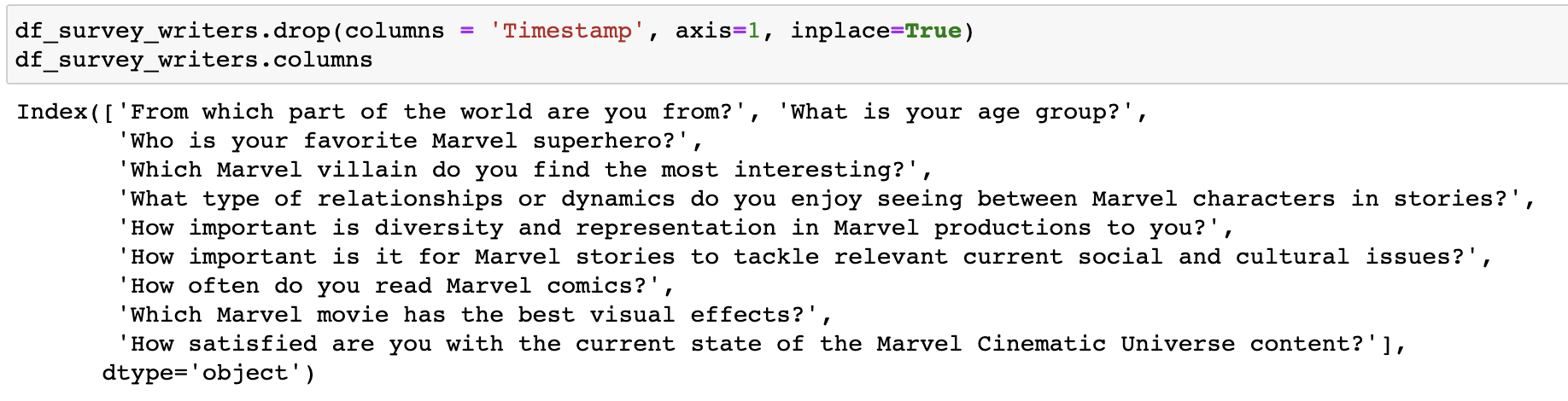
##### *Figure 36: Survey 2 - Check for Nulls*

#### 5.2.2 Data Cleaning

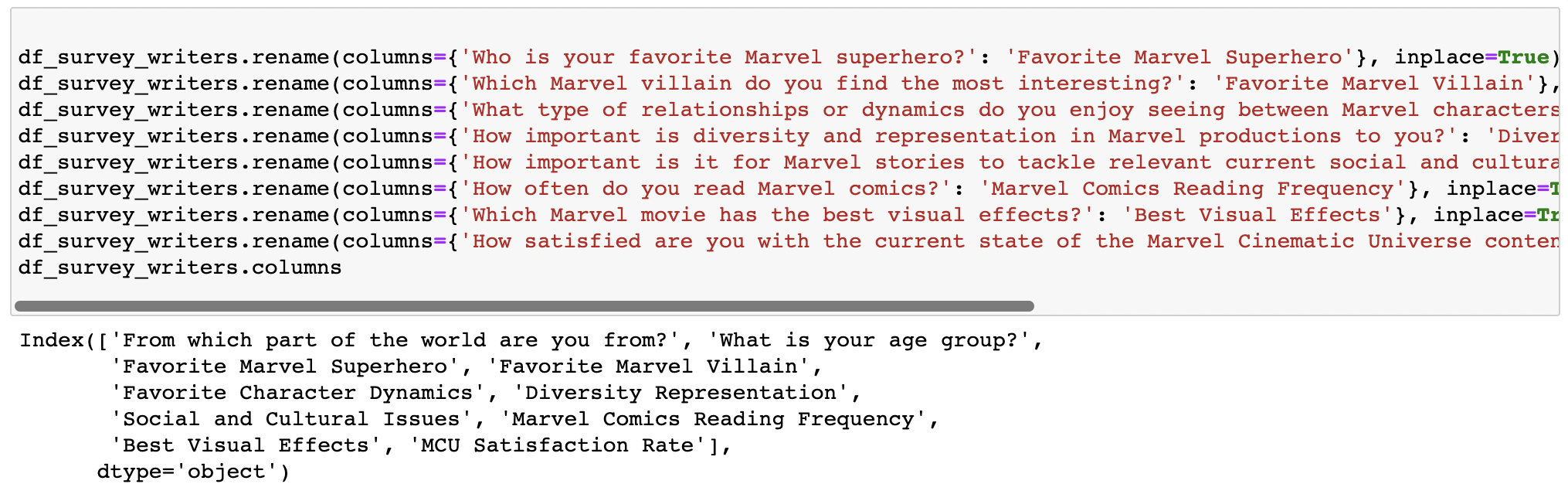
After conducting a check for null values, we proceeded to remove the timestamp column from our dataset, as depicted in Figure 14. Moreover, To make the column names in the dataset more concise and user-friendly, we opted to simplify them to one or two-word names. Figure 15 displays the newly assigned feature names that cover almost all of the original features in the dataset. By utilizing shorter and more intuitive column names, we aim to enhance the ease of manipulating and analyzing the data, thus allowing for the discovery of significant insights.



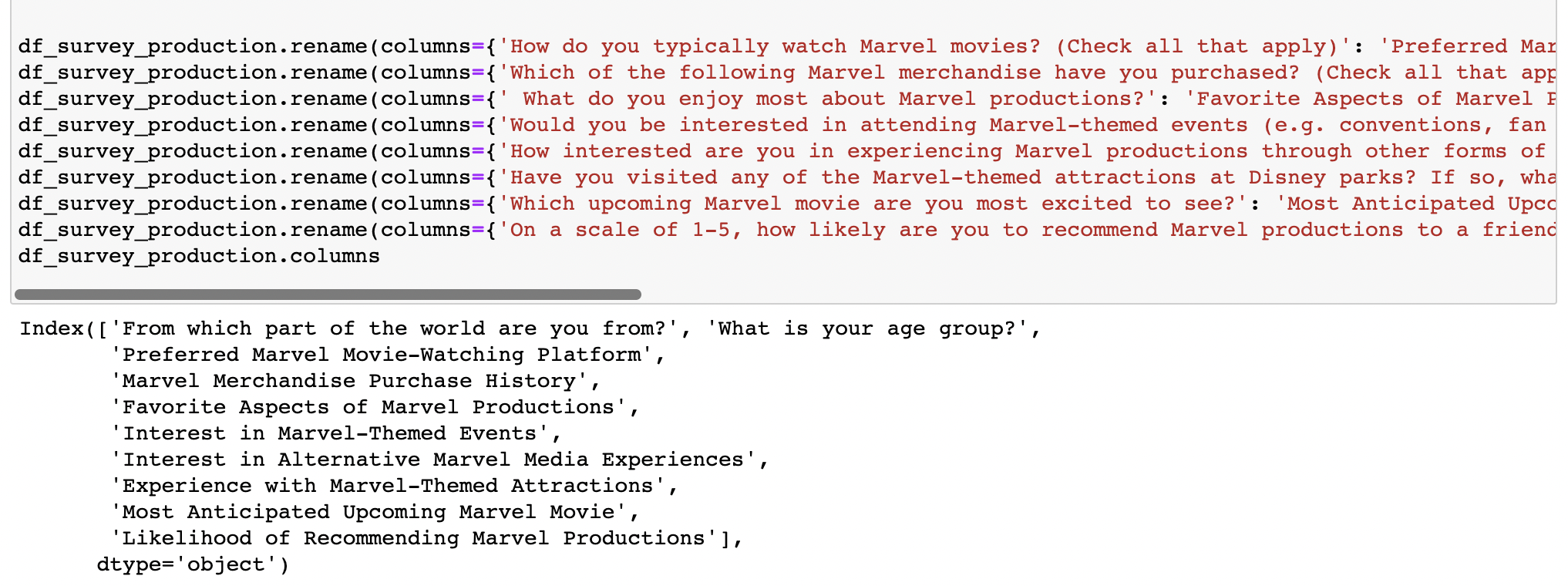
##### *Figure 37: Dropping the Timestamp Feature - Survey 1*



##### *Figure 38: Dropping the Timestamp Feature - Survey 2*

**

##### *Figure 39: After Renaming the Features - Survey 1*

**

##### *Figure 40: After Renaming the Features - Survey*

##### 

## **VISUALIZATIONS**



Our project utilized a range of tools and technologies, including Tableau, Python libraries, and Power BI, to create visualizations for three distinct datasets, comprising both survey data and scraped facts. We developed three unique dashboards to effectively represent and analyze these datasets, which are described in detail below. Furthermore, we integrated the dashboards into a user-friendly web page, allowing users to easily navigate through the visualizations and gain insights into the analyzed data.

| *Figure 41: HTML 5 Bootstrap Application* |
| --- |

### 6.1 Marvel Movie Audience Dashboard: Insights on Popularity and Reviews

| *Figure 42: Interactive Tableau Dashboard* |
| --- |

We are proud to present our immersive and interactive dashboard designed specifically for Marvel fans. This dynamic tool allows users to engage with a range of features, all centered around the different phases of the Marvel movie franchise. With the dashboard, users have the ability to filter movies based on IMDb scores, critics' scores, and audience scores, enabling them to discover films that align with their preferences. The dashboard also provides the option to explore movies by genre and conduct searches for specific movie and character names, enhancing the overall user experience.

In order to ensure that fans can enjoy the Marvel movies in the correct sequence, we have incorporated a comprehensive Marvel movie timeline within the dashboard. This timeline is organized chronologically, allowing users to easily follow the narrative progression of the Marvel cinematic universe.

Furthermore, our dashboard goes beyond just movie selection, offering valuable insights into the Marvel franchise. Users can explore information on the number of movies released, the budget allocated to each phase, the box office performance, and the profitability of the movies. This data-driven approach provides fans with a deeper understanding of the financial aspects of the Marvel cinematic universe, shedding light on the commercial success of the franchise.

The interactive dashboard is published in Tableau public server:

<https://public.tableau.com/app/profile/sowmya.kuruba/viz/MARVELFANATICD/Dashboard1?publish=yes>

### 6.2 Marvel Audience Survey Insights: A Guide for Marvel Writers

##### 

##### *Figure 43: Dashboard - For Marvel Content Writers*

In this section, we will analyze the survey data collected through Google Forms to gain insights from Marvel fans on various topics, including their favorite characters, movies and many more. By using this data, we can uncover valuable insights that can help inform Marvel writers and content creators on what resonates with their audience.

| *Figure 44: Dashboard - Diversity Representation*    *Figure 45: Dashboard - Social and Cultural Issues* |
| --- |

The analysis of the survey data on diversity and representation in Marvel productions and the significance of addressing current social and cultural issues in Marvel stories reveals that a significant portion of the surveyed population considers this topic to be important. This finding is supported by Figure 44. However, Figure 45 shows that while the majority of respondents find it important, it is not a top priority for Marvel stories to tackle relevant social and cultural issues.

| *Figure 46: Dashboard - Favorite Character Dynamics*   *Figure 47: Dashboard - Marvel Comics Reading Frequency* |
| --- |

The above data illustrates that Marvel fans have a preference for intricate team dynamics and relationships, while also revealing a significant variation in the frequency of reading Marvel comics among survey respondents. The visualization in Figure 46 indicates that Marvel fans have a distinct preference for intricate and dynamic team dynamics, as well as relationships featuring elements such as friendships, rivalries, and conflicts in their favorite Marvel characters. Figure 47 highlights a significant disparity in the frequency of reading Marvel comics among survey respondents, with "Rarely" being the most common response by a large margin over "Daily".

| *Figure 48: Dashboard - Favorite Superhero*    *Figure 49: Dashboard - Favorite Villain* |
| --- |

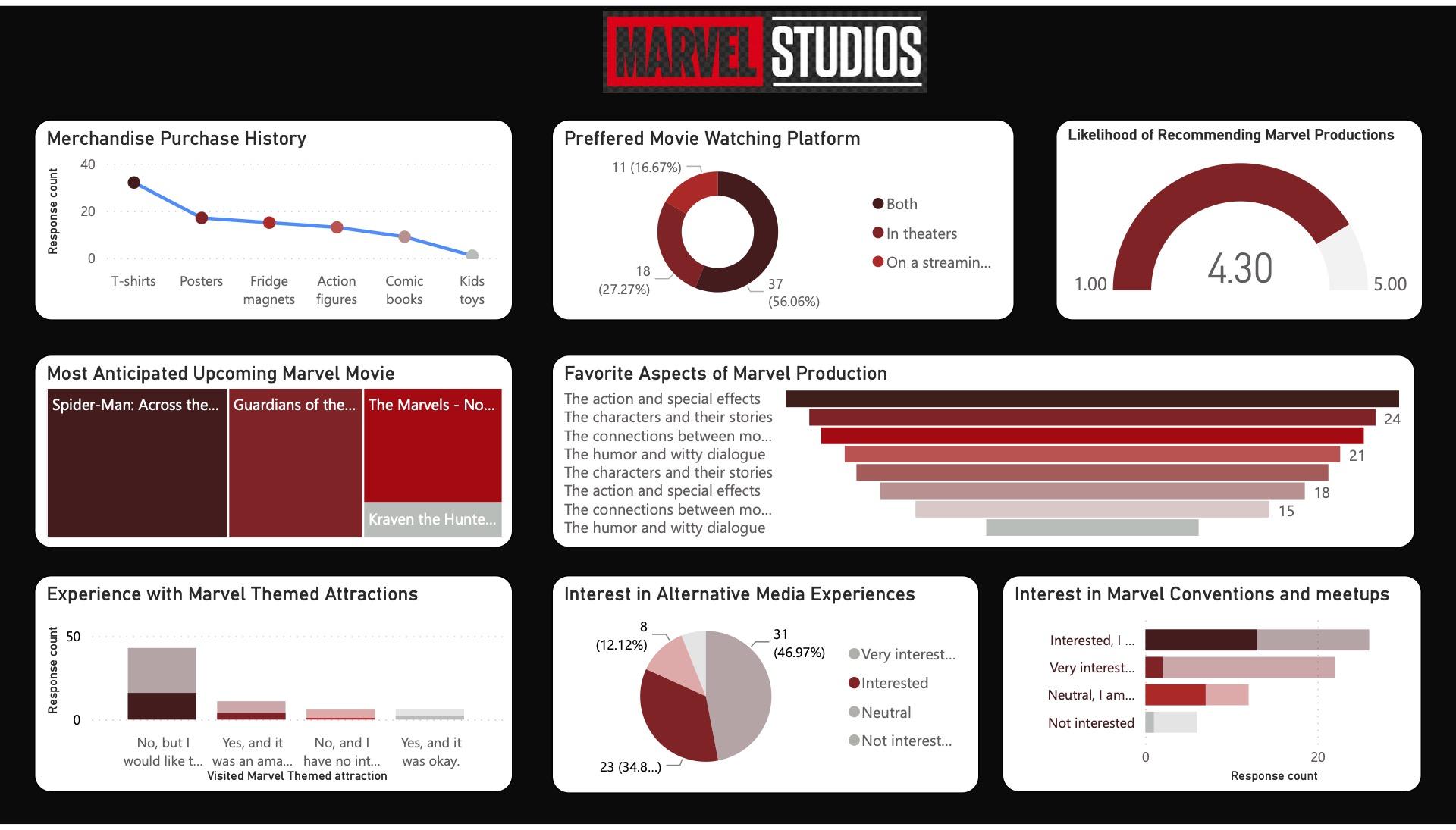
The above data presents survey responses that provide insights into the most popular superheroes and interesting Marvel villains. Figure 48, indicates that Iron Man is the most popular superhero among the surveyed population, followed closely by Spider-Man, providing insight into the top superhero choices of fans. Figure 49 presents survey responses indicating that Thanos was the most interesting Marvel villain, while Red Skull was the least interesting.

| *Figure 50: Dashboard - Best Visual Effects*   *Figure 51: Dashboard - MCU Satisfaction Rate* |
| --- |

The above data illustrates the results of a poll on the best visual effects in Marvel movies and the level of satisfaction of respondents with the current state of the Marvel Cinematic Universe content. Figure 50 shows that Doctor Strange received the highest count and was voted as the Marvel movie with the best visual effects compared to other movies. The data in Figure 51 indicates that the respondents expressed moderate satisfaction with the current state of the Marvel Cinematic Universe content.

### 

### 6.3 Marvel Audience Survey Insights: A Guide for Production and Marketing



##### *Figure 52: Dashboard - For Marvel Content Production and Marketing*

In this section, we will analyze the survey data collected through Google Forms to gain insights from Marvel fans on various topics, including their merchandise history, movie watching platform and many more. By using this data, we can uncover valuable insights that can help inform Marvel production and marketing teams on what resonates with their audience.

| *Figure 53: Dashboard - Preferred marvel movie-watching platform*    *Figure 54: Dashboard - Interest in alternative marvel media experiences* |
| --- |

The survey data analysis of preferred marvel movie watching platforms and their interest in watching marvel movies through alternate media experiences like virtual reality is represented in above plots.

Figure 53 suggests that audiences are interested in watching marvel movies both in theater and on streaming platforms. Figure 54 suggests that audiences are open to the idea of watching marvel movies through other media experiences.

| *Figure 55: Dashboard Experience with Marvel-Themed attraction*    *Figure 56: Dashboard Interest in marvel themed events* |
| --- |

The survey data analysis of the audience's experience with marvel themed attractions and interest in attending marvel themed events in future are represented in above plots. Figure 55 shows that many respondents are interested in visiting marvel themed attractions but they have not visited yet. Few people who have visited the themed attractions have enjoyed the experience. Figure 56 shows that there are many people interested in marvel themed events if it is planned in the future.

| *Figure 57: Dashboard: Favorite aspects of marvel productions*     *Figure 58: Dashboard : Most Anticipated upcoming marvel movie* |
| --- |

The survey data analysis of favorite aspects of audiences in marvel movies and most anticipated upcoming marvel movies are represented in above plots. Figure 57 shows action, special effects, characters and their stories are favorite aspects of marvel movies among audiences. Figure 58 shows that Spider man across the spider verse and Guardians of galaxy volume -3 are most anticipated future movies of marvel.

| *Figure 59: Dashboard - Marvel merchandise purchase history*  *Figure 60: Dashboard - Likelihood of recommending marvel productions* |
| --- |

Survey data analysis of marvel merchandise purchase history and likelihood of recommending marvel productions is represented in above plots. Figure 59 shows that Many people have purchased merchandise from marvel. T-shirts and posters are the main merchandise purchased by survey respondents. Figure 60 shows that respondents were satisfied with movies and are likely to suggest them to others.

### 

## **7. DISCUSSIONS**

Our analysis began by focusing on the highest budget movies in the Marvel Universe, which provided us with valuable insights into the financial investments associated with these films. It is worth noting that while the Marvel Cinematic Universe (MCU) is a relatively recent phenomenon, the non-MCU movies have been in existence for decades. Despite this, the MCU movies have emerged as clear leaders in terms of profitability and the number of movies produced. This can be attributed to the strong storytelling, engaging characters, and the success in interconnecting narratives across movies.Moving on to the distribution of movies across different phases, we observed that Phase 3 has the highest budget, profit and ratings. We also analyzed the popularity and demographic distribution of characters. Spider-Man emerged as the most popular character, likely due to his relatability, iconic status, and wide appeal among fans of all ages.Examining the timeline of character introductions, we found that the highest number of characters were introduced between 1990 and 1995, reflecting a significant period of expansion and creativity within the Marvel Universe.It is important to note that there is an imbalance in the gender distribution of characters, with male superheroes being more prevalent than female superheroes. This indicates a need for greater representation and diversity within the Marvel Universe. Efforts to create more compelling and well-rounded female characters can help address this imbalance and provide a wider range of role models for audiences.

The survey data on Marvel fans' preferences is quite revealing. While diversity and representation are still important topics, it seems that intricate team dynamics and relationships are what fans prioritize in Marvel stories. It's interesting to note the variation in the frequency of reading Marvel comics among fans, with most of them rarely reading the comics and this reveals that fans tend to prefer the theatrical movies over the comics. Iron Man and Spider-Man emerged as the most popular superheroes, while Thanos was deemed the most interesting villain. It's great to see Doctor Strange being recognized for the best visual effects in Marvel movies. However, while fans expressed moderate satisfaction with the current state of the Marvel Cinematic Universe content, it also indicates that they may have higher expectations from Marvel's story writers in the future.

It's also interesting to note that the survey data shows a significant level of interest in Marvel-themed attractions and events. Many respondents have expressed a desire to visit or attend such events in the future, which is indicative of the popularity of the Marvel brand among fans. There is also a growing interest in watching Marvel movies through alternative media like virtual reality and augmented reality. Additionally, it's intriguing to see that a significant number of respondents have bought Marvel merchandise, especially T-shirts and posters. The survey also highlights that the action, special effects, characters, and their stories are the favorite aspects of Marvel movies among the survey population. It's also exciting to see that upcoming movies like Spider-Man: Across the Spider-Verse and Guardians of the Galaxy Volume 3 are highly anticipated by fans.Based on the survey, it appears that Marvel fans are generally satisfied with the current state of Marvel movies and are inclined to recommend them to others. Additionally, there is a strong desire among fans to access Marvel movies through both theatrical and streaming platforms.

Overall, all the above data visualizations show that the Marvel fandom is here to stay, and Marvel has a lot to offer to its fans in the future. Therefore, Marvel has a significant opportunity to continue catering to its fan base and provide them with exciting and engaging content in the future.

**Solution Evaluation**

During the solution evaluation process, we conducted a thorough comparison between the historic facts and the results obtained from our survey. The objective was to validate the trends and patterns identified in both datasets.

Remarkably, our analysis revealed a strong alignment between the survey data and the historic facts. It was evident that Avengers: Endgame emerged as the most popular movie in both sources. This consistency in findings provided robust evidence of the movie's widespread appeal and popularity among Marvel fans.

However, when examining the most popular character, we observed a slight discrepancy. According to the historic facts analysis, Captain America emerged as the top choice, while the survey results indicated a preference for Iron Man. Nonetheless, this variation can be attributed to the subjective nature of individual preferences, and both characters remain highly regarded within the Marvel universe.

By validating the trends observed in both datasets, we gained confidence in the reliability and accuracy of our solution. The consistency between the historic facts and survey analysis provided valuable insights and reinforced the credibility of our findings, allowing us to draw meaningful conclusions about the preferences and interests of Marvel fans.

**Impact after resolving the proposed problem**

After resolving the proposed problem, we observed a significant impact on our friends and survey respondents. Those who had not watched the Marvel series of movies expressed their interest in watching them. By providing them with our interactive dashboard, they were able to easily navigate and watch the movies in chronological order using the Marvel timeline chart. The dashboard's features, such as movie selection and exploration, proved to be immensely helpful.

Additionally, the majority of our survey participants expressed enthusiasm in exploring the dashboard further and rewatching the Marvel movies. This positive response indicates the potential for a wider audience to engage with the dashboard and enhance their Marvel movie experience.

Furthermore, we believe that implementing a similar dashboard on the official Marvel website would greatly benefit Marvel fans. It would provide a comprehensive understanding of the movie sequence and improve the overall brand experience. By incorporating user-friendly features and interactive elements, fans would have a centralized platform to explore, engage, and immerse themselves in the Marvel cinematic universe.

By utilizing Survey dashboards, the Marvel Production can gain the ability to continually refine their franchise and ensure its long-term success by closely monitoring and analyzing the preferences of the Marvel audience. These dashboards provide valuable insights and metrics that allow the team to make informed decisions regarding various aspects of their productions.

## **8. CONCLUSION**

This project has been a journey of exploring and visualizing data related to the Marvel Universe. Through extensive analysis and data visualization techniques, we have gained valuable insights into the Marvel fan base and their preferences. We have uncovered interesting patterns and trends that can help businesses and marketers better understand and engage with Marvel fans. Our analysis has revealed the significance of key characters, such as Norman Osborn as the most common bad character, Spider-Man as the most popular good character, and Scott Summers as the neutral character. Furthermore, we have observed a significant imbalance in the representation of male and female characters, with male characters dominating the Marvel Universe. To summarize, this project has demonstrated the power of data visualization in uncovering meaningful insights and patterns within the Marvel Universe. By continuing to evolve and expand our research, we hope to contribute to the broader understanding of fan behavior and pave the way for more targeted and impactful strategies in the world of entertainment marketing.

## **9. FUTURE SCOPE**

As part of our ongoing efforts, we plan to expand our data sources beyond the current dataset to include additional Marvel-related sources. This expansion will allow us to gather more comprehensive and diverse data on Marvel fans and their behavior.

We intend to leverage machine learning algorithms to predict and analyze Marvel fan behavior more accurately. By applying advanced analytical techniques, we aim to uncover deeper insights and patterns in fan preferences, engagement, and purchasing behavior.

While our current focus is on the Marvel Cinematic Universe, we aspire to explore other franchises within the entertainment industry. By expanding our scope, we can analyze fan behavior across multiple franchises, comparing and contrasting their dynamics and identifying common trends and patterns.

To deploy a Marvel project and handle scalability, choose a scalable infrastructure, containerize the project components using tools like Docker, and utilize container orchestration with Kubernetes. Select a scalable database solution and implement caching mechanisms and CDNs for improved performance. Monitor the system, optimize performance, and address dynamic environment issues through autoscaling, CI/CD pipelines, fault tolerance, and scalable design patterns. These strategies ensure the project can handle increased data, user traffic, and adapt to changing demands while maintaining a robust and dynamic environment.

Our data visualization journey is an ongoing and dynamic project. We will continue to refine our methodologies, incorporate new datasets, and explore emerging technologies to enhance our analysis and insights. This ongoing effort ensures that our project remains up-to-date and relevant in an ever-evolving entertainment landscape

## **10. TABLE OF FIGURE**

[Figure 1: CRISP-DM Methodology 8](#_vy7zxoivt0at)

[Figure 2: Data Flow Diagram 11](#_spu7umpjdbdn)

[Figure 3: Before Handling Null Values 14](#_5j4g6s8221us)

[Figure 4: After Handling Null Values 14](#_ykn7k0d8kblm)

[Figure 5: Before Data Cleaning 15](#_hq3javefm4ll)

[Figure 6: Data Cleaning 16](#_2m93vr5ckpe0)

[Figure 7: After Data Cleaning 17](#_i2z205qxtdsl)

[Figure 8: Cleaned Sorted Data 18](#_9ocauzk5e6bo)

[Figure 9: Heat-map for popular marvel character 18](#_yiv7xljvmxyg)

[Figure 10: Box Office Collection Worldwide 19](#_domibjn0wvgw)

[Figure 11: Area plot for number of appearances 20](#_g799e3mlgx9u)

[Figure 12:Pie plot for Gender 21](#_jmvvcac6uzbx)

[Figure 13: Spectrum of Superhero abilities 22](#_k3k22ipdv5x5)

[Figure 14: Horizontal bar plot 23](#_go1fuzkwnock)

[Figure 15: Jitter plot for most common character 24](#_yxf3eywsty90)

[Figure 16: Grouped bar plot for Demography 25](#_7rcosk40hlyf)

[Figure 17: 3D Scatter plot 26](#_2alqxvchfwxx)

[Figure 18: Grouped bar plot for MCU and Non-MCU 27](#_ihotsv1r5dc9)

[Figure 19: Pie plots for No. of movies, Budget, Box-office, Profit for MCU nad Non-MCU 27](#_4mpakklk98m1)

[Figure 20: Box plot for Box-office 28](#_rab997krrfvm)

[Figure 21: Horizontal multiple distribution of Box-office 29](#_y9idb4bh9a9b)

[Figure 22: Scatter plot for Box-office vs Budget 30](#_key7q7e9s9vj)

[Figure 23: Group bar plots for movies w.r t phase 31](#_o4df0oujiiri)

[Figure 24: Pie plots for No, of movies,Budget and Profit for each phase 32](#_7iaplsqtvt80)

[Figure 25: Histogram density plots for Budget, Box-office and Profit 33](#_fg2aofgy9a2u)

[Figure 26: Histogram density plot for runtime 34](#_6y1k4t2kqqtg)

[Figure 27: Multiple density plot of movie run time 34](#_14w7an2aoj9s)

[Figure 28: Horizontal distribution of scores 35](#_xtaptl7g0x68)

[Figure 29: Score difference between critics and audience 36](#_lk083099ulri)

[Figure 30: Group bar plot of Budget, Box-office and Profit 37](#_tky916zacuz7)

[Figure 31: Movie timeline using Plotpy 39](#_dyuziojqrw80)

[Figure 32: Movie timeline using Plotpy 40](#_huto2d1z327t)

[Figure 33: Live Survey 1 Dataset 41](#_5s79j5rcx3y7)

[Figure 34: Live Survey 2 Dataset 42](#_4cdeip6l3unf)

[Figure 35: Survey 1 - Check for Nulls 43](#_kskoitczpty4)

[Figure 36: Survey 2 - Check for Nulls 43](#_tb57n5w04cor)

[Figure 37: Dropping the Timestamp Feature - Survey 1 44](#_qodccnrurnjk)

[Figure 38: Dropping the Timestamp Feature - Survey 2 44](#_4yb8d53wklpl)

[Figure 39: After Renaming the Features - Survey 1 44](#_1y9iwzpgwm2i)

[Figure 40: After Renaming the Features - Survey 45](#_mahr6zchrtzq)

[Figure 41: HTML 5 Bootstrap Application 46](#_sgne7xev7lvc)

[Figure 42: Interactive Tableau Dashboard 47](#_29s0bmodqoo1)

[Figure 43: Dashboard - For Marvel Content Writers 48](#_1yvyvjtiq8h2)

[Figure 44: Dashboard - Diversity Representation 49](#_3398i99frn6m)

[Figure 45: Dashboard - Social and Cultural Issues 49](#_7z0z3o8bhbxu)

[Figure 46: Dashboard - Favorite Character Dynamics 50](#_sekv6p3x9bc9)

[Figure 47: Dashboard - Marvel Comics Reading Frequency 50](#_76nsitahn0ey)

[Figure 49: Dashboard - Favorite Villain 51](#_45mkd7jfnhq5)

[Figure 50: Dashboard - Best Visual Effects 52](#_tned12rtmai9)

[Figure 51: Dashboard - MCU Satisfaction Rate 52](#_phmnx4u1puvb)

[Figure 52: Dashboard - For Marvel Content Production and Marketing 53](#_phbxb6xyqqgc)

[Figure 53: Dashboard - Preferred marvel movie-watching platform 54](#_nudufykciv07)

[Figure 54: Dashboard - Interest in alternative marvel media experiences 54](#_ejtey5nwulgh)

[Figure 55: Dashboard Experience with Marvel-Themed attraction 55](#_37yxvtrnmsam)

[Figure 56: Dashboard Interest in marvel themed events 55](#_ht07zc2ifz44)

[Figure 57: Dashboard: Favorite aspects of marvel productions 56](#_4s5qdpboclb5)

[Figure 58: Dashboard : Most Anticipated upcoming marvel movie 56](#_enu0vhz14m8x)

[Figure 59: Dashboard - Marvel merchandise purchase history 57](#_1qexde2ylhlj)

[Figure 60: Dashboard - Likelihood of recommending marvel productions 57](#_x01s6e4qbs76)

## **11. REFERENCES**

[1] Podo, L., & Velardi, P. (2022) “[Plotly.plus, an Improved Dataset for Visualization Recommendation](https://dl.acm.org/doi/abs/10.1145/3511808.3557669)” in Proceedings of the CIKM '22: The 31st ACM International Conference on Information and Knowledge Management

[2]A. Jović and Stančin, "[An overview and comparison of free Python libraries for data mining and big data analysis](https://ieeexplore.ieee.org/abstract/document/8757088?casa_token=C4yNUYYLd_EAAAAA:U7hB0-rqyY2zpnXXQyCTOkfRtfreNyu-oDoGCRfX0zwweyOR_6atiyAV_OsupRSn4jcqtUmE)" in Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 20-24, 2019

[3]María Teresa Rodríguez,Sérgio Nunes,Tiago Devezas,”[Telling Stories with Data Visualization](https://dl.acm.org/doi/abs/10.1145/2804565.2804567?casa_token=5BSInW2jrnYAAAAA:0aMxZ4pPoeZWTdHVK6ihpStKSV4iOIZkvjvgi6rFp3xf9MZZkP_5PBd5q1PWpX8v42zCQwIlPbs)” in Proceedings of the 2015 Workshop on Narrative & Hypertext,September 2015

[4] Jamie Hoelscher,Amanda Mortimer “[Using Tableau to visualize data and drive decision-making](https://www.sciencedirect.com/science/article/pii/S0748575117301951?casa_token=YeEoX_JSvJgAAAAA:vj3LyBfq4SJsuz3cTPACOEoVRLu4rWGjWNWLuZBms6gT8Ttclbb1zuSssyUdpzajlXkxx1Mz)” in Proceedings of the [Journal of Accounting Education](https://www.sciencedirect.com/journal/journal-of-accounting-education) [Volume 44](https://www.sciencedirect.com/journal/journal-of-accounting-education/vol/44/suppl/C), September 2018, Pages 49-59.

[5]Louis T. Becker (Contributor) & Elyssa M. Gould (Column Editor) (2019) “[Microsoft Power BI: Extending Excel to Manipulate, Analyze, and Visualize Diverse Data](https://www.tandfonline.com/doi/full/10.1080/00987913.2019.1644891?casa_token=zhu7dO5tMW0AAAAA%3AVXc2YE8nzJyyDLpueT-GX28H3tGYWFHXkSw0jRM0qZlGH3CnFLiV1mBgHzeHmNlbpv_JRMaKUts)”, Serials Review, 45:3, 184-188, DOI: [10.1080/00987913.2019.1644891](https://doi.org/10.1080/00987913.2019.1644891)

[6]Jaehoon Lee, Giseop Noh and Chong-kwon Kim, "[Analysis & visualization on movie's popularity and reviews](https://ieeexplore.ieee.org/abstract/document/6741434?casa_token=GTCIc-RGLaUAAAAA:F50uxjM_Fz900q3GHtq1VZMWNjLp61PK7b3SCDF095ZI-RE9tpPLNHtZQtmhPn65os_PcjExbsw)" 2014 International Conference on Big Data and Smart Computing (BIGCOMP), Bangkok, Thailand, 2014, pp. 189-190, doi: 10.1109/BIGCOMP.2014.6741434

**OTHER SUPPORTING RESOURCES**

| **GitHub link** | [**https://github.com/sowmyakuruba20/MARVEL/**](https://github.com/sowmyakuruba20/MARVEL/) |
| --- | --- |
| **Blog** | [**https://medium.com/@nikhil.thota\_81762/data-visualization-course-project-marvel-ous-insights-on-marvel-data-f784614902b3**](https://medium.com/@nikhil.thota_81762/data-visualization-course-project-marvel-ous-insights-on-marvel-data-f784614902b3) |
| **Youtube** | [**https://www.youtube.com/watch?v=qyRh80vkI18**](https://www.youtube.com/watch?v=qyRh80vkI18) |