# The Simpsons Data Analysis


| | |
|---|---|
| AKSHAYA MOTAMARRI | 20BCE1371 |
| LENIN VASAN | 20BCE1892 |
| G NITHISH KANNA | 20BCE1304 |
| SANJAY R R | 20BCE1537 |

# 1. Introduction

## *1.1 Existing Problem*

The Simpsons show is an American animated sitcom created by Matt Groening. The series is a satirical depiction of American life, epitomised by the Simpson family, which consists of Homer, Marge, Bart, Lisa, and Maggie. Since its debut on December 17, 1989, 750 episodes of the show have been broadcast as of May 21 2023. It is the longest-running American animated series, the longest-running American sitcom, and the longest-running American scripted primetime television series, both in terms of seasons and number of episodes.

## *1.2 Purpose*

The achievements accomplished by this show are the dreams of many other show creators and actors. Also for the Simpsons creators need to know how the show is performing and how the viewers are responding to the show. This will help them to even improve the show and reach even more heights. This project wishes to help these requirements, by analysing the data generated by the show through visualisations aids in understanding the reach of the show over the years, and the response it got from the audience and critics. The outcomes of this analysis wish to help aspiring actors and content creators gain knowledge on the characteristics of the audience and accomplish their goals.

# 2. Literature Survey

## *2.1 Existing Problem:*

Existing research on "The Simpsons" has explored character analysis, episode and season analysis, and location analysis to gain insights into the show's dynamics.

**2.1.1 Character Analysis:** Thompson and Rodriguez (2017) visualized character networks to analyze relationships and character evolution. Our solution expands on this by incorporating location analysis and exploring character distribution across different show locations.

**2.1.2 Episode and Season Analysis:**

Nguyen and Chen (2019) focused on episode counts per season. Our solution adds location analysis within episodes, visualizing the occurrence of different settings to understand spatial dynamics.

#### 2.1.3 Location Analysis:

Location analysis is relatively unexplored in "The Simpsons" research. Our approach draws from studies on other shows, highlighting the significance of locations in storytelling and character development.

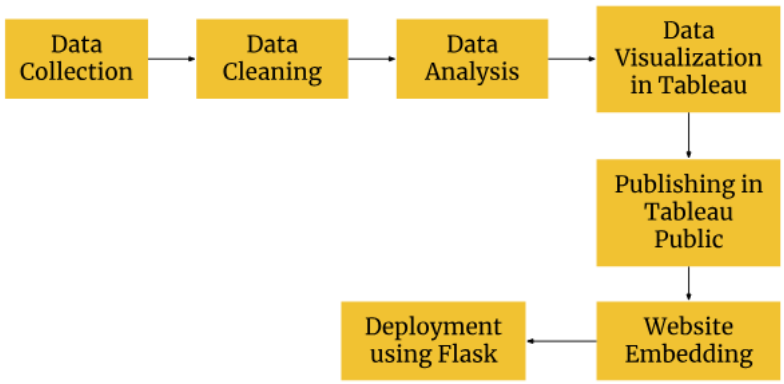
### *2.2 Proposed Solution:*

Our proposed solution combines data analytics and visualization techniques to provide engaging insights into "The Simpsons." We analyze character distribution, and episode counts, and incorporate location analysis to uncover spatial patterns and enhance the understanding of the show's dynamics.

# 3. Theoretical Analysis

### *3.1 Block Diagram*

The block diagram provides a step-by-step overview of the data analysis and visualization process for the project based on "The Simpsons" show. The key stages of the process are as follows:




**1. Data Collection:** Gather relevant data about "The Simpsons" show, including characters, episodes, locations, and other attributes.

**2. Data Cleaning:** Clean the collected data by handling missing values, removing duplicates, and transforming it for analysis.

**3. Data Analysis:** Apply analytical techniques like statistical analysis, network analysis, and text mining to uncover patterns and relationships.

**4. Data Visualization:** Create visual representations such as charts, graphs, and interactive dashboards in Tableau to effectively communicate the insights.

**5. Publishing:** Publish the finalized visualizations on Tableau Server or Tableau Public for easy sharing and access.

**6. Embedding:** Embed the visualizations into a website using HTML and JavaScript for a seamless user experience.

**7. Flask Deployment:** Deploy the website with embedded visualisations using Flask, a Python web framework, for hosting and integration.

## *3.2 Hardware / Software Designing*

The project requires the following hardware and software components:

### Hardware:

- Computer system with an Intel Core i5 processor or higher
- Minimum of 8GB RAM
- Adequate storage capacity for the dataset and software applications

### Software:

- Microsoft Excel: Used for data cleaning, preprocessing, and initial analysis tasks.
- MySQL: Serves as the relational database management system (RDBMS) for storing the cleaned and processed data.
- Tableau: Utilized for advanced data visualization and dashboard creation.
- Tableau Server or Tableau Public: Platforms for publishing and sharing visualizations online.
- HTML and JavaScript: Used for embedding the visualizations into a website.

- Flask: Python web framework for website deployment and integration of embedded visualizations.

By leveraging these hardware and software components, the project enables effective data analysis, visualization, and deployment of the visualizations on a website. The visual representations provide insights into various aspects of "The Simpsons" show, making the findings more engaging and accessible to the audience.

# 4. Experimental Investigations

During the course of working on the data analysis and visualization project based on "The Simpsons" show, several investigations were conducted to gain deeper insights into the dataset and validate the effectiveness of the solution. The following are some of the key analyses and investigations that were performed:

**1. Character Analysis:** A comprehensive analysis of the characters in "The Simpsons" was conducted to understand their characteristics, relationships, and roles within the show. This involved exploring the frequency of appearances, analyzing character interactions using network analysis, and identifying the most prominent and central characters in the series.

**2. Episode Analysis:** An investigation into the episodes of "The Simpsons" was carried out to uncover patterns and trends. This included analyzing episode durations, exploring the distribution of genres across episodes, and examining the popularity of different episodes based on IMDb ratings or viewer ratings.

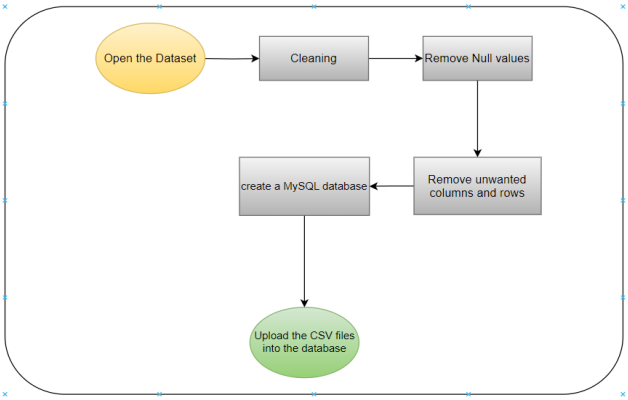
**3. Season Analysis:** A detailed examination of the seasons of "The Simpsons" was performed to identify any patterns or changes over time. This involved analyzing the number of episodes per season, tracking changes in IMDb ratings over seasons, and investigating the presence of recurring themes or storylines throughout the series.

**4. Location Analysis:** The dataset was analyzed to gain insights into the different locations featured in "The Simpsons." This investigation involved identifying the most frequently visited locations, examining location-based trends or patterns across episodes or seasons, and visualizing the geographical distribution of locations within the show.

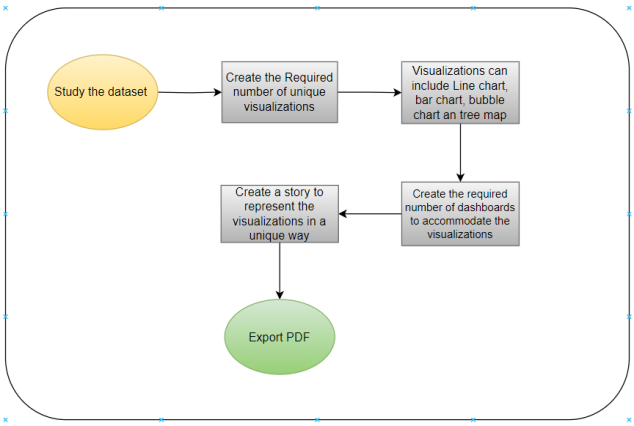
**5. Talkativeness Analysis:** An investigation into the talkativeness of characters was carried out to understand the distribution of dialogue among different characters. This involved analyzing the number of lines spoken by each character, identifying the most talkative and least talkative characters, and exploring any correlations between talkativeness and character popularity or screen time.These experimental investigations provided valuable insights into various aspects of "The Simpsons" show. By analyzing characters, episodes, seasons, locations, and talkativeness, a deeper understanding of the show's dynamics, narratives, and popularity was achieved. The findings from these investigations were incorporated into the data visualizations and dashboards, enhancing the overall storytelling and presentation of the project.

# 5. Flowchart

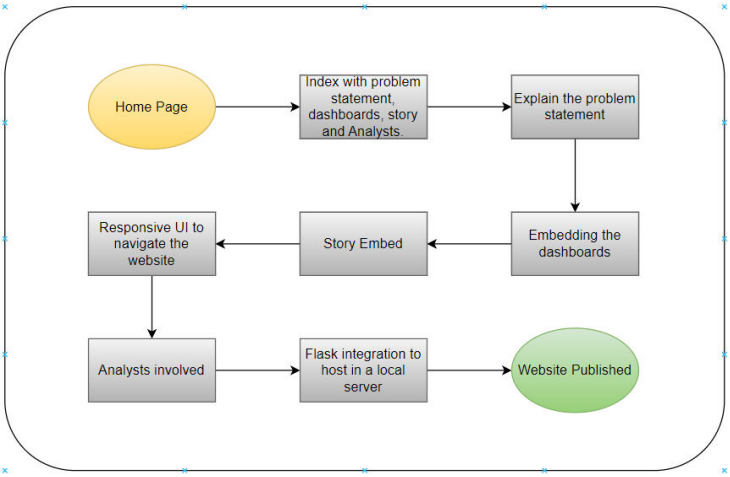
Dataset Cleaning and Deployment




Data Visualisations and Dashboard Creation




Web integration




# 7. Advantages and disadvantages

## 7.1. Advantages

- ***Tableau Integration:*** Tableau seamlessly integrates multiple data sources for a comprehensive analysis of "The Simpsons" show.
- ***Real-time Analysis:*** Perform real-time analysis by connecting to live data sources or scheduling automatic refreshes.
- ***Storytelling and Presentations****:* Use Tableau's storyboard tool to create compelling narratives with annotated visualizations for effective presentation and sharing of insights.
- ***Data-driven Decision Making:*** Gain valuable insights for better decision making, such as identifying popular characters, understanding viewer preferences, and analyzing factors influencing episode ratings.
- ***User-friendly Interface:*** Tableau's intuitive interface allows for easy drag-and-drop functionality, filters, and customization options without extensive coding skills.

## 7.2. Disadvantages

- Price: Tableau is a commercial program with varying prices based on the selected version, licensing fees, maintenance costs, and training expenses.
- Dependency on Data Quality: The quality of the underlying data impacts the accuracy and reliability of Tableau's analyses and visualizations.
- Data Processing Limitations: Tableau's data processing capabilities may be limited compared to specialized statistical software or programming languages, requiring additional tools or software integration for advanced computations or statistical analysis

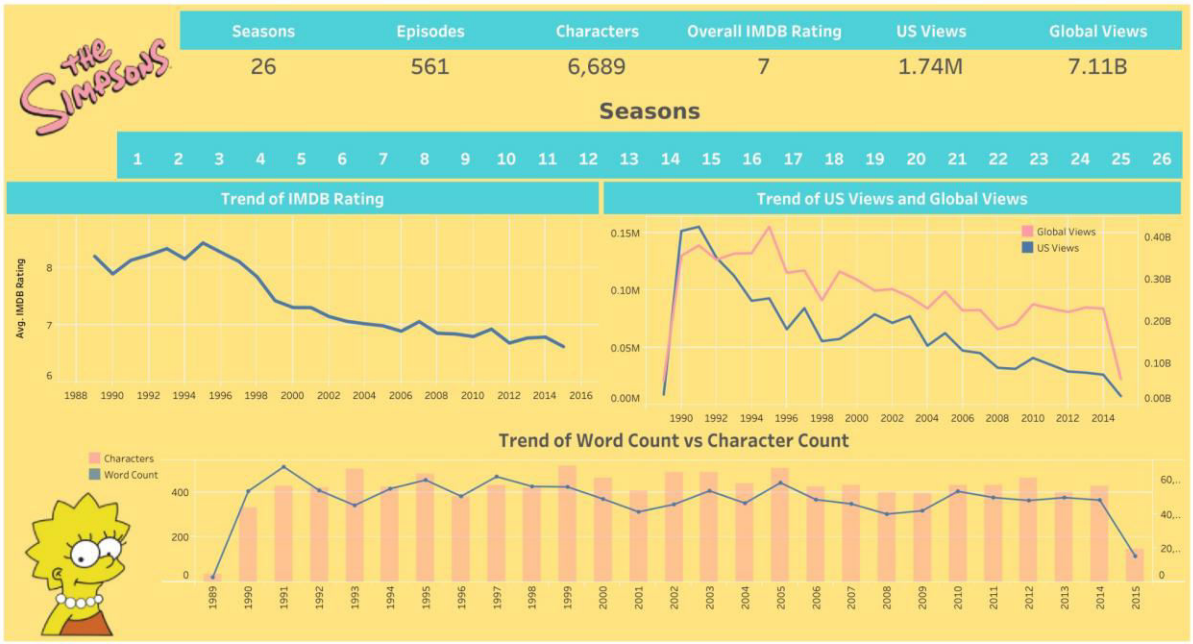
# 8. Applications

- ***Viewership Analysis:*** Understand the popularity and engagement of "The Simpsons" through ratings, viewing patterns, and demographic data.
- ***Character Analysis:*** Analyze character prominence, appeal, and development using variables like screen time and fan sentiment.
- ***Episode Analysis:*** Evaluate episode success, identify viewer involvement factors, and spot trends in themes and storytelling techniques.
- ***Location Analysis:*** Gain insights into the relationship between locations, characters, and episodes in "The Simpsons."

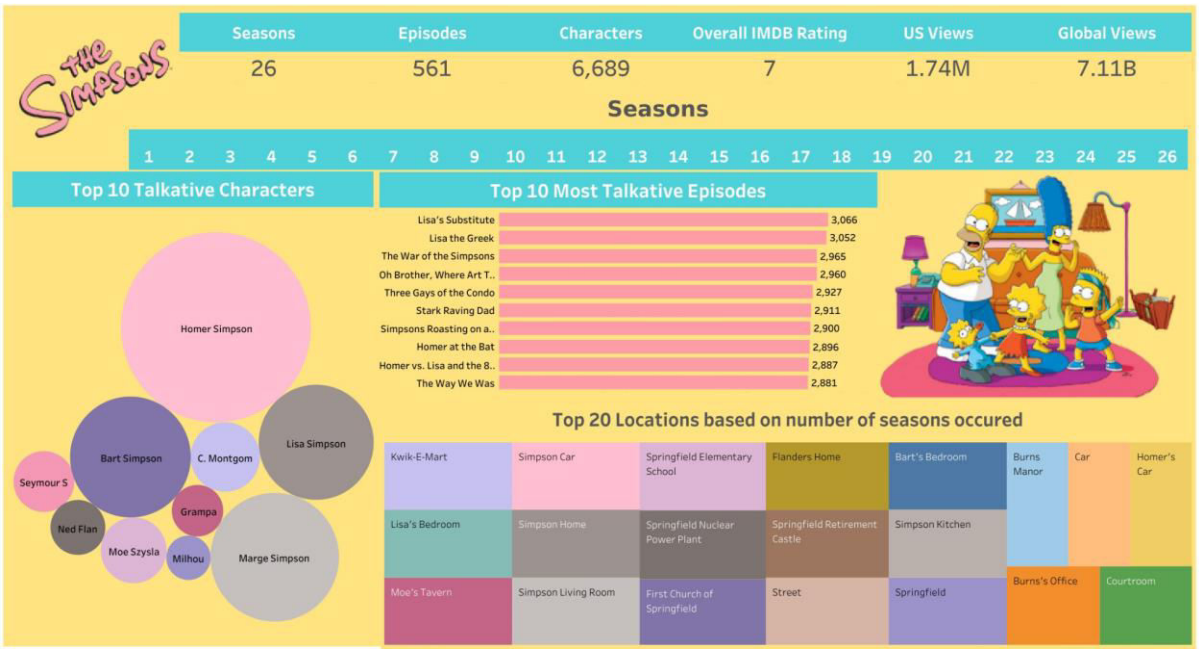
# 9. Results

## 9.1 Dashboards

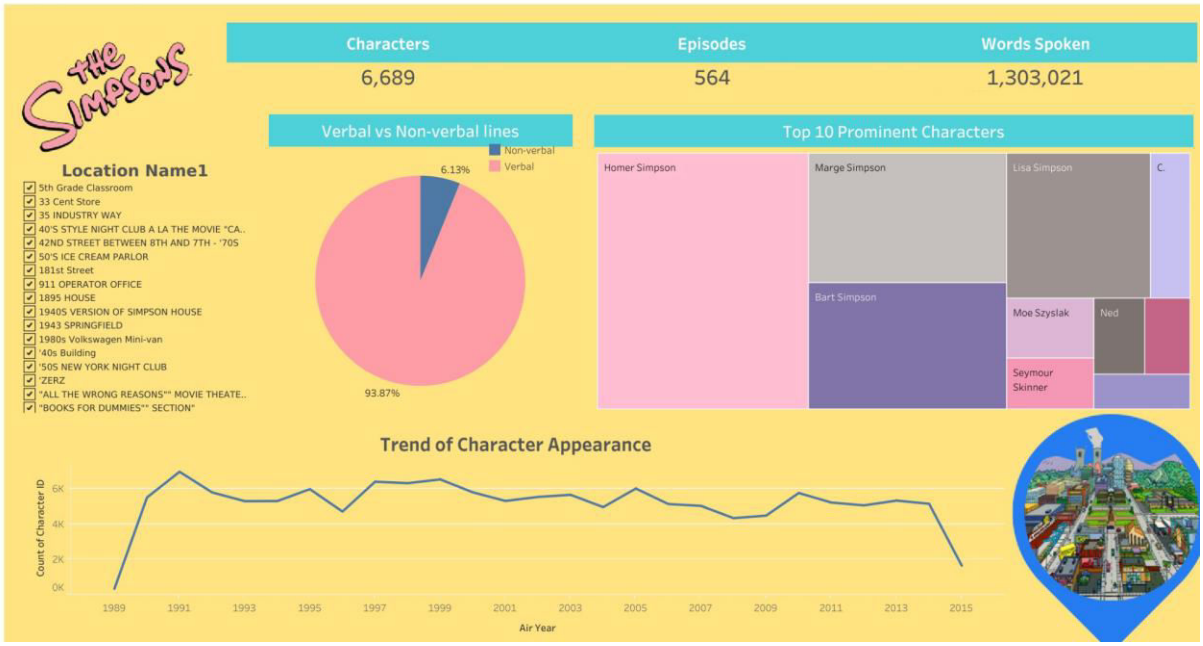
### *Overall show Analysis*



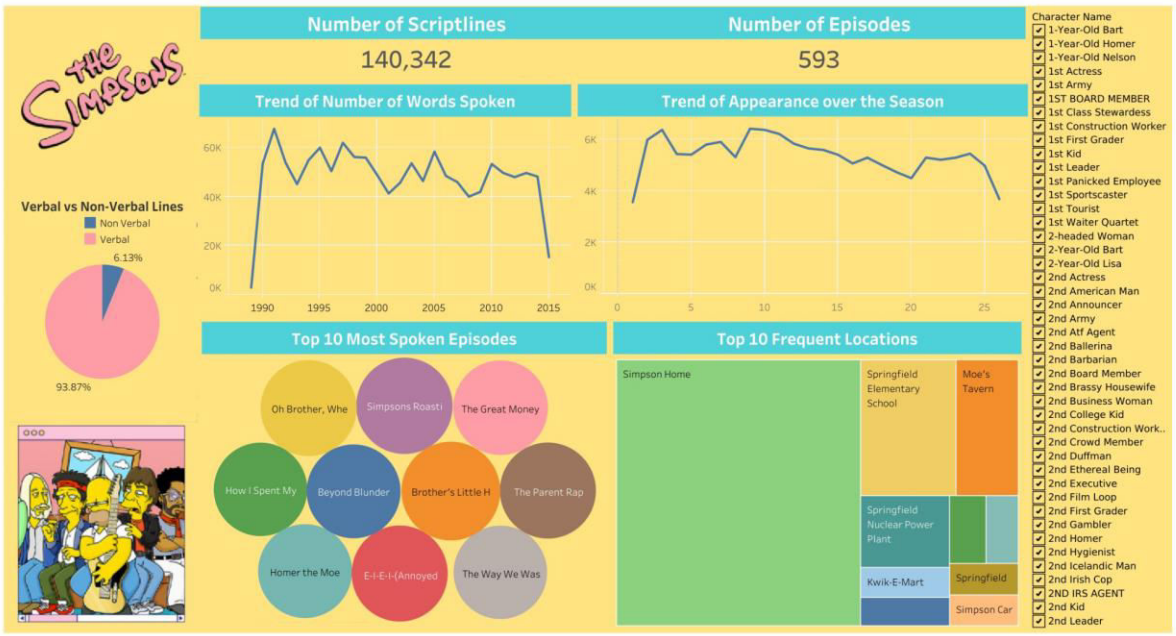

### *Show's top analysis*



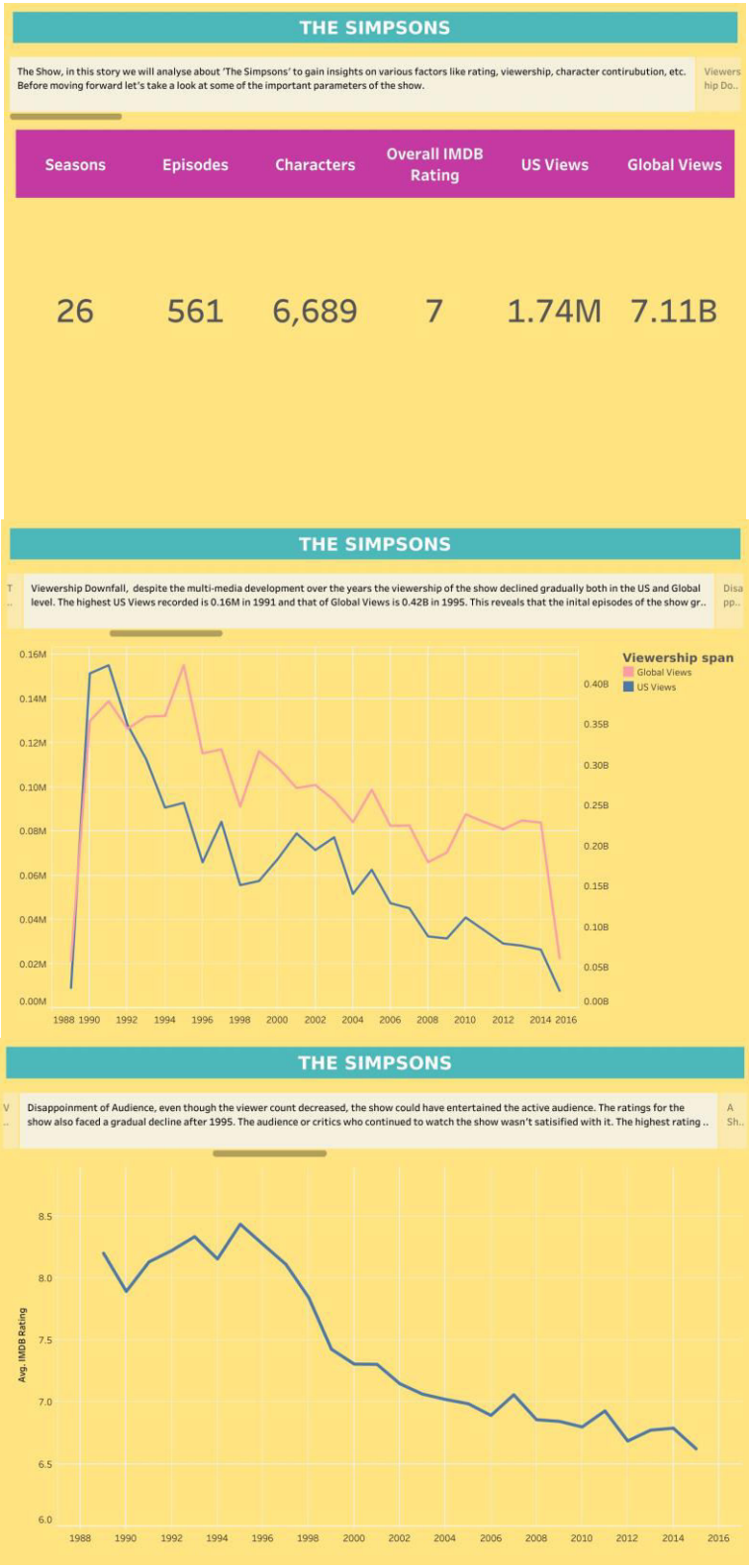
## *Location Analysis*

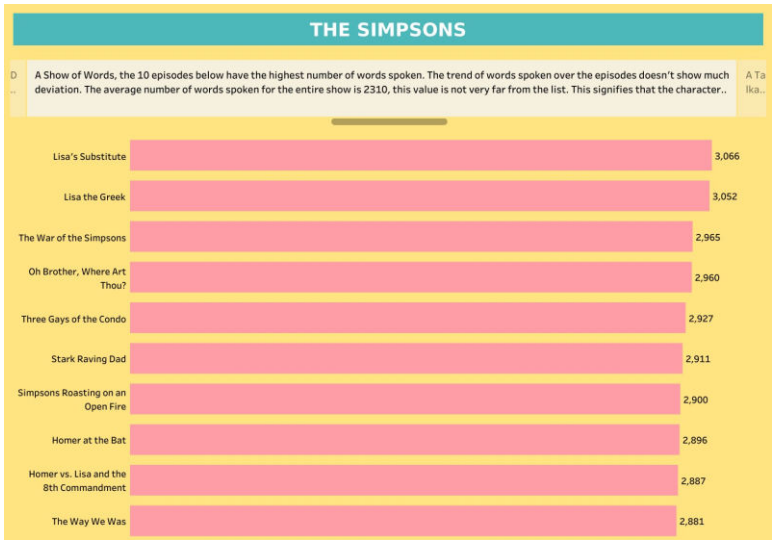


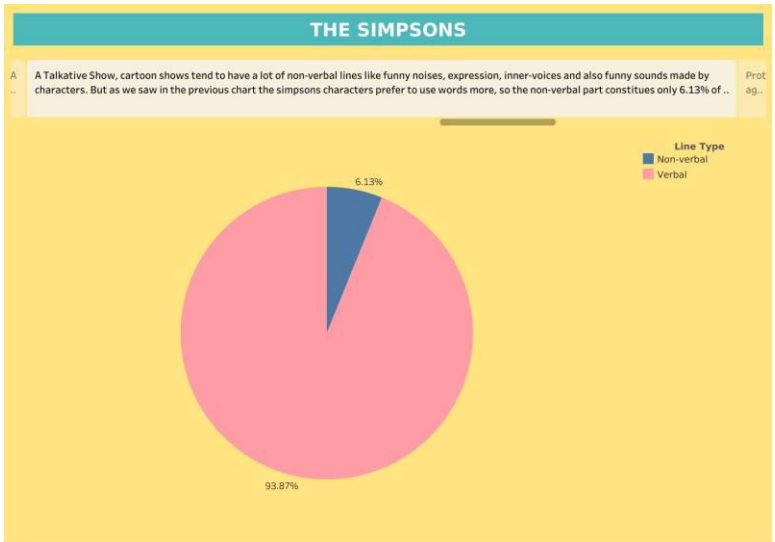

## *Character Analysis*

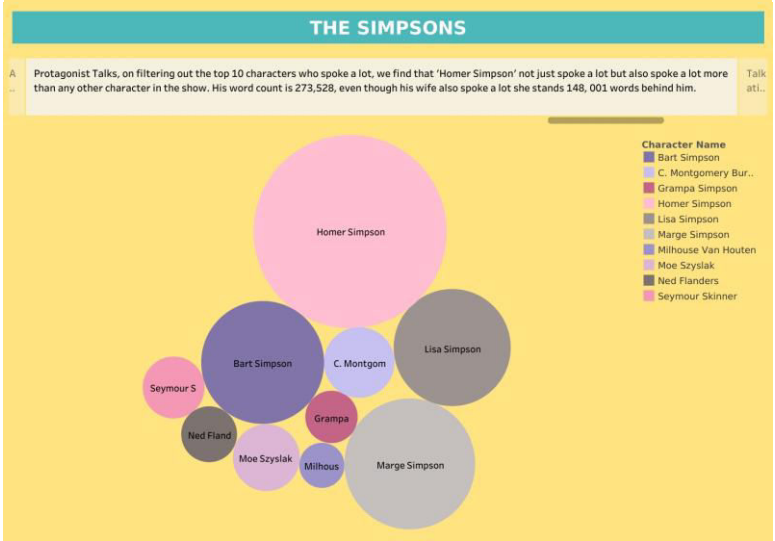


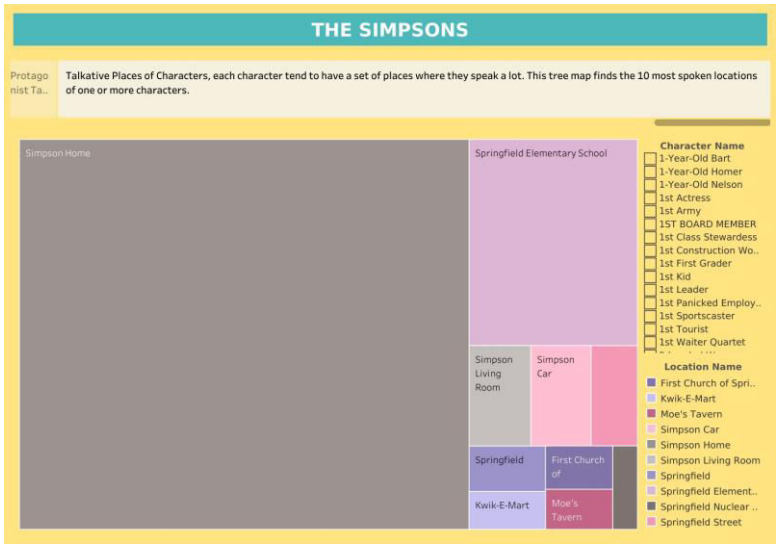
## 9.2 Story


THE SIMPSONS
The Show, in this story we will analyse about 'The Simpsons' to gain insights on various factors like rating, viewership, character contirubution, etc. Before moving forward let's take a look at some of the important parameters of the show.
Seasons
Episodes
Characters
Overall IMDB Rating
US Views
Global Views
26
561
6,689
7
1.74M
7.11B
THE SIMPSONS
Viewership Downfall, despite the multi-media development over the years the viewership of the show declined gradually both in the US and Global level. The highest US Views recorded is 0.16M in 1991 and that of Global Views is 0.42B in 1995. This reveals that the inital episodes of the show gr..
Viewership span
Global Views
US Views
0.16M
0.14M
0.12M
0.10M
0.08M
0.06M
0.04M
0.02M
0.00M
0.40B
0.35B
0.30B
0.25B
0.20B
0.15B
0.10B
0.05B
0.00B
1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016
THE SIMPSONS
Disappointment of Audience, even though the viewer count decreased, the show could have entertained the active audience. The ratings for the show also faced a gradual decline after 1995. The audience or critics who continued to watch the show wasn't satisified with it. The highest rating ..
Avg. IMDB Rating
8.5
8.0
7.5
7.0
6.5
6.0
1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016


THE SIMPSONS
A Show of Words, the 10 episodes below have the highest number of words spoken. The trend of words spoken over the episodes doesn't show much deviation. The average number of words spoken for the entire show is 2310, this value is not very far from the list. This signifies that the character..
Lisa's Substitute 3,066
Lisa the Greek 3,052
The War of the Simpsons 2,965
Oh Brother, Where Art Thou? 2,960
Three Gays of the Condo 2,927
Stark Raving Dad 2,911
Simpsons Roasting on an Open Fire 2,900
Homer at the Bat 2,896
Homer vs. Lisa and the 8th Commandment 2,887
The Way We Was 2,881


THE SIMPSONS
A Talkative Show, cartoon shows tend to have a lot of non-verbal lines like funny noises, expression, inner-voices and also funny sounds made by characters. But as we saw in the previous chart the simpsons characters prefer to use words more, so the non-verbal part constitues only 6.13% of ..
Line Type
Non-verbal
Verbal
6.13%
93.87%


THE SIMPSONS
Protagonist Talks, on filtering out the top 10 characters who spoke a lot, we find that 'Homer Simpson' not just spoke a lot but also spoke a lot more than any other character in the show. His word count is 273,528, even though his wife also spoke a lot she stands 148, 001 words behind him.
Character Name
Bart Simpson
C. Montgomery Bur..
Grampa Simpson
Homer Simpson
Lisa Simpson
Marge Simpson
Milhouse Van Houten
Moe Szyslak
Ned Flanders
Seymour Skinner
Homer Simpson
Bart Simpson
C. Montgom
Lisa Simpson
Seymour S
Grampa
Ned Fland
Moe Szyslak
Milhous
Marge Simpson




# 10. Conclusion

The analysis of "The Simpsons" show reveals several interesting observations. Over the years, both the viewership and ratings of the show have gradually declined, indicating a concerning trend on both a national and global level. This decline in popularity is surprising considering the advancements in the multimedia industry during the same period. Another parameter of observation is the word count and character count. It was observed that as the word count increases, the number of characters appearing in a year also increases, suggesting a fair distribution of script lines among characters in each episode. A horizontal bar chart of the top 10 talkative characters shows that there is not a significant difference in the number of words spoken between episodes. A dedicated pie chart reveals that verbal lines constitute 93.87% of the total lines, highlighting the show's talkative nature, which caters to its adult audience's preference for content-rich dialogue. The analysis of different locations and characters provides insights into their contributions, communication styles, popular episodes and locations, and trends in their appearances and word count over the years.

# 11. Future Scope

The Simpsons show is still on air and is expected to grow over the coming years. In such cases the analysis model we proposed can be used with the future data. This analysis model is created in such a way that it doesn’t depend on the time or other external factors. The model relies solely on the response given by the audience as rating and on the various show

attributes, which allows the model to be used even in the coming years. And also the dataset used for this project was limited to the year 2015, hence further improvements in data collection will allow this model to be applied to the current as well future.

# 12. Bibliography

**Data Source :** https://www.kaggle.com/datasets/prashant111/

# Appendix

Source code for Web Interface Implementation :

https://drive.google.com/drive/folders/1nzqSIWLV4f0rFa3qW5RinHvNMANc7Sne?usp=sharing

Demonstration videos of Project Steps :

https://drive.google.com/drive/folders/1K7B1aTT6lBiQ9tKedRU3LAp1piNhiERN?usp=sharing

Overall Project Demo video : :

https://drive.google.com/file/d/1unSKd2zhFQlWGvATTAKQKuW3T6aQQlry/view?usp=sharing