

The Simpsons Data Analytics

Team no. 232

Team Members:

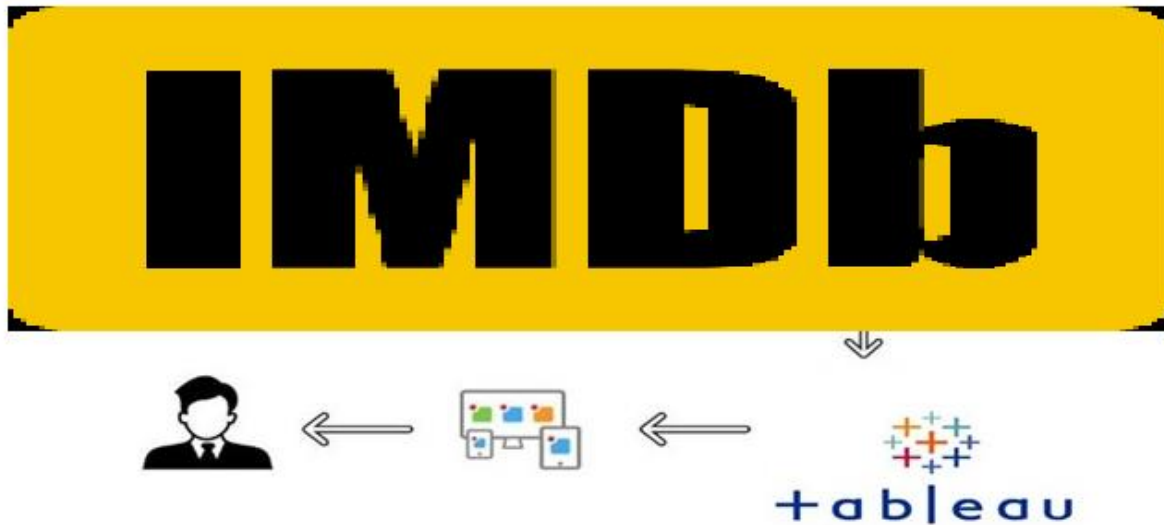
Name	Registration Number	Campus
Rishabh Agarwal	20BCE0802	VIT VELLORE
Nachiket Dandegaonkar	20BCE10465	VIT BHOPAL
Ameya Dhanorkar	20BCE10406	VIT BHOPAL
Rajveer Heera	20BCE0921	VIT VELLORE

About our Project:

The business problem at hand is the successful implementation of a data visualization project centered around "The Simpsons Data Analytics." The challenge lies in effectively harnessing and presenting the vast amount of data available from the iconic TV series, "The Simpsons," in a visually engaging and insightful manner. The project requires mining and analyzing various aspects of the show, such as character interactions, episode themes, and cultural references, to uncover meaningful patterns and trends. However, the complexity arises from transforming this raw data into intuitive and comprehensible visualizations that cater to a diverse audience, including both avid fans and data enthusiasts.

The team must tackle issues like data cleaning, ensuring data accuracy, and deciding on the appropriate visualization techniques to encapsulate the richness of "The Simpsons" universe. Additionally, addressing the challenge of data scalability and creating an interactive platform for users to explore and interact with the visualizations further adds to the complexity of the project. Ultimately, the aim is to deliver a data visualization solution that enhances understanding and appreciation of "The Simpsons" through data-driven insights.

Technical Architecture:



Project Flow

To accomplish this, we have to complete all the activities listed below,

Define Problem / Problem Understanding

- Specify the business problem
- Business requirements
- Literature Survey
- Social or Business Impact.

Data Collection & Extraction from Database

- Collect the dataset,
- Connect IBM DB2 with IBM cognos

Data Preparation

- Prepare the Data for Visualization

Data Visualizations

- No of Unique Visualizations

Dashboard

- Responsive and Design of Dashboard

Story

- No of Scenes of Story

Report

- Creating a report

Performance Testing

- Amount of Data Rendered to DB ‘
- Utilization of Data Filters
- No of Calculation Fields
- No of Visualizations/ Graphs

Web Integration

- Dashboard and Story embed with UI With Flask

Project Demonstration & Documentation

- Record explanation Video for project end to end solution
- Project Documentation-Step by step project development procedure

The Business Problem

The business problem at hand is the successful implementation of a data visualization project centered around "The Simpsons Data Analytics." The challenge lies in effectively harnessing and presenting the vast amount of data available from the iconic TV series, "The Simpsons," in a visually engaging and insightful manner. The project requires mining and analyzing various aspects of the show, such as character interactions, episode themes, and cultural references, to uncover meaningful patterns and trends. However, the complexity arises from transforming this raw data into intuitive and comprehensible visualizations that cater to a diverse audience, including both avid fans and data enthusiasts. The team must tackle issues like data cleaning, ensuring data accuracy, and deciding on the appropriate visualization techniques to encapsulate the richness of "The Simpsons" universe. Additionally, addressing the challenge of data scalability and creating an interactive platform for users to explore and interact with the visualizations further

adds to the complexity of the project. Ultimately, the aim is to deliver a data visualization solution that enhances understanding and appreciation of "The Simpsons" through data-driven insights.

Business Requirements

Business Requirements for "The Simpsons Data Analytics" Project:

1. Data Collection:

- Gather comprehensive information about episodes, including titles, air dates, and plot summaries.
- Collect detailed character data, including names, relationships, and personalities.
- Capture cultural references, parodies, and celebrity guest appearances.
- Monitor social media platforms and online forums for fan discussions and reactions.
- Ensure data quality, accuracy, consistency, and relevancy.
- Comply with data governance and privacy regulations.
- Maintain clear documentation of data sources and transformations.

2. Data Cleaning and Preparation:

- Identify and handle missing, inconsistent, or erroneous data.
- Standardize data formats, naming conventions, and units of measurement.
- Remove duplicate records and resolve any data conflicts.
- Perform data integration from multiple sources.
- Transform and normalize data for analysis.
- Apply appropriate data anonymization techniques, if required.

3. Data Analysis:

- Apply statistical and analytical methods to identify patterns and trends.
- Explore relationships between characters, episodes, and cultural references.
- Conduct sentiment analysis on fan discussions and reactions.
- Perform clustering or classification techniques for character profiling.

- Utilize data visualization to present insights effectively.

4. Report Creation:

- Generate comprehensive reports on various aspects of "The Simpsons" data.
- Summarize key findings and insights from the data analysis.
- Include visualizations, charts, and graphs to enhance understanding.
- Tailor reports for different stakeholders, considering their specific requirements.
- Ensure the reports are accessible, user-friendly, and visually engaging.
- Provide recommendations based on the data analysis for future strategies or improvements.

These business requirements cover the critical aspects of the project, from data collection to report creation. By fulfilling these requirements, the project team can deliver a robust data analytics solution that uncovers valuable insights from "The Simpsons" data and facilitates informed decision-making.

Literature Survey

This literature survey aims to provide an in-depth exploration of the existing research and literature pertaining to data analytics applied to The Simpsons television series. As one of the longest-running animated sitcoms, The Simpsons has not only captured a massive following but has also had a significant cultural impact. Analyzing the extensive amount of data generated by the show can offer valuable insights into diverse aspects, including character development, humor, social commentary, and audience reception.

The survey begins by offering an overview of The Simpsons and its cultural significance, establishing the foundation for understanding the relevance of data analytics in the context of this iconic show. It then delves into the various methodologies employed in conducting research related to The Simpsons data analytics, outlining the research methods, sources, databases, and inclusion and exclusion criteria utilized in the survey.

Social Or Business Impact

Social Impact:

1. **Cultural Understanding:** The project can provide a deeper understanding of the cultural references, parodies, and themes present in "The Simpsons." It allows fans and researchers to explore the show's social commentary, its portrayal of various cultural phenomena, and its influence on popular culture.
2. **Fan Engagement and Appreciation:** The data analytics project can enhance fan engagement by offering new perspectives and insights into their favorite show. It can

spark discussions, encourage fan theories, and foster a sense of community among fans. This deeper appreciation can contribute to the show's enduring popularity and fan loyalty.

3. **Academic Research:** The project's findings can serve as a valuable resource for academic research on topics like television studies, cultural studies, media analysis, and data analytics. Researchers can leverage the project's methodologies and insights to explore broader themes within the field.

Business Impact:

1. **Content Creation and Strategy:** The project's findings can inform content creators and showrunners in developing new episodes, storylines, and character arcs. By understanding fan preferences, sentiment, and engagement levels, the project can guide decision-making in content creation, potentially leading to increased viewership and audience satisfaction.
2. **Marketing and Branding:** The insights from the project can be utilized in marketing campaigns and brand strategies related to "The Simpsons." By understanding the show's cultural impact and audience sentiment, marketers can tailor messaging, promotions, and merchandise to resonate with fans effectively.
3. **Data-Driven Decision-Making:** The project's data analytics can empower decision-makers within the entertainment industry. Studios and production companies can leverage the insights to evaluate the success of different episodes, seasons, or characters, making informed decisions regarding future investments, licensing agreements, and syndication deals.
4. **Fan-driven Experiences:** The project can inspire the development of interactive platforms, mobile applications, or websites that allow fans to explore and engage with "The Simpsons" data. These fan-driven experiences can create additional revenue streams and enhance the overall fan experience.
5. **Industry Innovation:** The project's methodologies and techniques can stimulate innovation within the entertainment industry. It can inspire other TV shows, studios, and content creators to adopt data analytics and visualization approaches to gain insights into their own productions, leading to more data-informed and successful content strategies.

In summary, the "The Simpsons Data Analytics" project has the potential to generate both social and business impacts. It enhances cultural understanding, fosters fan engagement, facilitates academic research, informs content creation and marketing strategies, enables data-driven decision-making, and drives industry innovation. By leveraging data analytics in the context of this beloved TV series, the project can contribute to a richer fan experience and provide valuable insights for the entertainment industry as a whole.

Data Collection & Extraction From Database

Data collection is the process of gathering and measuring information on variables of interest in an established, systematic fashion that enables one to answer stated research questions, test hypotheses, evaluate outcomes, and generate insights from the data.

Collect The Dataset

Activity 1.1: Understand the data

Check out the below link to understand the dataset in detail:

https://www.kaggle.com/datasets/prashant111/the-simpsons-dataset?select=simpsons_episodes.csv

Data Preparation

Data preparation for Tableau involves the process of organizing, cleaning, and transforming raw data into a format that can be effectively visualized and analyzed within the Tableau software. This includes tasks such as data cleaning, data integration, data formatting, and data aggregation. The goal is to ensure that the data is accurate, consistent, and structured in a way that enables meaningful insights and visualizations in Tableau.

Prepare The Data For Visualization

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency.

Data Visualization

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex datasets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

No Of Unique Visualizations (Filters Applied)

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the data include bar charts, line charts, heat maps, scatter plots, pie charts, maps, etc. These visualizations can be used to compare performance, track changes over time, show distribution, and show relationships between variables.

Activity 1.1: Treemap of Seasons vs IMDB Rating This treemap visualization presents a comparative view of different seasons of "The Simpsons" based on their IMDB ratings. It allows users to identify trends and variations in season ratings at a glance.

Activity 1.2: Treemap of Seasons vs US Viewership in Millions This treemap visualization showcases the viewership data for each season of "The Simpsons" in the United States, providing a visual representation of the relative popularity of different seasons.

Activity 1.3: Horizontal Bar Chart of Seasons vs IMDB Rating The horizontal bar chart visualizes the IMDB ratings of each season of "The Simpsons" in a straightforward manner, enabling easy comparison and identification of the highest and lowest rated seasons.

Activity 1.4: Horizontal Bar Chart of Seasons vs US Viewership in Millions Similar to Activity 1.3, this horizontal bar chart focuses on the viewership data, allowing users to quickly identify the most and least-watched seasons in the United States.

Activity 1.5: Amount of Dialogues by Characters Throughout Different Ages/Versions This visualization illustrates the distribution of character dialogues across different ages or versions within "The Simpsons." It provides insights into the evolution and prominence of characters throughout the series.

Activity 1.6: Popularity of Locations by Number of Dialogues Hosting/Referencing the Location This visualization presents the popularity of different locations within "The Simpsons" based on the number of dialogues that reference or take place in those locations. It highlights the significance of various settings in the show.

Activity 1.7: Gender Distribution of Characters This visualization depicts the gender distribution of characters within "The Simpsons." It provides an overview of the representation of male and female characters throughout the series.

Activity 1.8: Episodes vs Number of Dialogues This visualization showcases the relationship between the number of dialogues and the number of episodes in "The Simpsons," providing insights into dialogue intensity and episode length.

Activity 1.9: Most Lines Spoken by Characters by Location This visualization identifies the characters who have the most lines of dialogue in specific locations within "The Simpsons." It allows users to explore character prominence in different settings.

Explanation video link :

<https://drive.google.com/drive/folders/1Dv50gSDEwCjld2hpQFyFKe8l7G2wILOE?usp=sharing>

Dashboard

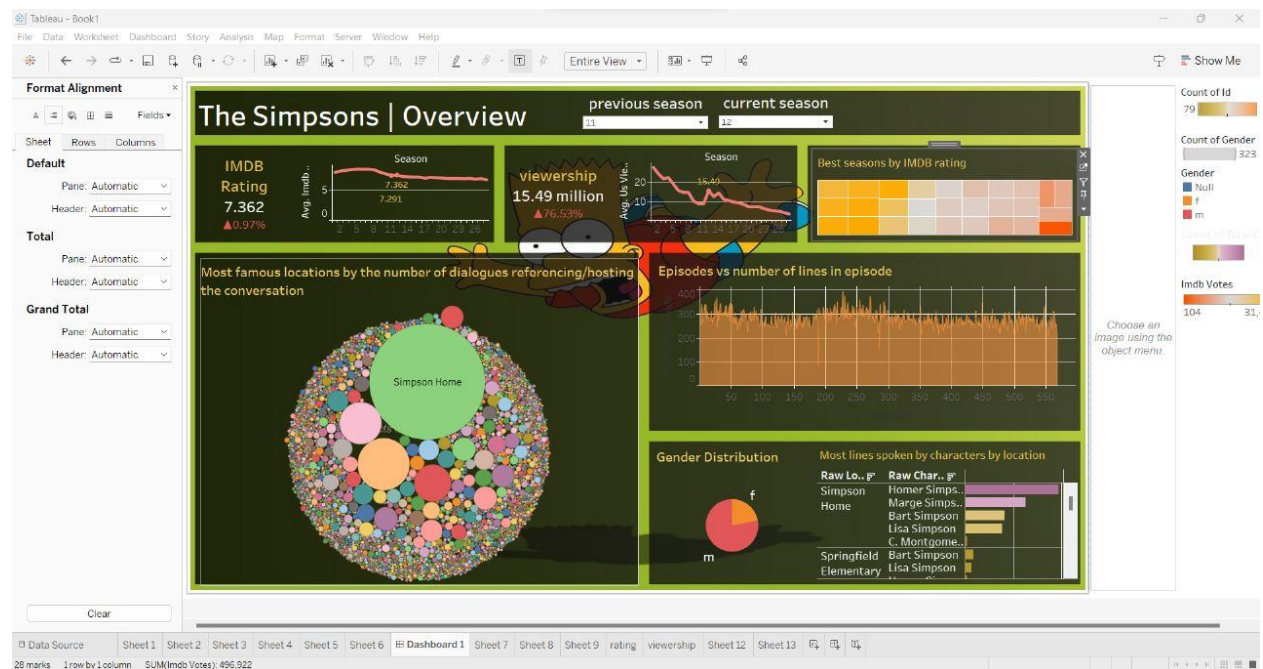
A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case.

Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Responsiveness And Design Of Dashboard

The dashboard for the "The Simpsons Data Analytics" project incorporates nine visualizations that offer valuable insights into various aspects of the TV series. These visualizations provide a comprehensive overview of different factors such as season ratings, viewership, character dialogues, and more.

Explanation video link: <https://drive.google.com/file/d/1TuiffZz9E3eW-Fm5OUGq4BRpiXMXbF8X/view?usp=sharing>



Story

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that

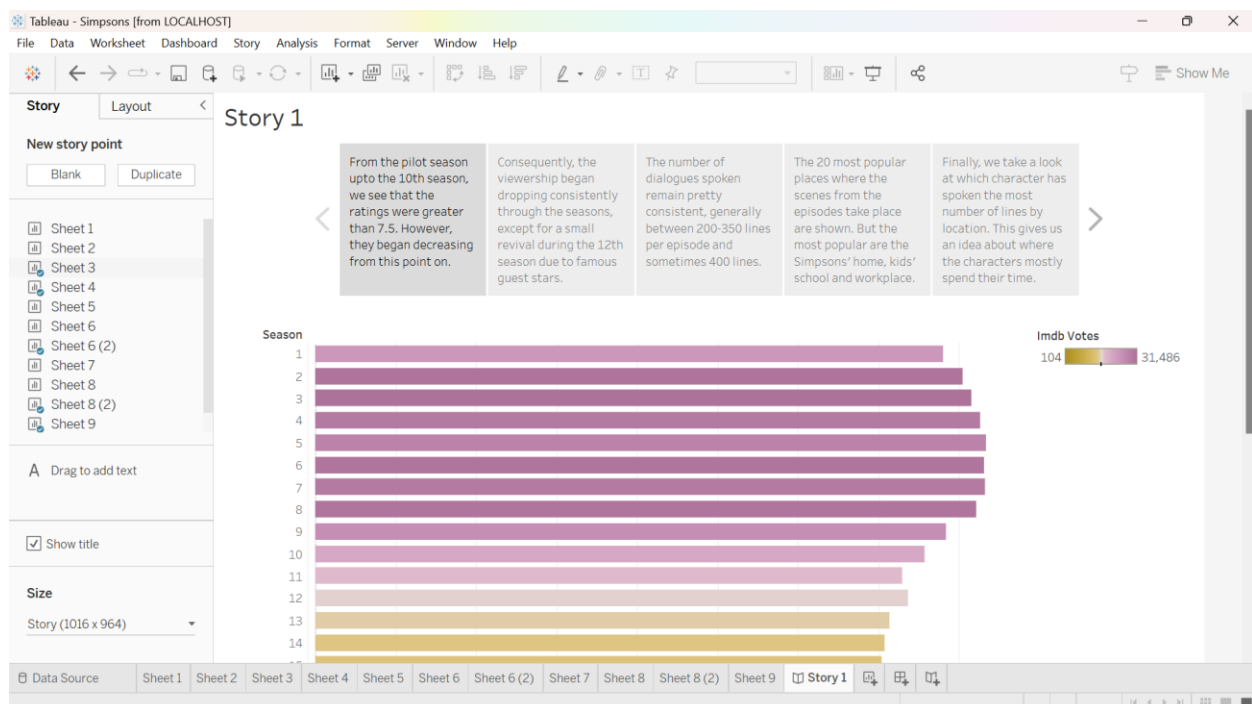
summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

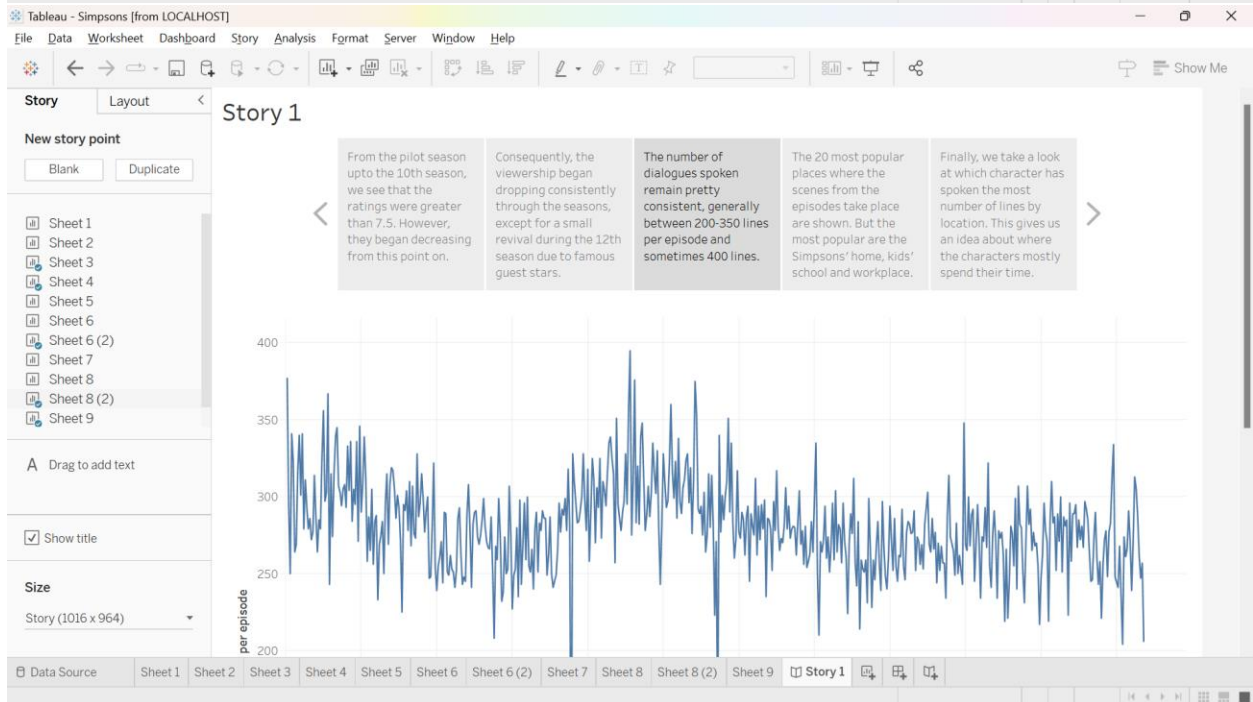
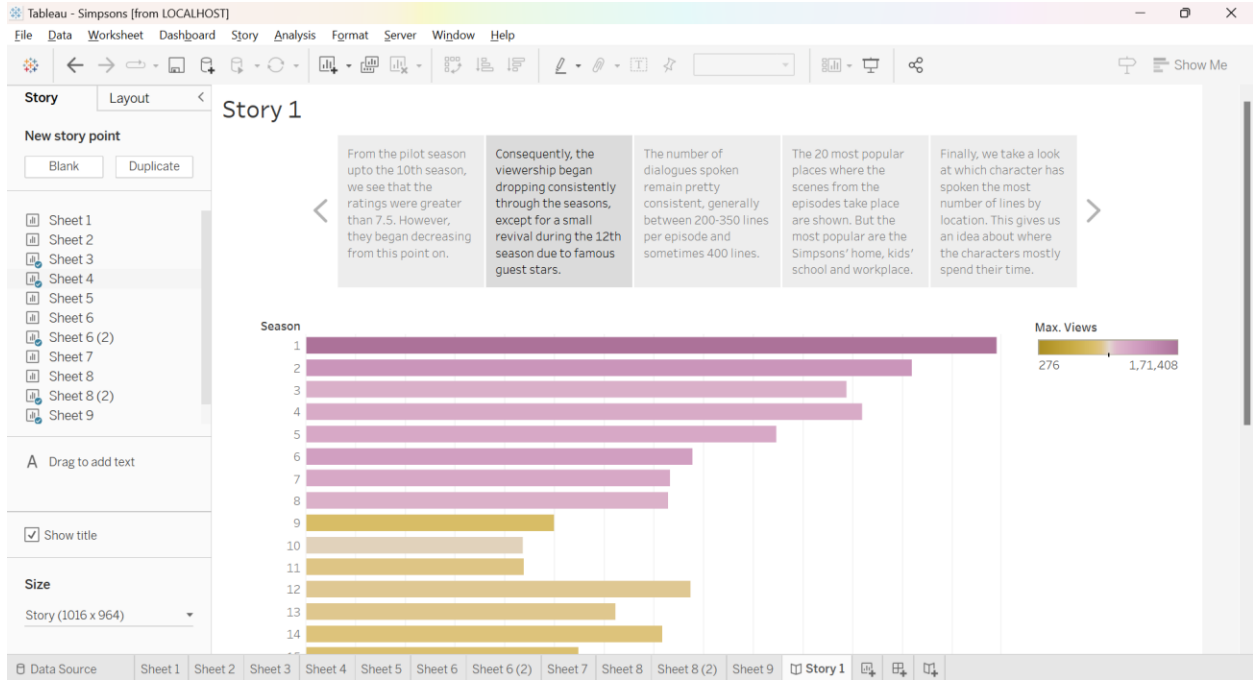
No. of Scenes Of Story

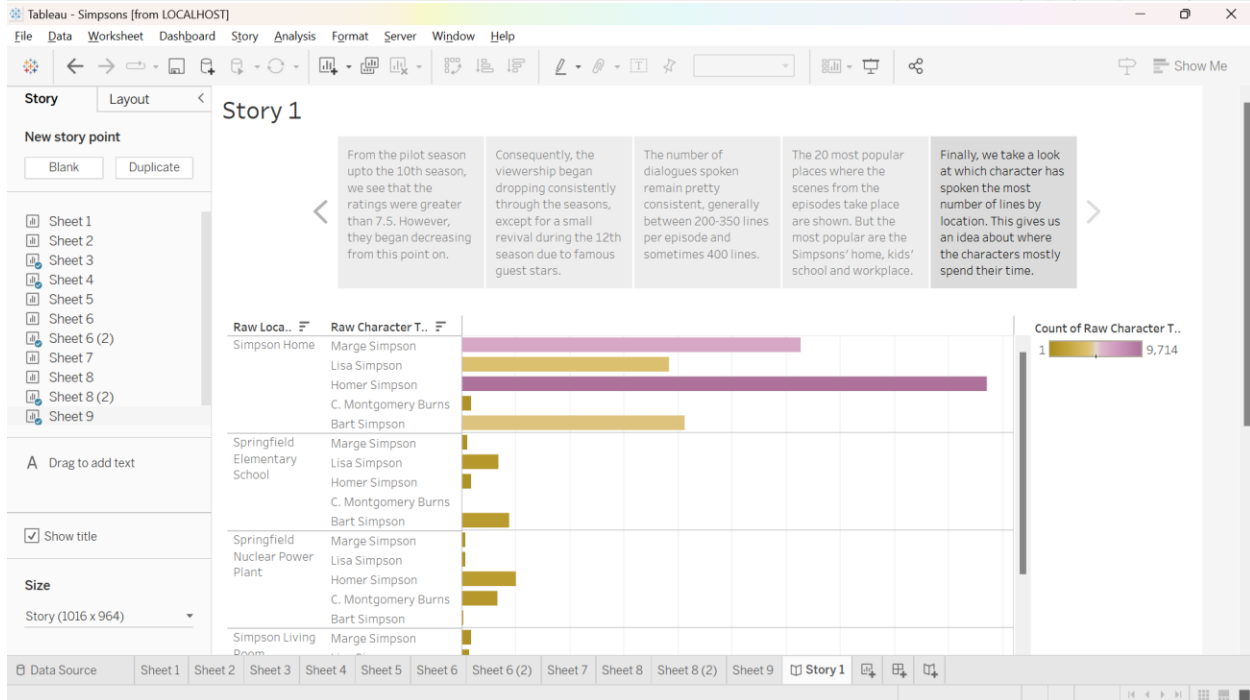
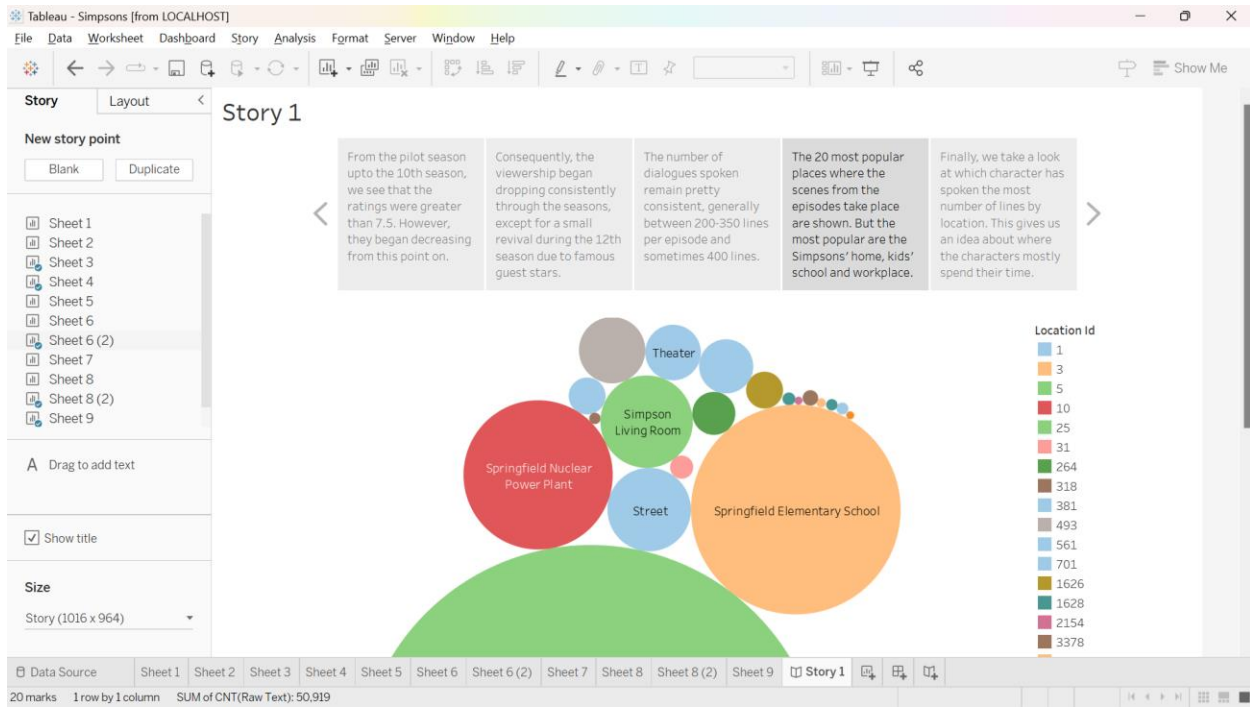
The number of scenes in a storyboard for a data visualization analysis of the factors affecting the insights of IT Sector Salaries, will depend on the complexity of the analysis and the specific insights that are trying to be conveyed. A storyboard is a visual representation of the data analysis process, and it breaks down the analysis into a series of steps or scenes.

Explanation video link:

<https://drive.google.com/drive/folders/1b2x5qmVVLytxuB-2b3SJoPBEGzB8YRIW?usp=sharing>







Performance Testing

Performance testing for Tableau focuses on evaluating the software's speed, responsiveness, and scalability under various conditions and workloads. It involves measuring and analyzing key performance indicators such as query response time, data loading speed, dashboard rendering time, and concurrent user handling capacity. The testing process helps identify any performance bottlenecks, optimize system configurations, and ensure that Tableau can handle the expected workload efficiently, providing users with a smooth and responsive experience while working with large datasets and complex visualizations.

Web Integration

Publishing helps us track and monitor key performance metrics, to communicate results and progress. Help a publisher stay informed, make better decisions, and communicate their performance to others. The project is made using Flask as backend framework.

Flask is a lightweight and flexible Python web framework that enables the development of web applications. It provides the necessary tools and libraries to build a robust backend that can handle user requests and serve dynamic content.

The Flask backend consists of the following components:

Route: We defined a single route in Flask to handle all incoming requests from users. This route is responsible for rendering the HTML templates and processing user interactions.

Templates: We have used a bootstrap template from bootstrapmade. Flask supports the use of templates for rendering dynamic content. We utilized this popular templating engine, to generate HTML pages dynamically based on the data retrieved from the backend. The templates contain the necessary HTML, CSS, and JavaScript code to create an interactive user interface.

Replit Hosting

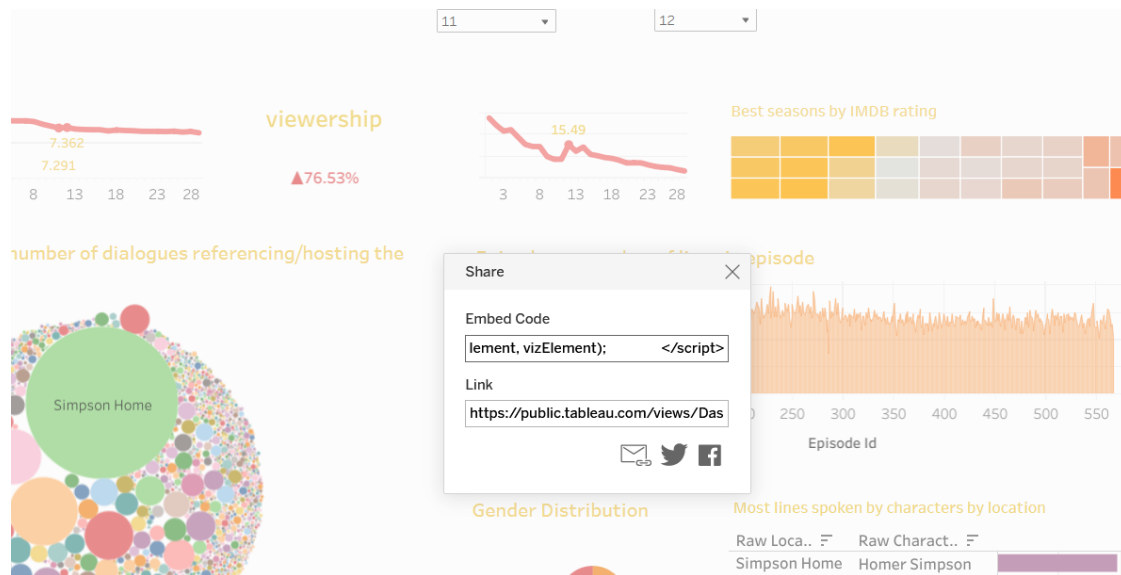
Replit is an online coding platform that provides an integrated development environment (IDE) and hosting capabilities. We chose Replit to host our Simpson data analytics application as it offers a seamless and convenient hosting solution for Flask applications.

By deploying our Flask backend on Replit, we were able to make our application accessible to users over the internet without the need for setting up complex server infrastructure. Replit provides a dedicated URL for our application, allowing users to access the Simpson data analytics platform with ease.

Additionally, Replit offers collaborative features that facilitated the development and deployment process. Our team members could work on the project simultaneously, making it convenient to collaborate and share code changes in real-time.

Integrating dashboard/reports/stories to web

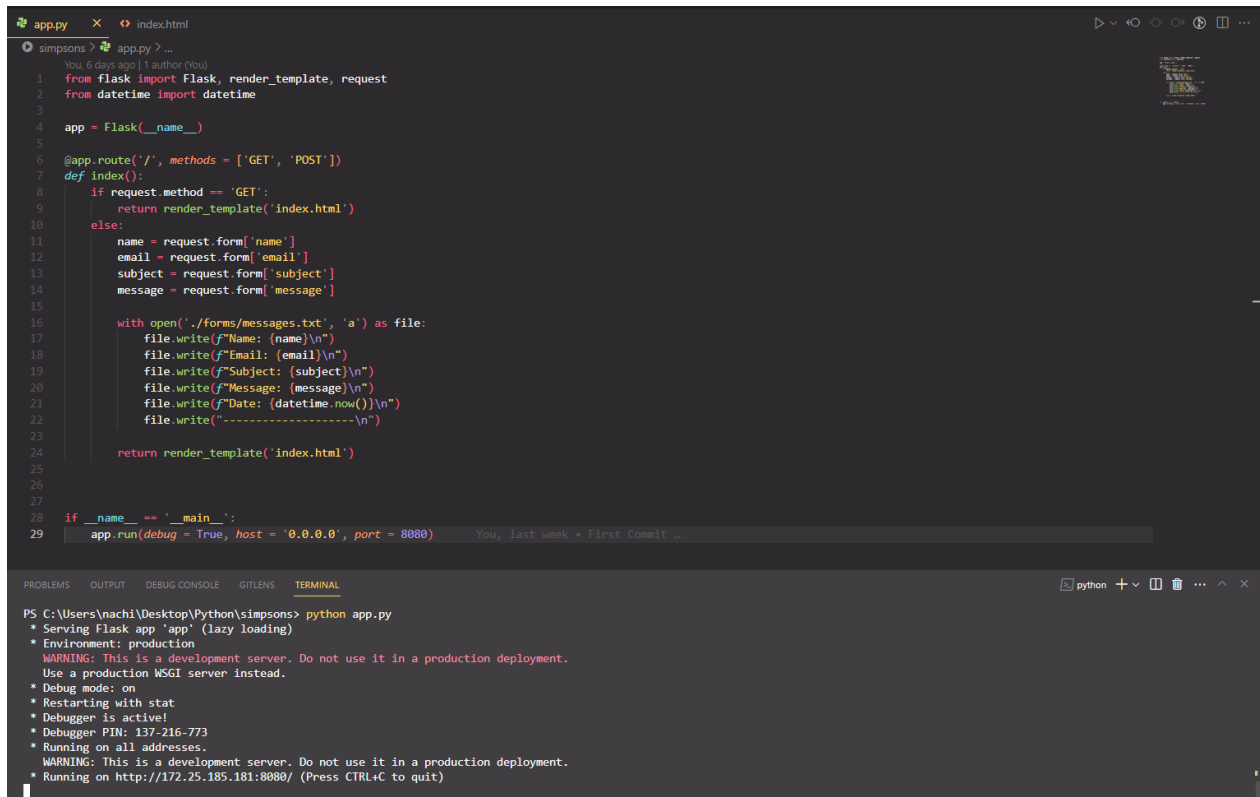
Step 1: Go to Dashboard/story/report, click on share button on the top ribbon



Step 2: Copy the embed Code and paste in index.html

```
app.py index.html X
dex.html > html > body > main#main > section#contact.contact > div.container > div.row.justify-content-center.align-items-center.mt-5 > div.col-lg-8 > form.php-email-form
170 </div>
171 </div>
172 </section>
173
174
175
176
177
178 <div class='tableauPlaceholder' id='viz1687519623274' style='position: relative'><noscript><a href='#'><img
179 alt='Dashboard 1 (2)'
180 src='https://public.tableau.com/static/images/Da/Dashboard_16875180852620%47;Dashboard128
181 style='border: none' /></a></noscript><object class='tableauViz' style='display: none;'
182 <param name='host_url' value='https://public.tableau.com/' />
183 <param name='embed_code_version' value='3' />
184 <param name='site_root' value='' />
185 <param name='name' value='Dashboard_16875180852620%47;Dashboard12' />
186 <param name='tabs' value='no' />
187 <param name='toolbar' value='yes' />
188 <param name='static_image'
189 value='https://public.tableau.com/static/images/Da/Dashboard_16875180852620%47;Dashboard128
190 <param name='animate_transition' value='yes' />
191 <param name='display_static_image' value='yes' />
192 <param name='display_spinner' value='yes' />
193 <param name='display_overlay' value='yes' />
194 <param name='display_count' value='yes' />
195 <param name='language' value='en-GB' />
196 <param name='filter' value='publish=yes' />
197 </object></div>
198 <script
199 type='text/javascript'> var divElement = document.getElementById('viz1687519623274'); var vizElement
200
201
202
203
204
205
206
```

Step 3: Run the Flask code:



The screenshot shows a code editor with a file named `app.py` and a terminal window below it. The code in `app.py` is a Flask application that listens on `/` and handles GET and POST requests. It uses `render_template` to serve `index.html`. For POST requests, it extracts form data (name, email, subject, message) and writes it to a file named `messages.txt` in the `./forms` directory. The terminal output shows the command `python app.py` being executed, and the Flask development server starting on `http://172.25.185.181:8080/`. The terminal also displays several warnings and status messages, including the environment (production), debug mode (on), and the server's IP and port.

```
1 from flask import Flask, render_template, request
2 from datetime import datetime
3
4 app = Flask(__name__)
5
6 @app.route('/', methods = ['GET', 'POST'])
7 def index():
8     if request.method == 'GET':
9         return render_template('index.html')
10    else:
11        name = request.form['name']
12        email = request.form['email']
13        subject = request.form['subject']
14        message = request.form['message']
15
16        with open('./forms/messages.txt', 'a') as file:
17            file.write(f"Name: {name}\n")
18            file.write(f"Email: {email}\n")
19            file.write(f"Subject: {subject}\n")
20            file.write(f"Message: {message}\n")
21            file.write(f"Date: {datetime.now()}\n")
22            file.write("-----\n")
23
24        return render_template('index.html')
25
26
27
28 if __name__ == '__main__':
29     app.run(debug = True, host = '0.0.0.0', port = 8080)
```

```
PS C:\Users\nachi\Desktop\Python\simpsons> python app.py
* Serving Flask app 'app' (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: on
* Restarting with stat
* Debugger is active!
* Debugger PIN: 137-216-773
* Running on all addresses.
  WARNING: This is a development server. Do not use it in a production deployment.
* Running on http://172.25.185.181:8080/ (Press CTRL+C to quit)
```

Activity 1: Implementing Flask

Activity 2: Integrating with Tableau Public

Activity 3: Hosting on Replit

Explanatory video: https://drive.google.com/file/d/17daBtSIPFSrH5G5Lm5OJSRY-Xb0m4gda/view?usp=drive_link