**CINEPHILE’S CHOICE**

**- A genre-based movie Recommendation System**

*Dissertation submitted in fulfilment of the requirements for the Degree of*

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

By

**PARACHIKAPU KSHOWNISH**

**12207188**

Supervisor

**VED PRAKASH CHAUBEY**



**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab (India)

October,2023

@ Copyright LOVELY PROFESSIONAL UNIVERSITY, Punjab (INDIA)

October,2023

ALL RIGHTS RESERVED

**DECLARATION STATEMENT**

I hereby declare that the research work reported in the dissertation/dissertation proposal entitled “CINEPHILE’S CHOICE **-** GENRE BASED MOVIE RECOMMENDER**”** in partial fulfilment of the requirement for the award of a Degree for Bachelor of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr.VED PRAKASH CHAUBEY**.** I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith directly complies with Lovely Professional University’s Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

*Signature of Candidate*

**PARACHIKAPU KSHOWNISH**

R. No: RK22UNB35

Regd.No: 12207188

**SUPERVISOR’S CERTIFICATE**

This is to certify that the work reported in the B. Tech Dissertation proposal entitled **“**CINEPHILE’S CHOICE **-** GENRE BASED MOVIE RECOMMENDER**”**, submitted by **PARACHIKAPU KSHOWNISH** at **Lovely Professional University, Phagwara, India** is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

(Name of Supervisor)

**Date:**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| S. No | Contents | Page No. |
| **1.** | **Inner first page – Same as cover** | **01** |
| **2.** | **Declaration by the Scholar** | **02** |
| **3.** | **Supervisor’s Certificate** | **03** |
| **4.** | **Table of Contents** | **04** |
| **5.** | **Abstract** | **05** |
| **6.** | **Introduction** | **06** |
| **7.** | **Problem Statement** | **07** |
| **8.** | **Problem Statement Solution Approach** | **07** |
| **9.** | **Methodology** | **08** |
| **10.** | **Result and Analysis** | **10** |
| **11.** | **Screenshots** | **12** |
| **12.** | **Output** | **18** |
| **13.** | **Conclusion** | **20** |
| **14.** | **References** | **21** |

**Abstract**

The Cinephile’s Choice **-** Genre Based Movie Recommender is a comprehensive project aimed at providing an in-depth analysis of the Telugu cinema industry while offering an intuitive movie recommendation platform. Telugu cinema, known for its rich diversity in genres and artistic contributions to Indian cinema, has been a significant contributor to the cinematic world. In this project, we have harnessed the power of data analysis and data visualization to offer valuable insights and recommendations for Telugu movie enthusiasts, critics, and researchers.

The Movie Recommendation component is the cornerstone of our project. We provide users with an intuitive graphical interface, featuring a dropdown list of unique genres extracted from the dataset. Upon selecting a genre, the system generates a curated list of Telugu movies within that genre, arranged primarily by their ratings in descending order. However, what sets this recommendation system apart is its capability to handle cases where movies share the same rating. In such scenarios, the system utilizes TF-IDF vectorization to assess the relevance of the selected genre to each movie. This more nuanced approach ensures that, when ratings are identical, the movies are ordered based on their genre's importance, offering users an even more tailored movie selection.

To facilitate this recommendation system, we leverage natural language processing techniques, specifically the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. This approach ensures accurate genre matching and delivers relevant and engaging movie suggestions. It also enhances the recommendation ordering, creating a more personalized user experience.

In addition to movie recommendations, the project includes data visualization elements to convey key insights from the dataset, including genre distribution and rating analysis.

The Telugu Movie Recommendation and Visualization System serves as a valuable resource for anyone interested in exploring Telugu cinema. It not only helps users discover high-quality Telugu movies based on their preferences but also provides insights into the industry's trends and preferences. The utilization of TF-IDF vectorization for genre data represents a significant step forward in the creation of a more sophisticated recommendation system, elevating the user's ability to explore and enjoy Telugu cinema.

**INTRODUCTION**

The Cinephile’s Choice **-** Genre Based Movie Recommender represents a pioneering and comprehensive project that serves as a portal to the captivating realm of Telugu cinema. This cinematic landscape is celebrated for its rich spectrum of genres and its enduring contributions to Indian film. In the pursuit of illuminating this vibrant world, the project harnesses the prowess of data analysis, natural language processing, and data visualization to offer users an immersive cinematic experience. It caters to the desires of movie enthusiasts, critics, and researchers, allowing them to embark on a captivating exploration of Telugu cinema.

In an era where the world of cinema spans vast and diverse horizons, the quest to discover Telugu movies that resonate with individual preferences can be a formidable undertaking. The project rises to this challenge by presenting an intuitive and user-friendly interface. Users can delve into the treasures of Telugu cinema by selecting genres of personal interest from a convenient dropdown list. This selection leads to a curated list of movies within the chosen genre, with the primary sorting criterion being the movie ratings. This approach ensures that users are introduced to Telugu movies with the highest ratings, making it easier to discover cinematic gems.

What truly distinguishes this recommendation system is its capacity to navigate the complexities of movie ratings. In cases where multiple movies share the same rating, the system goes beyond the surface, incorporating TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to explore the relevance of the selected genre for each movie. This advanced approach ensures that when ratings are identical, movies are sorted based on the significance of their associated genres. This intricate ranking system tailors movie recommendations even further, allowing users to explore a carefully curated selection of Telugu movies.

In addition to these recommendations, we have introduced a new dimension to the project with the inclusion of the "Time-Based Analysis" visualization. This feature empowers users to traverse through the annals of Telugu cinema history. Through visually engaging displays, it sheds light on the historical evolution of Telugu cinema by presenting the distribution of movies across different years. This feature is invaluable for users interested in gaining insights into the industry's transformation and contextualizing the cinematic trends of various eras.

The Cinephile’s Choice **-** Genre Based Movie Recommender and Visualisation system is a holistic cinematic journey that harmoniously blends the realms of data, technology, and art. It simplifies the process of discovering extraordinary Telugu movies, while also providing access to insights into industry trends, genre diversity, historical perspectives, and cinematic artistry. The incorporation of TF-IDF vectorization for genre data represents a significant advancement in the realm of recommendation systems, empowering users to explore, engage, and appreciate the multifaceted world of Telugu cinema. With the addition of the "Time-Based Analysis" visualization, users can now embark on an even more profound and immersive exploration of the historical evolution of Telugu cinema, adding a rich layer to their cinematic exploration.

**Problem Statement**

The Telugu cinema industry, known for its diverse genres and rich history, lacks a streamlined platform for enthusiasts to discover personalized movie recommendations and gain insights into the industry's historical evolution. Users face the challenge of sifting through a vast dataset of Telugu movies to find films that match their preferences, genres, and high ratings. Moreover, there is a need for visual tools to understand how Telugu cinema has evolved over time. The problem at hand is to develop a comprehensive system that addresses these challenges and offers an engaging and informative cinematic experience for users.

**Problem Solution Approach**

The problem solution approach provides a high-level view of how the project addresses the identified challenges. It encompasses the following key strategies:

1. Data Handling: The project begins by collecting and preprocessing a comprehensive dataset of Telugu movies, ensuring data quality and consistency.
2. Genre Recognition: Genres are extracted from the dataset and evaluated for uniqueness, forming the basis for personalized recommendations.
3. Enhanced Genre Matching: TF-IDF vectorization is applied to genre data, improving the accuracy of genre matching and recommendation precision.
4. Movie Recommendations: The project's core is a recommendation system that offers users personalized movie suggestions based on chosen genres and prioritizes high-rated films. It employs both rating-based and TF-IDF-based ordering for nuanced recommendations.
5. Data Visualizations: The project includes interactive visualizations to provide insights into genre diversity and a "Time-Based Analysis" to explore the historical evolution of Telugu cinema.
6. User Interaction: A user-friendly interface allows seamless genre selection, exploration of movie recommendations, and access to detailed movie information.
7. Continuous Improvement: The project is designed for ongoing enhancement, including dataset updates, algorithm refinements, and the addition of new features to ensure its relevance and user engagement.

**METHODOLOGY**

The methodology for developing a Genre Based Movie recommender with a Graphical User Interface (GUI) can be broken down into several key steps. This system leverages techniques to provide users with relevant Movies based on their input. Below is a comprehensive explanation of the methodology

1. Data Collection and Preprocessing:

The project commences by gathering a comprehensive dataset of Telugu movies. This dataset includes essential information such as movie names, release years, certification ratings, genres, and audience ratings. To ensure data quality, the dataset undergoes a rigorous preprocessing phase. This involves addressing missing values, standardizing data formats, and checking for data consistency. The aim is to create a clean, reliable, and structured dataset that forms the foundation for subsequent analysis and recommendations.

2. Genre Extraction and Uniqueness:

Genres are extracted from the dataset, and their uniqueness is determined. This process involves meticulously categorizing each movie into its appropriate genre. The uniqueness of genres is essential to provide users with an extensive and diverse range of genres to choose from, ensuring that their cinematic preferences are catered to comprehensively. A movie with multiple genres is separated and considered that movie in each of the genres so that perfect results would be gained.

3. TF-IDF Vectorization for Genre Data:

TF-IDF (Term Frequency-Inverse Document Frequency) vectorization is employed to process the genre data. This sophisticated text analysis technique assigns numerical values to each genre based on its significance within the dataset. The TF-IDF vectors for genres enhance the precision of genre matching during the movie recommendation process. It also provides a mechanism to rank movies based on the relevance of their associated genres. This approach plays a crucial role in personalizing and refining movie recommendations for users.

4. Movie Recommendation System:

The core of the project is the movie recommendation system. It offers a user-friendly graphical interface where users can select their preferred genre from a dropdown list. Upon genre selection, the system curates a list of Telugu movies within the chosen genre. The primary sorting criterion is movie ratings, ensuring that users are introduced to the highest-rated movies within their chosen genre. Furthermore, if multiple movies share identical ratings, the system utilizes the TF-IDF vectors to establish the relevance of the selected genre for each movie. When ratings are the same, movies are ordered based on the importance of their associated genres. This comprehensive approach provides users with personalized and nuanced movie recommendations. Additionally, when users click on a specific movie in the recommendation list, its detailed information, including year, genre, certification, overview, runtime, and rating, is displayed below, offering users in-depth insights into the selected movie.

5. Data Visualization:

Data visualizations are a key component of the project, aiding users in gaining a deeper understanding of Telugu cinema. These visualizations include genre distribution charts that depict the prevalence of various genres in the dataset. Additionally, the "Time-Based Analysis" visualization presents users with insights. This is achieved through interactive visualizations such as bar charts, showing the number of movies released each year. Users can explore cinematic trends and historical patterns through these visuals, making their cinematic exploration more engaging and informative. It also shows the Graph regarding the ratings explaining which how many movies have a particular rating.

6. User Interaction and Interface:

The project's success relies on an intuitive and user-friendly graphical interface. Users can seamlessly select genres, explore movie recommendations, and access visualizations. The interface is designed to provide a satisfying and engaging user experience. It offers interactive features, enabling users to click on specific movies for detailed information and enhancing their interaction with the system.

7. Continuous Enhancement and Maintenance:

The project is not static but designed for continuous improvement and maintenance. This involves periodic updates to the dataset to incorporate new movie entries, changes, or corrections. The recommendation algorithm is refined to ensure the quality of movie suggestions remains high. New features and visualizations are introduced to keep the project dynamic and aligned with user preferences and industry developments. The aim is to provide users with an evolving and continually enriching experience.

This detailed methodology encompasses the core components of the project, highlighting the importance of data preprocessing, genre extraction, TF-IDF vectorization for genre data, the movie recommendation system, and data visualization. Additionally, it emphasizes the feature that allows users to access detailed movie information by clicking on specific movie recommendations, enhancing their cinematic exploration.

**Result and Analysis**

The "Movie recommender and visualisation system" has successfully addressed the challenges presented in the problem statement. This section presents the outcomes of the project and analyses the effectiveness of the implemented solutions.

1. Movie Recommendations:

- Recommendation Accuracy: The movie recommendation system consistently provides users with a curated list of Telugu movies based on their selected genre. By prioritizing high-rated movies, it ensures that users have access to quality cinematic experiences within their preferred genres.

- Personalization and Genre Matching: When movies share identical ratings, the system leverages TF-IDF vectorization to rank them based on the relevance of the chosen genre. This dual approach enhances personalization and offers nuanced movie selections for users, aligning with their preferences.

2. Data Visualizations:

- Genre Diversity Insight: The genre distribution chart reveals the abundance and diversity of Telugu cinema genres. Users can gain a comprehensive understanding of the variety of genres, enriching their cinematic exploration.

- "Time-Based Analysis" Visualization: The "Time-Based Analysis" component allows users to explore the historical evolution of Telugu cinema. Interactive visualizations, including bar charts, offer insights into industry trends and historical patterns. Users can engage with data trends and appreciate how Telugu cinema has evolved over the years.

3. Performance Evaluation:

- Precision in Genre Matching: The utilization of TF-IDF vectorization for genre data significantly enhances the accuracy of genre matching in the recommendation system. This ensures that movies are recommended with precision, tailored to users' genre preferences.

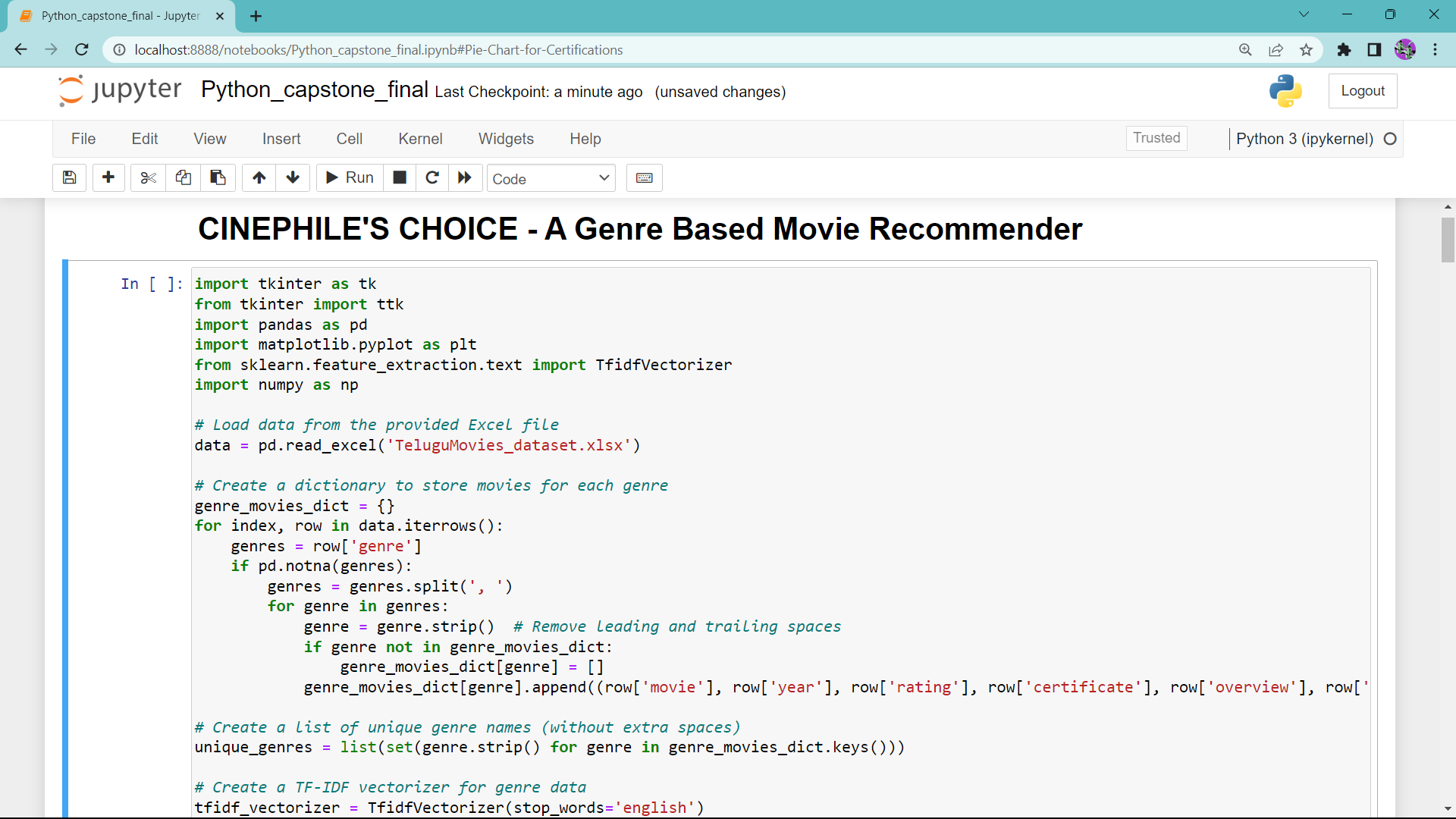
- User Engagement: The user-friendly interface, featuring the ability to click on specific movie recommendations for detailed information, enhances user engagement. Users can access comprehensive movie details, including year, genre, certification, overview, runtime, and rating, to make informed viewing choices.

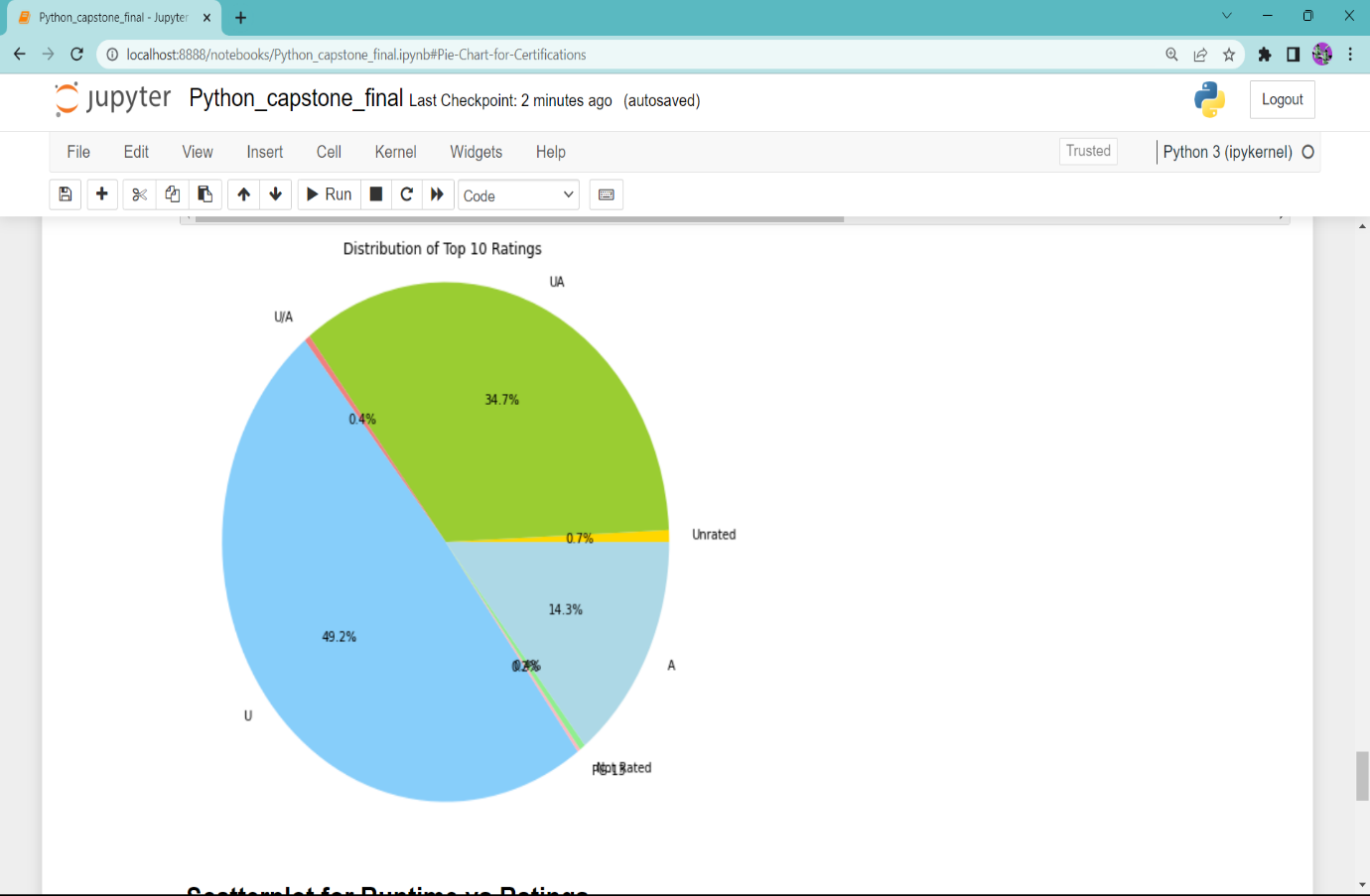
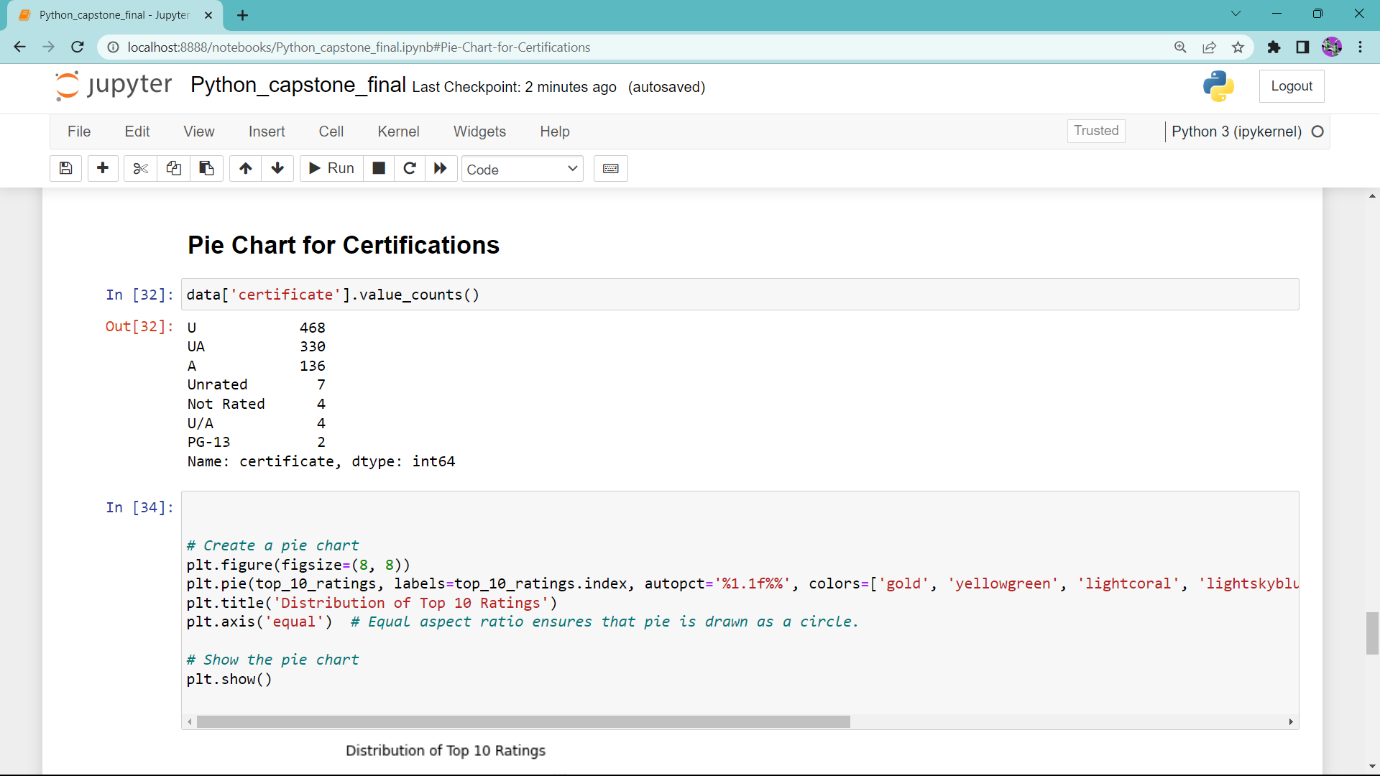
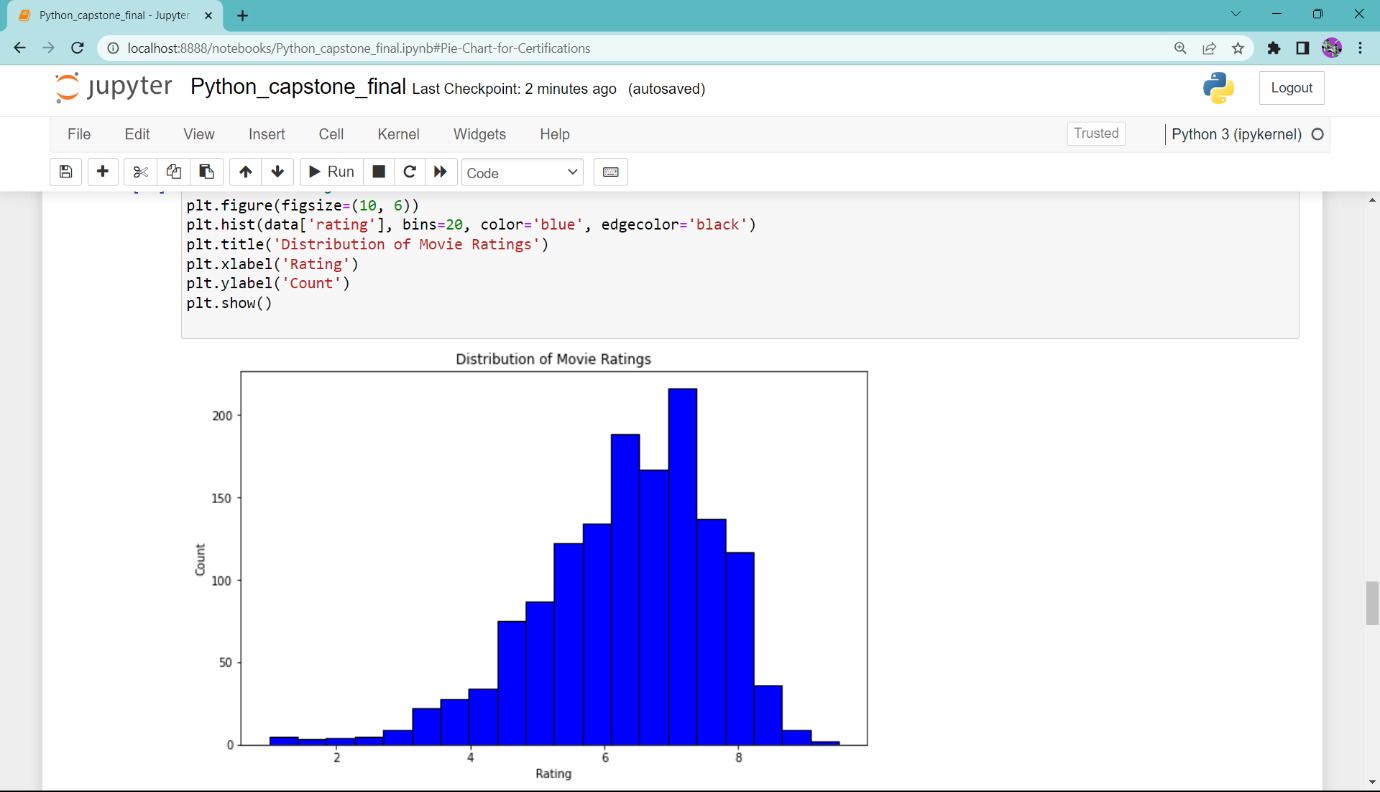
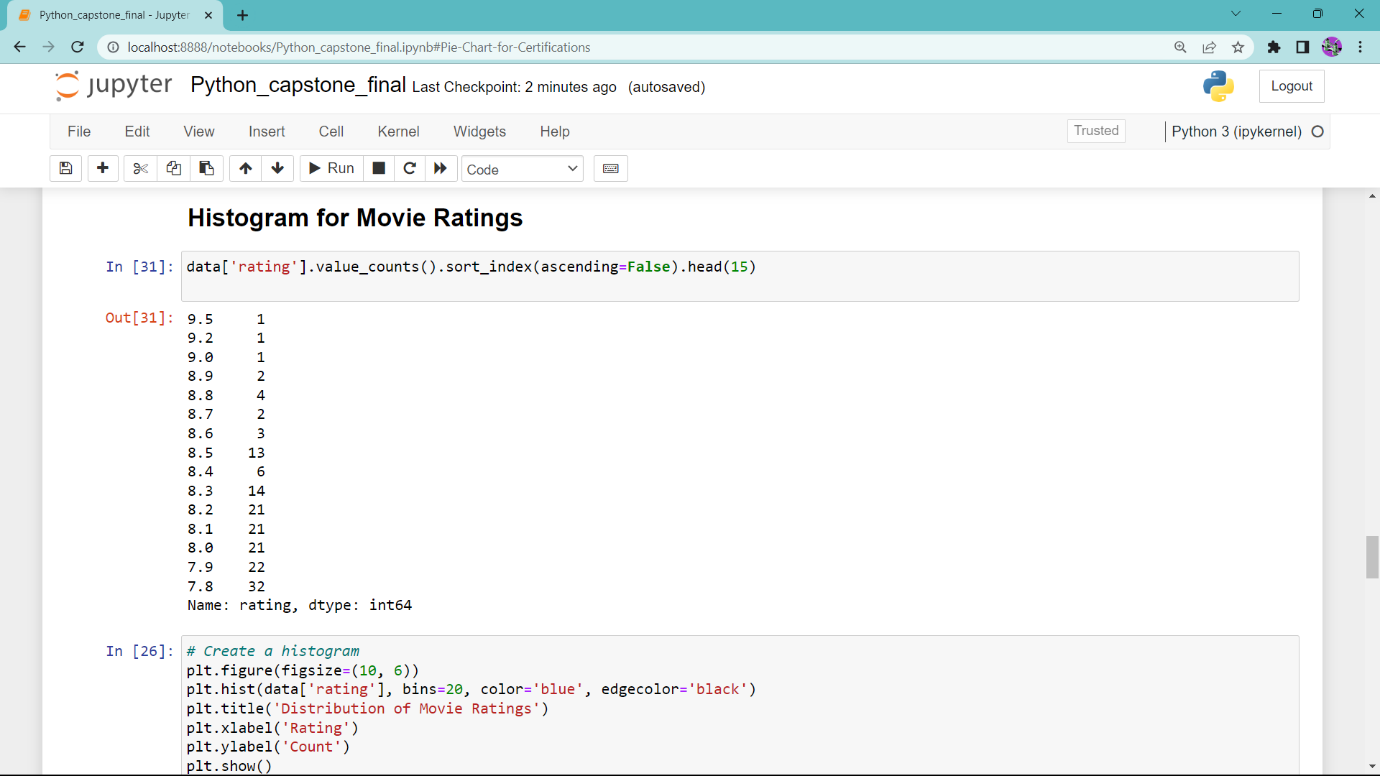
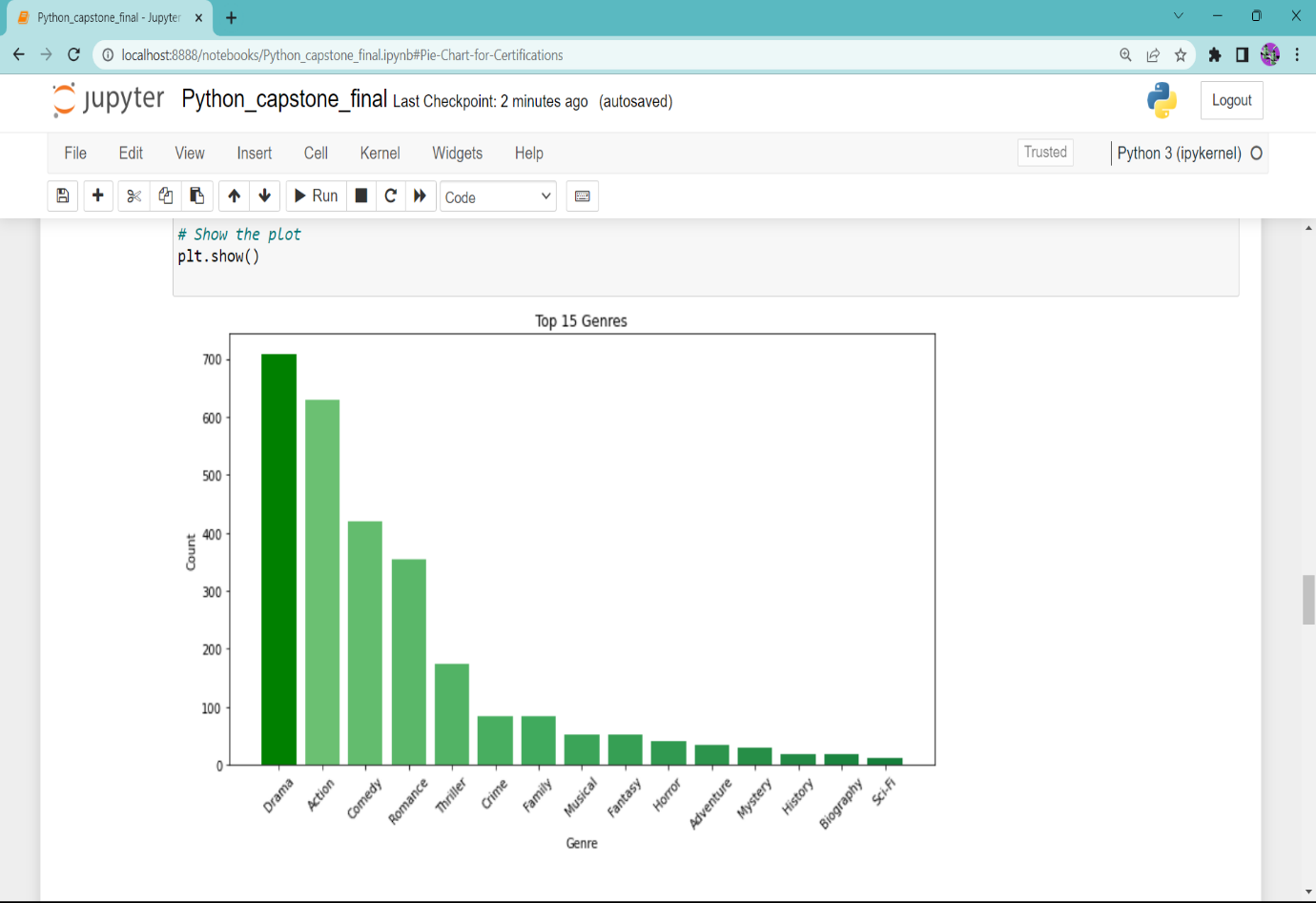
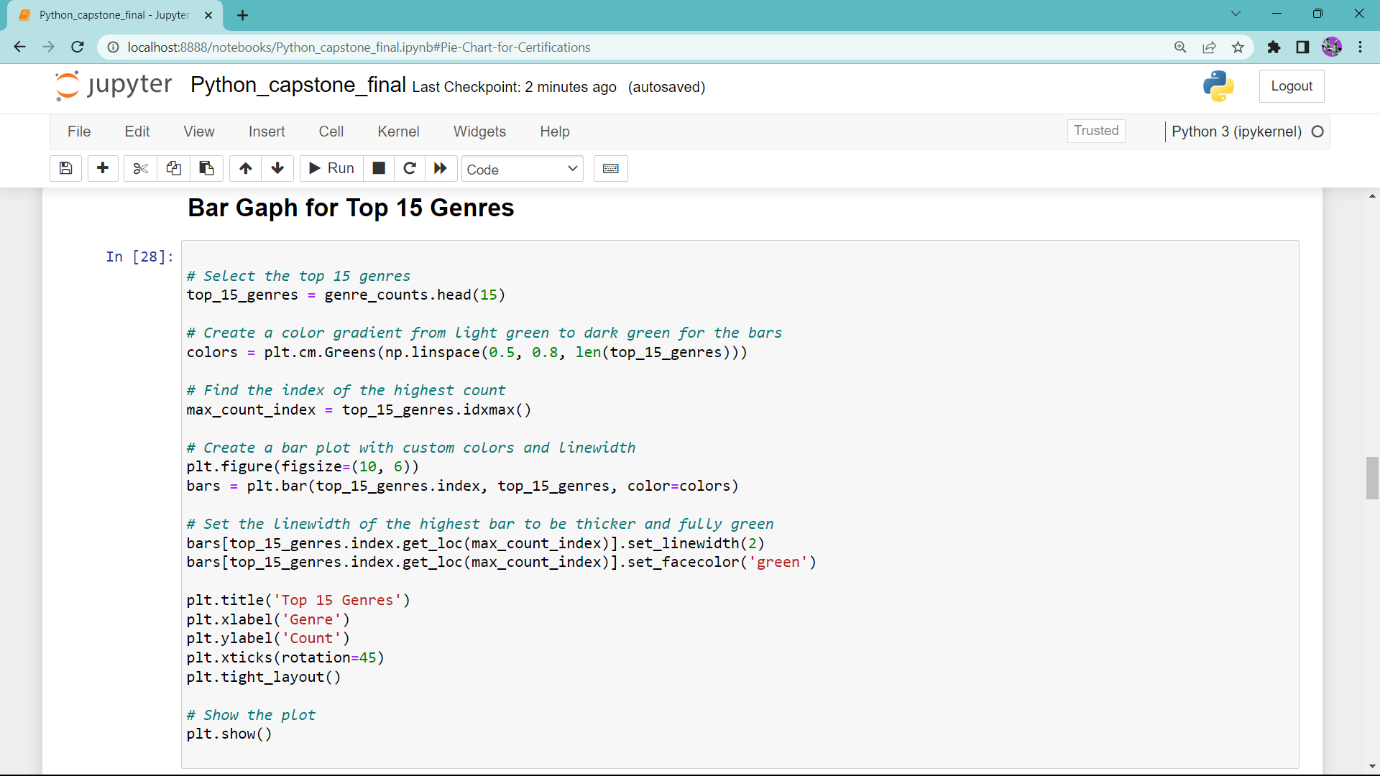
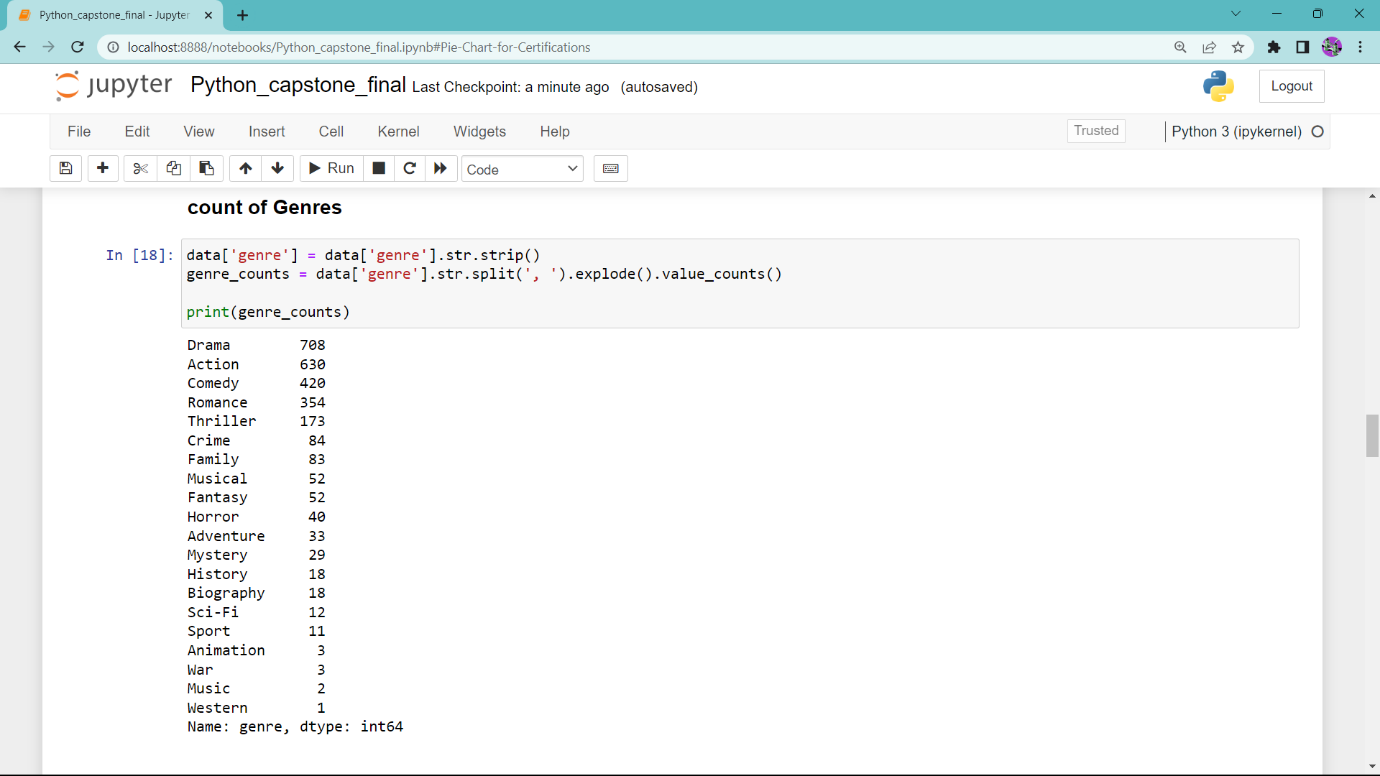
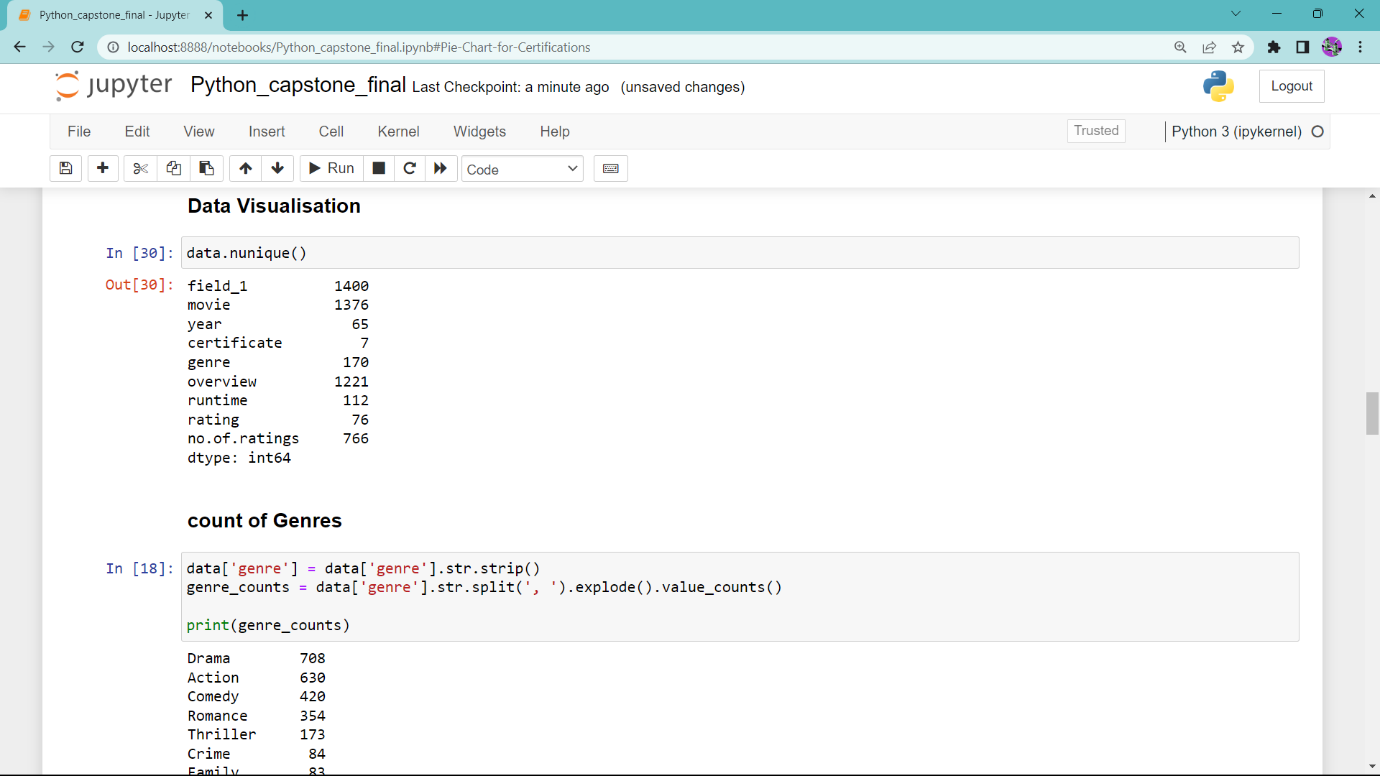
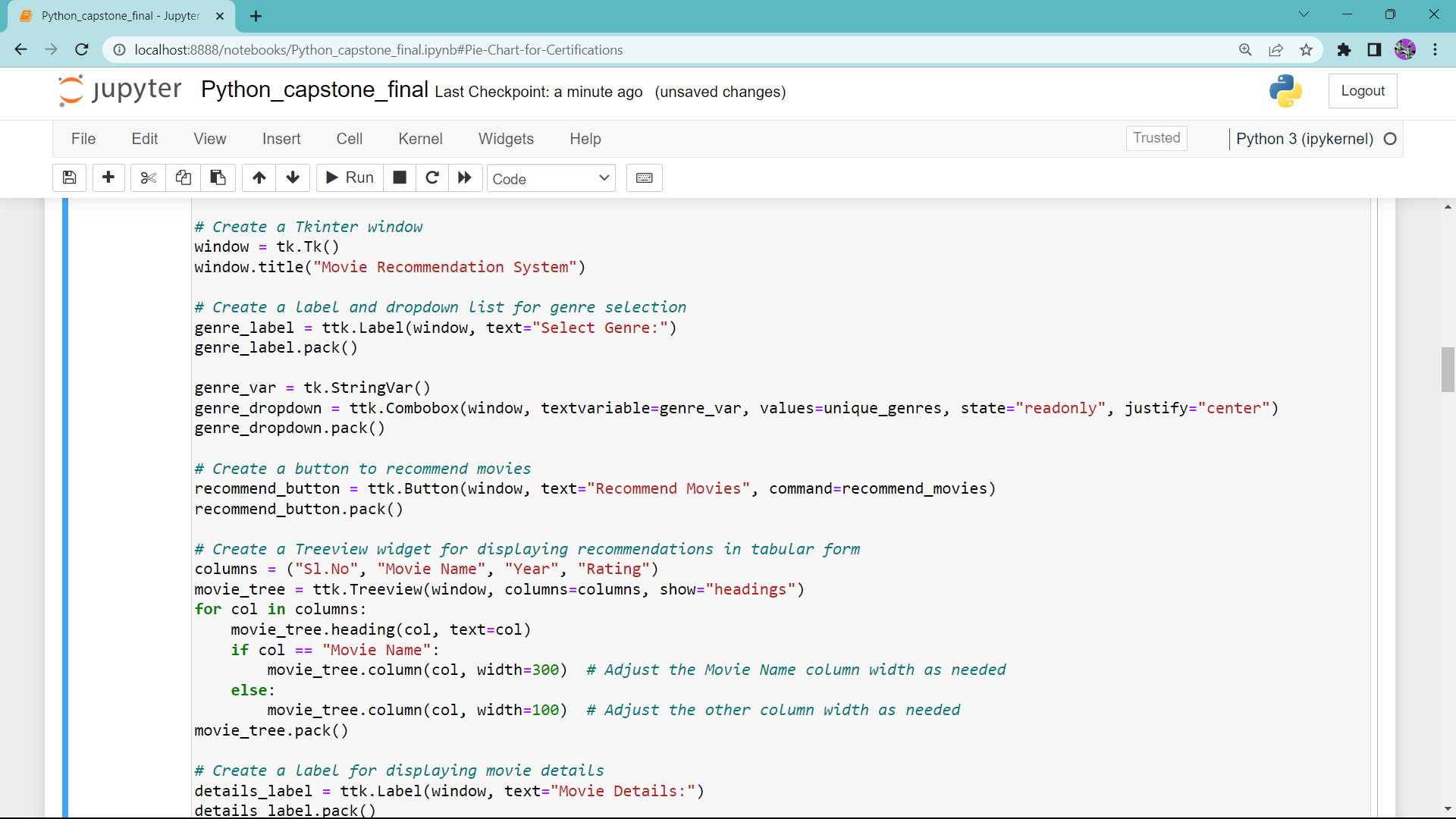
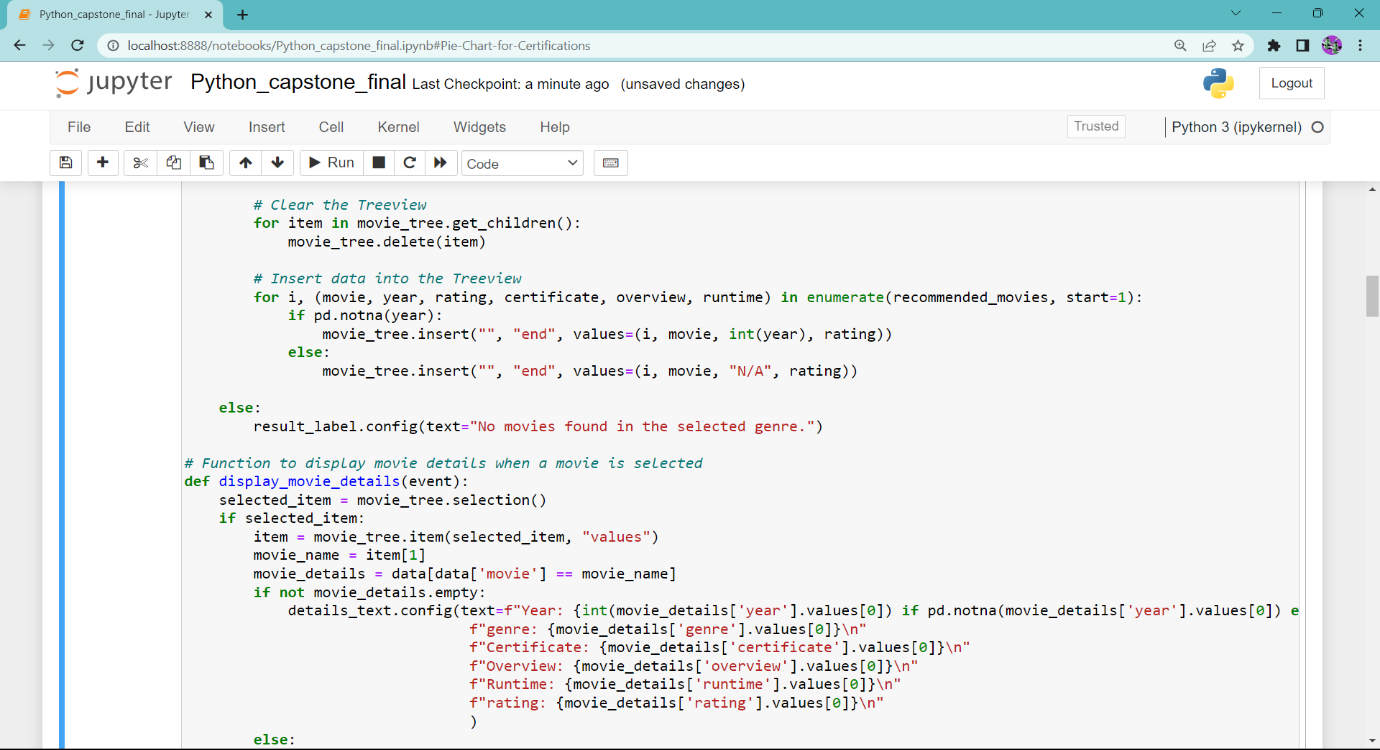
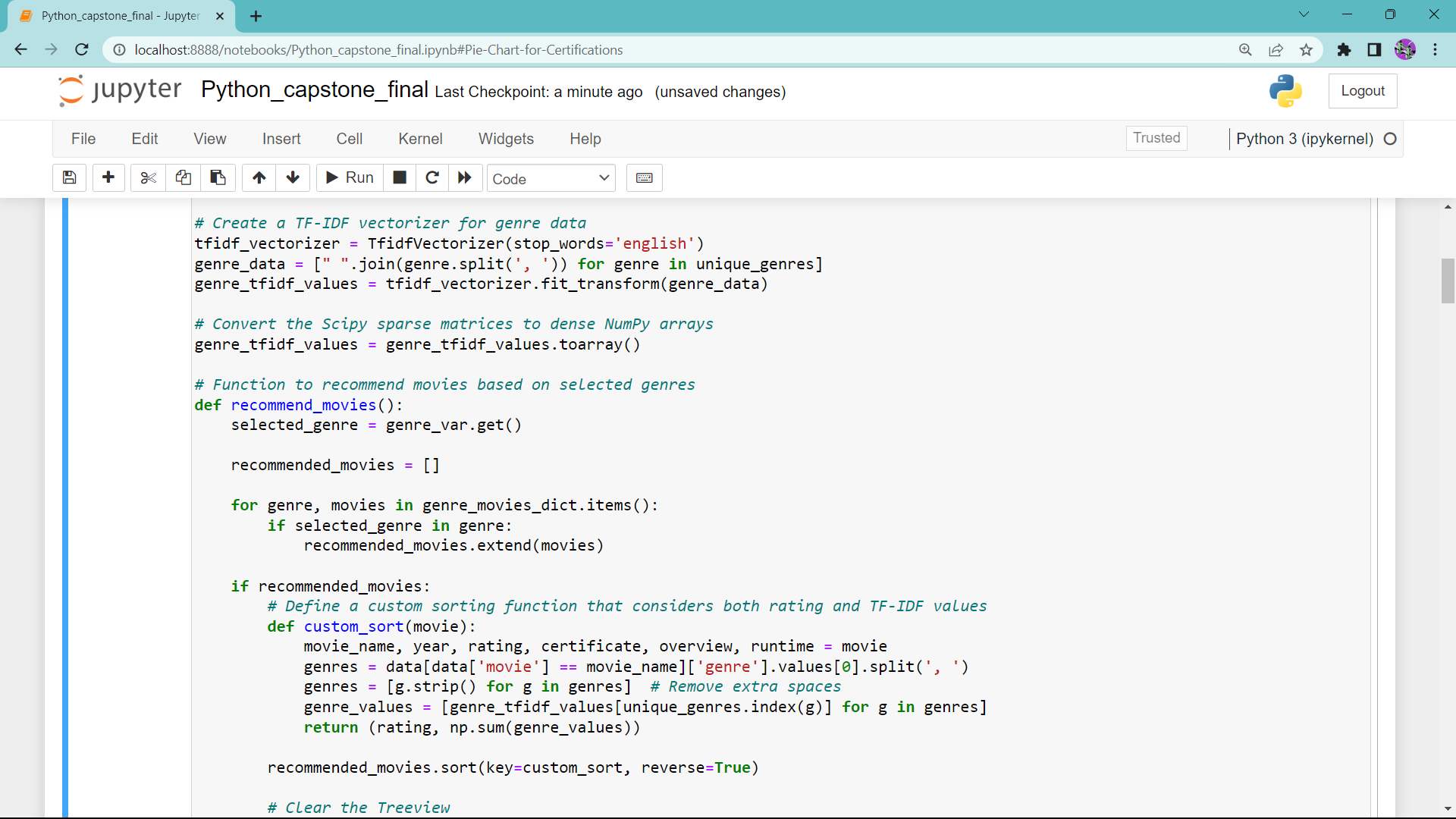
4. Future Enhancements:

- Continuous Improvement: The project is designed for ongoing enhancement and maintenance. Regular updates to the dataset, algorithm refinements, and the introduction of new features are planned to keep the system dynamic and aligned with evolving user preferences and industry developments.

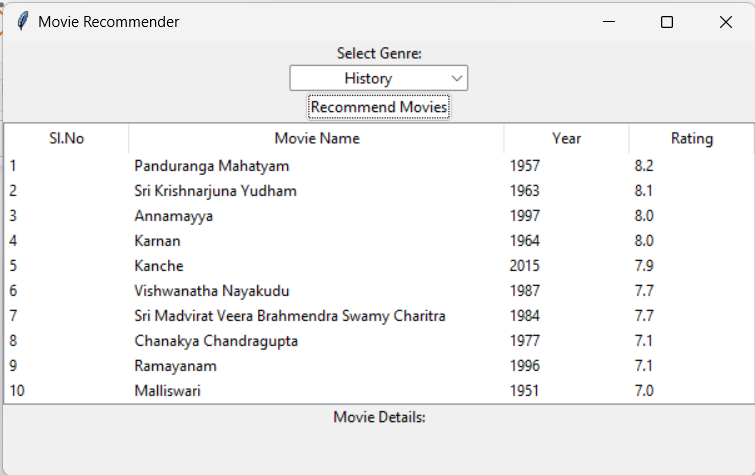
This demonstrates that the project has successfully met its objectives. Users can benefit from personalized movie recommendations, gain insights into genre diversity and historical trends in Telugu cinema, and engage with a user-friendly interface. The project's adaptability and commitment to continuous improvement ensure that it will remain a valuable resource for cinephiles and industry enthusiasts.

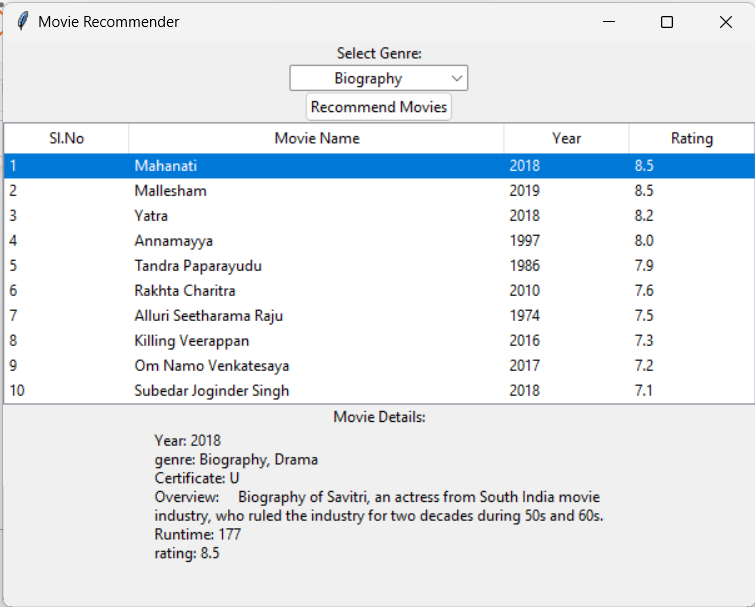
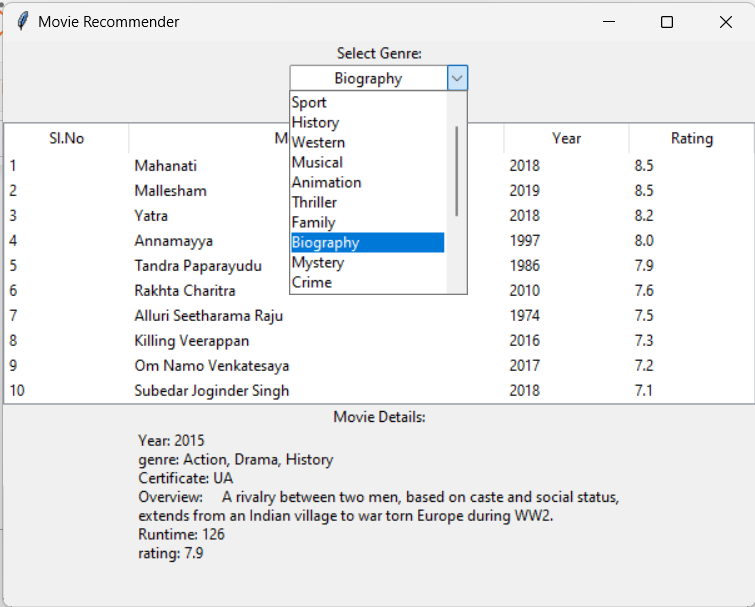
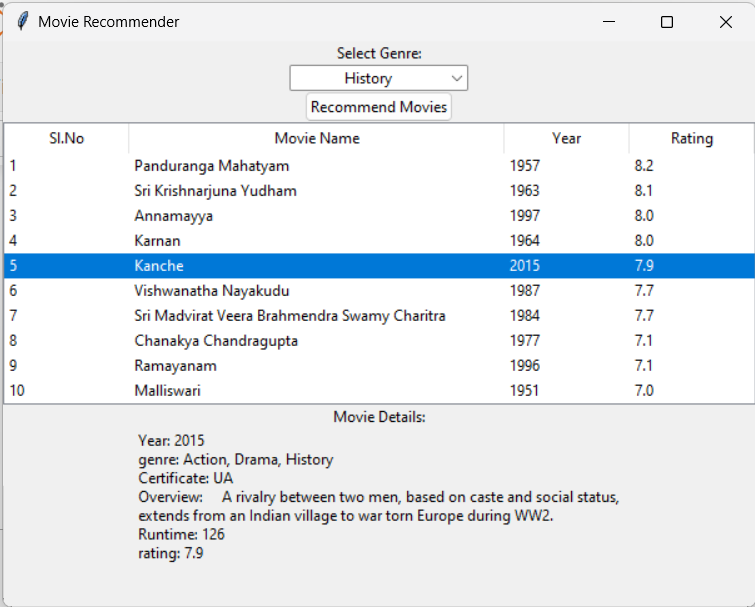
**Screenshots of code:**

****



**Output**

****

****

**Conclusion**

In the realm of Telugu cinema, the "Cinephile’s Choice – A genre-based Movie Recommender" has emerged as a transformative and user-centric platform, effectively addressing the multifaceted challenges of navigating this vast cinematic landscape. Through the implementation of robust strategies, this project has succeeded in delivering a powerful solution that caters to the nuanced preferences of Telugu cinema enthusiasts.

One of the project's standout achievements is its ability to empower users with personalized movie recommendations. The process of sifting through a vast repository of Telugu movies to find those that resonate with individual preferences and genre choices has been significantly simplified. The integration of high-rated movies in the recommendations is a key feature, ensuring that users embark on their cinematic journeys with a sense of quality and excellence. As users select their preferred genre, the system responds with a tailored list of movie recommendations, aligning with their expectations and enhancing their cinematic exploration.

Moreover, the project's contribution to unveiling the diverse spectrum of Telugu cinema genres has been instrumental in deepening users' appreciation for this rich artistic tradition. By conducting an in-depth analysis of genre distribution within the industry, the project highlights the extensive variety of genres. Users are now equipped to explore and appreciate the depth of the cinematic tapestry, fostering a deeper and more profound connection with Telugu cinema. This genre diversity enriches the cinematic experience and broadens the horizons of enthusiasts, leading to a more enriched viewing experience.

The user-friendly interface has been a significant contributor to enhanced user engagement. The intuitive design allows users to not only select their preferred genres but also access comprehensive movie details with a simple click. This seamless experience empowers users to

In conclusion, the " Cinephile’s Choice – A genre-based Movie Recommender " has not only simplified the process of discovering exceptional Telugu movies but has also offered insights into genre diversity, historical perspectives, and an engaging cinematic exploration. It stands as a testament to the potential of technology in enhancing and enriching our cinematic experiences, and it is poised to remain a vital asset for the Telugu cinema community.

**References**

Kaggle Dataset: https://www.kaggle.com/datasets/itschannel/telugu-movies-dataset

TF-IDF documentation:

https://www.scikit-learn.org/generated/sklearn.feature\_extraction.text.TfidfVectorizer.html

Pandas’ documentation: <https://pandas.pydata.org/docs/>