International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

NETFLIX DATASETS ANALYSIS USING PANDAS

[1]R.shirisha,[2]T.Vikash varma,[3]P.Shruthi,[4]J.Praveen Reddy

[1]Assistant Professor, [234]Students

Department of CSE(Data Science)

Siddhartha institute of technology & sciences,narapally

shirisharangu.cse@siddhartha.co.in, 23TQ1A6718@siddhartha.co.in, 23TQ1A6735@siddhartha.co.in, 23TQ1A6758@siddhartha.co.in

ABSTRACT

Netflix is one of the world's leading subscription-based streaming service providers, offering a vast library of movies, television series, documentaries, and original content across more than 190 countries. As of 2023, Netflix has accumulated over 238 million paid subscribers globally, generating enormous amounts of data about its content catalog, release patterns, genre distributions, and international production trends. The analysis of such data can provide valuable insights into the strategic direction of the platform, audience preferences, and emerging trends in the entertainment industry. This Mini Project presents a comprehensive Exploratory Data Analysis (EDA) of the Netflix dataset sourced from Kaggle. The dataset contains detailed information about approximately 8,807 titles available on Netflix, including attributes such as title name, type (Movie or TV Show), director, cast members, country of production, date added to the platform, release year, rating, duration, genre (listed_in), and a brief description of each title. The primary objective of this project is to analyze the Netflix content library to uncover meaningful patterns, trends, and insights that can inform content strategy, audience targeting, and regional content distribution decisions. The analysis focuses on several key dimensions including the temporal distribution of content additions, the balance between movies and TV shows, geographic content production patterns, genre popularity analysis, content rating distributions, and the identification of most prolific directors and actors on the platform.

I INTRODUCTION

In the modern digital world, streaming platforms have become one of the most popular sources of entertainment. Among these platforms, Netflix is one of the largest streaming services that provides a wide variety of movies, TV shows, documentaries, and web series.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Netflix generates a large amount of data related to its content such as genres, release years, ratings, and countries. Analyzing this data can help understand trends in entertainment, content popularity, and viewer preferences. Netflix was founded in 1997 and initially started as a DVD rental service. Over the years, it transformed into a global streaming service that provides on-demand entertainment content to millions of users. Today, Netflix operates in more than 190 countries and offers thousands of movies and television shows in multiple languages. This project focuses on analyzing the Netflix dataset using the programming language Python. Python is widely used in the field of data science because of its powerful libraries and tools that simplify data processing and visualization. In this project, the Pandas library is used for data cleaning and manipulation, while Matplotlib is used for generating graphical visualizations.

II LITERATURE SURVEY

The literature survey for this project spans multiple domains including streaming media analytics, exploratory data analysis methodologies, content recommendation systems, and data-driven entertainment strategy. A review of relevant prior work helps contextualize this project's contributions and situate it within the broader academic and industrial discourse on streaming platform analytics.

Streaming Platform Analytics Gomez-Uribe and Hunt (2015) presented a seminal study on Netflix's recommendation system, highlighting how data-driven personalization drives content discovery and reduces subscriber churn. Their work demonstrated that over 80% of content watched on Netflix is discovered through the recommendation engine, establishing the foundational importance of content data analysis for streaming platforms.

Hallinan and Striphas (2016) examined how Netflix's algorithmic culture shapes content production and consumption, introducing the concept of 'algorithmic culture' as a framework for understanding how data analytics influences not just how content is recommended but also what content is produced. Lobato (2019) conducted a comprehensive study of the global streaming industry in 'Netflix Nations', examining how Netflix navigates cultural, linguistic, and regulatory differences across its 190+ country operations. The study highlighted how regional content data analysis informs Netflix's localization strategy and international original content commissioning decisions. 2

Exploratory Data Analysis Methodologies Tukey (1977), in his foundational work 'Exploratory Data Analysis', established the principles of EDA as a systematic approach to understanding data before formal modeling. His advocacy for visual data exploration and pattern discovery without predetermined hypotheses forms the methodological backbone of

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

this project. Wickham and Grolemund (2017) in 'R for Data Science' articulated the data science workflow as consisting of data import, tidy, transform, visualize, model, and communicate stages — a framework closely aligned with the methodology employed in this project. McKinney (2017) in 'Python for Data Analysis' provided the foundational reference for Pandas based data manipulation, which serves as the primary data processing tool in this project. The text's treatment of DataFrame operations, missing value handling, and time-series analysis directly informed the preprocessing methodology.

III SYSTEM ANALYSIS

system analysis is the process of examining a system in detail to understand its components, functionality, and performance. it involves identifying the current problems, gathering requirements, and analyzing how different parts of the system interact with each other. the main goal of system analysis is to improve efficiency, productivity, and overall system performance by proposing better solutions or modifications. it plays a crucial role in software development and helps in designing systems that meet user needs effectively. system analysis is the process of studying and evaluating a system in detail to understand how it works and to identify areas for improvement. it involves collecting and interpreting data about the current system, understanding user requirements, and examining the interactions between different components. analysts focus on identifying problems, inefficiencies, and limitations in the existing system the process also includes feasibility analysis, which determines whether the proposed improvements are technically, economically, and operationally possible. based on the findings, system analysts suggest solutions or enhancements that can improve performance, reduce costs, and increase user satisfaction. overall, system analysis serves as a foundation for designing effective and efficient systems that align with organizational goals.

Existing system

The existing system refers to the current method or process that is being used before any improvements or new systems are introduced. it may be manual, semi-automated, or fully automated, depending on the organization. in many cases, the existing system may face issues such as inefficiency, time consumption, higher chances of errors, and difficulty in managing large amounts of data.

due to these limitations, users may experience delays and reduced productivity. analyzing the existing system helps in identifying these problems and understanding its strengths and weaknesses. this understanding is important for developing a better and more efficient proposed system that can overcome the drawbacks of the current one.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

Disadvantages of existing system

- time consuming processes
- higher chances of human errors
- lack of automation
- difficult to manage and store large data
- low efficiency and productivity

Proposed system

The proposed system is an improved and advanced version designed to overcome the limitations of the existing system. it focuses on automation, accuracy, and efficiency by using modern technologies and tools. the proposed system reduces manual work, minimizes errors, and allows faster processing of data.it provides better data management, improved security, and easy access to information. users can generate reports quickly and make better decisions based on accurate data. overall, the proposed system aims to increase productivity, save time, and deliver a more reliable and user-friendly experience.

Advantages of proposed system

- increases efficiency and productivity
- minimizes human errors
- faster data processing
- improved data accuracy

IV METHODOLOGY

This project employs a descriptive research design using quantitative data analysis methods. The methodology follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, adapted for EDA-focused analytics projects. The CRISP-DM framework provides a structured, iterative approach to data analytics that is widely adopted in both academic and industrial settings.

Business Understanding The initial phase involves defining the analytical objectives in terms of business questions. Key questions framing this analysis include: What is the composition of Netflix's content library? How has the library evolved over time? Which countries and genres are most represented? Who are the most prolific creators? What audience demographics does the content target? These questions translate directly into the analytical tasks performed in subsequent phases.

Data Understanding In this phase, the Netflix dataset is loaded and subjected to initial inspection. Key activities include:

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

• Loading the CSV file using pandas.read_csv() with appropriate parameter configurations.
• Inspecting the DataFrame structure using df.info(), df.shape, and df.dtypes. • Examining the first and last records using df.head() and df.tail().
• Computing basic descriptive statistics using df.describe().
• Identifying missing values using df.isnull().sum() and visualizing the missingness pattern.
• Checking for duplicate records using df.duplicated().sum().

System Architecture

it explains the organization of hardware, software, data, and network elements within a system. the architecture provides a blueprint that guides the development and implementation process.it includes various layers such as presentation layer, application layer, and data layer, each having specific responsibilities. system architecture ensures scalability, reliability, and performance by defining clear communication between components. it also helps in maintaining and upgrading the system easily by providing a well-structured framework. overall, system architecture theory is essential for building efficient, flexible, and robust systems.

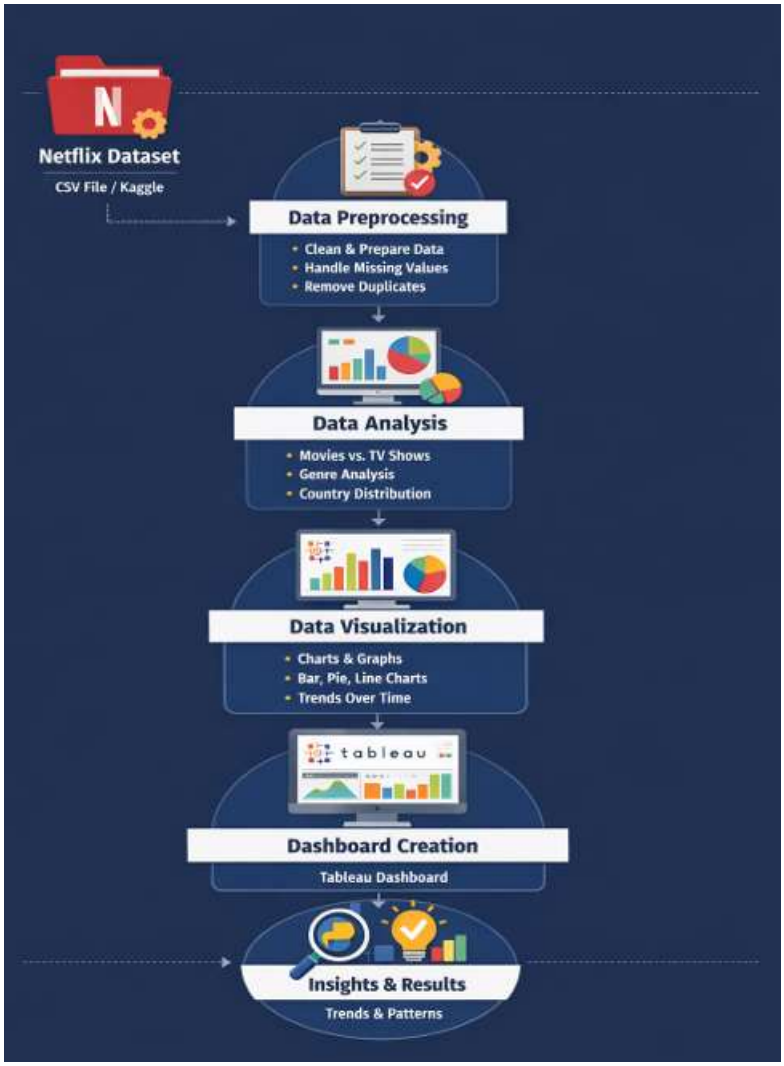
system architecture theory also focuses on defining the overall structure, behavior, and more detailed design of a system. it helps in selecting appropriate technologies, frameworks, and design patterns to ensure the system meets both functional and non-functional requirements such as performance, security, and scalability.it involves different architectural styles like client-server architecture, layered architecture, microservices architecture, and monolithic architecture. each style has its own advantages and is chosen based on the system needs. good system architecture improves communication between components, reduces complexity, and makes the system easier to test, deploy, and maintain.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



V RESULTS&OUTPUT

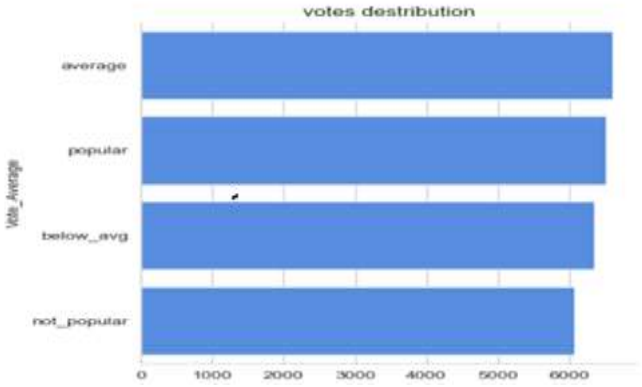
.

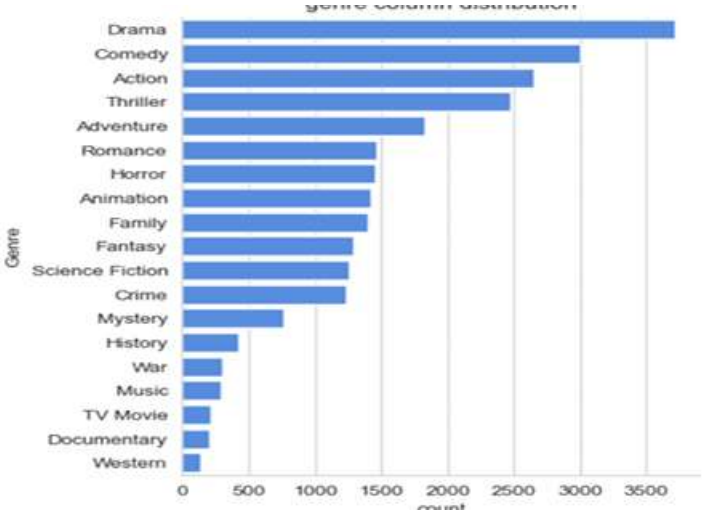
International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



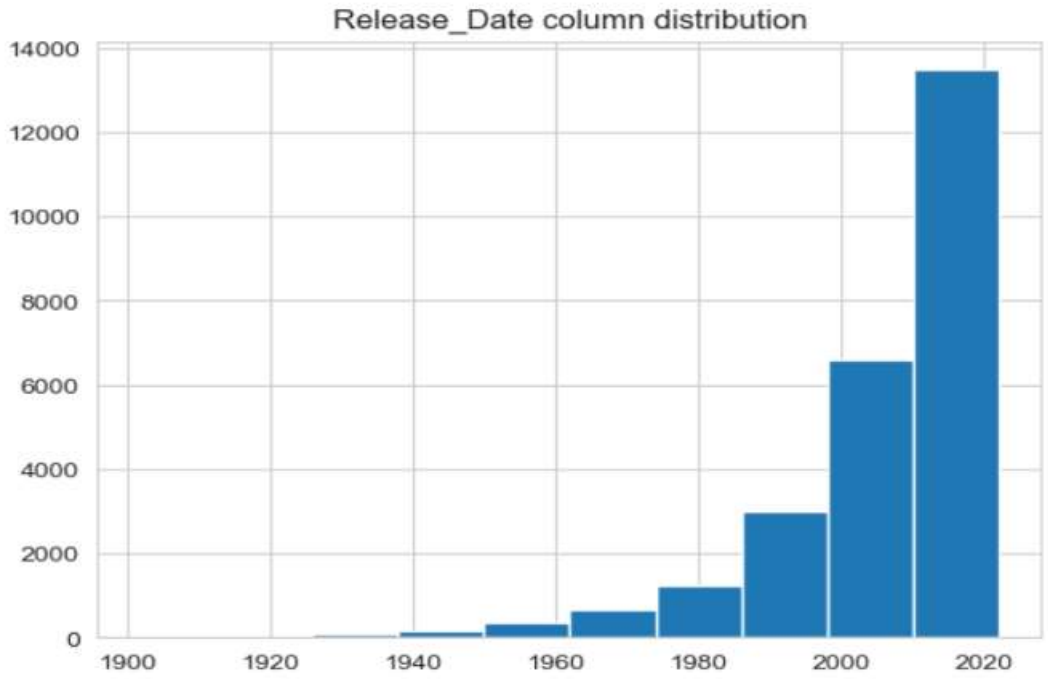


International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper



VI CONCLUSION

This Mini Project has successfully conducted a comprehensive Exploratory Data Analysis of the Netflix content dataset comprising 8,807 titles. The analysis followed a systematic data science pipeline encompassing data acquisition from Kaggle, extensive preprocessing and feature engineering, multi-dimensional exploratory analysis, and production of informative visualizations. The project has demonstrated the power of Python's data science ecosystem — particularly Pandas, Matplotlib, and Seaborn — in extracting meaningful insights from real-world entertainment data.

The analysis has yielded the following key findings:

• Netflix's content library is predominantly composed of Movies (69.6%) with TV Shows representing 30.4% of the total catalog, reflecting a diverse content strategy that combines a broad movie catalog with selective, high-quality TV Show investments.

• The United States dominates content production on Netflix, followed by India and the United Kingdom. The strong representation of Indian content reflects Netflix's significant investment in one of the world's largest and fastest-growing streaming markets.

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

• Netflix's content library experienced its most rapid expansion during the period from 2016 to 2019, with 2019 marking the peak year for new content additions.

• The platform predominantly targets adult and mature teen audiences, with TV-MA and TV-14 ratings accounting for over 55% of all content.

• International Movies and Dramas are the most represented genre categories, reflecting Netflix's dual strategy of acquiring global international content while investing in emotionally engaging dramatic storytelling across markets.

• July and December are the peak months for content additions, suggesting a deliberate calendar strategy aligned with summer and holiday viewership peaks.

REFERENCE

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies:

International Journal of

DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X ww.ijdim.com Original Research Paper

From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.