Journal of Intelligent Decision Support System (IDSS) 6 (4) (2023), pp. 216-222

Published by: Institute of Computer Science (IOCS)

Journal of Intelligent Decision Support System (IDSS)

Journal homepage: www.idss.iocspublisher.org

Analysis streaming application viewership with EDA

Aminuddin Indra Permana

University of Pembangunan Panca Budi, Medan, Indonesia

Article Info

Article history:

Received Nov 13, 2023
Revised Nov 29, 2023
Accepted Nov 30, 2023

Keywords:

Application;
EDA;
Prediction;
Streaming;
Quality.

ABSTRACT

Streaming applications require a high amount of bandwidth to deliver high-quality media content to users. However, bandwidth is not always available or consistent, especially in remote or congested areas. This can result in buffering, lagging, or poor quality of the streaming content, which can frustrate users and affect their satisfaction and retention. Streaming applications need to minimize the delay between the source and the destination of the media content, especially for live or interactive streaming. However, latency can be affected by many factors, such as network congestion, server load, routing, encoding, etc. Predictive analysis can help to forecast the future outcomes or behaviors of the streaming data, such as the demand, the popularity, the retention, the churn, etc. For example, one can use predictive analysis to estimate the optimal pricing strategy for a streaming service, or to predict the likelihood of a viewer to cancel their subscription. Streaming application with EDA can also help to detect and resolve any issues or errors that may affect the streaming quality, such as network congestion, server load, device compatibility, etc. Streaming application with EDA can help to understand and predict the user behavior, such as the viewing duration, frequency, preference, rating, feedback, etc., of the media content consumed by the users.

Corresponding Author:
Aminuddin Indra Permana,
Faculty of Science and Technology,
University of Panca Budi,
Jln. Gatot Subroto, Medan Sunggal, Sumatera Utara, Medan, 20122, Indonesia.
Email: aminuddin@dosen.pancabudi.ac.id

Introduction

EDA is a process of examining, summarizing, and visualizing data to gain insights and understanding (Mukhiya & Ahmed, 2020; Reid et al., 2020; Thakkar & Vikas, 2022). EDA can help to identify patterns, trends, outliers, and relationships in data, as well as to formulate hypotheses and test assumptions (Velleman & Hoaglin, 2023). EDA can also help to prepare your data for further analysis, such as machine learning or statistical modeling (Kumar & Joshna, 2021; Mukhiya & Ahmed, 2020; Peng et al., 2021). There are many sources of information and examples of EDA on streaming application viewership, such as documentary, thriller, sci-fi, comedy, one of the leading streaming platforms globally (Amier & Setiawan, 2019; Darapaneni et al., 2020; Zaveri et al., 2023).

(Gu et al., 2023) This article provides a comprehensive data analysis blog that delves deep into the world of Netflix. It uses Python and its data analysis libraries to explore the vast collection of Netflix's offerings and uncover valuable information that sheds light on content additions, duration distributions, genre correlations, and more. It also analyzes release patterns, seasonal trends, and audience preferences to better understand the content dynamics within Netflix's vast universe.

(Sanjana et al., 2023) This article discusses the essential elements of modern event-driven solutions, such as the ability to process continuous event streams to derive real time insights and intelligence. It also describes the capabilities and technology choices that are available to provide streaming analytics, geospatial analysis, text analytics, rules, and complex event processing as part of the event-driven architecture.

(Sahoo et al., 2019) This article presents an exploratory analysis and visualization of the Netflix dataset consisting of both movies and shows. It compares the total number of movies and shows in the dataset, the distribution of ratings, the most popular genres, the most frequent actors and directors, and the growth of Netflix content over the years. It also uses interactive plots and charts to illustrate the findings.

Streaming applications are software platforms that allow users to watch, listen, or interact with various types of media content, such as videos, music, podcasts, games, etc (Chukwu, 2023; Ma, 2023; Srimahalap, 2020). Streaming applications can offer many benefits, such as convenience, accessibility, personalization, and interactivity. However, streaming applications also face many challenges (Criollo-C et al., 2021; Khan et al., 2022), such as: (a). Streaming applications require a high amount of bandwidth to deliver high-quality media content to users. However, bandwidth is not always available or consistent, especially in remote or congested areas. This can result in buffering, lagging, or poor quality of the streaming content, which can frustrate users and affect their satisfaction and retention (Kesavan et al., 2021). (b). Streaming applications need to minimize the delay between the source and the destination of the media content, especially for live or interactive streaming. However, latency can be affected by many factors, such as network congestion, server load, routing, encoding, etc. High latency can cause synchronization problems, missed events, or reduced interactivity, which can diminish the user experience and engagement (Bouraqia et al., 2020; Yaqoob et al., 2020). (c). Streaming applications need to support a wide range of devices and platforms, such as smartphones, tablets, laptops, smart TVs, gaming consoles, etc. However, different devices and platforms may have different specifications, capabilities, formats, and standards, which can make it difficult to ensure a consistent and optimal streaming quality across all devices and platforms. Streaming applications need to adapt to the device and platform characteristics, such as screen size, resolution, aspect ratio, codec, etc., and provide the best possible streaming quality for each device and platform (Falkowski-Gilski & Uhl, 2020; Johnson, 2020).

Method

There are different methods of streaming application viewership with EDA, depending on the type and purpose of the analysis. Some of the common methods are:

- This method involves summarizing and visualizing the basic characteristics of the streaming data, such as the number of viewers, the duration of viewing, the genres of content, the ratings, the demographics, etc. Descriptive analysis can help to understand the overall patterns and trends in the streaming data, as well as to identify outliers and anomalies. For example, one can use descriptive analysis to compare the viewership of different streaming platforms, such as documentary, thriller, sci-fi, comedy or to explore the seasonal variations in the streaming behavior of the viewers.
- This method involves using statistical or machine learning techniques to model the streaming data and make predictions or estimates based on the historical or current data. Predictive analysis can help to forecast the future outcomes or behaviors of the streaming data, such as the demand, the popularity, the retention, the churn, etc. For example, one can use predictive analysis to estimate the optimal pricing strategy for a streaming service, or to predict the likelihood of a viewer to cancel their subscription.
- This method involves using optimization or simulation techniques to provide recommendations or actions based on the streaming data and the desired goals or objectives. Prescriptive analysis can help to improve the decision making or performance of the streaming data, such as the content recommendation, the personalization, the marketing, the customer service, etc. For example, one can use prescriptive analysis to generate personalized content suggestions for each viewer, or to optimize the advertising campaigns for a streaming service.

Results and Discussions

Streaming application viewing results and discussions that can be taken from the EDA streaming application viewership:

- One result is that documentary, thriller, sci-fi, comedy has a dominant position in the streaming industry, with a large and diverse user base, a global presence, and a vast library of content. The EDA can show how documentary, thriller, sci-fi, comedy has grown over the years, what are the most popular genres, ratings, and countries of its content, and how it has adapted to the local preferences and languages of its viewers. The discussion can focus on the strategies and challenges that documentary, thriller, sci-fi, comedy faces to maintain its competitive edge, such as content production, pricing, recommendation, and personalization.
- Another result is that streaming applications can benefit from event-driven architectures, which enable them to process continuous streams of data and derive real-time insights and intelligence. The EDA can show how streaming applications can use various technologies and capabilities, such as streaming analytics, geospatial analysis, text analytics, rules, and complex event processing, to enhance their performance and user experience. The discussion can focus on the design and implementation of event-driven solutions, such as the event sources, the event processing, the event consumers, and the event feedback.
- A third result is that streaming applications can use visualization techniques to explore and communicate their data effectively. The EDA can show how streaming applications can use different types of plots and charts, such as bar charts, pie charts, line charts, scatter plots, heat maps, etc., to illustrate the patterns and trends in their data, as well as to identify outliers and anomalies. The discussion can focus on the best practices and principles of data visualization, such as choosing the appropriate visual form, encoding the data accurately, highlighting the key messages, and providing interactivity and context.

Table 1. Duration Watch

Device_ID	Video_ID	Duration_Watched (minutes)	Age	Ratings	Interaction_Events	
count	6214	6214	6214	6214	6214	6214
mean	496.76215	50.323785	60.378389	39.665594	2.987448	50.776794
std	287.681163	28.671178	34.597273	17.571317	1.403304	29.009363
min	1	1	0.055809	10	1	0
25%	248	26	30.413538	24	2	26
50%	493.5	50	60.765637	40	3	51
75%	742	75	90.874141	55	4	76
max	1000	100	119.999972	70	5	100

The classification based on age is as follows

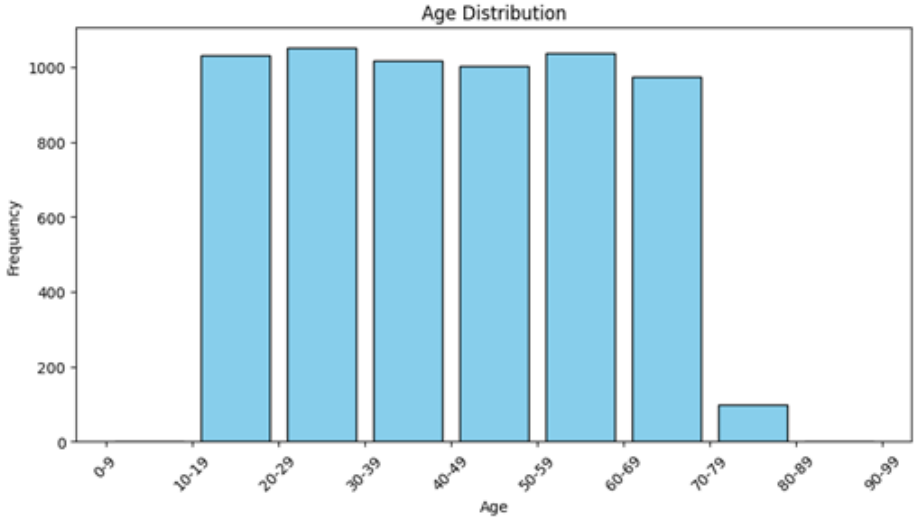
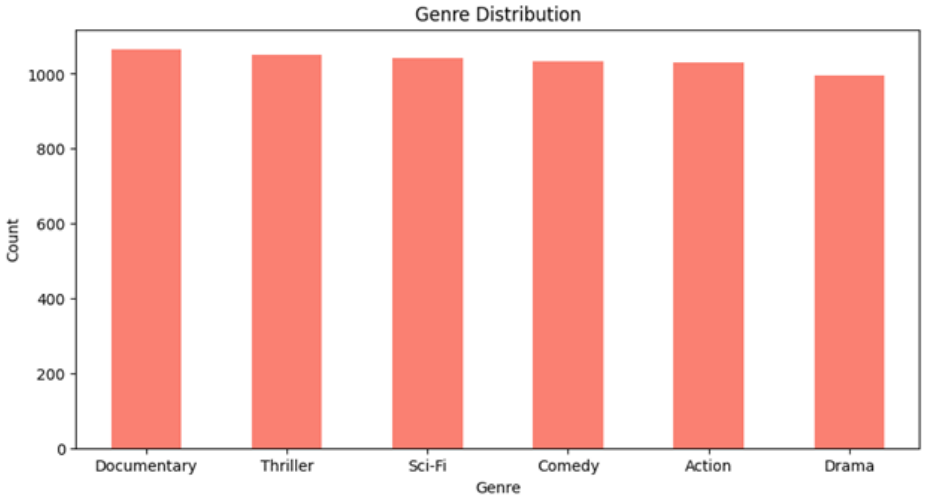
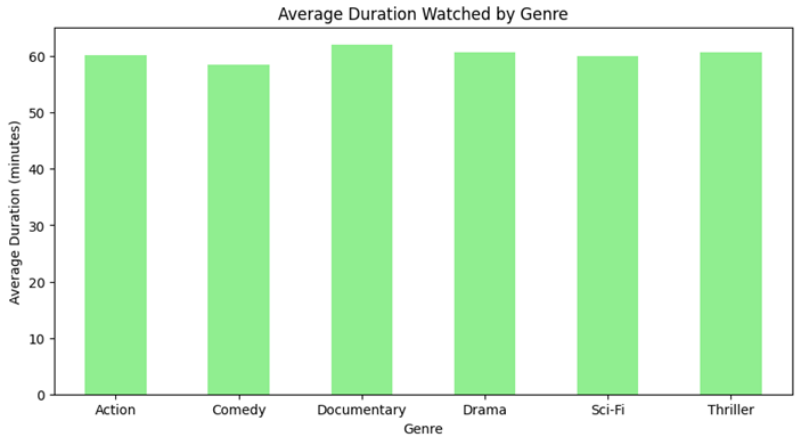


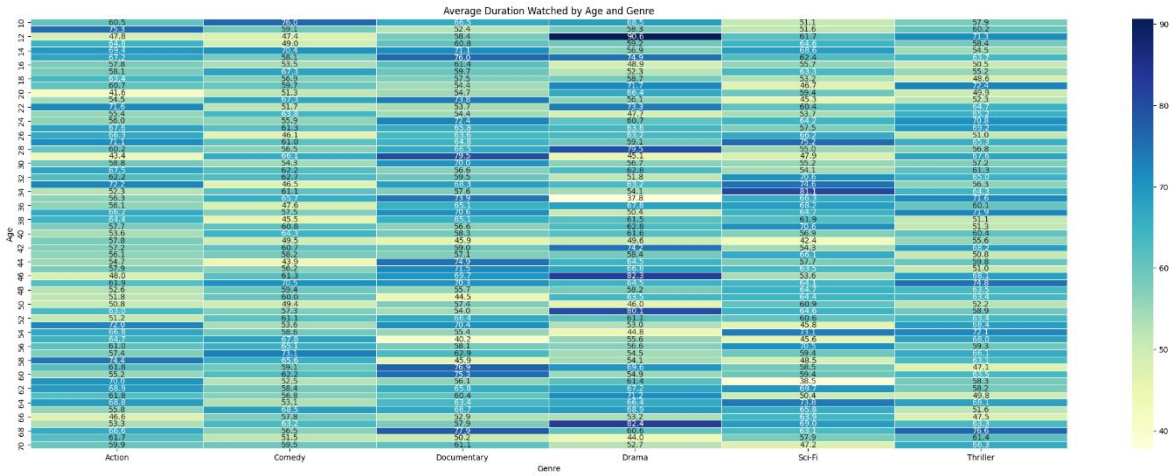
Figure 1. Age Distribution



Figures 2. Genre Distribution



Figures 3. Average Duration Watched by Genre



Figures 4. Average Duration Watched by Age and Genre

Discussions

Exploratory Data Analysis (EDA) on streaming apps reveals that documentaries, thrillers, science fiction and comedies dominate the industry with a wide user base and global variety. EDA allows a deep understanding of the evolution of these genres, their ranking, and their popularity by region and country. Discussions can focus on the strategies employed by streaming platforms to maintain a competitive advantage, such as proper content production, effective pricing, and thorough personalization. In addition, event-driven architectures open up opportunities for streaming applications to improve performance and user experience through continuous data processing and real-time intelligence. EDA results show that the use of technologies such as streaming analytics, geospatial analysis and complex event processing can provide valuable insights. Discussions can deepen understanding of the design and implementation of event-driven solutions, including event sources, processing, and consumers, as well as their influence on event feedback.

Data visualization techniques are emerging as an important element in presenting information effectively. EDA pointed out that streaming applications can use various types of plots and graphs to illustrate patterns and trends in their data, while identifying outliers and anomalies. Discussions can explore best practices and principles of data visualization, including the selection of appropriate visual forms, accurate encoding of data, and providing interactivity and context. In addition, this analysis also underscores the importance of globalization and locality strategies in tailoring content to the preferences and languages of viewers around the world. Overall, EDA on streaming apps opens the door to an in-depth understanding of the dynamics of the industry, helping platforms to improve their strategies, optimize user experience, and stay relevant in a competitive market.

Conclusions

Streaming application with EDA can help to measure and improve the streaming quality, such as the bandwidth, latency, resolution, format, etc., of the media content delivered to the users. Streaming application with EDA can also help to detect and resolve any issues or errors that may affect the streaming quality, such as network congestion, server load, device compatibility, etc. Streaming application with EDA can help to understand and predict the user behavior, such as the viewing duration, frequency, preference, rating, feedback, etc., of the media content consumed by the users. Streaming application with EDA can also help to forecast the future demand, popularity, retention, and churn of the users, and to segment and target the users based on their characteristics and interests. Streaming application with EDA can help to enhance and personalize the content recommendation, such as the suggestions, notifications, promotions, etc., of the media content offered to the users. Streaming application with EDA can also help to optimize the content production, acquisition, and distribution, and to create more engaging and interactive content experiences for the users. Future research development could focus on the integration of new technologies such as artificial intelligence (AI) and machine learning in streaming applications with exploratory data analysis (EDA). Research can explore how predictive models can be used to accurately project user behavior, identify content consumption patterns, and improve personalization of recommendations. Research can expand the understanding of how real-time analytics can help detect and address streaming quality issues more effectively. The shift towards a more holistic approach, encompassing content ecosystem understanding, user preferences, and infrastructure optimization, will provide a foundation for innovation in creating streaming experiences that are more dynamic, satisfying, and in line with user expectations.

References

Amier, R. H., & Setiawan, J. (2019). Visualization and Prediction of Film Award Nominations by Using of Visual Data Mining (VDM) and Exploratory Data Analysis (EDA) Method. *2019 5th International Conference on New Media Studies (CONMEDIA)*, 84–88.

Bouraqia, K., Sabir, E., Sadik, M., & Ladid, L. (2020). Quality of experience for streaming services: measurements, challenges and insights. *IEEE Access*, *8*, 13341–13361.

Chukwu, O. J. (2023). Interrogating the Online Internet-Based Broadcast Media Stations: Platforms, Implications and Emerged Paradigms. *Journal of Management and Science*, *13*(3), 74–81.

Criollo-C, S., Guerrero-Arias, A., Jaramillo-Alcázar, Á., & Luján-Mora, S. (2021). Mobile learning technologies for education: Benefits and pending issues. *Applied Sciences*, *11*(9), 4111.

Darapaneni, N., Bellarmine, C., Paduri, A. R., Entoori, S., Kumar, A., Vybhav, S. V, & Mondal, K. (2020). Movie success prediction using ml. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 869–874.

Falkowski-Gilski, P., & Uhl, T. (2020). Current trends in consumption of multimedia content using online streaming platforms: A user-centric survey. *Computer Science Review*, *37*, 100268.

Gu, K., Shang, R., Althoff, T., Wang, C., & Drucker, S. M. (2023). How Do Analysts Understand and Verify AI-Assisted Data Analyses? *ArXiv Preprint ArXiv:2309.10947*.

Johnson, C. (2020). The appisation of television: TV apps, discoverability and the software, device and platform ecologies of the internet era. *Critical Studies in Television*, *15*(2), 165–182.

Kesavan, S., Saravana Kumar, E., Kumar, A., & Vengatesan, K. (2021). An investigation on adaptive HTTP media streaming Quality-of-Experience (QoE) and agility using cloud media services. *International Journal of Computers and Applications*, *43*(5), 431–444.

Khan, S. A. R., Godil, D. I., Yu, Z., Abbas, F., & Shamim, M. A. (2022). Adoption of renewable energy sources, low-carbon initiatives, and advanced logistical infrastructure—an step toward integrated global progress. *Sustainable Development*, *30*(1), 275–288.

Kumar, A. S., & Joshna, K. (2021). *Student's Performance Analysis with EDA and Machine Learning Models*.

Ma, X. (2023). Introduction to Digital Content. In *Social Influence on Digital Content Contribution and Consumption: Theories, Empirical Analyses, and Practices* (pp. 3–14). Springer.

Mukhiya, S. K., & Ahmed, U. (2020). *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd.

Peng, J., Wu, W., Lockhart, B., Bian, S., Yan, J. N., Xu, L., Chi, Z., Rzeszotarski, J. M., & Wang, J. (2021). Dataprep. eda: Task-centric exploratory data analysis for statistical modeling in python. *Proceedings of the 2021 International Conference on Management of Data*, 2271–2280.

Reid, C., Keighrey, C., Murray, N., Dunbar, R., & Buckley, J. (2020). A novel mixed methods approach to synthesize EDA data with behavioral data to gain educational insight. *Sensors*, *20*(23), 6857.

Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12), 4727–4735.

Sanjana, N., Raj, S., & Sandhya, S. (2023). Real-time Event Streaming for Financial Enterprise System with Kafka. *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 1–6.

Srimahalap, W. (2020). *Exploring why Thai people listen to podcasts.*

Thakkar, J. C., & Vikas, S. C. (2022). A Pragmatic Approach on Adoption of EDA to Make Intelligent Business Decisions. *International Journal of Wireless Network Security*, *8*(2), 30-42p.

Velleman, P. F., & Hoaglin, D. C. (2023). *Exploratory data analysis.*

Yaqoob, A., Bi, T., & Muntean, G.-M. (2020). A survey on adaptive 360 video streaming: Solutions, challenges and opportunities. *IEEE Communications Surveys & Tutorials*, *22*(4), 2801–2838.

Zaveri, A. A., Mashood, R., Shehmir, S., Parveen, M., Sami, N., & Nazar, M. (2023). AIRA: An Intelligent Recommendation Agent Application for Movies. *Journal of Informatics and Web Engineering*, *2*(2), 72–89.