

Research Article

Comparative Study of SQL-Based Data Processing and Python-R Analytics for Advanced Data Visualization Using Tableau

Dr. Emilia Hoffmann ¹

¹Faculty of Computer Science and Data Processing Technologies, University of Stuttgart, Stuttgart, Germany



Received: 15 April 2026
Revised: 01 May 2026
Accepted: 30 May 2026
Published: 01 June 2026

Copyright: © 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the Creative Commons Attribution License 4.0 (CC-BY), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

Abstract

In recent years, data visualization has become a crucial component of data analytics pipelines, enabling stakeholders to interpret complex datasets and make informed decisions. Tools such as Tableau have significantly simplified visualization tasks, while backend data processing approaches like SQL-based systems and Python-R analytics frameworks provide the computational foundation for transforming raw data into meaningful insights. This paper presents a comprehensive comparative study of SQL-based data processing and Python-R analytics for advanced data visualization using Tableau. The study evaluates both approaches in terms of performance, scalability, flexibility, and suitability for visualization-centric workflows. It also explores how datasets such as Netflix movies and IMDb datasets can be processed using both paradigms and visualized effectively in Tableau dashboards. The findings highlight that SQL-based processing is highly efficient for structured and large-scale querying tasks, while Python-R analytics provide superior flexibility for statistical modeling, machine learning integration, and advanced preprocessing. Tableau acts as a unifying visualization layer that bridges both approaches. The study concludes that hybrid pipelines combining SQL and Python-R provide the most effective solution for modern data-driven environments.

Keywords: SQL-based data processing, Python analytics, R analytics, Tableau visualization, data visualization, comparative study, big data analytics, Netflix dataset analysis, IMDb dataset, dashboard development, business intelligence, data preprocessing, exploratory data analysis, interactive dashboards, data storytelling, hybrid analytics pipeline, data warehousing, SQL-on-Hadoop, statistical analysis, machine learning integration, visual analytics, data transformation, structured query language, predictive analytics, data science workflow .

INTRODUCTION

Data visualization has evolved into a central pillar of modern data science workflows, enabling organizations to interpret complex datasets and communicate insights effectively. With the exponential growth of data, traditional analysis methods are no longer sufficient, leading to the emergence of advanced data processing paradigms such as SQL-based systems and Python-R analytics environments. According to Patel (2021), Tableau has become one of the most widely used tools for transforming processed data into interactive dashboards that support decision-making processes.

The increasing demand for real-time analytics, particularly in domains like entertainment, healthcare, and e-commerce, has encouraged researchers to explore efficient backend processing techniques. SQL-based systems such as PostgreSQL and Hive offer structured query processing capabilities, while Python and R provide flexible analytical environments capable of statistical modeling and machine learning integration

(Batt et al., 2020). Meanwhile, visualization platforms like Tableau bridge the gap between raw data processing and human interpretation by enabling interactive and dynamic dashboards (Stirrup et al., 2016).

The purpose of this study is to compare SQL-based data processing with Python-R analytics in the context of advanced visualization using Tableau. The paper also examines how datasets such as Netflix content data and IMDb movie datasets can be used to demonstrate the effectiveness of both approaches in real-world scenarios (Singh, 2023; Bansal, 2023).

LITERATURE REVIEW

The concept of data visualization and analytics has been widely studied in academic and industrial research. Murray (2013) emphasized the importance of Tableau as a tool for fast and intuitive visual analytics. Similarly, Ryan (2018) highlighted the role of storytelling in data visualization, arguing that meaningful narratives improve decision-making efficiency.

SQL-based systems have been extensively used in large-scale data processing environments. Stonebraker and Cetintemel (2005) discussed the limitations of one-size-fits-all database systems and emphasized specialized architectures for efficient processing. Hive, introduced by Thusoo et al. (2010), provided a scalable solution for distributed data warehousing using Hadoop ecosystems. Floratou et al. (2014) further expanded this discussion by analyzing SQL-on-Hadoop systems and their role in distributed computing environments.

On the other hand, Python and R-based analytics frameworks have gained popularity due to their flexibility and rich ecosystem of statistical libraries. Chen and Lin (2014) highlighted the importance of deep learning and advanced analytics techniques in handling big data challenges. J. Hwang and Yoon (2021) emphasized Tableau's role in integrating analytical outputs from Python and R for visualization purposes.

Netflix and IMDb datasets have been widely used in academic research for visualization case studies. Roy (2019) demonstrated how movie datasets can be used to analyze trends in content production. Ojediran's Netflix Tableau dashboard further illustrates how streaming data can be visualized effectively using Tableau interfaces. Similarly, Gedam's GitHub project provides an end-to-end implementation of Netflix data analysis using Tableau dashboards.

METHODOLOGY

The methodology of this research is based on a comparative experimental framework where datasets are processed using two different paradigms: SQL-based processing and Python-R analytics. The processed outputs are then visualized using Tableau to evaluate differences in performance, flexibility, and visualization quality.

The datasets used in this study include Netflix movies and TV shows dataset (Bansal, 2023), IMDb dataset (Singh, 2023), and supplementary visualization datasets from public Tableau repositories. These datasets include structured fields such as title, genre, release year, rating, duration, and country of origin.

The workflow consists of three major stages:

First, data preprocessing is performed using SQL queries and Python-R scripts separately. SQL is used for filtering, aggregation, joins, and grouping operations, while Python-R is used for statistical cleaning, missing value imputation, and advanced transformations.

Second, processed datasets are exported into Tableau-compatible formats such as CSV or direct database connections.

Third, Tableau dashboards are created to visualize trends, distributions, and comparative insights across datasets.

According to Stirrup et al. (2016), Tableau's ability to connect directly with SQL databases and Python scripts makes it highly suitable for hybrid analytics workflows.

SQL-BASED DATA PROCESSING

SQL-based data processing is widely used due to its structured query language capabilities and optimized execution in relational databases. Systems such as PostgreSQL and Hive provide scalable environments for handling large datasets (PostgreSQL Documentation, 2023).

In this study, SQL is used to perform operations such as filtering Netflix content based on release year, grouping movies by genre, and calculating average ratings. SQL queries are highly efficient for such tasks because they leverage indexed data structures and optimized query execution plans.

For example, grouping Netflix titles by country and calculating content distribution can be achieved using simple GROUP BY queries. Similarly, IMDb datasets can be filtered to extract top-rated movies using ORDER BY and LIMIT clauses.

Stonebraker and Cetintemel (2005) argue that SQL systems are highly effective for structured data but lack flexibility in handling unstructured or semi-structured data. However, Floratou et al. (2014) highlight that modern SQL-on-Hadoop systems have significantly improved scalability for big data applications.

One of the key advantages of SQL-based processing is its ability to handle large-scale datasets with minimal computational overhead. However, it is less suitable for advanced statistical modeling and machine learning tasks compared to Python-R environments.

PYTHON-R ANALYTICS FOR DATA PROCESSING

Python and R provide powerful ecosystems for data analysis, statistical modeling, and machine learning integration. Python libraries such as Pandas, NumPy, and SciPy, along with R's statistical packages, enable advanced data transformations that are not easily achievable using SQL alone.

According to Chen and Lin (2014), Python-R frameworks are particularly effective in deep learning and predictive analytics tasks. In this study, Python is used to clean Netflix and IMDb datasets, handle missing values, normalize data distributions, and perform exploratory data analysis.

R is used for statistical summarization and visualization preprocessing, enabling deeper insights into dataset distributions and correlations. For instance, correlation analysis between movie ratings and release years can be easily performed using R statistical functions.

Unlike SQL, Python-R environments support complex data manipulation workflows such as time-series analysis, clustering, and regression modeling. This makes them highly suitable for advanced analytics pipelines before visualization in Tableau.

However, Python-R workflows are computationally heavier compared to SQL-based processing, especially when dealing with large datasets. Batt et al. (2020) emphasize that

despite this limitation, Python-R remains essential for advanced analytics due to its flexibility and extensibility.

TABLEAU FOR ADVANCED DATA VISUALIZATION

Tableau serves as the final visualization layer in this study, enabling interactive dashboards and dynamic data exploration. According to Patel (2021), Tableau simplifies complex datasets into intuitive visual representations that support decision-making processes.

Netflix and IMDb datasets are visualized using Tableau dashboards that include bar charts, heatmaps, line graphs, and geographical maps. These visualizations help in understanding content distribution, genre popularity, and rating trends over time.

DataScience RoadMap tutorials (2022, 2023) demonstrate how dashboards such as Netflix content distribution and movie analytics can be built within minutes using Tableau's drag-and-drop interface. These practical implementations highlight Tableau's ease of use and flexibility.

Ryan (2018) emphasizes that storytelling is a key aspect of Tableau dashboards, where data narratives guide users through insights rather than presenting raw numbers. This approach enhances user engagement and analytical comprehension.

Tableau's ability to connect with both SQL databases and Python/R scripts makes it a central tool in modern analytics ecosystems.

COMPARATIVE ANALYSIS

The comparison between SQL-based processing and Python-R analytics reveals distinct strengths and limitations for each approach.

SQL-based systems are highly efficient for structured querying, aggregation, and large-scale data retrieval. They are optimized for performance and are widely used in enterprise environments. However, they lack flexibility in handling advanced analytics tasks.

Python-R analytics provide superior capabilities in statistical modeling, machine learning, and complex data transformations. They are highly flexible but require more computational resources and development effort.

When integrated with Tableau, both approaches complement each other effectively. SQL provides clean and structured datasets, while Python-R adds analytical depth. Tableau then transforms these datasets into interactive visual dashboards.

According to Hwang and Yoon (2021), hybrid systems combining SQL and Python-R offer the most efficient workflow for modern data analytics pipelines.

CASE STUDY: NETFLIX AND IMDB DATA VISUALIZATION

Netflix and IMDb datasets are widely used for demonstrating data visualization techniques. Netflix data analysis focuses on content distribution, genre trends, and regional availability (Bansal, 2023). IMDb datasets provide insights into movie ratings, popularity, and production trends (Singh, 2023).

Using SQL, Netflix data is grouped by genre and country to analyze content distribution. Python is used to clean rating inconsistencies and perform statistical analysis. Tableau

dashboards then visualize these insights through interactive charts.

Gedam's GitHub project demonstrates a complete workflow for Netflix data analysis using Tableau, while Thesni's Tableau dashboard showcases content distribution visualization.

IMDb analysis similarly uses SQL for filtering and Python for statistical ranking of movies. These results are then visualized in Tableau storyboards, as demonstrated by Neware's IMDb workbook.

These case studies highlight the effectiveness of combining SQL, Python-R, and Tableau in real-world analytics scenarios.

DISCUSSION

The findings of this study indicate that no single approach is sufficient for end-to-end data analytics workflows. SQL-based systems excel in structured data processing, while Python-R provides analytical depth and flexibility. Tableau acts as the visualization layer that unifies both approaches.

Osuri (2016) highlights the growing importance of data-driven platforms like Netflix, which rely heavily on backend analytics systems. Similarly, Roy (2019) emphasizes the role of visualization in interpreting complex movie datasets.

The integration of SQL and Python-R with Tableau represents a hybrid analytics model that is increasingly adopted in industry environments. This model supports scalability, flexibility, and interpretability simultaneously.

CONCLUSION

This paper presented a comparative study of SQL-based data processing and Python-R analytics for advanced data visualization using Tableau. The study demonstrated that SQL is highly efficient for structured data operations, while Python-R provides advanced analytical capabilities. Tableau effectively bridges both approaches by enabling interactive and insightful visualizations.

The use of Netflix and IMDb datasets illustrated practical applications of both methodologies. The study concludes that hybrid analytics pipelines combining SQL, Python-R, and Tableau offer the most effective solution for modern data-driven environments.

Future research can explore the integration of machine learning models directly within Tableau dashboards and the use of real-time streaming data analytics for enhanced visualization systems.

REFERENCES

1. A. Ojediran, "Netflix Movies And Tv Shows Tableau Dashboard," Available: <https://oyasalofa.medium.com/netflix-movies-and-tvshows-tableau-dashboard-12fe06359b6b>
2. A. Patel. "Data Visualization Using Tableau." 2021.
3. A. Singh, "IMDb movies dataset," Retrieved October 20, 2023 from <https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>.
4. DataScience RoadMap. Create an Amazing Dashboard Using Tableau in 22 minutes. (Jun 3, 2023). Accessed: October 25, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=14eWTssItXQ&t=3s&ab_channel=DataScienceRoadMap.

5. DataScience RoadMap. Create Netflix dashboard with Tableau in 30 minutes. (September 9, 2022). Accessed: October 20, 2023. [Online Video]. Available: <https://www.youtube.com/watch?v=BTArwS4ljC4&abchannel=DataScienceRoadMap>
6. D. G. Murray, Tableau your data!: fast and easy visual analysis with tableau software : John Wiley & Sons, 2013.
7. F. Thesni, "Netflix Content Distribution," Available: <https://public.tableau.com/app/profile/faheena.thesni/viz/NetflixContentDistribution17024787172750/Practice?publish=yes>.
8. G. Gedam, "Netflix Data Analysis using Tableau," Available: <https://github.com/gulshang7/NetflixDataAnalysisusingTableau>.
9. G. Neware, "IMDB Movies Tableau Workbook ", Available: <https://public.tableau.com/app/profile/govindram.neware/viz/IMDBMoviestableauworkbook/IMDbStory>
10. J. Hwang and Y. Yoon, Data Analytics and Visualization in Quality Analysis Using Tableau : CRC Press, 2021.
11. J. Stirrup, A. Nandeshwar, A. Ohmann, and M. Floyd, Tableau: Creating Interactive Data Visualizations : Packt Publishing Ltd, 2016.
12. L. Osur, "Netflix and the development of the internet television network," 2016.
13. L. Patrao. "Needs and Benefits of Data Visualization ", Available: <https://www.edureka.co/blog/needs-and-benefits-of-data-visualization/>.
14. L. Ryan, Visual data storytelling with tableau: story points, telling compelling data narratives : Addison-Wesley Professional, 2018.
15. N. Roy, "Data Visualization on Movies Dataset using Tableau," International Journal of Engineering Trends and Technology (IJETT), Volume 67 Issue 2, February 2019.
16. P. N. Srivastava, "Netflix Dashboard:A Step-by-Step Guide to Creating Engaging Visuals with Tableau ", Available: <https://www.numpyninja.com/post/netflix-dashboard-a-step-by-stepguide-to-creating-engaging-visuals-with-tableau>
17. R. Sleeper, Practical tableau: 100 tips, tutorials, and strategies from a Tableau zen master : O'Reilly Media, Inc., 2018.
18. S. Bansal, Netflix Movies and TV Shows, Retrieved October 20, 2023 from <https://www.kaggle.com/datasets/shivamb/netflix-shows/data>
19. S. Batt, T. Grealis, O. Harmon, and P. Tomolonis, "Learning Tableau: A data visualization tool," The Journal of Economic Education, vol. 51, pp. 317–328, 2020.
20. T. Guillevin, Getting Started with Tableau 2019.2: Effective data visualization and business intelligence with the new features of Tableau 2019. 2 : Packt Publishing Ltd, 2019.
21. V. Chawin, "An End-to-End Netflix Case Study Using Tableau Desktop," Available: <https://medium.com/@v.chawin/an-end-to-endnetflix-case-study-using-tableau-desktop-c07c2c68ed71>.