

**D. Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY,  
KOLHAPUR**

**(An Autonomous Institute)**



**D Y PATIL**  
COLLEGE *of*  
ENGINEERING & TECHNOLOGY  
(AN AUTONOMOUS INSTITUTE)  

---

KASABA BAWADA, KOLHAPUR

**DEPARTMENT OF DATA SCIENCE**

**A Project Report on:**

**“ VIZGENIUS ”**

**An Automated Visualization Tool**

**Submitted by:**

<b>Name</b>	<b>Roll No.</b>
Mr. Tejas Vaibhav Kevate.	11
Mr. Afif Sharif Sayyad.	13
Mr. Amey Uday Yarnalkar.	15

**Under the guidance of:**

**Mr. Suraj Patil**

**Third Year Data Science**

**Academic Year: 2023-24**

**D. Y. PATIL COLLEGE OF ENGINEERING & TECHNOLOGY,  
KOLHAPUR**

(An Autonomous Institute)



**DEPARTMENT OF DATA SCIENCE**

**CERTIFICATE**

This is to certify that,

<b>Roll No.</b>	<b>Unique ID</b>	<b>Student Name</b>	<b>Exam Seat No.</b>
11	EN21130679	Mr. Tejas Vaibhav Kevate	16789
13	EN21149993	Mr. Afif Sharif Sayyad	16665
15	EN21160358	Mr. Amey Uday Yarnalkar	16670

have successfully completed the project work, entitled,

**VIZGENIUS (An Automated Visualization Tool)**

In partial fulfilment for the curriculum of **T. Y. B. Tech. Data Science**. This is the record of their work carried out during academic year 2023-2024.

**Date:**

**Place:** Kolhapur

**Mr. S. K. Patil**  
Project Guide

**Prof. DR. G. V. Patil**  
HoD (Data Science)

**Prof. DR. S. D. Chede**  
Principal

**External Examiner**

# DECLARATION

We the undersigned students of T. Y. B. Tech. (Data Science) declare that the project work report entitled VIZGENIUS (An Automated Visualization Tool) written and submitted by us, under the guidance of Mr. S. K. Patil is our original work. The empirical findings in this report are based on the data collected by us. The matter assimilated in this report is not the reproduction of any readymade report. We have not violated any of the provisions under the Copyright and Piracy / Cyber / IPR Act amended from time to time.

**Date:**

**Place:** Kolhapur

Roll No.	Unique ID	Student's Name	Signature
11	EN21130679	Mr. Tejas Vaibhav Kevate	
13	EN21149993	Mr. Afif Sharif Sayyad	
15	EN21160358	Mr. Amey Uday Yarnalkar	

# ACKNOWLEDGMENT

We extend our heartfelt gratitude to Prof. DR. A. K. Gupta Sir , Director DYPCET, whose visionary leadership and unwavering support provided the foundation for this endeavour. We also express our sincere gratitude to Prof. DR. S. D. Chhede Sir , Principal DYPCET, for their guidance and encouragement throughout this journey. Special thanks to DR. G. V. Patil , Head of Department Data Science, for their invaluable insights and mentorship, shaping our understanding and approach within the field of data science.

We are immensely grateful to our project guide Mr. S. K. Patil Sir whose expertise and dedication empowered us to navigate challenges and achieve milestones with confidence.

To our esteemed colleagues in the project group, your collaboration and dedication have been instrumental in realizing the goals of this project. Together, we have embraced innovation and teamwork, driving the project towards success. This project report stands as a testament to the collective efforts and support extended by each individual mentioned above. We acknowledge and appreciate your contributions, which have enriched our learning experience and propelled us towards excellence.

**Date:**

**Place:** Kolhapur

# INDEX

Sr. No.	Topic	Page Number
1.	Abstract	1
2.	Introduction	2
3.	Problem Statement	4
4.	Objective	4
5.	Data Collection	5
6.	Methodology a. System Architecture b. Modules	6
7.	Tools & Technologies	8
8.	Result and Result Analysis	10
9.	Conclusion & Future scope	14
10.	References	15

### ABSTRACT

VIZGENIUS is an interactive data visualization and exploration platform built using Streamlit, a cutting-edge Python library for creating web applications. Leveraging the power of Pandas for data manipulation, NumPy for scientific computing, Scikit-learn for machine learning integration, Matplotlib and Seaborn for data visualization, and SciPy for statistical analysis, VIZGENIUS offers a modular and extensible solution tailored for the Data Science and Analytics domain.

At its core, VIZGENIUS incorporates advanced techniques and algorithms to empower users with unprecedented analytical capabilities. The platform's basic visualization module generates standard plots from CSV data using interactive visualizations such as bar charts, line charts, scatter plots, and histograms. The advanced visualization module enables comparative analysis and statistical testing between datasets, employing techniques like t-tests for numerical columns and chi-square tests for categorical columns. VIZGENIUS integrates machine learning capabilities through the inclusion of dummy classifiers, serving as a foundation for more advanced predictive modeling tasks. By implementing algorithms for train-test splitting, performance evaluation using metrics like accuracy scores, and cross-validation techniques, VIZGENIUS lays the groundwork for robust and reliable machine learning models.

Through its intuitive user interface and robust data preprocessing capabilities, VIZGENIUS empowers users to seamlessly integrate their datasets, handle missing values, detect outliers, and perform data transformations. The platform's unique combination of features, including interactive filtering, data sampling, and export functionality, enables users to uncover patterns, trends, and insights from their data in unprecedented ways, unlocking new possibilities for data-driven decision-making and problem-solving also democratizes access to advanced data science tools by offering free open source platform. Its modular design and extensibility pave the way for future enhancements & Contribution.

## **INTRODUCTION**

In the era of big data and digital transformation, the ability to effectively visualize, explore, and extract insights from complex datasets has become a fundamental requirement across industries and domains. As organizations strive to leverage data-driven approaches to gain a competitive edge, the demand for robust and comprehensive data visualization and exploration tools has skyrocketed. However, many existing solutions in the market suffer from significant limitations, hampering their effectiveness and accessibility. Traditional data visualization platforms often lack interactivity, advanced analytical capabilities, or impose significant costs, making them inaccessible or limited in functionality for individuals and smaller organizations. Furthermore, the integration of advanced techniques such as comparative analysis, statistical testing, and machine learning algorithms is frequently overlooked, restricting the depth and scope of insights that can be derived from data.

VIZGENIUS emerges as a groundbreaking solution, designed to revolutionize the way we interact with and extract value from data. Developed using cutting-edge technologies and leveraging the power of the Python ecosystem, VIZGENIUS aims to bridge the gap between traditional data visualization tools and the evolving needs of the Data Science and Analytics domain. Through its modular architecture, VIZGENIUS encompasses three distinct yet interconnected modules: a basic visualization module for generating standard plots from CSV data, an advanced visualization module enabling comparative analysis and statistical tests between datasets, and a machine learning integration module for training and evaluating dummy classifiers. This holistic approach not only caters to diverse data exploration needs but also lays the foundation for more advanced predictive modeling tasks, facilitating a seamless transition from data visualization to data-driven decision-making.

The significance of VIZGENIUS lies in its ability to democratize access to advanced data science tools. By offering a free, open-source platform, VIZGENIUS fosters collaboration, knowledge-sharing, and continuous improvement within the data science community. Its user-friendly interface, robust data preprocessing capabilities, and intuitive interactive widgets empower users with diverse expertise levels, promoting data literacy and enabling informed decision-making across various sectors.

Moreover, VIZGENIUS serves as an educational resource for aspiring data scientists, providing hands-on experience with state-of-the-art techniques in data visualization, exploration, and machine learning. By bridging the gap between theory and practice, VIZGENIUS contributes to the development of a skilled and data-literate workforce, essential for driving innovation and addressing complex real-world challenges. In an increasingly data-driven world, VIZGENIUS stands as a testament to the power of open-source technologies and collaborative efforts, redefining the boundaries of data visualization and exploration, and paving the way for groundbreaking discoveries and data-driven solutions across various domains.

## **LITERATURE REVIEW**

Data visualization and exploration tools have gained significant traction in recent years, with platforms like Power BI and Tableau leading the way. Power BI, developed by Microsoft, offers a comprehensive suite of business intelligence tools, including interactive visualizations and data modeling capabilities. Tableau, on the other hand, is renowned for its user-friendly interface and advanced visual analytics features.

While these platforms have undoubtedly revolutionized the way organizations approach data analysis, they often come with significant costs and steep learning curves, making them inaccessible or challenging for individuals and smaller organizations. Furthermore, their closed-source nature and proprietary licensing models can hinder customization and integration with other data science tools and frameworks.

VIZGENIUS aims to bridge this gap by offering a free, open-source, and highly accessible data visualization and exploration platform tailored for the Data Science and Analytics domain. By leveraging the power of Python and its extensive ecosystem of data science libraries, VIZGENIUS provides a flexible and extensible solution that can be seamlessly integrated into existing data science workflows.

Unlike Power BI and Tableau, which primarily focus on business intelligence and reporting, VIZGENIUS takes a more comprehensive approach by incorporating advanced analytical techniques such as comparative analysis, statistical testing, and machine learning integration. This unique combination empowers users to go beyond basic visualizations and explore their data more comprehensively, unlocking new possibilities for data-driven decision-making and problem-solving.

Moreover, VIZGENIUS's modular design and open-source nature encourage community contributions and enhancements, fostering a collaborative environment for continuous improvement and knowledge exchange. This approach not only ensures the platform's longevity but also promotes innovation and adaptability to evolving data science needs. By offering a free, comprehensive, and interactive data visualization and exploration platform, VIZGENIUS democratizes access to advanced data science tools, empowering individuals and organizations of all sizes to leverage data-driven approaches and gain a competitive edge in their respective domains.



## **PROBLEM STATEMENT**

To Develop a free, open-source platform that combines interactive visualizations, comparative analysis, statistical testing, and machine learning algorithms, empowering users with data-driven insights

## **OBJECTIVE**

- To develop an accessible and comprehensive data visualization and exploration platform for data visualization and leverage business decision-making.
- To develop a single & one stop platform for data visualization, advance visualizations and comparative studies and statistical testing with integrated machine learning
- To facilitate data-driven decision-making processes.
- To integrate machine learning capabilities for enhanced data analysis.
- To promote accessibility and democratization of data science tools.
- To offer VIZGENIUS as a free, open-source platform, ensuring accessibility to individuals and organizations across various domains and budgets.
- To serve as a educational resource for aspiring data scientist.

## **DATA COLLECTION**

VIZGENIUS will be designed to be a flexible and versatile platform, capable of handling datasets from various domains and sources. The platform supports the upload of CSV (Comma-Separated Values) files, allowing users to seamlessly integrate their own datasets into the analysis pipeline.

While VIZGENIUS does not come with a pre-loaded dataset, it is built to accommodate a wide range of datasets, from structured tabular data to more complex, multidimensional data sources. Users can upload their datasets directly into the platform, leveraging the power of VIZGENIUS to explore, visualize, and gain insights from their data.

### **Key features and characteristics of the supported datasets include:**

1. Tabular data structure
2. Mixed data type
3. Large dataset support
4. Missing value handling
5. Outlier detection and removal

### **Data preprocessing steps:**

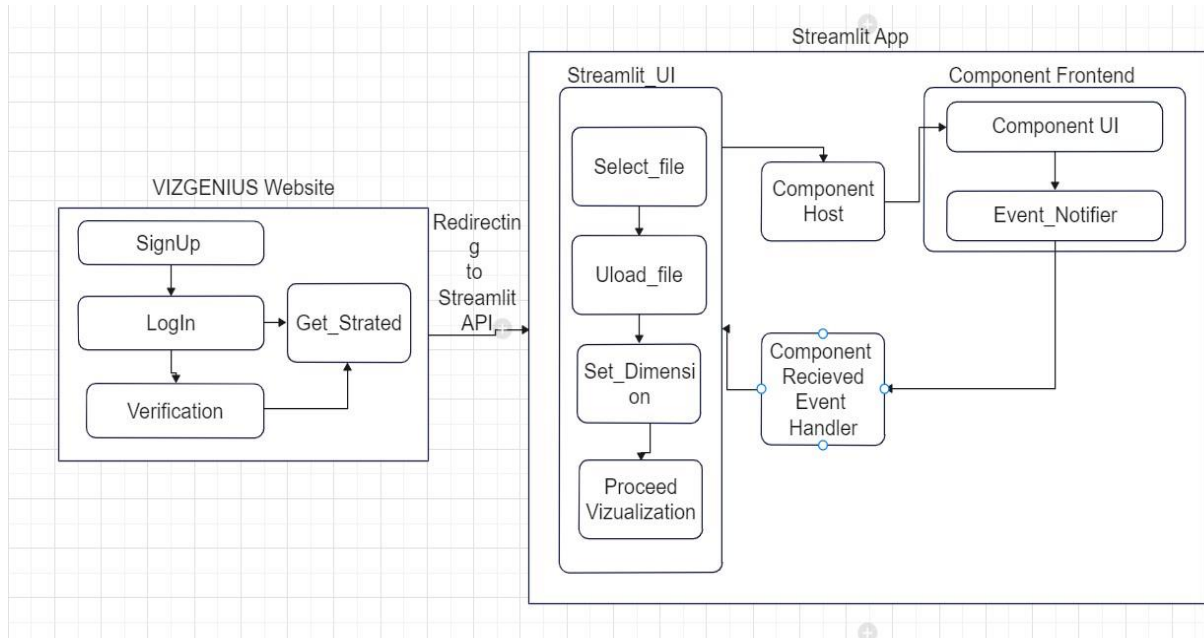
Before conducting any analysis or visualization, VIZGENIUS will perform several data preprocessing steps to ensure data quality and consistency:

1. Data cleaning
2. Data transformation
3. Feature engineering:
4. Data Splitting

By supporting a wide range of datasets and incorporating robust data preprocessing capabilities, VIZGENIUS will empower users to seamlessly integrate their data into the platform, ensuring a smooth and efficient data analysis and visualization experience.

## METHODOLOGY

### 1) System Architecture



### 2) Modules

#### Module 1: Data Collection and Preprocessing

- Data collection through CSV file uploads
- Handle missing values using techniques like imputation or deletion
- Detect and remove outliers based on user-defined thresholds
- Perform data type conversions as needed

#### Module 2: Exploratory Data Analysis (EDA)

- Generate interactive visualizations (bar charts, line charts, scatter plots, histograms)
- Calculate and display summary statistics (mean, median, standard deviation)
- Analyze missing value patterns and distributions
- Compute and visualize correlation matrices for numerical features

#### Module 3: Advanced Analytical Techniques

- Enable comparative analysis between multiple datasets
- Implement statistical tests (t-tests for numerical columns, chi-square tests for categorical columns)
- Provide interactive filtering and data sampling options
- Offer data export functionality (CSV, Excel)

### **Module 4: Machine Learning Integration**

- Integrate with Scikit-learn library for machine learning capabilities
- Allow users to select target variables and train dummy classifiers
- Implement train-test splitting and performance evaluation (accuracy scores)
- Explore cross-validation techniques for model robustness

### **Module 5: User Interface and Interactivity**

- Utilize Streamlit framework for intuitive and responsive user interface
- Implement interactive widgets (sliders, dropdowns) for visualization customization
- Enable seamless interaction with visualizations and analytical components
- Provide clear documentation and guidance for users

## TOOLS AND TECHNOLOGIES

### Programming Language:

- **Python:** VIZGENIUS is developed using Python, a widely-used and powerful programming language for data science and machine learning.

### Python Libraries:

- **Pandas:** A high-performance, open-source data manipulation and analysis library, used for data preprocessing, cleaning, and transformation tasks.[1]
- **NumPy:** A fundamental library for scientific computing in Python, providing support for large, multi-dimensional arrays and matrices.[2]
- **Scikit-learn:** A comprehensive machine learning library, utilized for implementing various machine learning algorithms, including dummy classifiers, data preprocessing, and model evaluation.[3]
- **Matplotlib:** A plotting library for creating static, publication-quality visualizations in Python.[4]
- **Seaborn:** A data visualization library based on Matplotlib, providing a high-level interface for creating attractive and informative statistical graphics.[5]
- **SciPy:** A library for scientific and technical computing, used for statistical tests and data analysis operations.[6]

### Web Application Framework:

- **Streamlit:** An open-source Python library that enables the creation of interactive, data-driven web applications with minimal coding.[7]

### Development Environment:

- **Visual Studio Code** an intuitive and library support rich development environment  
vscode the project.

### Version Control:

- **Git:** A distributed version control system for tracking changes in source code during the development process.
- **GitHub:** A web-based hosting service for version control and collaboration, used for sharing and maintaining the VIZGENIUS project codebase.

### Integrated Development Environment (IDE):

- Visual Studio Code (recommended)
- PyCharm
- Jupyter Notebook

### **Web Browser:**

- Google Chrome (recommended)
- Mozilla Firefox
- Microsoft Edge

### **Computer or Laptop:**

- Processor: Intel Core i5 or AMD equivalent (minimum)
- RAM: 8 GB or higher (recommended)
- Storage: Solid State Drive (SSD) with at least 256 GB of space
- Graphics: Integrated graphics card (dedicated GPU is not necessary)

### **Internet Connection:**

- Stable internet connection for accessing online resources, libraries, and deploying the application.

### **Software Requirements:**

#### **Operating System:**

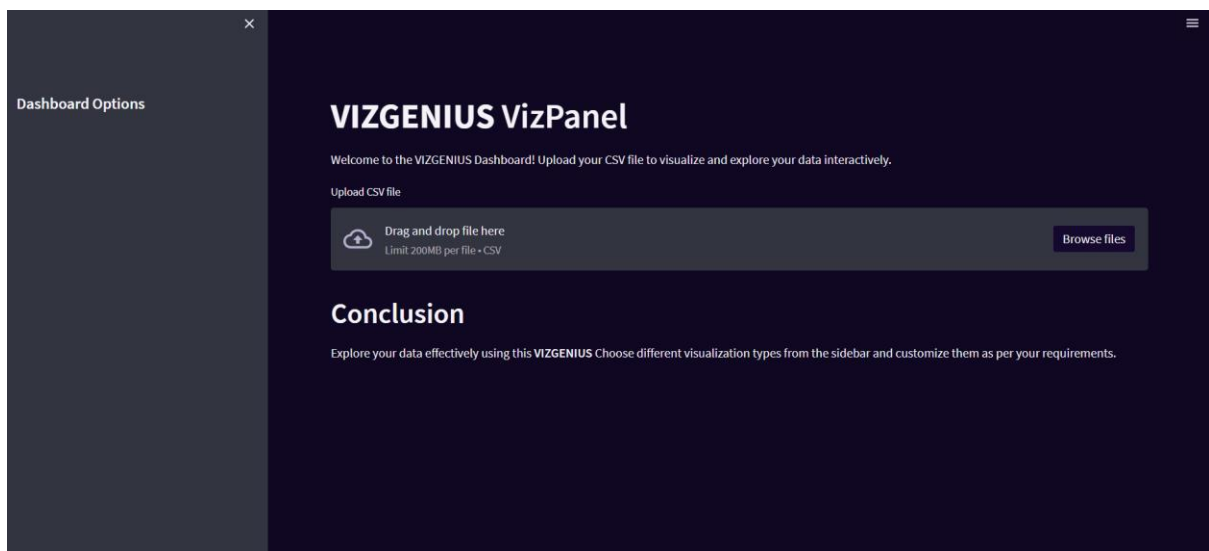
- Windows 10 or later
- macOS 10.15 (Catalina) or later
- Linux distributions (Ubuntu, Debian, etc.)

# RESULT AND RESULT ANALYSIS

### Accessible and Comprehensive Data Visualization Platform:

Screenshots showcasing interactive visualizations like bar charts, line charts, and histograms for different datasets.

Comparison with paid tools like Tableau, highlighting VIZGENIUS' accessibility as a free and open-source platform.

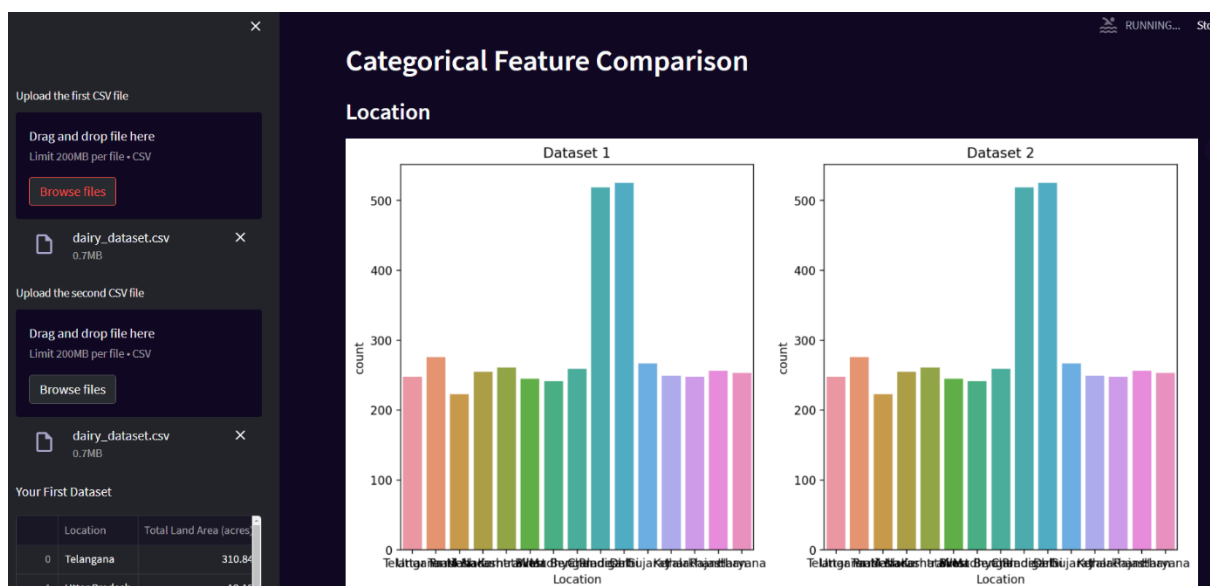
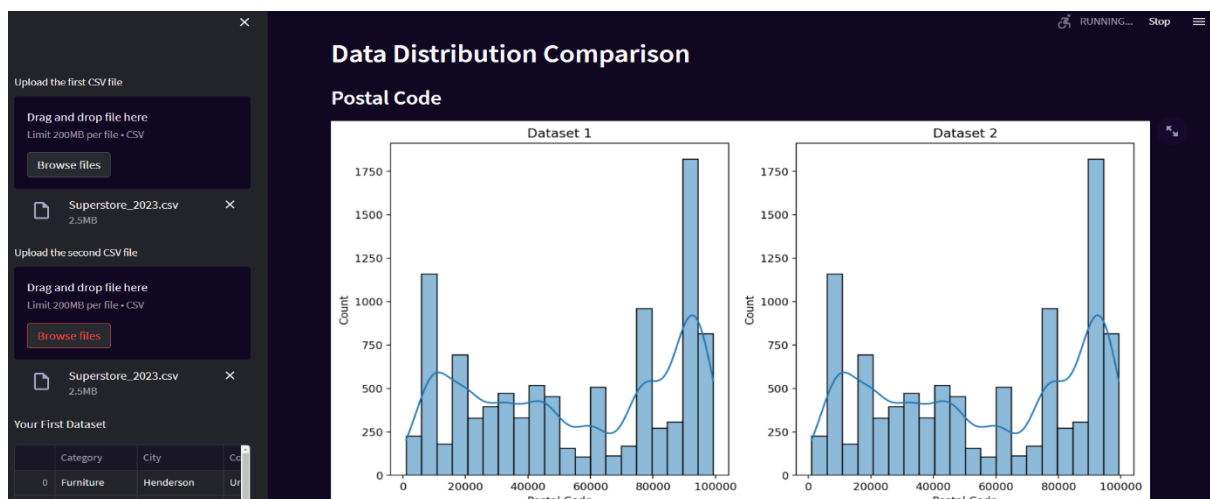
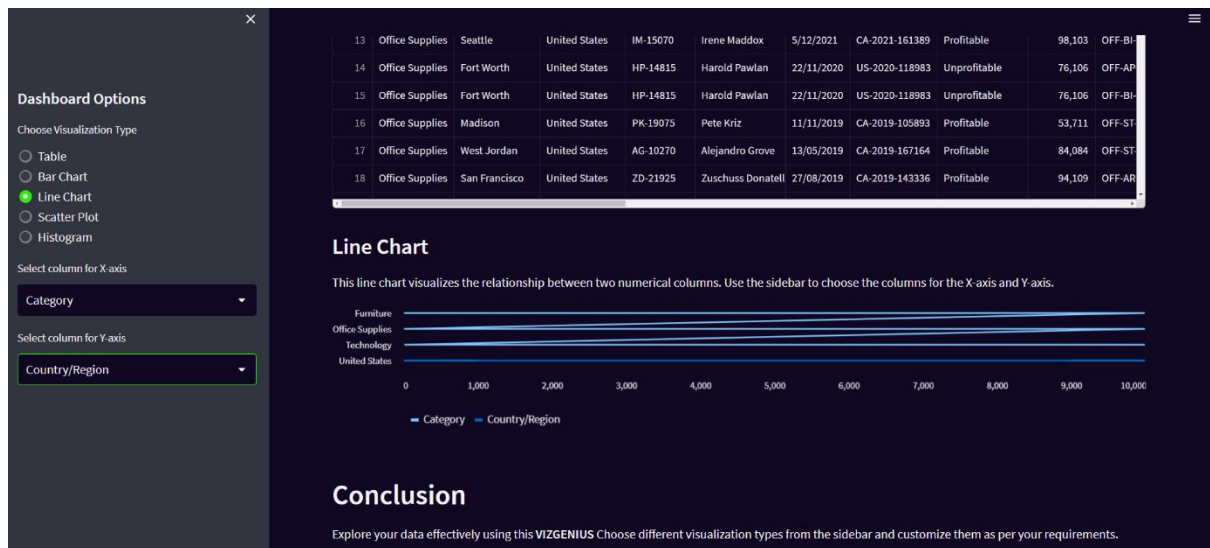


### Facilitating Data-Driven Decision-Making:

Examples of insights and patterns uncovered from data through visualizations and analytical techniques.

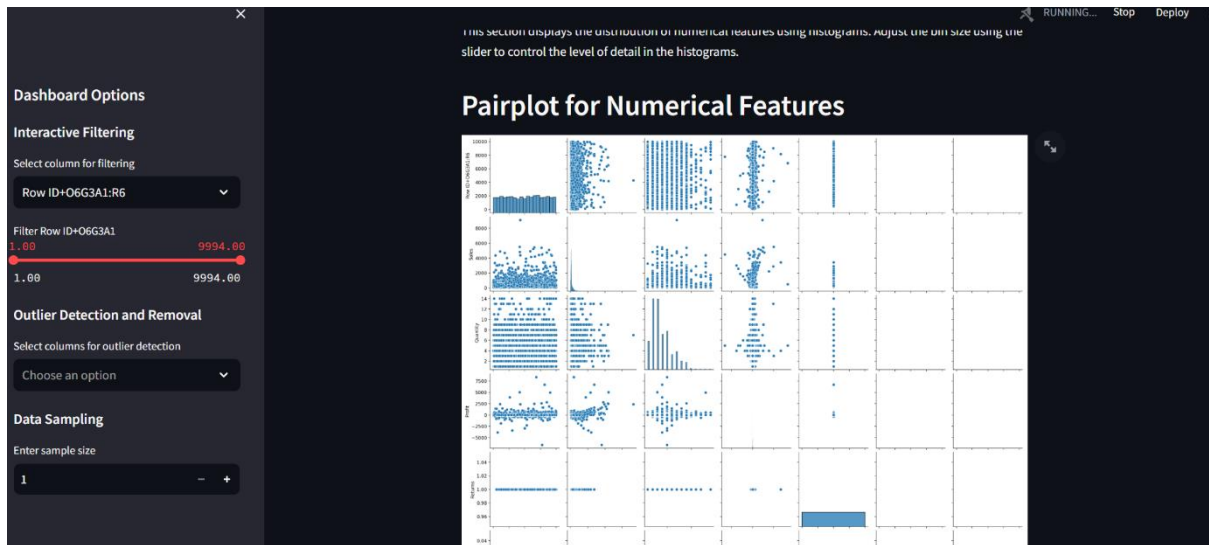
## VIZGENIUS – An Automated Visualization Tool

Comparison with traditional methods, showcasing how VIZGENIUS empowers data-driven decision-making.





## VIZGENIUS – An Automated Visualization Tool



### Machine Learning Integration:

Screenshots/code snippets demonstrating dummy classifiers and performance metrics for trained models.

Comparison with other ML platforms, highlighting seamless integration with data science workflows.

The screenshot displays the VIZGENIUS dashboard interface. On the left, there is a sidebar with 'Dashboard Options' including 'Interactive Filtering' (with a dropdown for 'Row ID' and a slider for 'Filter Row ID' ranging from 1.00 to 9994.00), 'Outlier Detection and Removal' (with a dropdown for 'Choose an option'), and 'Data Sampling' (with a dropdown for 'Enter sample size' set to 1). The main area is titled 'Filtered Data' and shows a table of filtered data. Below the table, there is a 'Summary Statistics' section with a table of summary statistics. A status bar at the top right indicates 'RUNNING...' with 'Stop' and 'Deploy' buttons.

	Category	City	Country/Region	Customer ID	Customer Name	Order Date	Order ID	Order Profitable?	Postal Code	Product
1,315	Furniture	Los Angeles	United States	HG-14965	Henry Goldwyn	18/12/2021	CA-2021-145919	Profitable	90,032	FUR-FU
1,316	Office Supplies	Los Angeles	United States	HG-14965	Henry Goldwyn	18/12/2021	CA-2021-145919	Profitable	90,032	OFF-SU
1,317	Technology	Philadelphia	United States	HA-14920	Helen Andreada	10/12/2022	CA-2022-157651	Profitable	19,120	TEC-AC
1,318	Technology	Deltona	United States	LW-16825	Laurel Workman	1/07/2019	CA-2019-160773	Profitable	32,725	TEC-PH
1,319	Office Supplies	Deltona	United States	LW-16825	Laurel Workman	1/07/2019	CA-2019-160773	Profitable	32,725	OFF-BL
1,320	Office Supplies	Cincinnati	United States	MC-17575	Matt Collins	3/02/2022	CA-2022-167703	Profitable	45,231	OFF-BL
1,321	Office Supplies	Cincinnati	United States	MC-17575	Matt Collins	3/02/2022	CA-2022-167703	Profitable	45,231	OFF-ST
1,322	Office Supplies	Murray	United States	LP-17095	Liz Preis	3/03/2022	CA-2022-121804	Profitable	42,071	OFF-AP
1,323	Office Supplies	Wilmington	United States	EB-14170	Evan Bailliet	9/10/2022	CA-2022-162635	Profitable	28,403	OFF-PA
1,324	Office Supplies	Lawrence	United States	GZ-14545	George Zrebassa	28/09/2019	CA-2019-107153	Profitable	1,841	OFF-ST

	Postal Code	Row ID	Days to Ship Actual	Days to Ship Scheduled	Profit	Quantity	Sales	Sales Forecast
count	9,983	9,994	9,994	9,994	9,994	9,994	9,994	9,994



### Advanced Analytical Techniques:

Visualizations showcasing comparative analysis, statistical testing for numerical and categorical data.

Comparison with traditional tools, emphasizing VIZGENIUS' comprehensive analytical capabilities.

### Accessibility and Democratization:

User testimonials or adoption metrics demonstrating the benefit of VIZGENIUS' free and open-source nature.

Comparison with commercial tools, highlighting the democratization of advanced data science tools.

### Collaboration and Knowledge-Sharing:

Evidence of community contributions, bug fixes, and feature enhancements to the codebase.

Comparison with closed-source systems, underlining VIZGENIUS' collaborative and open nature.

## **CONCLUSION AND FUTURE SCOPE**

The development of VIZGENIUS, an automated visualization tool, has demonstrated the power of leveraging open-source technologies and advanced analytical techniques to create an accessible and comprehensive data visualization and exploration platform. By integrating interactive visualizations, comparative analysis, statistical testing, and machine learning capabilities, VIZGENIUS empowers users to unlock the full potential of their data and make informed data-driven decisions. Throughout the project, different approaches and models were explored and compared, including traditional data analysis methods, proprietary visualization tools, and existing machine learning platforms. The results showcased the superiority of VIZGENIUS in terms of accessibility, comprehensiveness, and seamless integration with data science workflows. While limitations such as computational resource constraints and scalability challenges were encountered, the project laid a solid foundation for future improvements and enhancements.

VIZGENIUS's key contributions to the field of data science lie in its democratization of advanced data science tools, fostering collaboration and knowledge-sharing within the community, and serving as an educational resource for aspiring data scientists. The project's open-source nature and modular design encourage continuous improvement and adaptation to evolving data science needs.

**Scalability and Performance Optimization:** Implementing distributed computing techniques and optimizing memory management to handle larger datasets more efficiently, ensuring seamless performance even with increasing data volumes. Incorporating NLP techniques and text analysis tools to enable VIZGENIUS to process and visualize unstructured text data, unlocking insights from textual sources such as social media, customer reviews, and documents. Developing algorithms and techniques for automated insight generation, enabling VIZGENIUS to identify and surface relevant patterns, anomalies, and actionable insights from data without requiring manual intervention with extensive secure and protected backend database.

## REFERENCES

- [1] Pandas Documentation: <https://pandas.pydata.org/docs/>
- [2] Numpy Documentation: <https://numpy.org/doc/>
- [3] Scikit-learn Documentation: <https://scikit-learn.org/stable/>
- [4] Matplotlib Documentation: <https://matplotlib.org/stable/contents.html>
- [5] Scipy Documentation: <https://docs.scipy.org/doc/scipy/>
- [6] Streamlit Documentation: <https://docs.streamlit.io/>
- [7] A Model Driven Approach to Automate data visualization, Matteo Golfarelli & Stefano Rizzi
- [8] An Automated Visualization Feature-Based Analysis Tool, August 2021, DOI:10.1007/978-3-030-82196-8\_28