



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

DEPARTMENT OF DATA SCIENCE

●PROJECT PRESENTATION●

Academic Year: 2023-24

Third Year B.Tech (Data Science)

Presented By:

Mr. Tejas Vaibhav Kevate

Mr. Afif Sharif Sayyad

Mr. Amey Uday Yarnalkar

Under the Guidance of:

Mr. S. K. Patil

VIZGENIUS

AN AUTOMATED VIZUALISATION TOOL



| 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 inclusive platform for visualization
- One stop platform for advanced visualizations, comparative studies, statistical testing, and machine learning integration
- Incorporate machine learning capabilities to enhance data analysis and insights generation.
- AI Integration Tool Accessibility Open Access Educational Support

Requirement Specification



Flexible Platform: Handles diverse datasets from various domains.

Seamless Integration: Supports CSV file upload for user datasets.

Data Collection



Key Features

- Tabular data structure
- Mixed data type
- Large dataset support
- Missing value handling
- Outlier detection and removal

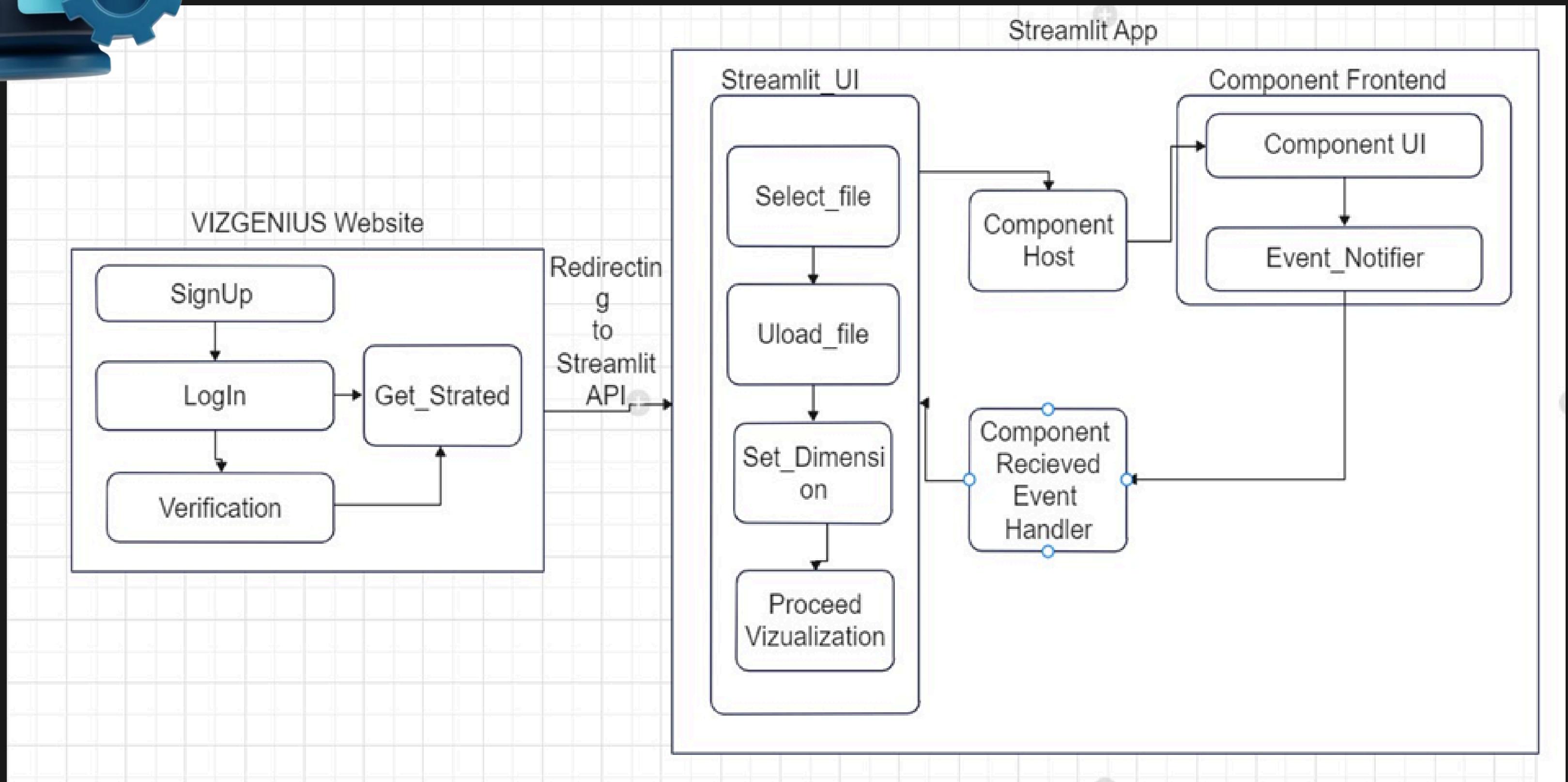


Data Preprocessing

- Cleaning
- Transformation
- Feature engineering
- Data splitting



SYSTEM ARCHITECTURE



Advanced Analytical Techniques

- Enabling comparative analysis between multiple datasets
- Implementing statistical tests (t-tests for numerical columns, chi-square tests for categorical columns)
- Providing interactive filtering and data sampling options
- Offering data export functionality (CSV, Excel)
- Enabling seamless interaction with visualizations and analytical components

Data Collection and Preprocessing

- Data collection through CSV file uploads
- Handling missing values using techniques like imputation or deletion
- Detecting and removing outliers based on user-defined thresholds
- Performing data type conversions as needed
- Utilizing Streamlit framework for intuitive and responsive user interface

Modules

Exploratory Data Analysis (EDA)

- Generating interactive visualizations (bar charts, line charts, scatter plots, histograms)
- Calculating and displaying summary statistics (mean, median, standard deviation)
- Analyzing missing value patterns and distributions
- Computing and visualizing correlation matrices for numerical features

Machine Learning Integration

- Integrating with Scikit-learn library for machine learning capabilities
- Allowing users to select target variables and train dummy classifiers
- Implementing train-test splitting and performance evaluation (accuracy scores)
- Exploring cross-validation techniques for model robustness

S/W & H/W REQUIREMENTS

- **Programming Language:** Python
- **Libraries:** Numpy, Pandas, Matplotlib, Scipy, Seaborn, Scikit-learn
- **Web- Framework:** Streamlit
- **Development Environment:** Visual Studio code, pycharm, Jupyter Notebook
- **Version control:** Git, Git-Hub
- **Web-Browser:** Chrome, Firefox, Edge
- **System:** Intel i5 & above core or AMD, 8 GB RAM, SSD 256GB+ , Integrated Graphics
- Stable Internet Connection

PROCESS

01

Select CSV File



02

Upload CSv File



03

Visualizations



04

Take Insights



RESULT ANALYSIS



OUR SERVICES

BASIC

PRO

ENTERPRISE





The screenshot shows the VizGenius homepage. At the top left is the VizGenius logo. The top right features a navigation bar with links for About, Vision, Programs, Blog, Log In, and Get Started. Below the navigation is a header section with the text "TRANSFORMING VISION INTO REALITY WITH AN AUTOMATED VISUALIZATION TOOL". The main title "VIZGENIUS" is displayed in large, bold, cyan letters. Below the title is a stylized infinity symbol and a network of blue dots. Three callout boxes are present: "Use Data to Get a 360-Degree View of Your Business", "Drive your World with Valuable & Insightful Data", and "Uncover Insights & Empower Decisions". A "Learn More" button and a red speech bubble icon are also visible.

VIZGENIUS

ABOUT HELP CONTACT

VizGenius

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.

Web View

Dashboard Options

VIZGENIUS VizPanel

Welcome to the VIZGENIUS Dashboard! Upload your CSV file to visualize and explore your data interactively.

Upload CSV file

Drag and drop file here
Limit 200MB per file • CSV

Browse files

Conclusion

Explore your data effectively using this VIZGENIUS Choose different visualization types from the sidebar and customize them as per your requirements.

Dashboard Options

Choose Visualization Type

- Table
- Bar Chart
- Line Chart
- Scatter Plot
- Histogram

Select column for X-axis

Ship Mode

Select column for Y-axis

Customer Name

Line Chart

This line chart visualizes the relationship between two numerical columns. Use the sidebar to choose the columns for the X-axis and Y-axis.



Aaron Bergman
Aaron Hawkins
Aaron Smayling
Adam Bellavance
Adam Hart
Adam Shillingsburg
Adrian Barton
Adrian Hane
Adrian Shami
Aimee Bixby
Alan Barnes
Alan Dominguez
Alan Haines
Alan Hwang
Alan Schoenberger
Alan Shoney
Alejandro Ballent...
Alejandro Grove
Alejandro Savely
Aleksandra Gann...
Alex Avila
Alex Grayson
Alex Russell
Alice McCarthy
Allen Arnold
Allen Goldenen
Allen Rosenblatt
Alyssa Crouse
Alyssa Tate
Amy Cox
Amy Hunt
Andrew Allen
Andrew Gjertsen
Andrew Roberts
Andy Reiter
Andy Yotov
Anemone Ratner

Deploy :

- Rerun R
- Settings
- Print
- Record a screencast
- About
- Developer options
- Clear cache C

Basic

Facilitating Data-Driven Decision-Making: Examples of insights and patterns uncovered from data through visualizations and analytical techniques. Comparison with traditional methods, showcasing how VIZGENIUS empowers data-driven decision-making.

Dashboard Options

Your VIZGENIUS Data Explorer

Welcome to the VIZGENIUS Data Explorer! Upload your CSV file to visualize and explore your data interactively.

Upload CSV file

Drag and drop file here
Limit 200MB per file - CSV

Browse files

Conclusion

Explore and analyze your data effectively using the VIZGENIUS Interactive Data Explorer! ** Upload your CSV file***; apply interactive filtering*, visualize data distributions, detect outliers, export data, and even train a dummy classifier. Enjoy exploring your data with ##VIZGENIUS!

Pro

Dashboard Options

Interactive Filtering

Select column for filtering

Row ID+O6G3A1:R6

Filter Row ID+O6G3A1
1.00 9994.00

Outlier Detection and Removal

Select columns for outlier detection

Choose an option

Data Sampling

Enter sample size

1

Dummy Classifier

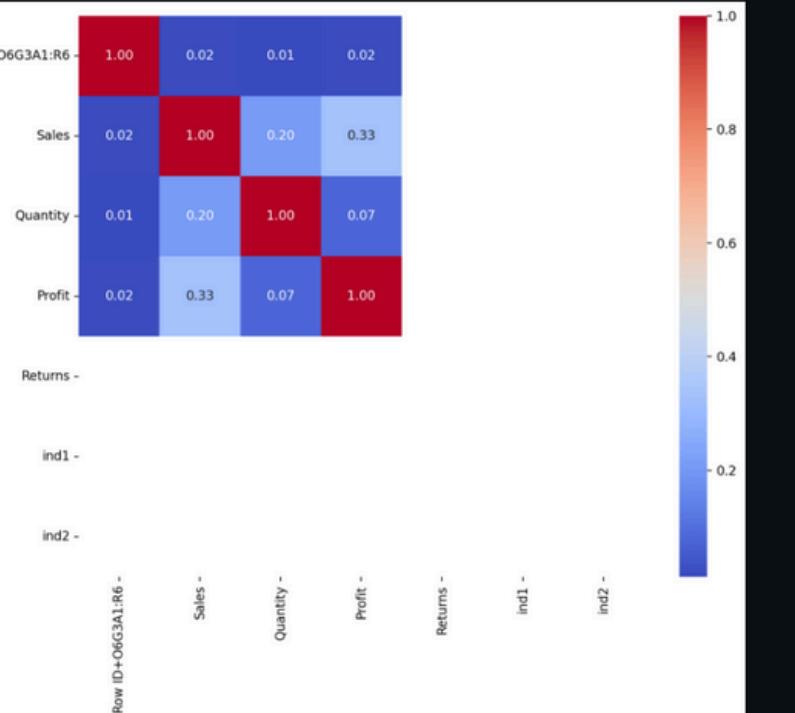
Select target column

Row ID+O6G3A1:R6

Train Dummy Classifier

Use the sidebar to specify the sample size, and the sampled data will be displayed below.

Correlation Heatmap

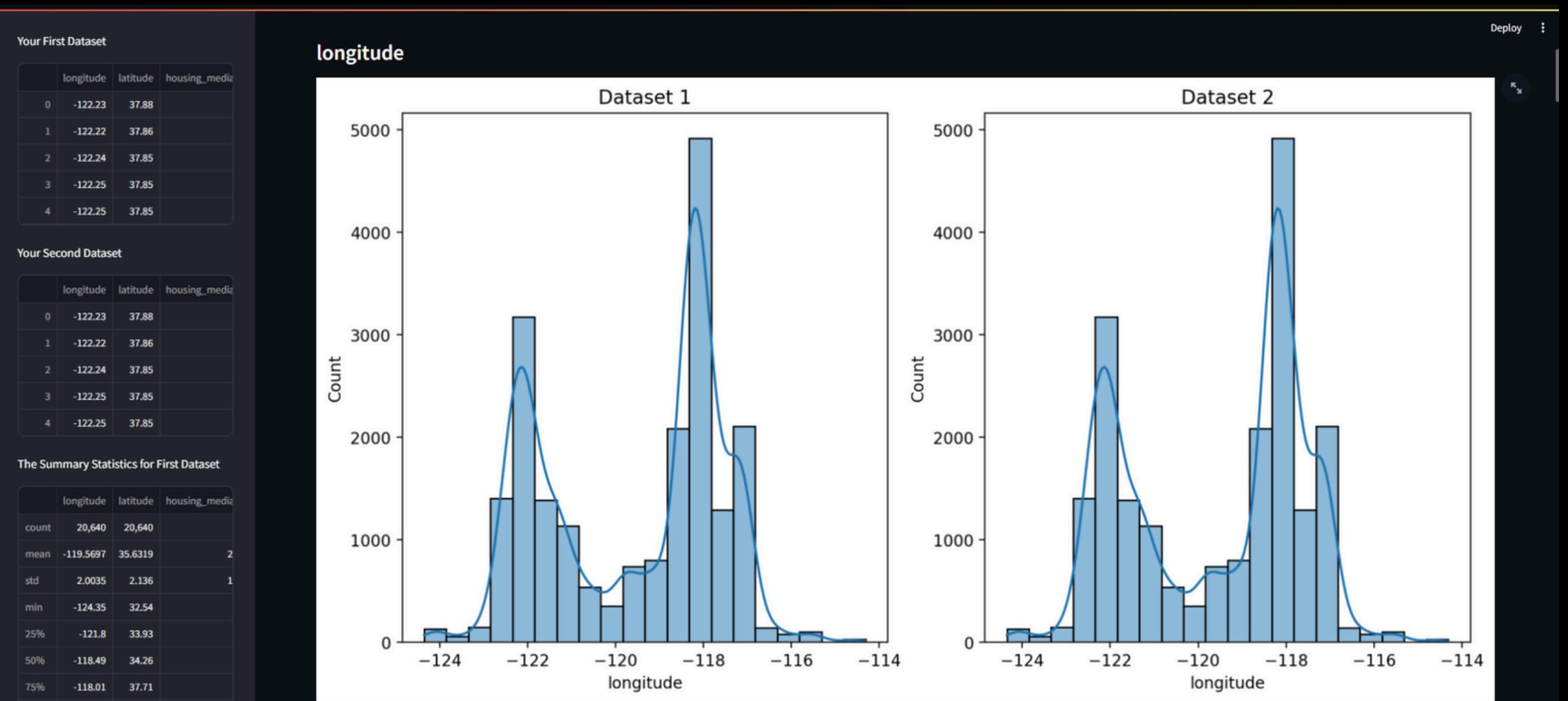
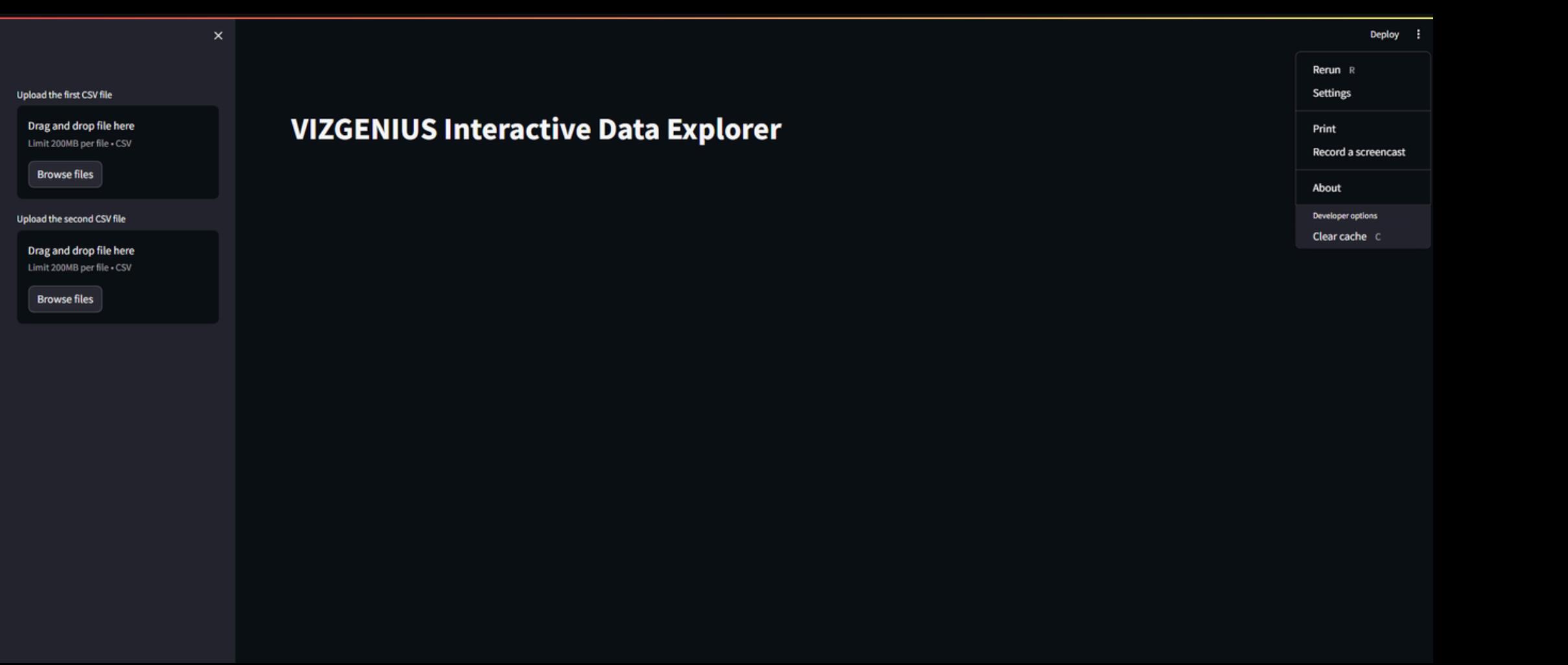


This section displays a correlation heatmap for numerical columns. Stronger correlations are indicated by colors closer to 1 or -1.

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.

Enterprise

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.



Conclusion & Future Scope

Conclusion:

VIZGENIUS leverages open-source tools to create a powerful, accessible data exploration platform. It integrates visualizations, analysis, and machine learning, empowering users to gain insights and make data-driven decisions.

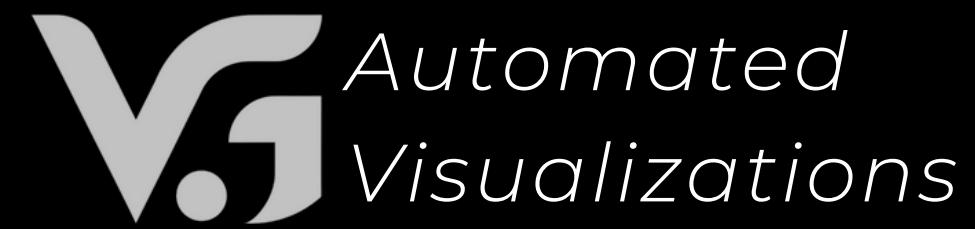
Future Scope:

- **Scalability & Performance:** Handle larger datasets efficiently and optimize performance.
- **Unstructured Data:** Analyze text data like social media posts and documents.
- **Automated Insights:** Develop algorithms to automatically identify patterns and insights from data.

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

THANK YOU!!



DATA IS THE FUEL OF MODERN WORLD!

