Data Analysis Suite – User Guide and Functionality Report

Introduction

The Data Analysis Suite is an interactive Shiny application, which is designed to simplify the process of data exploration, cleaning, feature engineering, and visualization. By using R's Shiny framework and leveraging popular packages such as shinythemes, DT, tidyverse, and corrplot, the app enables users to efficiently work with datasets by helping them through each critical step of the data analysis workflow.

Key Functionalities

1. Welcome

Overview and Instructions:

The landing page welcomes users and provides a step-by-step guide on how to navigate the app. It outlines the primary functionalities: data upload, preprocessing, feature engineering, and visualization.

Contributors & License:

The guide also credits the contributors and states that the application is distributed under the MIT License.

2. Data Upload

Input Options:

Users can either upload their own file with supported formats include CSV, Excel (.xls/.xlsx), JSON, and RDS, with a maximum file size of 1GB, or use an example dataset, including Iris, Motor Trends Cars, and the Penguins dataset.

Data Preview:

After the data upload or selection of an example dataset, a preview will be displayed using an interactive table. A summary box shows essential dataset details, such as the number of rows and columns or dataset description.

3. Preprocessing

Data Cleaning & Transformation:

Users can apply multiple preprocessing operations: Handle Missing Values, Remove Duplicates, Encode Categoricals, Handle Outliers, Normalize Features.

Processed Data Preview:

A data table displays the processed dataset, reflecting the changes based on selected preprocessing operations.

4. Feature Engineering

Creating New Features:

Users can select two numeric columns and choose an arithmetic operation (addition, subtraction, multiplication, or division) to create new features. If users adjust the preprocessing steps later, these new features will be recalculated automatically.
Engineered Data Preview:
The results are displayed in a separate data table, allowing users to verify that the new features have been correctly integrated.

5. Visualization & Exploratory Data Analysis (EDA)
Data Filtering:
A dynamic filter lets users select a numeric column and set a range for data filtering, ensuring visualizations of the desired subset of data. The app supports several types of plots: Scatter Plot, Histogram, Box Plot, Correlation Matrix. Also, users could customizable the plot size and download.
Summary Statistics:
The summary statistics (min, median, mean, max, standard deviation, and missing values count) for numeric columns will be displayed in a tidy table.

Deployment and Access
The application has been deployed on a hosting platform to facilitate easy access without the need for local setup. You can access the live app at:
http://127.0.0.1:5158/

Conclusion
The Data Analysis Suite is a comprehensive and user-friendly tool for performing a full spectrum of data analysis tasks, from initial data upload and preprocessing to advanced visualizations and feature engineering. Its design and interactive components make it an ideal application for both beginners and experienced data scientists who wish to streamline their analysis workflow.

Team Members Contributions:
Anqi Wu: Loading dataset, Data cleaning and Preprocessing, User Interface (UI) and User Experience (UX).
Keito Taketomi: Feature Engineering, EDA, User Interface (UI) and User Experience (UX), Web Application Functionality.
Ziyue Gao: Writing the report.
Yixin Xiao: Writing the report.