

ZANISTA: Key Features and Workflow

Saad H. Ellahie 2024545

Hamza Arif 2024206

April 21, 2025

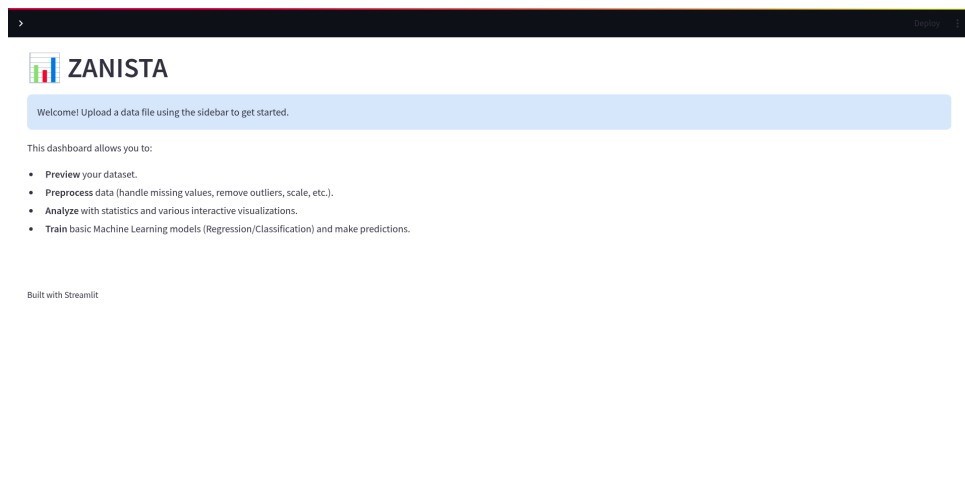


Figure: Conceptual Screenshot of the Dashboard Frontpage

Objective: Design a user-friendly, modular dashboard to streamline dataset ingestion, preprocessing, exploratory analysis, and basic machine learning, empowering users of any skill level to derive insights with minimal friction.

1. Data Ingestion

- Supports CSV, JSON, or Excel uploads; auto-converts to CSV for consistency.
- Validates file structure and reports parsing errors immediately.

2. Preprocessing Pipeline

Users toggle any (or all) of the following steps via checkboxes:

1. **Auto EDA Report:** Generates summary of data types, missingness, correlations, and outliers using Z-score (remove if $|z| > 3$).
2. **Missing Data Handling:** Prompt to impute default values per column or drop rows.
3. **Outlier Removal:** User defines the Z-score threshold and specifies the columns where outliers should be removed.

4. **Custom Row/Column Removal:** Interactive UI to deselect unwanted data slices.
5. **Scaling:** Choose numeric columns to apply scaling using a user-defined factor k .

3. Data Preview

- Quick view of head/tail (configurable number of rows).

4. Exploratory Analysis

- **Descriptive Statistics:** Mean, median, mode, variance, standard deviation, IQR, range. Option to display whisker plot.
- **Distribution Plots:** Choose dependent and independent variables; render bar, histogram, pie, or line charts.

5. Machine Learning Module

- **Model Training:** Predict a target variable using remaining features.
- **Options:** Linear Regression or Decision Tree. Instant train-and-evaluate with R^2 or accuracy metrics.

6. Deliverables

- Interactive web-based dashboard prototype.
- Demo video.