

Data Analytics Dashboard Documentation

May 23, 2025

1 Overview

The Data Analytics Dashboard is a web-based application designed to provide interactive data visualization and analysis for business datasets. Built using Streamlit, it allows users to upload datasets in CSV, XLSX, or JSON formats, visualize data through various chart types, and gain actionable insights using statistical and AI-driven techniques. The dashboard supports dynamic data exploration, enabling users to create custom charts, edit data, and rearrange visualizations via a drag-and-drop interface. It is tailored for business analysts and data professionals seeking to derive meaningful patterns, trends, and anomalies from sales or operational data.

2 Core Features

The dashboard offers the following key features:

- **Data Upload and Editing:** Supports uploading datasets in CSV, XLSX, and JSON formats. Users can preview and edit CSV data directly within the interface, ensuring flexibility in data preparation.
- **Automatic Chart Generation:** Automatically generates charts (Pie, Bar, Line, Scatter, and 3D Scatter) based on dataset characteristics, such as numerical, categorical, and timestamp columns.
- **Custom Chart Creation:** Users can select X, Y, and Z-axis columns and choose chart types, with intelligent suggestions based on data types and distributions.
- **AI and Statistical Insights:** Provides AI-driven insights using clustering (KMeans), anomaly detection (IsolationForest), and trend analysis (LinearRegression), alongside statistical summaries (e.g., missing data, variability, skewness).
- **Insight Management:** Users can toggle insights for individual charts or all charts, with downloadable summarized insights for business decision-making.
- **Responsive Visualizations:** Utilizes Plotly for interactive, responsive charts with a dark theme, optimized for various screen sizes.

3 Tech Stack

The dashboard is built using the following technologies:

- **Python 3:** Core programming language for backend logic and data processing.
- **Streamlit:** Framework for building the web-based user interface and handling interactivity.

- **Pandas:** For data manipulation and analysis, including data loading and preprocessing.
- **Plotly Express:** For generating interactive visualizations (Bar, Line, Scatter, Pie, and 3D Scatter charts).
- **Scikit-learn:** For machine learning tasks, including KMeans clustering, IsolationForest anomaly detection, and LinearRegression for trend analysis.
- **Transformers (Hugging Face):** For text summarization of AI insights using the T5-small model.
- **NumPy:** For numerical computations and data scaling.
- **JavaScript:** For implementing drag-and-drop functionality to reorder charts.

4 Limitations

The dashboard has the following limitations:

- **File Format Support:** Limited to CSV, XLSX, and JSON formats. Other formats (e.g., SQL databases, Parquet) are not supported.
- **AI Insight Generation:** Relies on the T5-small model for summarization, which may produce suboptimal summaries for complex datasets due to its lightweight nature and 512-token input limit.
- **Performance:** Large datasets may slow down chart generation and insight computation due to unoptimized processing, especially for clustering and anomaly detection.

5 Next Steps for Improvement

To enhance the Data Analytics Dashboard, consider the following improvements:

- **Expanded File Support:** Add support for additional file formats (e.g., SQL, Parquet) to broaden compatibility with diverse data sources.
- **Enhanced Data Editing:** Extend editing capabilities to XLSX and JSON files, potentially integrating a more robust data editor with validation checks.
- **Flexible Chart Layouts:** Allow cross-row chart reordering and customizable grid layouts to give users more control over the dashboard's appearance.
- **Interactive Filters:** Introduce dynamic filters for datasets to allow users to focus on specific subsets of data before generating charts.