

DATA ANALYZER AND VISUALIZER

Name : Rajkumar G. Shahu

Exam No : 30004

Project Guide : Prof. Hirav Joshi



Project Profile :

SARDAR PATEL UNIVERSITY

**C. P. Patel F. H. Shah Commerce (Autonomous) Collage, Anand
B.C.A, B.B.A (ITM), PGDCA & M.Sc. (Data Science) Department**

M.Sc. (DS) Semester - 3 PG03CMDS05 (2025 - 2026)

Full Name	RAJKUMAR GAURANGBHAI SHAHU
Project Title	DATA ANALYZER AND VISUALIZER
Roll No.	04
University Exam No.	30004
Front End Tool	STREAMLIT
Back End Tool	PYTHON
Project Guide	Prof. Hirav Joshi



Index

Sr. No	Description of Topic
1	<u>Acknowledgement</u>
2	<u>Project Overview</u>
3	<u>Features</u>
4	<u>Installation & Setup</u>
5	<u>Quick Start Guide</u>
6	<u>Supported File Formats</u>
7	<u>Visualization Types</u>
8	<u>User Interface</u>
9	<u>Conclusion</u>

Acknowledgement

- incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success.
- I am grateful to our project guide “**Prof. Hirav Joshi**” for the guidance, inspiration and constructive suggestion that helped us in the preparation of this project.
- I am also thankful to our Head of The Department “**Dr. Nayan Patel**” for his continuous guidance.
- I am also obliged to our Principal “**Dr. Tejal Tanna**” for creative support.
- I also thank my colleagues who have helped in successful completion of the project.



Project Overview

The **Data Analysis & Visualization Dashboard** is a comprehensive web-based application built with Streamlit that enables users to upload, analyze, and visualize data without writing any code. The dashboard provides an intuitive interface for exploratory data analysis (EDA), interactive visualizations, and data export capabilities.

Target Users

- Data analysts and scientists
- Business analysts
- Students and researchers
- Non-technical stakeholders who need data insights



Features

Core Functionality

- **Multi-format file upload** (CSV, Excel, JSON, Parquet)
- **Interactive data exploration** with filtering and sorting
- **Comprehensive statistical analysis** and summaries
- **12+ visualization types** with customization options
- **Real-time data filtering** and transformation
- **Export capabilities** (CSV, Excel formats)
- **Responsive design** for different screen sizes

Analysis Capabilities

- Descriptive statistics
- Data type detection and conversion
- Missing value analysis
- Correlation analysis
- Distribution analysis
- Categorical data analysis

Visualization Features

- Interactive charts with Plotly
- Customizable color schemes
- Dynamic filtering and grouping
- Responsive chart sizing
- Professional styling



Installation & Setup

- **Prerequisites :**
- Python 3.7 or higher
- pip package manager
- 2GB RAM minimum (8GB recommended)
- Modern web browser

- **Step 1: Environment Setup**

- # Create virtual environment (recommended)
- python -m venv streamlit_dashboard
- source streamlit_dashboard/bin/activate
- # On Windows: streamlit_dashboard\Scripts\activate
-
- # Or using conda
- conda create -n streamlit_dashboard python=3.9
- conda activate streamlit_dashboard



Installation & Setup

- **Step 2: Install Dependencies**

- # Install required packages
- pip install streamlit
- pip install pandas
- pip install plotly
- pip install seaborn
- pip install matplotlib
- pip install scipy
- pip install openpyxl
- pip install pyarrow

- **Step 3: Download and Run**

- # Save the dashboard code as 'r1.py'
- streamlit run r1.py



Quick start Guide

- **1. Launch the Application**
 - streamlit run r1.py
 - The dashboard will open in your default browser at <http://localhost:8501>
- **2. Upload Your Data**
 - Click "**Choose a file**" in the sidebar
 - Select your data file (CSV, Excel, JSON, or Parquet)
 - Wait for the upload confirmation
- **3. Explore Your Data**
 - Review the **Data Overview** section for basic statistics
 - Use the **Data Preview** to examine your dataset
 - Check **Data Types** and **Missing Values**



Quick start Guide

- **4. Create Visualizations**
 - Select a **Visualization Type** from the dropdown
 - Choose appropriate **columns** for X and Y axes
 - Customize colors, groupings, and aggregations
 - View your interactive chart
-
- **5. Filter and Export**
 - Use the **Data Filtering** section to subset your data
 - Export filtered results as **CSV** or **Excel**



Supported File Formats

- **CSV Files (.csv)**
 - # Reading parameters
 - df = pd.read_csv(file,
 - encoding='utf-8',
 - parse_dates=True,
 - infer_datetime_format=True
 -)
- **Excel Files (.xlsx, .xls)**
 - # Multi-sheet support
 - df = pd.read_excel(file, sheet_name=0) # First sheet
- **JSON Files (.json)**
 - # Nested JSON flattening
 - df = pd.json_normalize(json_data)
- **Parquet Files (.parquet)**
 - # High-performance columnar format
 - df = pd.read_parquet(file)



Visualization types

- **1. Scatter Plot**

- **Purpose:** Explore relationships between two numeric variables

- **Parameters:**

- X-axis: Numeric column

- Y-axis: Numeric column

- Color: Categorical column (optional)

- Size: Numeric column (optional)

- **Use Cases:**

- Correlation analysis

- Outlier detection

- Pattern identification

- **2. Line Chart**

- **Purpose:** Show trends over time or ordered categories

- **Parameters:**

- X-axis: DateTime or numeric column

- Y-axis: Numeric column

- Group by: Categorical column (optional)

- **Use Cases:**

- Time series analysis

- Trend visualization

- Performance tracking



Visualization types

- **3. Bar Chart**

- **Purpose:** Compare categories or show distributions

- **Parameters:**

- Category: Categorical column

- Value: Numeric column

- Aggregation: sum, mean, count, median

- **Use Cases:**

- Category comparison

- Distribution analysis

- Ranking visualization

- **4. Histogram**

- **Purpose:** Show distribution of a single numeric variable

- **Parameters:**

- Column: Numeric column

- Bins: Number of bins (10-50)

- Overlay: Normal distribution curve

- **Use Cases:**

- Distribution analysis

- Data quality assessment

- Statistical analysis



Visualization types

- **5. Box Plot**

- **Purpose:** Show distribution quartiles and outliers

- **Parameters:**

- Y-axis: Numeric column

- Group by: Categorical column (optional)

- **Use Cases:**

- Outlier detection

- Distribution comparison

- Statistical summary

- **6. Heatmap**

- **Purpose:** Show correlation between numeric variables

- **Parameters:**

- Correlation method: Pearson, Spearman

- Color scale: Various options

- Annotations: Show correlation values

- **Use Cases:**

- Feature selection

- Multicollinearity detection

- Pattern recognition



Visualization types

- **7. Pie Chart**
- **Purpose:** Show proportions of categorical data
- **Parameters:**
 - Category: Categorical column
 - Limit: Top N categories
- **Use Cases:**
 - Market share analysis
 - Category distribution
 - Budget allocation
- **8. Distribution Plot**
- **Purpose:** Combine histogram with density curve
- **Parameters:**
 - Column: Numeric column
 - KDE: Kernel density estimation overlay
- **Use Cases:**
 - Distribution shape analysis
 - Statistical modeling
 - Data exploration



Visualization types

- **9. Time Series**
- **Purpose:** Specialized line chart for temporal data
- **Parameters:**
- Date column: DateTime column
- Value column: Numeric column
- Resampling: Daily, weekly, monthly
- **Use Cases:**
- Temporal trend analysis
- Seasonal pattern detection
- Forecasting preparation



User Interface

The screenshot shows a web browser window titled "DATA ANALYZER" with the URL "localhost:8501". The interface has a dark background. On the left, there is a sidebar with a "Data Upload Here" section containing a file input field, a "Drag and drop file here" area with a 200MB limit for CSV, XLSX, XLS, JSON, and PARQUET files, and buttons for "Browse files" and "Use Sample Data". The main area features a logo with three vertical bars in red, green, and blue, followed by the text "DATA ANALYZER & VISUALIZER". A prominent message box says "Please upload a data file to get started!". At the bottom, the Windows taskbar is visible with various icons and a system tray icon for the application.

You are signed in as test

DATA ANALYZER

localhost:8501

Data Upload Here

Choose a file

Drag and drop file here
Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET

Browse files

Use Sample Data

DATA ANALYZER & VISUALIZER

Please upload a data file to get started!

Deploy

Search

ENG IN

07/08/2025

User Interface

The screenshot shows a web browser window titled "DATA ANALYZER" at the URL "localhost:8501". The main interface on the left is a "Data Upload Here" section with a "Choose a file" input field, a "Drag and drop file here" area with restrictions ("Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET"), and two buttons: "Browse files" and "Use Sample Data". A "Deploy" button is visible in the top right corner of this section.

A "File Explorer" dialog box is overlaid on the page, centered over the upload area. It shows a list of files and folders on "New Volume (E:)". The list includes:

Name	Date modified	Type
jyoti(project)	02/09/2025 13:38	File folder
Avnita(project)	01/08/2025 13:04	File folder
files		
jyoti(project)	25/08/2025 10:14	File folder
labsoft	30/06/2025 15:10	File folder
project	22/08/2025 11:25	File folder
if condition	22/10/2024 11:45	Microsoft Excel
if_analysis	21/10/2024 11:52	Microsoft Excel
Trial	13/09/2024 13:38	Microsoft Excel

At the bottom of the dialog, there are buttons for "File name:" (set to "if condition"), "Custom files", "Upload from mobile", "Open", and "Cancel".

The browser's taskbar at the bottom shows various pinned icons, including Microsoft Edge, File Explorer, and File History. A system tray icon for "ENG IN" is also visible.

User Interface

You are signed in as test

DATA ANALYZER

localhost:8501

Deploy :

Data Upload Here

Choose a file

Drag and drop file here
Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET

Browse files

Trial.xlsx 94.3KB

Selected file loaded successfully !

Selected File fetched successfully !

Shape: 38 rows × 14 columns

DATA ANALYZER & VISUALIZER

Data Overview

Total Rows	Total Columns	Numeric Columns	Text Columns
38	14	8	5

Data Preview

	Column1	Region	Manager	SalesMan	Item	Units	Unit_price	Sale_amt	Unnamed: 8	Unnamed: 9
0	2018-01-06 00:00:00	East	Martha	Alexander	Television	95	1,198	113,810	None	None
1	2018-01-23 00:00:00	Central	Hermann	Shelli	Home Theater	50	500	25,000	None	None
2	2018-02-09 00:00:00	Central	Hermann	Luis	Television	36	1,198	43,128	None	None
3	2018-02-26 00:00:00	Central	Timothy	David	Cell Phone	27	225	6,075	None	None

Thunderstorm w... In effect

Search

ENG IN

03:19 25

09:15

User Interface

You are signed in as test

DATA ANALYZER

localhost:8501

Deploy :

Data Upload Here

Choose a file ?

Drag and drop file here
Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET

Browse files

Trial.xlsx 94.3KB

Selected file loaded successfully !

Selected File fetched successfully !

Shape: 38 rows × 14 columns

Thunderstorm w... In effect

Search

ENG IN

03 10 2025

Data Types

	Column	Data Type	Non-Null Count	Null Count
Column1	Column1	datetime64	15	2
Region	Region	object	15	2
Manager	Manager	object	15	2
SalesMan	SalesMan	object	15	2
Item	Item	object	15	2
Units	Units	float64	15	2
Unit_price	Unit_price	float64	15	2
Sale_amt	Sale_amt	float64	15	2
Unnamed: 8	Unnamed: 8	float64	0	3
Unnamed: 9	Unnamed: 9	float64	0	3

Basic Statistics

	Units	Unit_price	Sale_amt	Unnamed: 8	Unnamed: 9
count	15	15	15	0	
mean	54.5333	828.9333	50,407.1333	None	N
std	27.7202	421.8094	37,856.8008	None	N
min	2	125	250	None	N
25%	33.5	500	27,500	None	N
50%	56	1,198	40,500	None	N
75%	78	1,198	78,469	None	N
max	95	1,198	113,810	None	N

Data Filtering



User Interface

The screenshot shows a web-based data analysis application with a dark theme. The top navigation bar includes tabs for 'You are signed in as test' and 'DATA ANALYZER'. The main interface is divided into two main sections: 'Data Filtering' on the left and 'Data Visualizations' on the right.

Data Filtering Section:

- Data Upload Here:** A section for uploading files. It includes a placeholder 'Drag and drop file here' and a note 'Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET'. A 'Browse files' button is available for manual selection.
- File Preview:** Shows a preview of the uploaded 'Trial.xlsx' file, which is 94.3KB in size.
- Success Messages:** Two green success messages: 'Selected file loaded successfully!' and 'Selected File fetched successfully!' followed by an exclamation mark.
- Shape Information:** Displays the data shape as '38 rows x 14 columns'.
- Data Filtering UI:** A form for filtering data. It starts with a magnifying glass icon and the title 'Data Filtering'. Below it is a dropdown menu labeled 'Region' containing 'East', 'Central', 'West', and 'nan'. A message below the dropdown states 'Filtered data: 38 rows (from 38 original rows)'.

Data Visualizations Section:

- Data Visualization Title:** A section titled 'Data Visualizations' featuring a bar chart icon.
- Visualization Selection:** A dropdown menu labeled 'Select Visualization Type' with 'Scatter Plot' selected.

User Interface

You are signed in as test

DATA ANALYZER

localhost:8501

Deploy

Data Upload Here

Choose a file

Drag and drop file here
Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET

Browse files

Trial.xlsx 94.3KB

Selected file loaded successfully!

Selected File fetched successfully!

Shape: 38 rows × 14 columns

Select Visualization Type

Scatter Plot

X-axis: Units

Y-axis: Units

Color by: None

Units vs Units

X (Units)	Y (Units)
0	0
10	25
15	30
20	35
25	38
30	40
35	42
40	50
50	55
60	60
70	75
80	80
90	90
95	95
98	98

News for you
TCS Announces...

Search

ENG IN

05/05/2025

A magnifying glass icon with a bar chart overlay.

User Interface

The screenshot shows a web browser window titled "DATA ANALYZER" with the URL "localhost:8501". The interface is dark-themed and includes the following components:

- Data Upload Area:** On the left, there's a section titled "Data Upload Here" with a file input field labeled "Choose a file". Below it is a "Drag and drop file here" area with a limit of "200MB per file • CSV, XLSX, XLS, JSON, PARQUET". A "Browse files" button is also present.
- File Preview:** A file named "Trial.xlsx" (94.3KB) is listed with an "X" icon to its right.
- Status Messages:** Two green success messages are displayed: "Selected file loaded successfully!" and "Selected File fetched successfully!".
- Shape Information:** A message stating "Shape: 38 rows × 14 columns" is shown in a blue box.
- Export Data Section:** In the center, there's a "Export Data" section with icons for "Download CSV" and "Download Excel".
- Footer:** At the bottom, a footer bar contains the text "0 2024 – 2026 ALL RIGHTS RESERVED BY RAJKUMAR SHAHU".
- System Tray:** The bottom right corner shows a system tray with icons for battery, signal, volume, and date/time (03 10 2024 09:17). A magnifying glass icon over a bar chart is also visible.
- Bottom Navigation:** The bottom of the screen features a navigation bar with various icons, including a search bar and system controls.

Conclusion :

- In this project, we developed an interactive **Data Analyzer and Visualizer** application using **Python** and **Streamlit**. The tool allows users to upload datasets, explore them through summary statistics, clean data, and generate a variety of insightful visualizations with ease.
- By integrating libraries such as **Pandas**, **Matplotlib**, **Seaborn**, and **Plotly**, we provided dynamic data exploration capabilities, including:
 - Viewing dataset structure and summary statistics
 - Handling missing values and filtering data
 - Creating visualizations like bar charts, histograms, scatter plots, heatmaps, and more
 - Exporting cleaned or filtered data for further use



Conclusion :

- This project demonstrates how data science tools can be made accessible through user-friendly web interfaces. It serves as a foundation for more advanced analytics applications, where users can gain insights from their data without writing any code.
- In future iterations, this tool can be enhanced with features like:
 - Machine learning model integration
 - Real-time data updates
 - Dashboard export capabilities
 - Collaboration tools or database integration
- Overall, the project showcases the power of combining Python's data stack with Streamlit's simplicity to democratize data analysis for both technical and non-technical users.





Thank You

Created By : Rajkumar G. Shahu