

# 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 PG03CMD505 (2025 - 2026)**

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



# Index

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

# 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 helpful 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



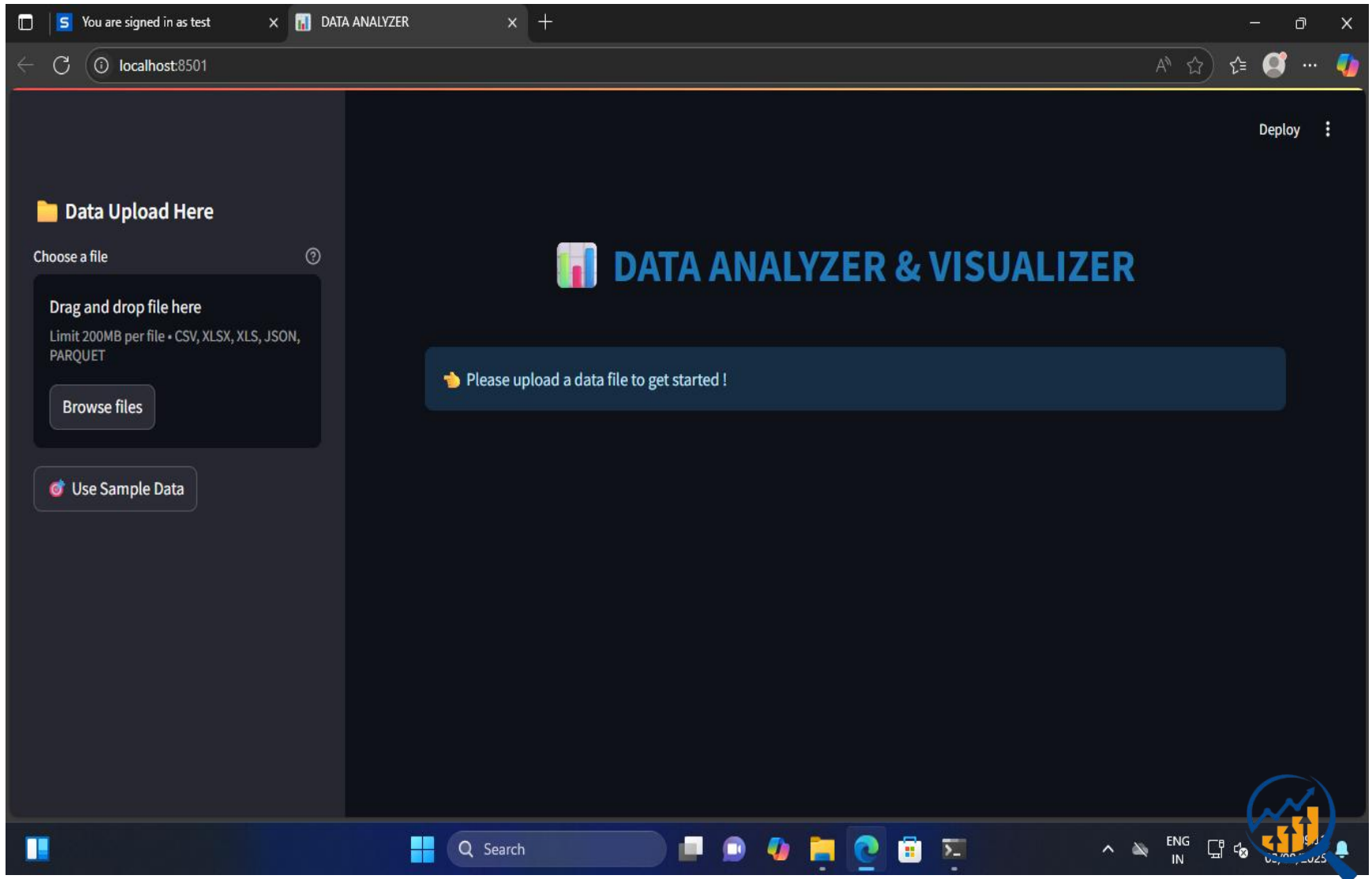
# Visualization types

- **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

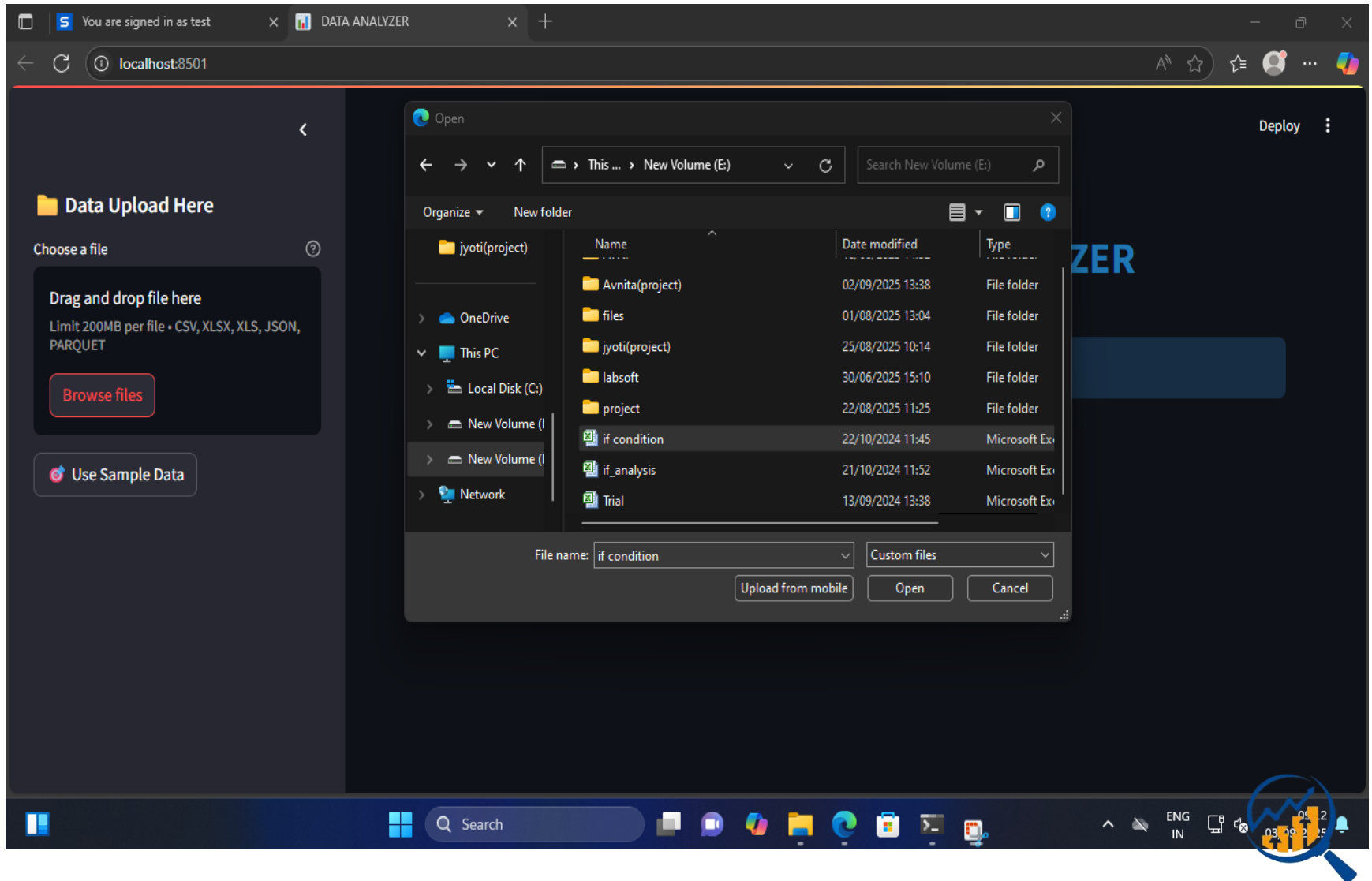




# User Interface



# User Interface



# User Interface

The screenshot shows a web browser window with the URL `localhost:8501`. The browser tabs include "You are signed in as test" and "DATA ANALYZER". The application interface is dark-themed and features a sidebar on the left for file uploads and a main content area for data analysis.

**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

# User Interface

The screenshot displays the DATA ANALYZER web application interface. The browser window shows the URL localhost:8501. The application has a dark theme and a sidebar on the left for file uploads. The main area is divided into two panels: 'Data Types' and 'Basic Statistics'.

**Data Upload Section:**

- Header: Data Upload Here
- Choose a file
- Drag and drop file here (Limit 200MB per file • CSV, XLSX, XLS, JSON, PARQUET)
- Browse files button
- File list: Trial.xlsx (94.3KB)
- Success messages: Selected file loaded successfully! and Selected File fetched successfully!
- Shape: 38 rows × 14 columns

**Data Types Table:**

	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

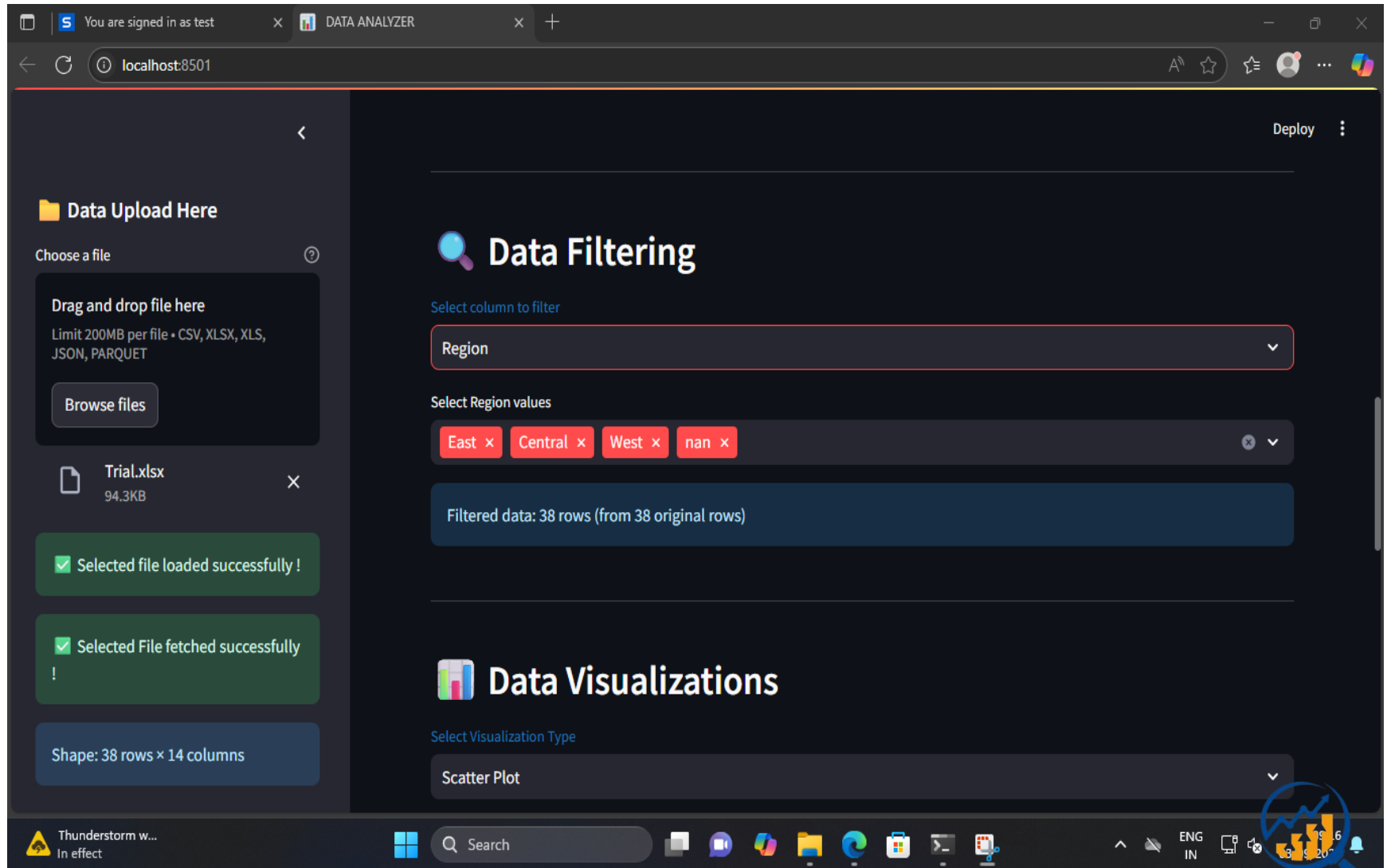
**Basic Statistics Table:**

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

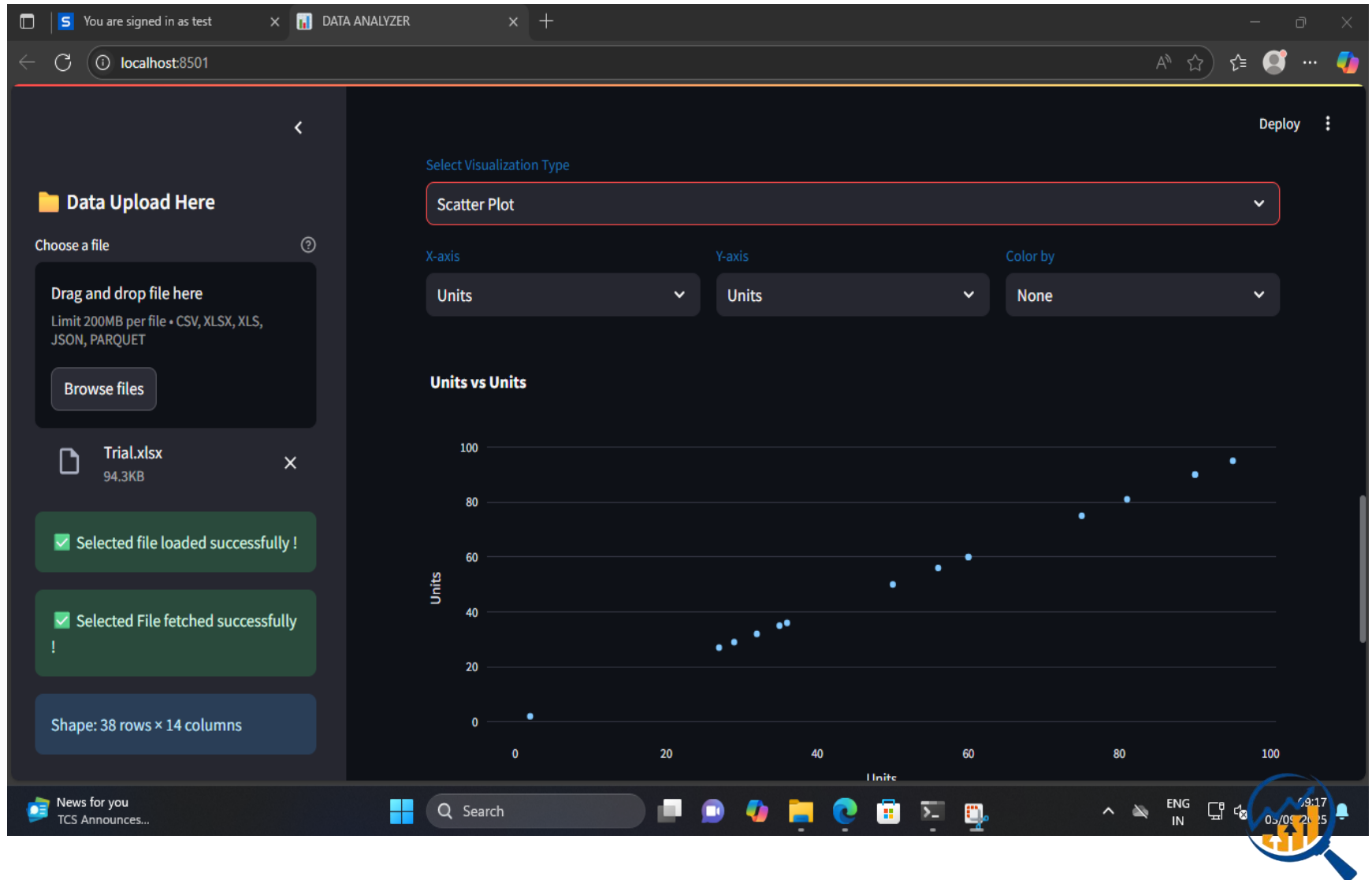
**Windows Taskbar:**

- Thunderstorm warning: In effect
- Search bar
- Taskbar icons: File Explorer, Microsoft Edge, etc.
- System tray: ENG IN, date/time (03/06/2023 05:15)

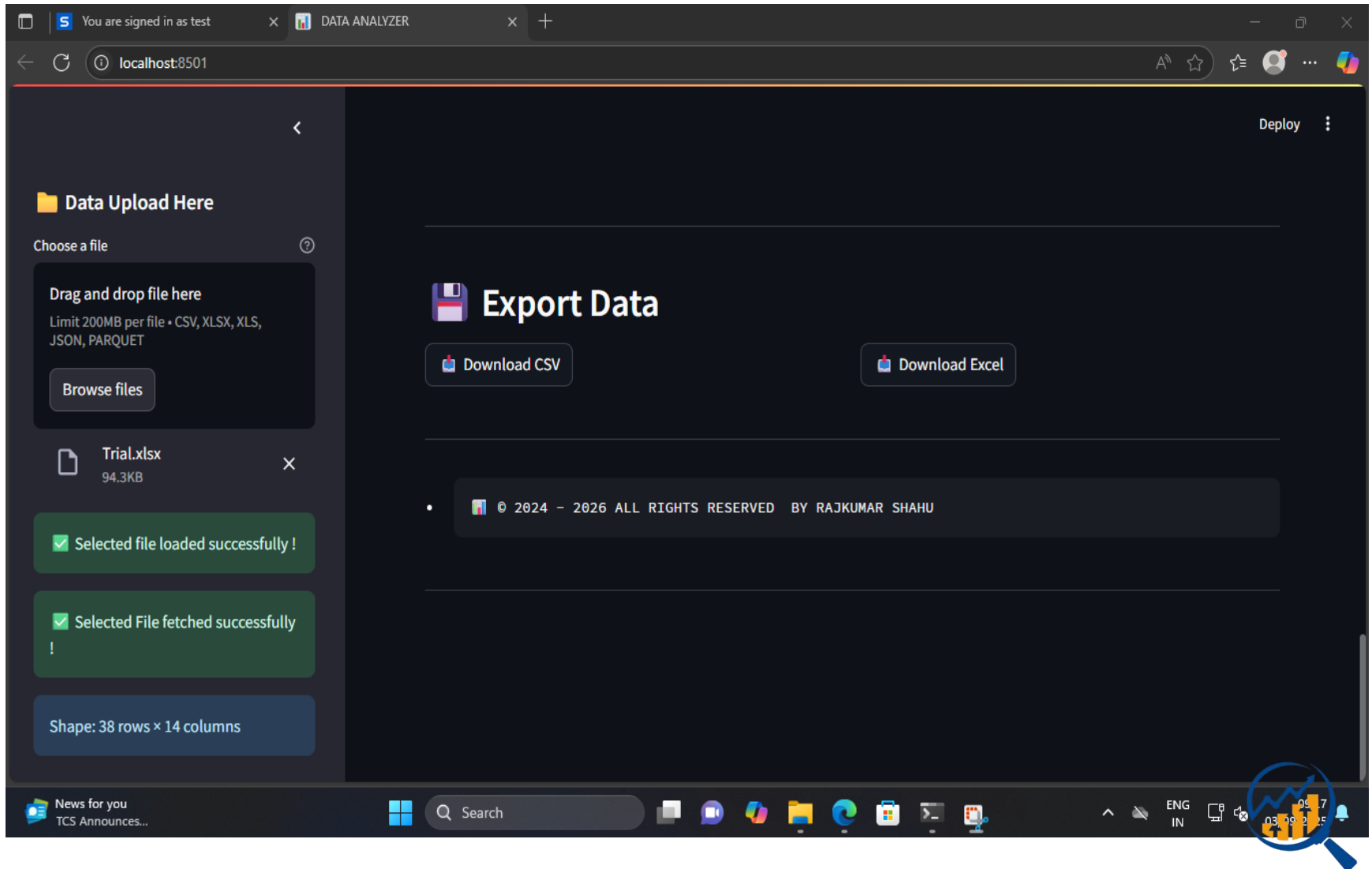
# User Interface



# User Interface



# User Interface



# 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