



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

School of Computer Science & Engineering

Department of Computer Science and Applications

2025-2026

Synopsis

On

Everything About You in a Drop: A Smart Analyzer Using Blood Test Datasets



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Mini Project

Course Code: COS4007B

**BSc Honors Data Science and Big Data
Analytics**

(BSc Hon DSBDA)



Group Id:

Team Leader: Brijesh Umesh Gupta

S.No.	Team Members	PRN
1	Brijesh Gupta	1272240145
2		
3		
4		

Project Title: Everything About You in a Drop: A Smart Analyzer Using Blood Test Datasets

Name of the Mentor: Rahul Sir

INDEX

Abstract	1
-----------------------	---

1. INTRODUCTION

1.1 Background	2
1.2 Challenges in Existing System	3

2. PROPOSED WORK

2.1 Overview of the System	4
2.2 Objectives of the Proposed System	5
2.3 Key Features	5

3. REQUIREMENT ANALYSIS

3.1 Purpose	6
3.2 Project Scope	7
3.3 Software & Hardware Requirements	8

4. DESIGN METHODOLOGY

4.1 System Architecture	9
4.2 UML / ER Diagram	10

5. IMPLEMENTATION DETAILS

5.1 Dashboard & Workflow Explanation	11
5.2 Sample Outputs (Screenshots of Graphs/Plots)	12

6. RESULTS & DISCUSSION

6.1 Interpretation of Blood Test Data	13
---	----

7. LIMITATIONS	14
-----------------------------	----

8. FUTURE SCOPE / PERSPECTIVE	15
--	----

9. CONCLUSION	16
----------------------------	----

10. BIBLIOGRAPHY / REFERENCES	17
--	----

Abstract

This project, “*Everything About You in a Drop*,” is an intelligent analyzer designed to interpret blood test datasets and visualize vital health parameters through interactive dashboards.

Using machine learning and data visualization techniques, the system identifies trends, abnormalities, and correlations in medical data such as glucose level, hemoglobin, cholesterol, and blood cell count.

The dashboard helps users, doctors, and lab technicians quickly interpret results with charts like line plots, scatter plots, histograms, box plots, pie charts, and heatmaps.

The goal is to transform raw medical data into meaningful, actionable insights that support health awareness and early disease detection.



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Introduction

Blood tests provide essential information about a person's health, but interpreting these values can often be confusing without medical expertise.

This project uses data science and analytics to simplify the interpretation process.

By analyzing patterns in blood test results, the system provides a visual and user-friendly interface for health understanding.

The analyzer can serve patients, diagnostic centers, and medical researchers by visualizing variations and trends over time, enabling proactive health monitoring.



Background

Traditional blood test reports are static and text-heavy, making it difficult to understand relationships between multiple parameters.

With the advancement of data science and visualization tools, we can now present medical data interactively and intelligently.

Using Python libraries such as Pandas, Matplotlib, and Plotly, combined with dashboards built in Streamlit or Dash, users can visualize large datasets in an intuitive manner.

The system focuses on simplifying data interpretation and providing valuable health patterns derived from raw blood test data.

Challenges in Existing System

- Manual interpretation of lab results often leads to errors or misunderstanding.
- Lack of interactive visualization tools in standard lab reports.
- Existing systems do not provide insights or pattern recognition based on multiple parameters.
- Difficulty in identifying abnormal values and health risks automatically.
- No centralized dashboard for visual comparison between patients or test periods.



Proposed Work

The proposed system introduces a *Smart Blood Test Analyzer* that reads blood test data, cleans and preprocesses it, and generates dynamic dashboards.

The tool allows users to upload CSV files containing blood test results and automatically plots insightful visualizations.

Conditional color formatting helps highlight abnormal readings.

The application also integrates a “Symptom Dictionary” to cross-reference common symptoms with abnormal test indicators.

Key features include:

- Automated data cleaning and analysis.
- Visual dashboards: line plot, scatter plot, histogram, box plot, pie chart, heatmap.
- Color-coded indicators for health conditions.
- Workflow integration for dataset upload, analysis, and result visualization.
- Friendly UI with a welcome screen: “*Welcome to Brijesh Gupta’s Everything About You in a Drop.*”



Requirement Analysis

Purpose

The purpose of this system is to develop a health analysis tool that can visualize and interpret blood test datasets using machine learning and data visualization methods. It aims to make complex data simple and insightful for both medical professionals and individuals.

Project Scope

This project is suitable for:

- Health monitoring applications
- Diagnostic laboratories
- Medical data science research
- Educational demonstrations in healthcare analytics

The project emphasizes on real-time visualization and analysis of medical data, making it valuable in both academic and practical healthcare environments.

Software & Hardware Requirements

Software Requirements:

- Python 3.x
- Streamlit / Dash (for dashboards)
- Pandas, NumPy, Matplotlib, Plotly, Seaborn
- Jupyter Notebook / VS Code
- Microsoft Excel / CSV files for datasets

Hardware Requirements:

- Laptop or PC with minimum 8GB RAM
- Intel i5 or higher processor
- Stable internet connection for dashboard deployment

Design Methodology

System Architecture

The architecture follows a modular approach:

1. **Data Input Module:** Reads the CSV file containing blood test results.
2. **Data Preprocessing Module:** Cleans and validates missing or inconsistent data.
3. **Analyzer Module:** Applies basic statistical operations and conditional checks.
4. **Visualization Module:** Displays results through various charts and plots.
5. **Dashboard Module:** Provides an interactive user interface with workflows and colorized health indicators.

UML / ER Diagram

The **ER Diagram** includes the following entities:

- **Patient** (Patient_ID, Name, Age, Gender)
- **Test_Result** (Test_ID, Patient_ID, Parameter, Value, Normal_Range, Date)
- **Analysis_Report** (Report_ID, Patient_ID, Observation, Risk_Level, Suggestion)

Relationships:

- One *Patient* → Many *Test_Results*
- One *Analysis_Report* → Belongs to One *Patient*



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Implementation Details

The project is implemented using Python and Streamlit for dashboard creation.

Workflows include:

1. Uploading dataset (CSV).
2. Data cleaning and feature extraction.
3. Generating visualizations automatically.
4. Displaying color-coded interpretations (e.g., red for abnormal, green for normal).
5. Exporting summarized insights.



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Dashboard & Workflow Explanation

- **Line Plot:** Tracks variation of test results over time.
- **Scatter Plot:** Shows relationships between two parameters (e.g., Hemoglobin vs. RBC).
- **Histogram:** Displays frequency distribution of test values.
- **Box Plot:** Highlights outliers and abnormal readings.
- **Pie Chart:** Represents proportion of parameters within normal/abnormal ranges.
- **Heatmap:** Correlates multiple blood parameters.



Sample Outputs (Screenshots of graphs/plots)

Welcome to Brijesh Gupta's Everything about you in a drop

This project analyses blood-test CSVs: It cleans the data automatically, trains a model on your chosen target column, predicts values for new samples, flags abnormal parameters, and suggests educational symptom lists and next steps – all offline.

Workflows

- Upload & Clean – Upload your CSV file (single header row). The app will auto-clean: trim headers, drop ID columns, impute numeric missing values, and label-encode categorical columns.
- Train & Predict – Choose a target column and train a RandomForest model. After training, predict on a new single-row CSV or enter values manually.
- Heuristic & Symptoms – If a parameter is unusually HIGH/LOW (based on $\pm 1.5\text{std}$), the app flags it and shows probable symptoms and suggested next steps from a local dictionary.
- Dashboards & Visuals – Explore the cleaned dataset using colorized tables, correlation heatmap, line/scatter/histogram/box plots.
- Save & Export – Download the cleaned CSV and the trained model for reuse.

Tips

- Keep column names clean (no trailing spaces).
- If a CSV fails to parse, open it in Excel and "Save as CSV (UTF-8)".
- The symptom suggestions are rule-based and educational only – not medical advice.

Medical disclaimer: Educational purposes only. Consult a qualified healthcare professional for any medical concerns.

Upload CSV & Auto-clean

Upload CSV

Drag and drop file here
Limit 20MB per file - CSV

Browse files

Diabetes Classification.csv 220.5KB

Raw preview (interactive)

	Unnamed: 0	Age	Gender	BMI	Chol	TG	HDL	LDL	Cr	BUN	Diagnosis
0	0	50	F	24	4.2	0.9	2.4	1.4	46	4.7	0
1	1	26	M	28	3.7	1.4	1.1	2.1	62	4.5	0
2	2	33	M	21	4.9	1	0.8	2	46	7.3	0
3	3	45	F	21	2.9	1	1	1.5	24	2.3	0
4	4	50	F	24	3.6	1.3	0.9	2.1	50	2	0

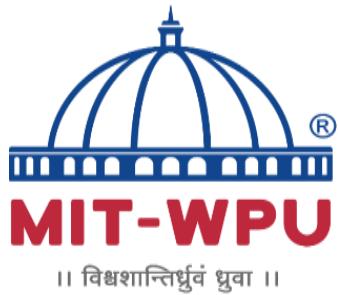
Shape: (5133, 11)

Auto-clean

Cleaned.

Clean preview (interactive)

	Unnamed: 0	Age	Gender	BMI	Chol	TG	HDL	LDL	Cr	BUN	Diagnosis
0	0	50	F	24	4.2	0.9	2.4	1.4	46	4.7	0
1	1	26	M	28	3.7	1.4	1.1	2.1	62	4.5	0
2	2	33	M	21	4.9	1	0.8	2	46	7.3	0
3	3	45	F	21	2.9	1	1	1.5	24	2.3	0



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Navigation

Go to

- Welcome
- Upload & Clean
- Train & Predict
- Dashboards
- Settings

Cleaned.

Clean preview (interactive)

	Unnamed: 0	Age	Gender	BMI	Chol	TG	HDL	LDL	Cr	BUN	Diagnosis
0	0	50	0	24	4.2	0.9	2.4	1.4	46	4.7	0
1	1	26	1	23	3.7	1.4	1.1	2.1	62	4.5	0
2	2	33	1	21	4.9	1	0.8	2	46	7.1	0
3	3	45	0	21	2.9	1	1	1.5	24	2.3	0
4	4	50	0	24	3.6	1.3	0.9	2.1	50	2	0

Colorized preview (values high/low highlighted):

	Unnamed: 0	Age	Gender	BMI	Chol	TG	HDL	LDL	Cr	BUN	Diagnosis
0	0.000	50.000	0	24.000	4.200	0.900	2.400	1.400	46.000	4.700	0.000
1	1.000	26.000	1	23.000	3.700	1.400	1.100	2.100	62.000	4.500	0.000
2	2.000	33.000	1	21.000	4.900	1.000	0.800	2.000	46.000	7.100	0.000
3	3.000	45.000	0	21.000	2.900	1.000	1.000	1.500	24.000	2.300	0.000
4	4.000	50.000	0	24.000	3.600	1.300	0.900	2.100	50.000	2.000	0.000
5	5.000	48.000	1	24.000	2.900	0.800	0.900	1.600	47.000	4.700	0.000
6	6.000	43.000	1	21.000	3.800	0.900	2.400	3.700	67.000	2.600	0.000
7	7.000	32.000	0	24.000	3.800	2.000	2.400	3.800	28.000	3.600	0.000
8	8.000	31.000	0	23.000	3.600	0.700	1.700	1.600	55.000	4.400	0.000
9	9.000	33.000	0	21.000	4.000	1.100	0.900	2.700	53.000	3.300	0.000
10	10.000	30.000	0	22.000	4.900	1.300	1.200	3.200	42.000	3.000	0.000
11	11.000	45.000	0	23.000	4.200	1.700	1.200	3.200	54.000	4.600	0.000
12	12.000	50.000	0	24.000	4.000	1.500	1.200	2.200	39.000	3.500	0.000

Upcoming Earnings

Deploy : 07:52 PM 12-11-2025

Navigation

Go to

- Welcome
- Upload & Clean
- Train & Predict
- Dashboards
- Settings

Train & Predict

Dataset shape: (5332, 11)

Select target column to predict:

Chol

Test size (%)

20

Train model

Regressor MSE: 0.051, R2: 0.951

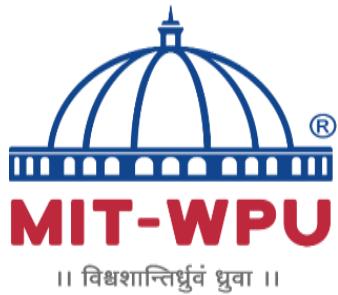
Feature importance (table & colorized bar)

feature	importance
6 LDL	0.5358
4 TG	0.1298
5 HDL	0.1085
0 Unnamed: 0	0.0711
8 BUN	0.0447
7 Cr	0.0419
3 BMI	0.031
1 Age	0.0295
2 Gender	0.0061
9 Diagnosis	0.0017

feature	importance
6 LDL	0.5358

Upcoming Earnings

Deploy : 07:53 PM 12-11-2025



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Navigation

Go to

- Welcome
- Upload & Clean
- Train & Predict
- Dashboards
- Settings

24°C Partly cloudy

Train model

Regressor MSE: 0.051, R2: 0.951

Feature importance (table & colorized bar)

feature	importance
6 LDL	0.5358
4 TG	0.1298
5 HDL	0.1085
0 Unnamed: 0	0.0711
8 BUN	0.0447
7 Cr	0.0419
3 BMI	0.031
1 Age	0.0295
2 Gender	0.0061
9 Diagnosis	0.0017

feature importance

feature	importance
6 LDL	0.535831
4 TG	0.129787
5 HDL	0.108505
0 Unnamed: 0	0.071083
8 BUN	0.044667
7 Cr	0.041888
3 BMI	0.031020
1 Age	0.029461
2 Gender	0.006092

Predict on a new sample

Upload a single-row CSV for prediction (optional)

Drag and drop file here
Limit: 200MB per file • CSV

one_row_diabetes.csv 96.08

Browse files

Deploy

Navigation

Go to

- Welcome
- Upload & Clean
- Train & Predict
- Dashboards
- Settings

24°C Partly cloudy

Prediction result

4.9348000000000002

Top contributing features (model importance)

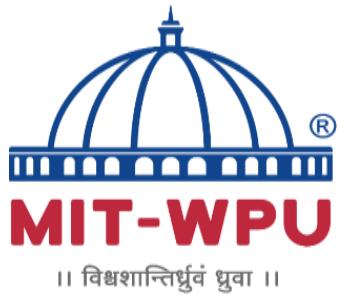
feature	importance
6 LDL	0.535831
4 TG	0.129787
5 HDL	0.108505
0 Unnamed: 0	0.071083
8 BUN	0.044667
7 Cr	0.041888
3 BMI	0.031020
1 Age	0.029461

Heuristic explanation & symptom suggestions

Observations:

- Unnamed: 0 is LOW (0.00 vs mean 2565.50)
- LDL is LOW (1.40 vs mean 2.91)

24°C Partly cloudy



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Navigation

- Go to
- Welcome
- Upload & Clean
- Train & Predict**
- Dashboards
- Settings

Index	Feature	Value
6	LDL	0.535831
4	TG	0.129787
5	HDL	0.108505
0	Unnamed: 0	0.071083
8	BUN	0.044667
7	Cr	0.041888
3	BMI	0.031020
1	Age	0.029461

Heuristic explanation & symptom suggestions

Observations:

- Unnamed: 0 is LOW (0.00 vs mean 2565.50)
- LDL is LOW (1.40 vs mean 2.91)

Symptom suggestions (educational, rule-based):

Unnamed: 0 – LOW

- Possible symptoms: lower-than-average value
- Suggested next steps: Consult specialist; Repeat test

LDL – LOW

- Possible symptoms: lower-than-average value
- Suggested next steps: Consult specialist; Repeat test

24°C Partly cloudy

Search

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	5132.000000	2565.500000	1481.625121	0.000000	1282.750000	2565.500000	3848.250000	5131.000000
Age	5132.000000	48.950312	14.048794	20.000000	36.000000	49.000000	59.000000	93.000000
Gender	5132.000000	0.634840	0.481926	0.000000	0.000000	1.000000	1.000000	2.000000
BMI	5132.000000	24.613406	4.277205	15.000000	22.000000	24.000000	27.000000	47.000000
Chol	5132.000000	4.866882	1.001052	0.000000	4.190000	4.800000	5.460000	11.650000
TG	5132.000000	1.719328	1.327057	0.000000	0.910000	1.380000	2.180000	32.640000
HDL	5132.000000	1.593305	1.038849	0.000000	1.090000	1.300000	1.590000	9.900000
LDL	5132.000000	2.914121	0.945423	0.300000	2.290000	2.790000	3.400000	9.900000
Cr	5132.000000	71.144800	28.494394	4.860753	58.000000	70.200000	81.600000	800.000000
BUN	5132.000000	4.890969	1.689206	0.500000	3.900000	4.720000	5.600000	38.900000
Diagnosis	5132.000000	0.388348	0.487422	0.000000	0.000000	0.000000	1.000000	1.000000

24°C Partly cloudy

Search



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Navigation

Go to

- Welcome
- Upload & Clean
- Train & Predict
- Dashboards
- Settings

Correlation heatmap (annotated)

	BUN	Diagnosis	Age	Gender	BMI	Chol	TG	HDL	LDL	Cr	BUN	Diagnosis
BUN	1.00	0.23	0.05	-0.20	0.06	-0.04	0.37	0.30	0.00	0.03	0.28	1.00
Diagnosis	5132.000000	0.388348	0.487422	0.050000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
Age	0.23	1.00	-0.01	0.28	0.21	0.18	0.17	0.22	0.02	0.21	0.57	-0.02
Gender	0.05	-0.01	1.00	0.15	-0.06	0.12	-0.06	-0.06	0.02	0.35	0.12	-0.02
BMI	-0.20	0.28	0.15	1.00	0.12	0.34	-0.04	0.03	0.06	0.11	0.43	-0.02
Chol	-0.06	0.21	-0.06	0.12	1.00	0.33	0.10	0.58	-0.01	0.07	0.13	-0.02
TG	-0.04	0.18	0.12	0.34	0.33	1.00	-0.00	0.09	0.06	0.05	0.24	-0.02
HDL	0.37	0.17	-0.06	-0.04	0.10	-0.00	1.00	0.55	-0.08	0.02	0.31	-0.02
LDL	0.30	0.22	0.02	0.03	0.58	0.09	0.55	1.00	-0.02	0.06	0.23	-0.02
Cr	0.00	0.02	0.35	0.06	-0.01	0.06	-0.08	-0.02	1.00	0.49	-0.04	-0.02
Diagnosis_BIN	-0.03	0.21	0.12	0.11	0.07	0.05	0.02	0.06	0.49	1.00	0.13	-0.02
Unnamed: 0	0.28	0.52	-0.02	0.43	0.13	0.24	0.31	0.23	-0.04	0.13	1.00	-0.02

Deploy :

Choose plots for exploration

Line plot (sequential)

Line plot (sequence)

Select numeric column for line plot

Chol

Line plot: Chol

Scatter plot

X-axis: Unnamed: 0

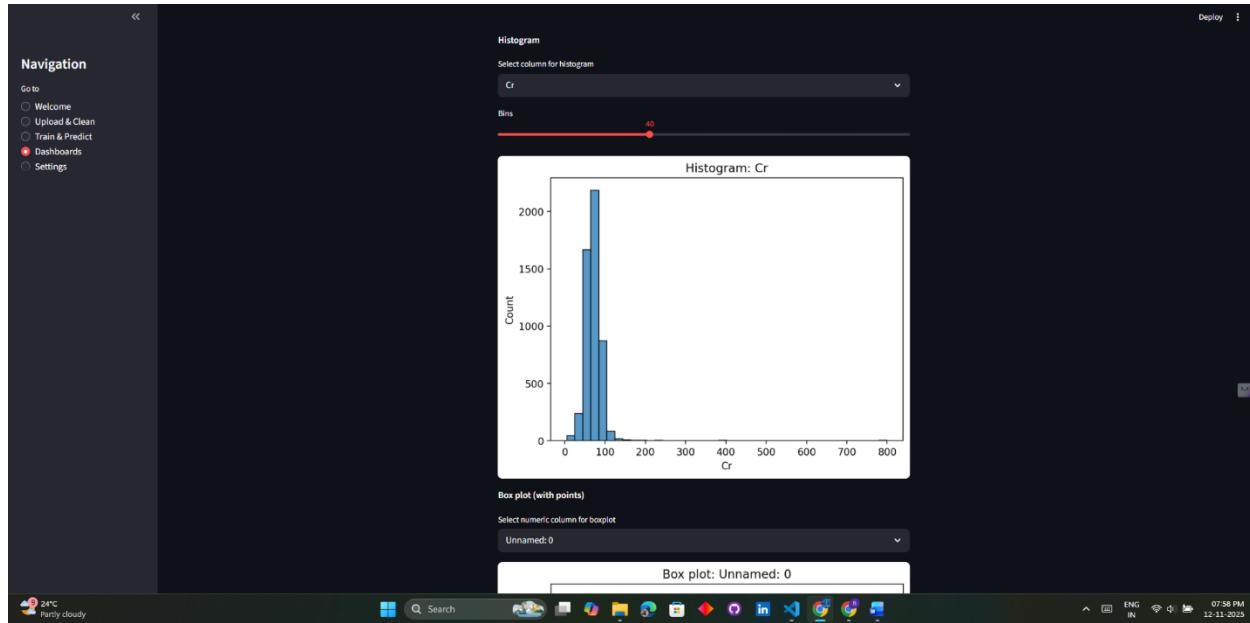
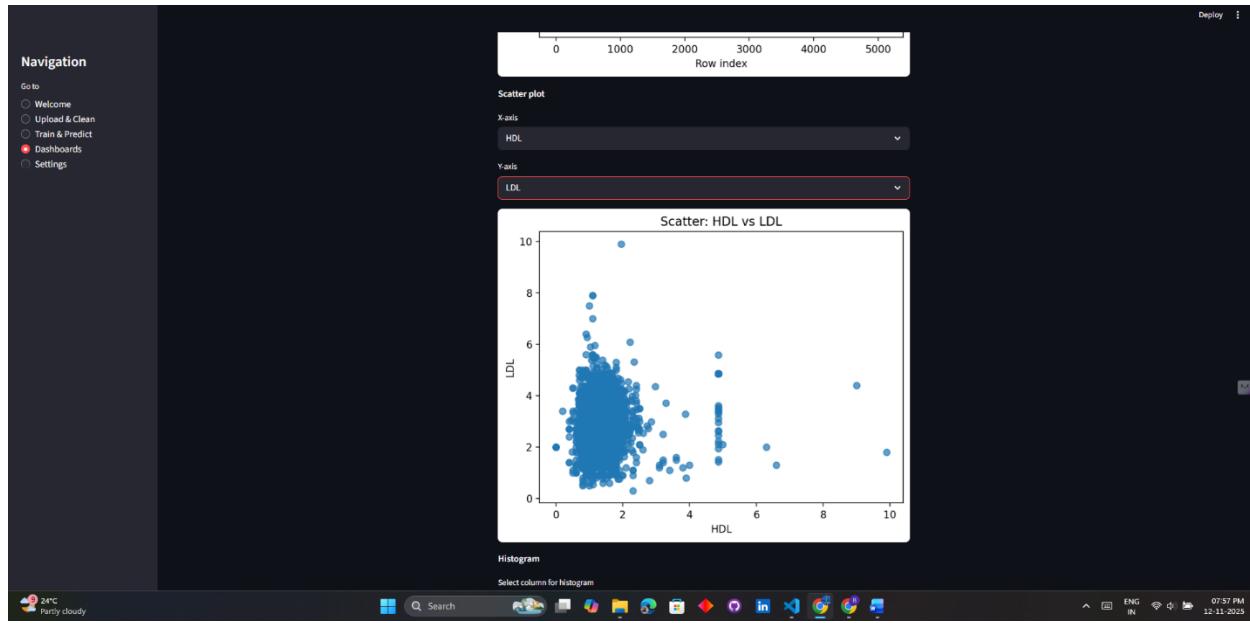
Y-axis: Age

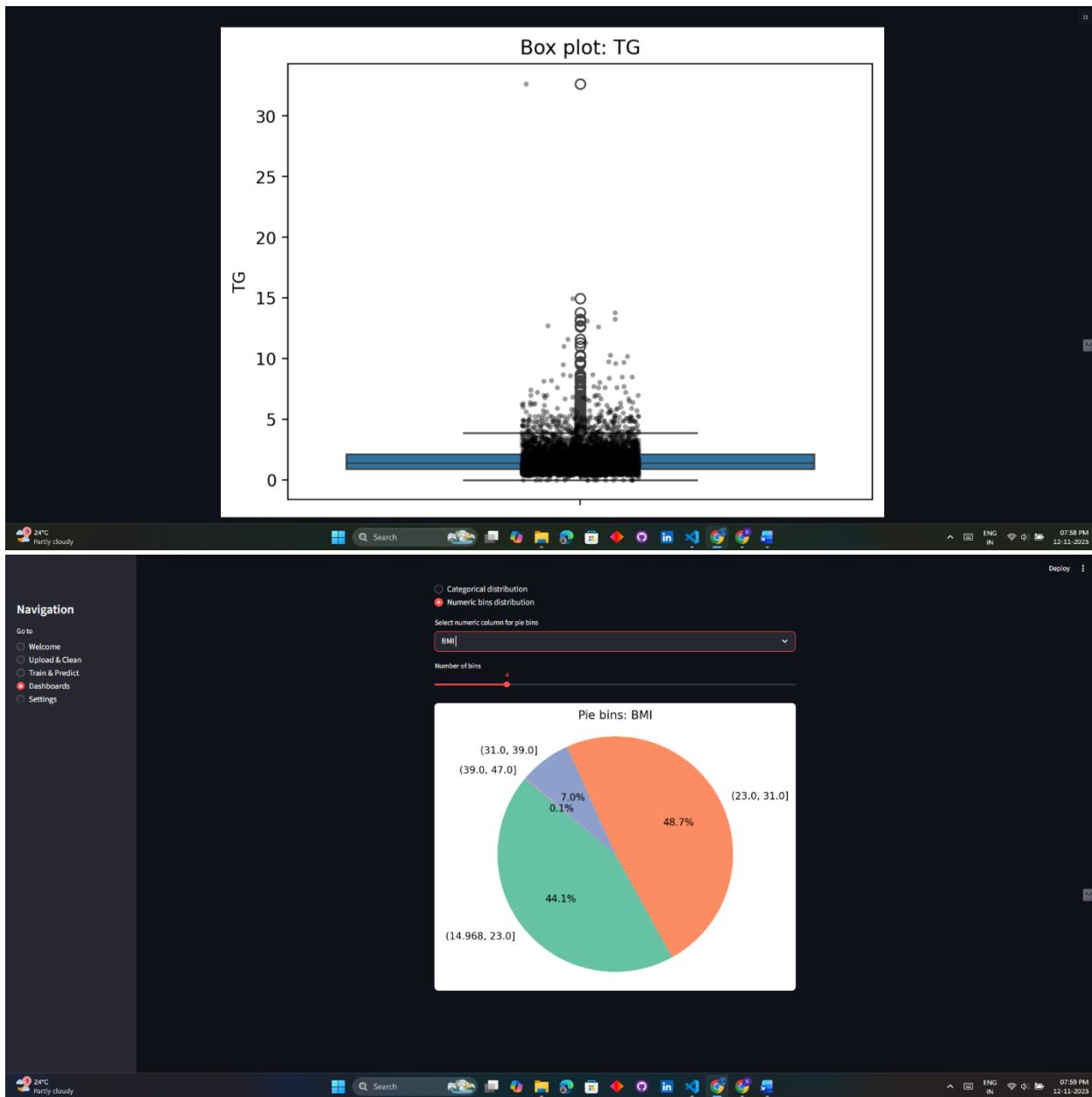
Deploy :



Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS







Results & Discussion

The analyzer successfully interprets uploaded blood datasets and visualizes all vital parameters.

Through conditional formatting, the system highlights abnormalities, helping in quick assessment of health conditions.

The correlation analysis and visual results support medical decision-making and improve the readability of data for non-experts.



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Interpretation of Blood Test Data

The system classifies and color-codes results:

- . **Green:** Normal range
- . **Yellow:** Slight deviation (monitor required)
- . **Red:** Critical value (medical attention suggested)

By comparing results across multiple parameters, users can easily identify health risks and focus areas.



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Limitations

- . Works only with structured dataset formats (CSV).*
- . Does not replace professional medical diagnosis.*
- . Currently analyzes limited types of blood parameters.*
- . Requires manual dataset updates for new records.*



Dr. Vishwanath Karad

**MIT WORLD PEACE
UNIVERSITY | PUNE**

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Future Scope / Perspective

- *Integration of predictive analytics using machine learning for early disease detection.*
- *Expansion to include other diagnostic test datasets (e.g., urine, ECG).*
- *Development of a mobile-friendly version for remote use.*
- *Integration with hospital or lab databases for real-time data visualization.*



Conclusion

The project “Everything About You in a Drop” demonstrates the potential of data science in healthcare.

By turning raw blood test data into interactive visuals, it bridges the gap between data and understanding.

The smart analyzer provides a modern, user-friendly approach to medical data interpretation, encouraging health awareness and data-driven medical decision-making.

Bibliography / References

1. *Python Documentation* —

<https://www.python.org/doc/>

2. *Streamlit Documentation* —

<https://docs.streamlit.io/>

3. *Pandas Library* —

<https://pandas.pydata.org/>

4. *Matplotlib Library* — <https://matplotlib.org/>

5. *WHO Blood Test Parameters* —

<https://www.who.int/>

6. *Kaggle Blood Test Dataset* —

<https://www.kaggle.com/>



Dr. Vishwanath Karad

MIT WORLD PEACE UNIVERSITY | PUNE

TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS