Project Title: Automated ETL + Dashboard System

Intern Name: KALIMI KANAKAVARAHALU

Date: June 2025

1.**Project Objective:**

To build an automated data science tool using Python and Streamlit which performs:

File upload (CSV/Excel)

ETL (missing value & outlier handling)

Summary statistics

Visual dashboards

Export cleaned data and visuals

2. Dataset Description:

Dataset: Iris Dataset

Total Rows: 150

Columns: sepal_length, sepal_width, petal_length, petal_width, species

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Figure 1: First 5 rows of the Iris dataset (df.head())

3. Missing Value Handling:

Strategy used: Mean / Median

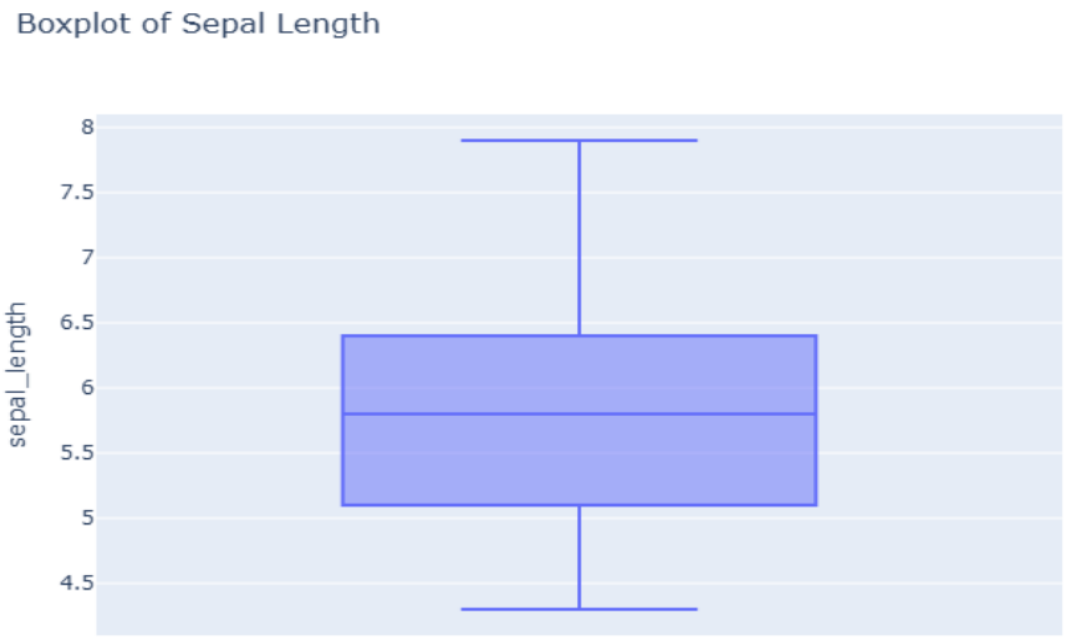
Code: df.fillna(df.mean())

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.848322	3.5	1.4	0.2	setosa
1	4.900000	3.0	1.4	0.2	setosa
2	4.700000	3.2	1.3	0.2	setosa
3	4.600000	3.1	1.5	0.2	setosa
4	5.000000	3.6	1.4	0.2	setosa

4. Outlier Detection:

Method: IQR Method

Columns checked: sepal_length, petal_length



5. Summary Statistics:

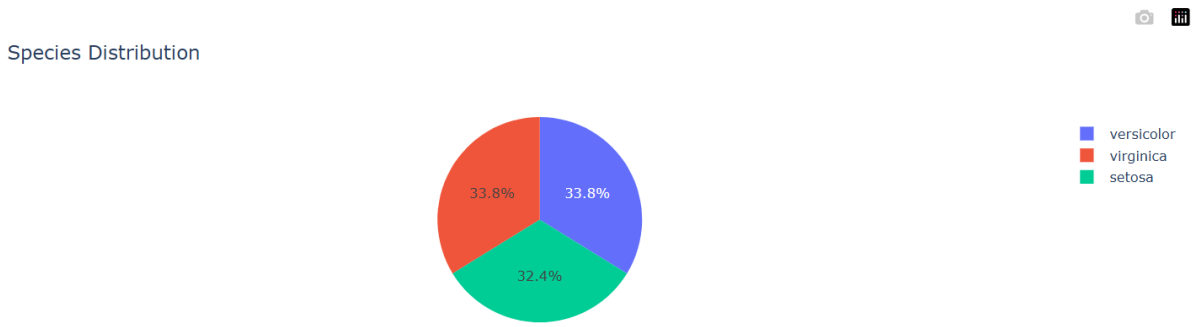
Use .describe() output:

Feature	Mean	Std	Skew	Kurtosis
sepal_length	**5.84**	**0.83**	**0.31**	**-0.61**

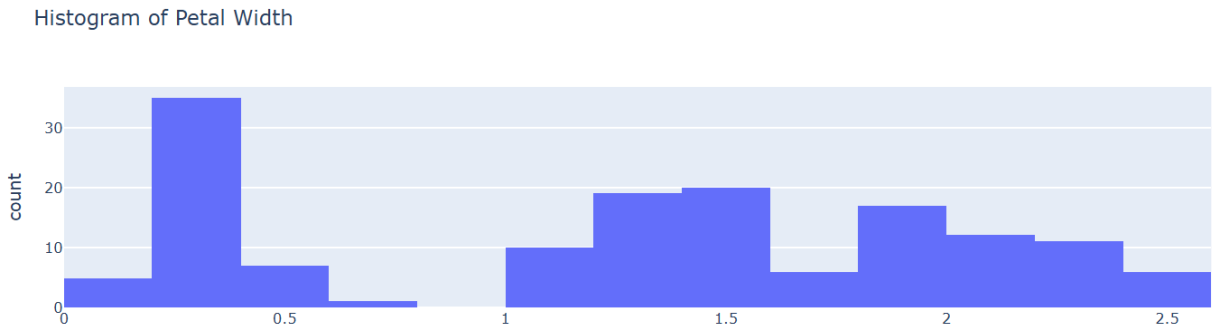
	sepal_length	sepal_width	petal_length	petal_width
count	149.000000	149.000000	149.000000	149.000000
mean	5.849318	3.048322	3.773154	1.211450
std	0.828504	0.423085	1.761435	0.757443
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.400000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.200000	6.900000	2.500000

6. Visualizations:

Pie chart for species



Histogram for sepal length



7. Exported Files:

- Cleaned data: cleaned_iris.csv

- Charts: stored in /assets/

8. Tools Used:

- Python 3.9
- Jupyter Notebook
- Pandas, NumPy, Plotly
- Streamlit (for app)
- MS Word (for report)

9. Learning Outcome:

How to clean and explore data automatically

How to generate plots using Plotly

How to present results in a report