Data Mining Lab-2(Graph Visualization)

DataSet used:

Iris DataSet :

https://drive.google.com/file/d/1SNvNLqlS889_BnExgj_WvZNCvb6EIbHH/view?usp=drive_link

Source Code:

```
df.hist(color='red',edgecolor='black',bins=10,cumulative=False)
plt.suptitle("HISTOGRAM")
plt.tight_layout()
plt.legend()
plt.show()
df.boxplot(color='red',figsize=(10,10))
import seaborn as sns
sns.pairplot(df, hue='species')
corr=df.select_dtypes(include='number').corr()
sns.heatmap(corr,annot=True,linecolor='red',linewidths=4)
list_sp = df['species'].unique()

for species_name in list_sp:
  species_df = df[df['species'] == species_name]
  print(species_name)
  for column in df.select_dtypes(include=np.number).columns:
    plt.figure(figsize=(5, 5))
    sns.histplot(data=species_df, x=column, kde=True, bins=10)
    plt.title(f'{column} for {species_name}')
    plt.show()
list_sp = df['species'].unique()

for species_name in list_sp:
  species_df = df[df['species'] == species_name]
  print(species_name)
  for column in df.select_dtypes(include=np.number).columns:
    plt.figure(figsize=(5, 5))
    sns.boxplot(y=species_df[column])
    plt.title(f'{column} for {species_name}')
    plt.show()

import matplotlib.gridspec as gridspec
```

```
def generate_species_plots_on_axes(species_df, hist_ax, box_ax,
heatmap_ax):
    numerical_cols = species_df.select_dtypes(include=np.number).columns
    species_name = species_df['species'].iloc[0]

    if numerical_cols.size > 0:
        sns.histplot(data=species_df, x=numerical_cols[0], kde=True,
bins=10, ax=hist_ax)
        hist_ax.set_title(f'Histogram - {numerical_cols[0]}')

    sns.boxplot(data=species_df[numerical_cols], ax=box_ax)
    box_ax.set_title(f'Boxplot')

    corr_matrix = species_df[numerical_cols].corr()
    sns.heatmap(corr_matrix, annot=True, linecolor='red', linewidths=4,
ax=heatmap_ax)
    heatmap_ax.set_title(f'Correlation Heatmap')
list_sp = df['species'].unique()

for species_name in list_sp:
    species_df = df[df['species'] == species_name].copy()

    fig = plt.figure(figsize=(15, 10))
    gs = gridspec.GridSpec(2, 2, figure=fig)

    ax_hist = fig.add_subplot(gs[0, 0])
    ax_box = fig.add_subplot(gs[0, 1])
    ax_heatmap = fig.add_subplot(gs[1, 0])

    generate_species_plots_on_axes(species_df, ax_hist, ax_box,
ax_heatmap)
    fig.suptitle(f'Analysis for {species_name}', y=1, fontsize=16)

    plt.tight_layout()
    plt.show()


    pair_plot = sns.pairplot(species_df, hue='species')
    pair_plot.fig.suptitle(f'Pair Plot for {species_name}', y=
```

Screenshots:

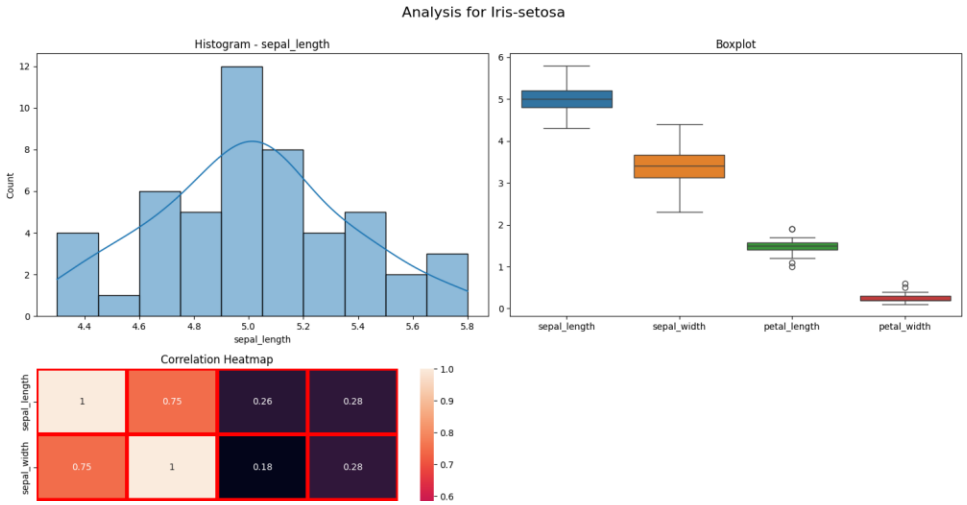
DM_Lab2.ipynb

File Edit View Insert Runtime Tools Help

Share Gemini

Commands + Code + Text ▷ Run all Connect

```
pair_plot = sns.pairplot(species_df, hue='species')
pair_plot.fig.suptitle(f'Pair Plot for {species_name}', y=1)
plt.show()
```



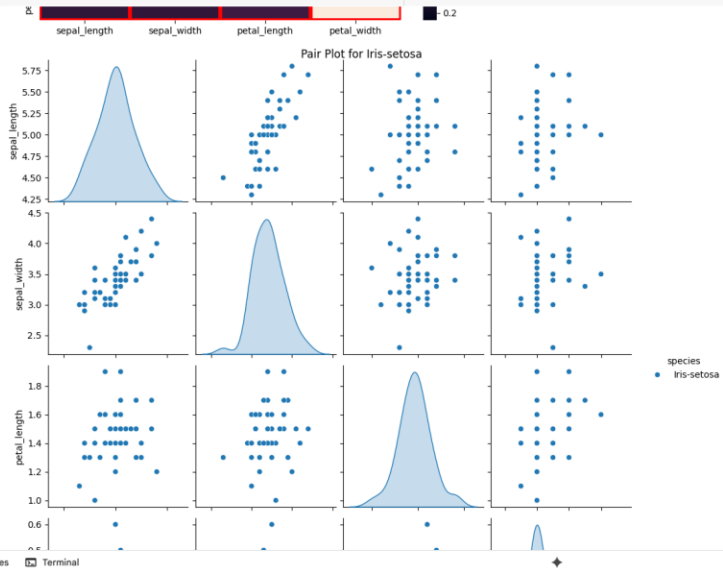
DM_Lab2.ipynb

File Edit View Insert Runtime Tools Help

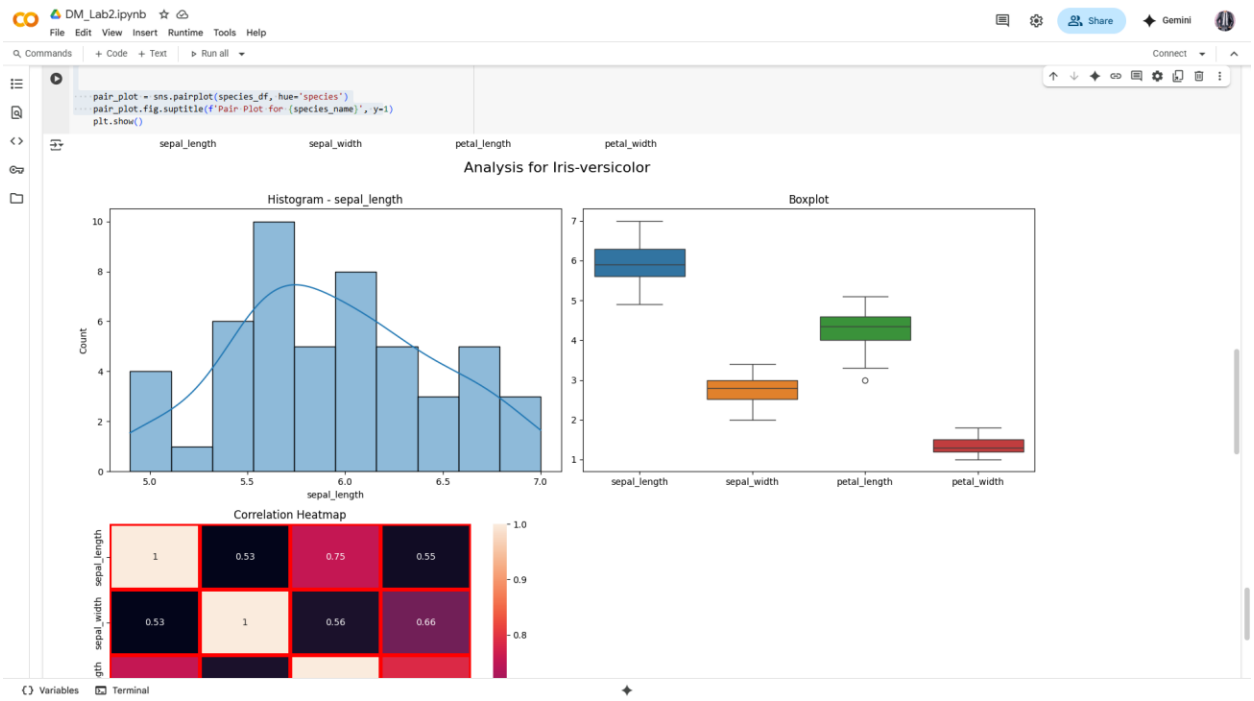
Share Gemini

Commands + Code + Text ▷ Run all Connect

```
pair_plot = sns.pairplot(species_df, hue='species')
pair_plot.fig.suptitle(f'Pair Plot for {species_name}', y=1)
plt.show()
```



Variables Terminal



Observations:

- **Histograms**

Show the **distribution of numeric features** (runs, wickets, strike rate, etc. if cricket OR petal/sepal if Iris).

You can observe **whether features are normally distributed, skewed, or uniform**.

Peaks in histogram indicate **common range**.

- **Boxplots**

Help identify **outliers** in the numeric features.

You can compare **spread and median values** across features.

E.g., in cricket dataset → "Some players have extremely high strike rates/wickets, visible as outliers."

- **Pairplot**

Shows **relationships between all pairs of numerical variables**.

Using `hue="species"`, you can see if groups (teams, player types, or Iris species) form **clusters**.

Example: In Iris, *setosa* separates clearly on petal length/width; in cricket, “bowlers” vs “batsmen” might separate by economy rate vs strike rate.

- **Correlation Heatmap**

Displays **strength of linear relationships** between numerical variables.