❖ Day-6 Data Visualization:

❖ Visualization:

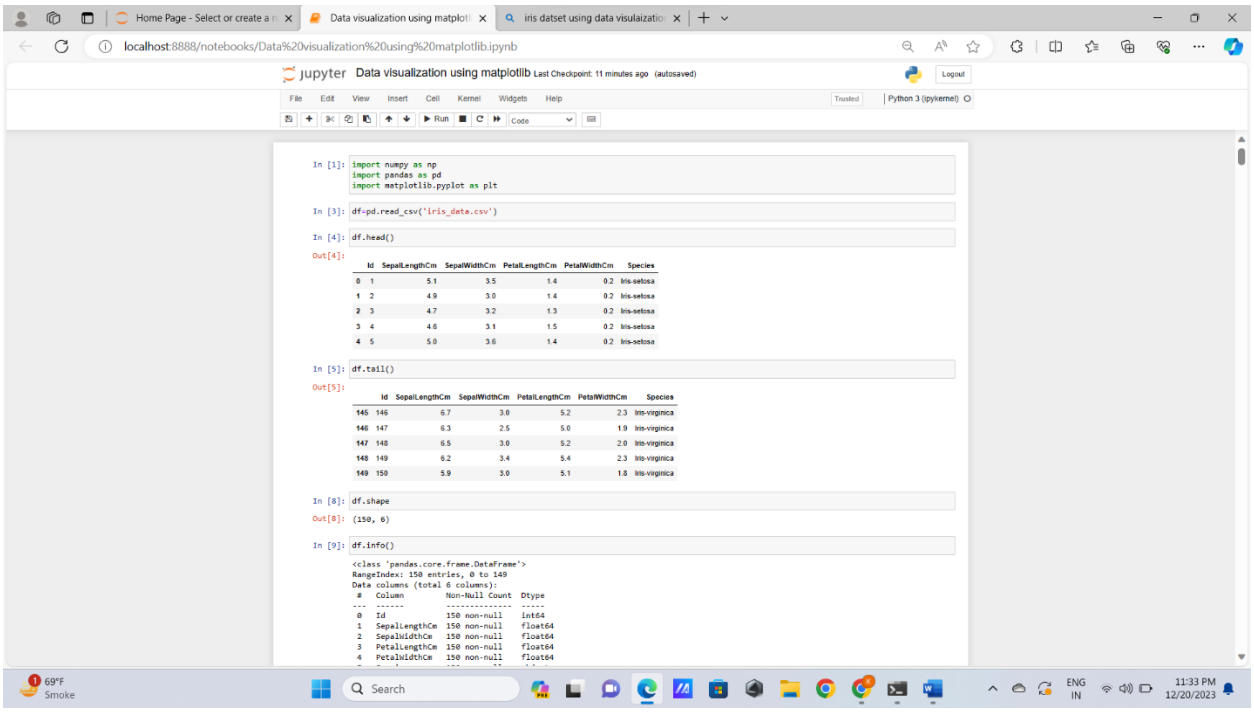
Data visualization is a crucial aspect of machine learning that enables analysts to understand and make sense of data patterns, relationships, and trends.

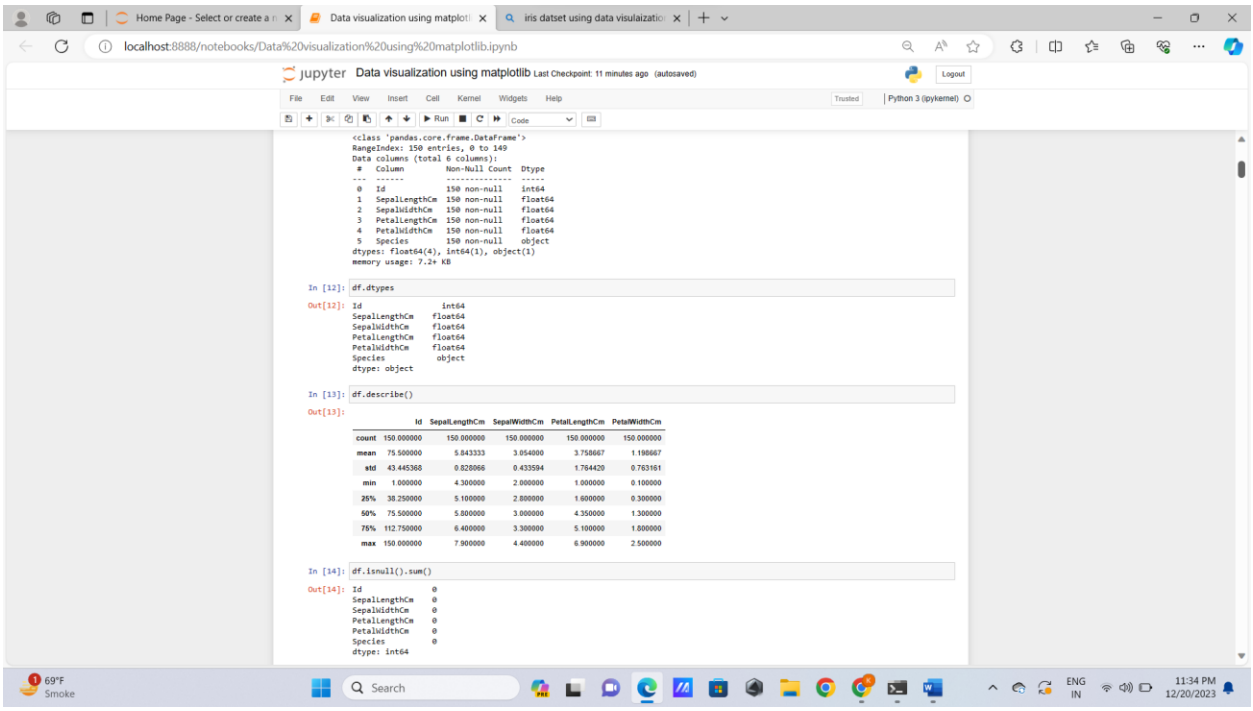
❖ advantages of data visualization include:

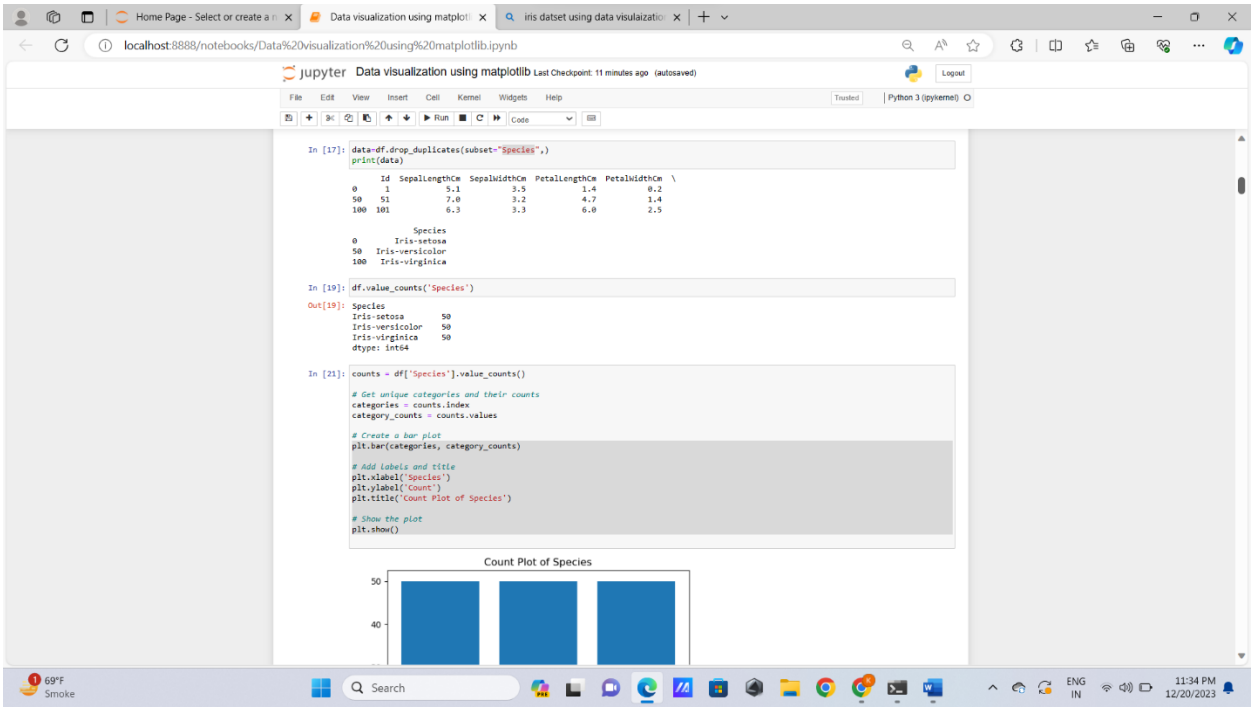
- Easily sharing information.
- Interactively explore opportunities.
- Visualize patterns and relationships.
-

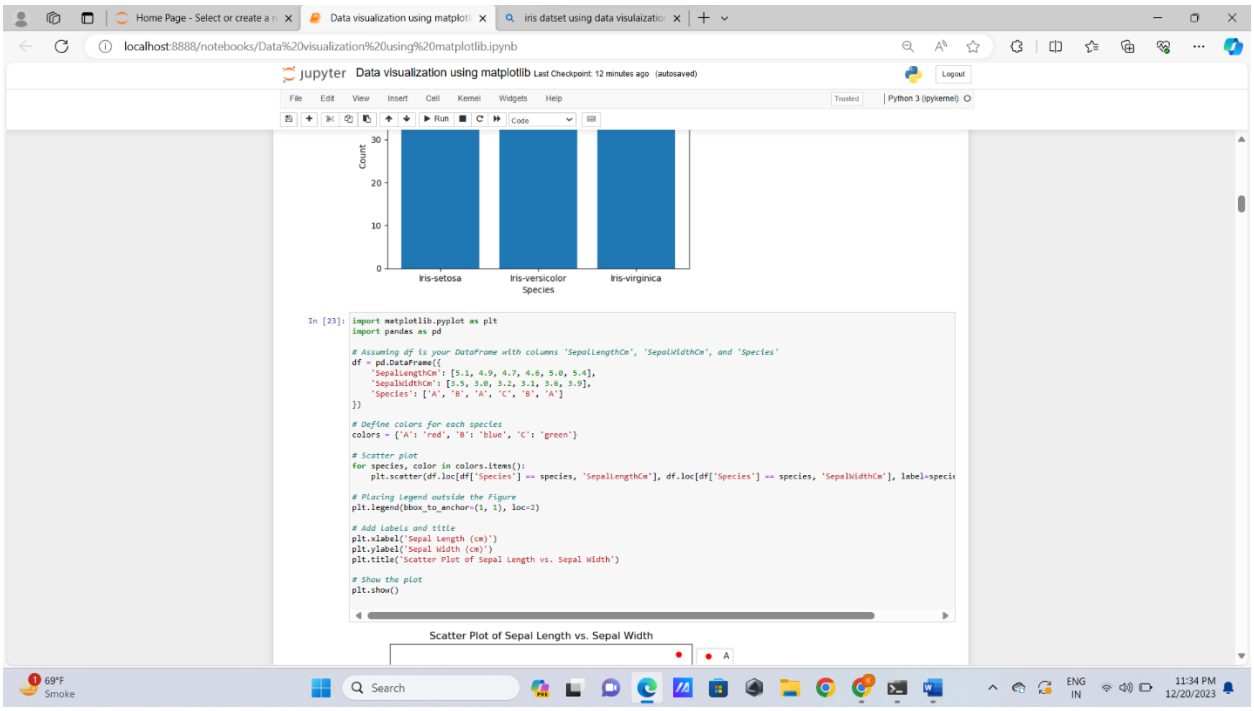
❖ disadvantages :

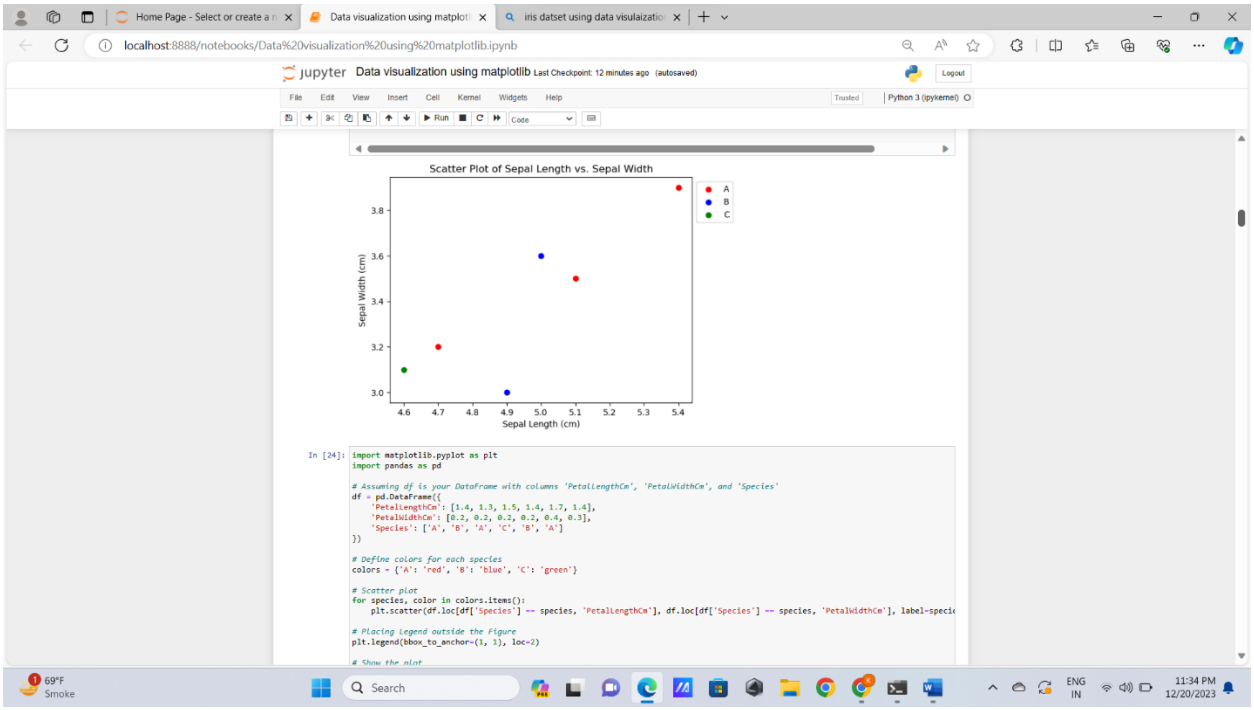
- Biased or inaccurate information.
- Correlation doesn’t always mean causation.
- Core messages can get lost in translation.

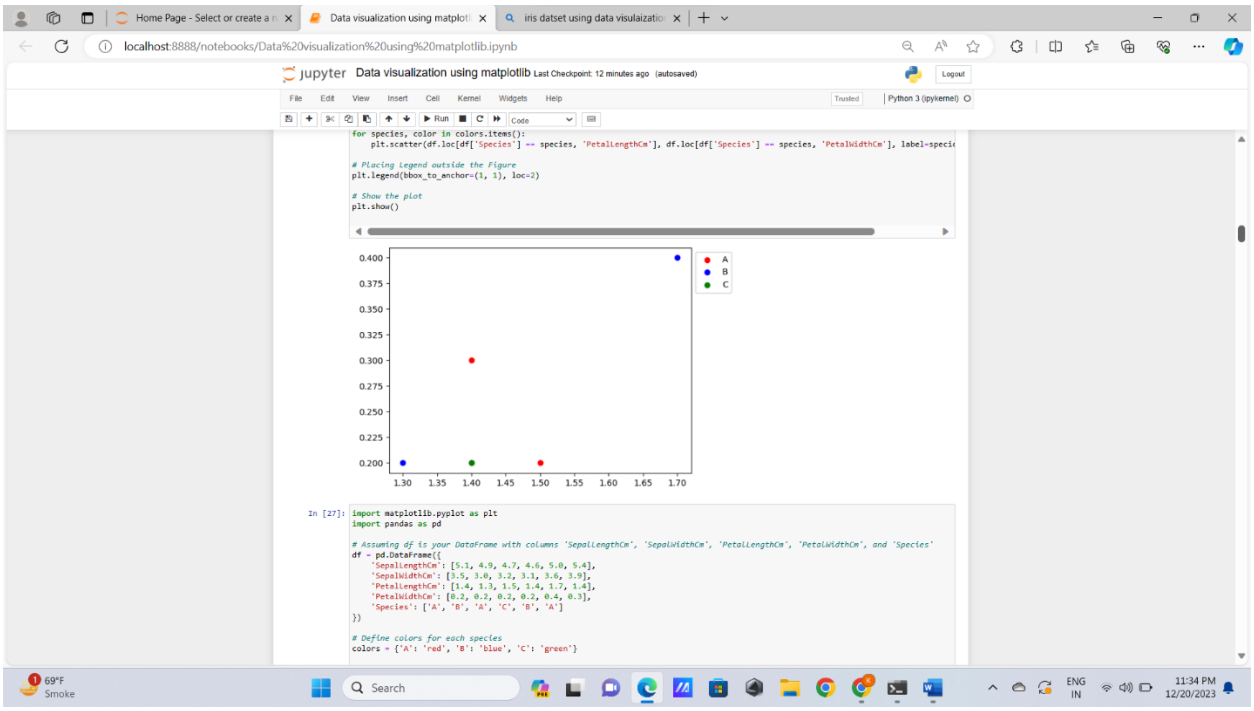


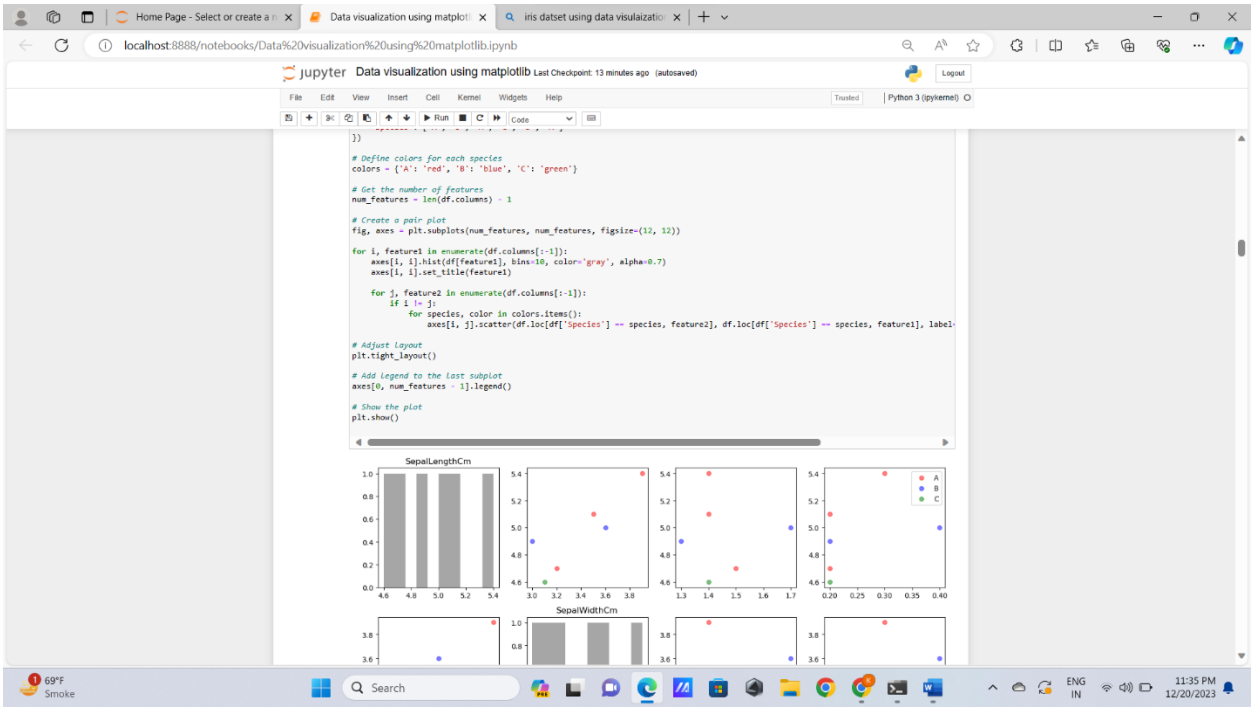

Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 11 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
Column Non-Null Count Dtype
0 Id 150 non-null int64
1 SepalLengthCm 150 non-null float64
2 SepalWidthCm 150 non-null float64
3 PetalLengthCm 150 non-null float64
4 PetalWidthCm 150 non-null float64
5 Species 150 non-null object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
In [12]: df.dtypes
Out[12]: Id int64
SepalLengthCm float64
SepalWidthCm float64
PetalLengthCm float64
PetalWidthCm float64
Species object
dtype: object
In [13]: df.describe()
Out[13]:
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
count 150.000000 150.000000 150.000000 150.000000 150.000000
mean 75.500000 5.843333 3.054000 3.758667 1.198667
std 43.445368 0.828066 0.433594 1.764420 0.763161
min 1.000000 4.300000 2.000000 1.000000 0.100000
25% 38.250000 5.100000 2.800000 1.600000 0.300000
50% 75.500000 5.800000 3.000000 4.350000 1.300000
75% 112.750000 6.400000 3.300000 5.100000 1.800000
max 150.000000 7.900000 4.400000 6.900000 2.500000
In [14]: df.isnull().sum()
Out[14]: Id 0
SepalLengthCm 0
SepalWidthCm 0
PetalLengthCm 0
PetalWidthCm 0
Species 0
dtype: int64
69°F
Smoke
Search
ENG
IN
11:34 PM
12/20/2023

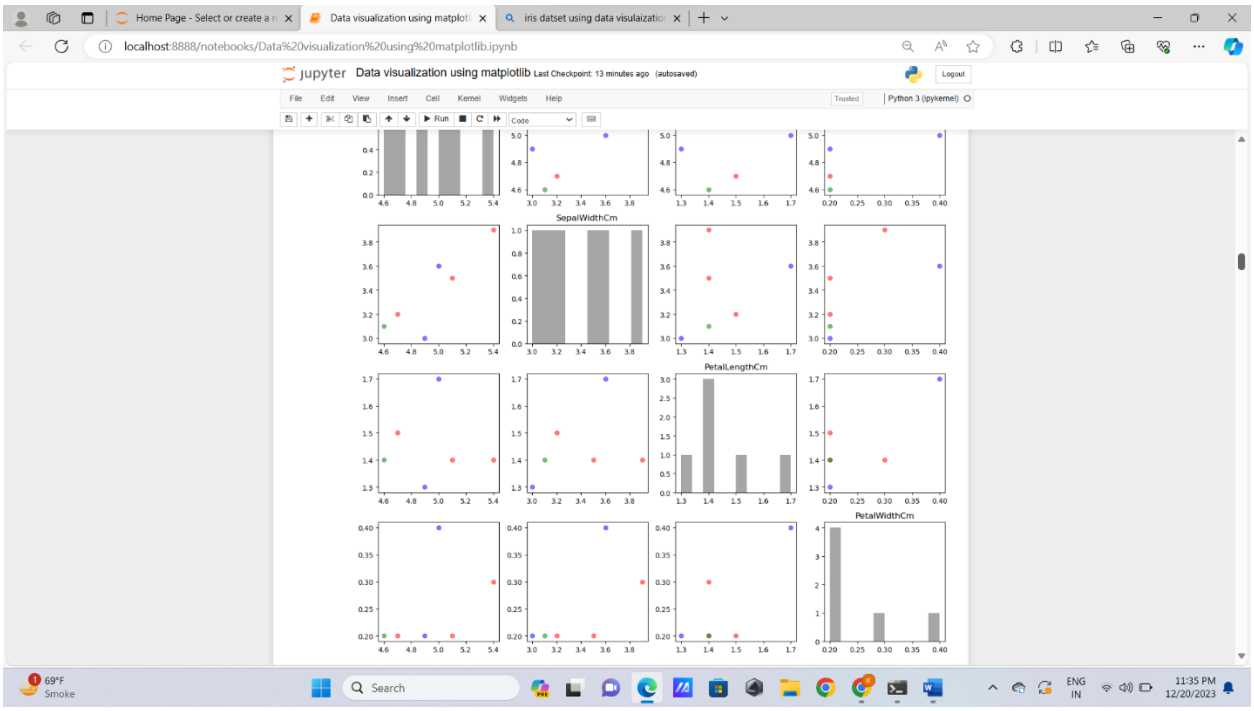

Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 11 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [17]: data=df.drop_duplicates(subset="Species",)
print(data)
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm \
0 1 5.1 3.5 1.4 0.2
50 51 7.0 3.2 4.7 1.4
100 101 6.3 3.3 6.0 2.5
Species
0 Iris-setosa
50 Iris-versicolor
100 Iris-virginica
In [19]: df.value_counts('Species')
Out[19]: Species
Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
dtype: int64
In [21]: counts = df['Species'].value_counts()
# Get unique categories and their counts
categories = counts.index
category_counts = counts.values
# Create a bar plot
plt.bar(categories, category_counts)
# Add labels and title
plt.xlabel('Species')
plt.ylabel('Count')
plt.title('Count Plot of Species')
# Show the plot
plt.show()
Count Plot of Species
50
40
69°F
Smoke
Search
ENG
IN
11:34 PM
12/20/2023

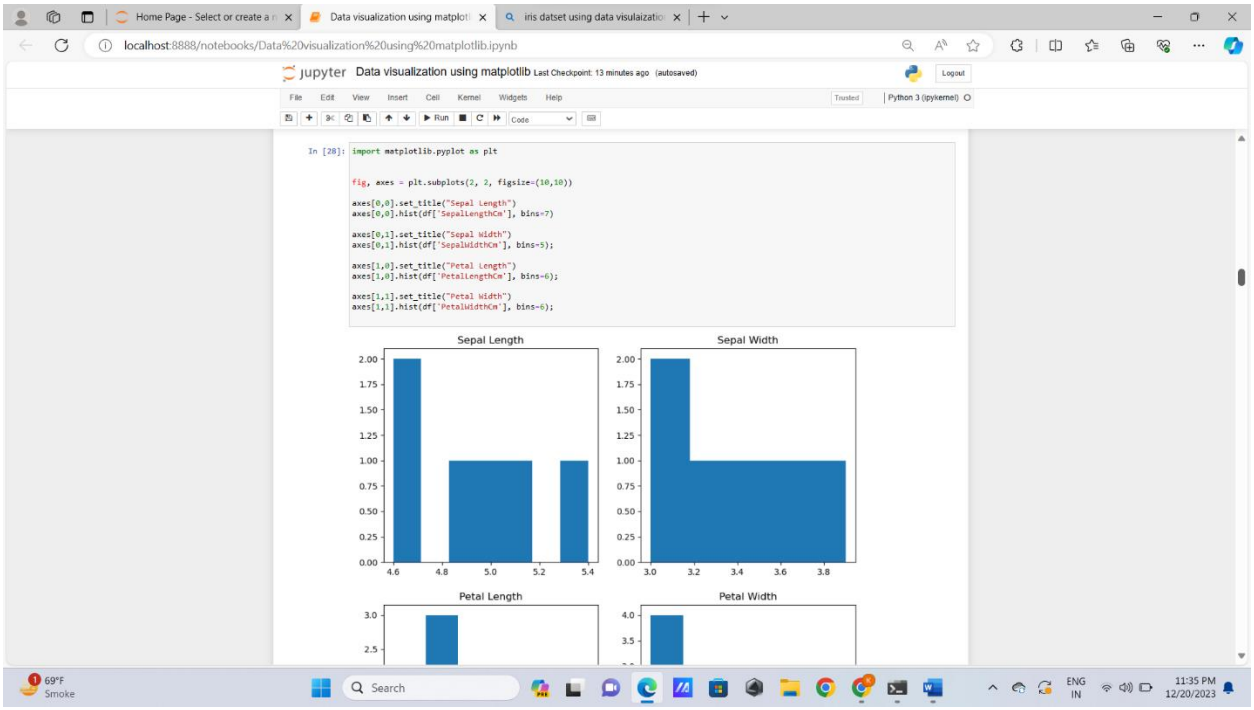

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 12 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Count
Iris-setosa
Iris-versicolor
Iris-virginica
Species
In [23]:
import matplotlib.pyplot as plt
import pandas as pd
# Assuming df is your DataFrame with columns 'SepalLengthCm', 'SepalWidthCm', and 'Species'
df = pd.DataFrame({
'SepalLengthCm': [5.1, 4.9, 4.7, 4.6, 5.0, 5.4],
'SepalWidthCm': [3.5, 3.0, 3.2, 3.1, 3.6, 3.9],
'Species': ['A', 'B', 'A', 'C', 'B', 'A']
})
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
# Scatter plot
for species, color in colors.items():
plt.scatter(df.loc[df['Species'] == species, 'SepalLengthCm'], df.loc[df['Species'] == species, 'SepalWidthCm'], label=speci
# Placing Legend outside the Figure
plt.legend(bbox_to_anchor=(1, 1), loc=2)
# Add labels and title
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.title('Scatter Plot of Sepal Length vs. Sepal Width')
# Show the plot
plt.show()
Scatter Plot of Sepal Length vs. Sepal Width
A
69°F
Smoke
Search
ENG
IN
11:34 PM
12/20/2023

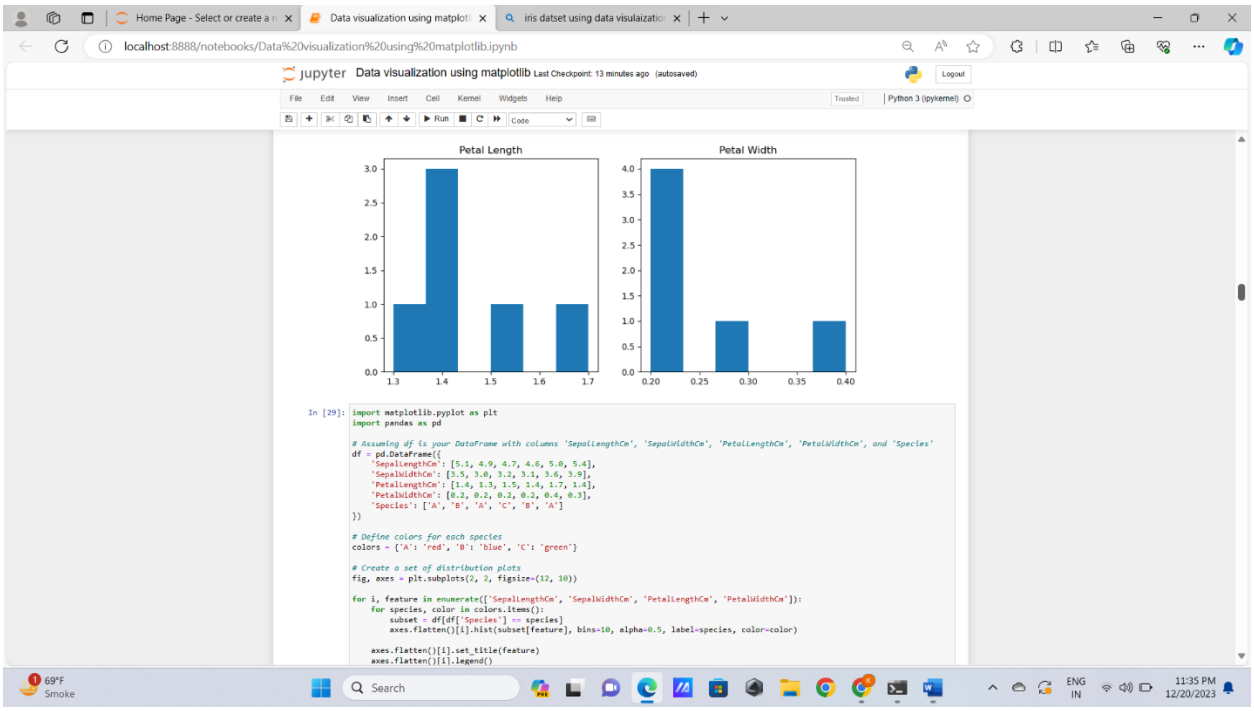

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 12 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Scatter Plot of Sepal Length vs. Sepal Width
A
B
C
Sepal Width (cm)
Sepal Length (cm)
In [24]:
import matplotlib.pyplot as plt
import pandas as pd
# Assuming df is your DataFrame with columns 'PetalLengthCm', 'PetalWidthCm', and 'Species'
df = pd.DataFrame({
'PetalLengthCm': [1.4, 1.3, 1.5, 1.4, 1.7, 1.4],
'PetalWidthCm': [0.2, 0.2, 0.2, 0.2, 0.4, 0.3],
'Species': ['A', 'B', 'A', 'C', 'B', 'A']
})
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
# Scatter plot
for species, color in colors.items():
plt.scatter(df.loc[df['Species'] == species, 'PetalLengthCm'], df.loc[df['Species'] == species, 'PetalWidthCm'], label=speci
# Placing Legend outside the Figure
plt.legend(bbox_to_anchor=(1, 1), loc=2)
69°F
Smoke
Search
ENG
IN
11:34 PM
12/20/2023

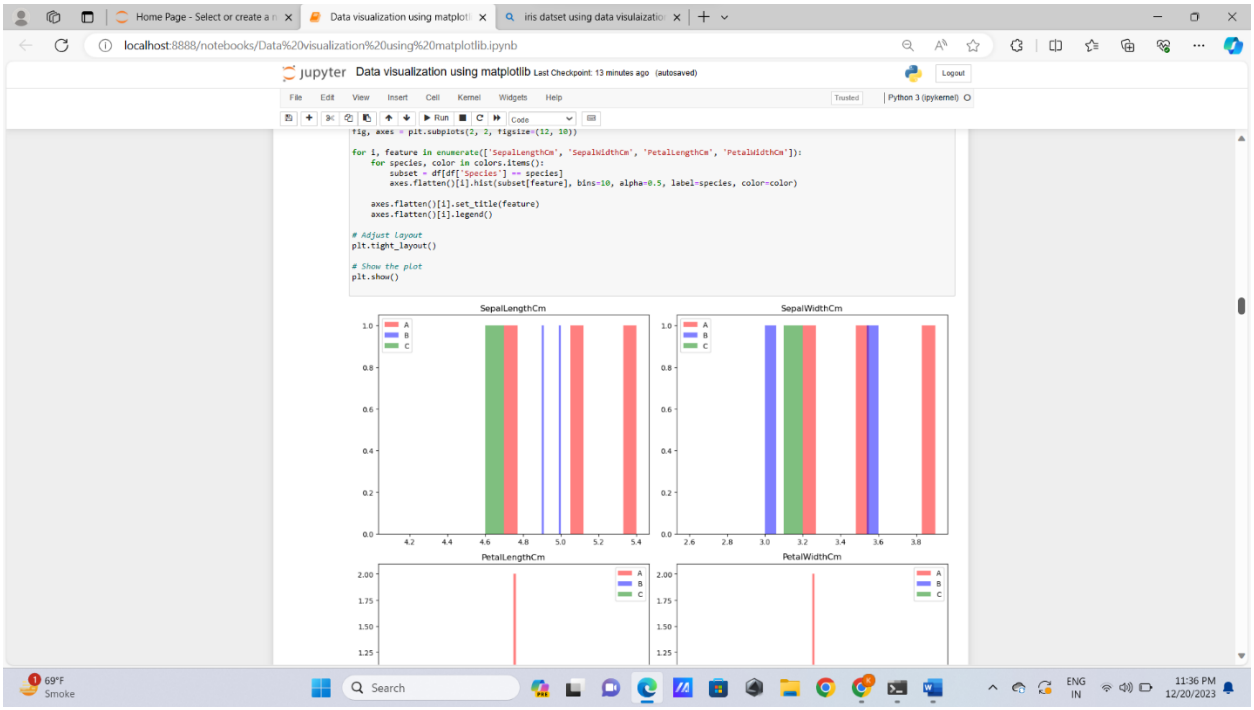

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 12 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
for species, color in colors.items():
plt.scatter(df.loc[df['Species'] == species, 'PetalLengthCm'], df.loc[df['Species'] == species, 'PetalWidthCm'], label=speci
# Placing Legend outside the Figure
plt.legend(bbox_to_anchor=(1, 1), loc=2)
# Show the plot
plt.show()
A
B
C
In [27]:
import matplotlib.pyplot as plt
import pandas as pd
# Assuming df is your DataFrame with columns 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', and 'Species'
df = pd.DataFrame({
'SepalLengthCm': [5.1, 4.9, 4.7, 4.6, 5.0, 5.4],
'SepalWidthCm': [3.5, 3.0, 3.2, 3.1, 3.6, 3.9],
'PetalLengthCm': [1.4, 1.3, 1.5, 1.4, 1.7, 1.4],
'PetalWidthCm': [0.2, 0.2, 0.2, 0.2, 0.4, 0.3],
'Species': ['A', 'B', 'A', 'C', 'B', 'A']
})
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
69°F
Smoke
Search
ENG
IN
11:34 PM
12/20/2023

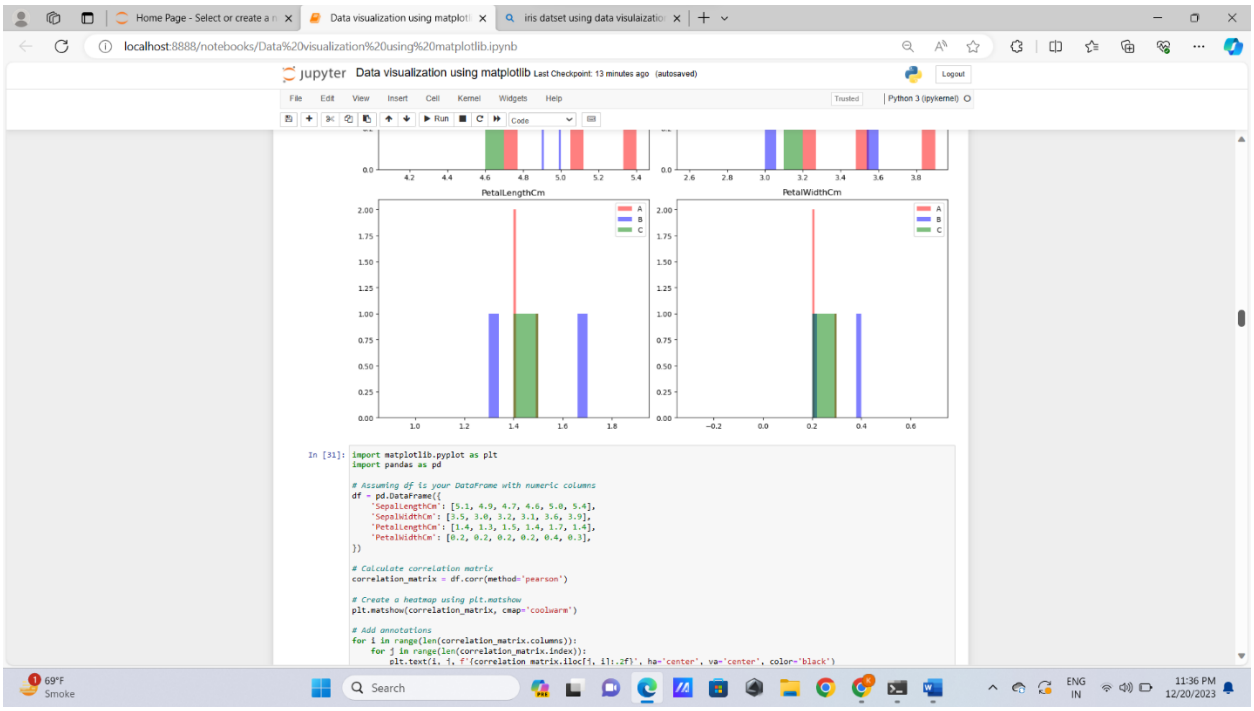

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
# Get the number of features
num_features = len(df.columns) - 1
# Create a pair plot
fig, axes = plt.subplots(num_features, num_features, figsize=(12, 12))
for i, feature1 in enumerate(df.columns[:-1]):
axes[i, i].hist(df[feature1], bins=10, color='gray', alpha=0.7)
axes[i, i].set_title(feature1)
for j, feature2 in enumerate(df.columns[:-1]):
if i != j:
for species, color in colors.items():
axes[i, j].scatter(df.loc[df['Species'] == species, feature2], df.loc[df['Species'] == species, feature1], label=
# Adjust layout
plt.tight_layout()
# Add legend to the last subplot
axes[0, num_features - 1].legend()
# Show the plot
plt.show()
SepalLengthCm
SepalWidthCm
A
B
C
69°F
Smoke
Search
ENG
IN
11:35 PM
12/20/2023

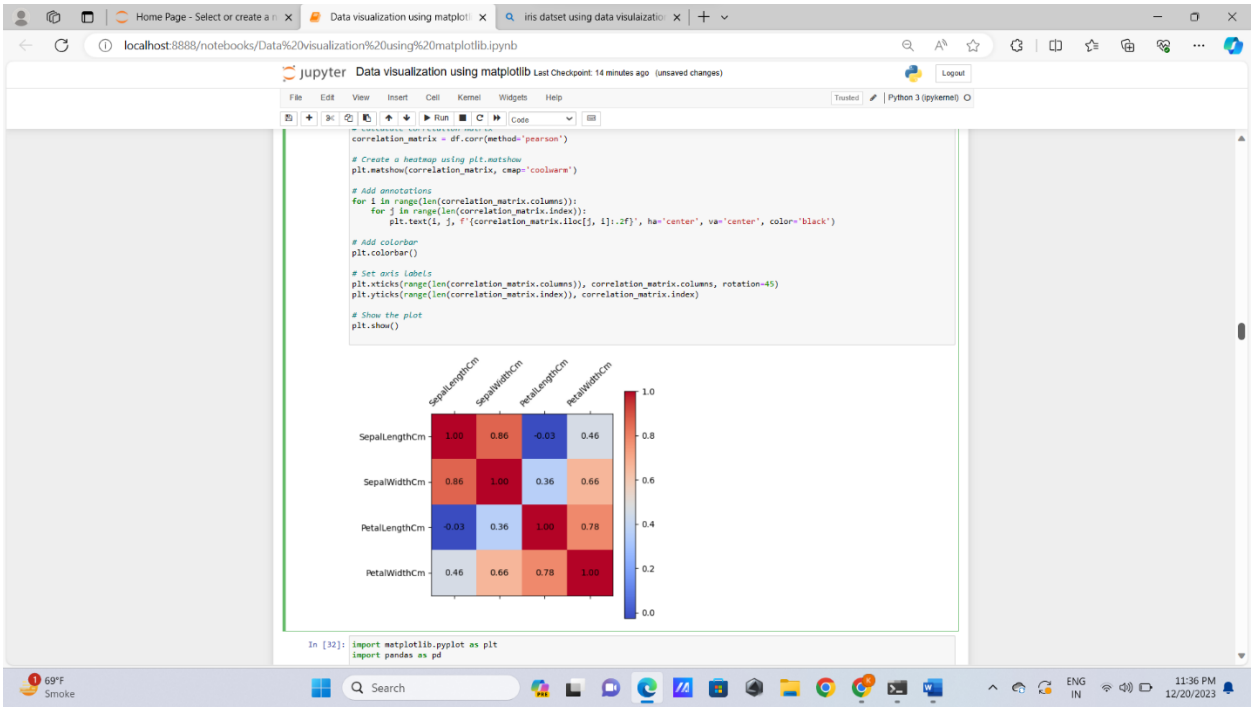

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
SepalWidthCm
PetalLengthCm
PetalWidthCm
69°F
Smoke
Search
ENG
IN
11:35 PM
12/20/2023

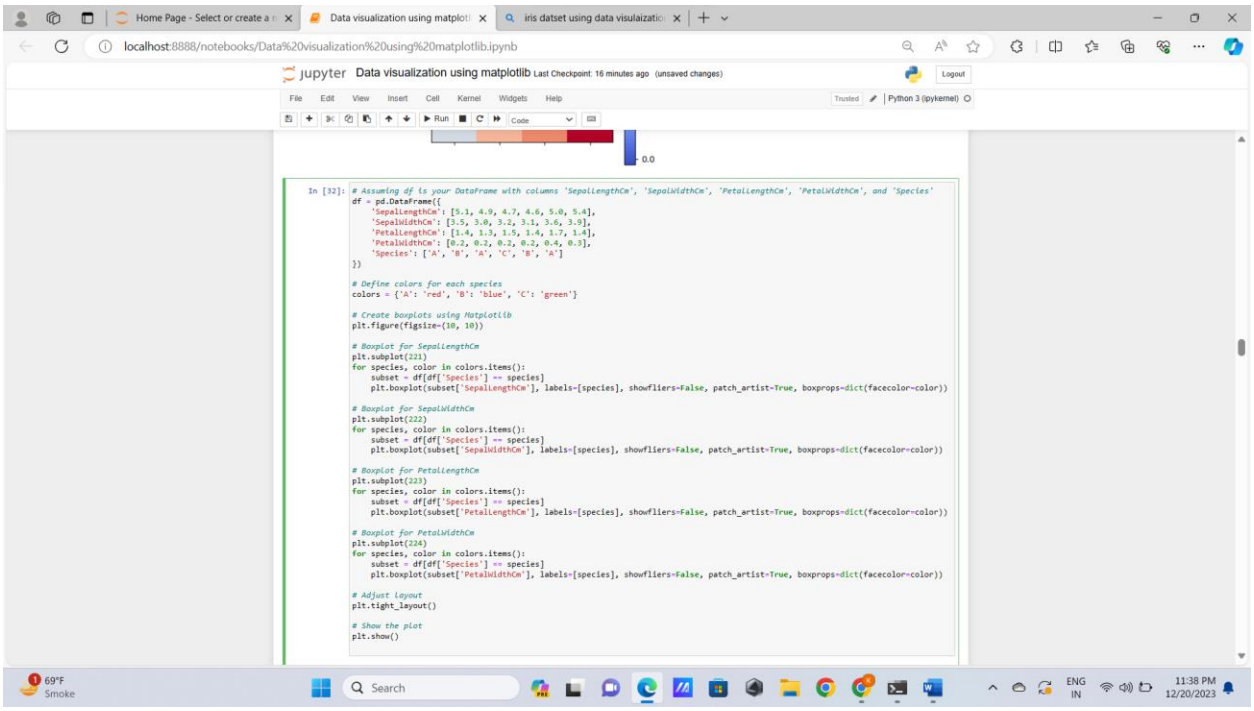

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [28]:
import matplotlib.pyplot as plt
fig, axes = plt.subplots(2, 2, figsize=(10,10))
axes[0,0].set_title("Sepal Length")
axes[0,0].hist(df['SepalLengthCm'], bins=7)
axes[0,1].set_title("Sepal Width")
axes[0,1].hist(df['SepalWidthCm'], bins=5);
axes[1,0].set_title("Petal Length")
axes[1,0].hist(df['PetalLengthCm'], bins=6);
axes[1,1].set_title("Petal Width")
axes[1,1].hist(df['PetalWidthCm'], bins=6);
Sepal Length
Sepal Width
Petal Length
Petal Width
69°F
Smoke
Search
ENG
IN
11:35 PM
12/20/2023

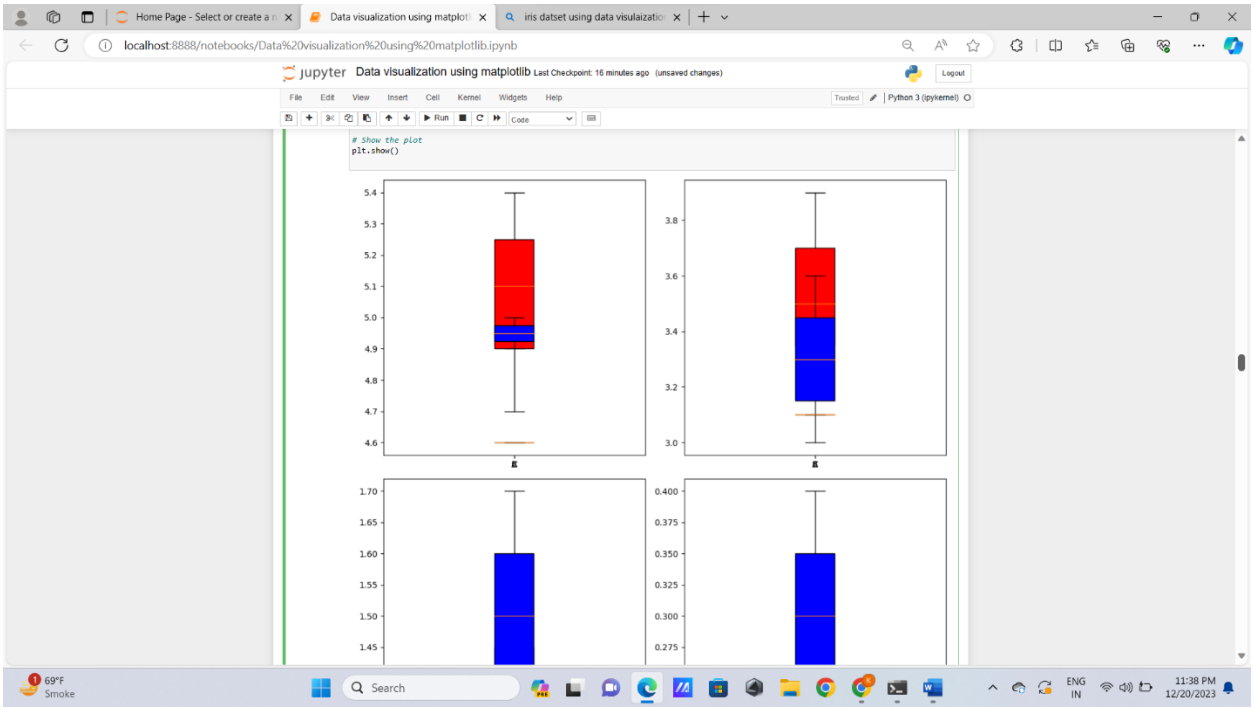

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Petal Length
Petal Width
In [29]:
import matplotlib.pyplot as plt
import pandas as pd
# Assuming df is your DataFrame with columns 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', and 'Species'
df = pd.DataFrame({
'SepalLengthCm': [5.1, 4.9, 4.7, 4.6, 5.0, 5.4],
'SepalWidthCm': [3.5, 3.0, 3.2, 3.1, 3.6, 3.9],
'PetalLengthCm': [1.4, 1.3, 1.5, 1.4, 1.7, 1.4],
'PetalWidthCm': [0.2, 0.2, 0.2, 0.2, 0.4, 0.3],
'Species': ['A', 'B', 'A', 'C', 'B', 'A']
})
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
# Create a set of distribution plots
fig, axes = plt.subplots(2, 2, figsize=(12, 10))
for i, feature in enumerate(['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']):
for species, color in colors.items():
subset = df[df['Species'] == species]
axes.flatten()[i].hist(subset[feature], bins=10, alpha=0.5, label=species, color=color)
axes.flatten()[i].set_title(feature)
axes.flatten()[i].legend()
69°F
Smoke
Search
ENG
IN
11:35 PM
12/20/2023

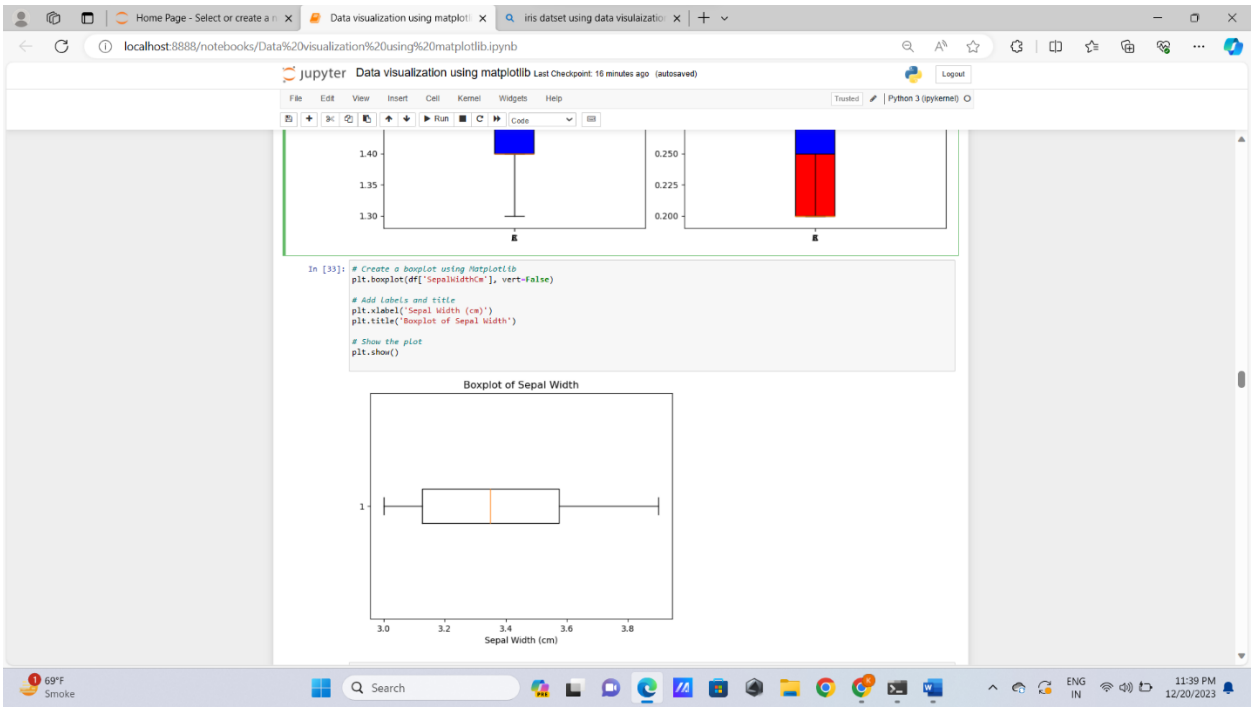

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
fig, axes = plt.subplots(2, 2, figsize=(12, 10))
for i, feature in enumerate(['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']):
for species, color in colors.items():
subset = df[df['Species'] == species]
axes.flatten()[i].hist(subset[feature], bins=10, alpha=0.5, label=species, color=color)
axes.flatten()[i].set_title(feature)
axes.flatten()[i].legend()
# Adjust layout
plt.tight_layout()
# Show the plot
plt.show()
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
A
B
C
69°F
Smoke
Search
ENG
IN
11:36 PM
12/20/2023

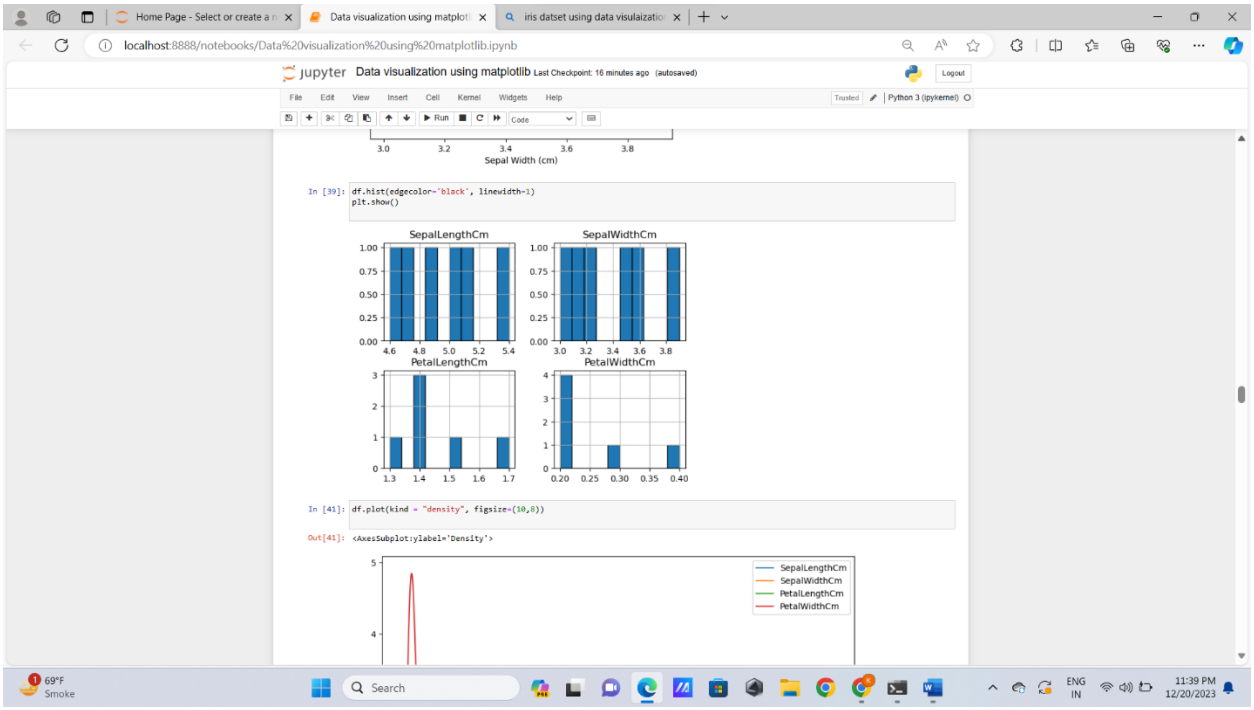

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 13 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
PetalLengthCm
PetalWidthCm
A
B
C
In [31]:
import matplotlib.pyplot as plt
import pandas as pd
# Assuming df is your DataFrame with numeric columns
df = pd.DataFrame({
'SepalLengthCm': [5.1, 4.9, 4.7, 4.6, 5.0, 5.4],
'SepalWidthCm': [3.5, 3.0, 3.2, 3.1, 3.6, 3.9],
'PetalLengthCm': [1.4, 1.3, 1.5, 1.4, 1.7, 1.4],
'PetalWidthCm': [0.2, 0.2, 0.2, 0.2, 0.4, 0.3],
})
# Calculate correlation matrix
correlation_matrix = df.corr(method='pearson')
# Create a heatmap using plt.matshow
plt.matshow(correlation_matrix, cmap='coolwarm')
# Add annotations
for i in range(len(correlation_matrix.columns)):
for j in range(len(correlation_matrix.index)):
plt.text(i, j, f'{correlation_matrix.iloc[j, i]:.2f}', ha='center', va='center', color='black')
69°F
Smoke
Search
ENG
IN
11:36 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 14 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
correlation_matrix = df.corr(method='pearson')
# Create a heatmap using plt.matshow
plt.matshow(correlation_matrix, cmap='coolwarm')
# Add annotations
for i in range(len(correlation_matrix.columns)):
for j in range(len(correlation_matrix.index)):
plt.text(i, j, f'{correlation_matrix.iloc[j, i]:.2f}', ha='center', va='center', color='black')
# Add colorbar
plt.colorbar()
# Set axis labels
plt.xticks(range(len(correlation_matrix.columns)), correlation_matrix.columns, rotation=45)
plt.yticks(range(len(correlation_matrix.index)), correlation_matrix.index)
# Show the plot
plt.show()
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
1.00
0.86
-0.03
0.46
0.86
1.00
0.36
0.66
-0.03
0.36
1.00
0.78
0.46
0.66
0.78
1.00
In [32]:
import matplotlib.pyplot as plt
import pandas as pd
69°F
Smoke
Search
ENG
IN
11:36 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 16 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
0.0
In [32]:
# Assuming df is your DataFrame with columns 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', and 'Species'
df = pd.DataFrame({
'SepalLengthCm': [5.1, 4.9, 4.7, 4.6, 5.0, 5.4],
'SepalWidthCm': [3.5, 3.0, 3.2, 3.1, 3.6, 3.9],
'PetalLengthCm': [1.4, 1.3, 1.5, 1.4, 1.7, 1.4],
'PetalWidthCm': [0.2, 0.2, 0.2, 0.2, 0.4, 0.3],
'Species': ['A', 'B', 'A', 'C', 'B', 'A']
})
# Define colors for each species
colors = {'A': 'red', 'B': 'blue', 'C': 'green'}
# Create boxplots using Matplotlib
plt.figure(figsize=(10, 10))
# Boxplot for SepalLengthCm
plt.subplot(221)
for species, color in colors.items():
subset = df[df['Species'] == species]
plt.boxplot(subset['SepalLengthCm'], labels=[species], showfliers=False, patch_artist=True, boxprops=dict(facecolor=color))
# Boxplot for SepalWidthCm
plt.subplot(222)
for species, color in colors.items():
subset = df[df['Species'] == species]
plt.boxplot(subset['SepalWidthCm'], labels=[species], showfliers=False, patch_artist=True, boxprops=dict(facecolor=color))
# Boxplot for PetalLengthCm
plt.subplot(223)
for species, color in colors.items():
subset = df[df['Species'] == species]
plt.boxplot(subset['PetalLengthCm'], labels=[species], showfliers=False, patch_artist=True, boxprops=dict(facecolor=color))
# Boxplot for PetalWidthCm
plt.subplot(224)
for species, color in colors.items():
subset = df[df['Species'] == species]
plt.boxplot(subset['PetalWidthCm'], labels=[species], showfliers=False, patch_artist=True, boxprops=dict(facecolor=color))
# Adjust layout
plt.tight_layout()
# Show the plot
plt.show()
69°F
Smoke
Search
ENG
IN
11:38 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 16 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
# Show the plot
plt.show()
5.4
5.3
5.2
5.1
5.0
4.9
4.8
4.7
4.6
A
3.8
3.6
3.4
3.2
3.0
A
1.70
1.65
1.60
1.55
1.50
1.45
0.400
0.375
0.350
0.325
0.300
0.275
69°F
Smoke
Search
ENG
IN
11:38 PM
12/20/2023


Data visualization using matplotlib
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
In [33]:
# Create a boxplot using Matplotlib
plt.boxplot(df['SepalWidthCm'], vert=False)
# Add labels and title
plt.xlabel('Sepal Width (cm)')
plt.title('Boxplot of Sepal Width')
# Show the plot
plt.show()
Boxplot of Sepal Width
Sepal Width (cm)


Data visualization using matplotlib
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
Sepal Width (cm)
In [39]:
df.hist(edgecolor='black', linewidth=1)
plt.show()
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
In [41]:
df.plot(kind = "density", figsize=(10,8))
Out[41]:
<AxesSubplot:ylabel='Density'>
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm

Home Page - Select or create a n | Data visualization using matplotli | iris datset using data visulaizatio

localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb

jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)

Logout

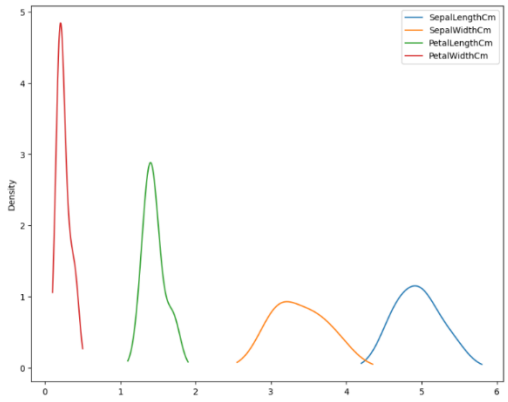
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [41]:
```
df.plot(kind = "density", figsize=(10,8))
```

Out[41]: <AxesSubplot:ylabel='Density'>



In [43]:
```
# Create a scatter plot
plt.scatter(df['SepalLengthCm'], df['SepalWidthCm'], alpha=0.5)

# Set labels and title
```

69°F Smoke | Search | ENG IN | 11:39 PM 12/20/2023

Home Page - Select or create a n | Data visualization using matplotli | iris datset using data visulaizatio

localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb

jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)

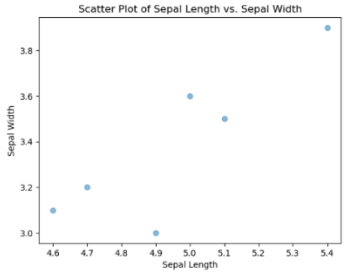
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

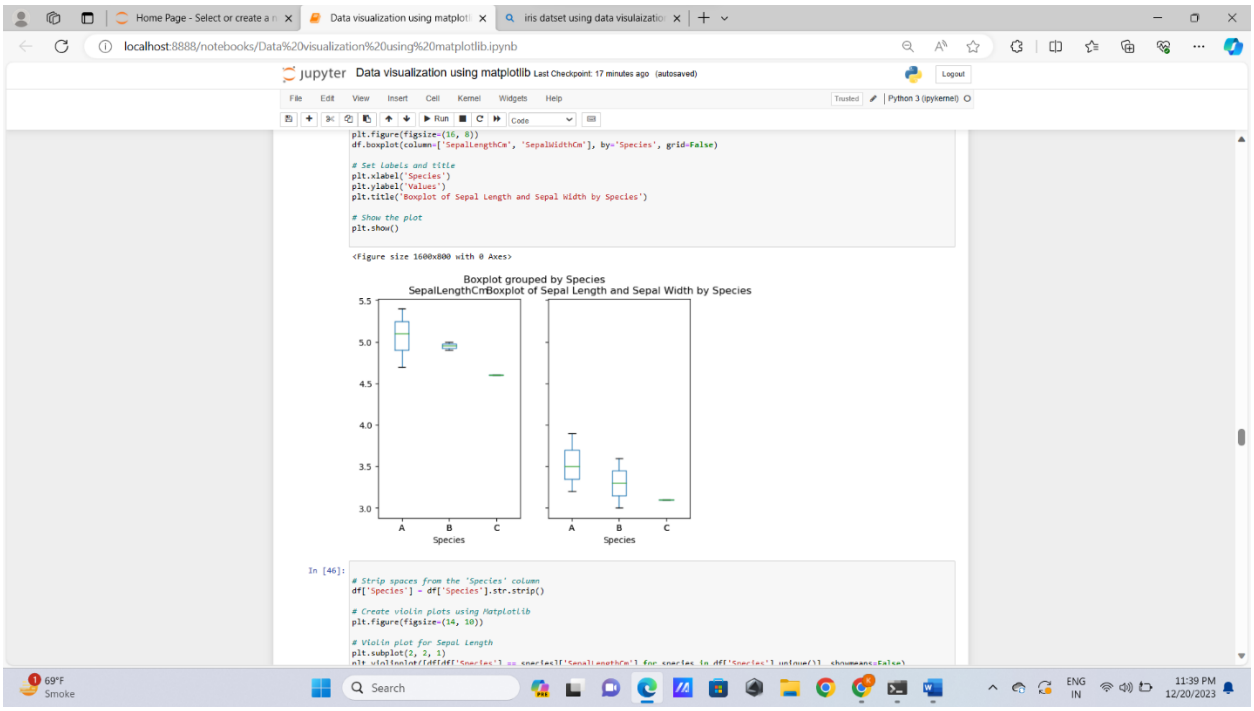
Run Code

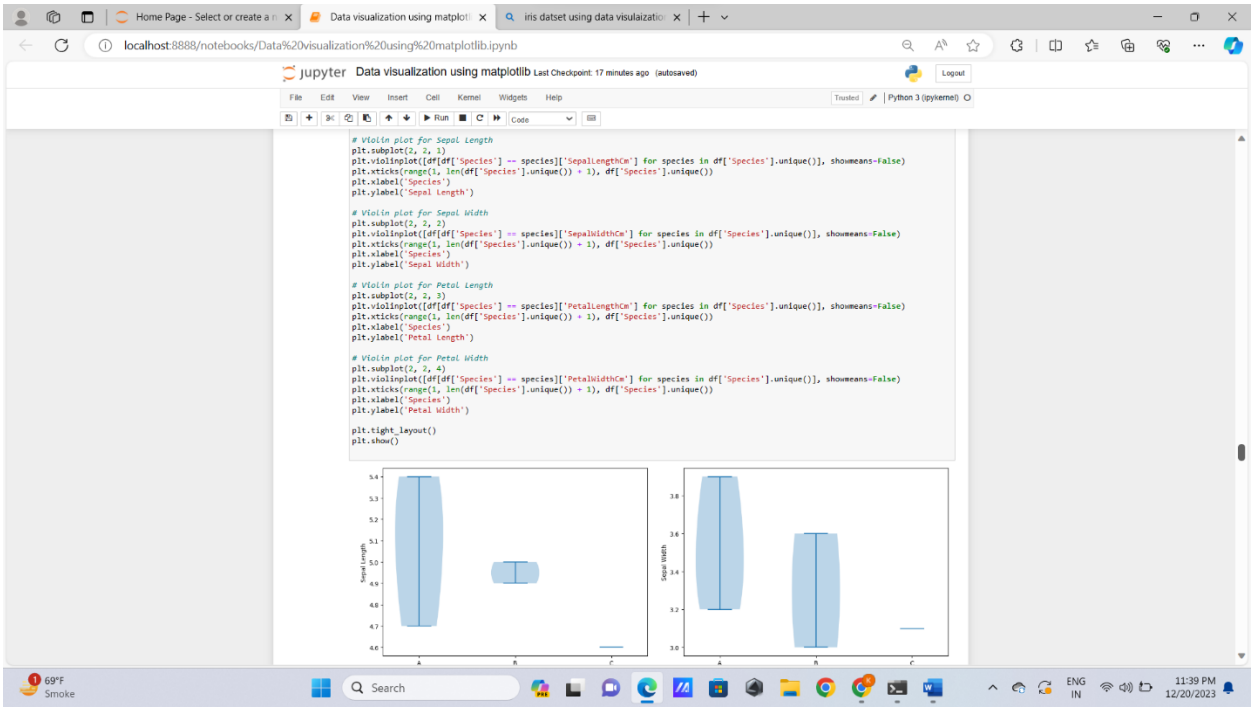
In [43]:
```
# Create a scatter plot
plt.scatter(df['SepalLengthCm'], df['SepalWidthCm'], alpha=0.5)

# Set labels and title
plt.xlabel('Sepal Length')
plt.ylabel('Sepal Width')
plt.title('Scatter Plot of Sepal Length vs. Sepal Width')

# Show the plot
plt.show()
```

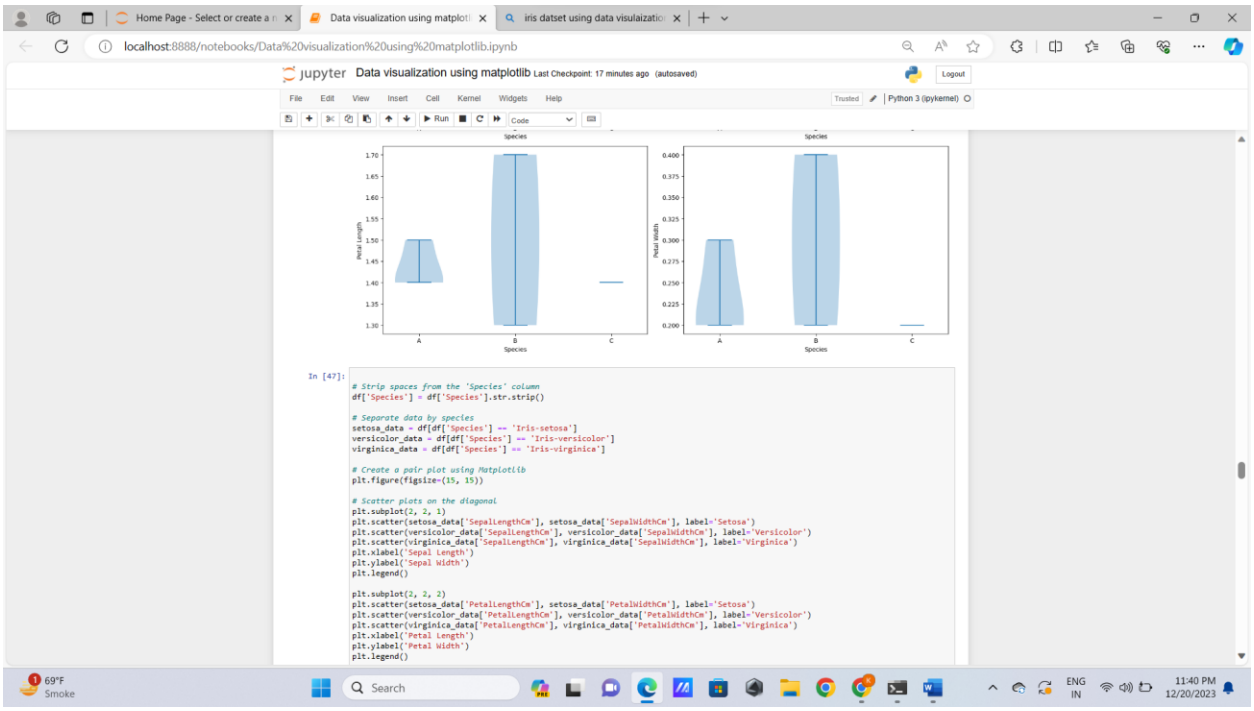


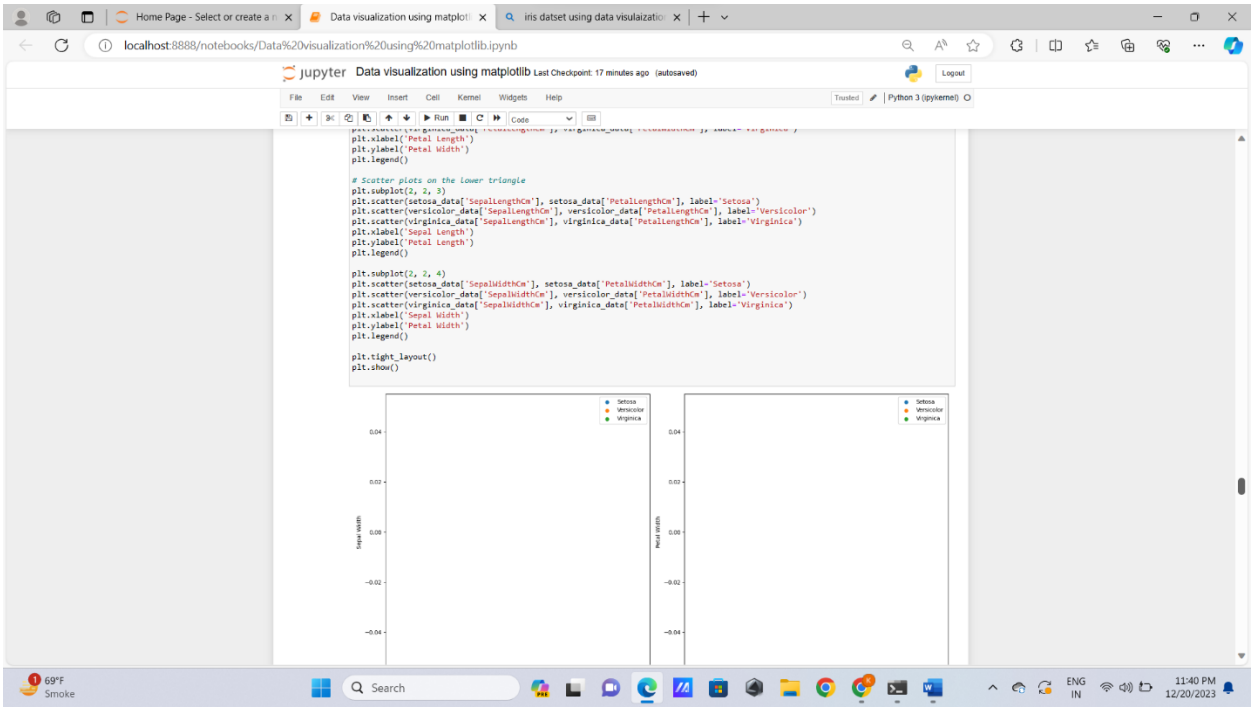
In [45]:
```
# Strip spaces from the 'Species' column
df['Species'] = df['Species'].str.strip()

# Create a boxplot using Matplotlib
plt.figure(figsize=(16, 8))
df.boxplot(column=['SepalLengthCm', 'SepalWidthCm'], by='Species', grid=False)

# Set labels and title
plt.xlabel('Species')
plt.ylabel('Values')
```

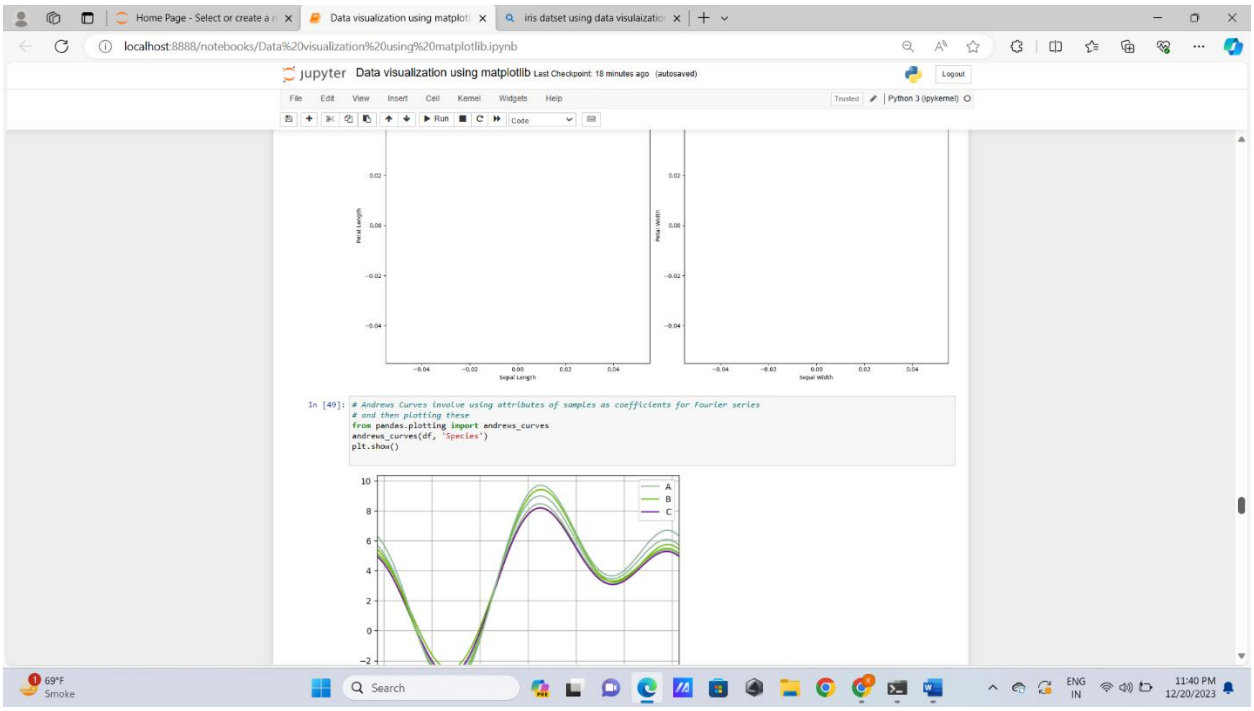
69°F Smoke | Search | ENG IN | 11:39 PM 12/20/2023

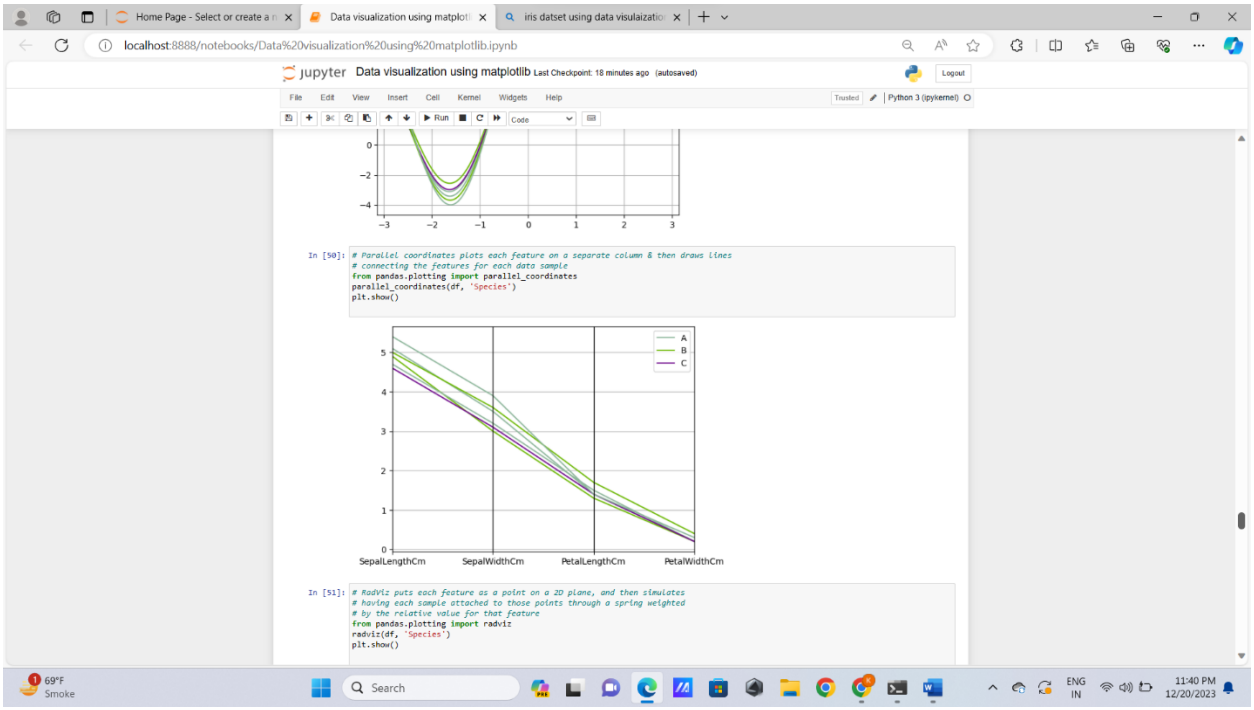

Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
plt.figure(figsize=(16, 8))
df.boxplot(column=['SepalLengthCm', 'SepalWidthCm'], by='Species', grid=False)
# Set labels and title
plt.xlabel('Species')
plt.ylabel('Values')
plt.title('Boxplot of Sepal Length and Sepal Width by Species')
# Show the plot
plt.show()
<Figure size 1600x800 with 0 Axes>
Boxplot grouped by Species
SepalLengthCmBoxplot of Sepal Length and Sepal Width by Species
5.5
5.0
4.5
4.0
3.5
3.0
A
B
C
Species
In [46]:
Strip spaces from the 'Species' column
df['Species'] = df['Species'].str.strip()
# Create violin plots using Matplotlib
plt.figure(figsize=(14, 10))
# Violin plot for Sepal Length
Search
69°F
Smoke
ENG
IN
11:39 PM
12/20/2023

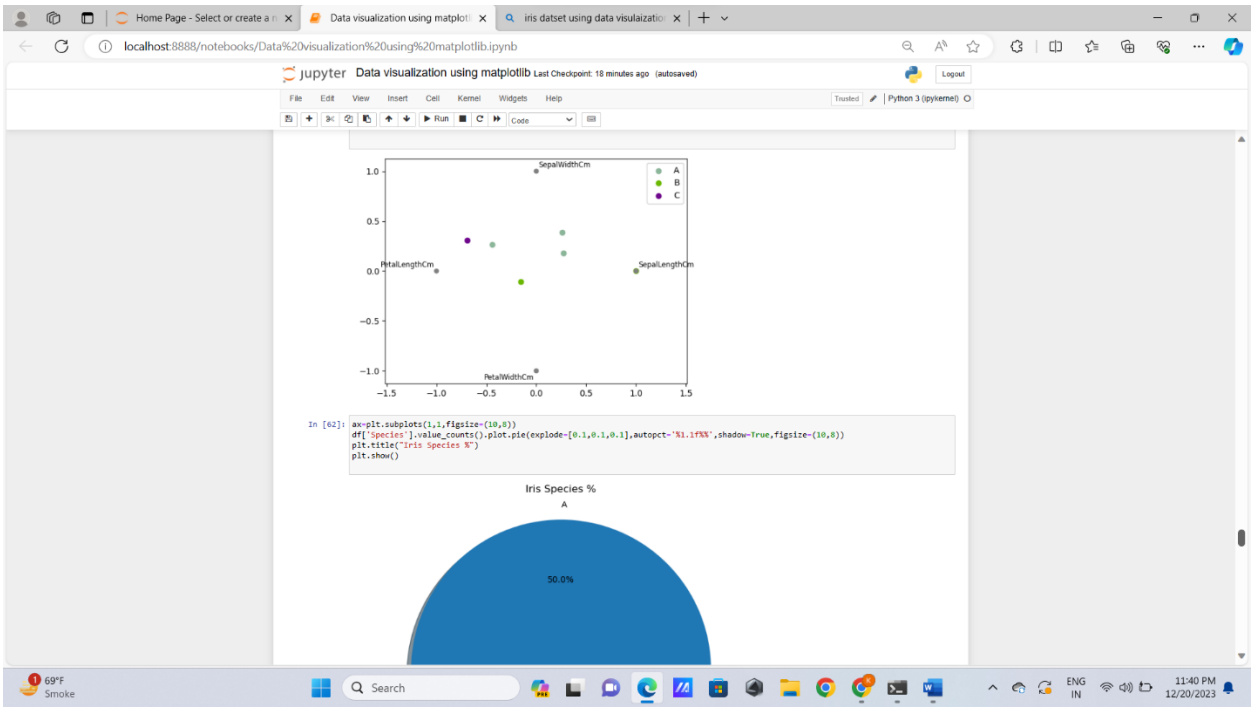

Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
# Violin plot for Sepal Length
plt.subplot(2, 2, 1)
plt.violinplot([df[df['Species'] == species]['SepalLengthCm'] for species in df['Species'].unique()], showmeans=False)
plt.xticks(range(1, len(df['Species'].unique()) + 1), df['Species'].unique())
plt.xlabel('Species')
plt.ylabel('Sepal Length')
# Violin plot for Sepal Width
plt.subplot(2, 2, 2)
plt.violinplot([df[df['Species'] == species]['SepalWidthCm'] for species in df['Species'].unique()], showmeans=False)
plt.xticks(range(1, len(df['Species'].unique()) + 1), df['Species'].unique())
plt.xlabel('Species')
plt.ylabel('Sepal Width')
# Violin plot for Petal Length
plt.subplot(2, 2, 3)
plt.violinplot([df[df['Species'] == species]['PetalLengthCm'] for species in df['Species'].unique()], showmeans=False)
plt.xticks(range(1, len(df['Species'].unique()) + 1), df['Species'].unique())
plt.xlabel('Species')
plt.ylabel('Petal Length')
# Violin plot for Petal Width
plt.subplot(2, 2, 4)
plt.violinplot([df[df['Species'] == species]['PetalWidthCm'] for species in df['Species'].unique()], showmeans=False)
plt.xticks(range(1, len(df['Species'].unique()) + 1), df['Species'].unique())
plt.xlabel('Species')
plt.ylabel('Petal Width')
plt.tight_layout()
plt.show()
Sepal Length
Sepal Width
A
B
C
Search
69°F
Smoke
ENG
IN
11:39 PM
12/20/2023

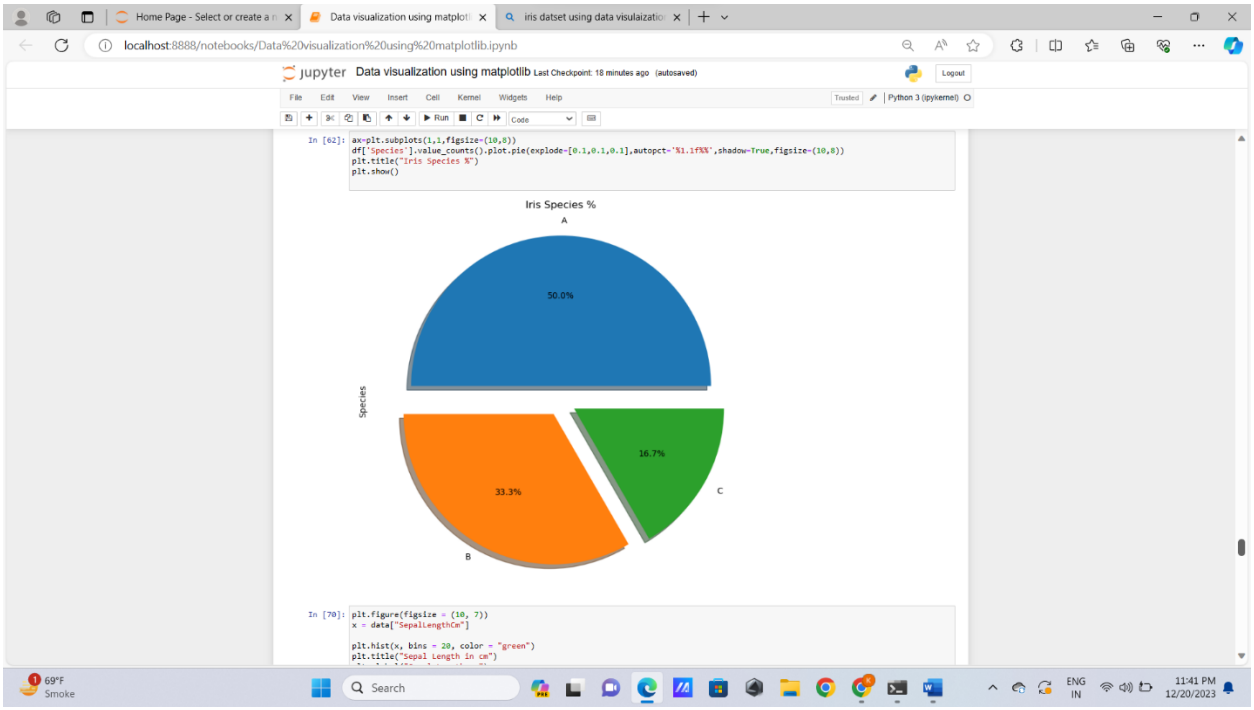

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
Species
Petal Length
Petal Width
A B C
In [47]:
Strip spaces from the 'Species' column
df['Species'] = df['Species'].str.strip()
# Separate data by species
setosa_data = df[df['Species'] == 'Iris-setosa']
versicolor_data = df[df['Species'] == 'Iris-versicolor']
virginica_data = df[df['Species'] == 'Iris-virginica']
# Create a pair plot using Matplotlib
plt.figure(figsize=(15, 15))
# Scatter plots on the diagonal
plt.subplot(2, 2, 1)
plt.scatter(setosa_data['SepalLengthCm'], setosa_data['SepalWidthCm'], label='Setosa')
plt.scatter(versicolor_data['SepalLengthCm'], versicolor_data['SepalWidthCm'], label='Versicolor')
plt.scatter(virginica_data['SepalLengthCm'], virginica_data['SepalWidthCm'], label='Virginica')
plt.xlabel('Sepal Length')
plt.ylabel('Sepal Width')
plt.legend()
plt.subplot(2, 2, 2)
plt.scatter(setosa_data['PetalLengthCm'], setosa_data['PetalWidthCm'], label='Setosa')
plt.scatter(versicolor_data['PetalLengthCm'], versicolor_data['PetalWidthCm'], label='Versicolor')
plt.scatter(virginica_data['PetalLengthCm'], virginica_data['PetalWidthCm'], label='Virginica')
plt.xlabel('Petal Length')
plt.ylabel('Petal Width')
plt.legend()
69°F Smoke
Search
ENG IN
11:40 PM 12/20/2023

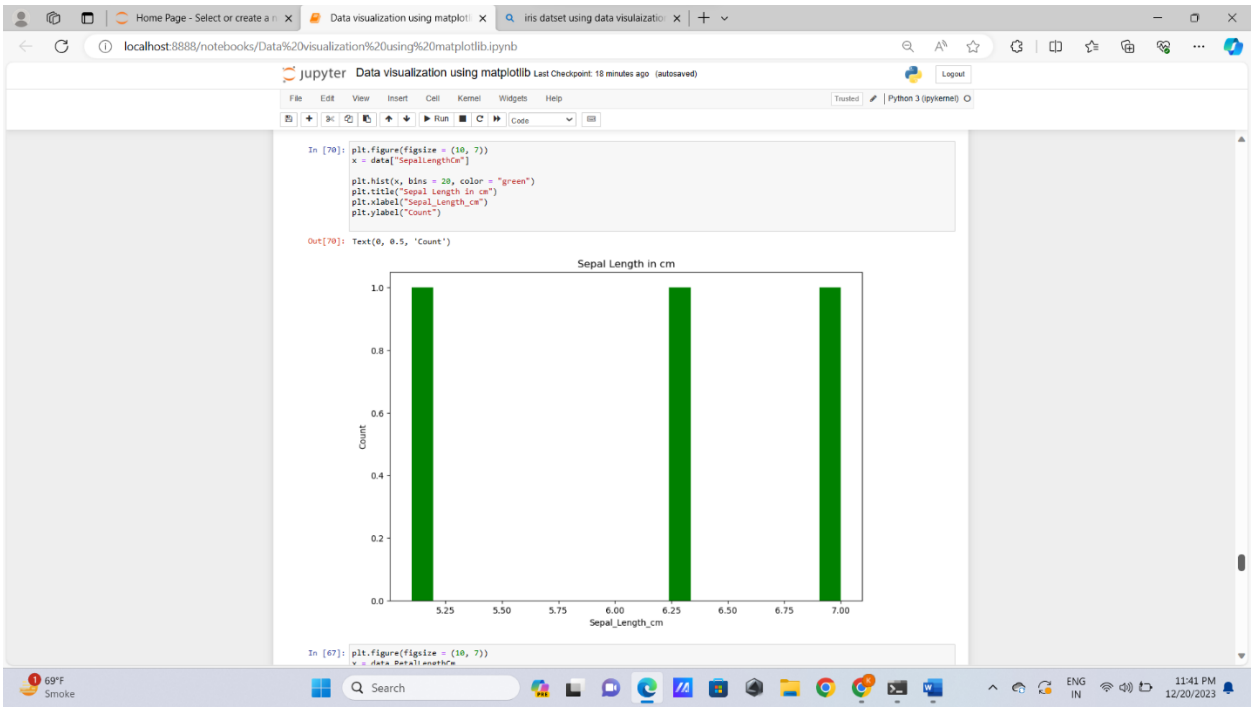

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 17 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
plt.xlabel('Petal Length')
plt.ylabel('Petal Width')
plt.legend()
# Scatter plots on the lower triangle
plt.subplot(2, 2, 3)
plt.scatter(setosa_data['SepalLengthCm'], setosa_data['PetalLengthCm'], label='Setosa')
plt.scatter(versicolor_data['SepalLengthCm'], versicolor_data['PetalLengthCm'], label='Versicolor')
plt.scatter(virginica_data['SepalLengthCm'], virginica_data['PetalLengthCm'], label='Virginica')
plt.xlabel('Sepal Length')
plt.ylabel('Petal Length')
plt.legend()
plt.subplot(2, 2, 4)
plt.scatter(setosa_data['SepalWidthCm'], setosa_data['PetalWidthCm'], label='Setosa')
plt.scatter(versicolor_data['SepalWidthCm'], versicolor_data['PetalWidthCm'], label='Versicolor')
plt.scatter(virginica_data['SepalWidthCm'], virginica_data['PetalWidthCm'], label='Virginica')
plt.xlabel('Sepal Width')
plt.ylabel('Petal Width')
plt.legend()
plt.tight_layout()
plt.show()
Setosa
Versicolor
Virginica
Sepal Width
Petal Width
69°F Smoke
Search
ENG IN
11:40 PM 12/20/2023

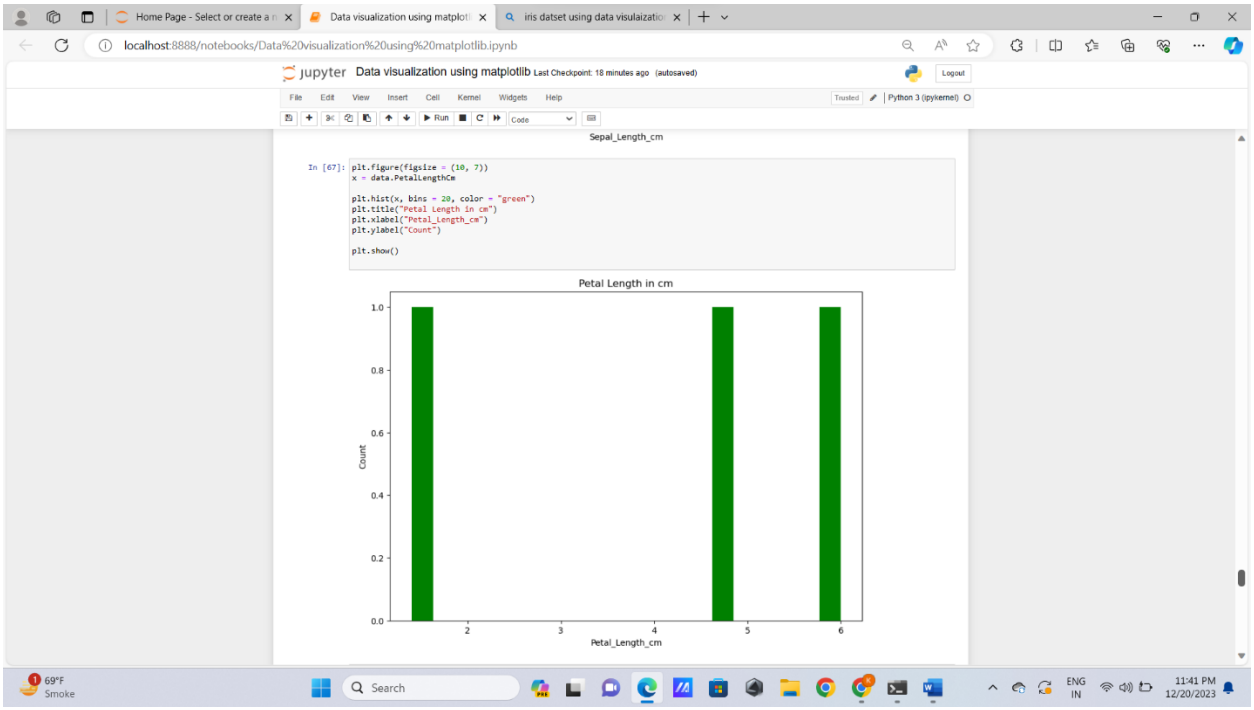

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
Petal Length
Petal Width
Sepal Length
Sepal Width
In [49]:
# Andrews Curves involve using attributes of samples as coefficients for Fourier series
# and then plotting these
from pandas.plotting import andrews_curves
andrews_curves(df, 'Species')
plt.show()
A
B
C
69°F
Smoke
Search
ENG
IN
11:40 PM
12/20/2023

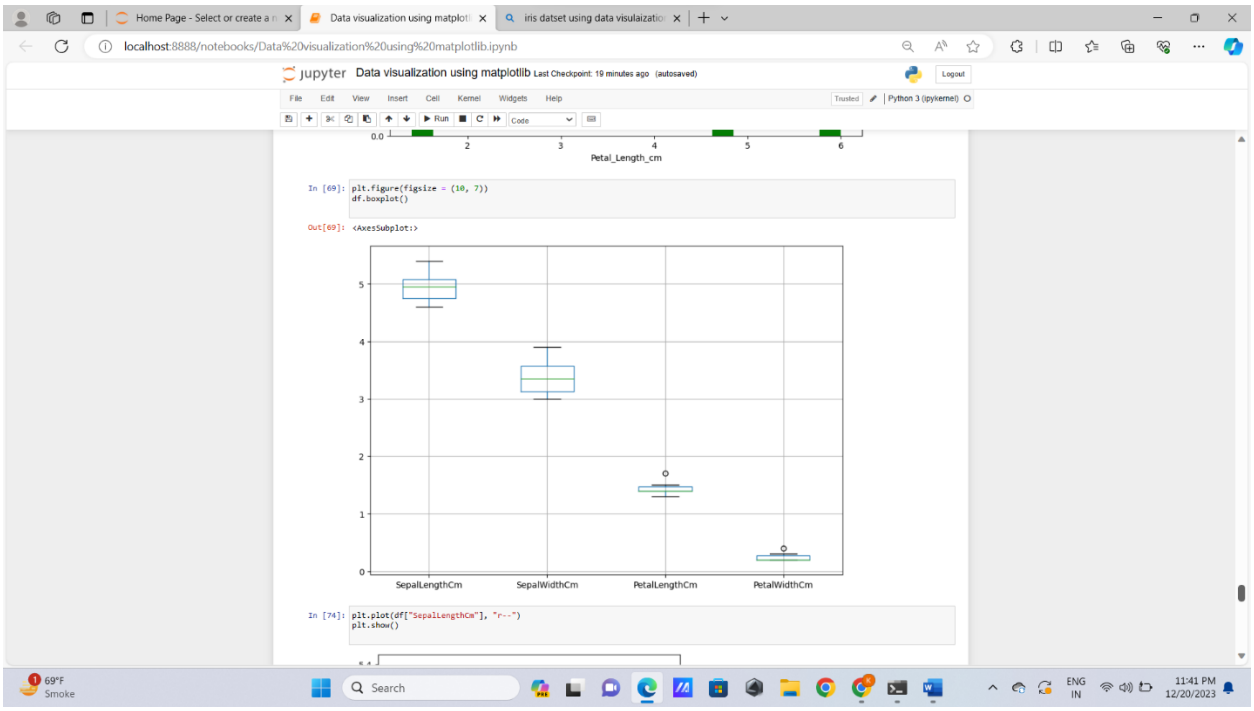

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
In [50]:
# Parallel coordinates plots each feature on a separate column & then draws lines
# connecting the features for each data sample
from pandas.plotting import parallel_coordinates
parallel_coordinates(df, 'Species')
plt.show()
A
B
C
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
In [51]:
# RadViz puts each feature as a point on a 2D plane, and then simulates
# having each sample attached to those points through a spring weighted
# by the relative value for that feature
from pandas.plotting import radviz
radviz(df, 'Species')
plt.show()
69°F
Smoke
Search
ENG
IN
11:40 PM
12/20/2023

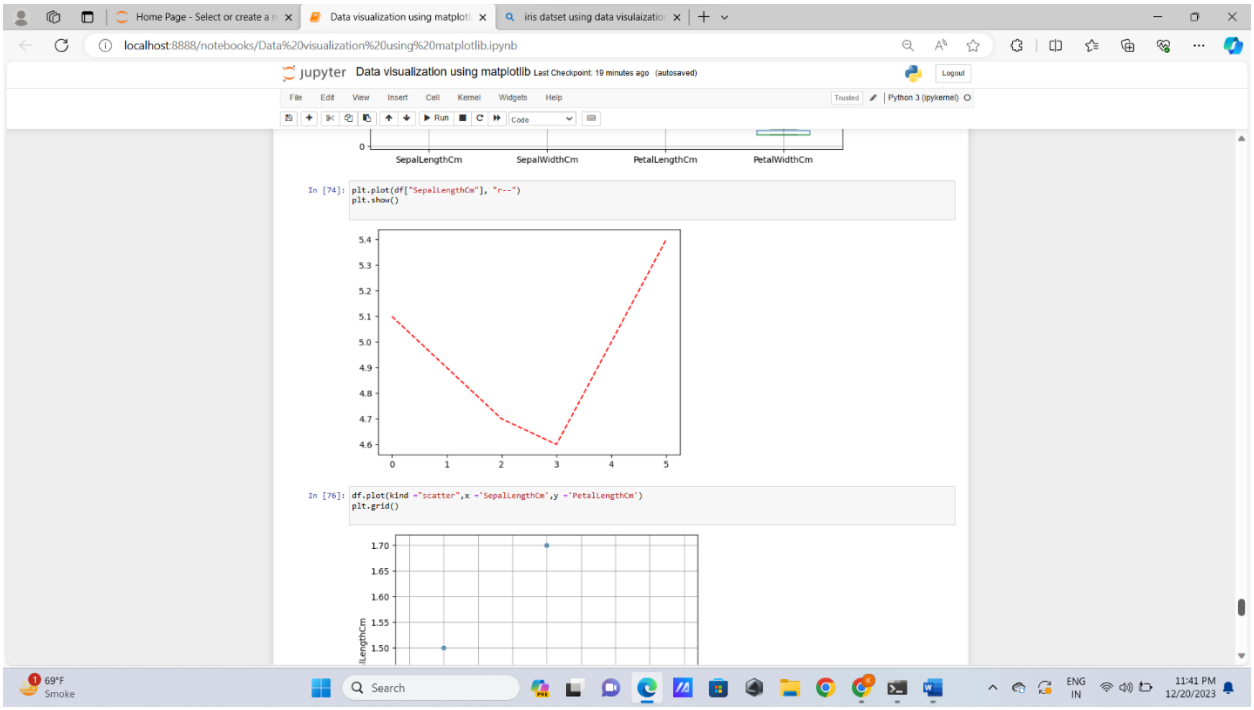

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
SepalWidthCm
PetalLengthCm
SepalLengthCm
PetalWidthCm
A
B
C
In [62]:
ax=plt.subplots(1,1,figsize=(10,8))
df['Species'].value_counts().plot.pie(explode=[0.1,0.1,0.1],autopct='%1.1f%%',shadow=True,figsize=(10,8))
plt.title("Iris Species %")
plt.show()
Iris Species %
A
50.0%
69°F
Smoke
Search
ENG
IN
11:40 PM
12/20/2023

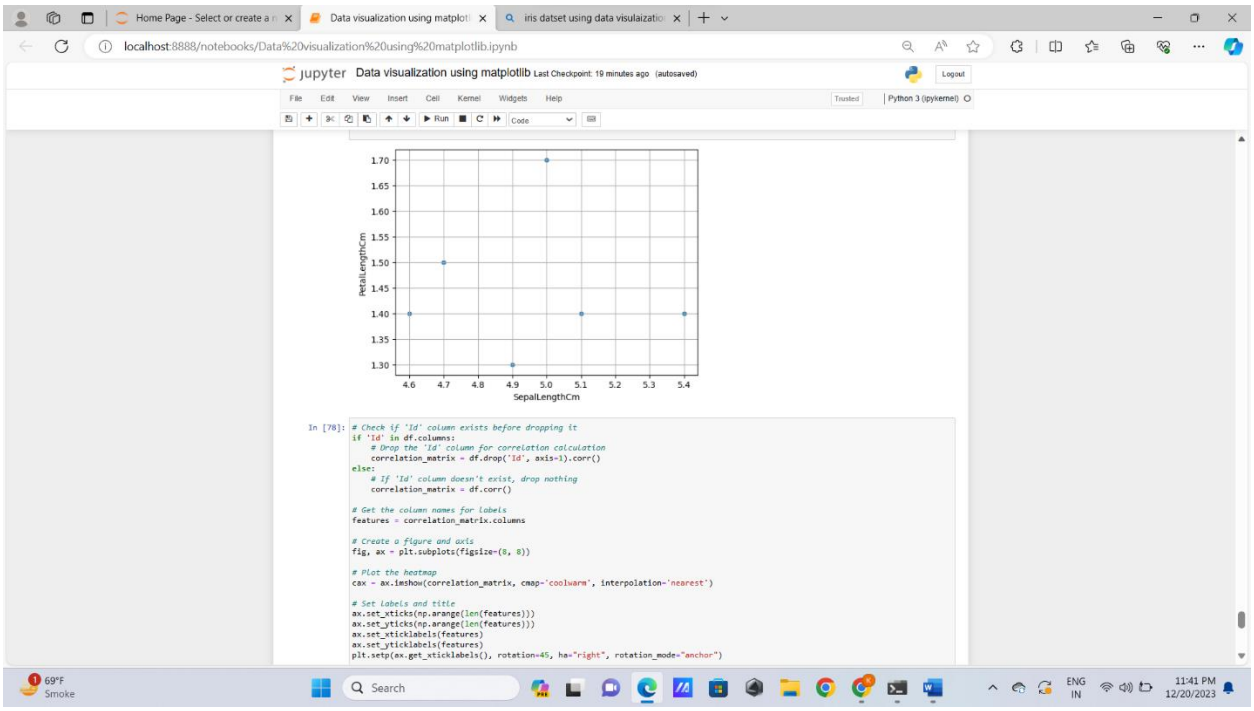

Home Page - Select or create a n
Data visualization using matplotl
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3 (ipykernel)
Run Code
In [62]:
ax=plt.subplots(1,1,figsize=(10,8))
df['Species'].value_counts().plot.pie(explode=[0.1,0.1,0.1],autopct='%1.1f%%',shadow=True,figsize=(10,8))
plt.title("Iris Species %")
plt.show()
Iris Species %
A
50.0%
Species
16.7%
C
33.3%
B
In [70]:
plt.figure(figsize = (10, 7))
x = data["SepalLengthCm"]
plt.hist(x, bins = 20, color = "green")
plt.title("Sepal Length in cm")
69°F
Smoke
Search
ENG
IN
11:41 PM
12/20/2023

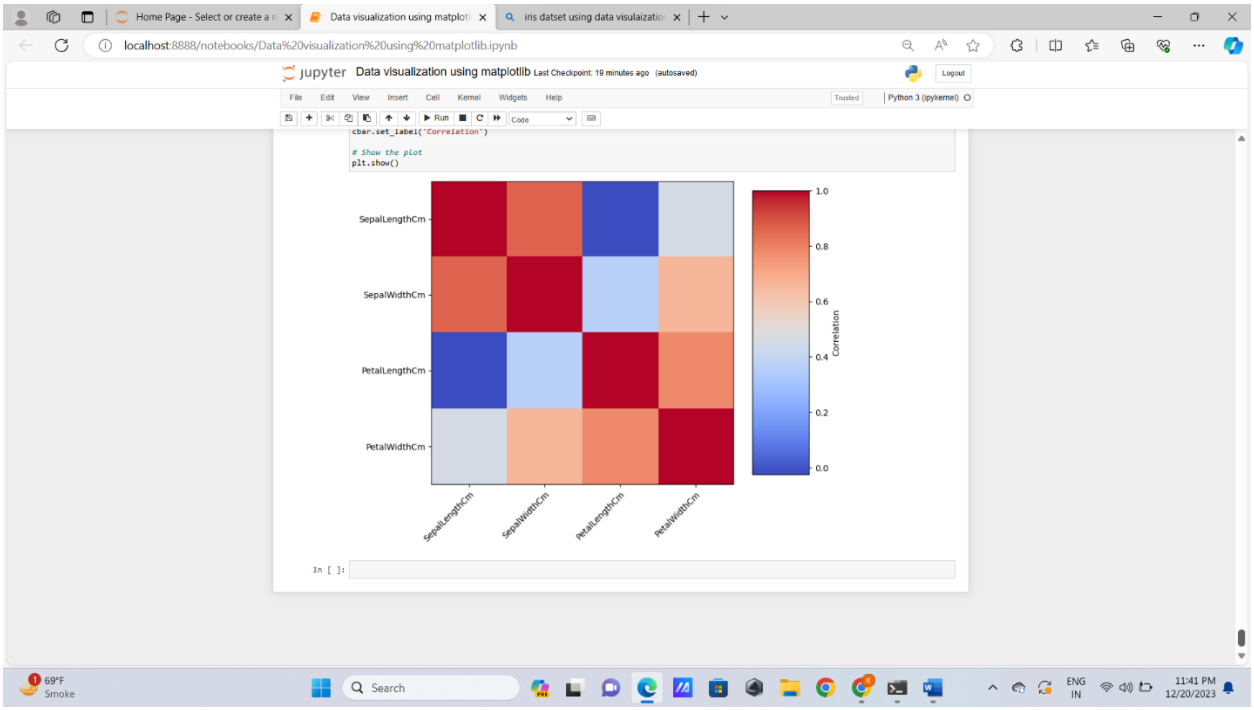

Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [70]:
plt.figure(figsize = (10, 7))
x = data["SepalLengthCm"]
plt.hist(x, bins = 20, color = "green")
plt.title("Sepal Length in cm")
plt.xlabel("Sepal_Length_cm")
plt.ylabel("Count")
Out[70]:
Text(0, 0.5, 'Count')
Sepal Length in cm
Count
Sepal_Length_cm
1.0
0.8
0.6
0.4
0.2
0.0
5.25
5.50
5.75
6.00
6.25
6.50
6.75
7.00
In [67]:
plt.figure(figsize = (10, 7))
69°F
Smoke
Search
ENG
IN
11:41 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 18 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Sepal_Length_cm
In [67]:
plt.figure(figsize = (10, 7))
x = data.PetalLengthCm
plt.hist(x, bins = 20, color = "green")
plt.title("Petal Length in cm")
plt.xlabel("Petal_Length_cm")
plt.ylabel("Count")
plt.show()
Petal Length in cm
Count
Petal_Length_cm
1.0
0.8
0.6
0.4
0.2
0.0
2
3
4
5
6
69°F
Smoke
Search
ENG
IN
11:41 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 19 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Petal_Length_cm
In [69]:
plt.figure(figsize = (10, 7))
df.boxplot()
Out[69]:
<AxesSubplot:>
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
In [74]:
plt.plot(df["SepalLengthCm"], "r--")
plt.show()
69°F
Smoke
Search
ENG
IN
11:41 PM
12/20/2023


Home Page - Select or create a n
Data visualization using matplotli
iris datset using data visulaizatio
localhost:8888/notebooks/Data%20visualization%20using%20matplotlib.ipynb
jupyter Data visualization using matplotlib Last Checkpoint: 19 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
In [74]:
plt.plot(df["SepalLengthCm"], "r--")
plt.show()
In [76]:
df.plot(kind ="scatter",x ='SepalLengthCm',y ='PetalLengthCm')
plt.grid()
69°F
Smoke
Search
ENG
IN
11:41 PM
12/20/2023





❖**Description**:

I'm thrilled to apply this newfound knowledge to real-world scenarios, and the prospect of further honing these skills through the Skill Boost Internship Program adds an extra layer of excitement. Here's to the

journey ahead and the exciting challenges awaiting in the Skill Boost Internship Program(www.Batweb.com).