

copy-of-untitled0-1

April 28, 2023

[]:

```
[1]: from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

1. Upload your data set to your drive (any mail id) and show how to access it into the Colab environment.

```
[2]: import pandas as pd

# Load the dataset
avila_data = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/avila-tr.txt",
    ↪header=None)

# Write the dataset to a CSV file
avila_data.to_csv("avila_data.csv", index=False)
```

Show the shape and size of the data set.

```
[4]: import pandas as pd

# Load the Avila data set
url = "/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv"
df = pd.read_csv(url)

# Show the shape and size of the data set
print("Shape of Avila data set:", df.shape)
print("Size of Avila data set:", df.size)
```

Shape of Avila data set: (10429, 11)

Size of Avila data set: 114719

2. Subset any 2 numeric columns from the data set and create a separate variable to store them.

```
[8]: import pandas as pd

# Load the Avila data set
```

```

url = "/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv"
df = pd.read_csv(url)

# Subset the data set to select two numeric columns
subset_df = df[['0.266074', '-0.16562']]

# Store the two numeric columns in a separate variable
numeric_cols = subset_df.values

# Show the first five rows of the numeric data
print(numeric_cols[:5])

```

```

[[ 0.130292  0.870736]
 [-0.116585  0.069915]
 [ 0.031541  0.2976  ]
 [ 0.229043  0.807926]
 [ 0.117948 -0.220579]]

```

3. Using appropriate graph technique

```

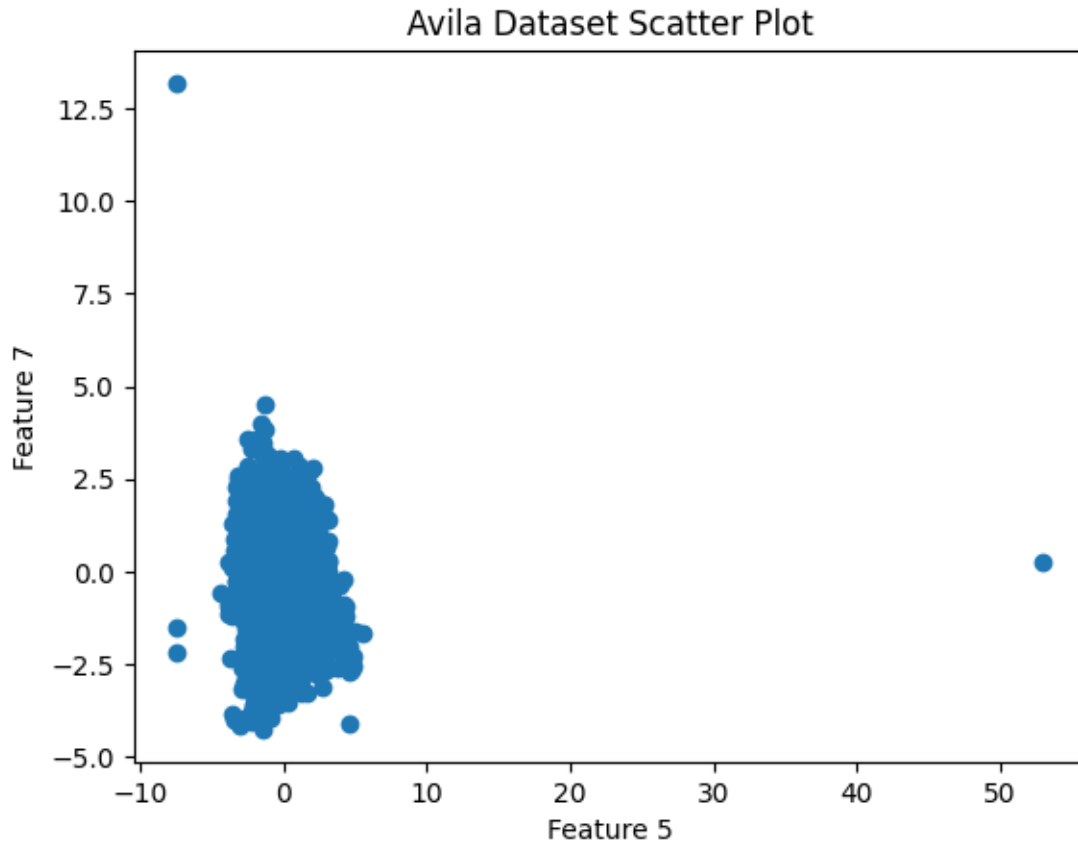
[9]: import pandas as pd
import matplotlib.pyplot as plt

# Load Avila dataset
url = "/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv"
avila_data = pd.read_csv(url, header=None)

# Select two features to plot
x_feature = 5
y_feature = 7

# Create scatter plot
plt.scatter(avila_data[x_feature], avila_data[y_feature])
plt.xlabel(f"Feature {x_feature}")
plt.ylabel(f"Feature {y_feature}")
plt.title("Avila Dataset Scatter Plot")
plt.show()

```



4. Draw a single graph to find weather two are correlated or not

```
[10]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load Avila dataset
url = "/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv"
avila_data = pd.read_csv(url, header=None)

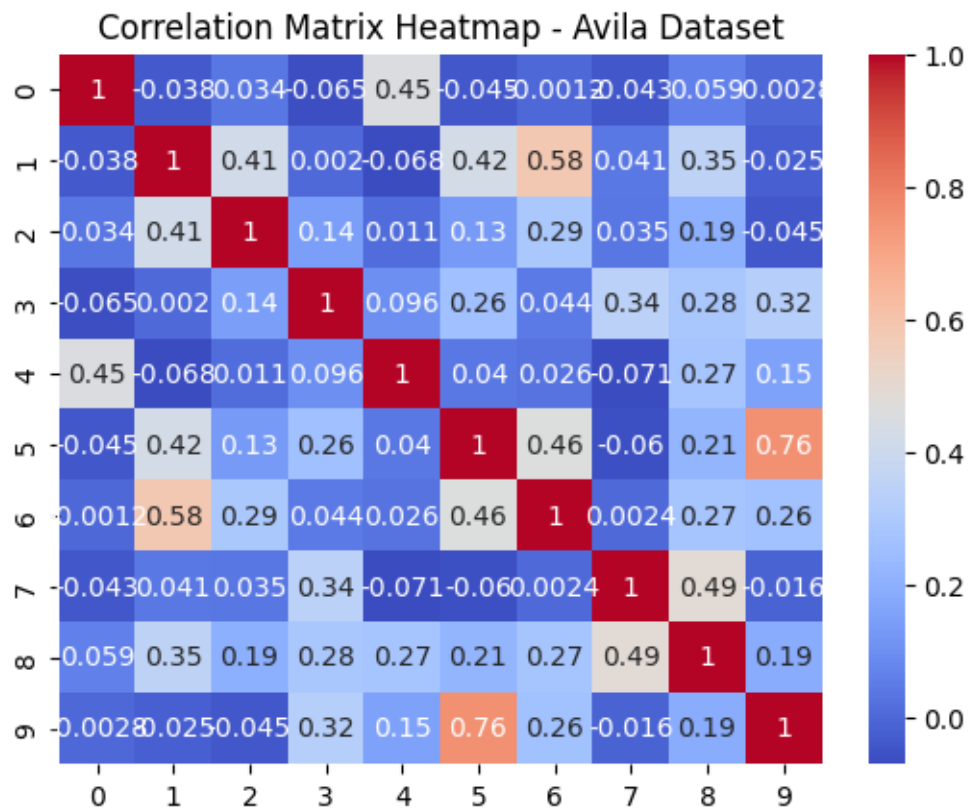
# Compute correlation matrix
corr_matrix = avila_data.corr()

# Create heatmap
sns.heatmap(corr_matrix, annot=True, cmap="coolwarm")
plt.title("Correlation Matrix Heatmap - Avila Dataset")
plt.show()
```

<ipython-input-10-223585dc299e>:10: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will

default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
corr_matrix = avila_data.corr()
```

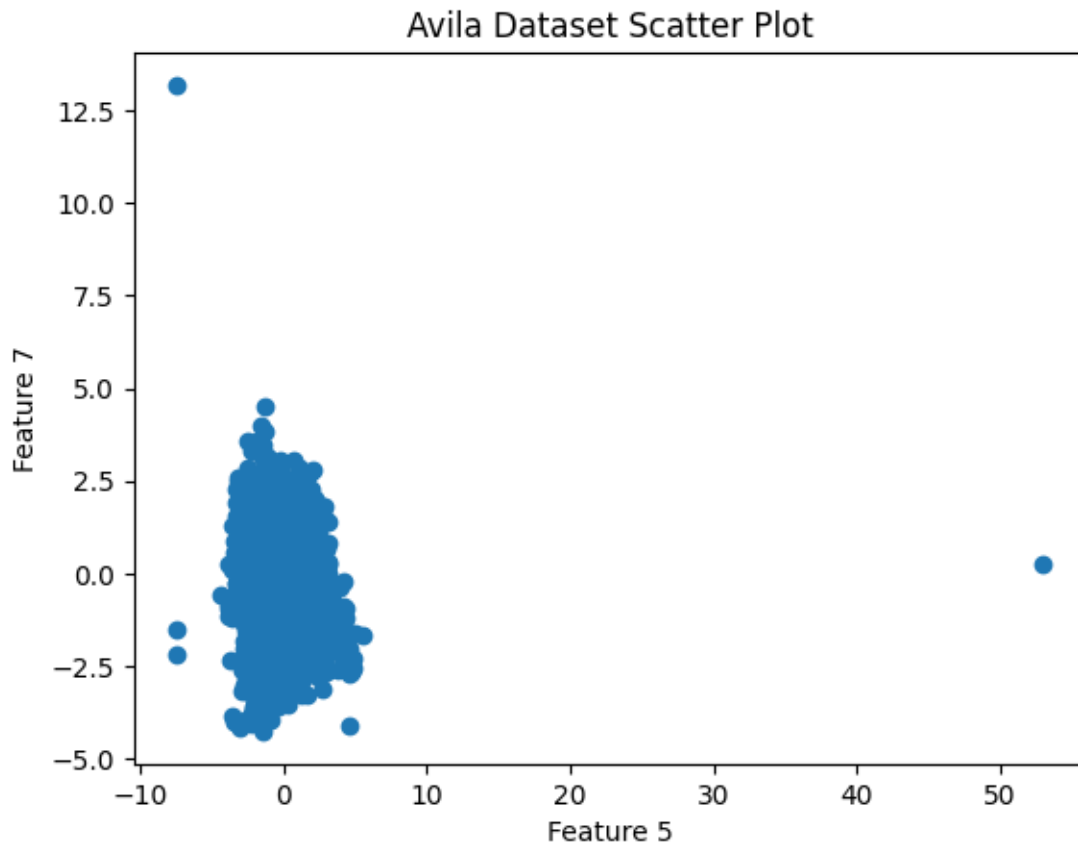


```
[11]: import pandas as pd
import matplotlib.pyplot as plt

# Load Avila dataset
url = "/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv"
avila_data = pd.read_csv(url, header=None)

# Select two features to plot
x_feature = 5
y_feature = 7

# Create scatter plot
plt.scatter(avila_data[x_feature], avila_data[y_feature])
plt.xlabel(f"Feature {x_feature}")
plt.ylabel(f"Feature {y_feature}")
plt.title("Avila Dataset Scatter Plot")
plt.show()
```



```
[16]: import pandas as pd
import matplotlib.pyplot as plt

# Load the Avila dataset
data = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/avila-tr.txt.csv")

# Extract a single variable, such as 'A'
h_values = data['A']

# Plot a histogram of 'H' values
plt.hist(h_values, bins=50)
plt.xlabel('A values')
plt.ylabel('Frequency')
plt.title('Histogram of Avila Dataset A values')
plt.show()
```

