

Lab 5

[Click here to Register Attendance](https://forms.gle/pesShFDe1ibVfS3L7)

|  |  |
| --- | --- |
| Name |  |
| Date |  |
| Student No |  |
| Student Email |  |

### **Databricks**

1. Create a new user account for Databricks Community Edition and insert a screenshot here:

|  |
| --- |
|  |

1. Create a New Cluster on Databricks and insert a screenshot here:

|  |
| --- |
|  |

1. Create a New Table on Databricks - using the Iris dataset used in previous weeks - insert a screenshot here:

|  |
| --- |
|  |

1. Create a New Notebook on Databricks and insert a screenshot here:

|  |
| --- |
|  |

1. Start a new PySpark session with the following code and show the Iris dataset:

import pyspark

from pyspark.sql import SparkSession

# Start Spark Session

spark = SparkSession.builder.appName("Operations").getOrCreate()

# Load data from the Iris dataset

df=sqlContext.sql("select \* from iris\_data")

df.show()

Insert a screenshot here:

|  |
| --- |
|  |

* 1. Use the following code to print a row of data:  
     for row in df.head(5):

print(row)

print('\n')

Insert a screenshot here:

|  |
| --- |
|  |

1. Describe your Iris dataset using the following code:

df.describe().show()

Insert a screenshot here:

|  |
| --- |
|  |

1. Describe the dataset schema using this code:  
    df.describe().printSchema()  
   Insert a screenshot here:

|  |
| --- |
|  |

1. Find the Average Petal Width for each class using this code:  
   display(df.select("Class","PetalWidth").groupBy("Class").agg(avg("PetalWidth")))  
   Insert a screenshot here:

|  |
| --- |
|  |

1. Generate summary graphs of each class of flower using this code:  
   import seaborn as sns

sns.set(style="white")

df1 = sns.load\_dataset("iris")

g = sns.PairGrid(df1, diag\_sharey=False)

g.map\_lower(sns.kdeplot)

g.map\_diag(sns.kdeplot, lw=3)

g.map\_upper(sns.regplot)

display(g.fig)  
Insert a screenshot here:

|  |
| --- |
|  |

1. Research some of the differences between ordinary Python dataframes and PySpark dataframes and identify 3 key differences

|  |
| --- |
|  |