University College of Applied Sciences

Quiz 5-Mar-22

Student name: [ رشا علي ابوركاب ]

Student ID: [ 220210069 ]

We will be using the **iris data set** for this tutorial. The data we will use are in a file called **Iris\_Data.csv**

# Question 1

Load the data from the file using the techniques learned today.

Determine the following:

* The number of data points (rows). (*Hint:* check out the dataframe .shape attribute.)

| [in] | import pandas as pd  iris\_data = pd.read\_csv(r"C:\Users\r4092\Downloads\iris\_data.csv",",")  print(iris\_data.shape[0])  or  import seaborn as sns  data = sns.load\_dataset("iris")  print(data.shape[0]) |
| --- | --- |
| output | 150 |

* The column names. (*Hint:* check out the dataframe .columns attribute.)

| [in] | **import** pandas **as** pd  iris\_data **=** pd.read\_csv(r"C:\Users\r4092\Downloads\iris\_data.csv",",")  print(iris\_data.columns)  or  **import seaborn as sns**  **data = sns.load\_dataset("iris")**  **print(iris\_data.columns.tolist())** |
| --- | --- |
| output | Index(['sepal\_length', 'sepal\_width', 'petal\_length', 'petal\_width',  'species'], dtype='object')  or  ['sepal\_length', 'sepal\_width', 'petal\_length', 'petal\_width', 'species'] |

* The data types for each column. (*Hint:* check out the dataframe .dtypes attribute.)

| [in] | import pandas as pd  iris\_data = pd.read\_csv(r"C:\Users\r4092\Downloads\iris\_data.csv",",")  print(iris\_data.dtypes) |
| --- | --- |
| output | sepal\_length float64  sepal\_width float64  petal\_length float64  petal\_width float64  species object  dtype: object |

# Question 2

Examine the species names and note that they all begin with 'Iris-'. Remove this portion of the name so the species name is shorter.

| [in] | iris\_data['species'] = iris\_data.species.str.replace('Iris-', '')  iris\_data.head() |
| --- | --- |
| output |  |

# Question 3

Determine the following:

* The number of each species present. (*Hint:* check out the series .value\_counts method.)

| [in] | num = pd.value\_counts(['sepal\_length' , 'sepal\_width' , 'petal\_length' , 'petal\_width' , 'species'])  print(num) |
| --- | --- |
| output | sepal\_length 1  petal\_width 1  sepal\_width 1  species 1  petal\_length 1  dtype: int64 |

* The mean, median, and quantiles and ranges (max-min) for each petal and sepal measurement.

*Hint:* for the last question, the .describe method does have median, but it’s not called median. It’s the *50%* quantile. .describe does not have range though, and in order to get the range, you will need to create a new entry in the .describe table, which is max - min.

| [in] | num = pd.value\_counts(['sepal\_length' , 'sepal\_width' , 'petal\_length' , 'petal\_width' , 'species'])  iris\_data.mean(axis=1) |
| --- | --- |
| output | 0 2.550  1 2.375  2 2.350  3 2.350  4 2.550  ...  145 4.300  146 3.925  147 4.175  148 4.325  149 3.950  Length: 150, dtype: float64 |

# Question 4

Calculate the following **for each species** in a separate dataframe:

* The mean of each measurement (sepal\_length, sepal\_width, petal\_length, and petal\_width).

| [in] | iris\_data.groupby('species').mean()  sepal\_length = iris\_data['sepal\_length'].mean()  sepal\_width = iris\_data['sepal\_width'].mean()  petal\_length = iris\_data['petal\_length'].mean()  petal\_width = iris\_data['petal\_width'].mean()  print("sepal\_length", sepal\_length , "\n" , "sepal\_width",  sepal\_width , "\n" ,"petal\_length", petal\_length,"\n" , "petal\_width", petal\_width,"\n" ) |
| --- | --- |
| output | sepal\_length 5.843333333333335  sepal\_width 3.0540000000000007  petal\_length 3.7586666666666693  petal\_width 1.1986666666666672 |

* The median of each of these measurements.

*Hint:* you may want to use Pandas groupby method to group by species before calculating the statistic.

If you finish both of these, try calculating both statistics (mean and median) in a single table (i.e. with a single groupby call).

| [in] | iris\_data.groupby('species').median()  sepal\_length = iris\_data['sepal\_length'].median()  sepal\_width = iris\_data['sepal\_width'].median()  petal\_length = iris\_data['petal\_length'].median()  petal\_width = iris\_data['petal\_width'].median()  print("sepal\_length", sepal\_length , "\n" , "sepal\_width",  sepal\_width , "\n" ,"petal\_length", petal\_length,"\n" , "petal\_width", petal\_width,"\n" ) |
| --- | --- |
| output | sepal\_length 5.8  sepal\_width 3.0  petal\_length 4.35  petal\_width 1.3 |

# Question 5

Make a plot of sepal\_length vs sepal\_width using Matplotlib. Label the axes and give the plot a title.

| [in] | import matplotlib.pyplot as plt  import numpy as np  import math  plt.xlabel('Sepal Length' )  plt.ylabel('sepal width' )  plt.title('Sepal Length vs sepal width' )  plt.plot(data["sepal\_length"], data["sepal\_width"], 'b.' ) |
| --- | --- |
| output |  |

That’s all,

Best wishes