

Write-up	Correctness of Program	Documentation of Program	Viva	Timely Completion	Total	Dated Sign of Subject Teacher
4	4	4	4	4	20	

Group A Assignment No. 3

Title of the Assignment: Descriptive Statistics - Measures of Central Tendency and variability

Perform the following operations on any open source dataset (e.g., data.csv)

1. Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative (categorical) variables. For example, if your categorical variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.
2. Write a Python program to display some basic statistical details like percentile, mean, standard deviation etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris-versicolor' of iris.csv dataset.

Provide the codes with outputs and explain everything that you do in this step.

Output :

notebook195b65595b | Kaggle

Explore and run machine learning code with Kaggle Notebooks. Logged out sessions end after 15 minutes. Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Draft Session (14m)

```
[1]: import numpy as np # linear algebra
import pandas as pd

df=pd.read_csv("../input/mall-customers/Mall_Customers.csv")
print("read the csv file")
```

read the csv file

```
[2]: df.head()
```

```
[2]:
```

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Console

Type here to search

7:38 PM 16/03/2022

notebook195b65595b | Kaggle

Logged out session ends in 43 seconds. To save and continue working. Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Draft Session (15m)

```
[3]: df.mean()
```

`/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.`

```
[3]:
```

	CustomerID	Age	Annual Income (k\$)	Spending Score (1-100)
dtype: float64	100.50	38.85	60.56	50.20

```
[4]: df.loc[:, 'Age'].mean()
```

```
[4]: 38.85
```

```
[6]: df.mean(axis=1)[0:4]
```

`/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.`

```
[6]:
```

0	18.50
1	29.75
2	11.25

Console

Type here to search

7:38 PM 16/03/2022

My Quick Converter x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 21 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Run All Code

Draft Session (15m)

[6]:

```
df.mean(axis=1)[0:4]
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
***Entry point for launching an IPython kernel.

[6]:

```
0 18.50  
1 29.75  
2 11.25  
3 30.00  
dtype: float64
```

[7]:

```
df.median()
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
***Entry point for launching an IPython kernel.

[7]:

```
CustomerID    100.5  
Age           36.0  
Annual Income (k$)  61.5  
Spending Score (1-100)  50.0  
dtype: float64
```

[8]:

```
df.median(axis=1)[0:4]
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
***Entry point for launching an IPython kernel.

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 4 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Run All Code

Draft Session (15m)

[8]:

```
df.median(axis=1)[0:4]
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
***Entry point for launching an IPython kernel.

[8]:

```
0 17.0  
1 18.0  
2 11.0  
3 19.5  
dtype: float64
```

[9]:

```
df.mode()
```

[9]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Female	32.0	54.0	42.0
1	2	NaN	NaN	78.0	NaN
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
...
195	196	NaN	NaN	NaN	NaN
196	197	NaN	NaN	NaN	NaN

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

notebook195b65595b Draft saved

File Edit View Run Help

+ Run All Code

Draft Session (15m)

[8]:

```
df.median(axis=1)[0:4]
```

/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
***Entry point for launching an IPython kernel.

[8]:

```
0 17.0  
1 18.0  
2 11.0  
3 19.5  
dtype: float64
```

[9]:

```
df.mode()
```

[9]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Female	32.0	54.0	42.0
1	2	NaN	NaN	78.0	NaN
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
...
195	196	NaN	NaN	NaN	NaN
196	197	NaN	NaN	NaN	NaN

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

My Quick Converter x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 0 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Draft Session (16m)

[9]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Female	32.0	54.0	42.0
1	2	NaN	NaN	78.0	NaN
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
...
195	196	NaN	NaN	NaN	NaN
196	197	NaN	NaN	NaN	NaN
197	198	NaN	NaN	NaN	NaN
198	199	NaN	NaN	NaN	NaN
199	200	NaN	NaN	NaN	NaN

200 rows x 5 columns

[11]:

```
df.loc[:, 'Age'].mode()
```

[11]: 0 32
dtype: int64

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 0 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Draft Session (16m)

[12]:

```
df.min()
```

[12]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
	1	Female	18	15	1

dtype: object

[13]:

```
df.loc[:, 'Age'].min(skipna=False)
```

[13]: 18

[14]:

```
df.max()
```

[14]:

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
	200	Male	70	137	99

dtype: object

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

My Quick Converter x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 0 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Run All Code

Draft Session (17m)

195 1.0 0.0
196 1.0 0.0
197 0.0 1.0
198 0.0 1.0
199 0.0 1.0

200 rows x 2 columns

```
[21]: df_encode=df_u.join(enc_df)
df_encode
```

	CustomerID	Genre	Age	Income	Spending Score (1-100)	0	1
0	1	Male	19	15	39	0.0	1.0
1	2	Male	21	15	81	0.0	1.0
2	3	Female	20	16	6	1.0	0.0
3	4	Female	23	16	77	1.0	0.0
4	5	Female	31	17	40	1.0	0.0
...
195	196	Female	35	120	79	1.0	0.0

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

kaggle.com/scratchpad/notebook195b65595b/edit

Logged out session ends in 0 seconds To save and continue working Sign in or Register

notebook195b65595b Draft saved

File Edit View Run Help

+ Run All Code

Draft Session (17m)

```
[21]: df_encode=df_u.join(enc_df)
df_encode
```

	CustomerID	Genre	Age	Income	Spending Score (1-100)	0	1
0	1	Male	19	15	39	0.0	1.0
1	2	Male	21	15	81	0.0	1.0
2	3	Female	20	16	6	1.0	0.0
3	4	Female	23	16	77	1.0	0.0
4	5	Female	31	17	40	1.0	0.0
...
195	196	Female	35	120	79	1.0	0.0
196	197	Female	45	126	28	1.0	0.0
197	198	Male	32	126	74	0.0	1.0
198	199	Male	32	137	18	0.0	1.0
199	200	Male	30	137	83	0.0	1.0

200 rows x 7 columns

Console

Type here to search

New Tab x kaggle - Saferbrowser Yahoo Indi x notebook195b65595b | Kaggle x +

notebook3ebb305451 Failed to save draft.

File Edit View Run Add-ons Help

+ Draft Session (9m)

```
[1]: import numpy as np # linear algebra
import pandas as pd
df=pd.read_csv("../input/irisass3/iris.data")

[3]: col_names=['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width', 'Species']

[4]: iris=pd.read_csv("../input/irisass3/iris.data", names=col_names)

[5]: irisSet=(iris['Species']=='Iris-setosa')

[6]: print('Iris-setosa')
print(iris[irisSet].describe())
```

Console

iris.data

Share Save Version 0

Data + Add data

Input

irisass3 iris.data

Output (44.1MB / 19.6GB)

/kaggle/working

Settings

Schedule a notebook run

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

notebook3ebb305451 Failed to save draft.

File Edit View Run Add-ons Help

+ Draft Session (9m)

```
[6]: print('Iris-setosa')
print(iris[irisSet].describe())
```

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width
count	50.00000	50.000000	50.000000	50.00000
mean	5.00600	3.418000	1.464000	0.24400
std	0.35249	0.381024	0.173511	0.10721
min	4.30000	2.300000	1.000000	0.10000
25%	4.80000	3.125000	1.400000	0.20000
50%	5.00000	3.400000	1.500000	0.20000
75%	5.20000	3.675000	1.575000	0.30000
max	5.80000	4.400000	1.900000	0.60000

```
[10]: irisVer=(iris['Species']=='Iris-versicolor')

[8]: print('Iris-versicolor')

Iris-versicolor

[11]:
```

Console

iris.data

Share Save Version 0

Data + Add data

Input

irisass3 iris.data

Output (44.1MB / 19.6GB)

/kaggle/working

Settings

Schedule a notebook run

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

WhatsApp | Part A Assignment_No_3.docx.pdf | Redirecting | KAGGLE - Saferbrowser Yahoo Inc | notebook3ebb305451 | Kaggle

kaggle.com/kshitijsinde/notebook3ebb305451/edit

notebook3ebb305451 Failed to save draft. Share Save Version 0

File Edit View Run Add-ons Help

+ Draft Session (10m)

[11]:

```
print(iris[irisVer].describe())
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
count	50.000000	50.000000	50.000000	50.000000
mean	5.936000	2.778000	4.260000	1.326000
std	0.516171	0.313798	0.469911	0.197753
min	4.900000	2.000000	3.000000	1.000000
25%	5.600000	2.525000	4.000000	1.200000
50%	5.900000	2.800000	4.350000	1.300000
75%	6.300000	3.000000	4.600000	1.500000
max	7.000000	3.400000	5.100000	1.800000

[12]:

```
irisVir=(iris['Species']=='Iris-virginica')
```

[13]:

```
print('Iris-virginica')
```

Iris-virginica

[14]:

```
print(iris[irisVir].describe())
```

Console

iris.data

Show all

Type here to search

WhatsApp | Part A Assignment_No_3.docx.pdf | Redirecting | KAGGLE - Saferbrowser Yahoo Inc | notebook3ebb305451 | Kaggle

kaggle.com/kshitijsinde/notebook3ebb305451/edit

notebook3ebb305451 Failed to save draft. Share Save Version 0

File Edit View Run Add-ons Help

+ Draft Session (10m)

[12]:

```
irisVir=(iris['Species']=='Iris-virginica')
```

[13]:

```
print('Iris-virginica')
```

Iris-virginica

[14]:

```
print(iris[irisVir].describe())
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
count	50.000000	50.000000	50.000000	50.000000
mean	6.588000	2.974000	5.552000	2.026000
std	0.635880	0.322497	0.551895	0.274650
min	4.900000	2.200000	4.500000	1.400000
25%	6.225000	2.800000	5.100000	1.800000
50%	6.500000	3.000000	5.550000	2.000000
75%	6.900000	3.175000	5.875000	2.300000
max	7.900000	3.800000	6.900000	2.500000

Console

iris.data

Show all

Type here to search

Write-up	Correctness of Program	Documentation of Program	Viva	Timely Completion	Total	Dated Sign of Subject Teacher
----------	------------------------	--------------------------	------	-------------------	-------	-------------------------------