HOMEWORK
968-252 Data Science
Exploratory Data Analysis: General Information and Summary Statistics

ชื่อ-นามสกุล ธนวัฒน์ วิริยธรรมโสภณ รหัส 6530611033 สาขา COMP

จงเขียนคำสั่ง Python ที่ให้ผลลัพธ์ได้เช่นเดียวกับคำสั่ง R ต่อไปนี้

1) mydata <- read.csv("~/Desktop/Sample/Data/DataScience.csv")

```python
# 6530611033 Lab 2 ทำเป็น python

import csv
import numpy as np
import pandas as pd

#1
file_path = 'D:/ไว้ส่งงาน มอ/data sci/DataScience.csv'
with open(file_path, mode='r', encoding='utf-8') as mydata:
    print(mydata)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.e
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '56939' '
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
PS C:\Users\Admin\Downloads>
```

2) dim(mydata)

```python
#2 dim() in R
df = pd.read_csv(file_path)
dimensions = df.shape
print(dimensions)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! htt

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Micros
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../.
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mo
(19158, 14)
```

3) names(mydata)

```
18  #3 names() in R
19  df = pd.read_csv(file_path)
20  column = df.columns.tolist()
21  print(column)
22
23  # #4 view() in R
```

```
ROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS                Python Debug Console

indows PowerShell
opyright (C) Microsoft Corporation. All rights reserved.

nstall the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

S C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\Admin\.vscode\extensions\ms-
ython.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '56974' '--' 'C:\Users\Admin\Downloads\Untitled-2.p
'
_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
['enrollee_id', 'city', 'city_development_index', 'gender', 'relevent_experience', 'enrolled_university', 'education_level', 'major_discip
ine', 'experience', 'company_size', 'company_type', 'last_new_job', 'training_hours', 'target']
S C:\Users\Admin\Downloads>
```

4) View(mydata)

```
22
23  #4 view() in R
24  view = pd.read_csv(file_path)
25  print(view)
26
27  # #5 summary() in R
28  # df = pd.read_csv(file_path)
29  # summary = df.describe()
30  # print(summary)
31
32  # #6 head() in R
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS                Python Debug Console

PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launch
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
       enrollee_id      city  city_development_index gender  ...    company_type last_new_job training_hours target
0             8949  city_103                   0.920   Male  ...             NaN           1             36    1.0
1            29725   city_40                   0.776   Male  ...         Pvt Ltd          >4             47    0.0
2            11561   city_21                   0.624    NaN  ...             NaN       never             83    0.0
3            33241  city_115                   0.789    NaN  ...         Pvt Ltd       never             52    1.0
4              666  city_162                   0.767   Male  ...  Funded Startup           4              8    0.0
...            ...       ...                     ...    ...  ...             ...         ...            ...    ...
19153         7386  city_173                   0.878   Male  ...             NaN           1             42    1.0
19154        31398  city_103                   0.920   Male  ...             NaN           4             52    1.0
19155        24576  city_103                   0.920   Male  ...         Pvt Ltd           4             44    0.0
19156         5756   city_65                   0.802   Male  ...         Pvt Ltd           2             97    0.0
19157        23834   city_67                   0.855    NaN  ...             NaN           1            127    0.0
```

5) summary(mydata)

```
26
27  #5 summary() in R
28  df = pd.read_csv(file_path)
29  summary = df.describe()
30  print(summary)
31
32  # #6 head() in R
33  # df = pd.read_csv(file_path)
34  # header = df.head()
35  # print(header)
36
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

19155        24576  city_103                   0.920   Male  ...         Pvt
19156         5756   city_65                   0.802   Male  ...         Pvt
19157        23834   city_67                   0.855    NaN  ...

[19158 rows x 14 columns]
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' enc
       enrollee_id  city_development_index  training_hours        target
count  19158.000000            19158.000000    19158.000000  19158.000000
mean   16875.358179                0.828848       65.366896      0.249348
std     9616.292592                0.123362       60.058462      0.432647
min        1.000000                0.448000        1.000000      0.000000
25%     8554.250000                0.740000       23.000000      0.000000
50%    16982.500000                0.903000       47.000000      0.000000
75%    25169.750000                0.920000       88.000000      0.000000
max    33380.000000                0.949000      336.000000      1.000000
PS C:\Users\Admin\Downloads>
```

6) head(mydata)

```
31
32  #6 head() in R
33  df = pd.read_csv(file_path)
34  header = df.head()
35  print(header)
36
37  # #7 head(n=20) in R
38  # df = pd.read_csv(file_path)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS — Python Debug Console

```
std     9616.292592              0.123362       60.058462      0.432647
min        1.000000              0.448000        1.000000      0.000000
25%     8554.250000              0.740000       23.000000      0.000000
50%    16982.500000              0.903000       47.000000      0.000000
75%    25169.750000              0.920000       88.000000      0.000000
max    33380.000000              0.949000      336.000000      1.000000
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\pyt
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher'
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
   enrollee_id      city  city_development_index gender  ...    company_type last_new_job training_hours target
0         8949  city_103                   0.920   Male  ...             NaN           1             36    1.0
1        29725   city_40                   0.776   Male  ...         Pvt Ltd          >4             47    0.0
2        11561   city_21                   0.624    NaN  ...             NaN       never             83    0.0
3        33241  city_115                   0.789    NaN  ...         Pvt Ltd       never             52    1.0
4          666  city_162                   0.767   Male  ...  Funded Startup           4              8    0.0

[5 rows x 14 columns]
PS C:\Users\Admin\Downloads>
```

7) head(mydata, n=20)

```
37  # #7 head(n=20) in R
38  df = pd.read_csv(file_path)
39  header_count = df.head(n=20)
40  print(header_count)
41
42  # # 8 tail() in R
43  # df = pd.read_csv(file_path)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS — Python Debug Consol

```
    enrollee_id      city  city_development_index  gender  ...    company_type last_new_job training_hours target
0          8949  city_103                   0.920    Male  ...             NaN           1             36    1.0
1         29725   city_40                   0.776    Male  ...         Pvt Ltd          >4             47    0.0
2         11561   city_21                   0.624     NaN  ...             NaN       never             83    0.0
3         33241  city_115                   0.789     NaN  ...         Pvt Ltd       never             52    1.0
4           666  city_162                   0.767    Male  ...  Funded Startup           4              8    0.0
5         21651  city_176                   0.764     NaN  ...             NaN           1             24    1.0
6         28806  city_160                   0.920    Male  ...  Funded Startup           1             24    0.0
7           402   city_46                   0.762    Male  ...         Pvt Ltd          >4             18    1.0
8         27107  city_103                   0.920    Male  ...         Pvt Ltd           1             46    1.0
9           699  city_103                   0.920     NaN  ...         Pvt Ltd          >4            123    0.0
10        29452   city_21                   0.624     NaN  ...             NaN       never             32    1.0
11        23853  city_103                   0.920    Male  ...         Pvt Ltd           1            108    0.0
12        25619   city_61                   0.913    Male  ...         Pvt Ltd           3             23    0.0
13         5826   city_21                   0.624    Male  ...             NaN       never             24    0.0
14         8722   city_21                   0.624     NaN  ...             NaN       never             26    0.0
15         6588  city_114                   0.926    Male  ...         Pvt Ltd          >4             18    0.0
16         4167  city_103                   0.920     NaN  ...         Pvt Ltd       never            106    0.0
17         5764   city_21                   0.624     NaN  ...         Pvt Ltd           2              7    0.0
```

8) tail(mydata)

```
46
47  # # 9 tail(n=20) in R
48  # df = pd.read_csv(file_path)
49  # tail_count = df.tail(n=20)
50  # print(tail_count)
51
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS — Python Debug Con

```
16         4167  city_103                   0.920     NaN  ...         Pvt Ltd       never            106    0.0
17         5764   city_21                   0.624     NaN  ...         Pvt Ltd           2              7    0.0
18         2156   city_21                   0.624     NaN  ...         Pvt Ltd       never             23    1.0
19        11399   city_13                   0.827  Female  ...             NaN           1            132    1.0

[20 rows x 14 columns]
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsA
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\lau
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
       enrollee_id      city  city_development_index gender  ... company_type last_new_job training_hours target
19153         7386  city_173                   0.878   Male  ...          NaN           1             42    1.0
19154        31398  city_103                   0.920   Male  ...          NaN           4             52    1.0
19155        24576  city_103                   0.920   Male  ...      Pvt Ltd           4             44    0.0
19156         5756   city_65                   0.802   Male  ...      Pvt Ltd           2             97    0.0
19157        23834   city_67                   0.855    NaN  ...          NaN           1            127    0.0
```

9) tail(mydata, n = 20)

```
46
47  # # 9 tail(n=20) in R
48  df = pd.read_csv(file_path)
49  tail_count = df.tail(n=20)
50  print(tail_count)
51
52  #10 mydata[1:10, ] in R
53  # df = pd.read_csv(file_path)
54  # data = df.iloc[0:10]
55  # print(data)
56
57  # #11 mydata[1:10, 0:3] in R
58  # df = pd.read_csv(file_path)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS — Python Debug Console

```
19142        13750   city_40                 0.776    Male  ...  Early Stage Startup          1             26    0.0
19143        33047  city_103                 0.920    Male  ...             Pvt Ltd         >4             18    0.0
19144        17191   city_21                 0.624     NaN  ...                 NaN      never             48    1.0
19145          155   city_44                 0.725     NaN  ...             Pvt Ltd      never            190    0.0
19146        13167  city_103                 0.920    Male  ...             Pvt Ltd          1             51    0.0
19147        21319   city_21                 0.624    Male  ...             Pvt Ltd          1             52    1.0
19148         9212   city_21                 0.624     NaN  ...             Pvt Ltd          3             40    1.0
19149          251  city_103                 0.920    Male  ...             Pvt Ltd          1             36    1.0
19150        32313  city_160                 0.920  Female  ...      Public Sector          3             23    0.0
19151        11385  city_149                 0.689    Male  ...                 NaN          1             60    0.0
19152        29754  city_103                 0.920  Female  ...     Funded Startup          1             25    0.0
19153         7386  city_173                 0.878    Male  ...                 NaN          1             42    1.0
19154        31398  city_103                 0.920    Male  ...                 NaN          4             52    1.0
19155        24576  city_103                 0.920    Male  ...             Pvt Ltd          4             44    0.0
19156         5756   city_65                 0.802    Male  ...             Pvt Ltd          2             97    0.0
19157        23834   city_67                 0.855     NaN  ...                 NaN          1            127    0.0

[20 rows x 14 columns]
```

10) mydata[1:10,]

```
51
52  #10 mydata[1:10, ] in R
53  df = pd.read_csv(file_path)
54  data = df.iloc[0:10]
55  print(data)
56
57  # #11 mydata[1:10, 0:3] in R
58  # df = pd.read_csv(file_path)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS — Python Debug Con

```
[20 rows x 14 columns]
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsA
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\lau
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
   enrollee_id      city  city_development_index gender  ...    company_type last_new_job training_hours target
0         8949  city_103                   0.920   Male  ...             NaN          1             36    1.0
1        29725   city_40                   0.776   Male  ...         Pvt Ltd         >4             47    0.0
2        11561   city_21                   0.624    NaN  ...             NaN      never             83    0.0
3        33241  city_115                   0.789    NaN  ...         Pvt Ltd      never             52    1.0
4          666  city_162                   0.767   Male  ...  Funded Startup          4              8    0.0
5        21651  city_176                   0.764    NaN  ...             NaN          1             24    1.0
6        28806  city_160                   0.920   Male  ...  Funded Startup          1             24    0.0
7          402   city_46                   0.762   Male  ...         Pvt Ltd         >4             18    1.0
8        27107  city_103                   0.920   Male  ...         Pvt Ltd          1             46    1.0
9          699  city_103                   0.920    NaN  ...         Pvt Ltd         >4            123    0.0

[10 rows x 14 columns]
```

11) mydata[1:10, 1:3]

```
56
57  # #11 mydata[1:10, 0:3] in R
58  df = pd.read_csv(file_path)
59  data = df.iloc[0:10, 0:3]
60  print(data)
61
62  # # 12 unique(mydata[c("gender")]) in R
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
9          699  city_103                  0.920    NaN  ...         Pvt Ltd          >4

[10 rows x 14 columns]
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Loc
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adap
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
   enrollee_id      city  city_development_index
0         8949  city_103                   0.920
1        29725   city_40                   0.776
2        11561   city_21                   0.624
3        33241  city_115                   0.789
4          666  city_162                   0.767
5        21651  city_176                   0.764
6        28806  city_160                   0.920
7          402   city_46                   0.762
8        27107  city_103                   0.920
9          699  city_103                   0.920
```

12) unique(mydata[c("gender")])

```
61
62  # # 12 unique(mydata[c("gender")]) in R
63  df = pd.read_csv(file_path)
64  gender = df["gender"].unique()
65  print(gender)
66
67  # # 13 table(mydata$gender)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe'
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '57079' '--'
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
['Male' nan 'Female' 'Other']
PS C:\Users\Admin\Downloads>
```

13) table(mydata$gender)

```
67  # # 13 table(mydata$gender)
68  df = pd.read_csv(file_path)
69  gender_count = df["gender"].value_counts()
70  print(gender_count)
71
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '57079'
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
['Male' nan 'Female' 'Other']
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Loca
\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapt
\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
gender
Male      13221
Female     1238
Other       191
Name: count, dtype: int64
```

14) genderEducation <- table(mydata$education_level, mydata$gender)

```
70  # print(gender_count)
71
72  # # 14 genderEducation <- table(mydata$education_level, mydata$gender)   in R
73  df = pd.read_csv(file_path)
74  gender_count = pd.crosstab(df["education_level"], df["gender"])
75  print(gender_count)
76
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' '
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '57105' '--' '
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
gender           Female  Male  Other
education_level
Graduate            773  8144    112
High School          67  1395     33
Masters             339  2957     27
Phd                  47   280      4
Primary School        4   203      6
PS C:\Users\Admin\Downloads>
```

15) addmargins(genderEducation)

```
77   # #15 addmargins(genderEducation)
78   df = pd.read_csv(file_path)
79   gender_count = pd.crosstab(df["education_level"], df["gender"],margins= True)
80   print(gender_count)
81
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c
ython.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '57105' '--' 'C:

<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
gender           Female  Male  Other
education_level
Graduate            773  8144    112
High School          67  1395     33
Masters             339  2957     27
Phd                  47   280      4
Primary School        4   203      6
PS C:\Users\Admin\Downloads>
```

16) prop.table(genderEducation)

```
81
82   #16 prop.table(genderEducation)
83   df = pd.read_csv(file_path)
84   gender_count = pd.crosstab(df["education_level"], df["gender"])
85   prop_gender = gender_count.div(gender_count.sum().sum())
86   print(prop_gender)
87
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '571
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
gender            Female      Male     Other
education_level
Graduate        0.053714  0.565909  0.007783
High School     0.004656  0.096936  0.002293
Masters         0.023556  0.205476  0.001876
Phd             0.003266  0.019457  0.000278
Primary School  0.000278  0.014106  0.000417
PS C:\Users\Admin\Downloads>
```

17) testdata <- xtabs(~gender+education_level+major_discipline, data = mydata)

```
88   #17 17) testdata <- xtabs(~gender+education_level+major_discipline, data = mydata)
89   df = pd.read_csv(file_path)
90   testdata = pd.crosstab(index=[df["gender"], df["education_level"]], columns=df["major_discipline"])
91   print(testdata)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS Python Deb

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\Admin\Downloads>  & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users\
python.debugpy-2024.0.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '57144' '--' 'C:\Users\
y'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
major_discipline        Arts  Business Degree  Humanities  No Major  Other  STEM
gender education_level
Female Graduate           25               16          70        14     28   620
       Masters            10                9          42         2      8   268
       Phd                 1                0           6         0      0    40
Male   Graduate          137              169         283       143    176  7229
       Masters            22               54         138        20     65  2654
       Phd                 1                3           8         0      6   262
Other  Graduate            7                3           5         2      6    89
       Masters             0                0           2         1      0    24
       Phd                 0                0           1         0      1     2
PS C:\Users\Admin\Downloads>
```

18) ftable(testdata)

```
93    # #18ftable(testdata)
94    df = pd.read_csv(file_path)
95    ftable = testdata.stack()
96    print(ftable)
97
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
        Masters                0                0
        Phd                    0                0
gender  education_level  major_discipline
Female  Graduate         Arts                  25
                         Business Degree       16
                         Humanities            70
                         No Major              14
                         Other                 28
                         STEM                 620
        Masters          Arts                  10
                         Business Degree        9
                         Humanities            42
                         No Major               2
                         Other                  8
                         STEM                 268
        Phd              Arts                   1
                         Business Degree        0
                         Humanities             6
                         No Major               0
                         Other                  0
                         STEM                  40
Male    Graduate         Arts                 137
                         Business Degree      169
                         Humanities           283
                         No Major             143
                         Other                176
                         STEM                7229
        Masters          Arts                  22
                         Business Degree       54
```

19) install.packages("pastecs")
library(pastecs)
stat.desc(mydata)

20) stat.desc(mydata[,c("city_development_index","training_hours")])

```
108   # #21 mean(mydata$training_hours)
109   df = pd.read_csv(file_path)
110   mean = df['training_hours'].mean()
111   print("Mean = ", mean)
112
113   # #22 median(mydata$training_hours)
114   df = pd.read_csv(file_path)
115   median = df['training_hours'].median()
116   print("Median = ",median)
117
118   # #23 var(mydata$training_hours)
119   df = pd.read_csv(file_path)
120   var = df['training_hours'].var()
121   print("Var = ",var)
122
123   # #24 min(mydata$training_hours)
124   df = pd.read_csv(file_path)
125   min = df['training_hours'].min()
126   print("Min = ",min)
127
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไฟล์งาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
```

21) mean(mydata$training_hours)

```
108   # #21 mean(mydata$training_hours)
109   df = pd.read_csv(file_path)
110   mean = df['training_hours'].mean()
111   print("Mean = ", mean)
112
113   # #22 median(mydata$training_hours)
114   df = pd.read_csv(file_path)
115   median = df['training_hours'].median()
116   print("Median = ",median)
117
118   # #23 var(mydata$training_hours)
119   df = pd.read_csv(file_path)
120   var = df['training_hours'].var()
121   print("Var = ",var)
122
123   # #24 min(mydata$training_hours)
124   df = pd.read_csv(file_path)
125   min = df['training_hours'].min()
126   print("Min = ",min)
127
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไฟล์งาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
```

22) median(mydata$training_hours)

```
108   # #21 mean(mydata$training_hours)
109   df = pd.read_csv(file_path)
110   mean = df['training_hours'].mean()
111   print("Mean = ", mean)
112
113   # #22 median(mydata$training_hours)
114   df = pd.read_csv(file_path)
115   median = df['training_hours'].median()
116   print("Median = ",median)
117
118   # #23 var(mydata$training_hours)
119   df = pd.read_csv(file_path)
120   var = df['training_hours'].var()
121   print("Var = ",var)
122
123   # #24 min(mydata$training_hours)
124   df = pd.read_csv(file_path)
125   min = df['training_hours'].min()
126   print("Min = ",min)
127
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
```

23) var(mydata$training_hours)


```
108   # #21 mean(mydata$training_hours)
109   df = pd.read_csv(file_path)
110   mean = df['training_hours'].mean()
111   print("Mean = ", mean)
112
113   # #22 median(mydata$training_hours)
114   df = pd.read_csv(file_path)
115   median = df['training_hours'].median()
116   print("Median = ",median)
117
118   # #23 var(mydata$training_hours)
119   df = pd.read_csv(file_path)
120   var = df['training_hours'].var()
121   print("Var = ",var)
122
123   # #24 min(mydata$training_hours)
124   df = pd.read_csv(file_path)
125   min = df['training_hours'].min()
126   print("Min = ",min)
127
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
```


24) min(mydata$training_hours)


```
108   # #21 mean(mydata$training_hours)
109   df = pd.read_csv(file_path)
110   mean = df['training_hours'].mean()
111   print("Mean = ", mean)
112
113   # #22 median(mydata$training_hours)
114   df = pd.read_csv(file_path)
115   median = df['training_hours'].median()
116   print("Median = ",median)
117
118   # #23 var(mydata$training_hours)
119   df = pd.read_csv(file_path)
120   var = df['training_hours'].var()
121   print("Var = ",var)
122
123   # #24 min(mydata$training_hours)
124   df = pd.read_csv(file_path)
125   min = df['training_hours'].min()
126   print("Min = ",min)
127
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.exe' 'c:\Users
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
```

25) max(mydata$training_hours)

```python
# #25 max(mydata$training_hours)
df = pd.read_csv(file_path)
max = df['training_hours'].max()
print("Mean = ",max)

# #26 range(mydata$training_hours)
df = pd.read_csv(file_path)
range = df['training_hours'].max()-df['training_hours'].max()
print("Range = ",range)

# #27 quantile(mydata$training_hours)
df = pd.read_csv(file_path)
quantile = df['training_hours'].quantile()
print("Quantile = ",quantile)

# #28 max(table(mydata$gender))df = pd.read_csv(file_path)
maxtable = df['training_hours'].max()
print("Maxtable = ",maxtable)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
Mean =  336
Range =  0
Quantile =  47.0
Maxtable =  336
```

26) range(mydata$training_hours)

```python
# #25 max(mydata$training_hours)
df = pd.read_csv(file_path)
max = df['training_hours'].max()
print("Mean = ",max)

# #26 range(mydata$training_hours)
df = pd.read_csv(file_path)
range = df['training_hours'].max()-df['training_hours'].max()
print("Range = ",range)

# #27 quantile(mydata$training_hours)
df = pd.read_csv(file_path)
quantile = df['training_hours'].quantile()
print("Quantile = ",quantile)

# #28 max(table(mydata$gender))df = pd.read_csv(file_path)
maxtable = df['training_hours'].max()
print("Maxtable = ",maxtable)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
Mean =  336
Range =  0
Quantile =  47.0
Maxtable =  336
```

27) quantile(mydata$training_hours)

```python
# #25 max(mydata$training_hours)
df = pd.read_csv(file_path)
max = df['training_hours'].max()
print("Mean = ",max)

# #26 range(mydata$training_hours)
df = pd.read_csv(file_path)
range = df['training_hours'].max()-df['training_hours'].max()
print("Range = ",range)

# #27 quantile(mydata$training_hours)
df = pd.read_csv(file_path)
quantile = df['training_hours'].quantile()
print("Quantile = ",quantile)

# #28 max(table(mydata$gender))df = pd.read_csv(file_path)
maxtable = df['training_hours'].max()
print("Maxtable = ",maxtable)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
Mean =  336
Range =  0
Quantile =  47.0
Maxtable =  336
```

28) max(table(mydata$gender))


```python
# #25 max(mydata$training_hours)
df = pd.read_csv(file_path)
max = df['training_hours'].max()
print("Mean = ",max)

# #26 range(mydata$training_hours)
df = pd.read_csv(file_path)
range = df['training_hours'].max()-df['training_hours'].max()
print("Range = ",range)

# #27 quantile(mydata$training_hours)
df = pd.read_csv(file_path)
quantile = df['training_hours'].quantile()
print("Quantile = ",quantile)

# #28 max(table(mydata$gender))df = pd.read_csv(file_path)
maxtable = df['training_hours'].max()
print("Maxtable = ",maxtable)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
0
47.0
336
PS C:\Users\Admin\Downloads>  c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Microsoft\WindowsApps\python3.10.
ibs\debugpy\adapter/../..\debugpy\launcher' '57236' '--' 'C:\Users\Admin\Downloads\Untitled-2.py'
<_io.TextIOWrapper name='D:/ไว้ส่งงาน มอ/data sci/DataScience.csv' mode='r' encoding='utf-8'>
Mean =  65.36689633573442
Median =  47.0
Var =  3607.0188487579544
Min =  1
Mean =  336
Range =  0
Quantile =  47.0
Maxtable =  336
```