Machine Learning In Finance

**Lab Logbook Requirement:WEEK-2**

1) Determine a number (n) equal to the last digit of your SID. If the last digit of your SID is '0', then use 10.

2) Group by "relationship" and "hours-per-week". The first grouping should be by relationship.

3) Reduce all "hours-per-week" column values ••in the original DataFrame by the value 'n'. Use a function.

4) Group ••by "relationship" and the reduced "hours-per-week".

5) Add the code and result to your Lab Logbook.

**NOTE: DON'T FORGET TO SAVE AND BACK UP YOUR COMPLETED JUPYTER NOTEBOOK AND LAB LOGBOOK ON GITHUB OR ONEDRIVE.**

**SOLUTION:-**

**import** pandas **as** pd

**import** numpy **as** np

​

*#1Determine a number (n) equal to the last digit of your SID. If the last digit of your SID is '0', then use 10.*

StudentID **=** 2453375

z **=** StudentID **%** 100

x **=** z **+** 100 **if** z **<** 10 **else** z

s **=** np.arange(x)

print("x:", x)

print("s:", s)

​

​

df **=** pd.read\_csv('adult\_data\_mini.csv')

n**=**5

​

grp\_org**=**df.groupby(['relationship','hours-per-week']).size().reset\_index(name**=**'Count')

print("original grouping:",grp\_org)

​

**def** reduce\_hours(p):

**return** p**-**n

​

df['reduce\_hours'] **=** df['hours-per-week'].apply(reduce\_hours)

print("\nDataFrame with 'hours-per-week' reduce by", n)

print(df[['relationship','hours-per-week','reduce\_hours']])

​

grp\_reduce **=** df.groupby(['relationship','reduce\_hours']).size().reset\_index(name**=**'Count')

print("\nGrouping by 'relationship' and reduce 'hours-per-week':", grp\_reduce)

​

​

​OUTPUT:-

​

x: 75

s: [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47

48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71

72 73 74]

original grouping: relationship hours-per-week Count

0 Husband 13 1

1 Husband 40 4

2 Husband 45 1

3 Husband 80 1

4 Not-in-family 16 1

5 Not-in-family 40 2

6 Not-in-family 50 2

7 Own-child 30 1

8 Wife 40 2

DataFrame with 'hours-per-week' reduce by 5

relationship hours-per-week reduce\_hours

0 Not-in-family 40 35

1 Husband 13 8

2 Not-in-family 40 35

3 Husband 40 35

4 Wife 40 35

5 Wife 40 35

6 Not-in-family 16 11

7 Husband 45 40

8 Not-in-family 50 45

9 Husband 40 35

10 Husband 80 75

11 Husband 40 35

12 Own-child 30 25

13 Not-in-family 50 45

14 Husband 40 35

Grouping by 'relationship' and reduce 'hours-per-week': relationship reduce\_hours Count

0 Husband 8 1

1 Husband 35 4

2 Husband 40 1

3 Husband 75 1

4 Not-in-family 11 1

5 Not-in-family 35 2

6 Not-in-family 45 2

7 Own-child 25 1

8 Wife 35 2

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.