

LAB Logbook

Github repository link:

https://github.com/Salimboevm/ML_Finance

Lab 1

For the lab work in week 1, students were asked to create a vector using np.arange method and do some changes on the vector to practice Numpy and Python.

My SID is 1919019, and because of that created 19 elements(last two digits). Then, transformed this vector into a 2 dimensional array with 1 row using the reshape() method. Then, used NumPy's empty_like() method and slicing to create an independent array and save the values of the vector to that independent array. And finally, checked the shape attribute values of both arrays and printed all results at the end of each step.

Results:

```
[302]: vector = np.arange(19)
print(vector)
```

```
[ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18]
```

```
[306]: vector = vector.reshape(19,1)
print(vector)
```

```
[[ 0]
 [ 1]
 [ 2]
 [ 3]
 [ 4]
 [ 5]
 [ 6]
 [ 7]
 [ 8]
 [ 9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]]
```

```
[308]: new_array_2d = np.empty_like(vector)
new_array_2d[:, :] = vector
print(new_array_2d)
```

```
[[ 0]
 [ 1]
 [ 2]
 [ 3]
 [ 4]
 [ 5]
 [ 6]
 [ 7]
 [ 8]
 [ 9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]]
```

```
[310]: print(vector.shape)
print(new_array_2d.shape)
```

```
(19, 1)
(19, 1)
```

Lab 2

Lab 2 in Week 2, Pandas and its main functions were learnt. Requirements were using "adult_data_mini.csv" dataset and performing several operations.

Lab logbook requirement, n was determined as 9 (n=9) because of my student ID (last digit). Then data was grouped by "relationship" and "hours-per-week". Followed by "hours-per-week" column values were reduced by n=9. At this step, the function change_data(x) was created and used. To apply this function to the dataset, the apply() method was used and the original DataFrame was updated. Lastly, grouping by "relationship" and reduced "hours-per-week" was performed again.

Results:

```
Student ID: 1919019

[223]: group_by_hours = data.groupby(['relationship', 'hours-per-week'])
group_by_hours.size()

[223]:   relationship  hours-per-week
      Husband        13.0          1
                  40.0          2
                  45.0          1
                  80.0          1
    Not-in-family    16.0          1
                  40.0          2
                  50.0          2
  Own-child        30.0          1
    Wife           40.0          2
dtype: int64

[227]: def change_data(x):
        return x - 9

data['hours-per-week'] = data['hours-per-week'].apply(change_data)
data

[227]:   age  workclass  fnlwgt  education  education-num  marital-status  occupation  relationship  race  sex  capital-gain  capital-loss  hours-per-week  native-country  Answer  IsHomeles
      0  State-gov  77516.0  Bachelors       13.0  Never-married  Adm-clerical  Not-in-family  White  Male    2174.0     NaN    31.0  United-States  <=50K  False
      1  Self-emp-not-inc  83311.0  Bachelors       13.0  Married-civ-spouse  Exec-managerial  Husband  White  Male      0.0     0.0     4.0  United-States  <=50K  False
      2    Private  215646.0   HS-grad        9.0  Divorced  Handlers-cleaners  Not-in-family  White  Male      0.0     NaN    31.0  United-States  <=50K  False
      3    Private  234721.0      11th        7.0  Married-civ-spouse  Handlers-cleaners  Husband  Black  Male      0.0     NaN    31.0  United-States  <=50K  False
      4    Private  338409.0  Bachelors       13.0  Married-civ-spouse  Prof-specialty  Wife  Black  Female     0.0     NaN    31.0     Cuba  <=50K  False
      5    Private  284582.0    Masters       14.0  Married-civ-spouse  Exec-managerial  Wife  White  Female     0.0     NaN    31.0  United-States  <=50K  False
      6    Private  160187.0      9th        5.0  Married-spouse-absent  Other-service  Not-in-family  Black  Female     0.0     0.0     7.0  Jamaica  <=50K  False
      7  Self-emp-not-inc  209642.0   HS-grad       9.0  Married-civ-spouse  Exec-managerial  Husband  White  Male      0.0     0.0    36.0  United-States  >50K  False
      8    Private  45781.0     Masters       14.0  Never-married  Prof-specialty  Not-in-family  White  Female  14084.0     NaN    41.0  United-States  >50K  False
```

8	31	Private	45781.0	Masters	14.0	Never-married	Prof-specialty	Not-in-family	White	Female	14084.0	NaN	41.0	United-States	>50K	Fals
10	37	Private	280464.0	Some-college	10.0	Married-civ-spouse	Exec-managerial	Husband	Black	Male	0.0	NaN	71.0	United-States	>50K	Fals
12	23	Private	122272.0	Bachelors	13.0	Never-married	Adm-clerical	Own-child	White	Female	0.0	NaN	21.0	United-States	<=50K	Fals
13	32	Private	205019.0	Assoc-acdm	12.0	Never-married	Sales	Not-in-family	Black	Male	0.0	NaN	41.0	United-States	<=50K	Fals
14	40	Private	121772.0	Assoc-voc	11.0	Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0.0	NaN	31.0	?	>50K	Fals
15	25	Private	NaN	Some-college	NaN	NaN	NaN	NaN	White	Male	0.0	NaN	NaN	NaN	NaN	Tru


```
[229]: group_by_reduced_hours = data.groupby(['relationship', 'hours-per-week'])
group_by_reduced_hours.size()
```

```
[229]: relationship    hours-per-week
Husband          4.0              1
                 31.0             2
                 36.0             1
                 71.0             1
Not-in-family   7.0              1
                 31.0             2
                 41.0             2
Own-child       21.0             1
Wife            31.0             2
dtype: int64
```

```
[ ]:
```

Lab 3

Lab 4

Lab 5

Lab 6

Lab 7

Lab 8

Lab 9

Lab 10

Lab 11

Lab 12