

# 1 Index Explained

*If your system doesn't have pandas installed, use ! before your script, it can run the cell like a command window*

```
In [1]: ! pip install pandas
```

executed in 4.34s, finished 18:35:46 2019-02-19

...

**Our exercise dataset can be downloaded directly with pandas help from uci ml data repository**

Dataset Name: Adult

```
In [2]: import pandas as pd
```

```
adult = pd.read_csv("https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data",
                    names = ['age', 'workclass', 'fnlwgt', 'education', 'education_num', 'marital_status',
                              'occupation', 'relationship', 'race', 'sex', 'capital_gain', 'capital_loss',
                              'hours_per_week', 'native_country', 'label'], index_col = False)
```

```
print("Shape of data{}".format(adult.shape))
```

```
adult.head()
```

executed in 6.54s, finished 18:35:53 2019-02-19

Shape of data(32561, 15)

Out[2]:

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capital_loss	hours_per_week	native_country
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40	United-States
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13	United-States
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States