

**East West University**

**Assignment Report: 02**

**Course Title:** Data Mining

**Course Code:** CSE477

**Section:** 1

**Submitted To:**

Dr. Mohammad Rezwanul Huq

Assistant Professor,   
Dept. of Computer Science & Engineering

**Submitted By:**

Shaykh Siddique

ID: 2016-1-60-053

**Report:**

**Tool Used:** Anaconda Jupyter

* **Number of Records:**

The File adult.data contains 32561 records.

* **Number of features:**

With class feature it has 15 features.

* **Continuous, and nominal features:**

Six of features are continuous - age, fnlwgt, education-num, capital-gain, capital-loss and hours-per-week.

Seven features are nominal – workclass, education, marital-status, relationship, race and native-country.

Two are Binary sex and salary is greater than or less than 50K.

* **Calculation of average, median, maximum, minimum values and standard deviation:**

Average:-

Age: 38.58

Fnlwgt: 189778.37

Education Num: 10.08

Capital Gain: 1077.65

Capital Loss: 87.30

Hours-per-week: 40.44

As total number of data 32561, so the median will be:

|  |  |
| --- | --- |
| Age:-  Median : 37  Maximum : 90  Minimum : 17 | Capital Gain:-  Median : 0 (lots of zero)  Maximum : 99999  Minimum : 0 |
| Fnlwgt:-  Median : 178370  Maximum : 1484705  Minimum : 12285 | Capital Loss:-  Median : 0 (lots of zero)  Maximum : 4356  Minimum : 0 |
| Education Num:-  Median : 10  Maximum : 16  Minimum : 1 | Hours-per-week:-  Median : 40  Maximum : 99  Minimum : 1 |

Standard deviation:-

Age: 13.64

Fnlwgt: 105548.36

Education Num: 2.58

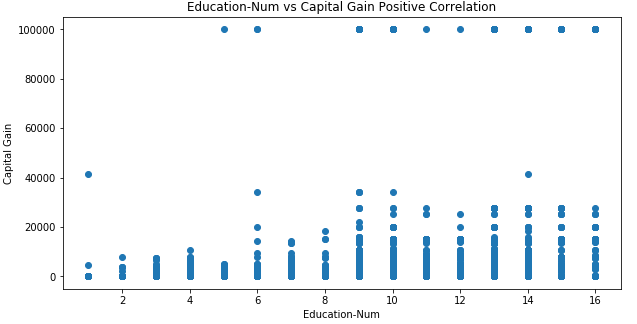
Capital Gain: 7385.18

Capital Loss: 402.95

Hours-per-week: 12.35

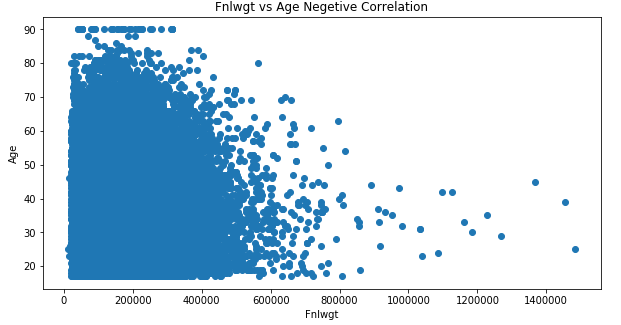
* **Scatter plots:**

Education-Num vs Capital Gain Scatter plot-



We can see here is positive correlation, so if the value of Education-Num increase the value of Capital Gain will also increase.

Fnlwgt vs Age Scatter plot-



If the age increase, the value of weight will decrease and vice versa. So these two variables are negatively correlated.

* **Dissimilarity between record 1 and record 53:**

Record1 - ['39', ' State-gov', ' 77516', ' Bachelors', ' 13', ' Never-married', ' Adm-clerical', ' Not-in-family', ' White', ' Male', ' 2174', ' 0', ' 40', ' United-States', ' <=50K']

Record2: ['47', ' Private', ' 51835', ' Prof-school', ' 15', ' Married-civ-spouse', ' Prof-specialty', ' Wife', ' White', ' Female', ' 0', ' 1902', ' 60', ' Honduras', ' >50K']

Dissimilarity: 0.286