**Executive Summary**

* This is representative sample of over 18000 individuals from 5000 families in the United State.
* There are many outliers are available and unnecessary data were removed because it helps to get an accurate analysis.
* After transformation of the data, we can do descriptive statistics for analysis of this data set.
* The below table illustrates the descriptive statistics of the data set.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Seq No** | **intnum** | **persnum** | **age** | **educatn** | **earnings** | **hours** | **kids** |
| **Mean** | **2428.5** | **4598.1** | **59.2** | **38.5** | **16.4** | **14244.5** | **1235.3** | **4.5** |
| **Median** | **2428.5** | **5464** | **4** | **38** | **12** | **11000** | **1517** | **2** |
| **SD** | **1402.0** | **2762.0** | **79.7** | **5.6** | **18.4** | **15985.4** | **947.2** | **14.9** |
| **Max** | **4856.0** | **9306.0** | **205.0** | **50.0** | **99.0** | **240000.0** | **5160.0** | **99.0** |
| **Min** | **1.0** | **4.0** | **1.0** | **30.0** | **0.0** | **0.0** | **0.0** | **0.0** |
| **Q1** | **1214.8** | **1905.0** | **2.0** | **34.0** | **12.0** | **85.0** | **32.0** | **1.0** |
| **Q3** | **3642.3** | **6655.0** | **170.0** | **43.0** | **14.0** | **22000.0** | **2000.0** | **3.0** |

* stats library of scipy is used to test the mean of earnings against the null hypothesis with the mean of **14244.5.** It can be seen that the p-value = 0.9999 is greater than alpha=0.05, hence we can’t say that the average is not equal to **14244.5.**
* Using is matplotlib library and searborn modules, we get different visualization.