[Ask a Manager Survey.csv]

Eng. AhMED nOUSIR

Ahmed Nehad Mohamed Shehata 42010704

Ali Hussien Sayed Ahmed 42010308

Abuzaid AbuElsaad Saad Abdul 42010650

Mariam Ahmed Mohamed 42020011

Fatma Abdel Fattah Shaaban 42010649

Fatma Gamal Esayed 42010622

2023

**Problem Definition:**

The main goal of this dataset is to represent salary of different jobs among the world, as its hard-to-get real world information about what jobs pay, where online salary websites are often inaccurate, moreover it’s hard to ask people directly about their salary, therefore this dataset contain different ranges of salaries from different people among the world, furthermore it was filled anonymously

Even thought this dataset contain much data that represent information from the responses as the age, industry, salary, years of professional work experience, gender, race, etc.

Even this dataset contains much of useful information, however it’s really hard to use this dataset is unclean, moreover, very random to make operations on

Therefore, this dataset needed much of preprocessing

to be used.

**Methods:**

* Renaming the columns into more meaningful and easier names to deal with
* Then data was reduced by removing the duplicated values and rows
* After that all the types of currency were integrated into one type of currency, so that dealing with it later become meaningful
* Then the all the countries were integrated to several countries only
* After That some columns which were written as ranges or into nonnumerical form, were transformed into numerical integrals
* Eventually some useless columns and rows were removed to be left with clean data to start processing on it

**Experiment:**

After preparing the data we started the visualization phase, where using some visualization tools we made some deductions, as:

* The boxplot which showed that the average age was between 30 and 40 years old
* Through a plot of statistical analysis, it was decided that the computing or tech was the most common industry among people also the most common job title was the software engineering
* Through a graph among the educational level the collage degree was the most common among people
* Also, it was deduced that the woman average salary is much large than the man salary (through a pie chart)
* However, through a statistical analysis it was found that men are more experienced than women
* Also, a barplot has shown that the people with the highest salary were those with six years of experience
* Another, barplot has shown that most of the respondents were American race

On the other hand, we made a machine learning model, using linear regression, that predict the average experience of the respondent in a specific field, through having the age, level of education, annual salary and years of experience as an input feature with an accuracy of 63%.

**References:**

<https://www.askamanager.org/2021/04/how-much-money-do-you-make-4.html>