Problem statement

In the recruitment domain, HR faces the challenge of candidates getting their salaries high by using fake experience. So we need to predict that by their experience they are really getting high salaries or not. We have to perform these predictions analysis.

Column names	Type of data	Relevance
Position of the employee	categorical	This column has a different types of positions names
no Years of Experience as an employee	Continuous	How much experience
monthly income of the employee	Continuous	monthly income

Data Pre-processing

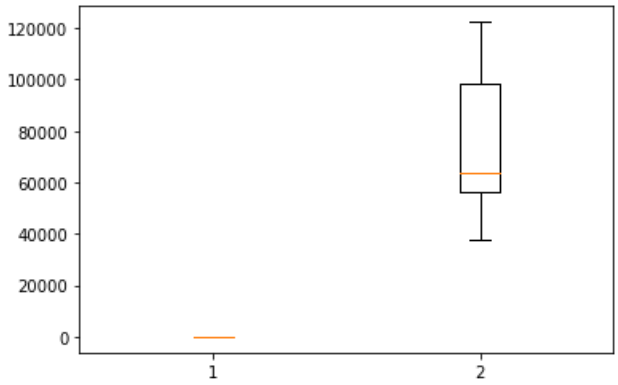
- Except position of employee remaining columns are numeric

- Checking if there are any null, nan, or duplicated values. the data type is numeric or objective
- Then I did normalize the data by using the norm function

Exploratory Data Analysis

Univariate Analysis

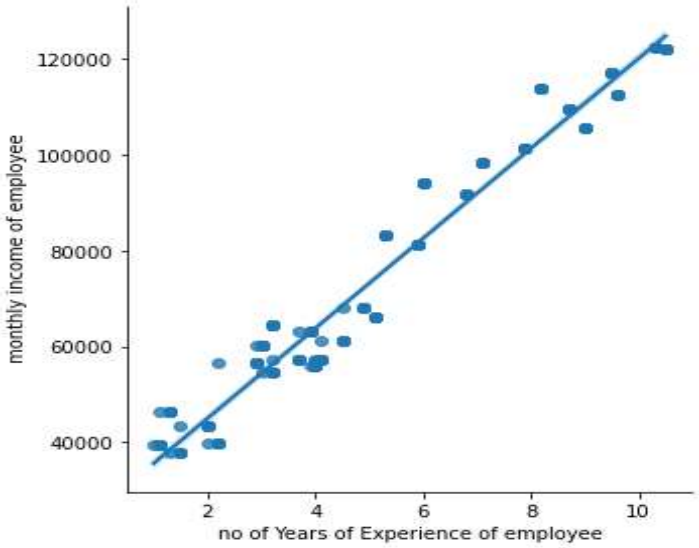
- By using a boxplot we can visually observe which columns have the outliers



- The hist plot to see the distribution of the dataset but this time we are using this visualization to see the changes that we can see after those null values are removed from the dataset and we can see the difference

Bivariate analysis.

- By using a regression plot we can see the employee& experiences



- Then I performed label encoding for the position column to know the total no of positions and the formation of numeric data

Model Building

- In the model building I took the target variable as a monthly income of employee
- column and the predicted variable as the remaining columns. Then I split it into training and testing it at 70% &30%.
- I buildeded the decision tree for a max depth is 5. I predicted the data.
- for training predicted accuracy data I got 85%
- for testing predicted accuracy is 62%

Model for Random forest

- Now after building the model let's check the accuracy of the model on the training dataset.
- In this model for the test predicted accuracy is 30%
- For training predicted accuracy was 98%

GridSearchCV

- by using the gridsearchcv model in the random forecast model classifier of 500 estimators
- for test accuracy of prediction is 83%
- for training predicted accuracy is 98%. So, this model has a high variance. This model is given good results in the prediction.

Conclusion:

- After performing the decision tree classifier and random forecast model. random forecast model gives good results in that grid search model given 83% accuracy of the data.

- 1st person was a business analyst, he has 1 year of exp and he got a 39k income which means he is not fake. He doesn't have 5 years of experience
- 87th person he is a manager he has more than 9 years of experience his actual salary is 1lakh so he can expect a more salary then previous.

Benefits:

- With help of this model we can easily say whether they are genuine or not because of this model they can take whomever they need an experienced person they can give salary as per that.