**CASE-STUDY-USING LINEAR REGRESSION**

**BANK GENDER DISCRIMINATION**

The Following Analysis has been performed on the data:

1. **Univariate Analysis:**

Checked the count of the values using bar charts in the categorical columns and for numerical columns checked its distribution

**Insights:**

Got insights about the distribution of the data and count of values in the given categorical columns

**2. Bivariate Analysis:**

Conducted an analysis of each independent variable with the salary using bar charts and scatter plots (for two continuous numerical columns).

Insights:

1. The average salary of employees rose with the increase in education level

2. With the Increase in the grade of the employees there is an increase in the average salary. ie. The higher was the grader higher the average salary.

3. There is a positive correlation between years 1 and salary but not strong enough. correlation between salary and years1 is 0.61.

4. There is very less positive correlation between age and salary.

5. The average salary for males was higher than for females.

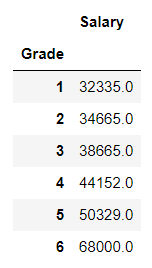
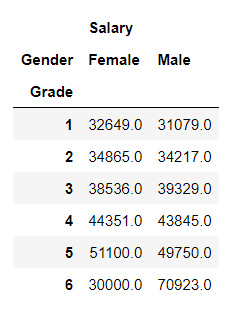
6. There isn’t much difference is salary whether there is PC job or not

**3. Multivariate Analysis:**

Performed multivariate analysis using pivoted tables and also analyzed the data by segregating the data into male and female.

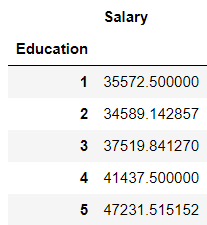
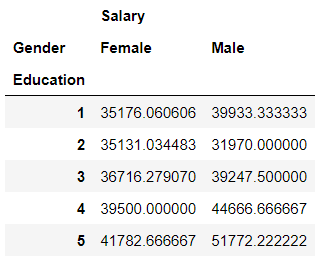
Insights:

1.The average salary for males and females for different grades. And the average salary for different grades.



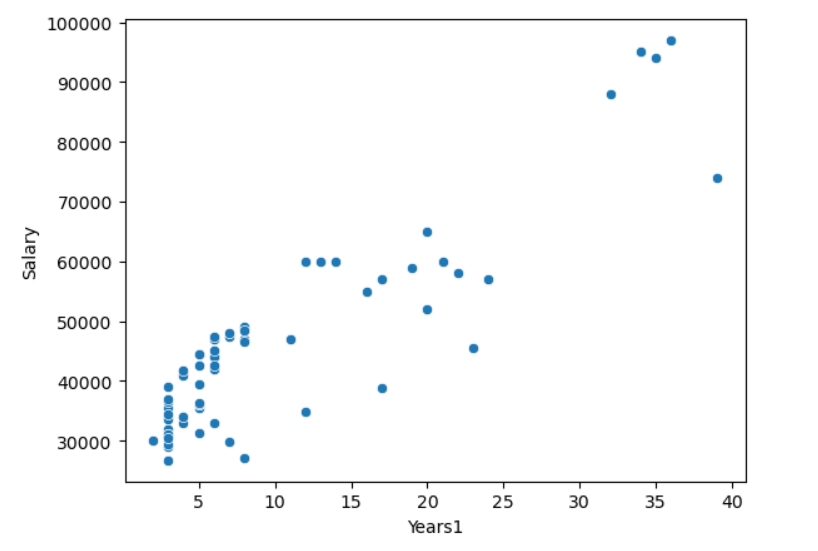
The salaries for male and female was the same for all levels when compared to the average salary as a whole except for level 6 where the average salary of men was higher when compared to women.

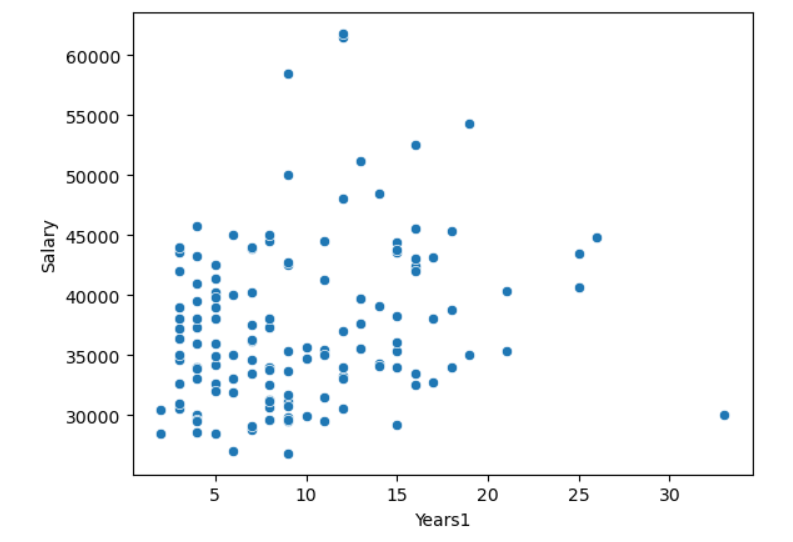
2.Comparing education level , salary and gender



Except for level 2, men had a higher salary than women at all education levels and their average were higher than the overall salary average for different education levels.

3.Scatter plot for male years1 and salary

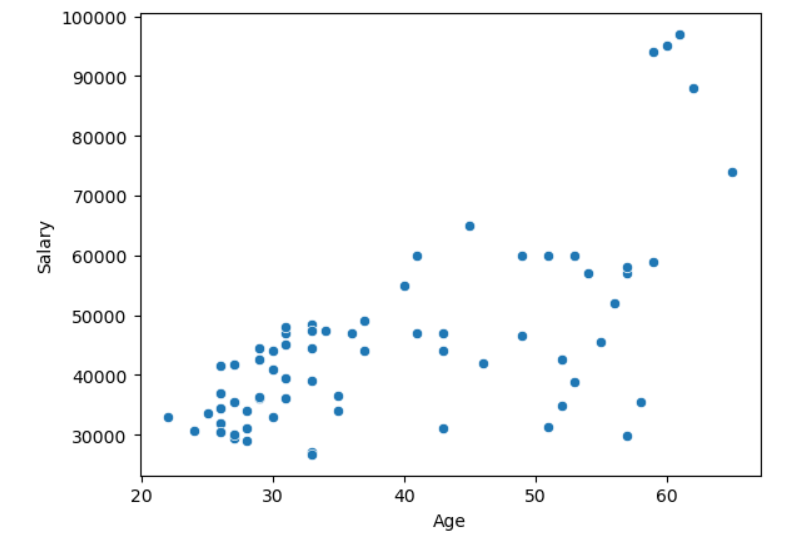


Previously we saw that years1 and salary had a positive correlation But when we check for only men there is a strong positive correlation between years 1 and salary. Correlation =0.88

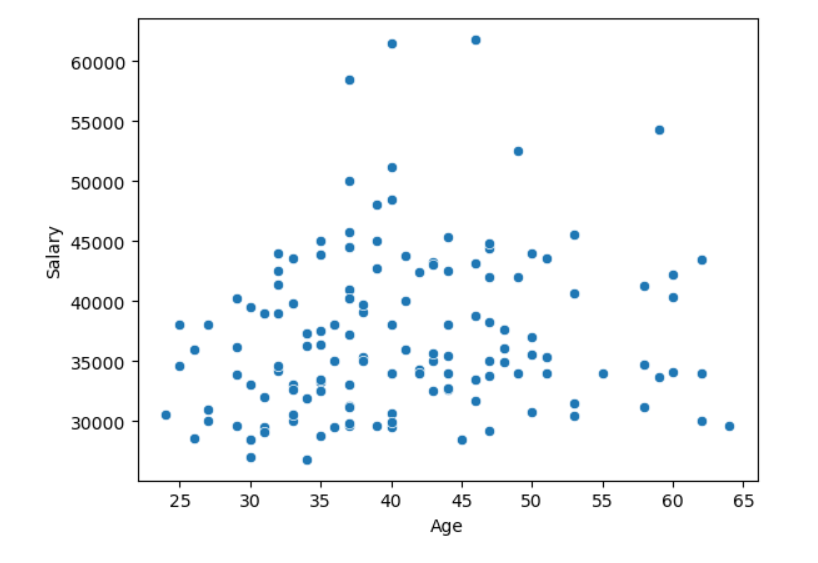
But for women, there isn’t a strong correlation between age and salary. Correlation =0.23

4. Age, Salary and Gender:

Male Age and Salary:

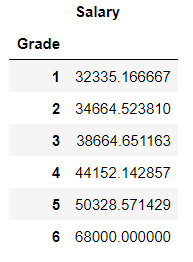
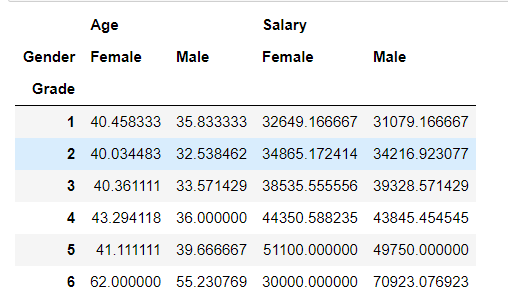


Female Age and Salary:



Male Age and Salary have a better correlation than female age and salary.

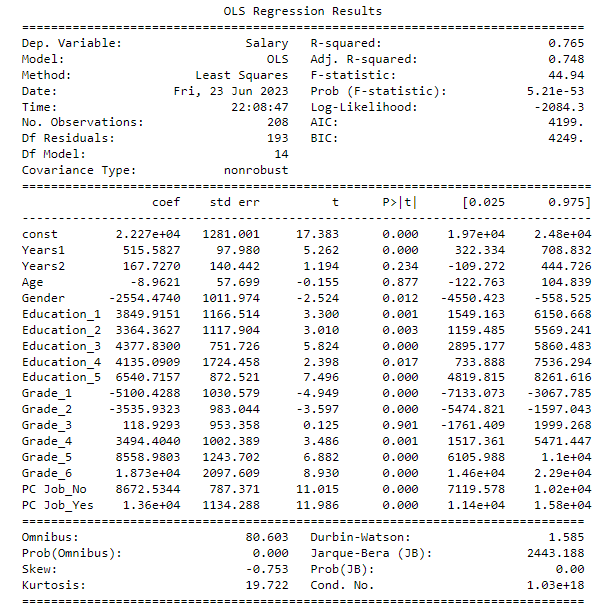
5. Gender, Age, Salary, and Job Grade:



The average salary for women and men is similar to the overall average salary for men and women combined. But the age of women in each education level is higher when compared to that of men who are receiving similar levels of salaries.

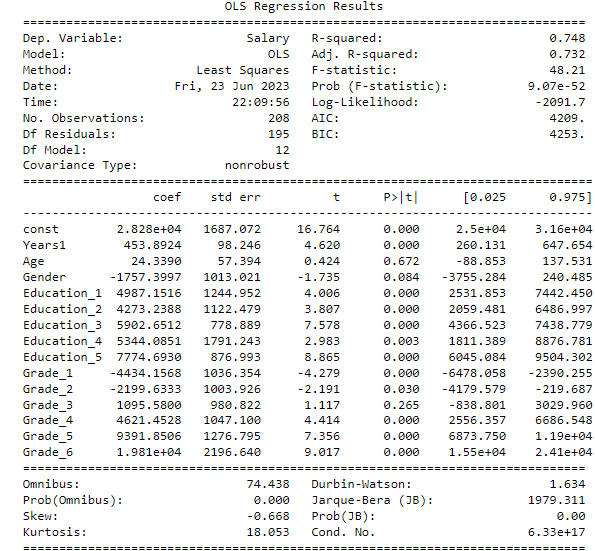
Summary statistic for Linear Regression with gender as a feature and a constant:

For gender I have encoded Female as 1 and male as 0:



The coefficient for gender is -2554(approx.) which means when the gender is female the target variable reduces by -8295 whereas for males the effect is nullified as the gender becomes zero.

Fitted a linear regression model with Gender, Education Level, Age, Years1, and Grade( on basis of eda):



The coefficient for gender is -1757(approx.) which means when the gender is female the target variable reduces by -1757 whereas for males the effect is nullified as the gender becomes zero.

**Observation and interpretation**

\* Yes, their is discrimination between gender, and we have evidence i.e thier difference in pay of salary

\* Their is also gender discrimination on basis of promotion as men tend to have lower age compared women have higher age, comparatively the avg age gap btw men and women is 10 years

\* With feature (Gender, Education-Level(1-5), Age, Years1, Grade-level(1-5), PC job)

The coefficient for gender is -2554(approx.) which means when the gender is female the target variable reduces by -2554 whereas for males the effect is nullified as the gender becomes zero

The coefficient for grade\_1 is -5100(approx.) which means when the grade\_1 is female the target variable reduces by -5100 whereas for males the effect is nullified as the gender becomes zero

The coefficient for grade\_2 is -3535(approx.) which means when the gender is female the target variable reduces by -3535 whereas for males the effect is nullified as the gender becomes zero

\* Best variation of model explain by above these set of of feature after removing insignificant feature (['Salary','Years2',"Age",'Grade\_3'])

\* With some other in combination of interaction effect

Gender with combination job grades(promotion) are also discriminated, as the gender and grades coefficient remain negative

Gender with combination education, years1, years2 and age are discriminated. These are shown by negative Coeff in model after combining them in interaction effect

**Conclusion:**

The important insights covered from visuals and the regression model regarding the presence /absence of gender discrimination:

1. For grade six there the male salary was much higher than the female salary
2. Men with higher years1 of experience had a higher chance of getting a better salary. but the same can’t be said for women as the years 1 and salary had less correlation.
3. Higher grades indicated a higher average salary, but the average age of women in each age is higher than the men but both were getting similar levels of salary.
4. One-way anova indicated that the average salary of women where significantly different from men. ( but does not indicate which is higher than which one.)
5. While fitting a linear model with gender as a feature and salary as the label (where gender F was 1 and M was 0, ) the value of the target variable decreased by a significant level indicating the decrease in salary in the presence of female )

From these above points, we can infer from the sample that there exists gender discrimination in the bank and that women are getting a lower salaries than men