Project : Glassdoor Gender pay gap

Description: This project is about the analysis of Gender Pay gap from Glassdoor data and apply linear regression model to predict the salary of a particular person.

Content: I have created 4 files for this project where every parts shows 4 types of analysis about the dataset.

1. Data

I have taken the data from Kaggle and data contains the information about 1000 people where their education, age, gender, seniority level, work performance, salary, bonus, job title are given.

2. Data Wrangling and Visualization

In this part I have done the data cleaning, added an extra column as total pay (basepay+bonus) and tried to analyse every column one by one and visualise it in order to see the variation of every column relation to the gender pay gap. In short how other features in the dataset influence the salary per gender.

3. Hypothesis Testing

In this part I have done some hypothesis testing in order to prove the statistically significance difference for the Gender pay gap.

4. Models

I have used linear regression model to predict the salary by using the features from the dataset and find the most significant features which influence the salary given to a particular employee according to gender. And satisfied all the assumptions.