

# Exploratory Analysis--Summary

Ria Pan, James Barbour, Fahmi Khalid

## What we have tried:

- We used the IBM HR Analytics Employee Attrition & Performance data from Kagle to explore the correlation between features and to find the common pattern in the dataset in order to uncover the factors that lead to employee attrition. By doing that, our goal is to build a predictive model that can predict employee attrition with high accuracy
- We tried to explore the dataset by answering the following questions:
  - The gender gap between workers of different fields
    - Does the salary(monthly income) between gender significantly difference?
    - Does male and female involved into different roles?
  - Does Marital Status have a significant effect on wages?
  - How does performance rating contribute to wage(monthly income)?
  - Does work life balance concur with years in position or years with company?  
How about with the monthly income?
- Created shiny app to display our results, as well as constructed the dashboard for final project submission
- Created github so everyone can push code on it

## Some results we found:

- In IBM, average monthly salary of female(6686.566) is higher than that of male (6380.508), which is out of what we expected
- Across all job roles, the different pay between women and men are marginal
- On higher position such as manager or manufacturing director, number of female and male are pretty much equal. But positions of sales executive and research scientist have attracted more male employees than females
- As for marital status, people married or divorced have higher income than people are single, which is interesting to see. It's also reasonable because they may have longer working years and higher position level. In this case, we plan to explore if job position concur with marital status
- The performance rating in the dataset is either 3 or 4 (good or excellent), maybe everyone at IBM is good at their work.
- There is no strong correlation between work life balance and years in company or years in position. Also monthly income has trivial influence on work life balance
- There is positive linear relationship between years at company and years since last promotion, with a significant adjusted  $R^2$  0.382

## What we'll do next

- Continue running on feature selection to delete irrelevant ones

- Continue on model selection and compare the accuracy between different models
- Perform logistic regression on employees' attrition
- Once we finish all analysis, we will put everything in shiny app to form a complete project