1. **Describe the categories of information that is collected, and how the data is reported (binary, continuous, categorical, etc). (10 points)**

There is a dataset of 1470 employees with 35 variables, which can be utilized by the head of R&D to analyze the possibility that experienced talents who have performed well intend to leave. In the dataset there are 961 employees from R&D department (others are from the department of sales (446 people) and the department of human resources (63 people)).

- Attrition

One of the most important variables is Attrition, which is a binary variable and denotes whether or not an employee is still working for the company. Among all 961 employees, 828 is still working for the company, while 133 employees have left. This variable can be used to group employees and compare the difference in different groups.

- Demographic variables
  - Gender: It is a binary variable, which has two categories: male and female
  - Over18: a binary variable denotes whether or not the employee is over 18 years old. All employees in the dataset are over 18.
  - Age: a continuous variable shows the age of all employees, ranging from 18 to 60.
  - Employee Number: a continuous variable shows the id of each employee and is distinct for each person.
  - Education: it is a categorical variable denotes all employees' education levels including Below College, College, Bachelor, Master, and Doctor.
  - Education Field: a categorical variable shows the education field of all employees, including Life Science, Medical, Technical Degree, and others.
  - Marital Status: it is a categorical variable shows the marital condition of all employees with three levels: Married, Single and Divorced.
- Working experience

Two continuous variables Number Companies Worked (ranging from 0 to 9) and Total Working Years (ranging from 0 to 40) are used to describe the working experience of employees.

- Working Condition

There are some **continuous variables** like **Training Times Last Year** (ranging from 0 to 6), **Years at Company** (ranging from 0 to 40), **Years in Current Role** (ranging from 0 to 18), **Years Since Last Promotion** (ranging from 0-15), **Years within Current Manager** (ranging from 0 to 17), along with some **categorial variables** like **Business Travel** (with three levels: Non-travel, Travel-rarely, Travel-frequently), **Work Life Balance** (4 levels), **Job Role** (6 levels: Research Scientist, Laboratory Technician, Manufacturing Director, Healthcare Representative, Research Director, Manager), **Job**

Level (5 levels), **describing** the working environment and content of their work inside the company.

Also, there is a **continuous variable Distance from Home** (ranging from 1 to 29) used to describe the convenience for them to go to work.

- Working Hour
  - Standard Hours: it is a variable to describe the standard working hours for each employee, it could be categorial or continuous. In this case all employees' standard working hours are 80.
  - Over Time: It is a binary variable to denote if the employee works in addition to their normal time. The overtime ratio of all employee is 28.2%, while the ratio of people who have left the company is 55.6%

- Benefits

There are some **continuous variables** (Daily Rate, Hourly Rate, Monthly Income, Monthly Rate, Percent Salary Hike) used to measure the salary information of all employees.

Also, there is a **categorical variable Stock Option Level** (4 levels: 0,1,2,3), which is used to describe their additional benefits.

- Assessment

The dataset has five categorical variables (Environment Satisfaction, Job Involvement, Job Satisfaction, Performance Rating, Relationship Satisfaction) which used to assess satisfaction and performance of employees. Each of the variables has four levels: Low, Medium, High, and Very High.

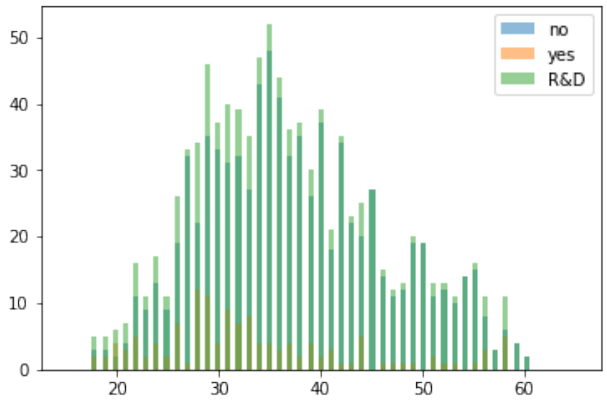
2. **Summarize the key metrics for the R&D division (e.g., total salary for R&D, number of employees, demographics, etc). Show your analysis. (10 points)**

Here are the KPI I chose for the R&D division.

- Demographic variables
  - Gender: It is a binary variable. Based on the statistics, we can find that male employees are more likely to leave the company.

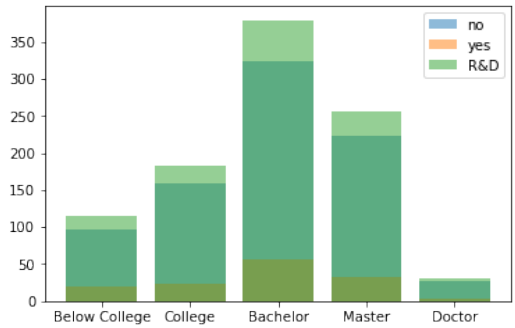
	Male	Female	Ratio(#Male/#Female)
R&D	582	379	1.54
Attrition = Yes	90	43	2.09
Attrition = No	492	336	1.46

  - Age: the age of all employees ranges from 18 to 60. From the graph, it is clear that most of the employees age from 25 to 45. Most of those who left the company are around 30.



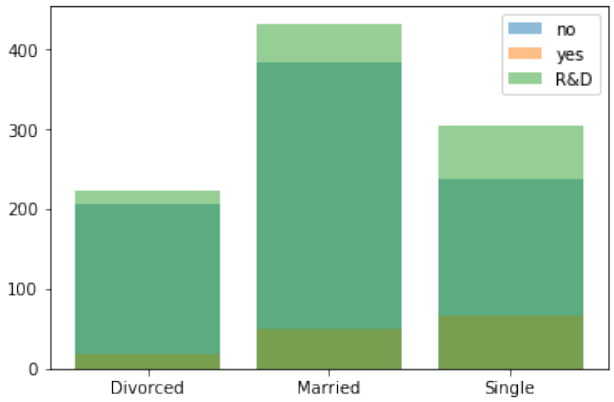
Distribution of age

  - Education: it is a categorical variable denotes all employees' education levels. People with bachelor degrees are more likely to leave the company.



Distribution of education

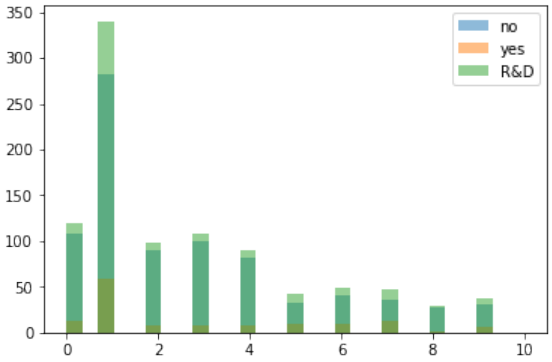
- Marital Status: it is a categorical variable shows the marital condition of all employees, and it is easy to find that single employees are more likely to leave the company.



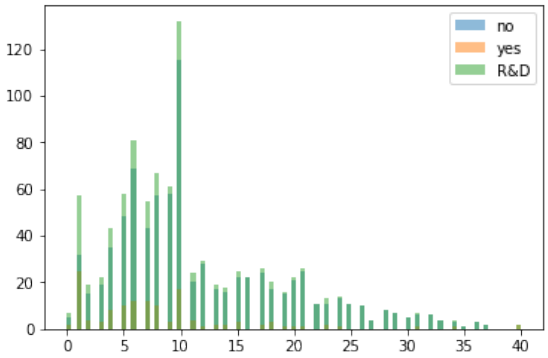
Distribution of Marital Status

- Working experience

Two continuous variables Number Companies Worked and Total Working Years are used to describe the working experience of employees. We can find that most employees' working experience is no more than ten year, and the companies they have worked for are no more than one.



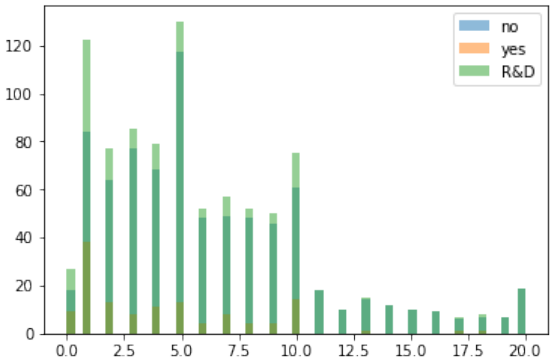
Number Companies Worked



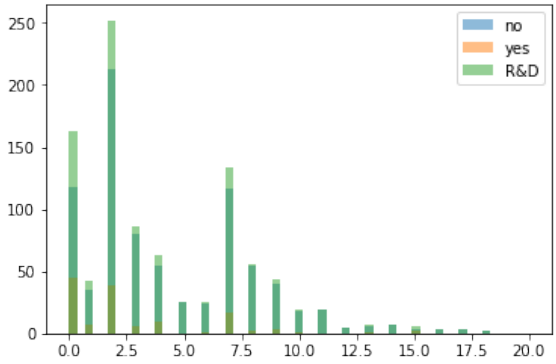
Total Working Years

- Working Condition
  - Years at Company. According to the plot, people who work for the company less than 3 years are more likely to leave the company.
  - Years in Current Role. Based on the plot, it is easy to find that people are less likely to leave the company, if they work for more than 2 years in their current position.
  - Years Since Last Promotion, because the plot shows that people are likely to leave, if they cannot get promoted in 2 years.

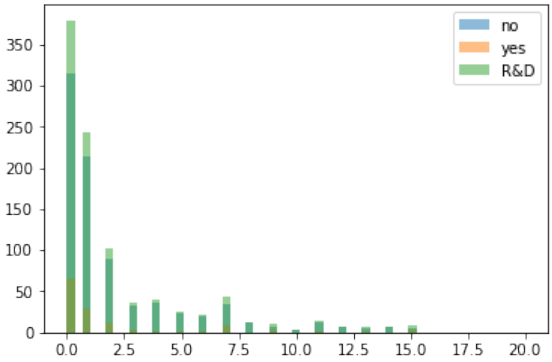
- Job Role. The plot shows that Research Scientists and Laboratory Technicians are more likely to leave.
- Job Level. Based on the graph, people whose work is in the lowest level.



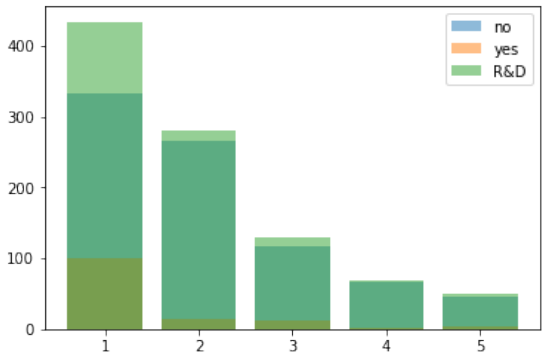
Years at Company



Years in Current Role

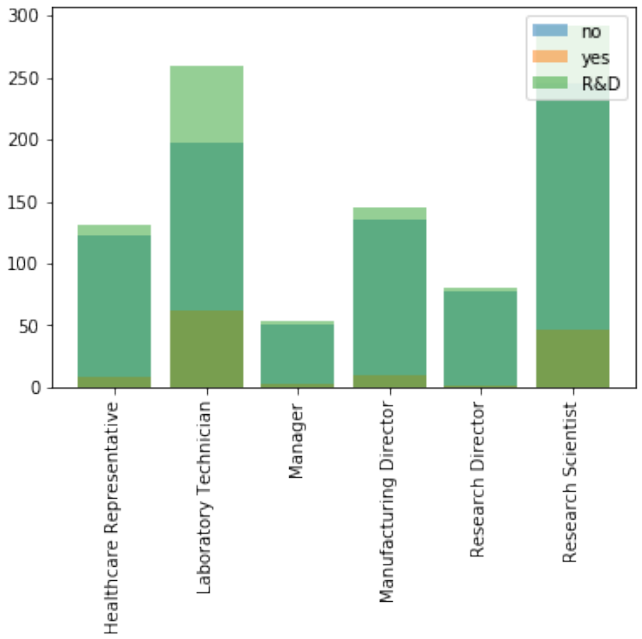


Years Since Last Promotion



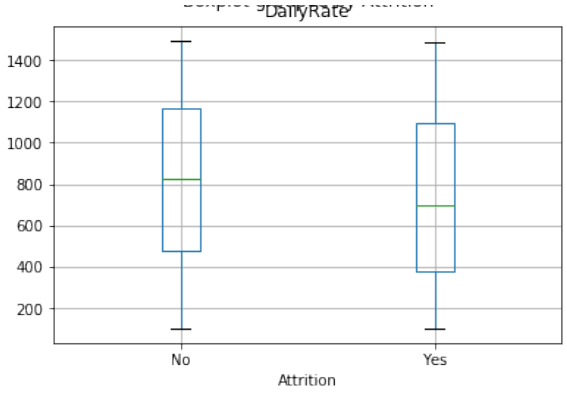
Job Level

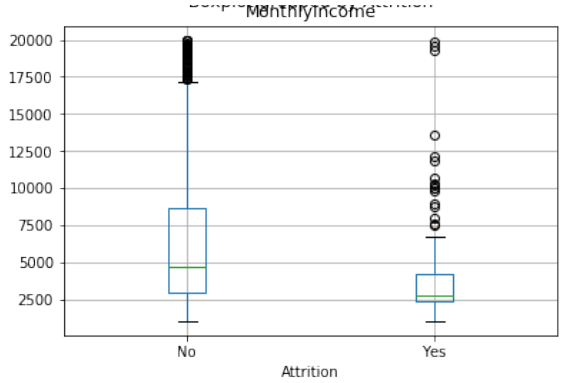
- Working Hour
  - Over Time: It is a binary variable to denote if the employee works in addition to their normal time. The overtime ratio of all employee is 28.2%, while the ratio of people who have left the company is 55.6%



Job Role

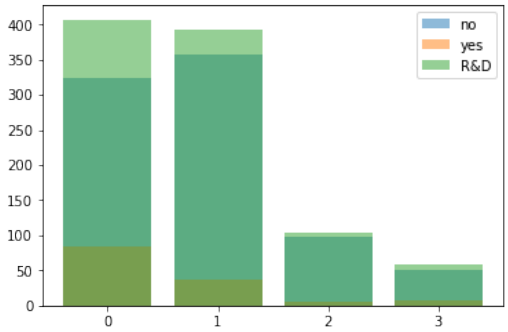
- Benefits
  - Two variables: Daily Rate and Monthly Income. Among all variables about salary, only the distribution of these two variables are significantly different between these two groups of people.





Daily Rate Monthly Income

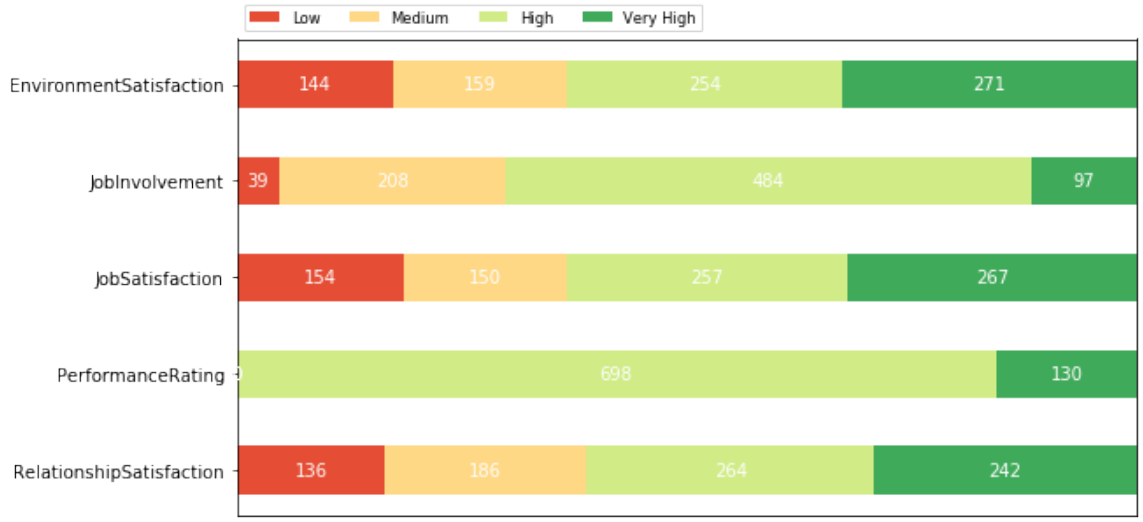
  - Stock Option Level, which is used to describe their additional benefits. It is obvious that people who don't have stock are more likely to leave the company.



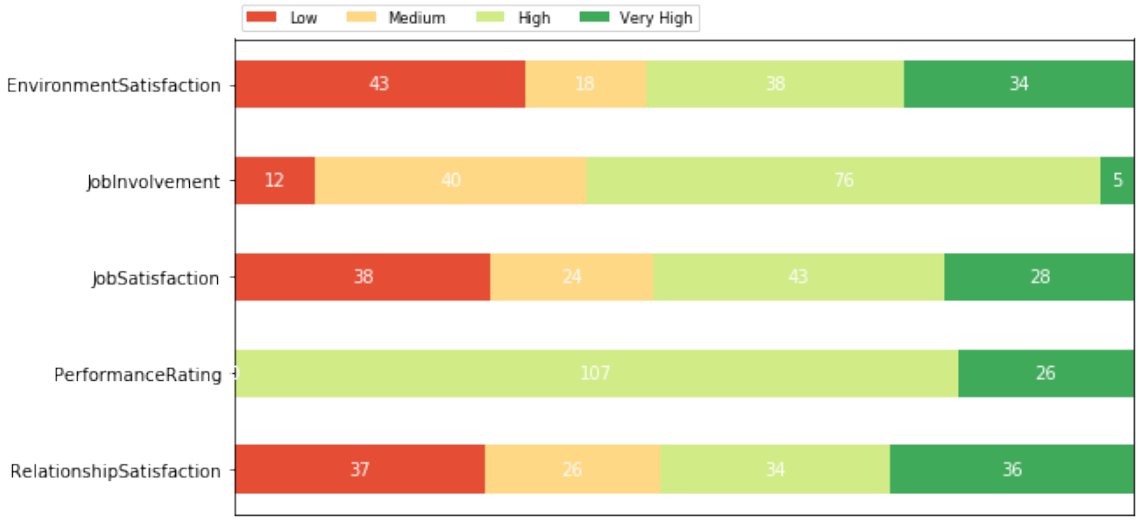
Stock Option Level

- Assessment

All five categorical variables (Environment Satisfaction, Job Involvement, Job Satisfaction, Performance Rating, Relationship Satisfaction). From the plots, we can find that the portion of high satisfaction and high performance among people who is stilling working for the company is slightly larger than the portion among people who have left.



Attrition = No



Attrition = Yes

3. **Do the data support the assertion of the head of R&D (i.e. the relationship between experience and performance)? Explain why or why not. (10 points)**

The head of R&D assert that those experienced talents who performance well and have been with the organization for a long time are more likely to leave the company.

However, I think the conclusion is incorrect.

Firstly, the performance of employees can be measured by salary, stock option, and the performance rating.

- The salary variables like Daily Rate and Monthly Income indicate that people whose salary is low are more likely to leave the company.
- The Stock Option Level variable also indicates that people who don't enjoy the stock benefits are more likely to leave.
- The performance rating shows that the average performance of people who has left is slightly better than the average of people who is still working for the company. However, the difference is very small, and also all employees are rated as 'high or 'very high', which is not a very persuasive variable.

Secondly, the working time can be measured by variables like Years at Company and Years in Current Role.

- However, the variable Years at Company shows that people who have worked for more than 10 years are less likely to leave. On the contrary, those who worked for only one year are more likely to leave.
- Also, the variable Years in Current Role has the similar conclusion: people are less likely to leave, if they have worked in their current role for more than 3 years.

4. **Identify at least one subset of employees who are at disproportionate risk for leaving based on the data provided. (Hint: It may be helpful to use pivot tables for your analysis.) (15 points)**

The most straightforward analysis is shown in the answer of Q2, and my conclusion is that people who have the following characteristics are more likely to leave the company along with my reasoning:

- Male
- Education: with a bachelor degree. Those people are competitive in job searching market and have more options, so they may want to change a job.
- Single. It is likely that those people don't have a future plan, so they are more likely to change their jobs.
- New in the company (less than 3 years). They might think the company does not meet their requirement.

- New in current role (less than 2 years). They are more likely to find the new position difficult or does not meet their requirement.
- Job Role is Research Scientist or Laboratory Technician. Maybe they are more competitive in the market.
- Low job level. They are more likely to search for better working opportunities.
- Work over time.
- Whose salary is low and cannot enjoy stock option benefits.
- Who are unsatisfied with working environment, relationship and their jobs.

5. What recommendations would you make to the head of R&D to address the risk factors for attrition that you identified in Q3? (15 points)

Suggestions

Firstly, arrange more training for employees who recently joined the company or switched to a new position. For one thing, those people might have more difficulties than those who have been working in this company for a long time. Necessary training can eliminate worry and pressure for them and help them get used to their new jobs. For another, communication can be helpful at the very beginning, because communication can eliminate misunderstanding. Also, if new employees are not satisfied with their current jobs, there still have chances for them to make some adjustment.

Secondly, for those who have make some excellent performance, the company can consider increase their salary. Even if it is not possible to offer stock option, there are still some other kind of benefits that can be given.

Thirdly, lay more emphasis on the education, for example offer some high-level professional training to both entry level and high-level employees, so that they can feel that they are making progress and their future is promising. This is also a good method to prevent high-level technical people leaving, who are important to the company.

Fourthly, recruit more employees to decrease the workload of current employees, so that they don’t have to work over time.

Lastly, organize some activities to add more interaction to improve the relationship between employees. So, they can work with harmony and efficiency.

6. **For one of your recommendations, how would you design a rapid experiment to learn whether your recommendation successfully mitigates attrition? What other information would you want to track and how would you track and report it? (15 points)**

I want to try my third recommendation, offering some high-level professional training to both entry level and high-level employees.

When doing this experiment, I want to deliver surveys each month to track the following information:

- Are you interested in the training?
- Are you tired of taking extra training?
- Do you think your workload increased?
- Do you think the training is inspiring for your further work?
- Do you have any other field that you are interested in?

After tracking this information each week, I will analyze and write a report on their responses to make sure that they feel rewarding, workload is not increased too much, they are interested in the topics we provide.

Besides, I will also deliver the survey about their satisfaction, and compare if they feel more satisfied than before.

Finally, my report will also compare the outcome (how many people left during this period, what are their roles, how many years they have worked, etc.) with the last few months' situation, competitors' situation and last year's situation to find out if this recommendation really helps.