

Research Proposal

Introduction: Attribution Can Be a Big Problem

Many companies are facing the problem of recruiting and retaining talents while at the same time having to deal with talent loss through attrition. Losing talents and employees result in performance losses and can have a long-term negative effect on the companies. Besides, attrition might leave gaps in its execution and human resource team, and causes loss of work team harmony and social goodwill. There are many factors that might result in attrition, issues of salary, lack of career mobility and challenges, working environment, just to name a few.

In this research, we will look into the data of IBM HR Analytics Employee Attrition & Performance. By studying employee profiles, we would like to answer what factors are more likely to cause attrition, and what potential methods can be applied to reduce attrition rate.

About the Data

- ✓ Data link:
- ✓ What the data is about?
- ✓ Why this data is interesting?

The data we will work on is titled "IBM HR Analytics Employee Attrition & Performance", which can be found at the following link: <https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>.

The dataset consists of 35 columns(variables) and 1471 rows(records). The sample size is large enough. The variables are likely to provide information for a meaningful study. Based on what we are mostly interested in, we might work on variables including "Attrition", "YearsAtCompany", "Gender", "Department", etc.

The data is interesting because it might answer what factors would cause attrition, and thus helps the company with its decision making. For example, if our study showed that a male employee has significantly higher attrition rate than a female employee, then the company might concern about this result when having two job candidates with everything else the same but one as male another female. Another scenario can be when the company has only one promotion opportunity and has to choose between one male and one female employee. In addition, the data might provide insights and point out those employees that are more likely to quit. So that, the company can intervene with such employees and prevent them from leaving.

About the Research Design

- ✓ What's the research/business question(s)?
- ✓ Identify the hypothesis?
- ✓ How to test the hypothesis?

In this study, we are interested in the question "Are male employees at the company more likely to quit the job?" To answer this question, we would study the attrition rate index, and compare it between male employees and female employees at the company. A higher attrition rate would indicate a higher probability to quit. A related question, but not the same and would require a different research, would be "Do male employees stay less years at the company?" To answer this question, we would look into

“year at the company” index. A smaller value of “year at the company” would indicate an employee would stay less year at the company. Note: the phrase “at the company” emphasizes that the population is the employees at the company we are studying, not the employees of every company.

Restate our first question: “Are male employees at the company more likely to quit the job?” Following that, the hypothesis would be the following:

Ho: male employees and female employees at the company have the same attrition rate.

Ha: male employees and female employees at the company do NOT have the same attrition rate.

Restate our second question: “Do male employees stay less years at the company?” Following that, the hypothesis would be the following:

Ho: the mean “year at the company” value of male employees and female employees at the company are the same.

Ha: the mean “year at the company” value of male employees and female employees at the company are NOT the same.

We can conduct A/B test for Both hypothesis no.1 and no.2. For hypothesis no.1, since the variable of interest is sample proportion, we can use T-test to test our hypothesis. Note that the default T-test is a two-tailed test, however, it might be more convincing to use one-tailed T-test for the sample proportion test.

For hypothesis no.2, we need to first check normality of the variable of interest in the sample dataset, then based on whether they follow normal distribution choose T-test or Kruskal-test.

Who Are the Audience/ Who Might Be Interested in This Study?

- ✓ Who are the users/stakeholders can be benefited?
- ✓ Why they can be benefited?

The ultimate goal of this study is to reduce employee attrition rate. So, anybody who can directly or indirectly benefited from this outcome would be our targeting audiences. To be more specific,

HRs: benefited from reducing the amount of time and effort on job selection process, and be more involved in other important tasks like payroll responsibility, employee development, employee engagement, workforce planning, etc.

Department managers/ team leaders: people at such positions need to concern about assigning work and promote those employees who are likely to make the most contribution to the company. And the bottom line of an employee can make contribution to the company is that he/she stays at the company.

Team workers (which are, likely, everybody): Team works like a chain. When one link of the chain is missing because of attribution, the team would not work at all. Besides, people do not feel motivated when having to do repeating work, and repeatedly training new employees is one of them.