**Group Name: Team Tenet**

**Group Members:**

WAMBUI NJOROGE

MERCY CHERUIYOT - Group Leader

JOSEPH NYINGI

IAN MUGAMBI

# **Business Understanding**

## Business Overview

### People are the lifeblood of an organization. Companies need employees who are fit to do their work. But there’s a reality that every organization, big or small, has to face: movement within the workforce and teams is inevitable. Employees come, and employees go.

Employee attrition refers to employees’ loss through a natural process, such as retirement, resignation, elimination of a position, personal health, or other similar reasons.

We will be looking into the factors that lead to employee attrition in a company using a dataset that was fictionally created by data scientists at IBM to explore similar cases.

### **How it affects businesses**

The major factor in negative attrition is costs. The process of losing and replacing an employee usually has both direct and indirect costs. Filling in one position can cost several thousand shillings between exit interviews, job postings, interview processes, new hiring paperwork, and training. Plus, you have indirect costs that result from lower production by a newer employee until he settles into the job. Additionally, newer, inexperienced employees can make mistakes and cost you customers and business.

On a more positive note, if the new employee is more qualified, capable, and productive than his predecessor, the change is positive. Along with talent upgrades, turnover can bring new blood to the workplace and prevent or combat staleness in the company culture. Companies that never have turnover can become set in their ways and be reluctant to consider and discuss new ideas or evolving trends. Losing a longtime employee and gaining a talented, knowledgeable, and forward-thinking employee can protect against a company becoming a dinosaur before its own eyes.

The key to success in any organization is attracting and retaining top talent. As an analyst, one of the critical tasks is determining which factors keep employees at the company and prompt others to leave.

Employee attrition over time can be a challenge for management. To decrease attrition, managers must understand the causes of employee turnover, the costs associated with attrition, and, finally, institute measures to reduce attrition rates.

### **Reasons for study**

The data is a set of data points on the employees currently working within the company or have resigned. The objective is to identify and improve these factors to prevent the loss of good employees.

## Business Objectives

1. To determine the monthly income means for attritioned employees with different education levels.
2. To establish the education level that had the most attritioned employees.
3. To determine the monthly income earned by different genders that got attritioned.

## Hypothesis

###### H0: There is a difference between monthly incomes made by attritioned employees with a master’s degree and those with a bachelor's degree.

###### H1: There is no difference between monthly incomes made by attritioned employees with a master’s degree and those with a bachelor's degree.

### **Assessing the Data**

1. **Resource Inventory**

* **Dataset** - IBM HR Analytics Employee Attrition and Performance

<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

* **Software** 
  + - Github
    - Google Colaboratory
    - Python
    - Trello
* **Personnel**
  + - Team Tenet

1. **Assumptions**

The available dataset was complete, and there were no data losses.

All possible factors that lead to employee attrition were recorded in the dataset.

1. **Constraints**

There are no constraints while working on the dataset.

# **2. Data Understanding**

Team Tenet is working on a fictional dataset created by IBM to explore similar cases in employee attrition.

IBM HR Analytics Employee Attrition and Performance - This dataset gives information on employee age, job satisfaction, gender, number of years at the current position, attrition, education, and environment satisfaction.

## Data Description

This dataset contains 35 columns and 1,471 rows sourced from Kaggle. Dataset link:

<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>

### **Verifying Data Quality**

The dataset had no missing values. There were also no known data errors in the dataset.

# **3. Data Preparation.**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the dataset from the CSV to a python notebook.

#### **Cleaning Data**

* + - The dataset had no null values. We started the dataset cleaning by checking the;
    - ***Validity*** - The columns dropped include; Over18, Employee count, relationship satisfaction, stock option level, business travel, daily rates, and years at company were done away with.
    - ***Accuracy***. The dataset had outliers in the Monthly Income column.
    - ***Completeness***. The dataset has no missing values and null values as well.
    - ***Consistency*** no duplicate found in the dataset.
    - ***Uniformity of data***- we renamed values in the Education column ( 1: ‘Below\_College,’ 2: ‘ College,’ 3: ‘Bachelor,’ 4: ‘Masters,’ 5: ‘ Doctor’). We dropped irrelevant columns to remain with variables related to income and attrition.

Our final dataset after cleaning had 1283 rows and 16 columns.

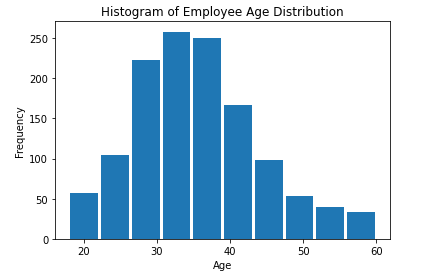
# 

# 

# **4. Analysis**

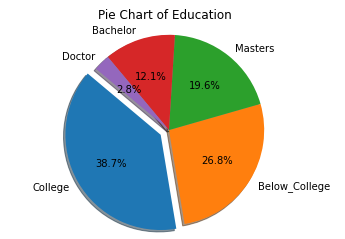
## Univariate analysis.

1. Histogram of Employee age distribution.



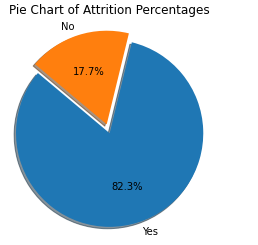
**Observation:** The median age is between 30 and 40 years, and it’s normally distributed.

1. Pie Chart to show the percentage distribution of education levels of employees.



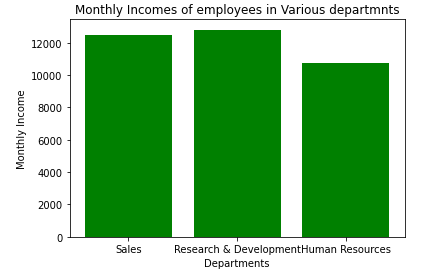
**Observation:** The highest percentage is 38.7% of the employees who happened to have attended college.

1. Pie Chart to show the distribution of the attrition count percentage.



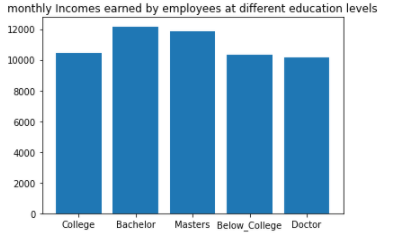
**Observation**: yes represents the bigger percentage with 82.3%, No = 17.7%

1. Bar chart for employees Monthly Income per Department.

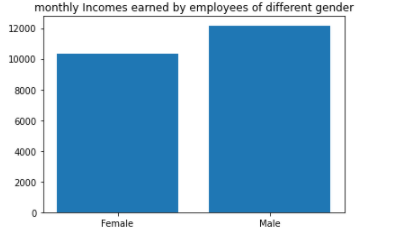


**Observation:** The research and development department had the highest number of employees.

1. Bar graph of Monthly Incomes earned by Attritioned employees at different education levels.

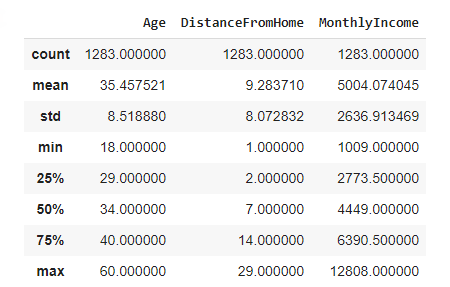


1. Bar graph of Monthly Incomes of employees by Gender.



**Observation:** Male employees earned more than the female employees in the company

Below is a diagram of central tendency measures of columns (Age, DistanceFromHome, and Monthly Income).



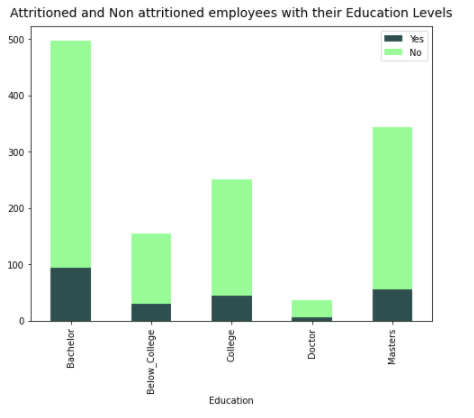
## Bivariate Analysis

Bivariate analysis has been initialized by performing a correlation between;

* Age and Attrition = -0.132
* Monthly Income and Attrition = -0.128

**Observation:** correlation is at -0.1, which is too low.

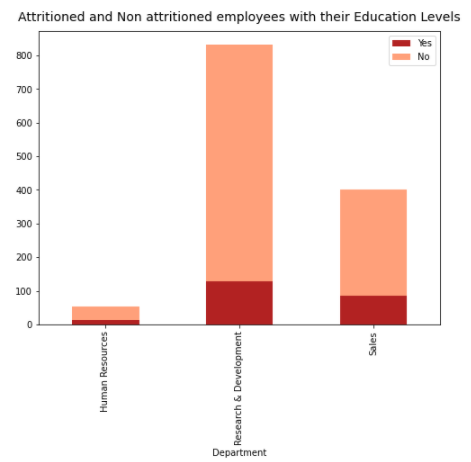
1. Stacked bar to show Education levels of employees, both attritioned and non-attritioned.



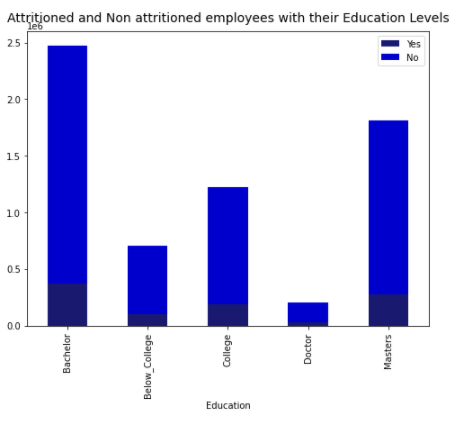
**Observation:** Bachelor had the highest number of employees followed by Master's, College, Below College, and finally, those with a Doctorate.

1. Stacked Bar Chart on Employees and Departments.

**Observation**: The research department had the highest number of employees with low attrition



1. Stacked Bar Chart showing the Monthly Income levels among Attritioned and Non-attritioned Employees.



**Observation:** The highest sum of income levels was observed among the employees with a Bachelor’s degree, and the lowest was those with a doctorate.

**Multivariate Analysis**

### **Principal Component Analysis**

Principal component analysis, or PCA, is a statistical technique to convert high dimensional data to low dimensional data by selecting the most important features that capture maximum information about the dataset.

1. **Preprocessing**

Divide the dataset into a feature set and corresponding labels. Where we declared our X and Y variables.

1. **Divide data into training and test sets.**

Splitting the dataset into the Training set and Test set

1. **Normalization**

PCA performs best with a normalized feature set. We conducted standard scalar normalization to normalize our feature set.

## Specifying the Hypothesis

Below were the hypothesis to be tested;

###### H0: There is no difference between monthly incomes made by attritioned employees with a master’s degree and those with a bachelor's degree.

###### H1: There is a difference between monthly incomes made by attritioned employees with a master’s degree and those with a bachelor's degree.

## Sampling Approach.

The sampling technique we used is a systematic sampling. Systematic sampling is a sampling technique where samples are chosen at regular intervals of a population.

***Why for this study?***

Because of its simplicity, and can be viewed as superior because it improves the potential for the sample units to be more evenly spread over the population.

## Hypothesis Testing

We used a two-sample z-test and p-value to either accept or reject our null hypothesis to perform hypothesis testing. Below are the reasons why we choose to use the z-test as the right test statistic:

• Data points are independent of each other.

• The sample size is greater than 30.

• The sample data has been randomly selected from a population, so there is an equal chance of being selected for each item.

The significance level is defined as the probability of rejecting the null hypothesis when it is true. For this analysis, the significance level chosen is 0.05 or 5%.

If the calculated P-value from the test statistic is less than 0.05, we reject the null hypothesis. If the value is greater than or equal to 0.05, then we accept the null hypothesis.

Since α = 0.05, and the test statistic is a one-tailed test, the critical value of z is 1.645, as per the z critical value table.

# **5. Summary and Conclusion**

The null and alternate hypothesis was successfully defined. Performed the sampling technique and carried out hypothesis testing, which resulted in the acceptance of the null hypothesis.

The P-value was 1.0. greater than 0.05; hence we fail to reject the null hypothesis. This confirms the claim that the monthly incomes of attritioned employees with a master’s degree were different from those with a bachelor's degree.

Tableau Link.

<https://public.tableau.com/profile/joseph.nyingi#!/vizhome/Team_TenetAttritionProject/Monthlyincomeearnedbyemployeesdifferenteducationlevels>

https://github.com/Mjcherono/Employee\_Attrition-\_hypothisis\_project.git