

**TASK**

**Exploratory Data Analysis on the Employee Attrition Dataset**

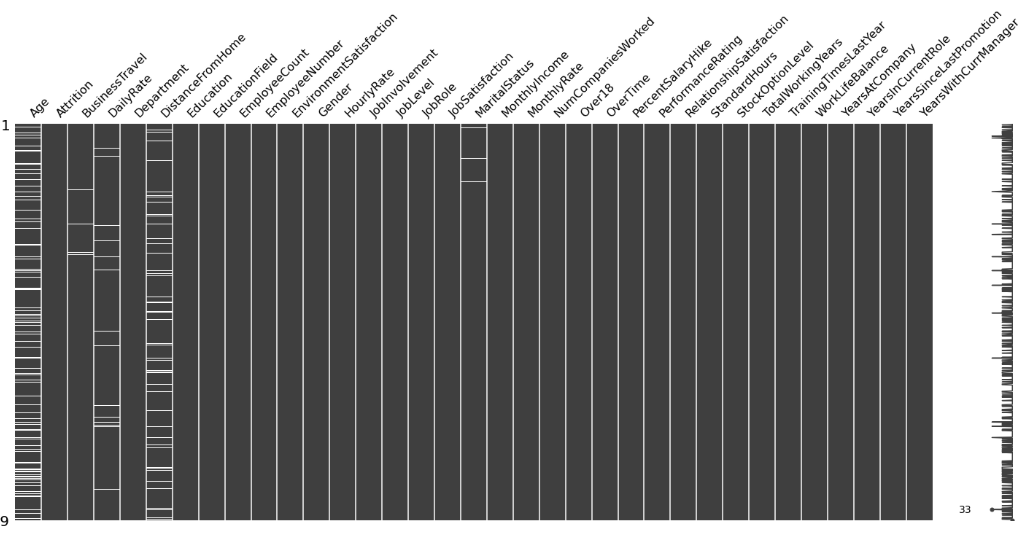
[](http://www.hyperiondev.com/portal/)

**Introduction**

Employee Attrition is a data set that contains information about employees and various characteristics related to their employment. The data includes details such as the employee's age, the total number of years they have worked, the number of years they have been in their current role, their department and job title, the distance from their home to work, and their income and salary. This data set can be used to analyze trends in employee retention and identify factors that may be contributing to employee turnover. It can also be used to develop strategies for improving employee retention and reducing attrition rates within an organization.

**MISSING DATA**

Cleaning the data in the Employee Attrition dataset involved several steps. First, I used the “missingno” library to identify how many values were missing in each column and to visualize the distribution of missing data. This helped me to understand which columns had a high percentage of missing values and which ones had relatively few missing values.



Next, I removed the "Age" and "Distance from Home" rows of data because their values could not be replaced without affecting the purity of the dataset. Removing these rows helped to ensure that the data was as accurate and representative as possible.

Finally, I used various techniques to fill in or impute missing values in the remaining columns, such as using the mean or median of the column to fill in missing values for numerical columns or using the mode to fill in missing values for categorical columns.

The missing values in the "Business Travel" and “Marital Status” columns were replaced by the most frequent value in their column using the mode method.

To replace the missing values in the "DailyRate" column I used the mean values of the daily rate depending on de "Department" column.

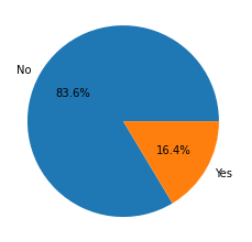
This helped to ensure that the data was complete and ready for analysis.

**DATA CLEANING**

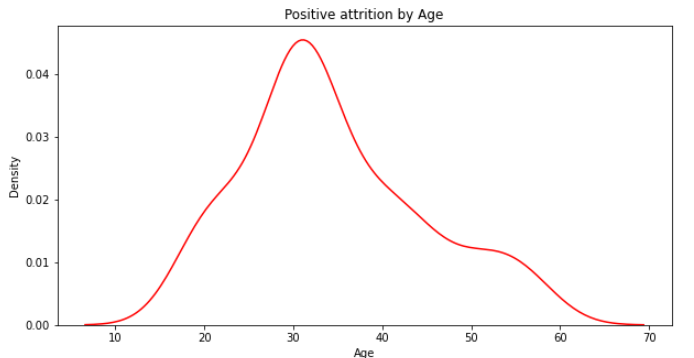
I identified and removed any columns irrelevant to the analysis or duplicate values for all rows. This included the "EmployeeCount" and "StandardHours" columns, which had only one value each for all their rows, and other columns such as employee ID numbers, which were not relevant to the analysis.

**DATA STORIES AND VISUALISATIONS**

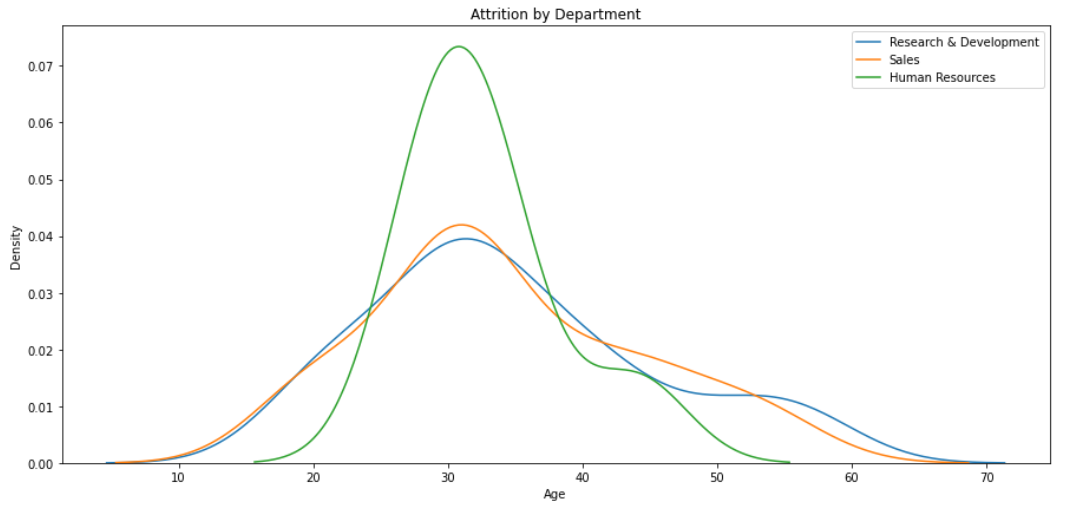
To start analyzing the cleaned Employee Attrition dataset, I used various visualization methods to gain insights from the data. I created a pie plot to view the difference in attrition rates between employees.

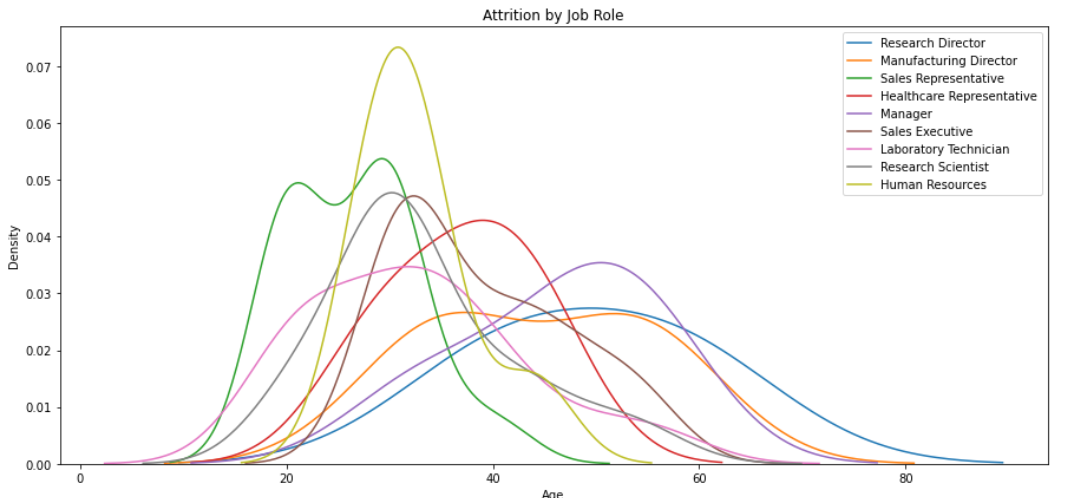


I also plotted the attrition rates for employees with "yes" values based on their age, department, and job role.



*The above plot shows that employees have an increasing rate of leaving their job until around 30*.

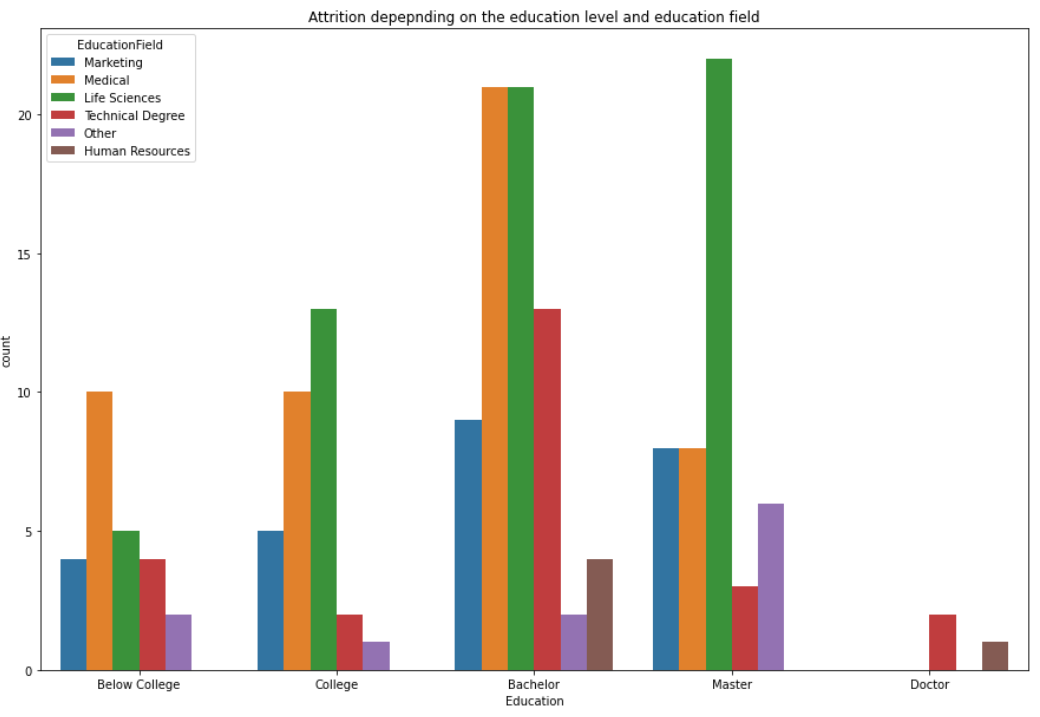




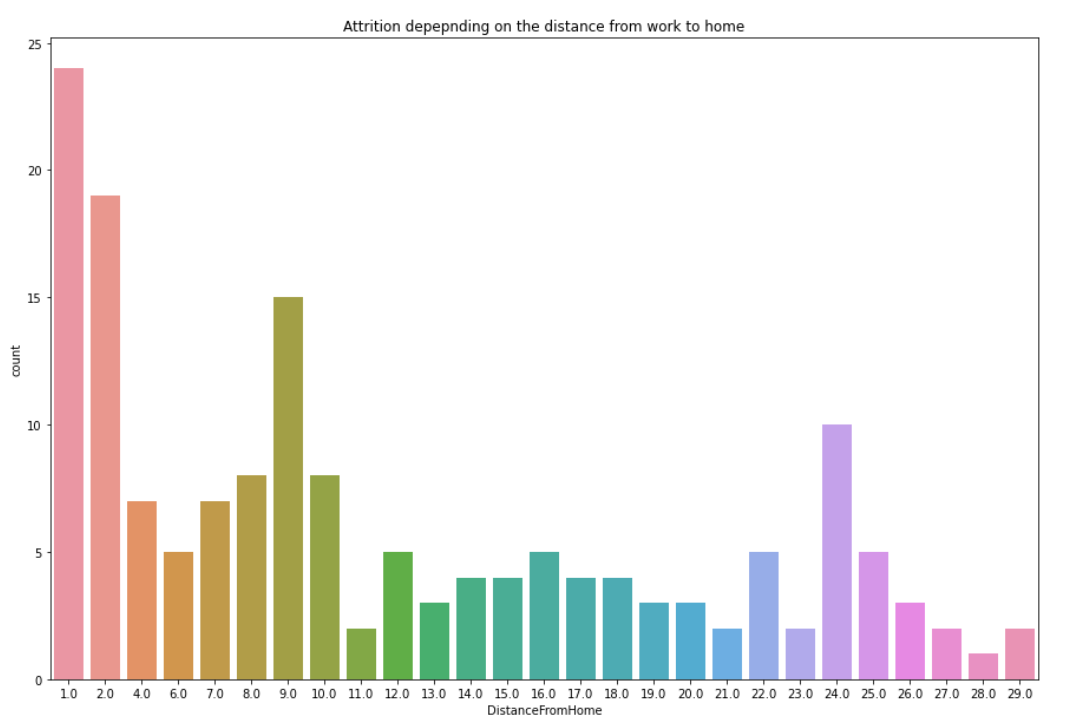
*The plots above show the Attrition per age depending on the department and the job role.*

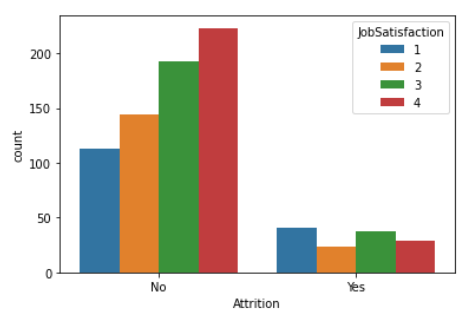
* *By departments and job role the highest attrition is achieved by employees in Human Resources*
* *By job role higher attrition is achieved by younger employees that work as Sales Representatives*
* *Managers and Healthcare Representatives’ jobs achieve a high attrition rate between 40 to 55.*

Additionally, I made a plot showing the attrition rates depending on the employee's education level and field of study.

 *The plot above shows the attrition rate depending on the education level in each education field*

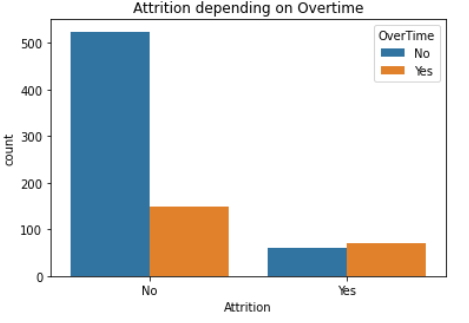
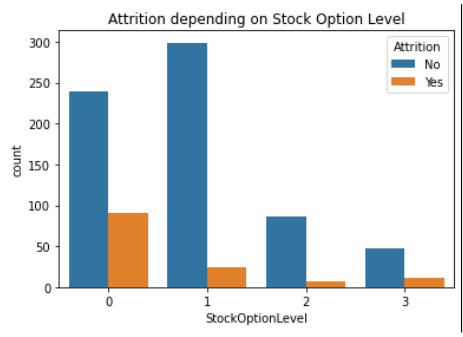
* *Highest attrition rates are achieved in the Life Science field and Medical Field.*
* *There is an ascending trend in the Life Science field that shows that the higher the education the higher the attrition rate*
* *Overall employees with a Bachelor’s or Master’s education have a higher attrition rate than the other levels of education*

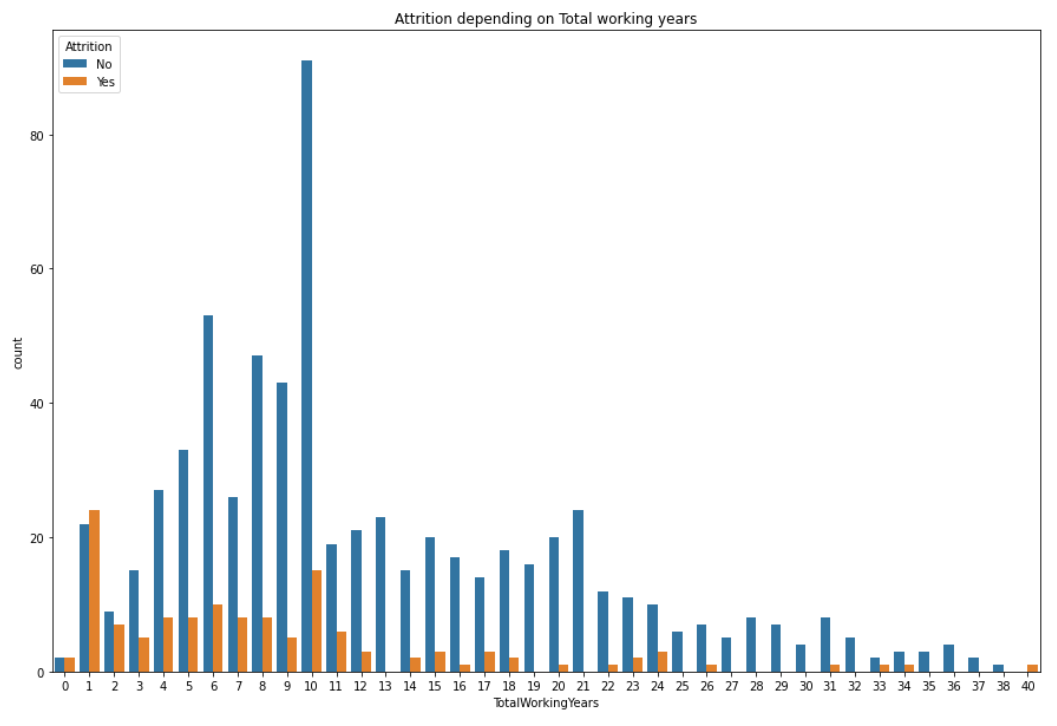




*In this plot is shown the employee’s attrition depends on their job satisfaction*

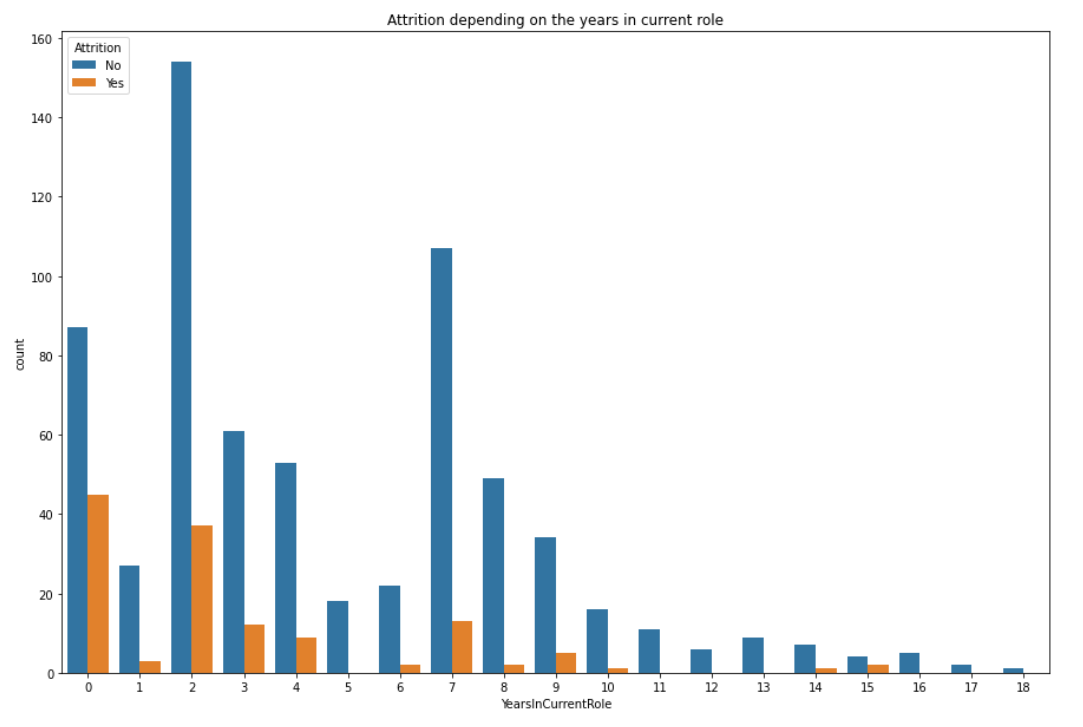
* *The higher the job satisfaction the lower of employees leaving*

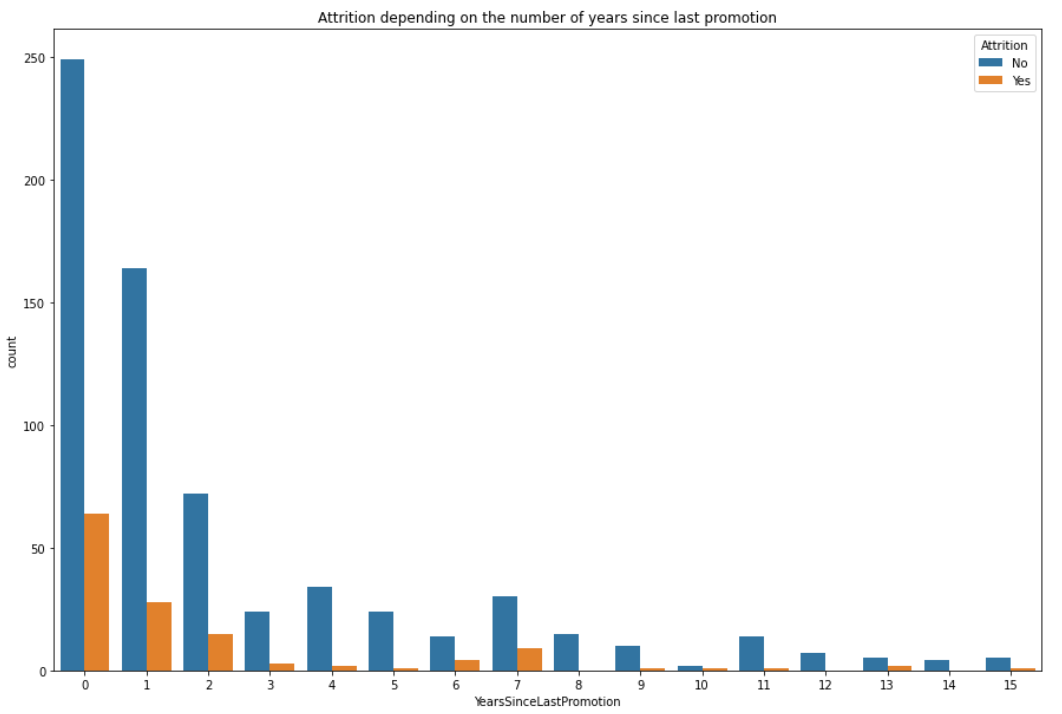
 



*The above plot shows the attrition depending on the total working years*

* *Employees have a higher rate of leaving their job in the first 10 years.*





These visualizations allowed me to understand the trends and patterns in the data and identify factors contributing to employee turnover. By using these visualizations, I was able to get a better understanding of the data and draw meaningful conclusions from it.

**THIS REPORT WAS WRITTEN BY : George Cireasa**

