Analysis Report

Insights / Trends from Employee Attrition data

Data Cleaning and Preparation

Some Categorical columns have been converted to numerical we will convert them to categorical features for EDA

Data Wrangling by Inspecting numerical and categorical features and removing unnecessary

EDA Overview

- For categorical columns we have found the attrition in respect to the element inside each of these categorical features to understand the attrition rate differences
- For Numerical columns we have plotted histograms and boxplots to understand the variance of these columns in terms of attrition

By doing so we have been able to extract a lot of insights in less amount of time from the data

Categorical Features and Insights

- **Business Travel : 71% of employees travel rarely the highest attrition rates are by employees who travel frequently 25% followed by rarely travelling employees**
- **Department : Highest attrition rates are from Sales department(20.6%) followed by HR .Substantial employees are from R&D**
- **Education :Highest attrition rates are from employees who are below bachelors and Bachelors and college students**
- **Education Field : HR, Marketing and Technical degree employees has higher attrition rate of 26%,22% and 24% each**
- **Employee Satisfaction : Employees with Low Satisfaction has highest attrition rate**
- **Gender :Male Employees has slightly higher attrition rate but not substantial**
- **Job Involvement : Employees with lowest job involvement has highest attrition rate**
- **Job Level : Entry Level employees has highest attrition rate**
- **Job role**
- **Job satisfaction : Employees with low satisfaction has highest attrition rate**
- **Marital Status : Single employees has highest attrition rate**
- **Over time : Employees who work over time has substantially more attrition rate**
- **Performance Rating : No significant difference seen**
- **Relationship Satisfaction :Employees with lower relationship satisfaction show higher attrition rate**
- **Work Life Balance : Employees voted for bad work life balance has higher attrition rate**

These are all the insights from categorical columns
Performance Rating will be removed in further analysis

Numerical Features and Insights

"When we plot numerical features as histograms and boxplots for both employees who have left the company and those who are still part of it, we can observe slight differences in variance. Only a few features show slightly significant influences, so I am mentioning them. You can find the plots for all the numerical features in the Jupyter Notebook file for this analysis."

These insights have been extracted by plotting swarm plots and have not been included in the jupyter notebook because the runtime for this plot is very high

- Age : Slight difference in variances but not significant
- Distance from Home :Slight difference in variances but not significant
- Monthly Income :Employees with higher salary tend to not leave the company
- No of Companies worked: Nothing significant just a slight variation among distribution
- Percent Hike :Employees with lower percent hike tend to leave the company more
- Total Working Years: Employees with higher total working years tend to not leave the company and vice versa
- Years at company :Employees that are at the company for less number of years leave more
- Years in current role: Employees that have been in the current role for higher number of years have less attrition
- Years with current manager :Nothing Significant just variance due to random factors
- Years Since Promotion : Nothing Significant just variance due to random factors

There is a slight variance in these numerical but nothing very significant the could influence the predicting to a larger extend but to some degree yes these variables does affect but certainly there will be multicolinearity between these variables so when we are building the model we will only use some features which could better predict attrition