**DATA ANALYTICS – IBM PROJECT**

Why do workers resign from the positions they hold? Can employee attrition be anticipated? Can companies take steps to prevent high employee turn-over?

Employee attrition is the reduction of staff by voluntary or involuntary reasons. These can be through natural means like retirement, or it can be through resignation, termination of contract, or when a company decides to make a position redundant. Excessive involuntary attrition can be detrimental to the overall business goals, and can lead to understaffing, increased job burden on remaining employees, and harmful consequences to company reputation.

You’ve just been hired by IBM to work in their Data Analytics sector. On your second day, you are given a dataset from HR that includes several records of employees that were with the company in the prior calendar year. Since the development of the dataset, some of those employees have since resigned from their positions. The company was surprised with the recent level of turnover and have tasked you with digging into the data to determine any insights about what leads to employee turnover, and alternatively, what aspects of the jobs can be protected to reduce employee turnover.

Using the ‘**IBM Employee Attrition**’ dataset – your role is to comprehensively explore the data and generate meaningful insights into what the data is presenting. You should conclude your assignment with 3 recommendations you will make to the HR department regarding employee attrition.

At the least, you should complete the following tasks:

* **Data prep and cleaning** – *prior to exploring the data, you should ensure that the dataset is clean, all variables are in their proper format, and the data is free of missing, abnormal, or outlier values.*
* **Exploratory data analysis** – *you should explore every aspect of the data to determine the features of the data, the range of values for all variables of interest, and the relationships between variables. Exploration should be conducted outside of predictive modeling – in other words, you are not only interested in how the dependent and independent variables interact – you should also spend some time exploring variations in the data among variables (i.e. how does salary differ between genders? Does time spent in the company differ by job role?)*
* **Modeling** – *after thoroughly exploring the data, create a logistic regression model that predicts employee attrition based on the other variables available to you. Clearly interpret your model findings. Re-run/tweak your model if you think you can find a better fit.*
* **Visualization –** *Create at least 3 visualizations that highlight some of the most interesting findings in your dataset. These can be findings from the regression model or any interesting relationships between variables.*

Good luck!

**IBM PROJECT – Dataset Descriptions**

|  |  |
| --- | --- |
| **Age** | *Age of employee in years* |
| **Attrition** | *Employee attrition status; yes = employee left position, no = employee did not leave position* |
| **BusinessTravel** | *Frequency of travel necessary for job; Travel Rarely = business travel not frequently required, Travel Frequently = business travel frequently required, Other = business travel required at other increments* |
| **Department** | *Department within company where employee works; research and development, sales, or other* |
| **DistanceFromHome** | *Distance job is from employees home (in Miles)* |
| **Education** | *Employee education level; 1 = below college, 2 = college, 3 = bachelor, 4 = masters degree, 5 = doctoral degree* |
| **EducationField** | *Education background of employee; life sciences, medical, marketing, technical degree, other* |
| **EmployeeNumber** | *Employee ID* |
| **EnvironmentSatisfaction** | *Employee satisfaction with job environment, 1 = low, 2 = medium, 3 = high, 4 = very high* |
| **Gender** | *Gender of employee* |
| **JobInvolvement** | *Level of involvement with job, 1 =low, 2 = medium, 3 = high, 4 = very high* |
| **JobLevel** | *Job level within company; 1 = level 1, 2 = level 2, 3 = level 3, 4 = level 4, 5 = level 5* |
| **JobRole** | *Job title; sales executive, research scientist, laboratory technician, manufacturing director, other* |
| **JobSatisfaction** | *Employee satisfaction with overall job, 1 =low, 2 = medium, 3 = high, 4 = very high* |
| **MaritalStatus** | *Marital status of employee; married, single, divorced* |
| **MonthlyIncome** | *Monthly income (in USD)* |
| **NumCompaniesWorked** | *Number of companies employee has previously worked with* |
| **Over18** | *Employee is over 18, true or false* |
| **OverTime** | *Employee is eligible to work overtime, true or false* |
| **PercentSalaryHike** | *Percent increase in salary over a two year period* |
| **PerformanceRating** | *Employee performance at current position; 1 = low, 2 = good, 3 = excellent, 4 = outstanding* |
| **RelationshipSatisfaction** | *Employee satisfaction with job-related relationships, 1 =low, 2 = medium, 3 = high, 4 = very high* |
| **StockOptionLevel** | *How many company stocks the employee owns for the company* |
| **TotalWorkingYears** | *Total years employee has worked in lifetime* |
| **TrainingTimesLastYear** | *Number of training/business development opportunities within past year* |
| **WorkLifeBalance** | *Quality of work/life balance; 1 = bad, 2 = good, 3 = better, 4 = best* |
| **YearsAtCompany** | *Total years employee has worked with company* |
| **YearsInCurrentRole** | *Total years employee has worked in their current role* |
| **YearsSinceLastPromotion** | *Number of years since the employees last promotion* |
| **YearsWithCurrManager** | *Number of years employee has been with their current manager* |