FINAL PROJECT WORK

[Link to Kaggle dataset](https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/)

The dataset I chose came from the large technology corporation IBM, and its data consists of information regarding employee attrition. There are thirty-five attributes, 26 of which are integers.

QUESTIONS

1. Which department suffered the most losses in employees?
2. Did People with more years at the company have more security? And are they happy with their position?
3. Do the performance ratings directly reflect who gets fired?
4. What effect does a good work life balance have on employee effort?
5. Do employees who work more hours have higher performance ratings? Is this reflected in their pay / promotions / bigger bonus?

My potential Client is IBM! Now, I know, IBM doesn’t need me doing this for them, but I am curious because my mom works for a giant corporation that operates like them, so I’ve always been curious what it’s like to have bosses whose bosses have bosses. This analysis will hopefully be able to shed some light on whether or not the attrition of these employees is justified or not.

data type: int64 Column 1: Age

data type: object Column 2: Attrition

data type: object Column 3: BusinessTravel

data type: int64 Column 4: DailyRate

data type: object Column 5: Department

data type: int64 Column 6: DistanceFromHome

data type: int64 Column 7: Education

data type: object Column 8: EducationField

data type: int64 Column 9: EmployeeCount

data type: int64 Column 10: EmployeeNumber

data type: int64 Column 11: EnvironmentSatisfaction

data type: object Column 12: Gender

data type: int64 Column 13: HourlyRate

data type: int64 Column 14: JobInvolvement

data type: int64 Column 15: JobLevel

data type: object Column 16: JobRole

data type: int64 Column 17: JobSatisfaction

data type: object Column 18: MaritalStatus

data type: int64 Column 19: MonthlyIncome

data type: int64 Column 20: MonthlyRate

data type: int64 Column 21: NumCompaniesWorked

data type: object Column 22: Over18

data type: object Column 23: OverTime

data type: int64 Column 24: PercentSalaryHike

data type: int64 Column 25: PerformanceRating

data type: int64 Column 26: RelationshipSatisfaction

data type: int64 Column 27: StandardHours

data type: int64 Column 28: StockOptionLevel

data type: int64 Column 29: TotalWorkingYears

data type: int64 Column 30: TrainingTimesLastYear

data type: int64 Column 31: WorkLifeBalance

data type: int64 Column 32: YearsAtCompany

data type: int64 Column 33: YearsInCurrentRole

data type: int64 Column 34: YearsSinceLastPromotion

data type: int64 Column 35: YearsWithCurrManager

Summary:

My final project involved digging into IBM’s employee data using Python libraries. I started by checking out the dataset and then explored some important questions about why employees leave. I looked at which departments were most affected, if people who had been with us longer felt more secure, and how performance ratings relate to attrition. I also checked if a good work-life balance influenced job involvement. The results were presented using mostly bar graphs, as the data ended up being more categorical than I initially thought. The project not only showed which departments had the most turnover but also revealed interesting patterns in how long people stick around and factors influencing their decisions.

ThIs project was a very insightful one for me, as it’s good to know what companies look at when they decide who to let go. The results seemed to show me that everything was relatively proportional when it came to Attrition, but when we look at who gets a raise, we see that the raise value almost exactly doubles with a higher performace rating. We can infer from this that your recorded performace matters deeply in how much you’ll end up getting paid, and that people get fired evenly around the board, so you should definetly be trying your hardest at work to stand out.

To do this analysis, I used libraries such as matplotlib, seaborn, and pandas. The analysis included a lot of value\_counts and comparisons between them, but that is the information the data provided best results for me. All In all, I enjoyed this project and all the insights gained from it.