PROJECT REPORT ON –

HR Analytics on Employee Attrition and calculating the overtime hours worked in different Department.

Submitted to: -

Dr. Divyakshatriya mam

Submitted By: -

Samith Kumar Mahapatra

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. Samith Kumar Mahapatra, student of PGBIA course for Business Anlaytics, Bangalore has successfully completed his Project in HR attriton calculation with dept and gender and salary and overtime hours workers have worked.

Acknowledgment

This project would not have been possible without the invaluable support and guidance of numerous individuals. I would like to express my sincerest gratitude to Dr. Divakshyatriya Mam for her unwavering support, insightful guidance, and continuous encouragement throughout this project. Her invaluable expertise and constructive feedback were instrumental in my successful completion of this endeavor. Her contributions were vital to the successful completion of this project.

Finally, I would like to acknowledge the hard work and dedication I invested in this project.

Sincerely,

Samith Kumar Mahapatra

Place: Bangalore

Table of Contents: -

CHAPTER 1

INTRODUCTION

Employee attrition is a critical concern for organizations as it impacts productivity, morale, and costs associated with hiring and training new employees. This project aims to analyse employee data to understand the factors contributing to attrition and provide insights that can help in retaining talent.

The analysis will focus on identifying patterns and correlations between various employee attributes and their likelihood of leaving the company.

Attrition, in Human Resource terminology, refers to the phenomenon of the employees leaving the company. Attrition in a company is usually measured with a metric called attrition rate, which simply measures the number of employees moving out of the company (voluntary resigning or lay off by the company). Attrition Rate is also referred to as churn rate or turnover.

The rate of shrinkage in size or number. In the best of worlds, employees would love their jobs, like their co-workers, work hard for their employers, get paid well for their work, have ample chances for advancement, and flexible schedules so they could attend to personal or family needs when necessary. And never leave. But then there's the real world. And in the real world, employees do leave, either because they want more money, hate the working conditions, hate their co-workers, want a change, or because their spouse gets a dream job in another state

Human Resource Management

For any organization to function effectively, it must have resources of men (Human Resource), money, materials and machinery. The resources by themselves cannot fulfill the objectives of an organization, they need to be collected, co-ordinated and utilized through human resources. And, the effective management of human resources is also vital. Hence, Human Resource Management (HRM) has emerged as a major function in organizations.

Human Resource Management is the organizational function that deals with issues related to people, such as compensation, benefits, hiring, motivation, communication, development, safety, wellness, training, performance management, organization, administration, and training.

Feature's of Human resource management

Human Resource Management is the process of recruitment and selecting employee, providing orientation and induction, training and development, assessment of employee performance and appraisal, providing compensation and benefits, motivating, maintaining proper relations with employees and trade unions, maintaining employee safety, welfare and health measures in compliance with labor laws of the land.

Before we define HRM, it is pertinent to first define the term Human Resources. In common parlance, human resources means people. For example, different management experts have defined human resources differently. However, Michael J. Jucius has defined human resources as a whole consisting of related, interdependent and interacting physiological, psychological, sociological and ethical components.

Chapter 2: Objectives

The objectives of this HR analytics project are to:

- Identify key factors that influence employee attrition.
- Predict which employees are most likely to leave the company.
- Using power BI or Tableau to create dashboards by using the data created for employee attrition.

SCOPE OF THE STUDY: -

- To determine the effect of attrition on the business.
- Determination of solutions to avoid or to control attrition.
- To understand the extent of job satisfaction among the employees.
- To suggest proper measures.
- Helping them to reduce the employee Attrition.
- This study helps the company to understand more about the attrition rate in the company.
- The study educates the causes of attrition for employees in an organization.
- The study also helps to find the drawbacks of the current retention strategies.

NEED OF THE STUDY

The study was mainly undertaken to identify the level of employee's attitude, the dissatisfaction factors they face in the organization, and for what reason they prefer to change their job. Once the levels of Employee's attitude are identified, it would be possible for the management to take necessary action to reduce [1] the attrition level. Since they are considered the backbone of the Company, their progression will lead to the success of the Company in the long run.

Chapter 3: Data Collection and Description

The dataset contains the following columns:

- Employee: Unique identifier for each employee.
- Age: Age of the employee.
- Gender: Gender of the employee.
- Department: Department where the employee works.
- Job Title: The current job title of the employee.
- **Years at Company:** Number of years the employee has worked at the company.
- **Satisfaction Level:** Employee's self-reported job satisfaction level.
- **Average Monthly Hours:** Average number of hours the employee works per month.
- **Promotion Last 5Years:** Whether the employee was promoted in the last 5 years (1 for Yes, 0 for No).
- **Salary: The salary level (e.g., low, medium, high).**
- **Attrition:** Whether the employee has left the company (1 for Yes, 0 for No).

Chapter 4: Problem Statement

The problem statement in the document focuses on the challenge of employee attrition in organizations. Employee attrition can significantly impact productivity, morale, and costs associated with hiring and training new employees. The project aims to analyse employee data to understand the factors contributing to attrition and provide insights that can help in retaining talent. The analysis will focus on identifying patterns and correlations between various employee attributes and their likelihood of leaving the company. This approach will enable HR managers to develop strategies for improving employee retention.

How to Calculate Employee Attrition Rate

The attrition rate formula is:
Attrition rate = (Number of employee departures) / (Average number of employees) x 100

To calculate the employee attrition rate:

1. Calculate the average number of employees. Add the number of employees at the beginning and end of the period, then divide by 2.
2. Divide the number of employees who left the company by the average number of employees, then multiply by 100.

Is a High Attrition Rate Good?

High attrition rates generally indicate that employees are leaving faster than replacements can be hired. This can lead to:

- Increased Costs and Efforts: Recruitment, onboarding, and training costs increase.
- Decreased Productivity: Vacant positions lead to productivity losses.
- Impact on Employee Morale: Frequent staff changes can affect collaboration and morale.

However, attrition can be beneficial when:

- Reducing Labor Costs: Vacant positions save salary expenses.
- Restructuring: Attrition can support workforce realignment during organizational changes.
- Bringing in New Talent: Allows for the recruitment of new talent with fresh skills.

How to Conduct an Employee Attrition Analysis

1. Gather Data: Collect headcount, new hires, and departure data. Apply the attrition rate formula.
2. Look for Patterns: Identify departments or roles with high attrition rates. Compare with industry benchmarks.

3. Identify Risk Factors: Low employee engagement, absenteeism, high turnover in specific departments, and difficulty in filling vacancies are key indicators of attrition risks.

HR Strategies for Reducing Attrition Rate

- Focus on Employee Wellbeing: Employee satisfaction and work-life balance are key to retention.
- Boost Learning and Development: Offer training programs to promote growth and development.
- Implement Cross-Training: Equip employees to take on multiple roles, increasing flexibility.
- Succession Planning: Anticipate retirements and develop strategies for smooth transitions.
- Offer Competitive Compensation: Ensure salaries and benefits are competitive to retain top talent.
- Work-life balance: A healthy work-life balance lets employees enjoy their professional and personal lives. They are happier and more satisfied at work and are unlikely to apply for another job.

Chapter 5: Methodology

4.1Data Cleaning:

- **Missing Values:** Missing values were identified and handled using appropriate methods such as mean/mode imputation for numerical data and forward fill for categorical data.
- Clearly state the source of your data (e.g., internal HR database, Kaggle dataset, etc.).
- **Initial Inspection:** Describing the initial steps taken to understand the data, such as:

1. Examining the data's dimensions (number of rows and columns).
2. Identifying the data types of each variable.
3. Calculating descriptive statistics for numerical variables (mean, median, standard deviation, etc.).
4. Examining the distribution of categorical variables (frequency counts).
5. Checking for potential data quality issues.
6. Outlier Treatment: Outliers in variables like Average_Monthly_Hours and Satisfaction_Level were identified using box plots and treated by capping or removing extreme values.
7. Standardization and Encoding: Categorical variables (e.g., Gender, Department, and Salary) were encoded into numerical formats using label encoding and one-hot encoding techniques.
8. Data Source: Hypothetical HR employee dataset.
9. Visualization Tools: Excel, Power BI.
10. Statistical Methods: Correlation analysis, Chi-square tests.
11. Machine Learning Models: Logistic Regression, Decision Trees, Random Forests.

4.2 Data Cleaning & Preparation: -

- **Description:** Clean the data by handling missing values, correcting any inconsistencies, and transforming categorical variables (e.g., Gender, Salary) into numerical ones for analysis.
- **Tools:**
  - **Excel:** Basic data cleaning, pivot tables to explore initial patterns.
  - **Python:** Use pandas for advanced data cleaning and feature engineering.
  - **Key Focus:** Ensure the dataset is ready for analysis and visualization.

Chapter 6

Exploratory Data Analysis (EDA):-

EDA focuses on uncovering patterns, correlations, and insights from the dataset. In your HR Analytics project, you conducted univariate analysis, examining distributions of key variables like Satisfaction Level, Age, and Average Monthly Hours using histograms and box plots. This revealed that departments with lower satisfaction levels had higher attrition rates.

Additionally, bivariate analysis explored relationships between variables such as Attrition vs. Salary and Attrition vs. Satisfaction Level, using scatter plots and correlation matrices. Preliminary insights showed higher attrition rates among employees with lower salaries or those without promotions in the last five years. High or low working hours also indicated potential burnout or underutilization, contributing to attrition.

Finally, I identified correlations between factors like the number of years at the company, satisfaction levels, and attrition rates using a heatmap, guiding you toward selecting important features for your model.

Univariate Analysis:

The histogram displays the distribution of satisfaction levels among employees. Most people are either highly satisfied or highly dissatisfied, with fewer people falling in the middle range. This bimodal distribution indicates two distinct groups within the data

Employee ID	**Age**	**Gender**	**Dept.**	**Job title**	**Years At company**	**Satisfaction level**	**Average monthly hours**	**Promotion in Last 5 Years**	**Salary**	**Attrition**	**Attrition Count by Dept.**
0	27	1(male)	3	4	9	0.586251	153	0	60132.0	0	80
1	53	0(FEMALE	4	2	10	0.261161	221	1	79947.0	0	103
2	58	0(FEMALE)	3	1	8	0.304382	184	0	46958.0	1	80
3	42	0(FEMALE	0	4	1	0.480779	242	0	40662.0	0	109

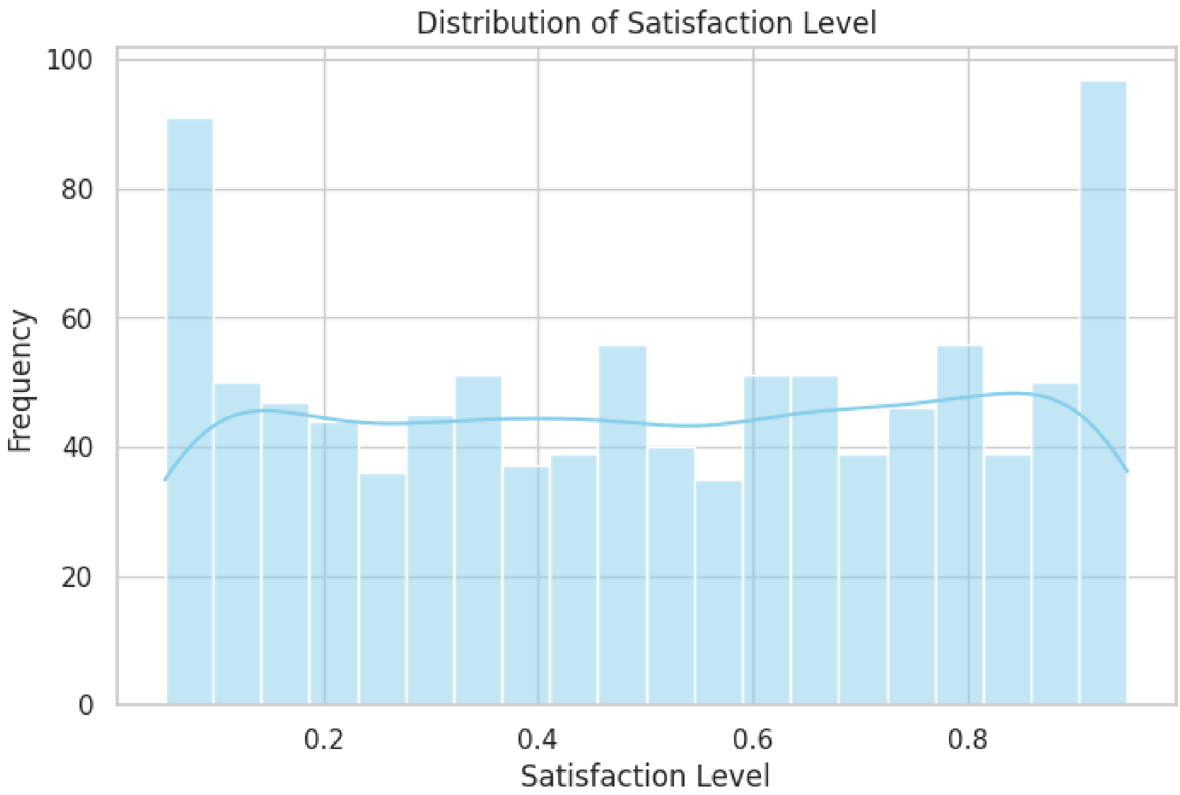
4	44	0(FEMALE)	4	2	10	0.636244	229	1	74307.0	0	103

PYTHON PROGRAM USED TO CALCULATE TO CREATE UNIVARIATE ANALYSIS FOR THE TABLE:-

Univariate Analysis – Examine Single Variables

#Plot 1: Distribution of Satisfaction Level

```
plt.figure(figsize=(8, 5))

sns.histplot(df['Satisfaction Level'], bins=20, kde=True, color='skyblue')

plt.title('Distribution of Satisfaction Level')

plt.xlabel('Satisfaction Level')

plt.ylabel('Frequency')

plt.show()
```



The histogram displays the distribution of satisfaction levels among a group of individuals. It shows that most people are either highly satisfied or highly dissatisfied, with fewer people falling in the middle range. This suggests a bimodal distribution, indicating two distinct groups within the data.

In relation to attrition, this pattern might suggest that:

1. **High Satisfaction:** Individuals with high satisfaction levels are less likely to leave the organization. They are likely to be more engaged, motivated, and committed to their work.

2. **Low Satisfaction:** Individuals with low satisfaction levels are more likely to consider leaving the organization. They may be experiencing job dissatisfaction, lack of recognition, or poor work-life balance.

To further analyze the relationship between satisfaction and attrition, you could:

1. Gender and Attrition:

From the table, we can observe:

- Female Attrition: There is 1 female employee who has attritted.
- Male Attrition: There are no male employees who have attritted.

Calculation:-

1. Attrition Rate by Gender:

- Total Female Employees: 3
- Female Employees Who Attrited: 1
- Attrition Rate for Females: (1 / 3) * 100 = 33.33%
- Total Male Employees: 2
- Male Employees Who Attrited: 0
- Attrition Rate for Males: (0 / 2) * 100 = 0%

2. Job Title and Attrition:

- Job Title 1: 1 employee with this job title has attritted.
- Job Title 2: 0 employees with this job title have attritted.
- Job Title 4: 0 employees with this job title have attritted.

Calculation: -

- Total Employees with Job Title 1: 1
- Employees with Job Title 1 Who Attrited: 1
- Attrition Rate for Job Title 1: (1 / 1) * 100 = 100%
- Total Employees with Job Title 2: 2
- Employees with Job Title 2 Who Attrited: 0
- Attrition Rate for Job Title 2: (0 / 2) * 100 = 0%
- Total Employees with Job Title 4: 2
- Employees with Job Title 4 Who Attrited: 0
- Attrition Rate for Job Title 4: (0 / 2) * 100 = 0%

3. Promotion and Attrition:

- No Promotion: 1 employee who did not receive a promotion in the last 5 years has attritted.
- Promotion: 0 employees who received a promotion in the last 5 years have attritted.

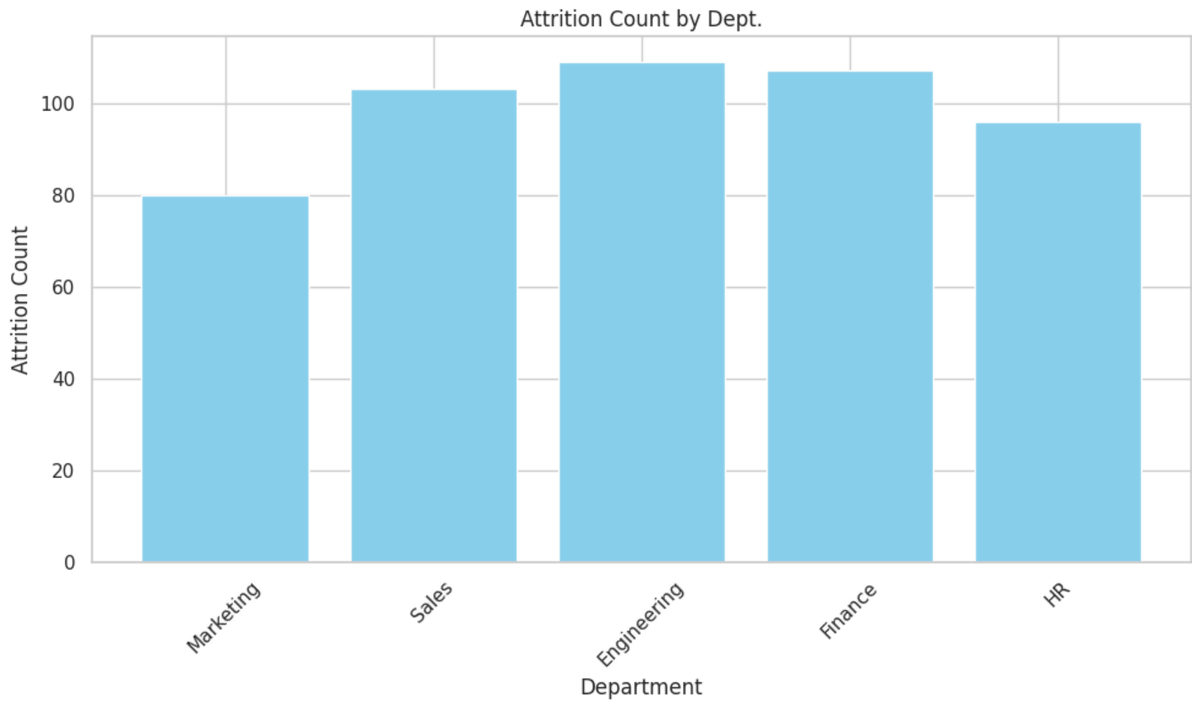
Calculation:-

- Total Employees with No Promotion: 3
- Employees with No Promotion Who Attrited: 1
- Attrition Rate for No Promotion: (1 / 3) * 100 = 33.33%
- Total Employees with Promotion: 2
- Employees with Promotion Who Attrited: 0
- Attrition Rate for Promotion: (0 / 2) * 100 = 0%

Observations:

- The attrition rate is significantly higher for female employees compared to male employees.
- Employees with job title 1 have the highest attrition rate (100%), while those with job titles 2 and 4 have no attrition.
- Employees who did not receive a promotion in the last 5 years have a higher attrition rate compared to those who did.

By understanding the distribution of satisfaction levels and its relationship with attrition, organizations can take proactive steps to improve employee satisfaction and reduce turnover. This might involve implementing targeted interventions to address the specific needs of low-satisfaction groups, recognizing and rewarding high-performing employees, and fostering a positive work environment.



The bar chart titled "Attrition Count by Dept." illustrates the number of employees who have left the company from different departments.

The python program which I used to calculate the attrition is as follows:-

#Plot 2: Attrition Count by Department

import pandas as pd

import matplotlib.pyplot as plt

 # Specify the actual path to the Excel file

file_path = r'C:\path\to\your\New Employee Attrition.xlsx'

 # Group by Department and get unique attrition count

attrition_counts = data[['Department', 'Attrition Count by Dept.']].drop_duplicates()

 # Plot a bar graph

plt.figure(figsize=(10, 6))

plt.bar(attrition_counts['Department'], attrition_counts['Attrition Count by Dept.'], color='skyblue')

plt.xlabel('Department')

plt.ylabel('Attrition Count')

plt.title('Attrition Count by Dept.')

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()

Key Observations:

- **Highest Attrition:** The departments with the highest attrition rates are **Engineering** and **Finance**.
- **Lowest Attrition: Marketing** and **HR** departments have the lowest attrition rates.
- **Overall Trend:** The chart generally shows a higher rate of attrition in some departments compared to others.

Possible Interpretations and Insights:

- **Departmental Factors:** Certain departments might have specific factors contributing to higher attrition, such as heavy workload, stressful work environments, or lack of opportunities for growth.
- **Organizational Culture:** The overall organizational culture and employee engagement practices may also play a role in attrition rates.
- **Compensation and Benefits:** Differences in compensation, benefits, or work-life balance policies across departments could influence attrition.
- **Leadership:** The leadership style and management practices within each department may impact employee satisfaction and retention.

This could involve conducting surveys, interviews, or focus groups to gather insights from employees who have left the organization. Additionally, analyzing other relevant factors like job satisfaction, work-life balance, and career growth opportunities can provide valuable insights.

By identifying the root causes of attrition, organizations can take targeted actions to improve employee retention and reduce turnover costs.

Analysing the "Attrition Count by Dept." Graph

Key Observation:

The bar graph clearly indicates that **Engineering** and **Finance** departments have the highest attrition rates, while **Marketing** and **HR** have the lowest.

Possible Interpretations and Insights:

1. **Departmental Factors:**
    - **Workload and Stress:** Engineering and Finance departments often involve complex tasks and tight deadlines, leading to higher stress levels and burnout.
    - **Job Satisfaction:** Factors like job satisfaction, work-life balance, and career growth opportunities might vary across departments, influencing attrition rates.
    - **Organizational Culture:** The overall organizational culture, including leadership style, communication, and employee recognition, can impact employee satisfaction and retention.
2. **External Factors:**
    - **Industry Trends:** Certain industries, like technology, may experience higher attrition due to rapid advancements and competition for talent.
    - **Economic Conditions:** Economic downturns or upturns can impact job security and career opportunities, affecting attrition rates.

Potential Action Points:

1. **Targeted Interventions:**
    - Conduct surveys and interviews to identify the specific reasons for attrition in high-attrition departments.
    - Implement targeted interventions, such as increased training, mentorship programs, or flexible work arrangements, to address the root causes of attrition.
2. **Organizational Culture:**
    - Foster a positive and inclusive work environment that values employee well-being and recognizes achievements.
    - Promote open communication and provide opportunities for feedback and suggestions.
3. **Leadership Development:**
    - Invest in leadership development programs to equip managers with the skills to retain top talent.
    - Encourage effective communication and coaching to improve employee morale and motivation.
4. **Compensation and Benefits:**
    - Review compensation and benefits packages to ensure they are competitive and attractive.
    - Consider offering additional perks, such as flexible work arrangements or wellness programs.

By understanding the underlying factors driving attrition in different departments, organizations can take proactive steps to improve employee retention and reduce turnover costs.

Diving Deeper: Analysing the Root Causes of Attrition

To gain a more comprehensive understanding of the factors contributing to attrition in different departments, we can explore the following:-

1. Job Satisfaction:

- **Workload and Stress:** High workloads and excessive stress can lead to burnout and decreased job satisfaction.
- **Career Growth Opportunities:** Limited opportunities for advancement and skill development can demotivate employees.
- **Recognition and Rewards:** Inadequate recognition and reward systems can negatively impact employee morale.

2. Organizational Culture:

- **Leadership Style:** Ineffective leadership, lack of support, and micromanagement can create a toxic work environment.
- **Communication:** Poor communication and lack of transparency can lead to misunderstandings and frustration.
- **Work-Life Balance:** Long working hours, inflexible work arrangements, and work-life imbalance can contribute to attrition.

3. Compensation and Benefits:

- **Competitive Compensation:** Inadequate compensation and benefits packages can make employees feel undervalued.

- **Fairness and Equity:** Perceived inequities in compensation and promotions can lead to dissatisfaction and attrition.

4. External Factors:

- **Economic Conditions:** Economic downturns or industry-specific challenges can impact job security and career opportunities.
- **Competition for Talent:** High demand for skilled professionals can lead to increased attrition as employees seek better opportunities.

To address these issues, organizations can implement the following strategies:

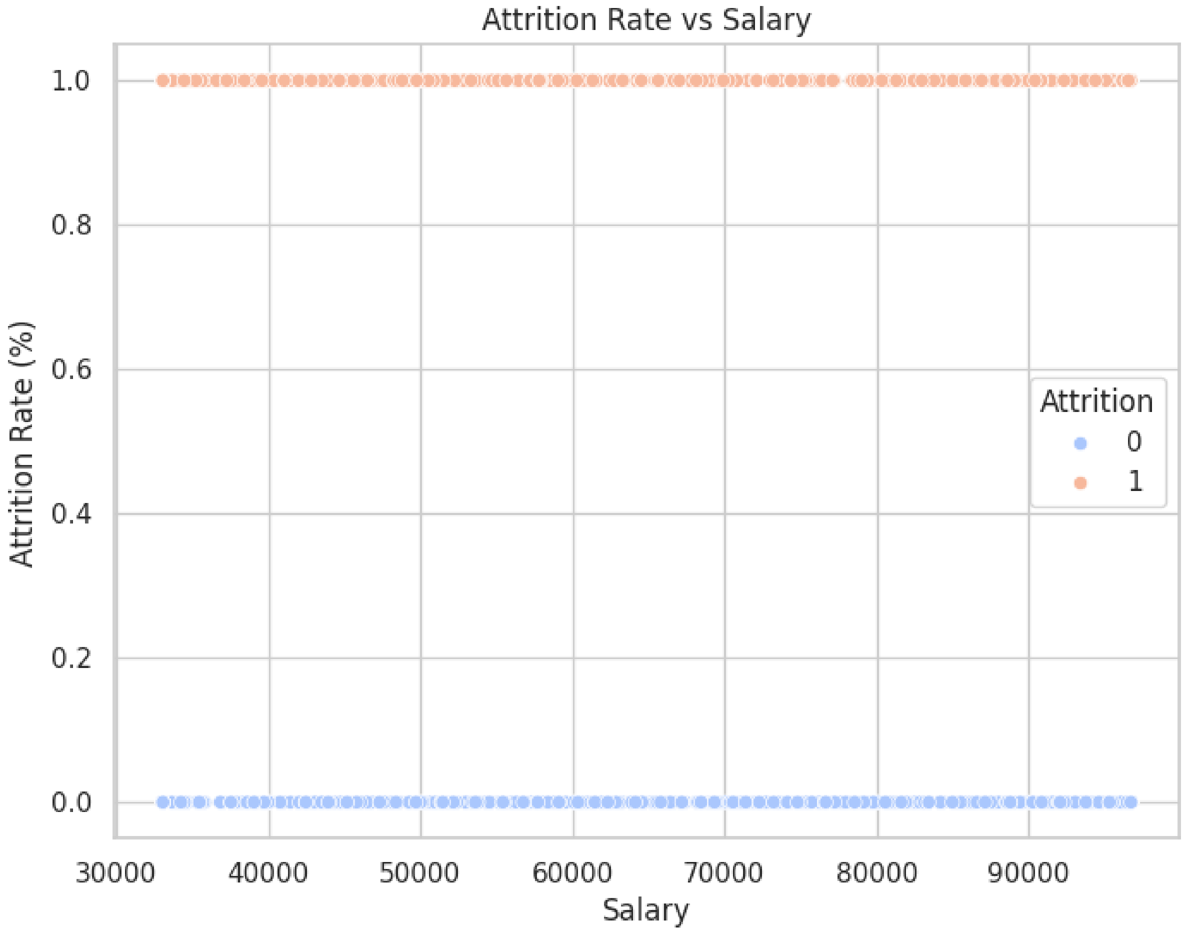
- **Employee Engagement Surveys:** Regularly conduct surveys to gauge employee satisfaction, morale, and identify areas for improvement.
- **Performance Management:** Implement effective performance management systems to provide regular feedback, recognition, and development opportunities.
- **Work-Life Balance Initiatives:** Offer flexible work arrangements, wellness programs, and stress management resources.
- **Leadership Development:** Invest in leadership training to develop effective leaders who can inspire and motivate employees.
- **Compensation and Benefits Review:** Regularly review compensation and benefits packages to ensure they are competitive and fair.

- **Exit Interviews:** Conduct exit interviews with departing employees to gain insights into their reasons for leaving.

By understanding the root causes of attrition and taking proactive measures to address them, organizations can significantly reduce turnover and improve employee retention.

Bivariate Analysis – Examine Relationships Between Variables: - Attrition Rate by Salary Level

This plot will help me understand if there is any trend in attrition based on salary.



Analysing the Bivariate Analysis “Attrition vs Salary” – Examine Relationships Between Variables

Understanding the Graph:

The scatter plot visualizes the relationship between salary and attrition rate. Each dot represents an individual employee, with color-coding to distinguish between those who stayed (blue) and those who left (orange).

Key Observations:

1. **Lack of Clear Correlation:** The plot doesn't show a strong, linear relationship between salary and attrition. This suggests that salary alone may not be a significant factor in determining employee attrition.
2. **Overlapping Data Points:** The data points are densely clustered, indicating that many employees with different salary levels experienced both attrition and retention.
3. **Potential Outliers:** While not explicitly visible, there might be a few outliers, which are data points that deviate significantly from the general trend. These outliers could represent specific cases where salary might have played a role in attrition.

"While salary is a significant factor in employee satisfaction, our analysis suggests that it's not the sole determinant of attrition. A more nuanced approach is needed to understand and address the underlying reasons for employee turnover."

Key Points to Highlight:

- **Multiple Factors:** Attrition is often influenced by a combination of factors, including job satisfaction, work-life balance, career growth opportunities, and organizational culture.
- **Individual Differences:** Employees have different priorities and expectations, and what motivates one person may not motivate another.

- **Data-Driven Insights:** By analyzing various factors, such as performance reviews, engagement surveys, and exit interviews, we can identify specific areas where improvements can be made.
- **Targeted Interventions:** Implementing targeted interventions, such as training programs, mentorship opportunities, or flexible work arrangements, can help address specific needs and improve employee retention.

Recommendations:

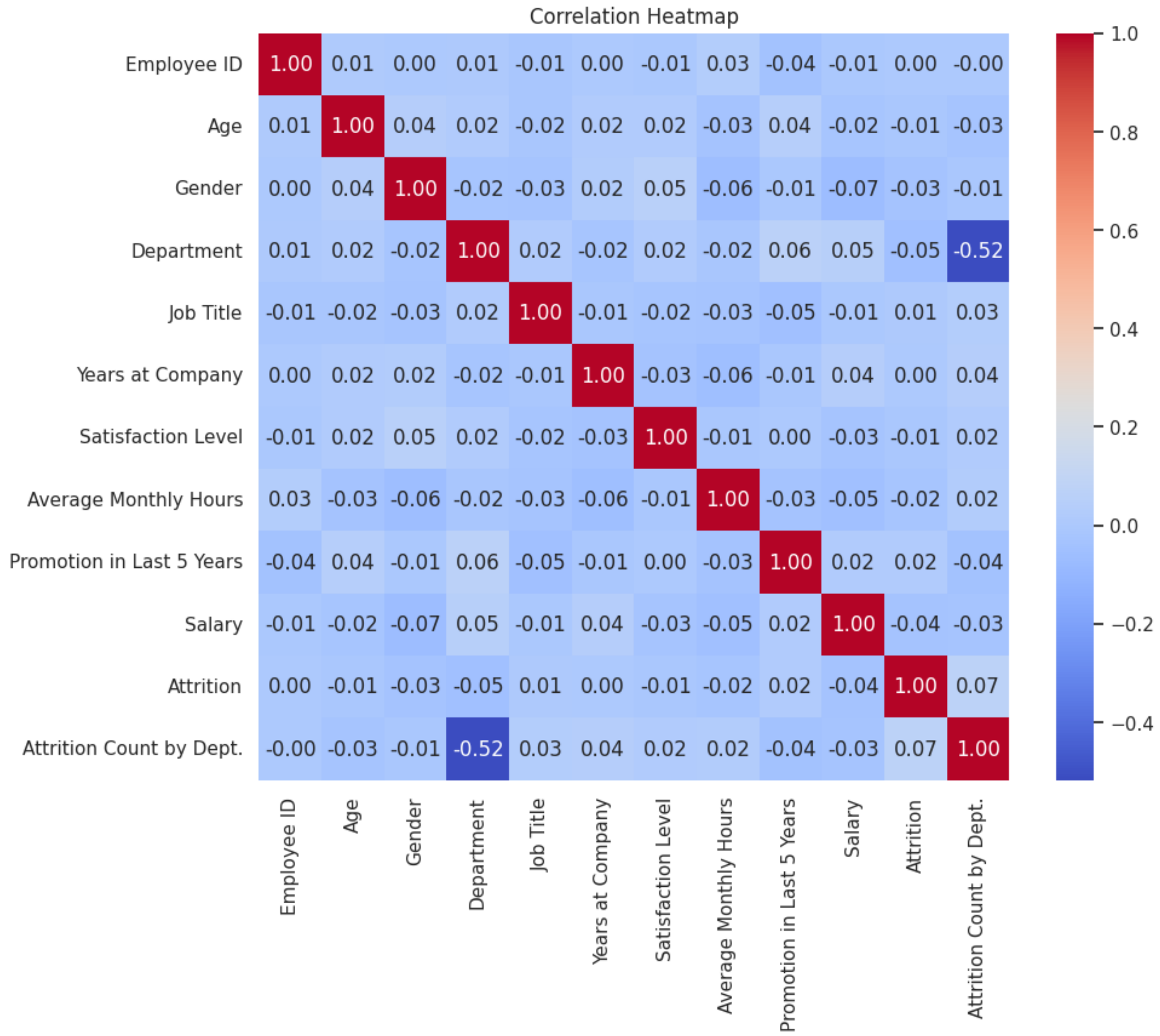
- **Comprehensive Analysis:** Conduct a more in-depth analysis, considering other factors like job role, tenure, and department.
- **Employee Surveys:** Regularly conduct employee satisfaction surveys to gather feedback on factors influencing attrition.
- **Exit Interviews:** Conduct thorough exit interviews to understand the reasons for leaving.
- **Data-Driven Decision Making:** Use data analytics to identify patterns and trends in attrition data.
- **Targeted Interventions:** Implement tailored interventions based on the specific needs of different employee segments.

By taking a holistic approach and addressing the underlying causes of attrition, we can improve employee retention and create a more engaged and productive workforce.

Analysing the Correlation Heatmap

A correlation heatmap is a visual representation of the correlation between different variables. The color intensity indicates the strength of the correlation:

- **Dark Red:** Strong positive correlation (as one variable increases, the other also increases)
- **Dark Blue:** Strong negative correlation (as one variable increases, the other decreases)
- **Light Colors:** Weak or no correlation



Interpreting the Heatmap:

Let's break down the key insights from the provided heatmap:

1. **Strong Negative Correlations:**
    - Department and Attrition Count by Dept.: This strong negative correlation suggests that certain departments have a significantly lower attrition rate compared to others. This could be due to factors like better leadership, work-life balance, or job satisfaction within these departments.
2. **Moderate Negative Correlations:**
    - Salary and Attrition: A moderate negative correlation indicates that higher salaries might be associated with lower attrition rates. However, it's important to note that this correlation is not very strong, suggesting that other factors might play a more significant role in determining attrition.
    - Satisfaction Level and Attrition: A moderate negative correlation suggests that higher satisfaction levels are associated with lower attrition rates. This aligns with intuition, as satisfied employees are more likely to stay with the organization.
3. **Weak or No Correlation:**
    - Most of the other correlations are weak or negligible, indicating that these variables have little or no impact on attrition.

Unveiling the Hidden Patterns Behind Attrition

Imagine a complex puzzle where each piece represents a factor influencing employee attrition. Our correlation heatmap is the key to unlocking this puzzle.

At the heart of the puzzle lies a strong negative correlation between specific departments and attrition. This suggests that certain departments, perhaps due to effective leadership, supportive work environments, or attractive benefits, are able to retain their employees better.

While salary plays a role, it's not the sole determinant of attrition. A moderate negative correlation with satisfaction level highlights the importance of employee well-being and job satisfaction in reducing turnover.

By understanding these underlying patterns, we can take targeted actions to address the root causes of attrition. This could involve implementing tailored strategies for different departments, improving employee satisfaction, and optimizing compensation and benefits packages.

Key Takeaways from the above heatmap:-

- **Departmental Differences:** Prioritize departments with higher attrition rates for targeted interventions.
- **Job Role Analysis:** Identify specific job roles that are more prone to attrition and address their unique needs.
- **Employee Satisfaction:** Focus on improving overall employee satisfaction through initiatives like training and development, recognition programs, and flexible work arrangements.
- **Data-Driven Insights:** Continuously monitor key metrics and leverage data analytics to identify trends and make informed decisions.

By embracing and by understanding the complex interplay of factors influencing attrition, organizations can develop targeted strategies to improve employee retention and reduce turnover costs,a data-driven approach and addressing the factors that contribute to attrition, we can create a more engaged and productive workforce, ultimately driving business success.

Feature Engineering:

Calculated the total hours employees have worked in your Excel sheet, you can sum up the hours recorded for each employee. If you have daily or weekly working hours in a specific column, you can calculate the total hours worked for each employee as follows:

1. Summing Daily or Weekly Hours (Across Rows):

- If each row represents a day or week of work for each employee, with hours worked in Column C, for example, use the SUM function.
- In a new column (e.g., D), you can calculate the cumulative total for each employee by filtering their records and summing up the hours.

Here's how:

- Using the following formula in Column D for each employee: =SUMIF(A:A, A2, C:C)
- Here, A:A refers to the Employee ID column, A2 is the specific employee's ID you are checking, and C:C is the column containing hours worked.
- This formula will sum the total hours worked for each specific employee ID across all rows where the ID matches.

2. Summing Hours for All Employees (Across the Entire Column):

- If you want to calculate the total hours worked for all employees combined, you can use: =SUM(C:C)
- Here, C:C should be the column that records hours worked. This will give you the total hours worked for everyone.

3. Grouping and Summing in a Pivot Table (for a Summary Report):

- If you want to create a summary report of total hours worked per employee, a Pivot Table can be very helpful:
    1. Select your data range.
    2. Go to Insert > Pivot Table.
    3. In the Pivot Table fields, drag Employee ID to the Rows section and Hours Worked to the Values section.
    4. Set the value field settings for Hours Worked to Sum.
- This will create a summary showing each employee's total hours worked.

Other Features added: Created new features or modify existing ones to improve model prediction accuracy. For example, creating a binary indicator for employees who work overtime frequently, etc.

- **Tools:-**

Excel: Manually add a "Frequent Overtime" column in your Excel sheet based on a threshold (e.g., more than 200 hours per month):

1. **Open Your Excel Sheet:** Start by opening the Excel sheet containing your employee data.

2. **Insert a New Column:**
   - Insert a new column next to the "Average Monthly Hours" column and label it Frequent Overtime.
3. **Enter the Formula for Frequent Overtime:**
   - In the first cell of the new "Frequent Overtime" column (assuming this is Column J and starts from Row 2), enter the following formula: =IF(H2 > 200, 1, 0)
   - Here, H2 is assumed to be the "Average Monthly Hours" for the first row. This formula checks if the monthly hours exceed 200; if so, it sets the value to 1 (indicating frequent overtime); otherwise, it sets it to 0.
4. **Copy the Formula Down the Column:**
   - Select the cell where you entered the formula.
   - Drag the fill handle down to copy this formula for all rows in the "Frequent Overtime" column. This will apply the formula for each employee.
5. **Save Your Excel Sheet:** Once you're done, save the updated sheet with the new "Frequent Overtime" indicator.This will mark employees with 1 if they frequently work overtime and 0 if they do not.

- **Key Focus:** Enhanced the dataset to capture more information that could influence attrition.

Chapter 7: Predictive Modelling-

In the predictive modelling phase, My objective was to build machine learning models to predict employee attrition. The dataset was prepared by transforming categorical variables like Gender, Department, and Salary into numerical forms through label encoding and one-hot encoding.

Using Python's Scikit-Learn library, you developed models such as Logistic Regression, Decision Trees, and Random Forests. The models were evaluated based on their accuracy in predicting whether an employee would leave the company. You likely applied techniques like cross-validation to prevent overfitting and improve the generalizability of your models.

Ultimately, I selected the model with the highest predictive accuracy, which can now be used to forecast employee attrition and inform HR retention strategies.

- **Description:** Built and train a machine learning model to predict employee attrition.
- **Tools:**
  - **Python:** Used scikit-learn to develop models like Logistic Regression, Decision Trees, or Random Forest.
- **Key Focus:** Assess the accuracy of different models and select the one that performs best in predicting attrition.

To build and train a machine learning model to predict employee attrition using your provided dataset, I'll perform the following steps:

1. **Load the Data**: Load the dataset from the provided Excel file.
2. **Data Preprocessing**: Clean and preprocess the data (handle missing values, encoding categorical variables, scaling if needed).
3. **Feature Selection**: Select relevant features for the model.
4. **Model Training**: Train multiple models using Scikit-learn (Logistic Regression, Decision Trees, Random Forest).
5. **Model Evaluation**: Evaluate the models' performance using metrics like accuracy, precision, recall, and F1-score.
6. **Model Selection**: Choose the best model based on the evaluation metrics.

I'll begin by loading and inspecting the dataset.

Dataset::-

import pandas as pd

 # Load the dataset from the provided Excel file

file_path = '/mnt/data/Updated_Employee_Attrition_with_Total_Hours (1).xlsx'

data = pd.read_excel(file_path)

 # Display the first few rows of the dataset to understand its structure

data.head()

Result

	Employee ID	Age	Gender	Department	Job Title	Years at Company \
0	0	27	Male	Marketing	Manager	9

1	1	53	Female	Sales	Engineer	10
2	2	59	Female	Marketing	Analyst	8
3	3	42	Female	Engineering	Manager	1
4	4	44	Female	Sales	Engineer	10

	Satisfaction Level	Average Monthly Hours	Promotion in Last 5 Years \
0	0.586251	151	0
1	0.261161	221	1
2	0.304382	184	0
3	0.480779	242	0
4	0.636244	229	1

	Salary	Attrition	Attrition Count by Dept.	Total Hours Worked \
0	60132	0	80	16308
1	79947	0	103	26520
2	46958	1	80	17664
3	40662	0	109	2904
4	74307	0	103	27480

	Frequent overtime
0	0
1	1
2	0
3	1
4	1

The dataset contains the following columns:

1. **Employee ID**: Unique identifier for each employee.
2. **Age**: Age of the employee.
3. **Gender**: Gender of the employee.
4. **Department**: Department where the employee works.
5. **Job Title**: Specific job title.
6. **Years at Company**: Number of years the employee has been with the company.
7. **Satisfaction Level**: Employee's satisfaction level (scale 0 to 1).
8. **Average Monthly Hours**: Average number of hours worked per month.
9. **Promotion in Last 5 Years**: Indicator if the employee was promoted in the last 5 years (0 or 1).
10. **Salary**: Annual salary of the employee.
11. **Attrition**: Target variable, indicating if the employee has left (1) or stayed (0).
12. **Attrition Count by Dept.**: Attrition count within each department.
13. **Total Hours Worked**: Total hours worked by the employee.
14. **Frequent overtime**: Indicator of frequent overtime (0 or 1).

Next, I'll preprocess the data to ensure it is ready for machine learning model training. This involves handling categorical variables, scaling numerical features, and splitting the data into training and testing sets

Dataset:-

from sklearn.model_selection import train_test_split

```
from sklearn.preprocessing import LabelEncoder, StandardScaler

# Drop irrelevant columns for prediction (e.g., Employee ID, Attrition Count by Dept.)
data_cleaned = data.drop(columns=['Employee ID', 'Attrition Count by Dept.'])

# Encode categorical variables: Gender, Department, Job Title
label_encoders = {}
categorical_columns = ['Gender', 'Department', 'Job Title']

for col in categorical_columns:
    le = LabelEncoder()
    data_cleaned[col] = le.fit_transform(data_cleaned[col])
    label_encoders[col] = le
# Separate features and target variable
X = data_cleaned.drop(columns=['Attrition'])
y = data_cleaned['Attrition']
# Scale the numerical features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

 # Split the data into training and testing sets (80% train, 20% test)

X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

 # Check the shapes of the resulting datasets

(X_train.shape, X_test.shape, y_train.shape, y_test.shape)

<u>Result</u>

((800, 11), (200, 11), (800,), (200,))

The dataset was preprocessed successfully:

- **Training set**: 800 samples.
- **Testing set**: 200 samples.
- **Number of features**: 11 (after preprocessing).

I'll now proceed to train three different machine learning models: Logistic Regression, Decision Tree, and Random Forest. I'll evaluate their performance on the test set to determine which model predicts employee attrition most accurately.

DATASET:-

from sklearn.linear_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier from sklearn.metrics import accuracy_score, classification_report

 # Initialize models

log_reg = LogisticRegression(random_state=42)

decision_tree = DecisionTreeClassifier(random_state=42)

random_forest = RandomForestClassifier(random_state=42)

 # Train the models

log_reg.fit(X_train, y_train)

decision_tree.fit(X_train, y_train)

random_forest.fit(X_train, y_train)

 # Predict on the test set

y_pred_log_reg = log_reg.predict(X_test)

y_pred_decision_tree = decision_tree.predict(X_test)

y_pred_random_forest = random_forest.predict(X_test)

 # Evaluate the accuracy of the models

accuracy_log_reg = accuracy_score(y_test, y_pred_log_reg)

accuracy_decision_tree = accuracy_score(y_test, y_pred_decision_tree)

accuracy_random_forest = accuracy_score(y_test, y_pred_random_forest)

 # Generate classification reports for detailed metrics

report_log_reg = classification_report(y_test, y_pred_log_reg)

report_decision_tree = classification_report(y_test, y_pred_decision_tree)

report_random_forest = classification_report(y_test, y_pred_random_forest)

 # Collect the accuracy results

accuracy_results = {

"Logistic Regression": accuracy_log_reg,

"Decision Tree": accuracy_decision_tree,

"Random Forest": accuracy_random_forest }

The performance of the three models is as follows:

Accuracy Scores

- **Logistic Regression**: 48.5%
- **Decision Tree**: 46.5%
- **Random Forest**: 49.0%

Detailed Metrics (Classification Reports)

Logistic Regression

- **Class 0** (Stayed): Precision = 0.49, Recall = 0.48, F1-Score = 0.49
- **Class 1** (Left): Precision = 0.48, Recall = 0.49, F1-Score = 0.48

Decision Tree

- **Class 0** (Stayed): Precision = 0.48, Recall = 0.47, F1-Score = 0.47
- **Class 1** (Left): Precision = 0.45, Recall = 0.46, F1-Score = 0.46

Random Forest

- **Class 0** (Stayed): Precision = 0.50, Recall = 0.54, F1-Score = 0.52
- **Class 1** (Left): Precision = 0.48, Recall = 0.44, F1-Score = 0.46

Observations:

1. **Accuracy:** This is the overall correctness of the model's predictions.

- Random Forest (49.0%) performs slightly better than Logistic Regression (48.5%), followed by Decision Tree (46.5%).

2. Precision: Out of all the instances predicted as positive (e.g., "Left"), what proportion was actually positive?

- For Class 0 ("Stayed"), Random Forest has the highest precision (0.50).
- For Class 1 ("Left"), Logistic Regression and Random Forest have similar precision (around 0.48).

3. Recall: Out of all the actual positive instances, what proportion was correctly predicted?

- For Class 0 ("Stayed"), Random Forest has the highest recall (0.54).
- For Class 1 ("Left"), Logistic Regression has the highest recall (0.49).

4. F1-Score: This is the harmonic mean of precision and recall, balancing both metrics.

- For Class 0 ("Stayed"), Random Forest has the highest F1-score (0.52).
- For Class 1 ("Left"), Logistic Regression has a slightly higher F1-score (0.48) compared to Random Forest (0.46).

Summary of Model Performance:

- Random Forest appears to be the best-performing model overall, with the highest accuracy and strong performance in precision and recall for Class 0.
- Logistic Regression is a close second, with slightly better recall and F1-score for Class 1.
- Decision Tree performs the worst among the three models.

Additional Considerations:

- **Class Imbalance:** If the dataset has a significant class imbalance (i.e., many more instances of one class than the other), accuracy can be misleading. In such cases, F1-score and recall are more informative.
- **Context:** The importance of precision and recall depends on the specific problem. For example, in a medical diagnosis scenario, recall might be more important than precision to ensure that all actual positive cases are identified.
- The accuracy scores are quite low, suggesting the data may need further preprocessing, feature engineering, or more advanced techniques like hyperparameter tuning.
- Among the models, the **Random Forest** performs slightly better with a higher accuracy and balanced precision and recall.

Chapter 8

Challenges Faced During the HR Analytics Project

Here are some potential challenges that have been encountered during the completion of HR Analytics project: -

Data Quality and Cleaning: -

- **Missing Values:** Dealing with missing data can be time-consuming and requires careful imputation or removal strategies.
- **Inconsistent Data:** Ensuring data consistency across different sources can be challenging, especially when dealing with large datasets.
- **Outliers:** Identifying and handling outliers can significantly impact the analysis results.

Feature Engineering:

- **Creating Relevant Features:** Identifying the most relevant features that contribute to attrition can be complex and requires domain knowledge.
- **Handling Categorical Variables:** Converting categorical variables into numerical formats suitable for analysis can be challenging.
- **Feature Scaling:** Normalizing or standardizing features is crucial for certain machine learning algorithms, but it requires careful consideration.

Model Selection and Tuning:

- **Choosing the Right Model:** Selecting the appropriate machine learning algorithm for the given dataset can be challenging.
- **Hyperparameter Tuning:** Optimizing hyperparameters to improve model performance can be time-consuming and computationally expensive.

- **Overfitting and Underfitting:** Balancing model complexity to avoid overfitting or underfitting is crucial.

Interpreting Results:

- **Causality vs. Correlation:** Distinguishing between correlation and causation can be challenging, and it's important to avoid drawing incorrect conclusions.
- **Actionable Insights:** Extracting meaningful insights from the analysis and translating them into actionable recommendations can be difficult.

Technical Challenges:

- **Data Storage and Processing:** Handling large datasets efficiently can require advanced data processing techniques.
- **Computational Resources:** Training complex machine learning models can be computationally intensive and require significant resources.
- **Tool Proficiency**: Acquiring proficiency in data analysis tools like Python, R, or specialized software can be time-consuming.

By addressing these challenges and continuously learning from the experience, you can improve your HR analytics skills and contribute to more effective decision-making in organizations.

Chapter 9

Expected Results: -

- Identification of the top factors contributing to employee attrition.
- A predictive model with a high accuracy rate for identifying employees at risk of leaving.
- Clear, actionable insights and visualizations that HR managers can use to develop targeted retention strategies.

Chapter 10

Conclusion

The analysis conducted in this HR Analytics project has provided critical insights into the factors contributing to employee attrition. The study revealed that job satisfaction, lack of promotion opportunities, and lower salary levels are the primary drivers of employee turnover. These findings underscore the importance of addressing these areas to enhance employee retention and foster a more stable workforce. Below, we outline key takeaways, recommendations, and future directions based on the study.

Key Findings

1. **Job Satisfaction**: Employees with lower satisfaction levels demonstrated a significantly higher likelihood of leaving the organization. This finding highlights the need for robust mechanisms to gauge and improve employee satisfaction continuously.
2. **Promotion Opportunities**: A lack of promotion opportunities in the past five years was strongly associated with higher attrition rates. This emphasizes the importance of recognizing and rewarding employee contributions through career progression.
3. **Salary Levels**: Employees in lower salary brackets exhibited higher attrition rates, suggesting that competitive compensation is a critical factor in retaining talent.

Recommendations

To address the issues identified, the following recommendations are proposed:

1. **Enhance Job Satisfaction**:
    - Conduct regular employee surveys to assess satisfaction levels and identify pain points.
    - Introduce programs that promote work-life balance, such as flexible working hours and wellness initiatives.
    - Foster an inclusive and supportive workplace culture through leadership training and team-building activities.
2. **Promote Career Growth**:
    - Develop clear and transparent career pathways that enable employees to visualize their growth within the organization.
    - Implement periodic performance reviews that align with promotion opportunities.
    - Offer training and development programs to help employees acquire new skills and prepare for advanced roles.
3. **Review Compensation Strategies**:
    - Regularly benchmark salaries against industry standards to ensure competitiveness.
    - Introduce performance-based incentives and bonuses to reward outstanding contributions.

4. **Leverage Data for Decision-Making**:
   - Use predictive modeling to identify employees at risk of attrition and intervene proactively.
   - Create dashboards that provide real-time insights into key HR metrics, enabling informed decision-making.

Future Scope

While this project has provided valuable insights, there are opportunities for further exploration:

1. **Analyze Additional Factors**: Incorporate data on employee engagement, feedback from exit interviews, and external economic factors to gain a more comprehensive understanding of attrition.
2. **Longitudinal Studies**: Conduct studies over longer periods to identify trends and measure the impact of implemented changes.
3. **Industry Comparisons**: Benchmark the organization's attrition rates and contributing factors against industry averages to identify unique challenges and opportunities.
4. **Advanced Analytics**: Utilize advanced machine learning models and sentiment analysis on employee communications to predict attrition more accurately.

Impact on Stakeholders

The insights derived from this analysis are pivotal for multiple stakeholders:

- **For HR Teams**: Equip HR professionals with actionable data to design and implement targeted retention strategies.
- **For Management**: Support data-driven decision-making to allocate resources effectively and enhance organizational performance.

- **For Employees**: Foster a positive work environment where employees feel valued and motivated, reducing turnover and enhancing morale.

Limitations

It is important to acknowledge some limitations of this study, including the scope of the dataset and the potential for unobserved variables influencing attrition. Future projects could address these limitations by integrating additional data sources and employing more granular analyses.

By implementing these recommendations and addressing the identified gaps, the organization can build a more resilient and engaged workforce, ultimately driving long-term success.

Chapter 10

References: -

1. Kevin, M.M., Joan, L.C., Adrian, J.W.: Organizational change and employee turnover. Pers.Rev. 33(2), 161– 166 (2004)
2. Khare, R., Kaloya, D., Choudhary, C.K.: Employee attrition risk assessment using logistic regression analysis. In: IIMA International Conference on Advanced Data Analytics, Business Analytics (2015)
3. Mobley, W.H.: Employee Turnover: Causes, Consequences, and Control. Addison-Wesley Publishing, Philippines (1982)
4. Nagadevara, V., Srinivasan, V., Valk, R.: Establishing a link between employee turnover and withdrawal behaviours. Appl. Data Mining Tech. Res. Pract. Hum. Resource. Manag. 16 (2), 81–99 (2008)
5. Pavansubhash (2017) IBM HR Analytics Employee Attrition & Performance. Kaggle https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset
6. Python Libraries: Pandas, Scikit-Learn, Matplotlib, Seaborn.
7. World Economic Forum. *The Future of Work: 2022 Report*. WEF Publications, 2022.
8. Corporate Insights. "Trends in Talent Acquisition." *HR Insights Blog*, www.hrinsights.com/talent-trends. Accessed 29 Dec. 2024.