POWER BI PROJECT REPORT (PRE)

BY- SRISHTI MISHRA

REG NO. – 2023JULB01233

TABLE OF CONTENTS

ABOUT THE TOPIC/DATASET

Understanding Employee Attrition Rate:

Employee Attrition Rate refers to the percentage of employees who leave a company over a specific period. It is a critical HR metric that reflects the stability and health of an organization's workforce. High attrition rates can indicate potential problems such as dissatisfaction, lack of engagement, or better opportunities elsewhere, while low attrition rates often signify a stable and satisfied workforce.

Why Analyse Employee Attrition?

1. Improving Employee Retention:
    - High attrition rates can lead to increased costs due to hiring, onboarding, and training new employees. Analyzing attrition helps identify root causes, allowing companies to address issues and retain talent.
2. Boosting Employee Morale:
    - Frequent departures can lead to a decline in morale among remaining employees. Understanding why employees leave can help prevent this and create a positive work culture.

Dataset Overview:

This dataset appears to focus on employee attributes and their relation to attrition (leaving the company). It contains various features that can help analyse the factors driving employee retention or attrition, making it useful for HR analytics and strategic decision-making. Here's a breakdown of the dataset:

- Employee: Unique identifier for each employee.
- Age: The age of the employee.
- Gender: The gender of the employee (e.g., Male, Female).
- Department: The department where the employee works (e.g., Sales, Marketing, Engineering).
- Job Title: The specific job role of the employee (e.g., Manager, Engineer).
- Years at Company: The number of years the employee has been with the company.
- Satisfaction Level: A measure of the employee's satisfaction at work, often represented as a score between 0 and 1.
- Average Monthly Hours: The average number of hours worked per month.
- Promotion_Last_5Years: Indicates whether the employee has been promoted in the last 5 years (0 = No, 1 = Yes).
- Salary: The salary of the employee.

PROBLEM STATEMENT

"To identify the key drivers of employee attrition in the company and develop predictive models to forecast attrition risk, focusing on factors such as satisfaction level, average monthly hours, promotion history, salary, and other demographic and job-specific attributes. The goal is to provide actionable insights for HR to improve retention strategies, enhance employee satisfaction, and optimize workforce stability."

DATA REQUIREMENTS

To effectively use this dataset in **Power BI** for analyzing and visualizing employee attrition, you should consider the following data requirements:

1. Data Preparation and Cleaning Requirements

- **Data Types**: Ensure all columns have appropriate data types. For example, numerical fields (e.g., Age, Years at Company) should be set as integer/decimal, and categorical fields (e.g., Gender, Department) should be text.
- **Missing Values Handling**: Check for and handle missing data, either by imputing values, removing incomplete records, or creating indicators for missing data.
- **Data Normalization**: If needed, normalize numerical data (e.g., Salary, Satisfaction Level) to make the values comparable.
- **Data Quality Check**: Remove duplicates and ensure data consistency across records (e.g., consistent spellings for departments).

2. Data Model Requirements

- **Fact Table**: The main table for analysis will be the employee dataset itself, with columns representing various attributes and the attrition flag.
- **Dimension Tables (If Available)**:
    - **Departments**: If available, use a dimension table to map and describe departments.
    - **Job Titles**: Create a dimension table to categorize job titles or roles.
    - **Time Dimension**: Consider using a timetable for aggregating attrition trends over time, if date-based data is available (e.g., join dates, attrition dates).

3. Columns for Analysis

- **Employee**: Unique identifier (can be used as a primary key in the data model).
- **Age**: Numeric field for demographic segmentation.
- **Gender**: Categorical data for diversity analysis.
- **Department**: Categorical data for departmental-level analysis.
- **Job Title**: Role-based categorization for role-level insights.
- **Years at Company**: Numeric data representing tenure; useful for attrition analysis by tenure.
- **Satisfaction Level**: Numeric data (likely a score between 0 and 1); useful for gauging employee engagement.
- **Average Monthly Hours**: Numeric data for workload analysis.
- **Promotion_Last_5Years**: Categorical or binary data to assess the impact of career growth on attrition.
- **Salary**: Numeric data for salary band analysis.
- **Attrition**: Binary target variable (0 = No, 1 = Yes) used to analyze the factors affecting attrition.

DATA COLLECTION

Typically involves gathering employee information from a variety of internal and external sources. Here are potential data collection sources:

1. Human Resource Information System (HRIS)

- **Employee Records**: Age, gender, job title, department, hire date, promotion history, and other demographics.
- **Compensation Data**: Salary, bonuses, and benefits data.
- **Performance Reviews**: Employee performance ratings, reviews, and related metrics.
- **Promotion History**: Records of promotions, transfers, and role changes within the organization.
- **Employee Satisfaction Surveys**: Satisfaction levels, engagement scores, and feedback.

2. Time and Attendance Systems

- **Work Hours**: Data on employee work hours, including overtime and average monthly hours.

- **Absenteeism Records**: Information on employee leaves, sick days, and overall attendance.

3. Exit Interviews and Surveys

- **Attrition Reasons**: Qualitative and quantitative data from employees who have resigned or been terminated, offering insights into reasons for leaving.
- **Feedback**: Responses related to work culture, management, compensation, and opportunities for growth.

4. Payroll Systems

- **Salary Data**: Monthly or annual salary figures.
- **Compensation History**: Changes in compensation, bonuses, or incentives.

5. Employee Satisfaction and Engagement Platforms

- **Employee Surveys**: Data on job satisfaction, engagement, workplace culture, and more.
- **Pulse Surveys**: Periodic surveys that gauge employee morale, challenges, and organizational alignment.

6. Performance Management Systems

- **KPIs and Goals**: Data on performance targets, key performance indicators (KPIs), and achievements.
- **Performance Trends**: Year-over-year performance data that might impact attrition.

7. Learning and Development Systems

- **Training Records**: Data on training completed, skill upgrades, certifications, and learning hours.
- **Career Development**: Information on courses and development programs employees participated in.

DATA VALIDATION

When preparing your employee attrition dataset for **Power BI dashboard creation**, it is critical to follow systematic **data validation steps**. These steps ensure that your data is accurate, clean, consistent, and ready for meaningful visualization. Here is a comprehensive list of data validation steps:

1. Data Integrity Checks

- **Unique Identifiers**: Ensure Employee_ID values are unique and non-null to correctly identify each record.

- **Data Completeness**: Check those key columns (e.g., Age, Department, Attrition) are not missing values, especially critical variables used for analysis.

2. Data Type Validation

- **Data Types**: Confirm that all fields have the appropriate data types (e.g., Age as integer, Satisfaction Level as float or decimal, Attrition as integer/binary).
- **Date and Time Fields**: Ensure date fields are in a recognized format for proper date-time calculations in Power BI.

3. Consistency Checks

- **Categorical Values**: Verify that categorical data (e.g., Gender, Department) has consistent spelling, capitalization, and formatting (e.g., no "M" and "Male" mixed).
- **Value Ranges**: Ensure numerical columns (e.g., Satisfaction Level should range from 0 to 1, Average Monthly Hours should be within a reasonable range) are valid.
- **Outlier Detection**: Identify and address outliers, such as unusually high Average Monthly Hours or Salary values that could skew results.

4. Handling Missing Data

- **Null Values**: Identify null or missing values in columns and decide on appropriate handling methods (e.g., removal, imputation, or creating a "missing" category).
- **Default Values**: Check if columns have inappropriate default values (e.g., 0 for Years at Company instead of missing) and correct them if necessary.

5. Data Transformation and Normalization

- **Standardization**: Standardize values for comparison across different records, such as scaling Satisfaction Level to a common range if necessary.
- **Data Formatting**: Ensure all date, currency, and numerical data is formatted consistently (e.g., currency fields show the same units).
- **Data Categorization**: Create meaningful categories from continuous data if applicable (e.g., age groups, salary ranges).

DATA CLEANING:

To prepare your employee attrition dataset for analysis and visualization in Power BI, it's important to follow data cleaning steps. These steps ensure the dataset is accurate, reliable, and ready for analysis. Here is a comprehensive list of data cleaning steps:

1. Remove Duplicates

- Identify Duplicate Rows: Check for duplicate records based on unique identifiers such as Employee ID.
- Remove Duplicates: Remove any duplicate rows to avoid overcounting and data inaccuracies.

2. Handle Missing Data

- Identify Missing Values: Identify columns with missing values using Power BI's "Data View" or by adding conditional columns to count nulls.
- Impute Missing Values:
    - For Numerical Data: Replace missing values with the mean, median, or mode, or use interpolation techniques.
    - For Categorical Data: Use the most common category or create a new category such as "Unknown" or "Not Specified."
- Drop Columns/Rows: If missing data is excessive and cannot be meaningfully imputed, consider dropping those rows or columns.

3. Correct Data Types

- Ensure Consistent Data Types: Convert columns to the appropriate data types (e.g., convert Age to integer, Satisfaction_Level to decimal/float, and Attrition to binary/integer).
- Date Fields: Ensure date columns are correctly formatted as dates in Power BI.

4. Trim Whitespaces

- Remove Leading and Trailing Whitespaces: Clean up text data in categorical columns (e.g., Department, Job Title) to avoid mismatches due to extra spaces.

5. Standardize Categorical Data

- Consistent Labelling: Ensure all values in categorical columns are consistently labelled (e.g., "HR" vs. "Human Resources" or "M" vs. "Male").
- Case Sensitivity: Convert text to lowercase or uppercase for consistent comparison.

6. Identify and Handle Outliers

- Visual Inspection: Use boxplots, scatterplots, or histograms in Power BI to identify potential outliers in columns like Salary, Age, or Average Monthly Hours.
- Treat Outliers: Decide on an approach to handle outliers:
    - Cap Extreme Values: Replace extreme values with a specified maximum/minimum.

- Remove: Drop rows with extreme outliers if they are considered noise.
- Transform Data: Apply transformations such as logarithmic or square root to reduce the impact of outliers.

7. Handle Zero and Null Values in Numerical Columns
   - Zero Values: Check for zero values in columns where they are unlikely (e.g., Years at Company should generally not be zero).
   - Replace or Flag: Decide to replace zero values with a more appropriate number or create a flag column indicating potentially problematic values.
8. Address Invalid or Inconsistent Values
   - Age Range Validation: Ensure Age values fall within a logical range (e.g., 18-65).
   - Department and Job Title Validation: Check for valid and known values. Correct or map inconsistent entries.
   - Satisfaction Scores: Ensure scores fall within the expected range (e.g., 0 to 1 for a score range).
9. Standardize Dates
   - Date Format: Ensure date columns (e.g., Hire Date) are standardized to a common format and recognized as a date in Power BI.
   - Calculate Tenure: Create a new column to calculate Tenure based on hire date if applicable, ensuring consistent values for further analysis.

TOOLS TO BE USED

To create a meaningful and effective dashboard for an employee attrition dataset, you can utilize a variety of tools and features for data processing, visualization, and analysis. Here's a list of recommended tools and their roles in building your Power BI dashboard:

1. Power BI Desktop
   - Primary Tool: Power BI Desktop is the primary tool for creating data models, developing visuals, and designing your dashboard.
   - Features:
     - Data Import: Import data from multiple sources (Excel, SQL, CSV, etc.) using the "Get Data" functionality.
     - Power Query Editor: Use Power Query for data transformation, cleaning, filtering, and merging data.

- Data Modeling: Create relationships between tables, define calculated columns, and create measures using DAX (Data Analysis Expressions).
- Visualizations: Create visual elements like bar charts, pie charts, histograms, scatter plots, line graphs, maps, and more for presenting data insights.
- Dashboard Design: Use layouts, themes, and slicers to build a coherent and interactive dashboard.

2. Power Query

- Data Transformation: Clean, transform, and prepare your data for analysis using features such as:
  - Remove/Transform Columns: Remove unnecessary columns, rename columns, split, or combine columns.
  - Data Type Changes: Convert data types to ensure consistency.
  - Handling Nulls and Missing Values: Fill or replace missing data directly in Power Query.
  - Data Filtering: Filter out irrelevant records, such as temporary employees.

3. DAX (Data Analysis Expressions)

- Creating Calculated Columns and Measures: Use DAX formulas to perform complex calculations like attrition rates, average tenure, or trends.
- KPI Calculations: Create Key Performance Indicators (KPIs) such as attrition percentage, average monthly working hours, and employee satisfaction levels.
- Data Aggregation: Build complex measures that aggregate data across multiple dimensions.

4. Power BI Visualizations

- Visual Charts and Graphs:
  - Bar/Column Charts: For visualizing counts of attrition across different departments or job roles.
  - Line Charts: For tracking trends over time, such as attrition rate changes.
  - Scatter Plots: For identifying relationships between variables, such as employee satisfaction and attrition.
  - Pie/Donut Charts: For displaying category distribution (e.g., reasons for attrition).

- Tree Maps: For hierarchical views of data (e.g., attrition rate by business unit).
- Heatmaps: To highlight patterns based on high or low attrition rates.

- Custom Visuals: Import and use custom visuals from the Power BI marketplace to create more engaging dashboards.

DASHBOARD

Creating dashboards in Power BI is crucial for organizations and data professionals as it offers significant benefits in terms of data visualization, analysis, and decision-making. Here are the key reasons why dashboard creation in Power BI is so important:

1. Enhanced Data Visualization

- Interactive Visuals: Power BI allows you to create interactive and visually engaging dashboards using charts, graphs, maps, and custom visuals. This makes complex data easier to understand at a glance, enabling users to spot trends, patterns, and anomalies quickly.
- Customizable Views: Dashboards can be customized to match the needs of different audiences, highlighting the most relevant insights for each stakeholder group.

2. Real-Time Data Analysis

- Live Data Updates: Power BI dashboards can be connected to live data sources, providing users with up-to-date insights. This is particularly useful for monitoring key performance indicators (KPIs), sales, customer metrics, and other time-sensitive information.
- Scheduled Refreshes: Power BI enables automatic data refresh schedules, ensuring dashboards reflect the latest data without manual intervention.

3. Improved Decision-Making

- Data-Driven Insights: Dashboards provide actionable insights by presenting data in a concise, clear, and visually compelling manner. Decision-makers can base their strategies on accurate and timely information.
- Quick Analysis: With the ability to filter, drill down, and slice data directly within the dashboard, users can quickly analyze the impact of specific factors on key metrics.

4. Centralized Data Access

- Single Source of Truth: Power BI dashboards integrate data from multiple sources (e.g., databases, Excel files, APIs, cloud services), creating a

centralized view of data. This ensures all stakeholders work with the same, consistent data set.

- Data Consolidation: By bringing data from various departments or business units into a unified view, dashboards provide a holistic understanding of organizational performance.

5. Interactive and User-Friendly Interface

- Ease of Use: Power BI's intuitive drag-and-drop interface makes it easy for non-technical users to create, modify, and explore dashboards without extensive programming knowledge.
- Exploration and Filtering: Users can interact with visuals by applying filters, slicers, and drill-throughs, exploring data dynamically to gain deeper insights.

6. Customizability and Flexibility

- Custom Dashboards: Dashboards can be tailored to individual users' needs, showing only the information relevant to their role, team, or project.
- Responsive Design: Power BI dashboards are responsive, ensuring that they look good and function well on different devices, including desktops, tablets, and mobile phones.