Employee Attrition Analysis and Prediction

This project aims to provide insights into the factors influencing employee attrition and predict which employees are likely to leave the company.

Problem Statement:

Acme Corporation, a leading tech company, is facing a significant challenge with employee turnover. The HR department is concerned about the increasing rate of attrition, as it negatively impacts team dynamics, project continuity, and overall company morale. To address this issue, Acme Corporation wants to leverage data analytics and machine learning to understand the factors influencing employee turnover and predict which employees are likely to leave in the near future.

Dataset:

Acme Corporation has provided historical data on employee demographics, job satisfaction, work environment, performance metrics, and turnover status. This dataset spans the last five years and includes information on employees who have left the company and those who are still currently employed.

The dataset typically includes several features that provide insights into employee characteristics, job satisfaction, and performance. While the exact features may vary, here's a general list of common features you might find in such a dataset:

Employee ID: A unique identifier for each employee.

Age: The age of the employee.

Attrition: A binary variable indicating whether the employee has left the company (1) or is still employed (0).

Business Travel: The frequency and nature of business-related travel (e.g., "Travel Rarely," "Travel_Frequently," "Non-Travel").

Department: The department to which the employee belongs (e.g., "Sales," "Research & Development," "Human Resources").

Distance From Home: The distance of the employee's residence from the workplace.

Education: The employee's level of education (e.g., "1: 'Below College'," "2: 'College'," "3: 'Bachelor'," "4: 'Master'," "5: 'Doctor').

Education Field: The field in which the employee's education lies (e.g., "Life Sciences," "Medical," "Marketing").

Environment Satisfaction: The level of satisfaction with the work environment on a scale.

Gender: The gender of the employee.

Job Involvement: The degree to which the employee is involved in their job.

Job Level: The level or rank of the employee's position.

Job Role: The specific role or title of the employee's job.

Job Satisfaction: The level of satisfaction with the job on a scale.

Marital Status: The marital status of the employee.

Monthly Income: The monthly salary of the employee.

Num Companies Worked: The number of companies the employee has worked for.

Over Time: Whether the employee works overtime or not.

Performance Rating: The performance rating of the employee.

Relationship Satisfaction: The level of satisfaction with relationships at the workplace.

Stock Option Level: The level of stock options provided to the employee.

Total Working Years: The total number of years the employee has been working.

Training Times Last Year: The number of training sessions the employee attended last year.

Work-Life Balance: The balance between work and personal life.

Years At Company: The number of years the employee has been with the current company.

Years In Current Role: The number of years the employee has been in their current role.

Years Since Last Promotion: The number of years since the last time the employee was promoted.

Years With Current Manager: The number of years the employee has been working under the current manager.

Please note that this is a general list, and the actual dataset might include additional features or variations. It's essential to explore the dataset thoroughly to understand the specifics of each feature and its relevance to the analysis.

Business Intelligence (BI) Analysis:

Data Exploration and Visualization:

Create interactive dashboards using BI tools to visualize trends and patterns in employee turnover.

Identify departments, roles, and specific projects with the highest turnover rates.

Descriptive Analytics:

Generate reports that highlight the primary reasons for attrition based on employee feedback, exit interviews, and other relevant sources.

Analyze the impact of factors like job satisfaction, workload, and career growth on employee turnover.

Predictive Analytics with BI:

Build predictive models within the BI tools to estimate the likelihood of turnover for current employees.

Implement scenario analysis to understand the potential impact of changes in satisfaction levels, compensation, or management practices.

Power BI:

Data Import: Power BI allows you to import data from various sources such as Excel, CSV, SQL Server, etc. You can import your dataset by clicking on "Get Data" and selecting the appropriate data source.

Data Cleaning and Transformation: After importing your data, you might need to clean it by removing duplicates, handling missing values, or correcting data types. Power BI provides a user-friendly interface for performing these tasks through the "Transform Data" option.

Data Modelling: In Power BI, you can create relationships between tables if your dataset contains multiple related tables. This step is crucial for proper data analysis and visualization.

Visualization: Power BI offers a wide range of visualization options such as bar charts, line charts, pie charts, maps, etc. You can create visualizations by dragging fields from your dataset onto the report canvas and selecting the appropriate visualization type.

Report Creation: Design your report layout by arranging the visualizations on different pages. You can add titles, headers, and footers to make your report more informative.

Interactivity: Power BI allows you to add interactivity to your report by adding filters, slicers, and drill-down capabilities. Users can interact with the report to explore the data further.

Formatting and Styling: You can customize the appearance of your visualizations by changing colors, fonts, and other formatting options to make them visually appealing and easy to understand.

Publishing: Once your report is ready, you can publish it to the Power BI service to share it with others. The report can be accessed through the Power BI web portal or mobile app.

Python:

Data Import: Use the panda's library in Python to import your dataset into a Data Frame, which is a tabular data structure similar to a spreadsheet.

Data Cleaning and Transformation: Clean and transform your data using pandas methods such as drop duplicates (), fillna(), and astype() to handle missing values, duplicates, and data type conversions.

Data Analysis: Use pandas and numpy libraries to analyze your data. You can calculate descriptive statistics, perform aggregations, group data, etc., to gain insights into your dataset.

Visualization: Use libraries like matplotlib or seaborn to create visualizations such as histograms, scatter plots, box plots, etc., to represent your data graphically.

Report Generation: You can generate reports in Python using libraries like Report Lab or matplotlib. These libraries allow you to create PDF reports with text, tables, and visualizations.

Interactive Reporting (Optional): If you need interactive reporting, you can use libraries like Plotly or Dash to create web-based interactive dashboards with Python.

Exporting: Export your reports to different formats such as PDF, HTML, or images using the appropriate functions provided by the libraries you are using.

Both Power BI and Python provide powerful tools for data analysis and reporting. The choice between them depends on factors such as your familiarity with the tools, specific requirements of the project, and the level of customization and control you need over the reporting process.