1. **Business problems**

IBM is a global technology company known for their leading position in several technology sectors, such as cloud computing, artificial intelligence, and modern devices. With over a century of innovation, IBM has a diverse workforce across multiple departments and job roles. Like many large organizations, IBM also faces the challenge of employee attrition when skilled employees leave the company for many reasons. A high attrition rate can lead to increased recruitment costs, loss of knowledge, and decreased organizational performance. Therefore, IBM is facing an issue that requires understanding the key factors driving attrition to develop effective retention strategies, which can be resolved by data visualization. Using charts to compare attrition rates, analyze employee satisfaction, and discover relevant data correlations can help decision-makers uncover hidden insights of employee attrition and performance. The visualization can support improving HR policies and reducing the speed of employee turnover.

1. **Project objectives**

The objective of this project is to analyze the IBM Analytics Employee Attrition dataset and use data visualization to demonstrate key factors that affect employee attrition. The project will focus on the following points:

* Develop at least 04 static visualizations and point out their business insights: Different visualization types (bar chart, line chart, heat map, scatter plot, etc.) will be created based on raw data such as education, positions, income, and years at the company, to create a first impression about IBM's human resources.
* Create an interactive dashboard in Power BI to tell a story of employee attrition and performance through visualizations.
* Make an analysis to explore correlations between data fields, and how their combination can drive employee attrition.

1. **Datasets overview**

Dataset name: IBM HR Analytics Employee Attrition & Performance (1470 records) *(please see the dataset file in the submission)*

This dataset contains data collected from the IBM HR system that reflects real-world characteristics and behaviors of IBM’s workforce. It has 35 columns and 1470 records. For each row, all personal and professional information related to human resources analysis is recorded for further analysis.

Key columns that describe their career at IBM include:

* Job Role, Department, Income, Job Satisfaction, Work-Life Balance...

Key columns that describe personal information include:

* Age, Gender, Education, Marital Status...

Key columns that describe their work experience include:

* Years at Company, Business Travel, Training Times...

This dataset is appropriate for analyzing business problems related to HR, especially employee attrition. First, it includes all essential data that influence an employee's decision to leave or stay. There are both quantitative data (e.g., Monthly Income, Age, Years at Company) and categorical data (e.g., Job Role, Marital Status, Department) to support multi-dimensional analysis. If we look only at the raw data, it would be hard to identify any valuable insights. However, when combining the data and visualizing it, meaningful insights into the employee experience at IBM will appear. For example, trends in job satisfaction, workload, and career development can reflect the overall attrition trend.

Data visualization makes complex relationships easier to interpret and communicate. While raw data only lists employees who left the company, visualizations can reveal clear patterns—such as which age groups or departments tend to experience higher attrition. Is it true that employee turnover mainly comes from those with lower income? Clear visualization helps the HR team move from observation to action. They will not only identify who is leaving but also discover the reasons behind these decisions. These insights may support HR in building new strategies to improve work-life balance or increase retention incentives for veteran employees. In this way, the dataset aligns with the project objectives by providing evidence for data-driven decisions to improve employee engagement at IBM.

1. **Visualization techniques**

I plan to integrate several visualization techniques to discover the data insights. However, in this section, I will only discuss and propose my ideas for the visualization, without drawing actual diagrams because the data needs to be cleaned before entering this step.

**Static visualization**

* **Bar chart:**

Bar charts can be used to compare the attrition rates across Job Role, Departments, or Education Level as they are effective for comparing categorical data.

Potential insights: Which job functions or departments have the highest turnover.

Business impact: HR can focus on the insights to develop retention strategies for these areas.

* **Scatter plot:**

This can be used to see if employees who earn lower salaries and have a low number of years at the company are more likely to resign.

Potential insights: The plot can show the relationship between income and turnover rate.

Business impact: HR and the employee development team can observe the scatter plot to perform periodic salary reviews or identify appropriate strategies to increase the retention rate of young employees.

* **Heat map:**

Similar to a scatter plot, we can use a heat map to illustrate attrition (color) by Job Role and Working Environment Satisfaction.

Potential insights: HR can detect high-risk combinations that often leave the company. For example, a group of technical employees with low satisfaction often has high attrition.

Business impact: If the root cause of high attrition is a poor working environment, company leaders can plan to improve office facilities or create more community events to promote a friendly working culture.

Other visualization techniques that can be considered: Line chart, pie chart, stacked columns…

**Interactive visualization**

Unlike static visualizations, when choosing visualizations for an interactive dashboard, we need to consider more factors, such as the purpose of the dashboard, the type of data to show, the target audience, key questions to answer, and interactivity features.

As the target audience consists of non-tech users, and most might come from the HR team, they would prefer simple and clean visualizations instead of complex ones.

Some proposed visualizations for the interactive dashboard are as follows:

* A slicer filter that can be used to filter users by department, gender, marital status, and education to help HR perform an interactive analysis for all employees.
* Line charts can show the attrition trends over time. We can reveal insights like how busy seasons can lead to increased resignations.
* A tree map can show attrition counts grouped by categorical data. This could quickly visualize where most attrition is concentrated.

We can also integrate the above static visualizations into the dashboard if they align with the dashboard’s purpose.