**IBM HR ANALYTICS REPORT**

**SITUATION**

* IBM’s HR department was facing a significant employee attrition problem and wanted to understand which factors contribute most to employees leaving the organization.
* I was given with **IBM HR Attrition dataset** containing employee demographic, performance, and work-related information.
* The challenge was to **clean, analyse, and visualize** the data to uncover insights that could help reduce attrition and improve retention.

**TASK**

My objective was to:

1. **Clean and prepare** the dataset using SQL and Python.
2. **Perform exploratory data analysis (EDA)** to identify key drivers of attrition.
3. **Build an interactive Power BI dashboard** to visualize insights clearly for decision-makers.

**ACTION**

I approached it systematically in three stages:

1. **SQL (Data Cleaning & Validation)**

* Imported the raw dataset into SQL Server.
* Standardized column names, corrected encoding errors (e.g., removing "ï»¿" from “Age”), and checked for duplicates and invalid values (Age < 18 or Monthly Income < 0).
* Removed columns that were irrelevant to data analysis (e.g., Standard Hours, Over 18, etc.).
* Ensured all numeric columns were in proper INT data types and text columns in TEXT.
* Verified no duplicates and confirmed consistent schema using DESCRIBE and COUNT(DISTINCT) checks.

**2. Python (EDA & Feature Engineering)**

* Loaded the cleaned data into **Pandas** for exploratory analysis.
* Checked for missing values, outliers, and distributions using **NumPy, Matplotlib, and Seaborn**.
* Created a new column Attrition\_num (1 for “Yes”, 0 for “No”) to support numeric aggregation.
* Identified top attrition drivers — **Over Time**, **Job Role**, **Years At Company**, and **Monthly Income**.
* Saved the processed data as Cleaned\_ibm\_hr\_attrition\_backup.csv for dashboard integration.

**3. Power BI (Visualization & Insights)**

* Imported the cleaned dataset and built an interactive dashboard titled **“IBM HR Analytics Dashboard.”**
* Added key KPIs:
  + **Total Employees:** 1470
  + **Attrition Count:** 237
  + **Attrition Rate:** 16.1%
  + **Average Monthly Income:** $6.5K
* Created meaningful visuals:
  + **Attrition by Monthly Income** (clustered bar chart showing peak attrition in lower-income ranges)
  + **Attrition by Job Role, Department, and Age Group**
  + **Attrition Rate by Work-Life Balance**
  + **Attrition by Over Time and Marital Status**
* Used calculated columns and DAX measures for accurate attrition percentages.
* Added slicers for gender, age group, department, job role, and marital status to make the dashboard fully interactive.
* Even though slicers were available for gender, age group, department, and other factors, I still created charts for the same categories. The reason was to give stakeholders a quick visual summary without needing to filter manually. Also, slicers are for filtering, but charts give an instant overall view. When someone selects a slicer, the related chart updates and shows the exact percentage or count, so it’s easier to compare before and after filtering.
* In short, Slicers are for **interaction** → “Show me only this group.”

Whereas, Charts are for **insight** → “Who is leaving more?”

**RESULT**

The final dashboard provided clear, actionable insights:

* **Over Time** employees showed significantly higher attrition (127 vs 110 non-overtime).
* Employees with **low income (1K–3K)** and **low work-life balance** had the highest attrition rates.
* Attrition was **most common among Sales Executives** and **younger employees (Under 30)**.
* Overall attrition rate: **16.1%**.

Demonstrated complete data analytics workflow — from raw data → cleaning → modelling → visualization.