**Project Description: Analyzing Excessive Absenteeism from Work**

**Introduction**  
This project focuses on analyzing the factors contributing to excessive absenteeism, defined as being absent from work for more than three hours. Understanding these factors is crucial for employers to identify patterns, mitigate risks, and enhance workforce productivity. The dataset contains various variables such as age, reasons for absence, transportation expenses, and family-related dynamics, which are explored to uncover insights into absentee behaviour.

**Project Workflow**

1. **Data Preprocessing**
   * Cleaning the dataset to handle missing values, inconsistencies, and outliers.
   * Encoding categorical variables for better compatibility with statistical and machine learning models.
   * Normalizing and scaling numerical features to improve model performance.
2. **Statistical Analysis**
   * Conducting exploratory data analysis (EDA) to identify trends and relationships between key variables.
   * Evaluating the distribution of absenteeism probabilities across demographics and job-related factors.
   * Testing for normality and significance to validate the impact of variables.
3. **Machine Learning**
   * Implementing predictive models to classify and estimate the probability of excessive absenteeism.
   * Comparing the performance of algorithms such as logistic regression, decision trees, and random forests.
   * Fine-tuning models for accuracy and generalization.
4. **Visualization on Tableau and Interpretation**
   * Creating interactive dashboards to showcase insights and highlight critical absenteeism patterns.
   * Visualizing factors such as age, reasons for absence, and transportation expenses to simplify communication of results.
   * Providing actionable recommendations based on visualization trends.

This comprehensive workflow ensures an in-depth analysis of excessive absenteeism, leveraging statistical rigor, predictive modeling, and impactful visualization to deliver meaningful insights.

**Files to be used**

1. Preprocessing.ipynb
2. Statistical\_Analysis\_and\_Machine\_Learning.ipynb
3. Visualization.twb
4. Absenteeism\_Integration.ipnyb : used to run machine learning model
5. Absenteeism\_module.py: module used to run machine learning model