Project Summary

1. Business Case

INX Future Inc. (INX), a global leader in data analytics and automation solutions for over 15 years, is renowned as a top employer worldwide. Despite this success, recent increases in employee attrition have raised concerns among senior management. These trends have coincided with a drop in client satisfaction scores by 8 percentage points and a rise in service complaints.

The leadership, cautious about impacting employee morale or company reputation, sought a data-driven approach to understand attrition's root causes. Under the guidance of CEO Mr. Brain, who has expertise in data science, the company embarked on a project to analyse employee data, identify key attrition drivers, and develop a predictive model to proactively manage retention and recruitment.

Project Objectives:

- Analyze attrition trends across different departments
- Identify the top 3 factors influencing employee attrition
- Build a predictive model to flag high-risk attrition cases for proactive action
- Provide actionable recommendations to reduce attrition and retain top talent

2. Requirements:

Employee data for INX was collected from a verified third-party source and is available in Excel format.

Dataset Link:

http://data.iabac.org/exam/p2/data/INX_Future_Inc_Employee_Performance_CDS_Project2_Data_V1.8

3. Data Analysis :

The dataset contains both numerical and categorical features. There are 19 numerical and 9 categorical columns. The target variable, Attrition, is binary (Yes/No), making this a classification problem.

Exploratory Data Analysis (EDA):
Univariate Insights:

- The average age of employees is around 36 years, with most employees falling between the 25-45 age group.
- A majority of the employees live close to the office, which may help with commute-related stress.
- Environment Satisfaction ratings of 3 are the most common.
- Most employees have limited prior job experience.
- The most frequent salary hike range is between 0-12%.
- Most employees have a moderate work-life balance rating.

Bivariate Insights:

- The Development department sees the lowest attrition, indicating better retention practices.
- Sales and Human Resources departments experience higher attrition compared to others.
- Employees in roles like Developer, Business Analyst, and Technical Lead tend to stay longer.
- Overtime workers are more likely to leave, indicating possible burnout.
- Male employees show a slightly lower attrition rate, potentially due to their larger presence in the workforce.
- Employees from Medical and Life Sciences backgrounds are more likely to stay.
- Employees who travel rarely are less likely to leave compared to frequent travelers.
- Attrition is higher when an employee has been with the same manager for too long, suggesting a need for managerial rotation.
- Employees who haven't been promoted in over 2 years are more likely to quit.
- Poor work-life balance and low relationship satisfaction strongly correlate with attrition.
- Employees with more than 10 years of total experience are at higher risk of leaving.
- A salary hike above 19% is a strong motivator to stay.
- High environment satisfaction is associated with better retention.
- Employees with low job satisfaction and involvement are more prone to attrition.

Additional Visualization (Multivariant Analysis) :

To better understand feature interactions, a Pair plot was generated showing relationships among Age, Hourly Rate, Years Since Last Promotion, and Attrition status. This visualization highlighted distinct clusters and trends correlating with attrition, aiding feature selection and model development.

4. Data Preprocessing :

- **Missing Values:** None found in the dataset
- **Encoding Categorical Data:** Applied Label Encoding to convert categories into numerical format
- **Outlier Treatment:** Used IQR and 3-sigma rule depending on data distribution
- **Feature Transformation:** Applied Square Root Transformation on skewed features such as YearsSinceLastPromotion
- **Feature Scaling:** Used StandardScaler to normalize numerical values

5. Feature Selection:

The dataset contains a wide range of features, but not all of them are equally important for predicting employee Attrition. To identify the most relevant features, we performed correlation analysis using correlation coefficients. This helped us quantify the strength of the relationship between each feature and the target variable "Attrition".

Based on the correlation values and domain relevance, the following features were selected as important contributors to employee attrition:

- Department
- Job Role
- Environment Satisfaction
- Last Salary Hike Percent
- WorkLife Balance

- Total Working Years (Experience)
- Years at Company
- Years in Current Role
- Years Since Last Promotion
- Years with Current Manager

These features demonstrated either a strong or moderate correlation with attrition (positive or negative) and thus were retained for model training.

Top 3 Factors Influencing Attrition:
- Environment Satisfaction
- Last Salary Hike Percent
- Work-Life Balance

These factors stood out with the most noticeable effect on whether an employee is likely to leave the company.

Feature Engineering & Preprocessing Steps:
- **Label Encoding:** Applied to categorical variables to convert them into numerical format suitable for modelling**.**
- **Standardization**: Used StandardScaler to bring numerical features to a common scale.
- **Feature Transformation:** Applied square root transformation to handle skewed features like Years Since Last Promotion**.**
- **Feature Dropping:**
  o Employee Number was dropped as it is a unique identifier and does not contribute to prediction.
  o Years Since Last Promotion was removed after transformation, as the derived feature was used instead.
  o

Correlation Heatmap:

A heatmap was created to visualize feature correlations. It showed no extreme multicollinearity, meaning that most features are independently informative.

Duplicate Check:

- No duplicate records were found in the dataset.

6. Machine Learning Model Creation & Evaluation:
Defining Features and Target:
- **Independent Features:** All selected numerical and encoded categorical features
- **Dependent Feature (Target):** PerformanceRating (2,3,4)

Handling Class Imbalance:

The target variable is imbalanced, To address this:

SMOTE (Synthetic Minority Oversampling Technique) was applied. It synthetically generates new examples of the minority class by interpolating between existing instances. This helps improve the model's ability to correctly predict attrition cases.

Train-Test Split:
- Data was split 75% for training and 25% for testing.

Algorithms Used:
- Logistic Regression
- Decision Tree Classifier
- Hyperparameter Tuned Decision Tree Classifier
- Random Forest Classifier
- Hyperparameter Tuned Random Forest Classifier
- Gradient Boosting
- Support Vector Machine (SVM)
- Artificial Neural Network (MLP Classifier)

Hyperparameter Tuning:

GridSearchCV was used to optimize key hyperparameters of the Decision Tree and Random Forest classifiers, including n_estimators, max_depth, and min_samples_leaf.

Model Results:

- **Logistic Regression:** Training Accuracy: 0.85, Testing Accuracy: 0.84
- **Decision Tree:** Training Accuracy: 1.00, Testing Accuracy: 0.93
- **Hyperparameter Tuned Decision Tree:** Training Accuracy: 0.98, Testing Accuracy: 0.90
- **Random Forest Classifier:** Training Accuracy: 1.00, Testing Accuracy: 0.97
- **Hyperparameter Tuned Random Forest Classifier:** Training Accuracy: 0.98, Testing Accuracy: 0.95
- **Gradient Boosting:** Training Accuracy: 1.00, Testing Accuracy: 0.96
- **SVC:** Training Accuracy: 0.90, Testing Accuracy: 0.89
- **Artificial Neural Network [MLP]:** Training Accuracy: 1.00, Testing Accuracy: 0.91

Model: The Hyperparameter Tuned Random Forest Classifier was selected as the final model due to its excellent balance of accuracy, generalization, and interpretability,Key for HR decision-making and actionable insights.

7. Recommendations to Reduce Employee Attrition:

Based on data insights and predictive modelling, the following recommendations aim to reduce attrition:
1. **Improve Work Environment:** Foster a more positive work culture, improve office facilities, and invest in team engagement activities.

2. **Offer Competitive Salary Hikes:** Employees receiving salary hikes below expectations are more likely to leave. A hike above 19% showed a strong retention effect.
3. **Promote Work-Life Balance:** Introduce flexible work hours, hybrid models, wellness programs, and mental health support initiatives.
4. **Manager Rotation :** Rotating managers every 2–3 years can improve employee engagement and prevent stagnation or dissatisfaction.
5. **Timely Promotions:** Consider promoting deserving employees every 1 to 2 years to reduce dissatisfaction due to career stagnation.
6. **Department-Specific Focus:** Focus on Sales and HR where attrition is highest..
7. **Support Female Employees:** Prioritize retention and recruitment where female attrition is low.
8. **Early Risk Identification:** Employees with low scores in Job Satisfaction, Work Involvement, or Relationship Satisfaction are more likely to leave. Create personalized retention plans for them.