Project Name: Employee Performance Analysis INX Future Inc

**March 10, 2024**

Code: 10281

Contents

[Business case](#_Toc160728271) 2

[problem statement 2](#_Toc160728272)

[Background 2](#_Toc160728273)

[Objective 2](#_Toc160728274)

[Exploratory Data Analysis 3](#_Toc160728275)

[Data Elements 3](#_Toc160728276)

[Analysis and Insights: Department wise performances 3](#_Toc160728277)

[Data Collection and Preparation](#_Toc160728278) 5

[Data Cleaning and Encoding](#_Toc160728279) 5

[Top 3 Important Factors effecting employee performance 5](#_Toc160728280)

[Model Development](#_Toc160728281) 6

ML [Model Result: 7](#_Toc160728282)

[Recommendations 8](#_Toc160728283)

[Conclusion 8](#_Toc160728284)

# Business Case

A data-driven exploration by INX Future Inc uncovers underlying factors and devises actionable strategies in response to declining employee performance metrics. The objective of this project is to elucidate departmental performances, identify top determinants of employee effectiveness, and propose targeted improvements to promote organizational excellence.

# Problem statement

INX Future Inc, referred to as INX, is a prominent data analytics and automation solutions provider with a global presence spanning over 15 years. Renowned for its employee-friendly policies and consistently ranked among the top 20 best employers in the past five years, INX has established itself as a leader in the industry. However, recent years have seen a decline in employee performance indexes, raising concerns among top management. This decline has been accompanied by increased escalations in service delivery and an 8% decrease in client satisfaction levels.

# Background

Despite its reputation as a top employer, INX faces challenges related to declining employee performance. CEO Mr. Brain recognizes the issues but is hesitant to take actions that could negatively impact employee morale and the company's reputation as a best employer. To address these concerns, Mr. Brain has initiated a data science project to analyze employee data and identify the underlying causes of performance issues. The project aims to provide actionable insights to help improve employee performance while minimizing adverse effects on morale and company perception.

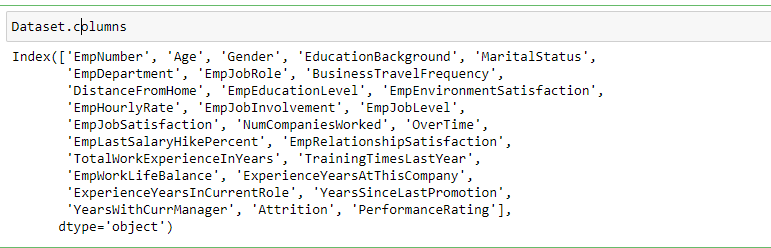
# Objective

The primary objectives of the project are as follows:

* Analyze department-wise performances to identify areas of improvement.
* Determine the top three factors influencing employee performance.
* Develop a trained model capable of predicting employee performance based on various factors.
* Provide recommendations for enhancing employee performance based on the analysis.

# Exploratory Data Analysis

## Dependent variables



In the exploratory data analysis (EDA), the handling of missing values is detailed. Furthermore, meticulous verification is conducted to ensure that the data types of features conform to the requirements of the analysis, ensuring accurate and reliable insights.

The Python program has been uploaded in Jupyter notebook format.

# Analysis and Insights: Department wise performances

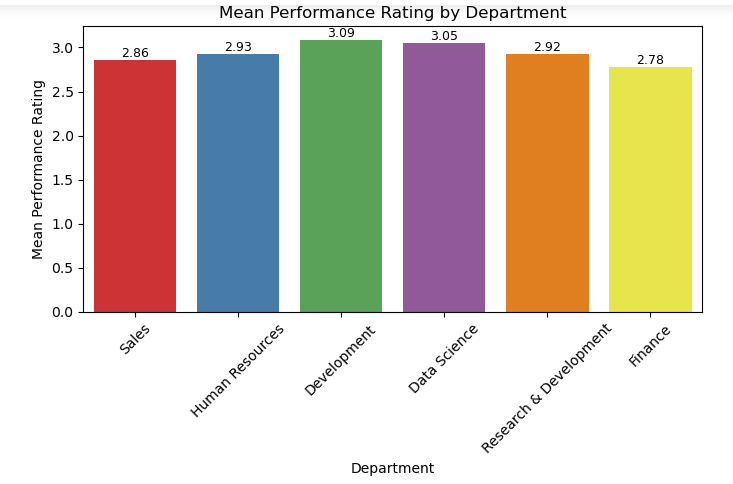
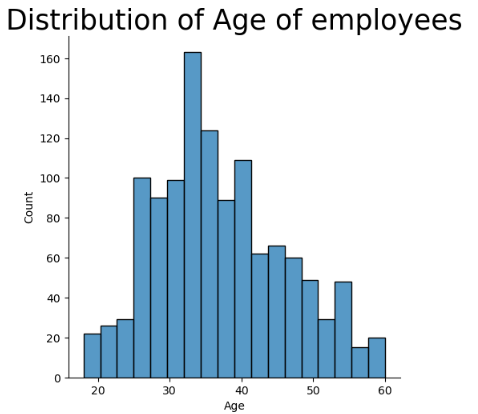
Analyzing departmental performance by utilizing descriptive statistics such as mean, standard deviation, and quartiles of performance ratings across different employee departments. By grouping the data by **'EmpDepartment' and 'PerformanceRating'**, insightful visualizations are generated to depict the distribution and comparative analysis of performance within each department.

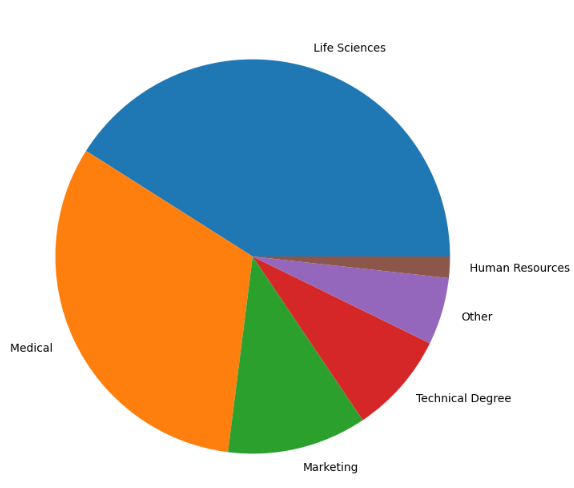
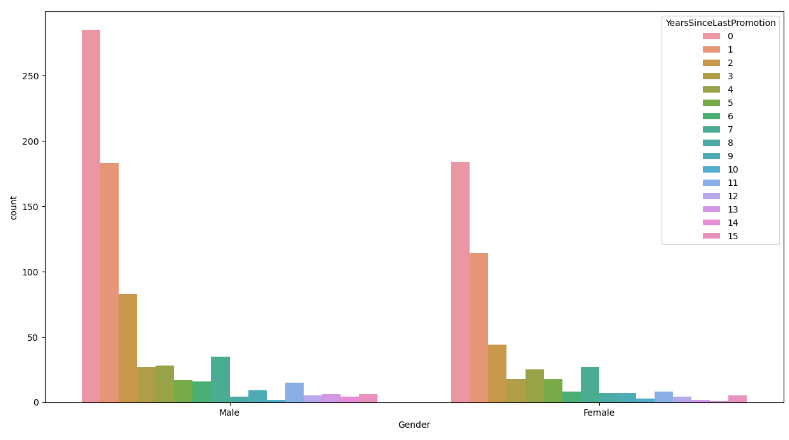
The Python program has been uploaded in Jupyter notebook format. Check the link or double click on file to access the project coding and visualization .

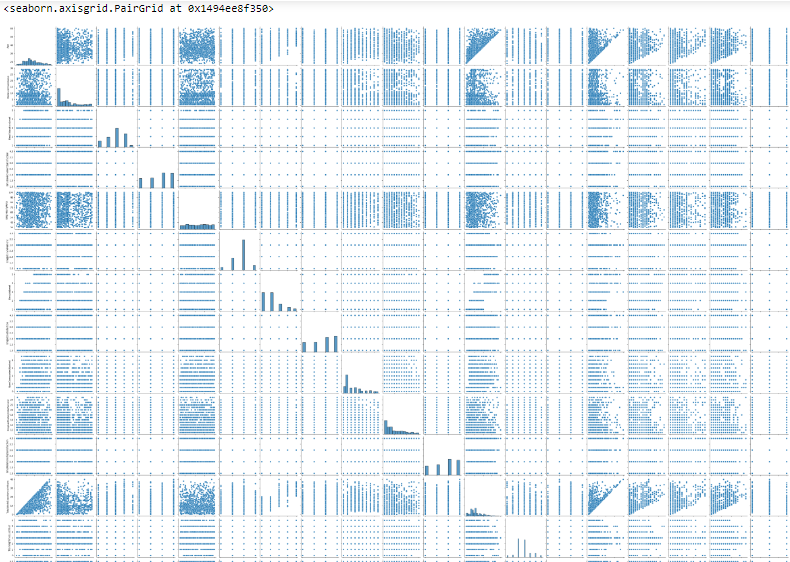
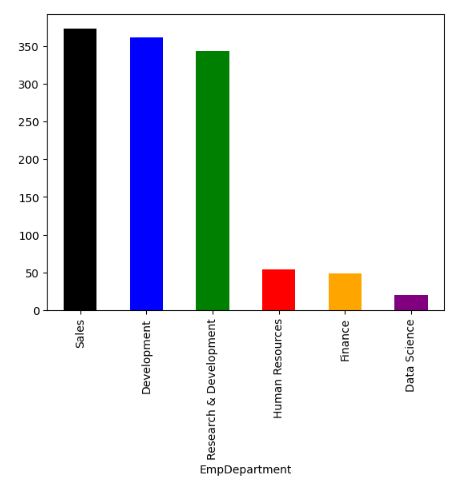
[**Employee performance analysis - Jupyter Notebook**](http://localhost:8889/notebooks/Employee%20performance%20analysis.ipynb)

****

**Visualization:**







# Data Collection and Preparation

The analysis utilizes a dataset containing employee data, including performance ratings and various factors potentially affecting performance. Data preprocessing steps include handling missing values, encoding categorical variables, and selecting relevant features for analysis.

# Data Cleaning and Encoding

The initial assessment of the raw data confirms its cleanliness, as no missing values were detected upon scrutiny. In preparation for accurate predictions, it is essential to perform label encoding on categorical values to ensure model compatibility and correlation plot.

The predictor variable encompasses numerous columns, prompting the calculation of correlation coefficients to pinpoint their significance. These coefficients guide the selection of features crucial for training models. Notably, key factors affecting performance include Department, Job Role, Environment Satisfaction, Last Salary Hike Percent, Work Life Balance, Experience Years At This Company, Experience Years In Current Role, Years Since Last Promotion, and Years With Current Manager, each exhibiting a correlation coefficient with Performance Rating exceeding 0.1. Feature transformation involved Standardization and Label Encoding.

Despite an exhaustive analysis involving all predictors, accuracy diminished. Similarly, Principal Component Analysis yielded reduced accuracy. The top three factors influencing employee performance are:

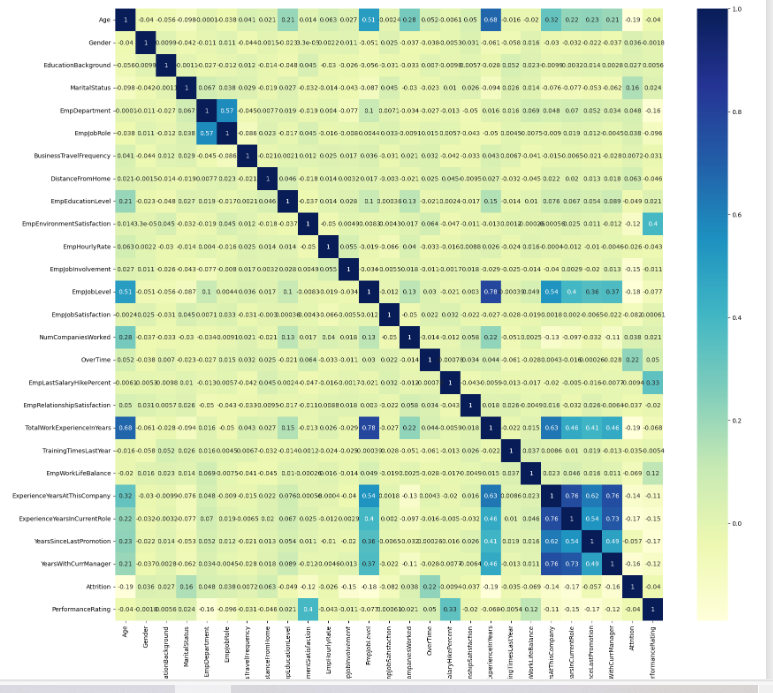
# Top 3 Important Factors effecting employee performance

## 1. Mean PerformanceRating with respect to EmpDepartment

.2. Employee Department with respect to gender.

3. Years since Last Promotion with respect to Gender

4. Employee Education level with respect to attrition.



# Model Selection

In the model development phase, various machine learning algorithms were employed to predict employee performance based on the provided dataset. Out of them I have selected The following models:

1. Random Forest with GridSearchCV

2. K-Nearest Neighbor (KNN)

3. Gradient Boosting

**Reason for selecting Model:**

By seeing the pairplot it is clear evident that linear regression, decision tree is not applicaple as the data points are randomly distributed.

Each model was trained and evaluated using the dataset , Random Forest with GridSearchCV, KNN, Gradient Boosting were assessed for their accuracy in predicting employee performance.

Training of each model involved fitting it to the training data and making predictions on the test data. The accuracy of each model was then calculated using the accuracy\_score metric. The accuracy scores were converted to percentages for better interpretation and visualization.

Finally, the accuracy scores of different models were plotted in a vertical bar plot to compare their performance. The top-performing model, as indicated by the highest accuracy score, was annotated for easy identification. This comprehensive approach allowed for the evaluation and selection of the most suitable model for predicting employee performance in the given context.

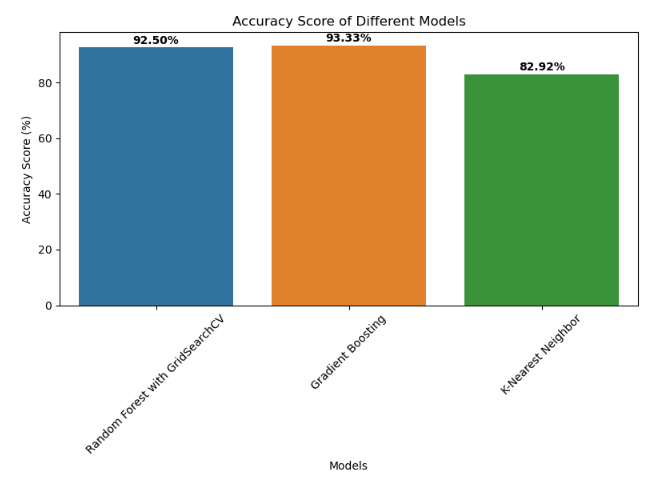
The Python program has been uploaded in Jupyter notebook format.

# ML Model Result:

After thorough evaluation and analysis of the machine learning models applied to predict employee performance, the final results showcase promising outcomes. Among the various models tested, the **Gradient Boosting** emerged as the top performer, achieving an impressive accuracy of 93.33%. This highlights the efficacy of ensemble learning techniques in accurately predicting employee performance based on the provided dataset.

Furthermore, both **KNN and Random forest with gridsearchcv** demonstrated strong performance, yielding an accuracy of ~82 AND 92%respectively. These results signify the robustness and versatility of **Gradient Boosting and Random forest with gridsearchcv**  in capturing complex patterns within the data and making accurate predictions.

Overall, the success of these models in predicting employee performance underscores the importance of employing sophisticated machine learning techniques in organizational decision-making processes. The high accuracy achieved by the **Gradient Boosting**.



# Recommendations

In summary, it is evident that enhancing the work environment significantly boosts performance. Moreover, regular salary increments and support for maintaining work-life balance are essential. Additionally, periodical rotation of managers can also positively impact performance.

# Conclusion:

In summary, our analysis underscores the effectiveness of ensemble learning techniques, particularly the **Gradient Boosting** model, which achieved an impressive accuracy of 93.33% in predicting employee performance. Key factors such as a conducive work environment, regular salary increments, and maintaining work-life balance emerged as critical drivers of employee performance.