

**INX FUTURE INC. EMPLOYEE PERFORMANCE PREDICTION MODEL**

**PROJECT SUMMARY**

In this project, I utilized a Random Forest Classification algorithm to predict employee performance ratings based on various factors. I chose the Random Forest due to its ability to handle complex datasets effectively by combining multiple decision trees.

For feature selection, I conducted an exploratory analysis of the dataset to identify influential variables. Key features considered included demographic information such as age, gender, education background, marital status, as well as job-related factors like department, job role, business travel frequency, and overtime. These features were selected based on their potential impact on employee performance, guided by domain knowledge and preliminary data analysis.

To prepare the data for model training, I employed label encoding for categorical variables. This transformation converts categorical variables into numerical format, facilitating processing by machine learning algorithms such as Random Forest.

Additionally, I utilized various techniques and tools throughout the project:

Exploratory Data Analysis (EDA): analyzing and visualizing data to understand its structure, patterns, and relationships. It helps identify trends, outliers, and potential insights that can guide further analysis and modeling decisions.

**Exploratory Data Analysis (EDA).**

- Show Dataset Info
- Number of Rows
- Column Names
- Data Types
- Missing Values
- Statistical Summary

Label Encoder: a technique to convert categorical variables into numerical format. It assigns a unique integer to each category in the variable. This transformation allows machine learning algorithms to handle categorical data as numerical data.

```

#encode the necessary columns
le = LabelEncoder()
df['Gender'] = le.fit_transform(df['Gender'])

le = LabelEncoder()
df['EducationBackground'] = le.fit_transform(df['EducationBackground'])

le = LabelEncoder()
df['MaritalStatus'] = le.fit_transform(df['MaritalStatus'])

le = LabelEncoder()
df['EmpDepartment'] = le.fit_transform(df['EmpDepartment'])

le = LabelEncoder()
df['EmpJobRole'] = le.fit_transform(df['EmpJobRole'])

le = LabelEncoder()
df['BusinessTravelFrequency'] = le.fit_transform(df['BusinessTravelFrequency'])

le = LabelEncoder()
df['Attrition'] = le.fit_transform(df['Attrition'])

le = LabelEncoder()
df['EmpNumber'] = le.fit_transform(df['EmpNumber'])

le = LabelEncoder()
df['OverTime'] = le.fit_transform(df['OverTime'])

```

GridSearchCV: a technique for hyperparameter tuning in machine learning models. It exhaustively searches through a specified grid of hyperparameters and evaluates the model's performance using cross-validation to find the best combination of hyperparameters.

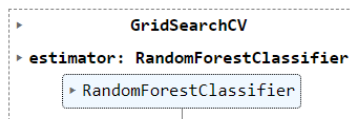
#### Tune & Boost My Model

✓ 0s [24] from sklearn.model\_selection import GridSearchCV

✓ 0s [25] #Define the parameter grid  
 param\_grid = {  
     'n\_estimators': [100, 200, 300],  
     'max\_depth': [None, 10, 20],  
     'min\_samples\_split': [2, 5, 10],  
     'min\_samples\_leaf': [1, 2, 4]  
 }

✓ 0s [26] #Initialize the GridSearchCV object  
 grid\_search = GridSearchCV(estimator=RandomForestClassifier(random\_state=42), param\_grid=param\_grid, cv=5, n\_jobs=-1)

✓ 2m [27] #Fit the GridSearchCV object to the data  
 grid\_search.fit(x\_train, y\_train)



Streamlit: a popular Python library to create an interactive web application that allows the users to visualize the data in question, evaluate the model performance, and user input for prediction.

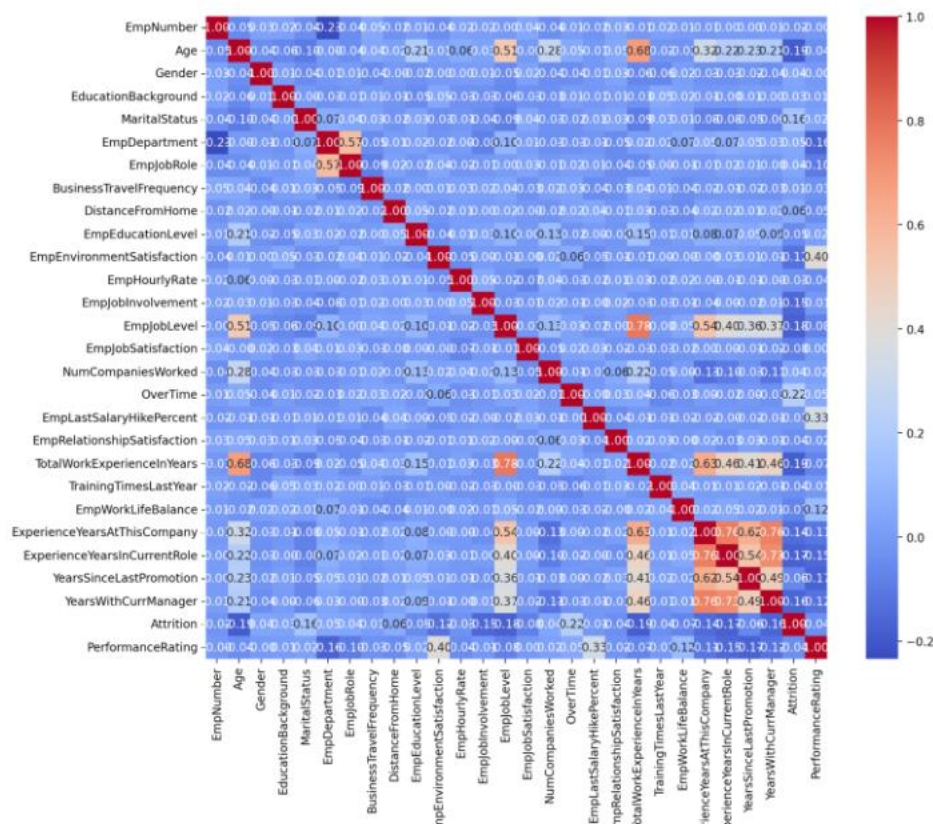
```
import streamlit as st
```

## Visualization of the Dataset.

Correlation Matrix Heatmap: Calculate the correlation coefficients between all pairs of variables and visualize them as a heatmap. This allows you to quickly identify strong correlations (positive or negative) between variables.

☒ Correlation Heatmap

Correlation Heatmap



Google Colab: Google Colab is a cloud-based platform provided by Google that allows users to write and execute Python code in a browser environment. It provides access to powerful hardware resources (such as GPUs and TPUs) and collaboration features, making it popular among data scientists and machine learning practitioners.

```
INX_Future_Inc_Employee_Performance_Prediction_Model.ipynb
File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

[1] #import the necessary libraries
import pandas as pd
import numpy as np

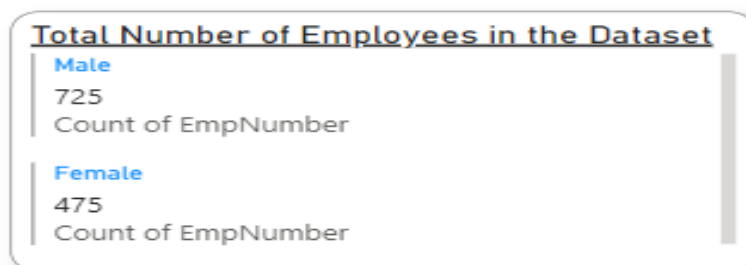
[2] # Load the data
url = ("http://data.iabac.org/exam/p2/data/INX_Future_Inc_Employee_Performance_CDS_Project2_Data_V1.8.xls")
df = pd.read_excel(url)
```

The combined use of the Random Forest algorithm, careful feature selection, and additional techniques and tools enabled the development of an effective model for predicting employee performance ratings with accuracy.

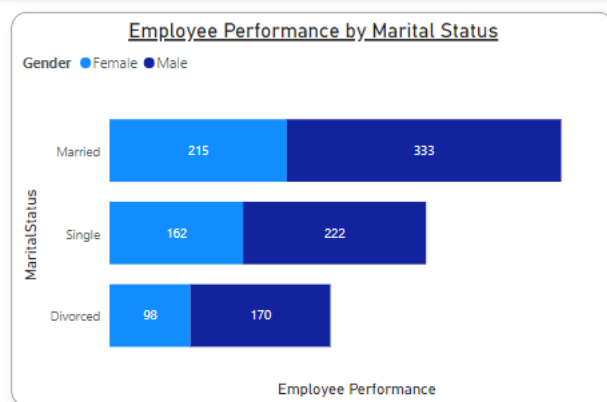
### OBSERVATIONS AND INSIGHTS

From the dataset, I explored the relationship the features and how they affect employee performance. Below, are some of the insights I gathered, with pictorial visualizations to illustrate the same:

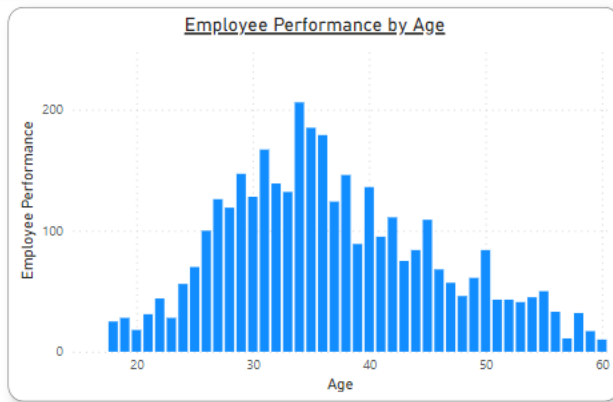
#### Total Number of Employees in the Dataset:



**Marital Status:** Divorced employees exhibit significantly lower performance compared to others. Marital status may affect performance due to personal factors such as stress or work-life balance issues.



**Age:** Employees between the ages of 26 and 45 tend to have the highest performance compared to younger and older employees. This suggests that experience and maturity positively impact performance.



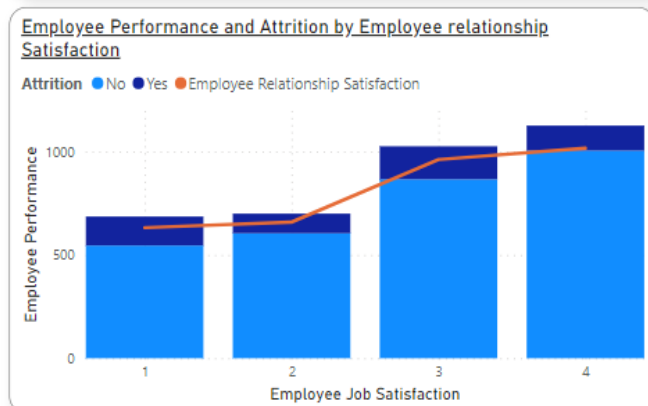
**Employee Job Satisfaction:** Higher job satisfaction is associated with higher performance. Satisfied employees are likely to be more motivated, committed, and productive

**Attrition:** Lower attrition rates are associated with higher employee performance. High turnover can disrupt workflows and morale, negatively impacting performance.

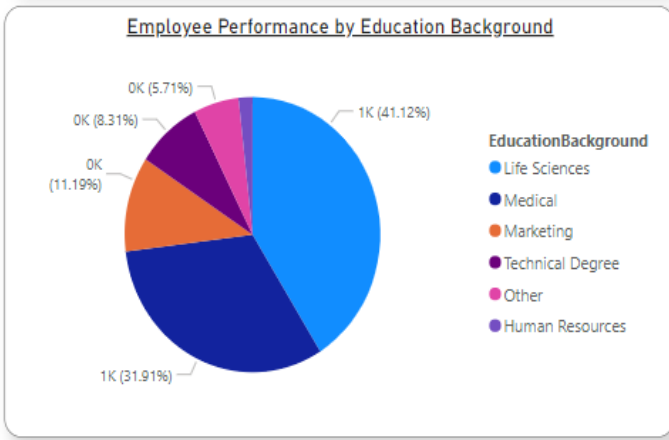
**Employee Relationship Satisfaction:** Higher relationship satisfaction correlates with higher performance. Positive relationships with colleagues and managers contribute to a supportive work environment.

**Employee Environment Satisfaction:** There is a positive relationship between employee environment satisfaction and performance. A positive work environment fosters motivation and engagement, leading to better performance.

**Years Since Last Promotion** and **Years With Current Manager:** Both variables show significant relationships with performance. Positive experiences such as promotions and stable management contribute to employee satisfaction and performance.

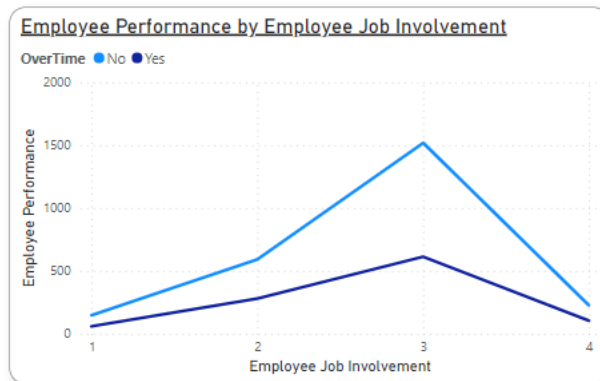


**Education Background:** Employees with education backgrounds in life sciences and the medical field, and taking up roles in the respective departments tend to have higher performance. This may be due to the nature of their jobs.



**Employee Job Involvement:** Moderately high job involvement correlates with higher performance. Employees who are engaged and invested in their work tend to perform better.

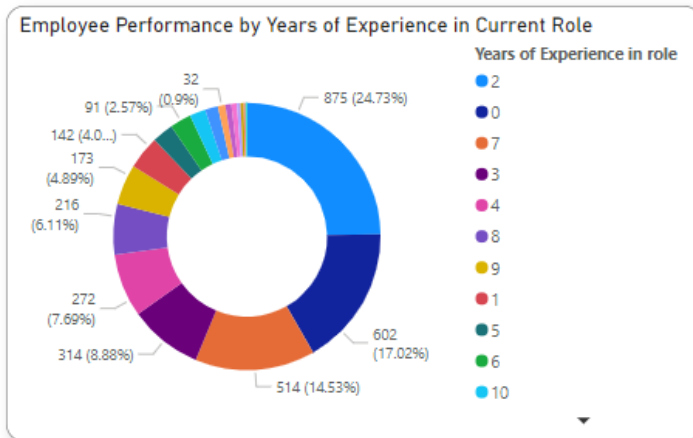
**Over Time:** Employees who do not work overtime tend to perform better. This suggests that excessive work hours may lead to burnout or decreased productivity.



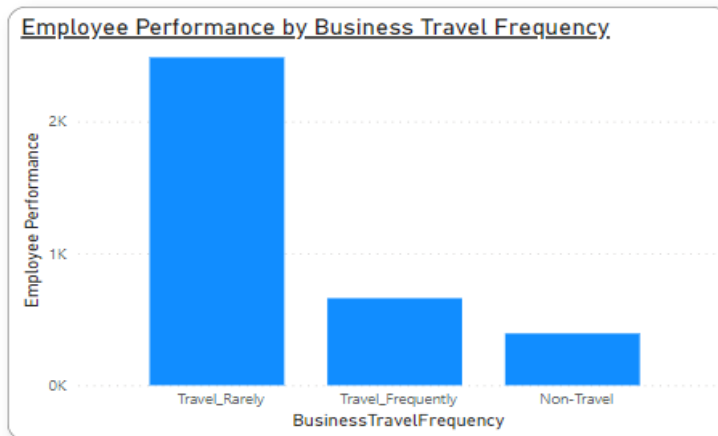
**Total Work Experience In Years:** There is a weak negative relationship between total work experience and performance. This suggests that extensive experience may not always translate to higher performance.

**Employee Hourly Rate:** No significant relationship was observed between hourly rate and performance. This suggests that compensation alone may not be a strong determinant of employee performance.

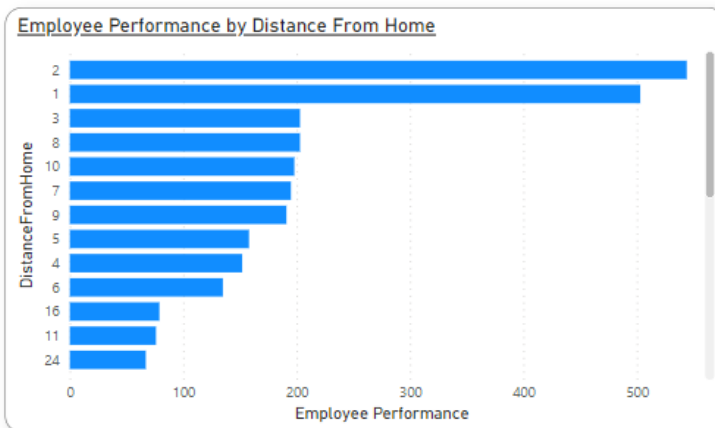
**Experience Years at This Company** and **Experience Years in Current Role:** Both variables show a strong negative relationship with performance. This could indicate that prolonged tenure in the same role or company may lead to stagnation or decreased motivation.



**Business Travel Frequency:** Employees who travel rarely tend to have significantly higher performance. Extensive travel may lead to fatigue or disruptions in work routines, affecting performance.



**Distance From Home:** Employees who live closer to the workplace tend to have higher performance. Commuting distance may impact punctuality, energy levels, and overall job satisfaction.



**Training Times Last Year:** No significant relationship was observed between training frequency and performance. This could indicate that the effectiveness of training programs may vary.

**Employee Last Salary Hike Percent:** A higher salary hike is associated with lower performance. This finding is unexpected and may warrant further investigation into potential underlying factors.

**Employee Work-Life Balance:** Moderately high work-life balance is associated with higher performance. A healthy balance between work and personal life promotes well-being and job satisfaction.

#### **Recommendations:**

**Tailored Training Programs:** Develop tailored training and development programs based on employees' education backgrounds and job roles. Providing relevant and targeted training opportunities can enhance employees' skills and competencies, leading to improved performance.

**Employee Engagement Initiatives:** Foster a positive work environment through employee engagement initiatives, such as team-building activities, recognition programs, and regular feedback sessions. Engaged employees are more likely to be motivated, committed, and productive, leading to higher performance levels.

**Career Development Opportunities:** Provide employees with opportunities for career growth and advancement within the organization. Offering clear career paths, mentorship programs, and opportunities for skill development can increase employee engagement and performance.

**Effective Communication Channels:** Establish open and transparent communication channels between employees and management. Encouraging regular communication, feedback, and dialogue can enhance collaboration, resolve conflicts, and improve overall performance.

**Recognition and Rewards:** Implement a comprehensive recognition and rewards program to acknowledge and celebrate employee achievements. Recognizing employees' contributions and efforts can boost morale, motivation, and performance.

**Performance Management Systems:** Implement robust performance management systems to set clear performance expectations, provide regular feedback, and track employee progress. Performance reviews should be fair, objective, and aligned with organizational goals to drive continuous improvement.

**Diversity and Inclusion Initiatives:** Promote diversity and inclusion within the workplace to foster a culture of belonging and respect. Embracing diversity of thought, backgrounds, and perspectives can lead to innovation, creativity, and improved performance.

**Workload Management:** Monitor and manage employees' workloads to prevent burnout and ensure a healthy work-life balance. Providing adequate resources, support, and assistance can help employees manage their responsibilities effectively and maintain high performance levels.

**Continuous Improvement Culture:** Cultivate a culture of continuous improvement where employees are encouraged to identify areas for growth, learn from failures, and innovate. Encouraging a growth mindset and embracing change can drive organizational excellence and sustained high performance.

By implementing these recommendations, organizations can create a supportive and empowering work environment that enables employees to thrive, contribute their best efforts, and achieve their full potential.