Project Summary:

Overall Summary:

- The project is about analyzing the features affecting the performance of the employee. Based on it, we are asked to develop a model for predicting the performance of an employee, as a rating of his overall performance in 2-4 scale.
- The data was given in an excel file. Based on basic checks, the data is observed to be clean
- Exploratory data analysis is conducted on the data. We observed the spread of each of the data features, and the effect it had on the overall performance of an employee.
- As a part of Task 1, Based on the analysis, insights were presented.
- As a part of Task2, Based on exploratory data analysis, as asked, department wise employee performance analysis were made.
- Next, as a part of Task3, a model that evaluates the performance of an employee and gives the possible rating of his performance based on the features is developed. This is a classification problem with 3 output classes (Multi-class classification problem).
- Based on correlation between features, few features seemed redundant. So, setting a threshold of 0.6 for correlation coefficient, few features are filtered out.
- We have used various classification algorithms like multi class logistic regression, KNN, decision tree, Random Forest, Gradient Boosting and XG Boosting algorithms to design the model. Based on the f1 scores yielded by the models, the best one is picked out.
- The data appeared imbalanced, and two different techniques of balancing the data, viz, SMOTE of imblearn library and minority resampling of sklearn library are chosen.
- Upon training and testing the model and evaluating the results, we figured out Random Forest Algorithm to be yielding best results in this case, with imbalanced data,, with an F1 score of 0.89. Hyperparameter tuning gave us slightly improved results, but the improvisation wasn't much. With the balanced data, Random Forest Algorithm gave similar results.
- The results of various models are presented in the src file.
- As a part of Task 4, the recommendations for improving the performance of employees are presented, based on the analysis of features of outperforming employees.

TASK WISE SUMMARY:

Task1: Department wise performances

- From the pie charts and data, it is evident that, development department and data science department perform similarly.
    - 84-85% of employees have a performance rating of 3, which is average. 10-12% of employees are high performers, with an overall performance rating of 4. 3.5-5% of employees are underperformers, going by their rating of 2.
- Human Resources and Research & Development departments are similar.
    - About 68-70% of employees in these departments have an overall performance rating of 3, About 11-12% of them have a rating of 4. About 18.5-20% of them have a rating of 2, which is the least performance.

- Sales department has 67% of employees rated at 3 (average), 23% rated at 2 (under performance) and only 9.4% rated at 4 (better performance).
- Finance department has 61% of average performers with rating of 3, 30.6% of employees with rating of 2, and 8% of employees with a rating of 4.
- Concluding,
    - Finance department has the most underperformers (rating 2), and least proportion of best performers(rating 4).
    - All other departments have 9.4-12% of best performers (rating 4).
    - However Development and data science have least number of underperformers with rating 2 (a proportion of only 3.6-5%)
    - Sales, HR and R&D have 18.5-23.3% of under performers (rating 2)

Task 2: Top 3 Important Factors affecting employee performance

Insights and Conclusions on factors affecting employee performance from EDA:

- Younger aged employees or the most aged ones are performing well. Younger employees have a higher proportion of either under or better performers than average performers. This can be attributed to age related aggression. As age increases, while the number of better performers have increased (at a senior level), average performers have been more in number. This seems to be the effect of midlife crisis (more responsibilities on personal front, as well as work front)
- Employees who had better job environment satisfaction and job satisfaction had a higher proportion of better performers. So, in other words, better environment, and job role, can be a factor that affects performance.
- Those who haven't worked in any company before had a higher proportion of average performers, while those who have worked in a company before had a higher proportion of better performers. Again, the performance is more of an average in those who have been hopping across companies.
- Higher salary hike employees have a higher proportion of better performers.
- Those who have experience within 10 years, have a higher proportion of better performers.
- Those who have been promoted newly, and have been for less than an year in the current job role or under the current manager have a higher proportion of better performers.
- We may attribute that higher performances could be related to the enthusiasm of promotion or newer challenges that the new job role or new manager offers. As the number of years pass, time induced or work induced monotony could have been affecting the performance.

  Therefore we could conclude that job satisfaction and work environment satisfaction, salary hike percentage, time induced or work induced monotony (increased years in the same job role or under the same manager) are the top factors that affect employee performance rating. Other factors like experience and age also have a significant impact.¶

TASK 3: MODEL TO EVALUATE AN EMPLOYEE PERFORMANCE

Conclusions from modeling

- Random Forest has turned out to be the best model with F1 score of 0.89 for unbalanced data and similar results for resampled balanced data. On hyper parameter tuning, it further improvised the F1 score, but with each run, the results haven't been consistent (I could made sure same results come each time, but this was a way i could check for better parameters)
- The results of various models have been presented in the dataframe Results

TASK 4: Recommendations to improve the employee performance based on insights from analysis.

- Higher proportion of underperformers or average performers is seen in those who have been in same role for more number of years, under same manager for more number of years. This may be interpreted as a kind of stagnation induced work monotony. Hence, occasional change of work, change of roles, might ensure the employee isn't finding job monotonous and deliver better performance.
- As years of experience in the job role increases, the proportion of better performers is coming down. This can be avoided, by using seniors into managerial roles, and new enthusiastic workers into job roles.
- Midlife crisis seems to be playing a role in overall performance as we see relatively lesser proportion of better performers in 35-45 aged people. We may address this issue by ensuring right work life balance for that segment of employees. It was also evident that better work life balance employees delivered better performance
- Employees who had better job environment satisfaction and job satisfaction had higher proportion of better performers. So, in other words, better environment, and job role, can be a factor that affects performance.
- Those who haven't worked in any company before had higher proportion of average performers, while those who have worked in a company before had higher proportion of better performers. Again, the performance is more of an average in those who have been hopping across companies. So it may turn out to be beneficial to hire those who worked in a company before rather than those who haven't worked anywhere before. Definitely, hoppers aren't great choice.
- Higher salary hikes and quicker promotions may promote the employee's performance, but subject to a matter of relevance. Since experienced (>10 years) employees have higher proportion of better performers, we my consider higher salary hikes and quicker promotions to experienced people to motivate their performance.