Notes : After Making EDA

HR (IBM) Project : Task03 – Meriskills

- After doing Data collection and Data prepare data is Cleaned
- By using EDA I found some results : first = correlations:

- Ages have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Good : In range [0.2 – 0.3] with columns (YearWithCurrentManager , YearSinceLastPromotion , YearInCurrentRole , YearsAtCompany , numberCompanyWorked , Education)
  - Strong: in Range [0.5 – 0.7] with Columns (TotalWorkingYears, MonthlyIncome, JopLevel)
- JopLevel have good or Strong relationship: good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Good : with columns (YearWithCurrentManager , YearSinceLastPromotion , YearInCurrentRole)
  - Strong : with columns (YearsAtCompany , TotalWorkingYears , MonthlyIncome (95%))
- MonthlyIncome have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Good : (YearWithCurrentManager , YearSinceLastPromotion , YearInCurrentRole)
  - Strong : with Columns (YearsAtCompany, TotalWorkingYear)
- PercentSalaryHeky have Strong relationship with PerformaceRating
- TotalWorkingYear have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Good : columns (YearWithCurrentManager , YearSinceLastPromotion , YearInCurrentRole , numberCompanyWorked)
  - Strong : with columns (YearsAtCompany)
- YearsAtCompany : have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Strong : with columns (YearWithCurrentManager , YearSinceLastPromotion , YearInCurrentRole)
- YearsInCurrentRole : have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Strong : with Columns (YearWithCurrentManager , YearSinceLastPromotion)
- YearsSinceLastPromotions : have good or Strong relationship : good [0.2 – 0.4] , strong [0.5 – 1.0]
  - Strong : with Columns (YearWithCurrentManager)

****

Second = Business Questions i asked to tring discovering some Insights :

1- How does the daily rate vary for employees who work overtime compared to those who don't ?
   - We found number of employes That Not Worked OverTime Have High Daily Rate by rate : 71.49% , number employees that have low dailyRate by rate : 28,51%
2- Do married employees tend to have higher or lower salaries compared to their single or divorced counterparts ?
   - I found the Married and Divorced employees have the same Rate of monthlyIncome by rate : 34.9 % , but in single have low of them by rate : 30.2% with difference 4.7%
3- Does the percentage of salary hike differ for employees with Jop roles and different marital statuses ?
   - No we found SalaryHike Not really differ between JopRoles and marital statuses
4- How does job satisfaction vary across different job roles ?

- I found the most role have high number of jop satisfactions [Sales Executive - Research scientists]
- But we found the average jop satisfactions in all JopRoles was nearly : 2.5 – 3 where the most employees give rate 2.5 these mean they have a problem with their JopRoles

5- Some Notes :
   - Some reasons that maybe cause emp left companys [distanceFromHome , Environment Satisfaction , Jop Satisfaction , monthlyIncome , worklifeBalance , YearLastPromotion ,
6- Remember DataSet isnot Imbalanced as I found more that 1200 emp not left company , the number of emp that left company was 200 there are a bias or imbalanced

Machine Learning Stage :

- I found my data is not balanced so I used a specific tec to make it balanced by using (RandomOverFitting)
- Now Data is Balanced we want to convert Categorical data to numeric data so we use Labal encoder
- After that I use Logistic Regression to classify but I found the Accuracy 80% but is not the best one so I used Feature Engineering and used a specific Tec called (Random Forest Classifier) this tec help me choose the importance features that will make my model better
- So after I using it I get a best accuracy (1.00)

```
print(classification_report(y_test,y_pred_selected))
```
✓ 0.0s

	precision	recall	f1-score	support
0	1.00	1.00	1.00	370
1	1.00	1.00	1.00	370
accuracy			1.00	740
macro avg	1.00	1.00	1.00	740
weighted avg	1.00	1.00	1.00	740

Thank You.@