**Steps while working on Projects:-**

1) Explain the requirement of hr analytics and why we are working on it

(note:- info as mention in the dataset)

2) News of IBM that tells about the kyndryl unit formation which is mention in the dataset

3)contents explaination that are mention and available in the Project

* Importing dataset, taking info of data,
* Finding out unique values from dataset and removing those columns having single value mentiones
* Checking % values by crosstab and finding out percentage of employee resigning
* splitting the columns as per datatype, and find out count of values of each column, and description of numeric columns.
* Splitting object column into dummies where 1 shows availability and 0 shoes non availability.
* Concatenate object and numeric clumn and then find the count of attrition value
* Values of yes is less than no so upsample the data so that data having yes will repeat its values till the count of no is reached.
* Now Visualisation:- starts where we have to mention following statements as result
* Prediction 1 :- From the above analysis we found that attrition is maximum in the age of 26 to 35. And the attrition rate keep decreasing adter age of 35 as people look for stability after some point of age.
* Prediction 2 :- In the given plot Attrition is high for those who have Salary<6000. It decreases as Salary is increasing. Hence we can say that Salary would be the first reason for Attrition
* Prediction 3:- Here We get that count of **Single** people resigning is more than other **Divorced** and **Married**.
* Prediction 4:- Hence from this we see clear impact of Business travel on Attrition that Employees who travel rarely are leaving company than other employees.This may be because of employee allowance for travelling or Travelling may be hectic.
* Prediction 5:- Hence from this the tendency to leave the company is more to those employee who have limited Stock availing Option level in Company. Since stock constitute huge amount of money while staying for few years, people donot want to loose the opportunity. People with very limited or no stock options have the freedom to leave the organization at will
* Prediction 6:- Hence the people with poor level of Work life balance have adjusted in their job but but as seen from above parameters those who have normal level of work life balance are not adjusting themselves thats why they are searching new platform.
* Prediction 7:- Hence It is predicted that those students from **Research and Development** department are leaving more than those of Other Department.
* Prediction 8:- Hence It is predicted that Employees from **LifeSciences** and **Medical** have max no of count than compares to other education field
* Prediction 9:- This Tells that attrition of male is more than female.
* Prediction 10:- This tells that People from **Laboratory Technician, Research Scientist, Sales Executive** are leaving their job than compare to other jobroles
* Prediction 11:- Here Employee having Overtime are changing job by attrition than those who are not having OverTime.
* Set correlation between attributes by corr matrix
* Lets start now ML modeling
* **Logistic Regression Modeling:-**

1. Splitting columns in train test by 70-30 partition
2. Scale columns by min max scaler with formula (X-Xmin)/(Xmax-Xmin)
3. Fitting the model to statistics model and take out the summary
4. Hence by GLM we interpret the coefficient value of model by Generalised Linear regression model also find the errors as
5. <https://www.statsmodels.org/devel/glm.html>, <https://www.mygreatlearning.com/blog/generalized-linear-models/>

<https://www.datasciencecentral.com/profiles/blogs/explaining-logistic-regression-as-generalized-linear-model-in-use/>

1. Use of recursive Feature elimination for eliminating the features not required

<https://machinelearningmastery.com/feature-selection-in-python-with-scikit-learn/>

<https://machinelearningmastery.com/rfe-feature-selection-in-python/>

<https://www.scikit-yb.org/en/latest/api/model_selection/rfecv.html>

7. model fitting, VIF calculation, confusion matrix, ROC- AUC curves, finding optimal cut off, finding probability of y\_pred by x\_test\_sm and then evaluating the score

* **Decision Tree Algorithm:-**

1.) Splitting of column in train and test by 70\_30 partition

2) running decision Tree classifier algorithm

3)predicting classification report, confusion matrix and accuracy score

4) plotting tree with gini index value

5) Grid search + cross validation done on dataset by max depth

6) Grid search + cross validation done on dataset by min no of leaves

7) Grid search + cross validation done on dataset by min no of sample split

8)finding best estimator and best accuracy

9) now again predicted classification report, confusion matrix and accuracy score

* **Random Forest Algorithm:-**

1)