1. import all packages

2. import data

3. remove duplicates

4. check data(head,shape,info,describe)

5. identify Target variable and check distribution (plot the graph,value\_count,heatmap for null values,bar plot,

histogram, kde plot etc)

6. treatment of outliers and null values

7. check all the categorical and numerical columns

8. check for null values and do treatement of null values

9. univariate analysis of categorical columns(pi chart, box plot)

10. univariate analysis of numerical columns(histogram,kde)

11. bivariate analysis wrt Target column:

\* numerical - numerical : scatter plot, corelation

\* numerical - categorical : boxplot, Anova

\* categorical - categorical : crossplot, chi-square

12. feature engineering

13. convert categorical to numerical (use dictionary or one\_hot encoding or other techniques)

14. feature selection

15. standardization or normalization

16. train\_test\_split

17. run the various models

18. perform cross validations of all the model (hyperparameter tunning: Greedserchcv,Random search, StratifiedKFold)

Ensembaling

19. use acuracy matrics as mape, mse, rmse for numerical column

20. use f1-score and confusion matrix for categorical column

21. finalise model

22. Create Pipeline of the model

23. run the model on entire data

24. model ready for deployment

1. hypthesis generation

2. teting of hypothesis

3. conclude the hypothesis

4. based on hypothesis, can perform bivariate analysis between all the variables