* For regression, use linear regression cz only algorithm in portions
* For classification:

1. KNN: if dataset is numerical, need to scale data for sure (no model only pred)
2. DT: if dataset is mostly categorical
3. Log Reg: categorical and satisfies the collinearity conditions of Lin Reg too
4. SVM: if dataset is small and complex
5. Naïve bayes: for multiclass problems, lot of categorical variables, features must be independent- less correlation (no model only pred)
6. Find problem domain
7. Imports
8. Load data, head, remove unique col
9. Info, describe
10. Preprocessing
    1. Null values
    2. Balanced dataset
    3. Multi collinearity (matrix, scatter plot)
    4. Encoding
    5. Scaling
    6. Test train split
11. Visualization (apply PCA to make it 2 dimensional if needed)
12. Algorithm selection and implementation with proper explanation
13. Model evaluation
    1. Classification: report, AUC, cross validation, under/overfitting
    2. Regression: MSE, R2, plot the line or smtg
14. Interpretation: for regression use the equation to interpret for classification based on evaluation and model parameters

Clustering:

* No of clusters given use k means
* Not given => hierarchical
* Elbow method or silhouette coefficient for KNN or hierarchical
* Dendrogram for hierarchical