1. Understanding data

* Data shape
* Checking for N/As
* Data types & categorical data levels
* Checking for unbalanced data (esp. in target variable)
* Checking for outliers as needed and replace with NaN

1. Univariate Analysis

* Target variable distribution

1. Correlations and distribution

* Correlations between target and features (corr map)
* Categorical – countplot with hue = target, independent on x, count on y + multiindex groupby then include % split
* Continuous – boxplot with x = target
* Multicollinearity (corr map)

1. Transform data

* Group continuous (eg. Age) into categorical groups as needed
* Dealing with ordered categorical data (LabelEncoder)
* Dealing with unordered categorical data (OneHotEncoder)

1. Model fitting + Feature importance (Precision, Recall, ROC, Feature Importances)

* Logistic Regression with SMOTE
* Random Forest
* Linear SVM