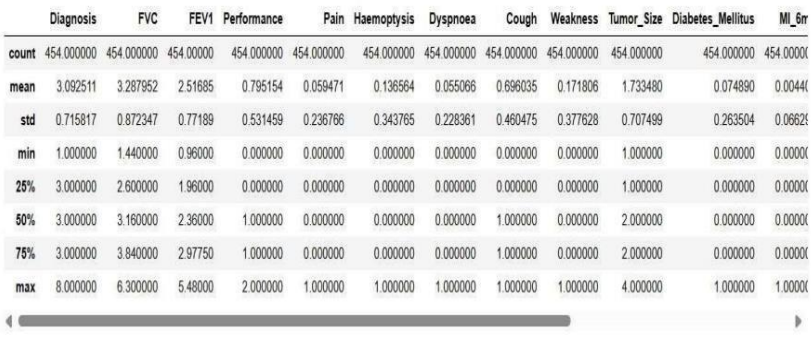
Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	739706
Project Title	One Year Life Expectancy post on Thoracic Surgery using Machine Learning
Maximum Marks	6 Marks

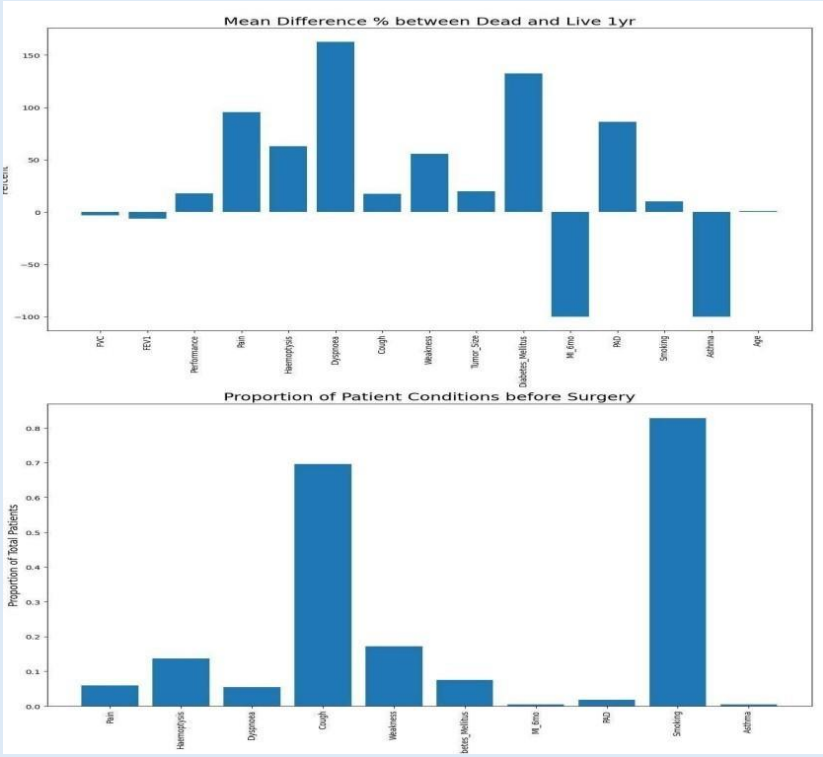
Data Exploration and Preprocessing Report

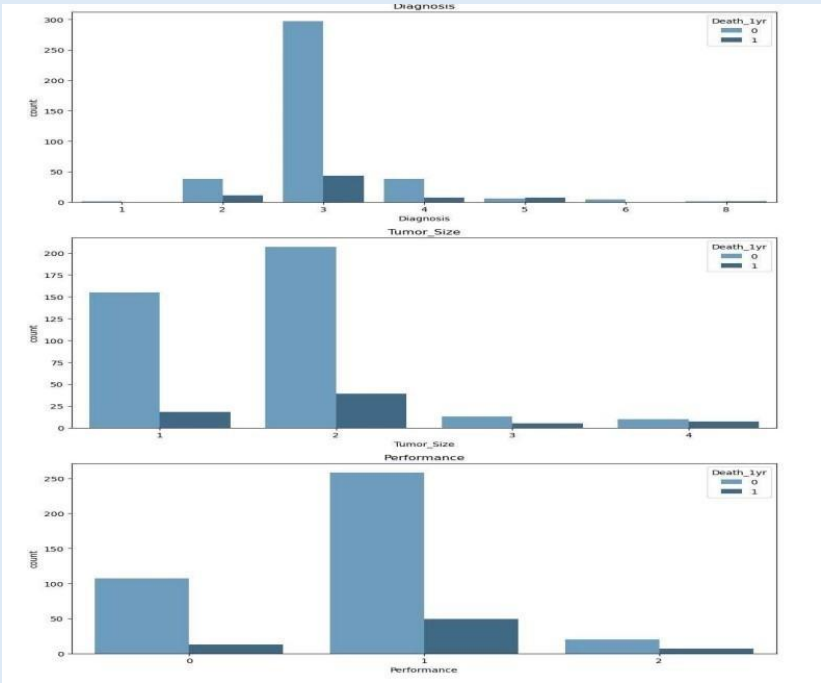
Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for pre-processing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

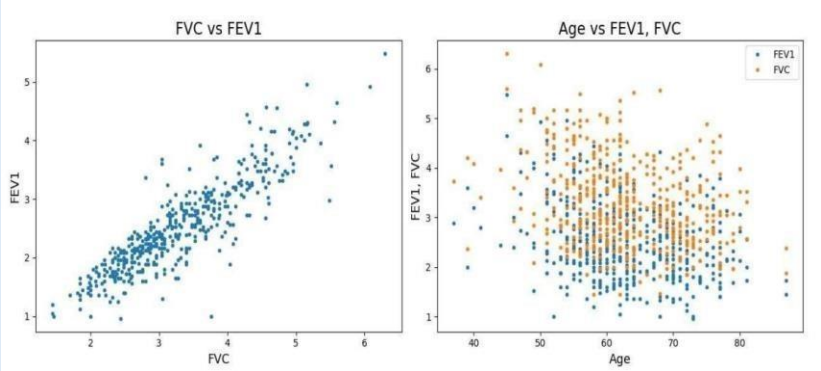
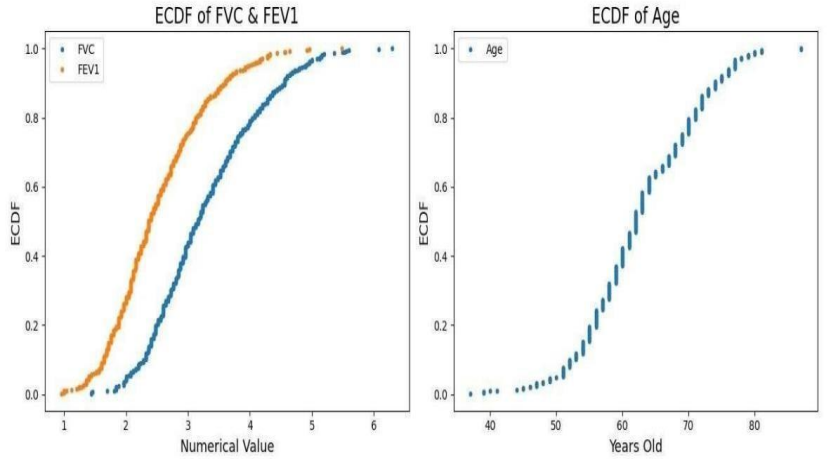
Section	Description
Data Overview	Dimension: 454 rows × 17 columns Descriptive statistics: 

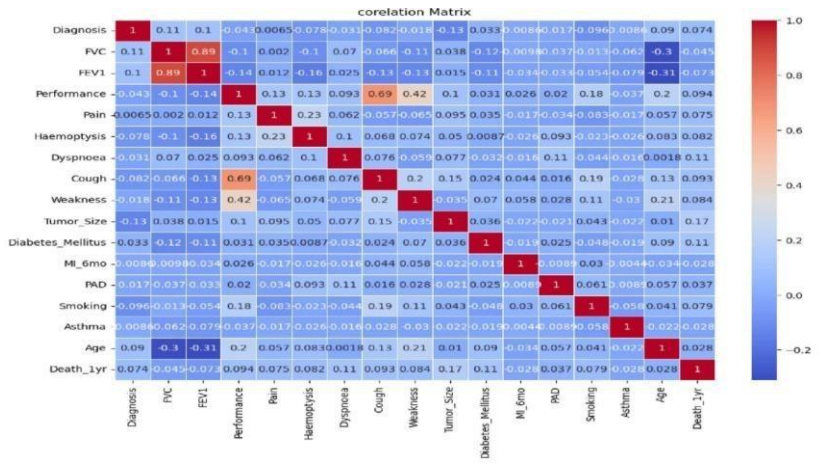
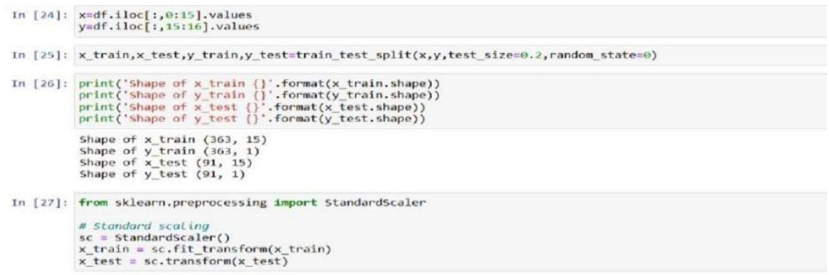
	Diagnosis	FVC	FEV1	Performance	Pain	Haemoptysis	Dyspnoea	Cough	Weakness	Tumor_Size	Diabetes_Mellitus	MI_6m
count	454.000000	454.000000	454.00000	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	454.0000
mean	3.092511	3.287952	2.51685	0.795154	0.059471	0.136564	0.055066	0.696035	0.171806	1.733480	0.074890	0.0444
std	0.715817	0.872347	0.77189	0.531459	0.236766	0.343765	0.228361	0.460475	0.377628	0.707499	0.263504	0.0662
min	1.000000	1.440000	0.96000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.0000
25%	3.000000	2.600000	1.96000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.0000
50%	3.000000	3.160000	2.36000	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000	0.0000
75%	3.000000	3.840000	2.97750	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	2.000000	0.000000	0.0000
max	8.000000	6.300000	5.48000	2.000000	1.000000	1.000000	1.000000	1.000000	1.000000	4.000000	1.000000	1.0000

Exploratory Data analysis





Correlation Coefficient	 

Data Preprocessing Code Screenshots	
Loading Data	
Corelation Matrix	
Data Transformation	
Feature Engineering	Attached the codes in final submission.

Save Processed Data	Data saved in the form of model .pkl file