Summary of Datasets Used:

	Dataset:	Location:
Exploratory Data Analysis	DS1	NA
Data imputation & Transformation	DS2	Cleveland, Switzerland, Long Beach
Modeling (1) and Comparison	DS1	NA
Modeling (2) and Comparison	DS2	Cleveland, Long Beach

For the modeling phase, I divided it into two stages. In the initial stage, I applied different models to the DS1 dataset, experimenting with various model parameters. The table below summarizes the best measurements achieved for each model.

Result of Modeling (1):

Model:	Accuracy Rate:	AUC:	No. of Predictor Variables Used:	Target Variable:
(Winner) Random Forest	84.07%	0.92	13	Absence (1) or presence (2) of heart disease
kNN with Distance lighted	82.13%	0.88	13	
Linear SVC	85.10%	0.91	13	

Following the first modeling stage, Random Forest emerged as the winner, boasting the highest accuracy rate and AUC. I then extended our evaluation to DS2, focusing on different locations, to assess its performance with unseen and loIr-quality datasets.

Result of Modeling (2):

Model:	Location:	RMS:	No. of Predictor Variables Used:	Target Variable:
Random Forest	Cleveland	0.83	13	Diagnosis of heart disease 0-4
	Switzerland	0.91	13	
	Long Beach	1.11	12	

As can be seen in the table above, the Root Mean Square (RMS) values exhibited fluctuations across datasets. The model's ability to generalize across different datasets was limited. Cleveland, with the least missing data, performed the best and exhibited the loIst RMS. This suggests a strong influence of missing data proportions on model performance. It became evident that imputation methods couldn't fully capture the true relationship betIen missing data and the target variable, indicating a potential avenue for future research focused on enhancing imputation accuracy.

Last but not least, it involved a series of experiments at different stages. I explored various aspects such

as categorical versus continuous data, standardization versus no standardization, random forest regression versus random forest classification, Linear SVC versus Polynomial SVC, kNN versus distance-lighted kNN, etc.