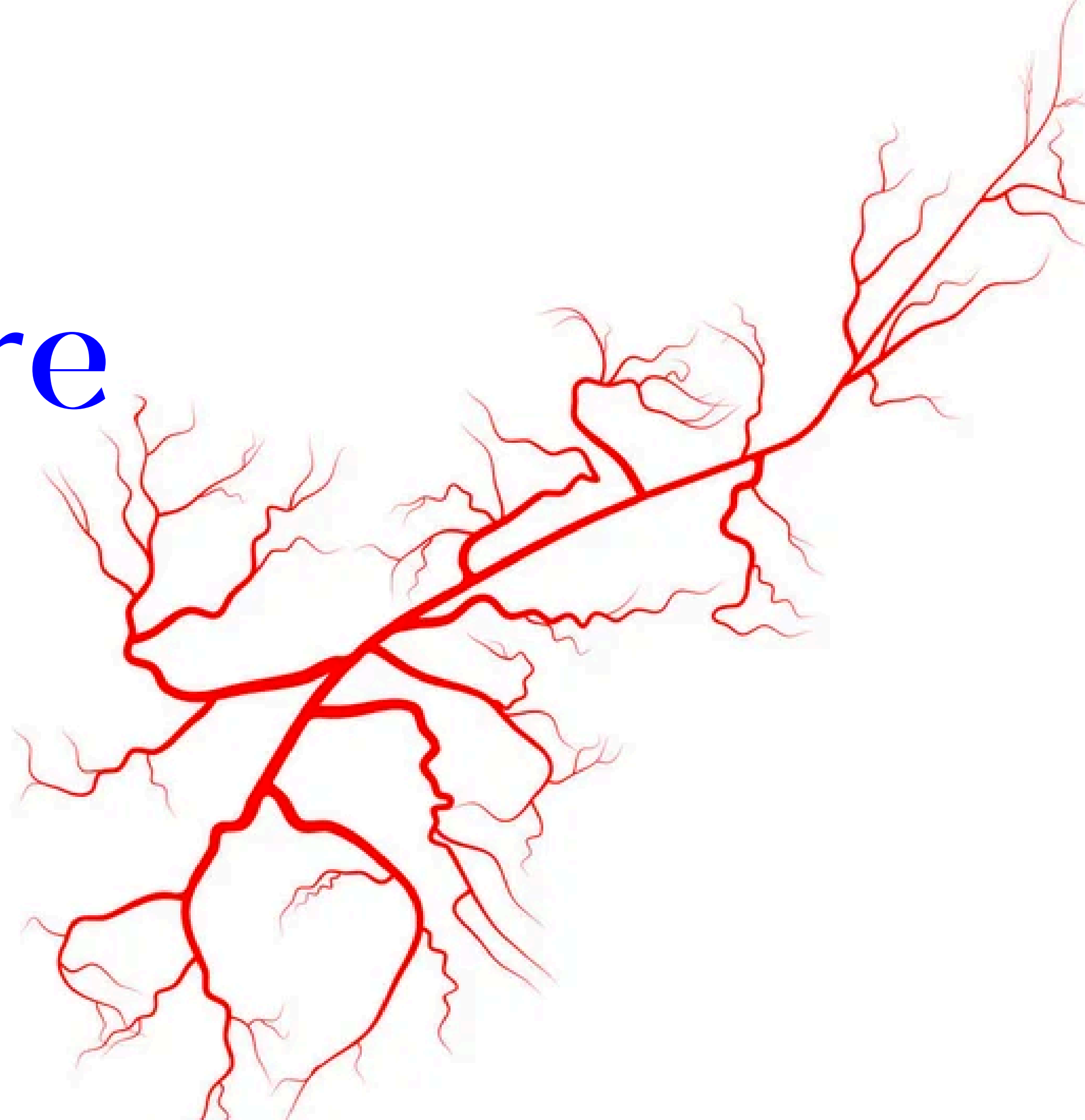


Heart Failure Seeds



Thomas

- 34 years old
- Data Scientist
- lives in the Bay Area
- eats healthy
- plays basketball
- occasional smoker

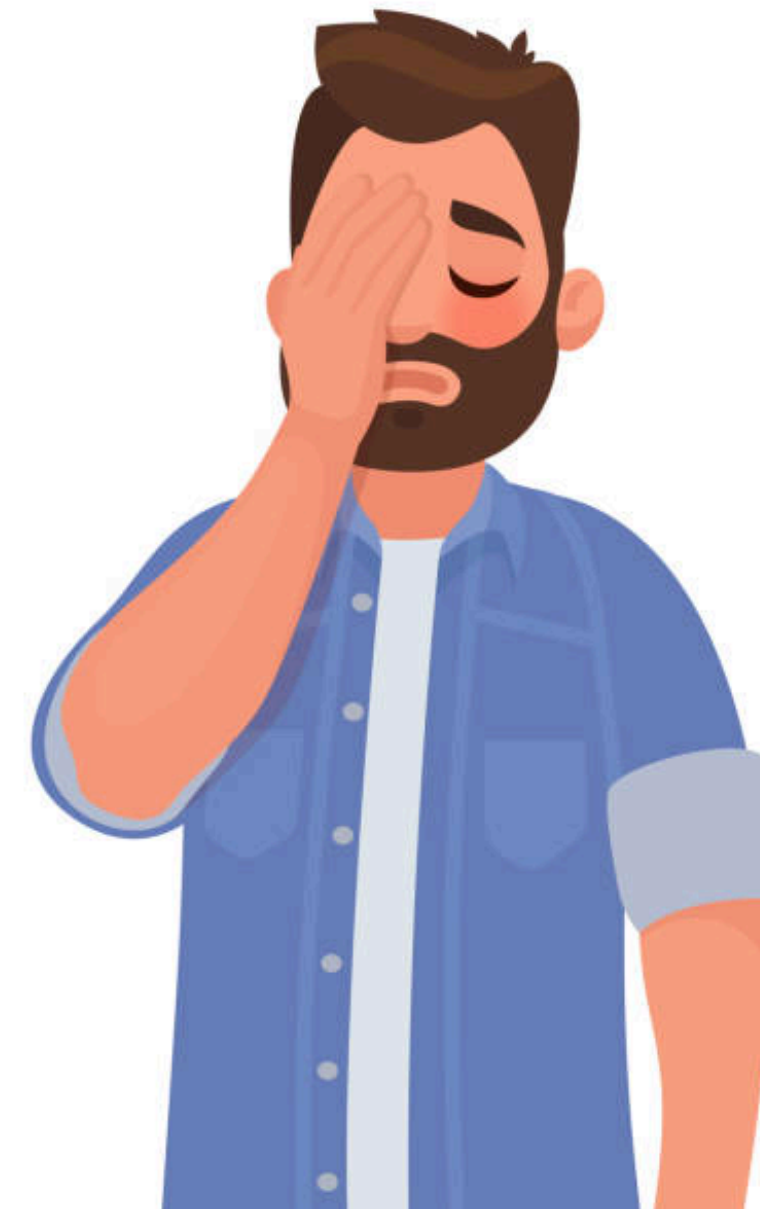


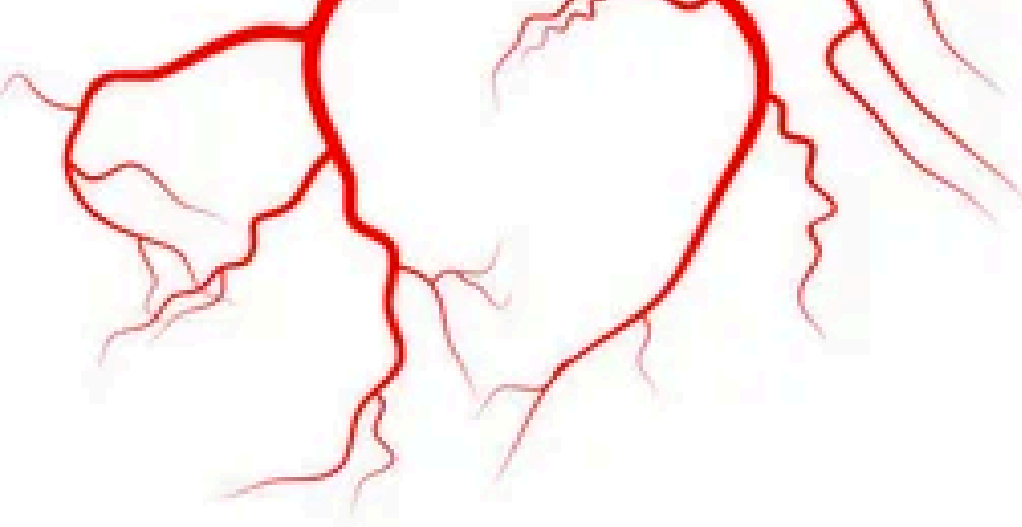
Objectives

- Help Thomas
- Find variables that are associated to heart failure and survival

Methodology

- Perform the following:
 - Chi-Squared test
 - Wilcoxin test





General Information

- Source: [Kaggle](#).
- The original data collectors recorded 13 different features for 299 patients who experienced heart failure.
- Of 299 patients, 203 survived, and 96 died.
- There are 7 numerical features and 6 binary features.
- Our primary variable of interest is the death event, which indicates if the patient died.

Summary of Numeric Variables:

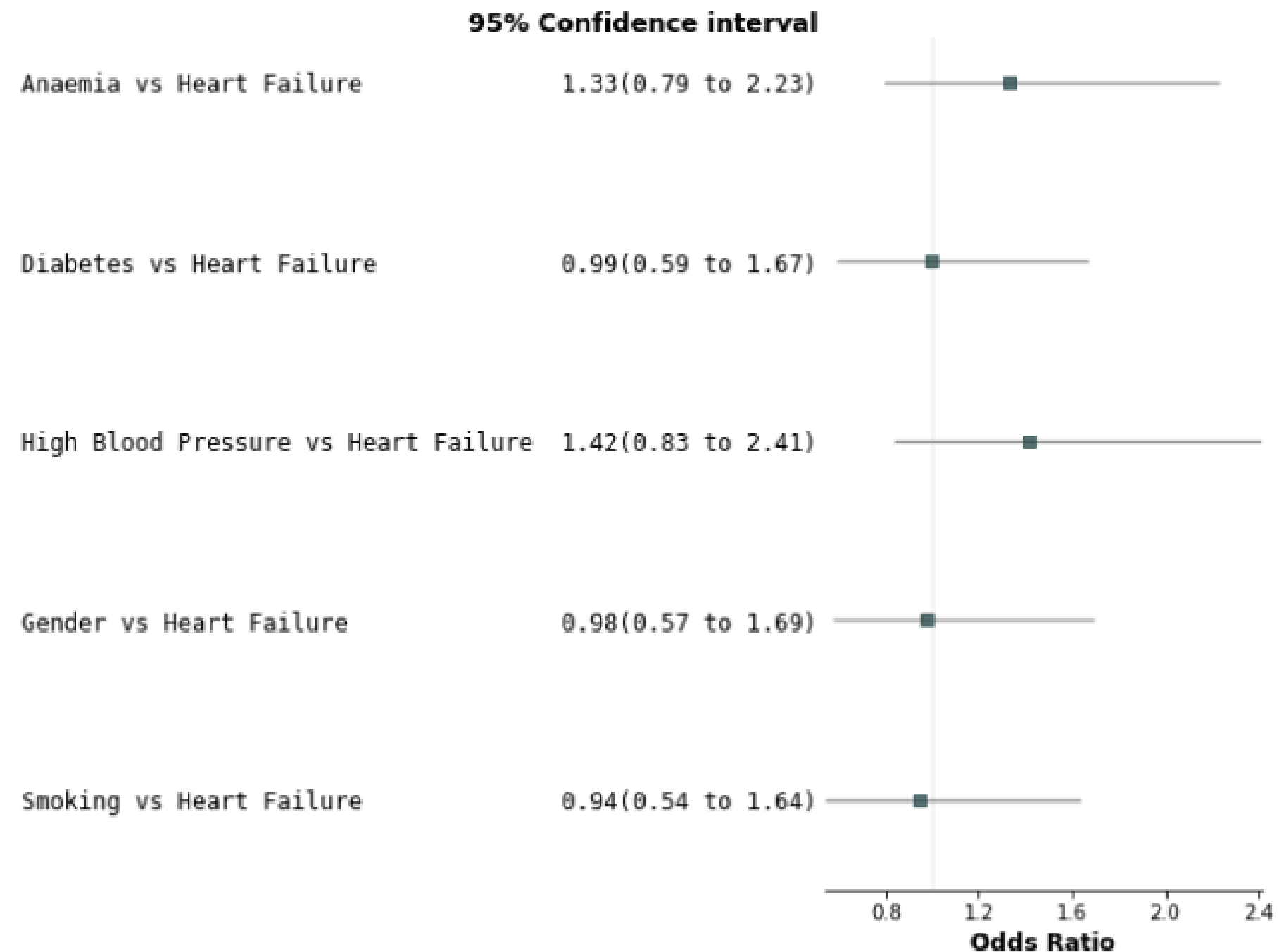
	age	creatinine_phosphokinase	ejection_fraction	platelets	serum_creatinine	serum_sodium	time
count	299.000000	299.000000	299.000000	299.000000	299.000000	299.000000	299.000000
mean	60.833893	581.839465	38.083612	263358.029264	1.39388	136.625418	130.260870
std	11.894809	970.287881	11.834841	97804.236869	1.03451	4.412477	77.614208
min	40.000000	23.000000	14.000000	25100.000000	0.50000	113.000000	4.000000
25%	51.000000	116.500000	30.000000	212500.000000	0.90000	134.000000	73.000000
50%	60.000000	250.000000	38.000000	262000.000000	1.10000	137.000000	115.000000
75%	70.000000	582.000000	45.000000	303500.000000	1.40000	140.000000	203.000000
max	95.000000	7861.000000	80.000000	850000.000000	9.40000	148.000000	285.000000

Summary of Binary Variables:

	Vairables and Level Interpretation:	Count of Level = 0	Count of Level = 1
0	Anaemia (1: Condition Present, 0: Not Present)	170	129
1	Diabetes (1: Condition Present, 0: Not Present)	174	125
2	High Blood Pressure (1: Condition Present, 0: Not Present)	194	105
3	Sex: (1: Male, 0: Female)	105	194
4	Smoking: (1:Smoker, 0:Non-Smoker)	203	96
5	Death Event: (1: Died, 0: Survived)	203	96

Odds Ratio Exploration

- The plot to the **right** shows the point estimate and 95% confidence interval for the Odds Ratio between a binary predictor and the Heart Failure Survival variable.
- The **value 1** is contained in each of the confidence interval shown.
- This indicates that we would **not** expect Chi-Sq test to show that there is association between any of these potential predictors and the Heart Failure Survival variable.



Smoking Status and Heart Failure Survival

We first explored if there is an association between Smoking and Heart Failure using a Chi-Sq Test.

The hypotheses for this test are:

- H0: There is no association between Smoking and Heart Failure
- H1: There is association between Smoking and Heart Failure

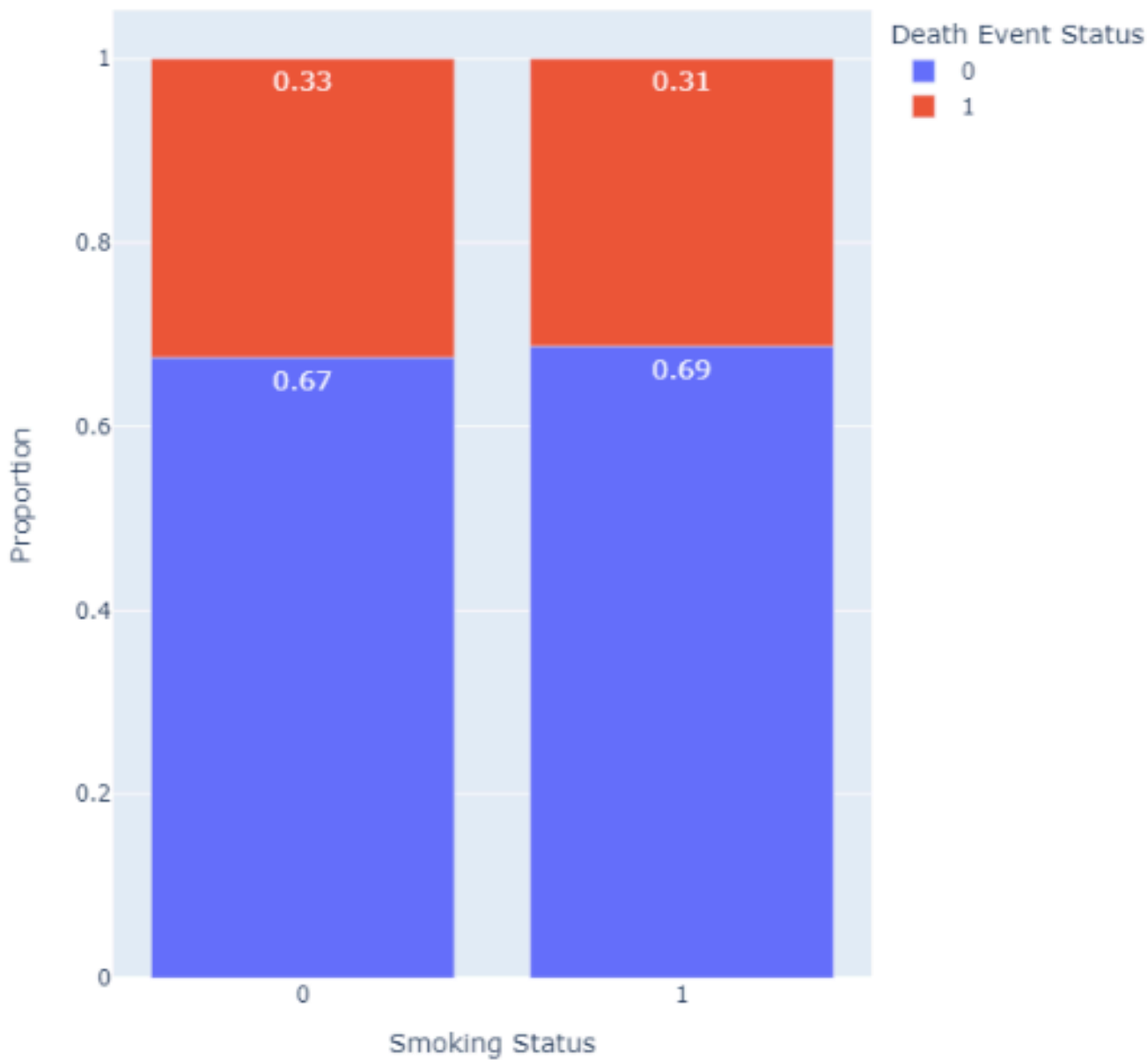
The test results are shown in the table to the bellow.

Based on the results of the Chi-Sq, we **fail to reject** the notion that there Smoking and Heart Failure Survival aren't associated.

Results of the Chi-Sq Test:

Variable At Focus:		Chi-Sq Test Statistic Value:	Degrees of Freedom	P-Value:
0	Smoking	0.007331	1	0.931765

Proportion of Death Event by Smoking Status



Serum Creatinine and Heart Failure Survival

We next explore if there is a relationship Serum Creatinine and Heart Failure variables using a Wilcoxon Rank Test (Based on the distribution of the variable).

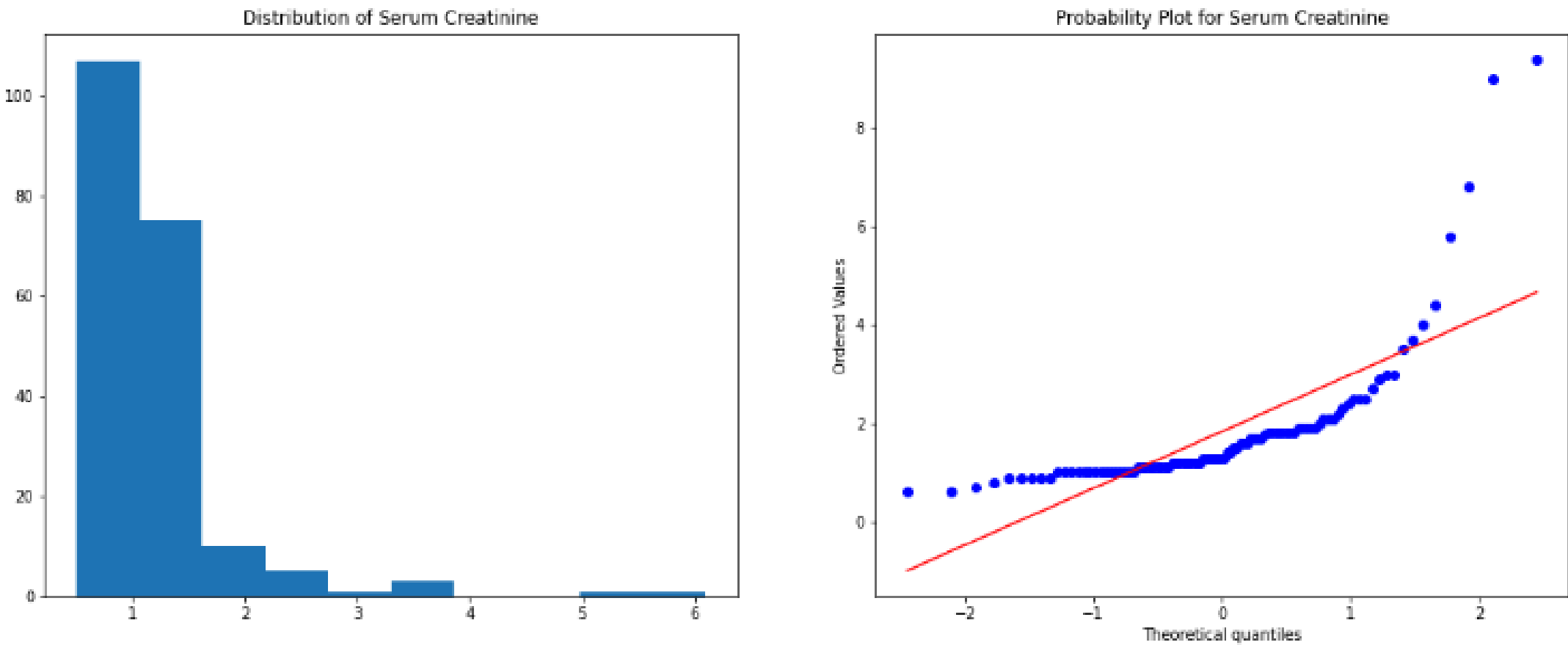
The hypothesis for this test are:

- H0: The Median Serum Creatinine Levels are same Heart Failure levels.
- H1: The Median Serum Creatinine Levels are different Heart Failure Survival levels.

The test results are shown in the table to the right.

Based on the results of the Rank Test performed, we reject H0 and conclude the median Serum Creatinine Levels are **different** Heart Failure levels.

Overall Distribution of Serum Creatinine:



Results of the Wilcoxon Rank Sum Test:

Wilcoxon Rank Sum Test Statistic		P-Value
0	-6.369654	1.894555e-10

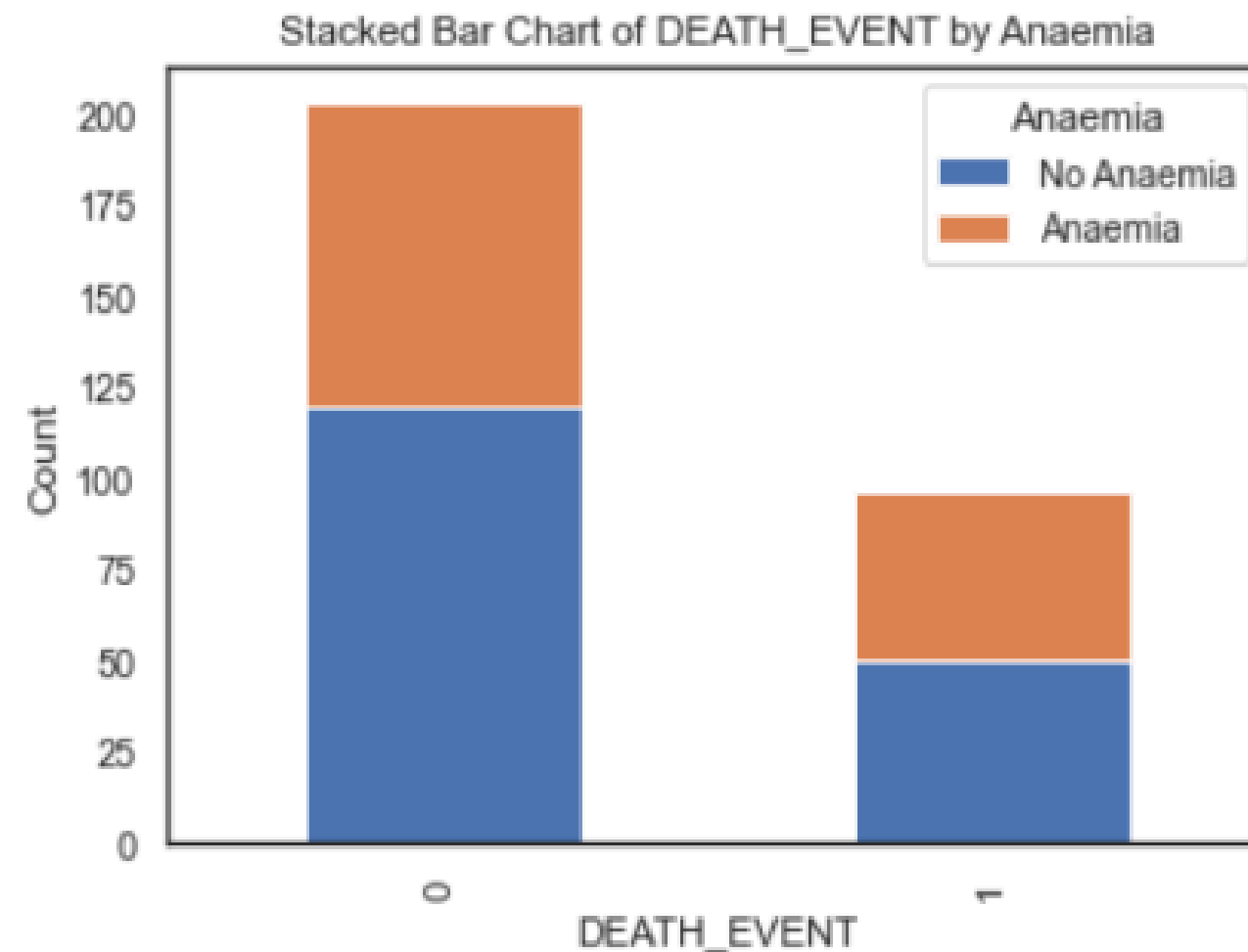
Anaemia and Heart Failure Survival

The hypothesis for this test are:

- H0: There is no association between anemia status and heart failure survival.
- H1: There is an association between anemia status and heart failure survival.

Applying the Chi-squared test yielded a test statistic of 1.042 and a p-value of 0.307. With a significance level of 0.05, we fail to reject the null hypothesis since the p-value is greater than 0.05.

Therefore, there is **not sufficient** evidence to conclude that there is a statistically significant association between anemia status and heart failure diagnosis. Further investigation may be warranted with a larger sample size.



Results of the Chi-Sq Test:

Chi-Squared Value: 1.042175485287005

P-value: 0.3073160508415107

Creatinine Phosphokinase and Heart Failure Survival

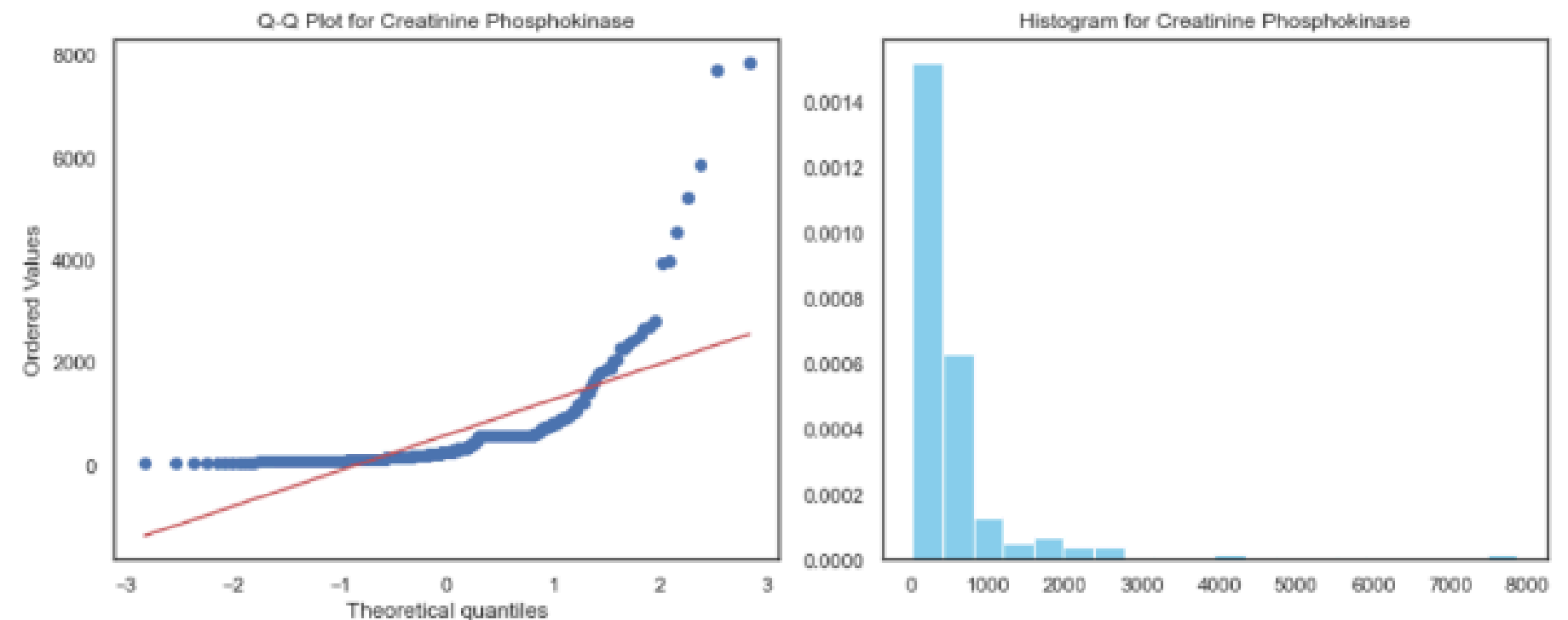
The hypothesis for this test are:

- H0: The median creatine phosphokinase levels are equal in heart failure survival and non-heart failure patients.
- H1: The median creatine phosphokinase levels differ between heart failure survival and non-heart failure patients.

Applying the Mann-Whitney U test yielded a test statistic of 9,420 and a p-value of 0.68. With a significance level of 0.05, we fail to reject the null hypothesis since the p-value is greater than 0.05.

Therefore, there is **not sufficient** evidence to conclude that median creatine phosphokinase levels differ significantly between heart failure and non-heart failure patients.

Overall Distribution of Creatinine phosphokinase:

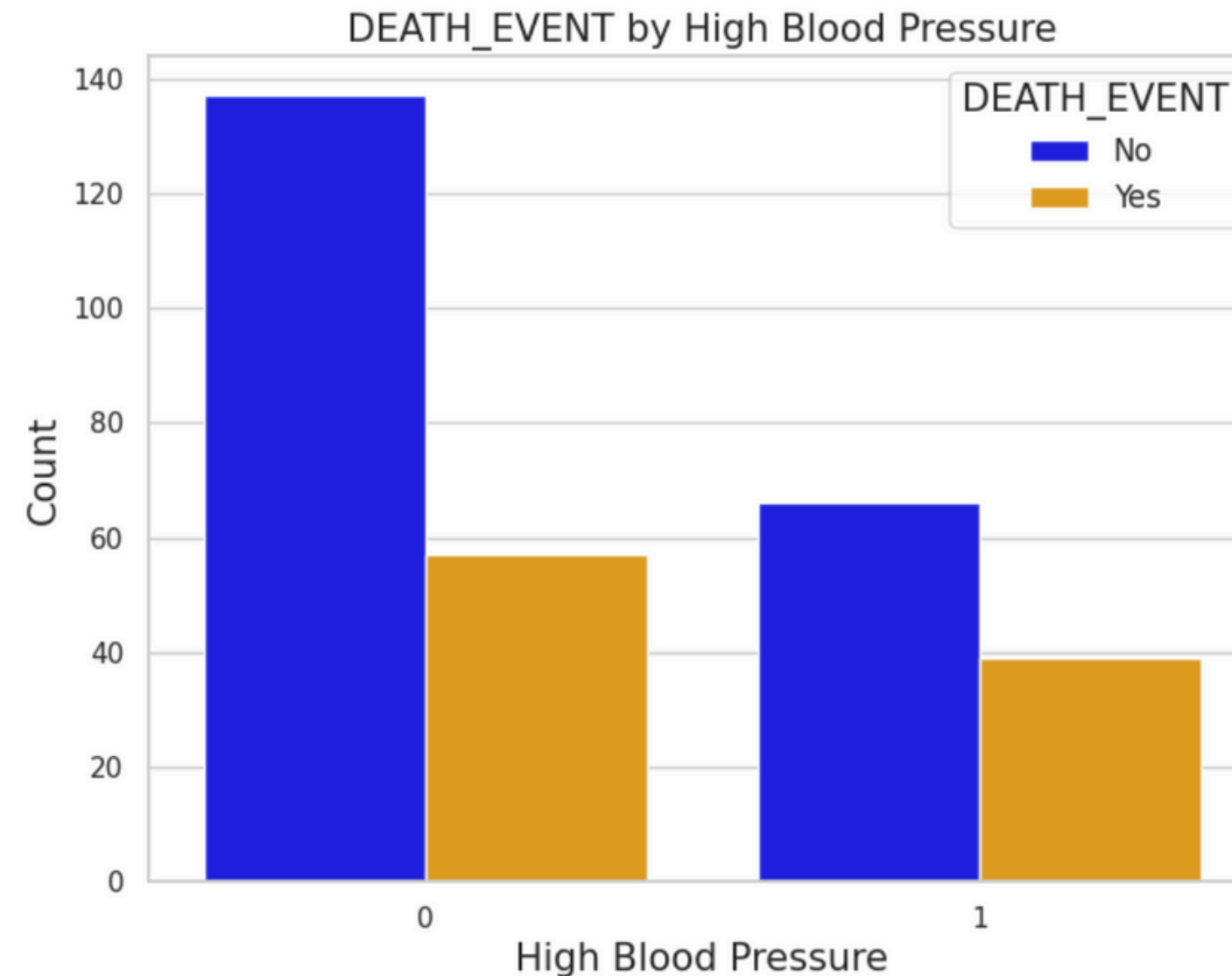


Results of the Mann-Whitney/Wilcoxon Rank Sum Test:

Mann-Whitney U Statistic: 9460.0
P-value: 0.6840400261440358

High Blood Pressure and Heart Failure Survival

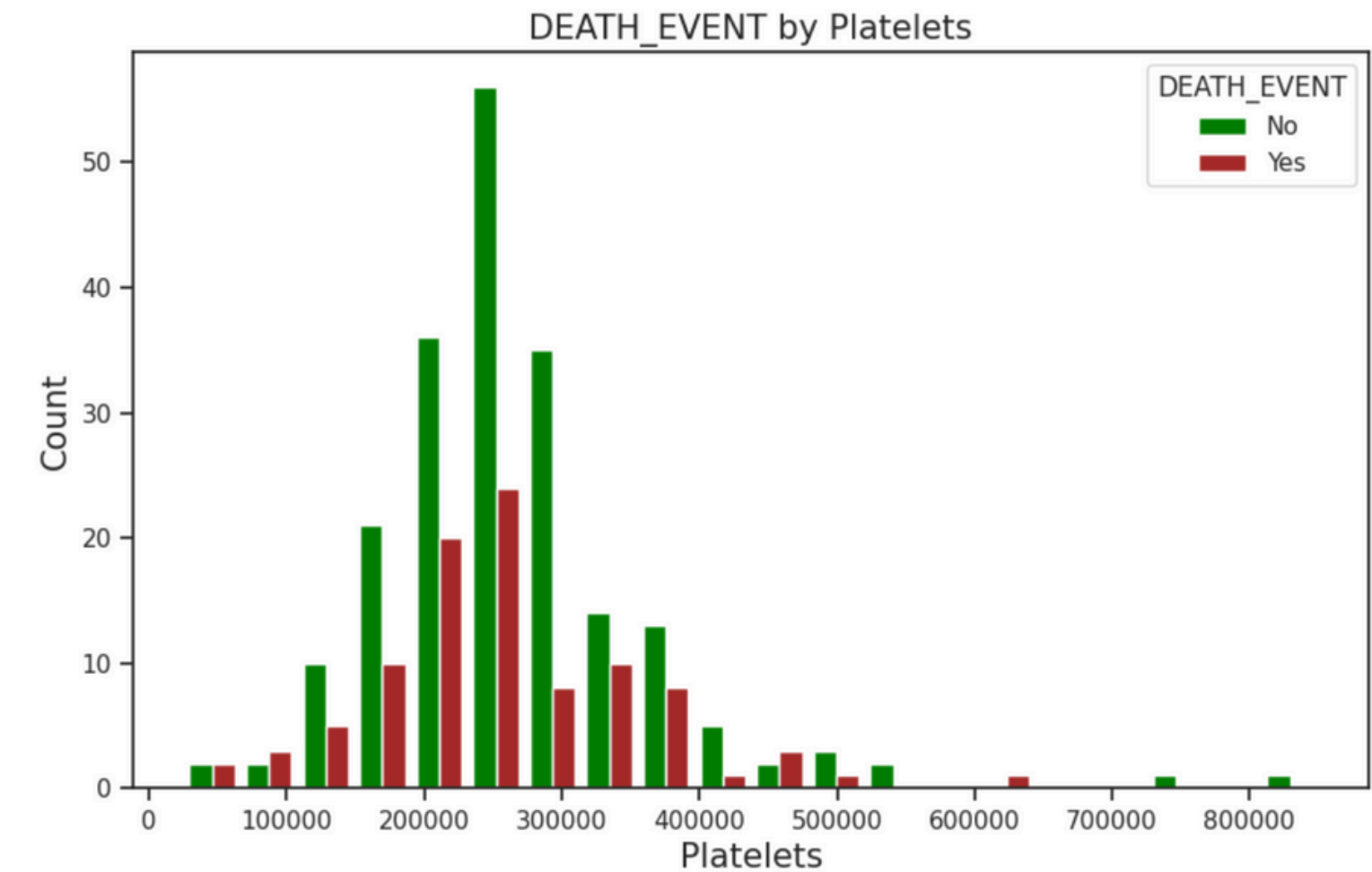
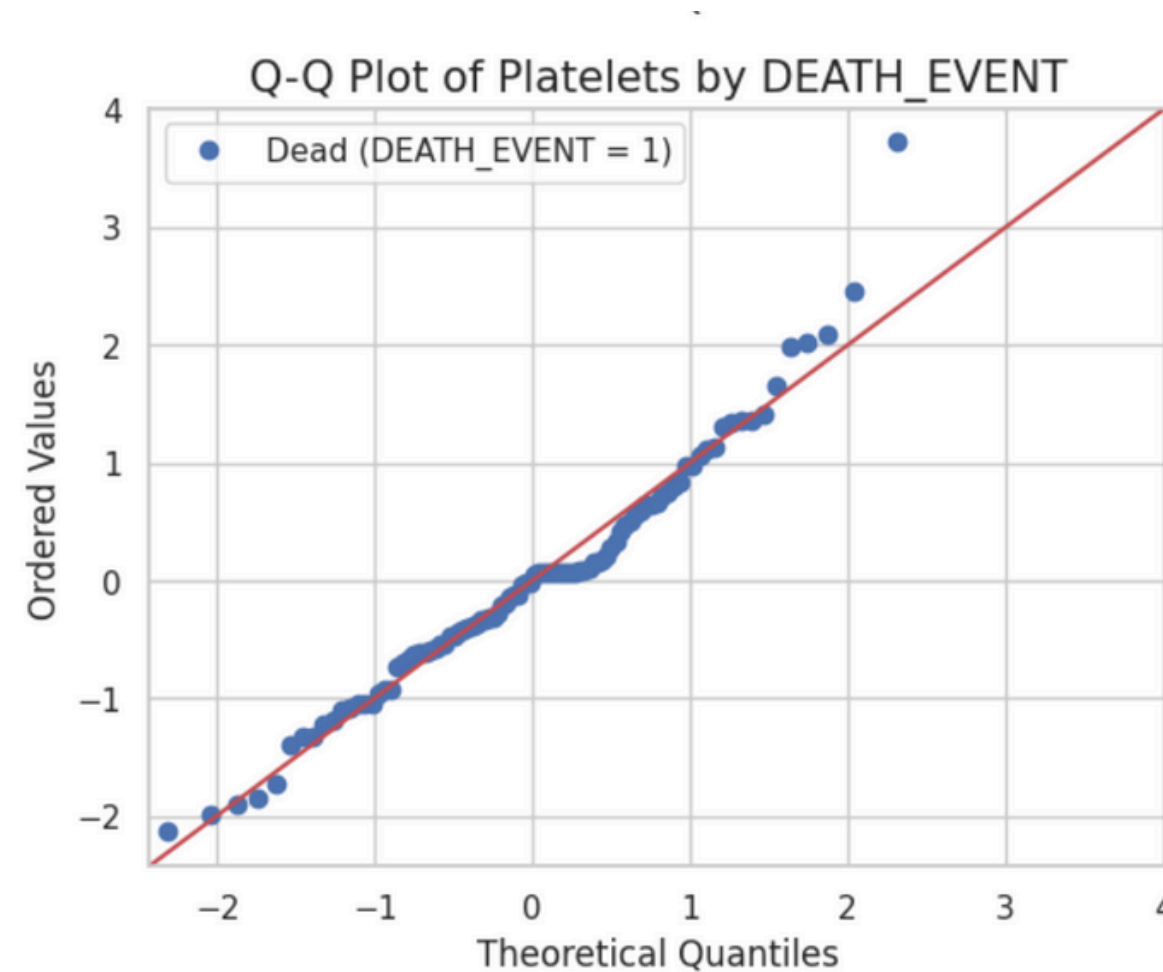
- **H0** - The distribution of HBP values is the same between individuals who died and survived
- **H1** - The distribution of HBP values is different between individuals who died and survived
- **Results:** There is no statistically significant association between 'high blood pressure' and the occurrence of 'DEATH_EVENT'



Chi-Square Test for 'high blood pressure' vs. 'DEATH_EVENT':
Chi-Square Statistic: 1.5434607612967581
P-Value: 0.21410341199416902

Platelets and Heart Failure Survival

- **H0:** There is no association between platelets and death
- **H1:** There is association between platelets and death
- **Results:** There is no statistically significant association between platelets and death



Mann-Whitney U test for platelets between the two groups:
Statistic: 9187.5
P-Value: 0.4255585098046608

Conclusion and Future work

- Thomas' condition improved
- Serum Creatinine was the only variable associated with heart failure survival.

In the Future

- We will use binary classifiers like Logistic Regression and KNN for survival prediction

