




Advanced EDA for Genomic Data Analysis

Upload your dataset to begin analysis


Choose a CSV file



Drag and drop file here

Limit 200MB per file • CSV

Browse files

 Breast_Cancer.csv 386.8KB ×

Dataset Preview

	Age	Race	Marital Status	T Stage	N Stage	6th Stage	differentiate	Grade	A Stage
0	68	White	Married	T1	N1	IIA	Poorly differentiated	3	Region
1	50	White	Married	T2	N2	IIIA	Moderately differentiated	2	Region
2	58	White	Divorced	T3	N3	IIIC	Moderately differentiated	2	Region
3	58	White	Married	T1	N1	IIA	Poorly differentiated	3	Region
4	47	White	Married	T2	N1	IIB	Poorly differentiated	3	Region

Dataset Structure

Shape of Dataset: (4024, 16)

Data Types:

	0
Age	int64
Race	object
Marital Status	object
T Stage	object
N Stage	object
6th Stage	object
differentiate	object
Grade	object
A Stage	object
Tumor Size	int64

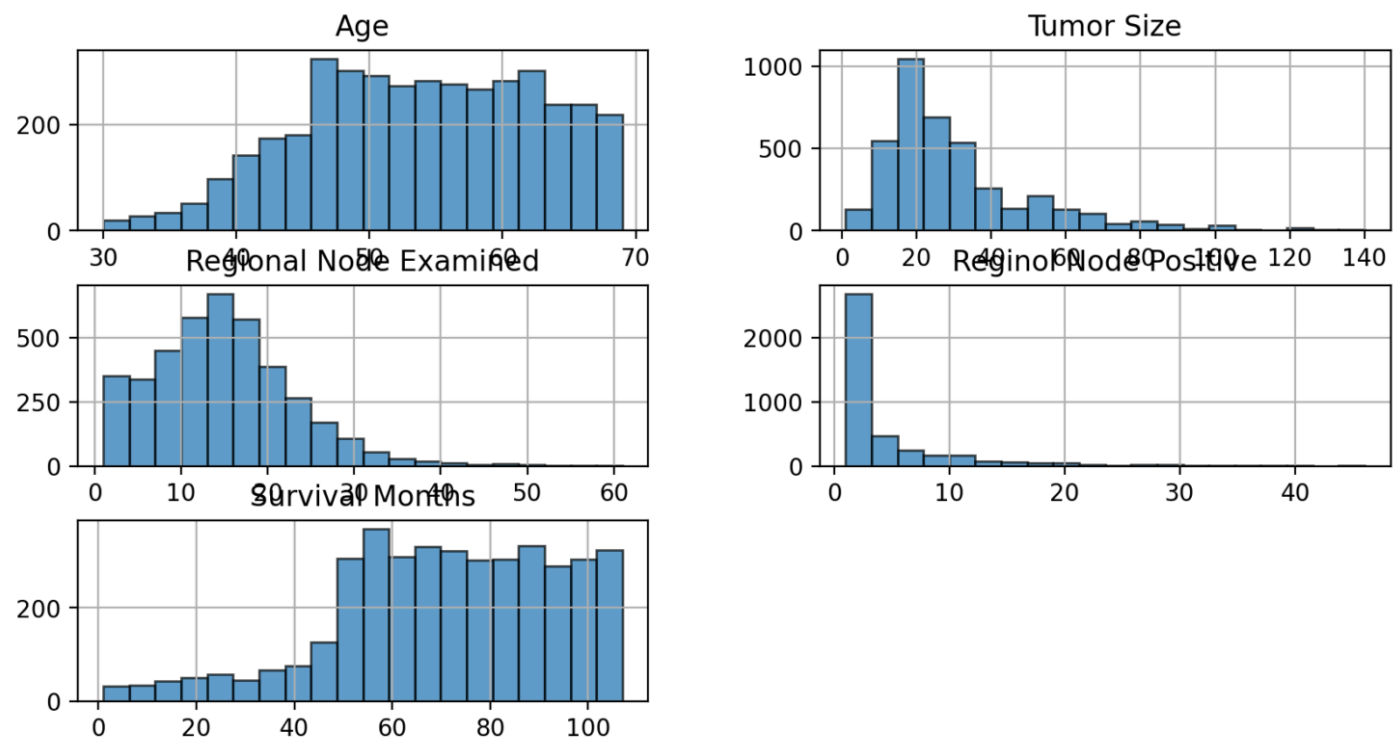
Data Quality Check

Missing Values:

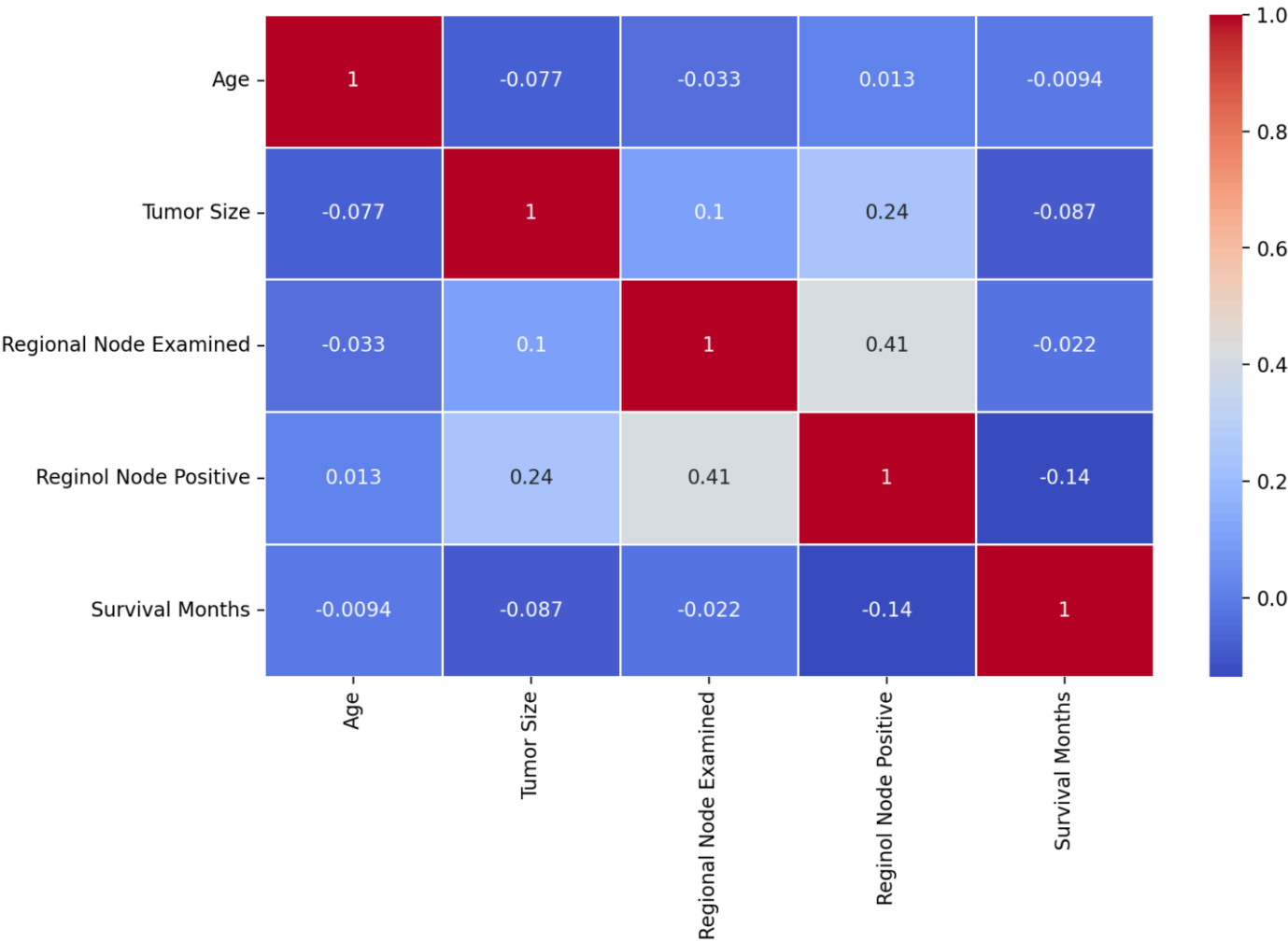
	0
Age	0
Race	0
Marital Status	0
T Stage	0
N Stage	0
6th Stage	0
differentiate	0
Grade	0
A Stage	0
Tumor Size	0

Duplicate Rows: 1

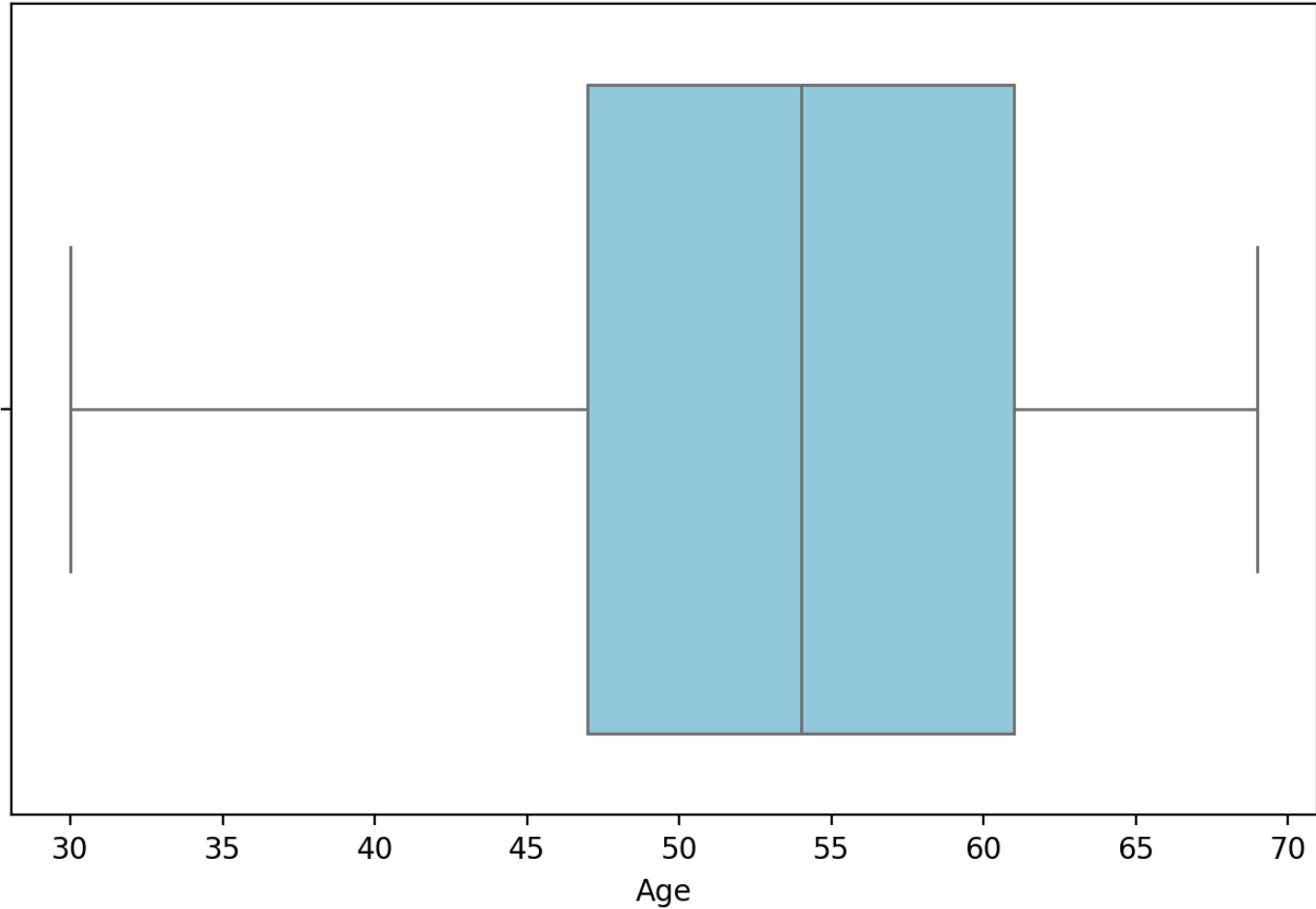
Feature Distribution

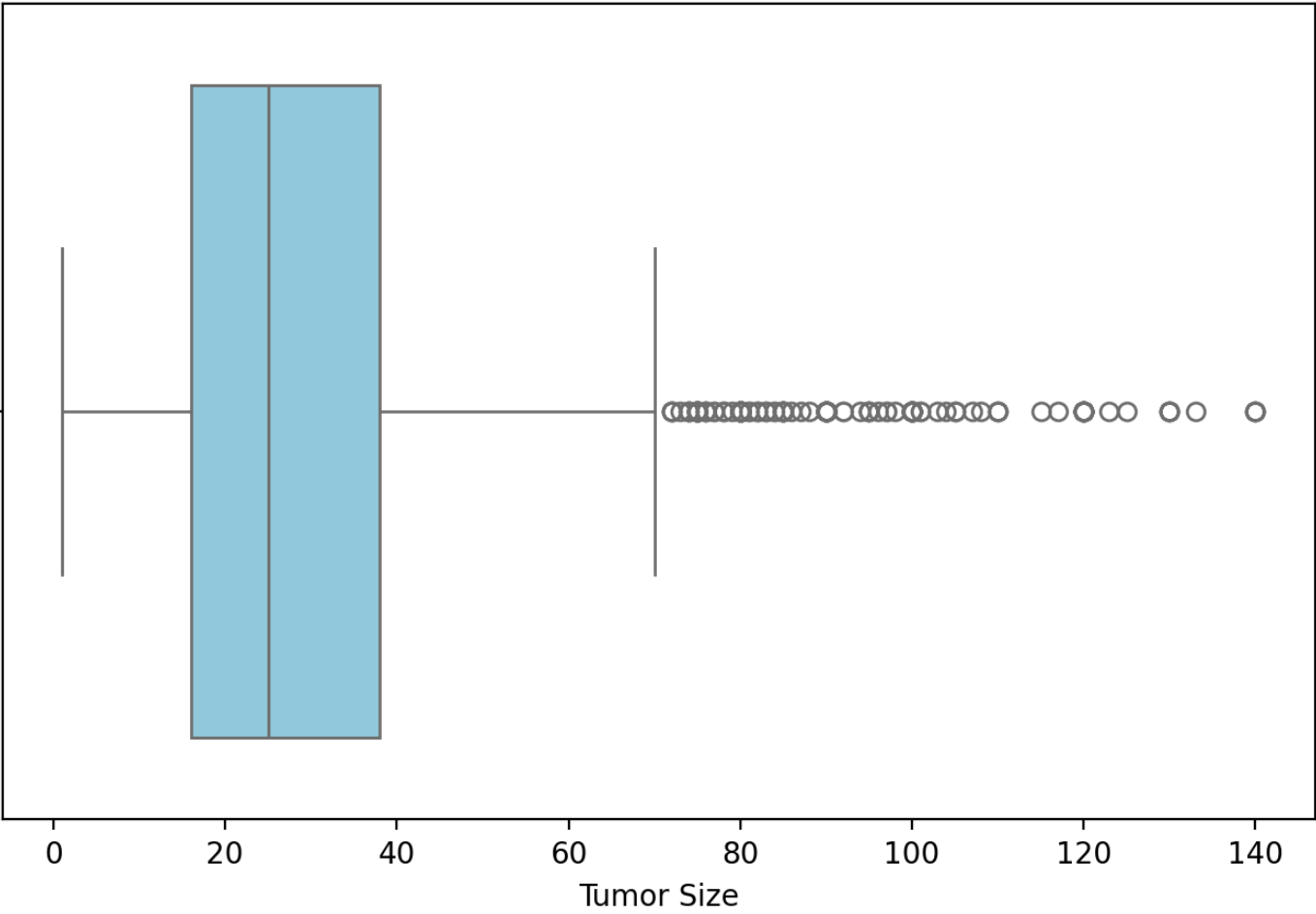


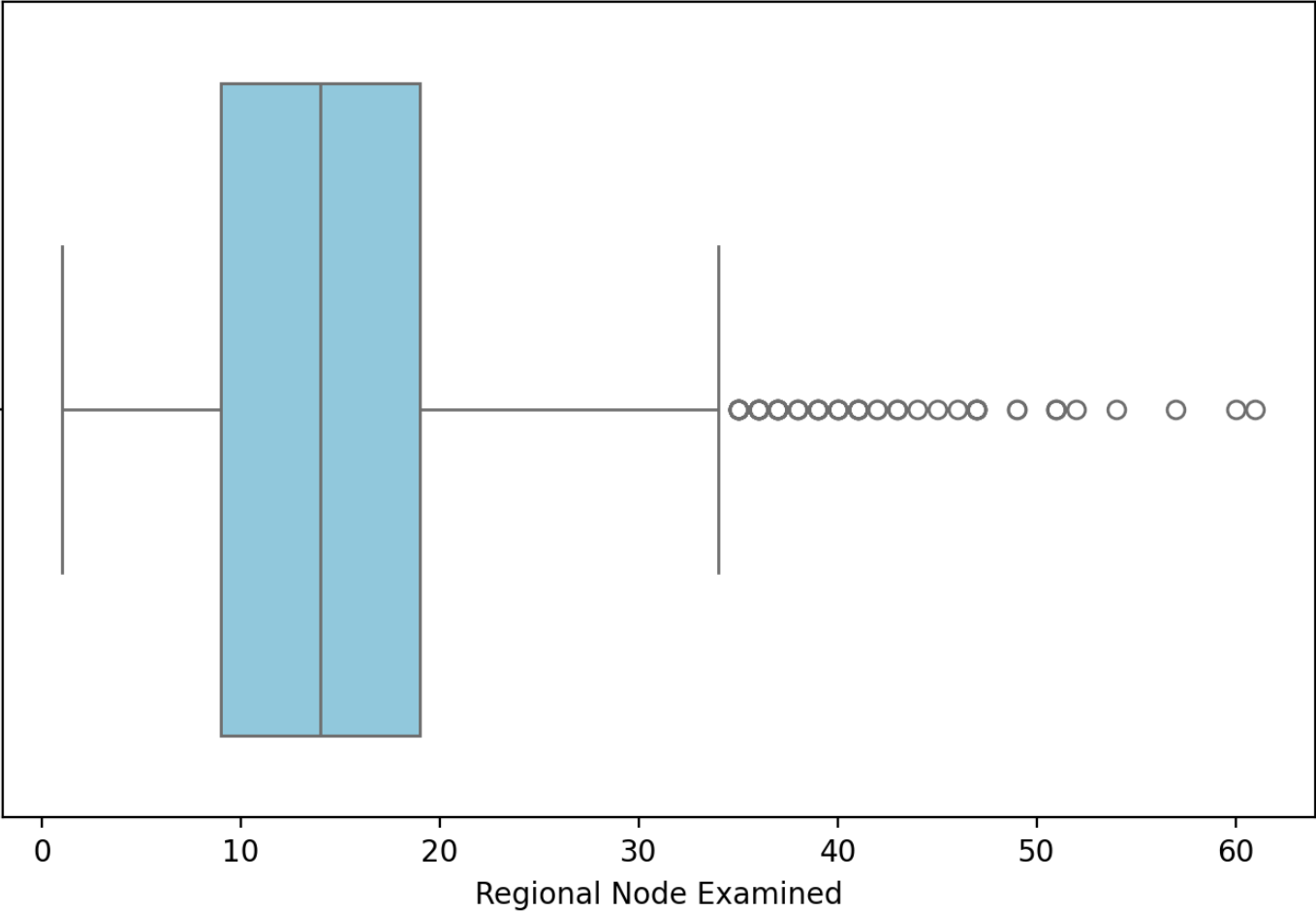
Feature Correlation

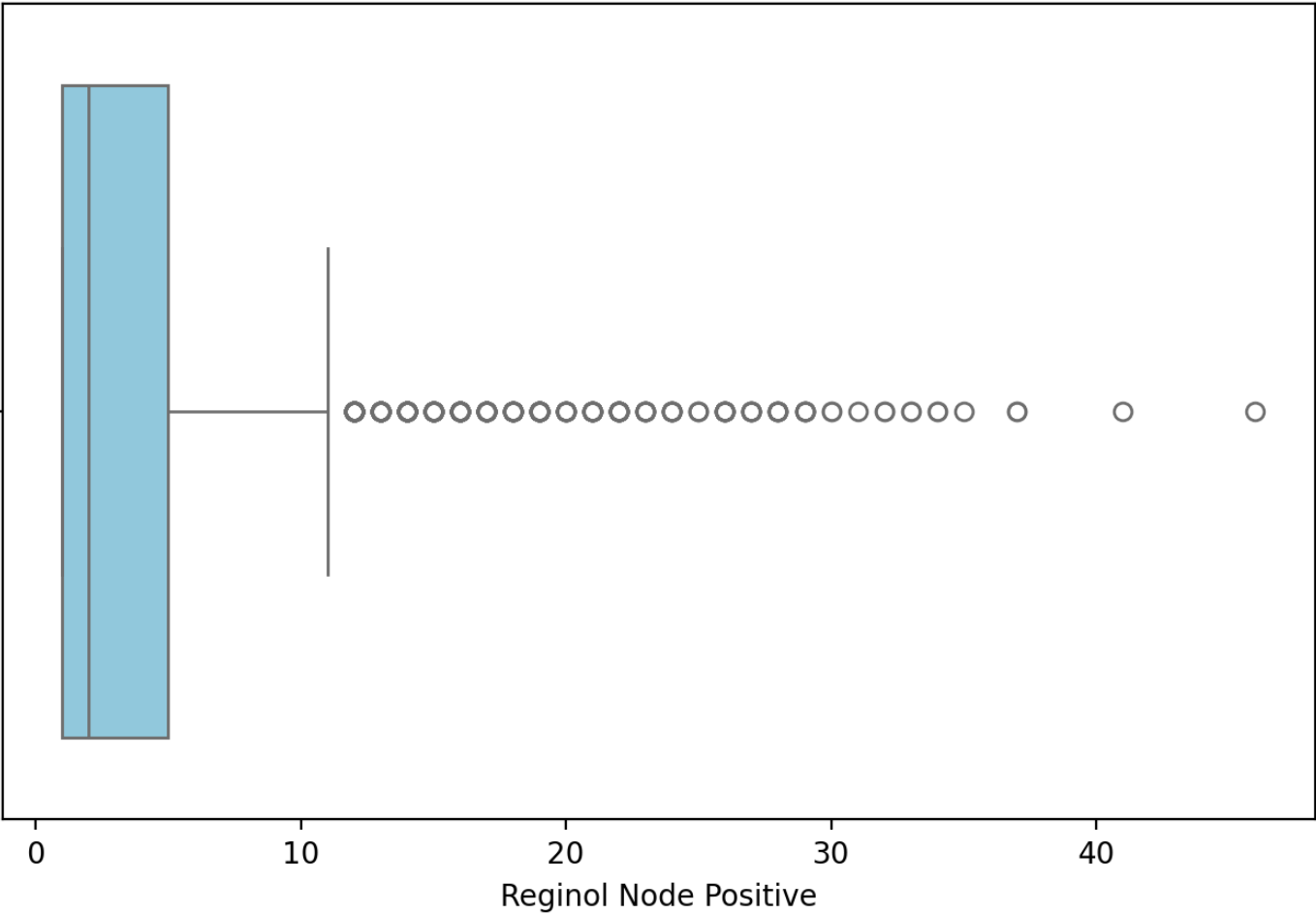


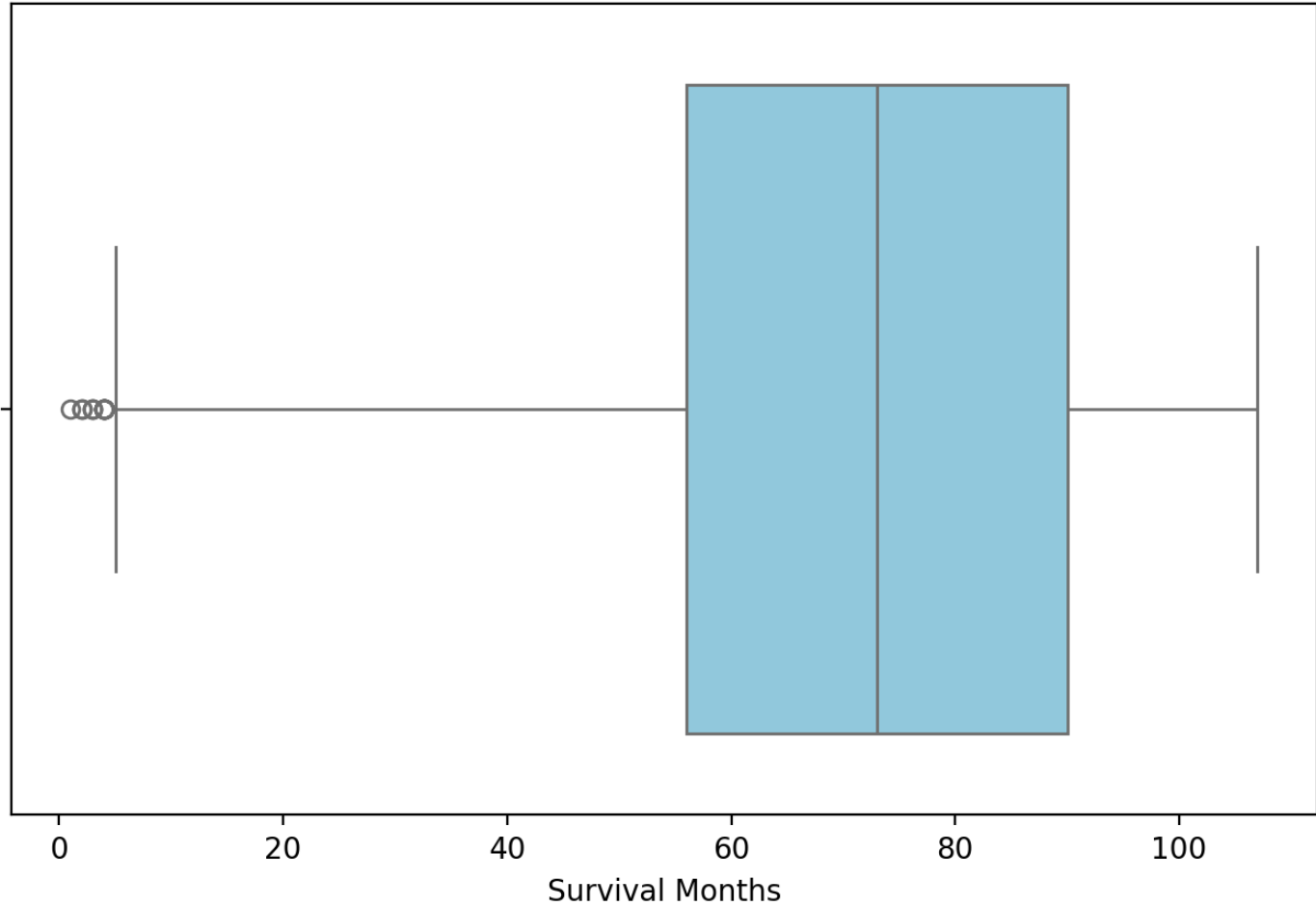
Outlier Detection



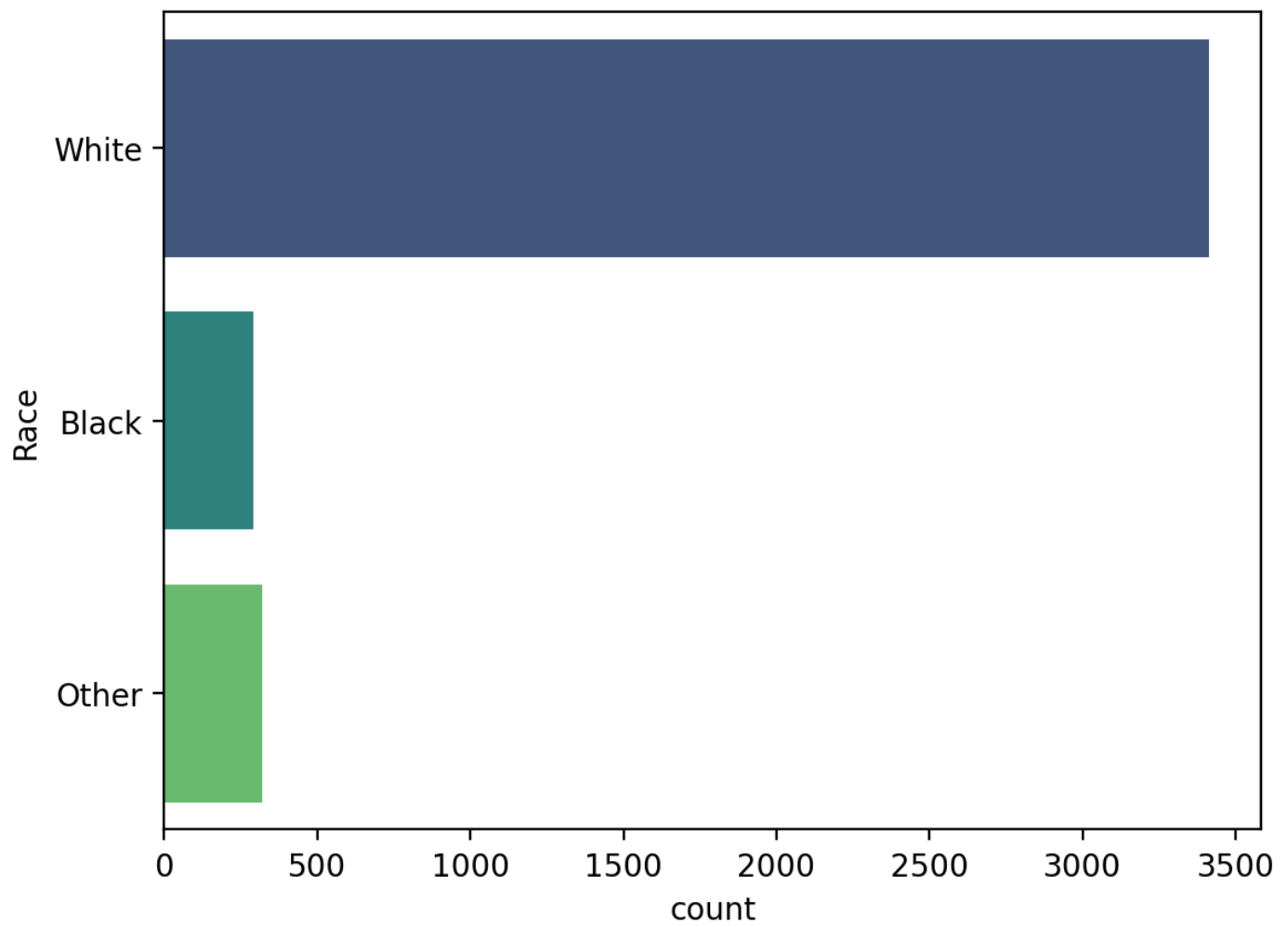


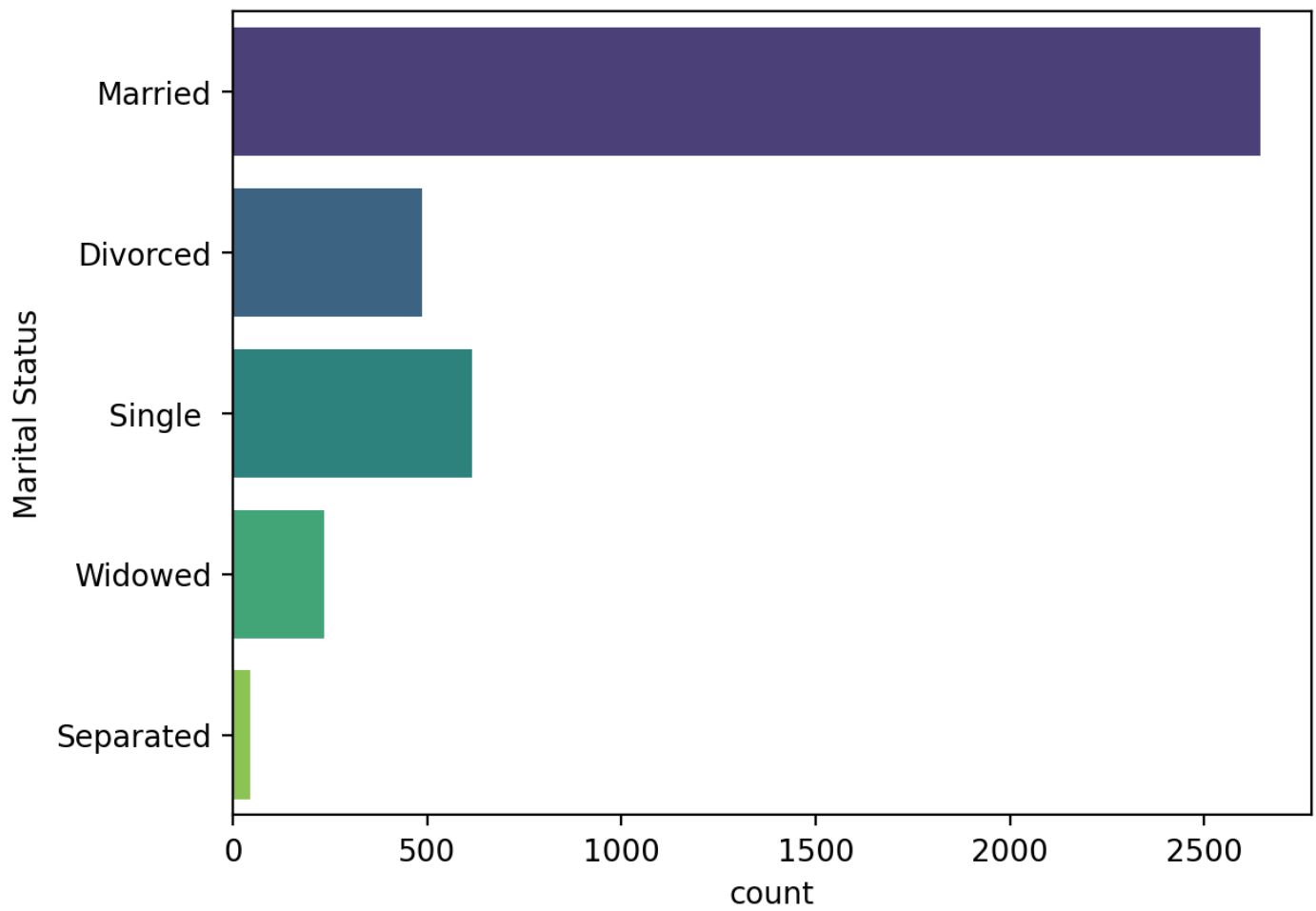


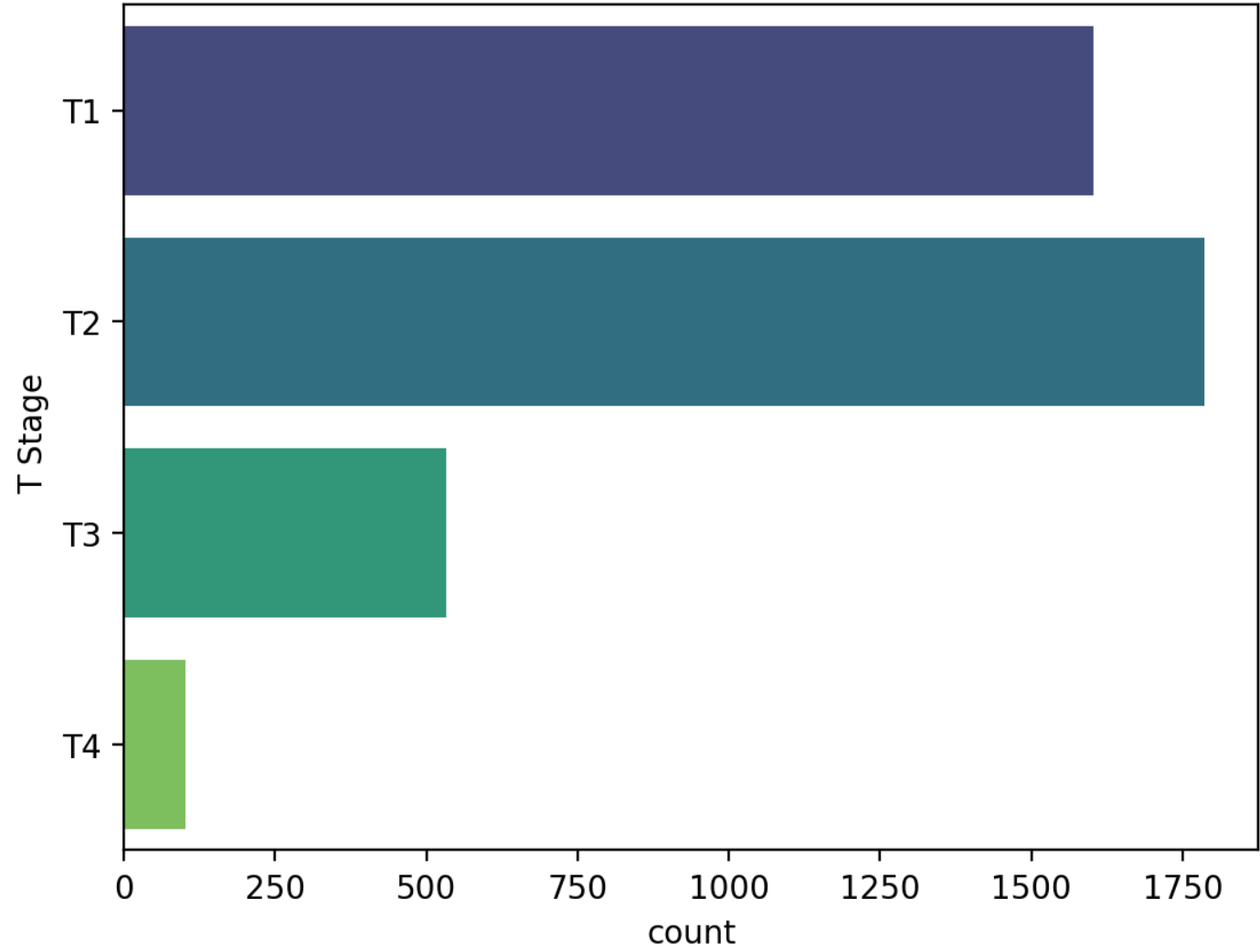


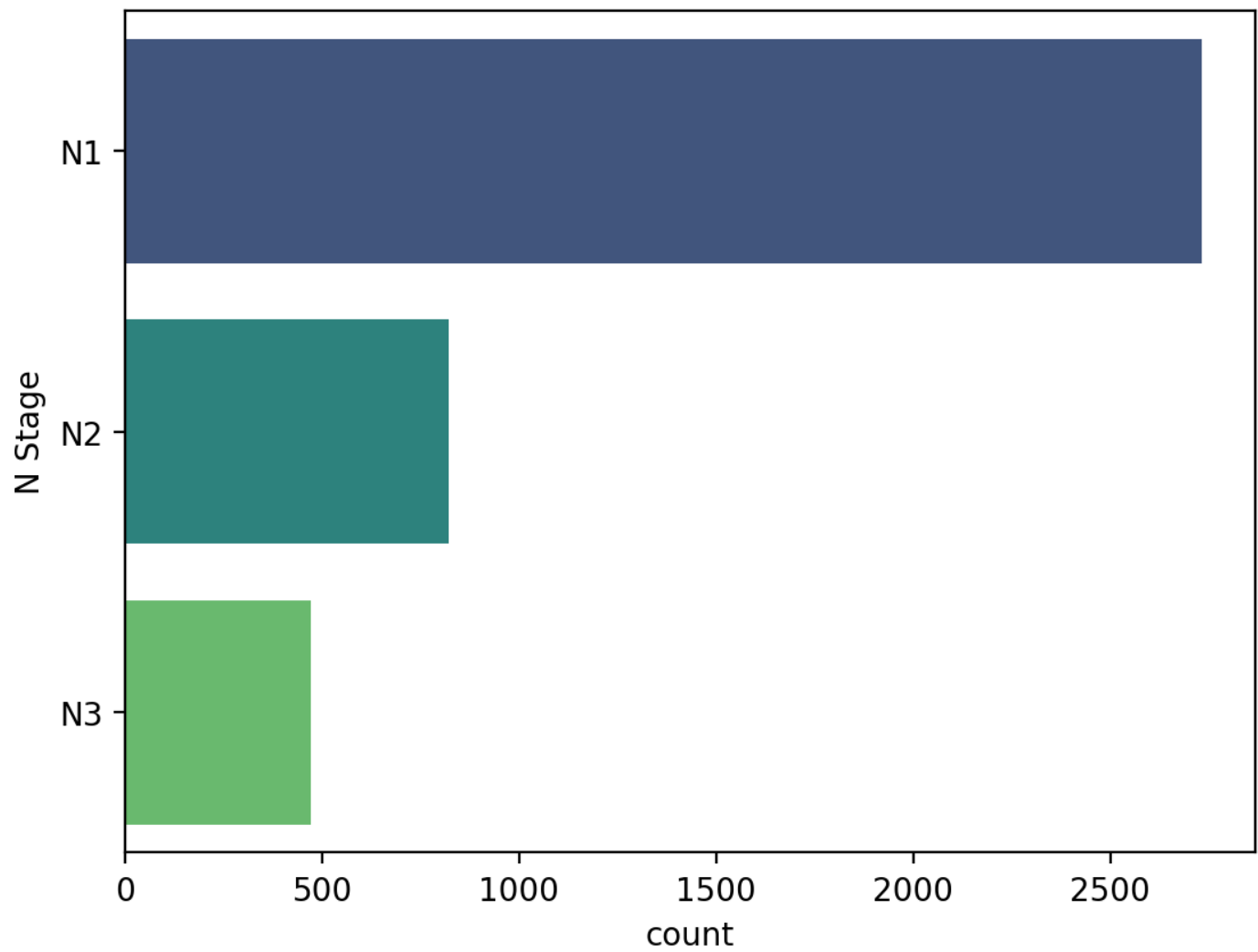


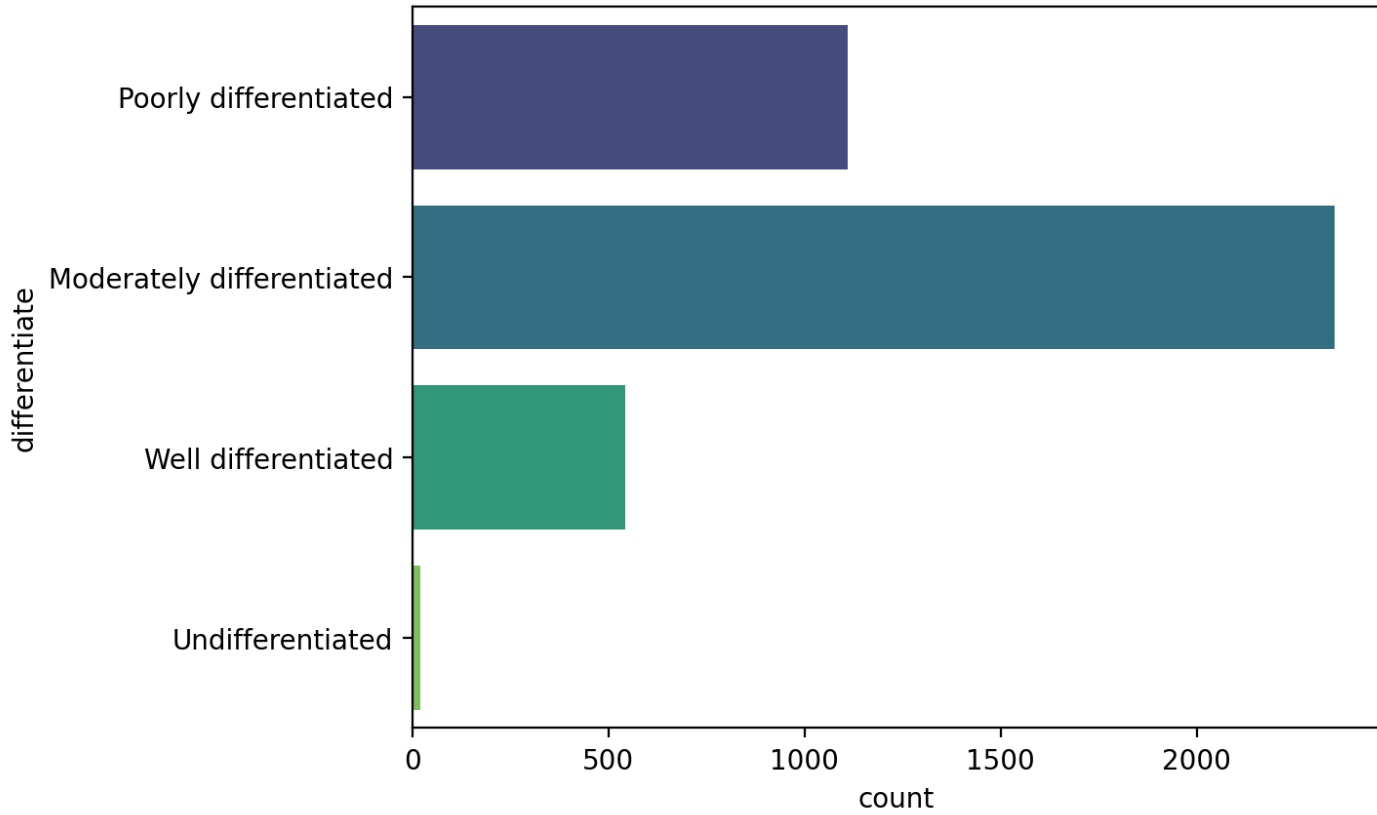
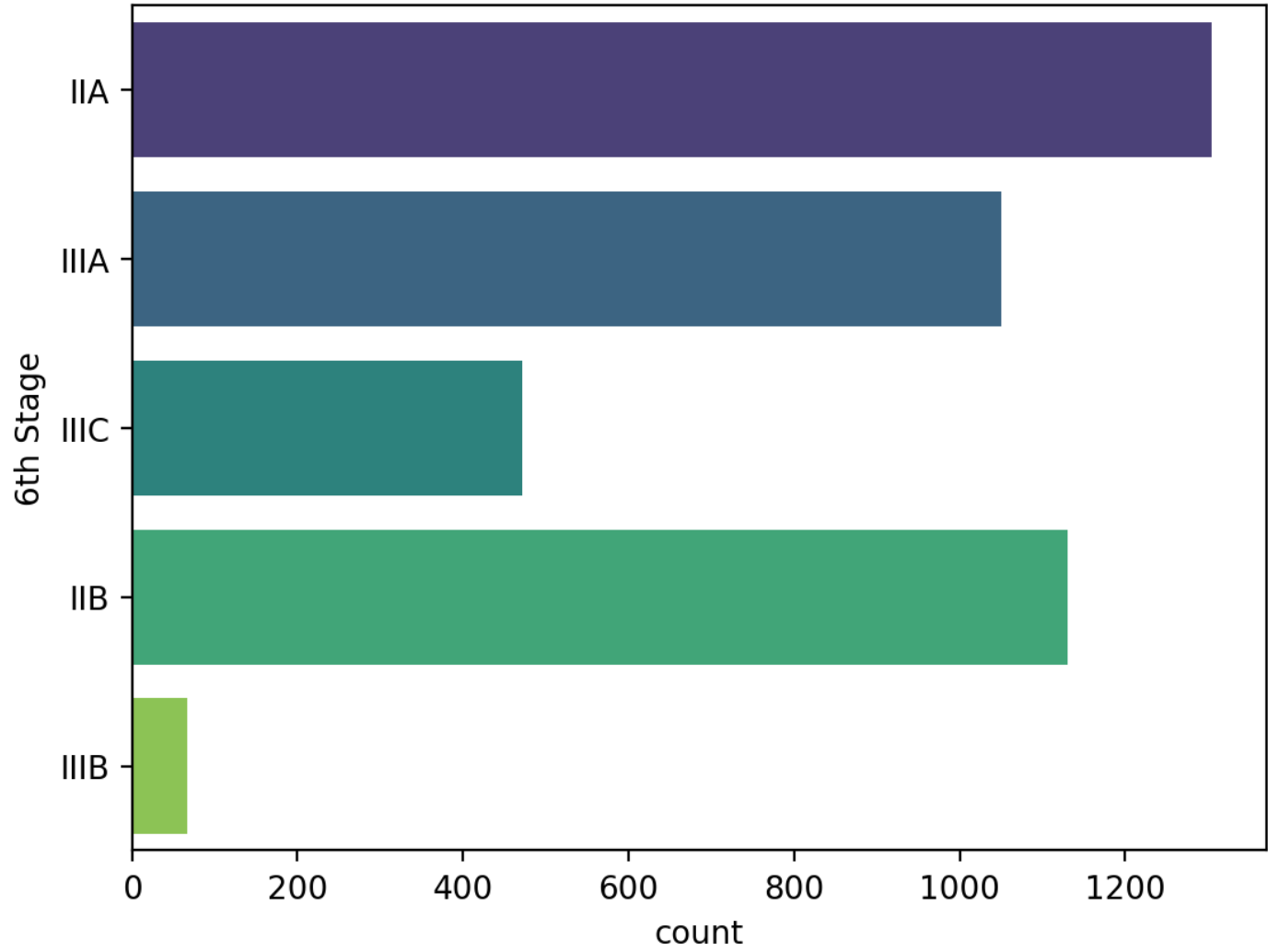
Categorical Feature Distribution

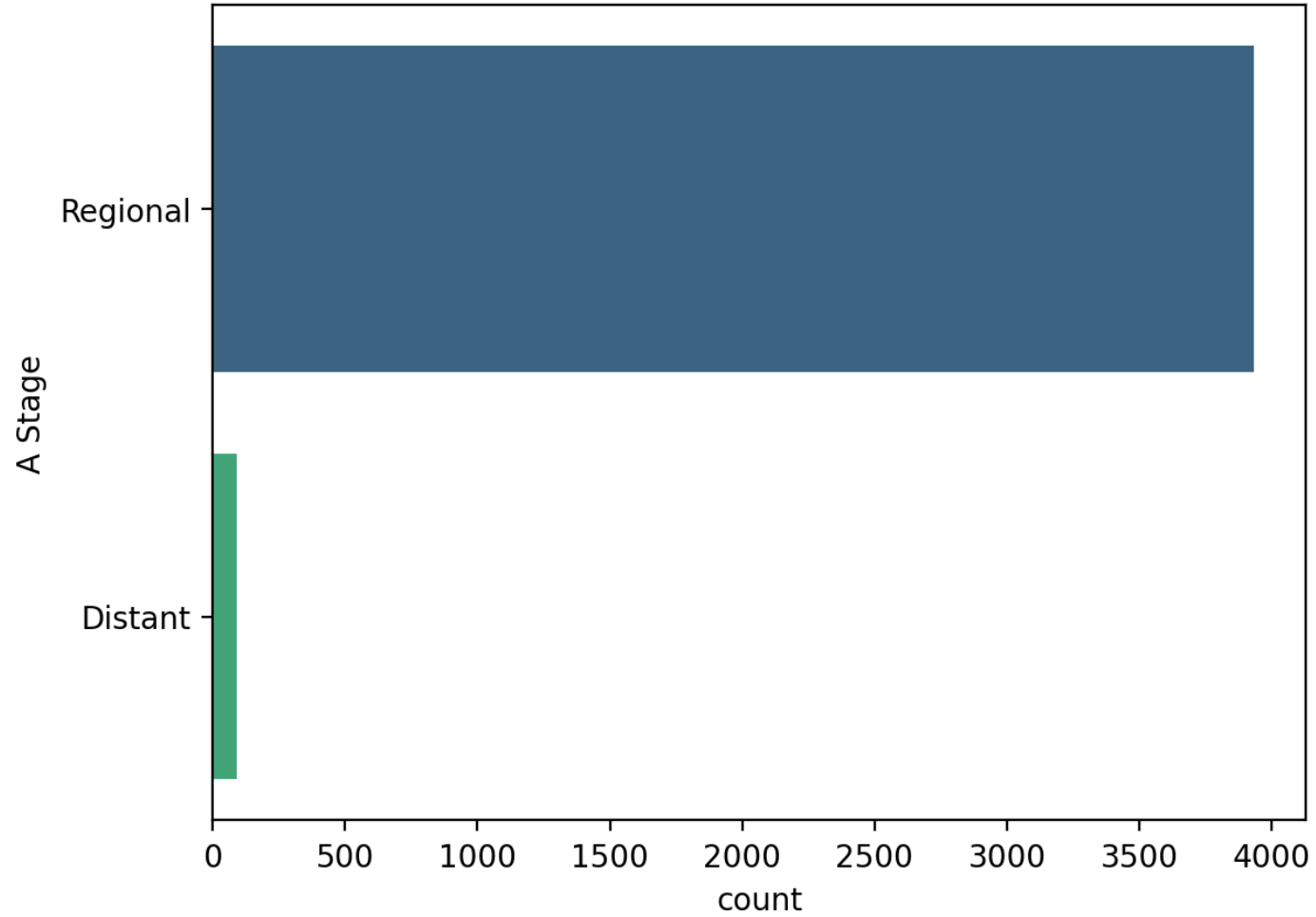
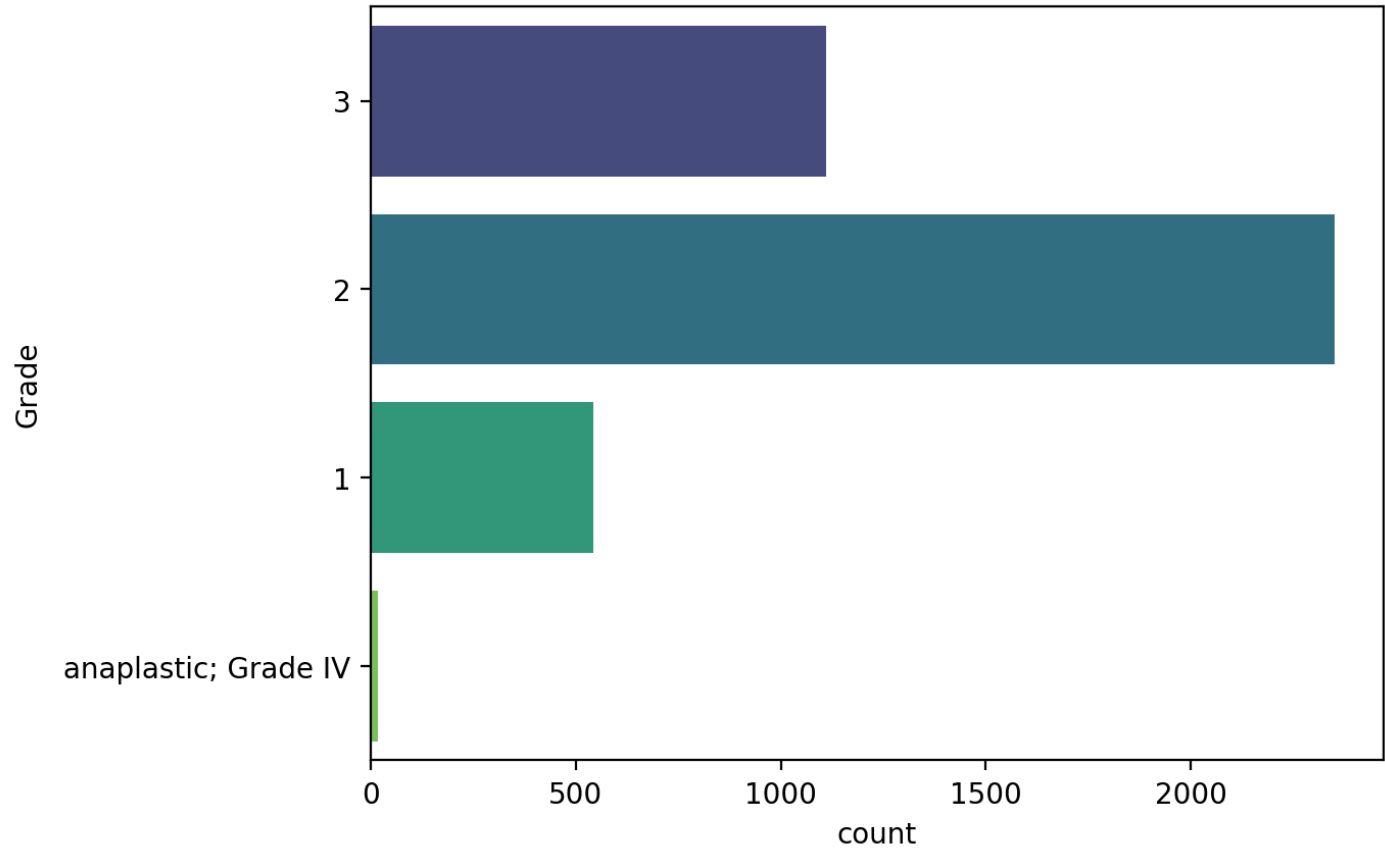


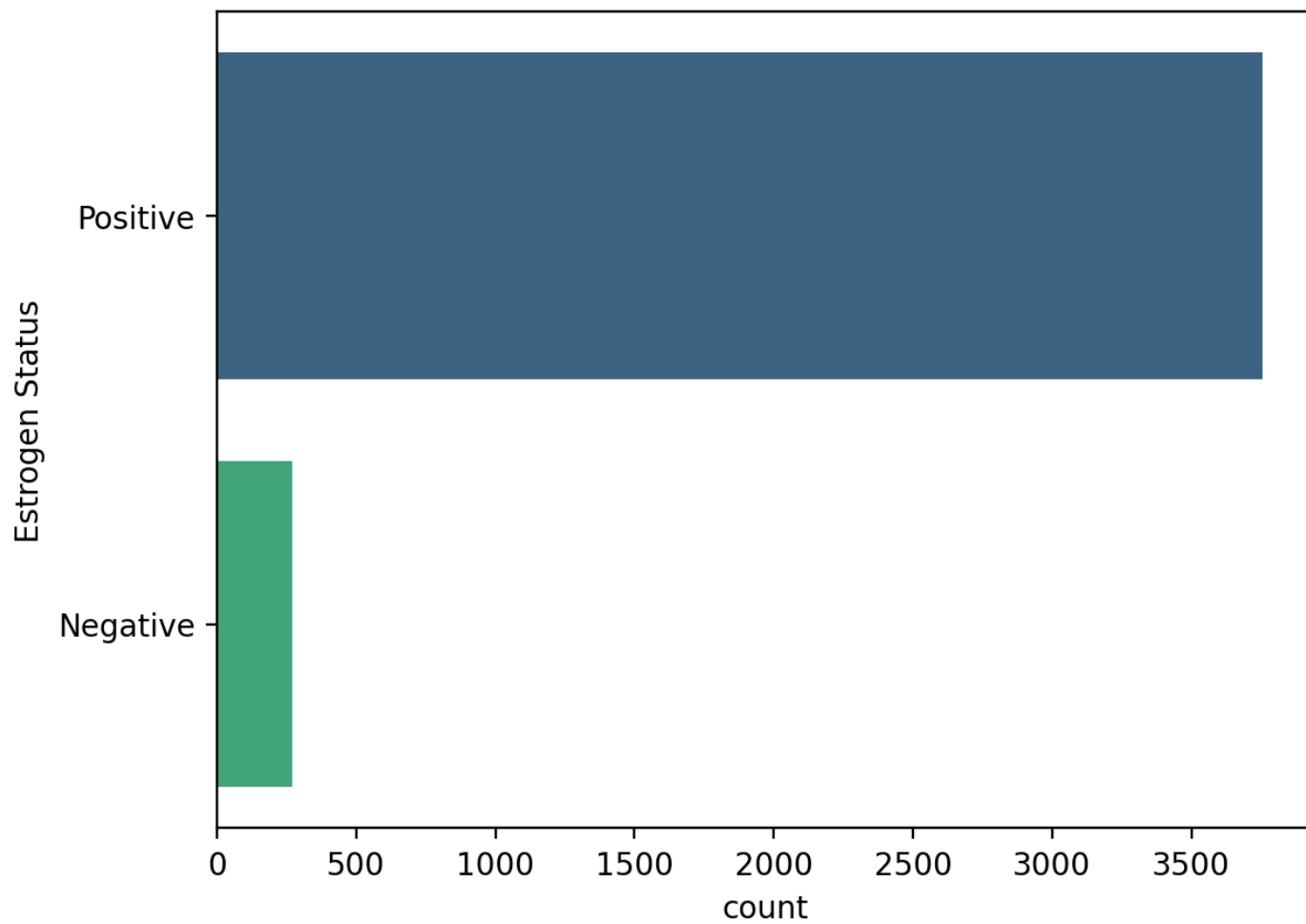


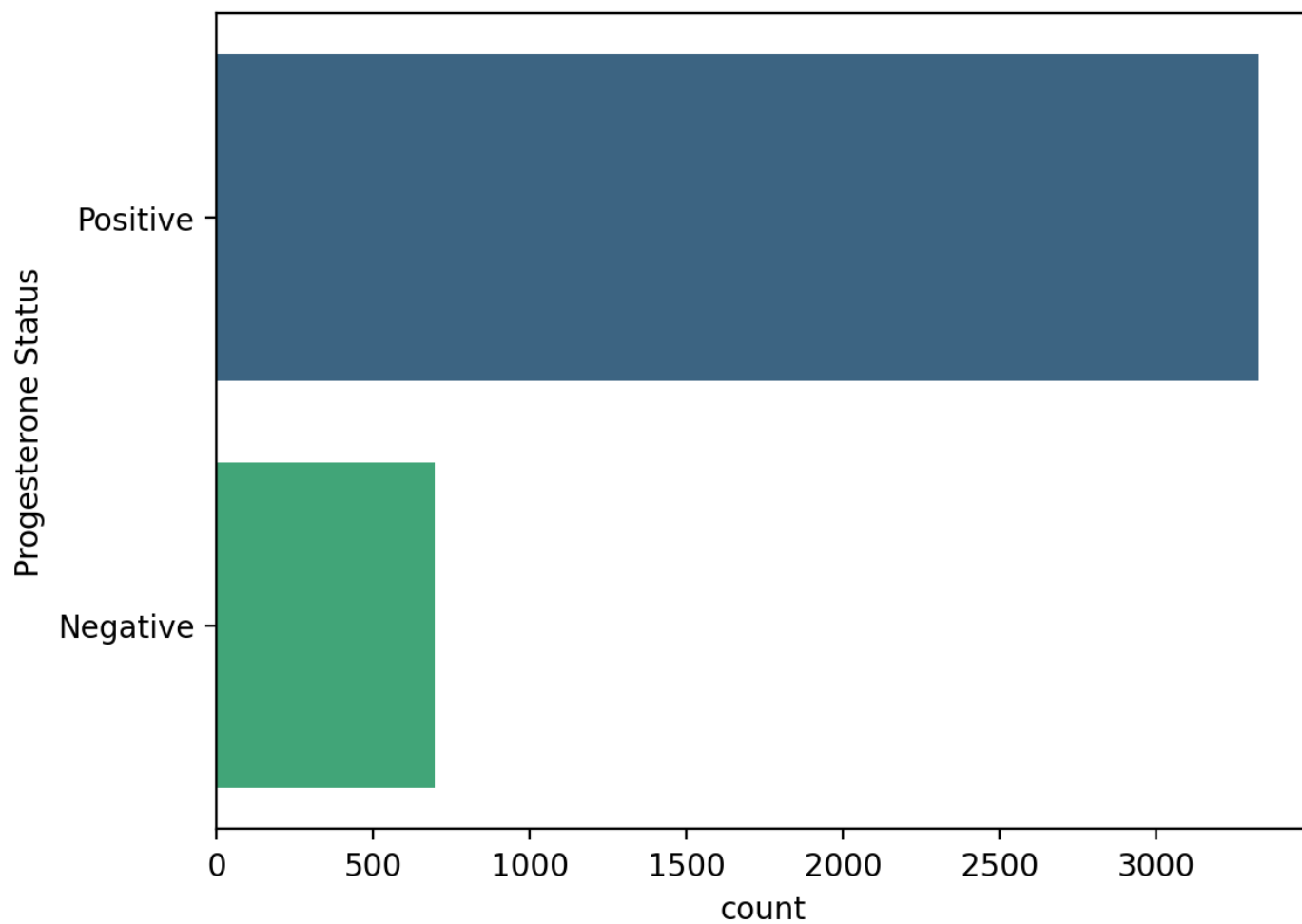


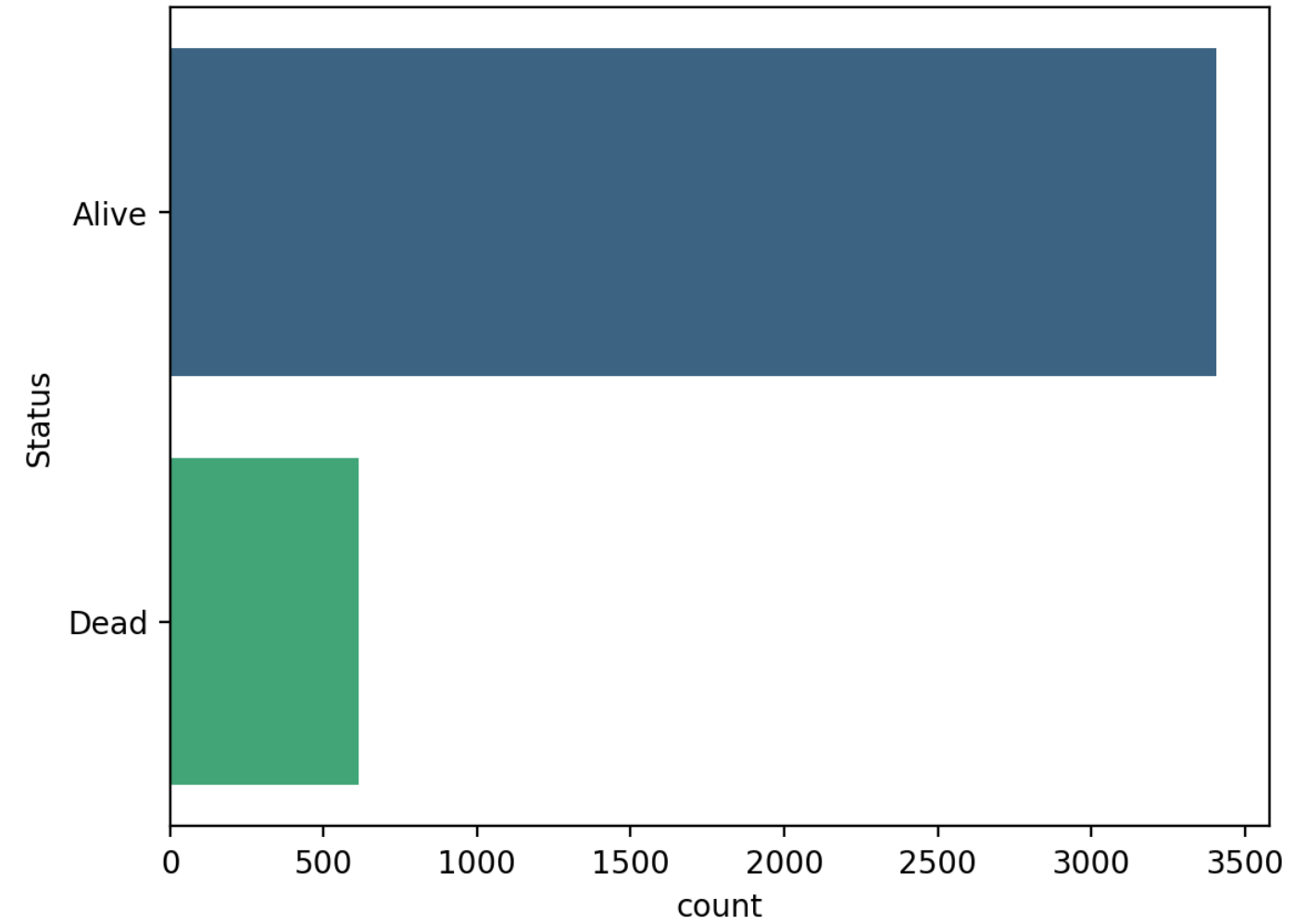












Data Summary Statistics

	Age	Tumor Size	Regional Node Examined	Reginol Node Positive	Survival Months
count	4,024	4,024	4,024	4,024	4,024
mean	53.9722	30.4737	14.3571	4.1581	71.298
std	8.9631	21.1197	8.0997	5.1093	22.9214
min	30	1	1	1	1
25%	47	16	9	1	56
50%	54	25	14	2	73
75%	61	38	19	5	90
max	69	140	61	46	107

EDA Completed Successfully! 🎉

