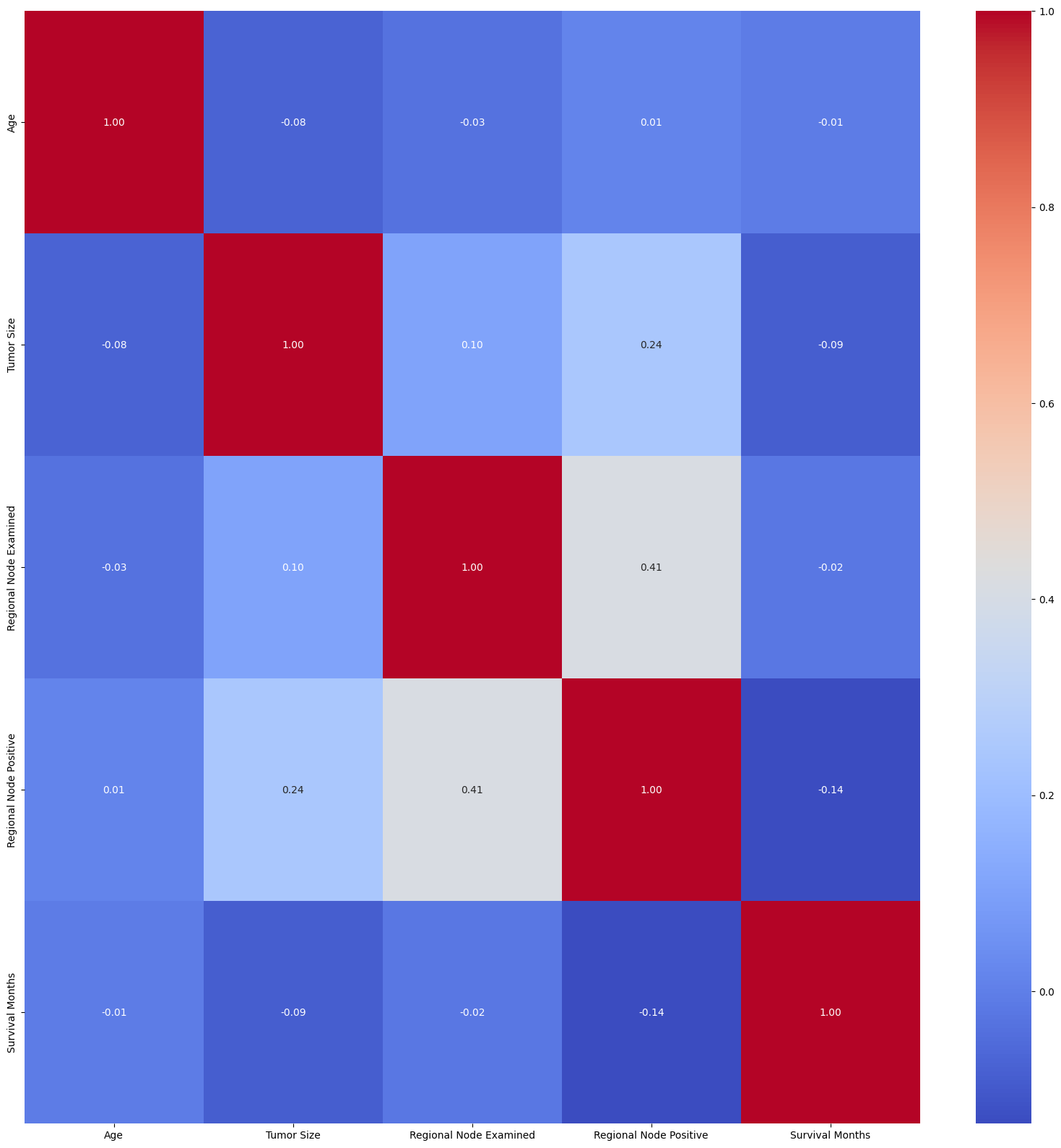
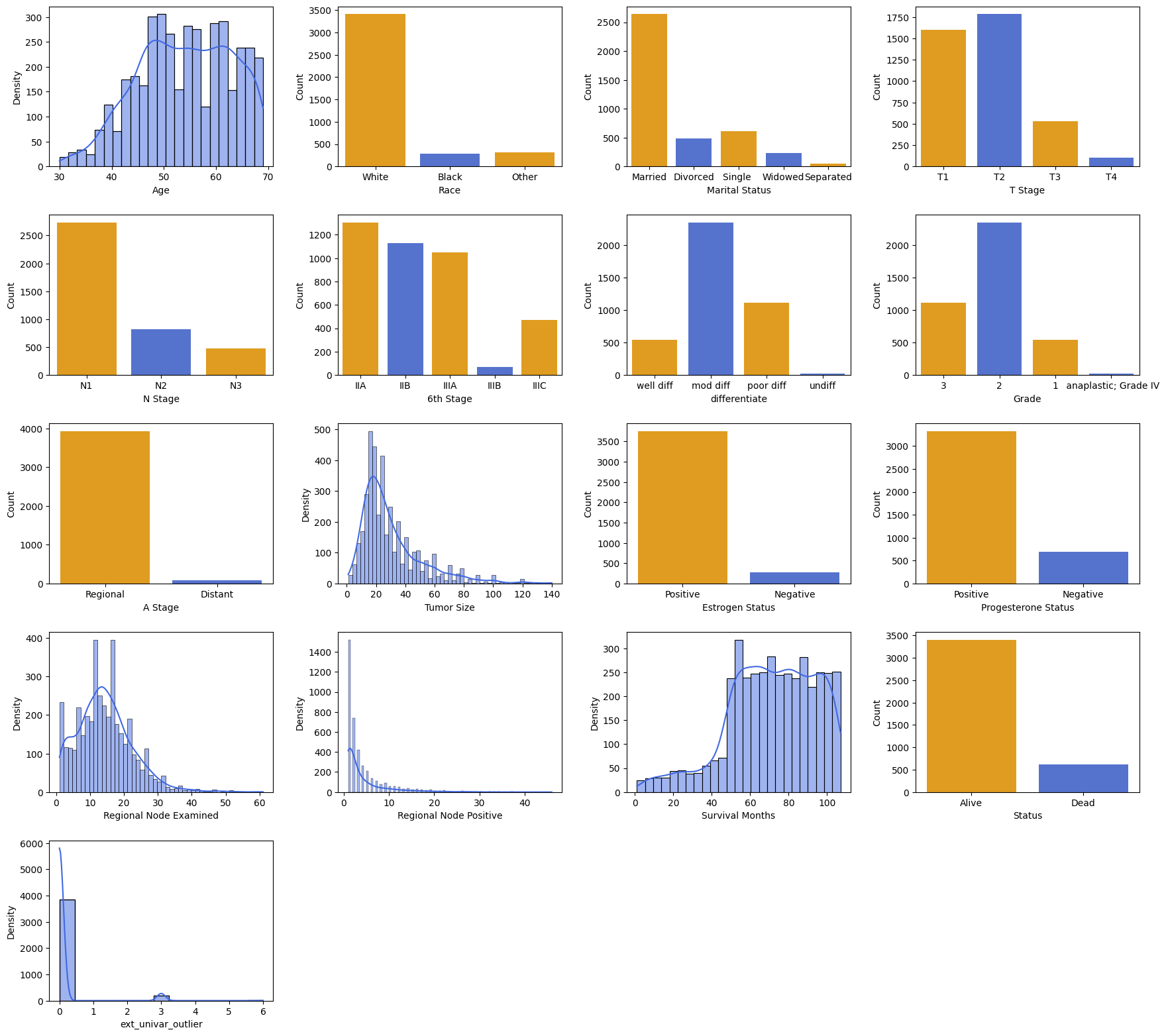
Appendix

RAW DATASET

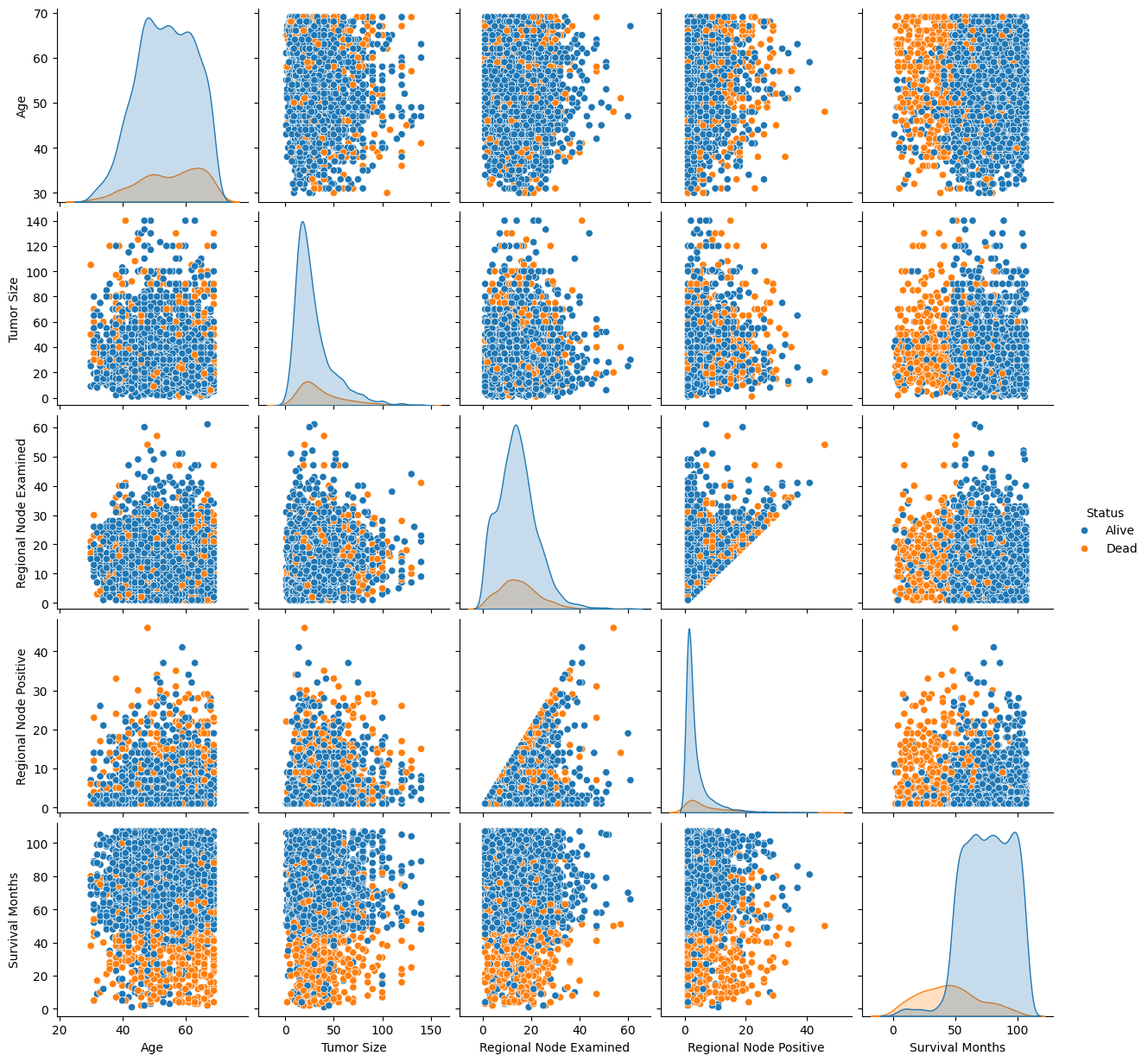
1. Correlation matrix
2. Univariate analysis



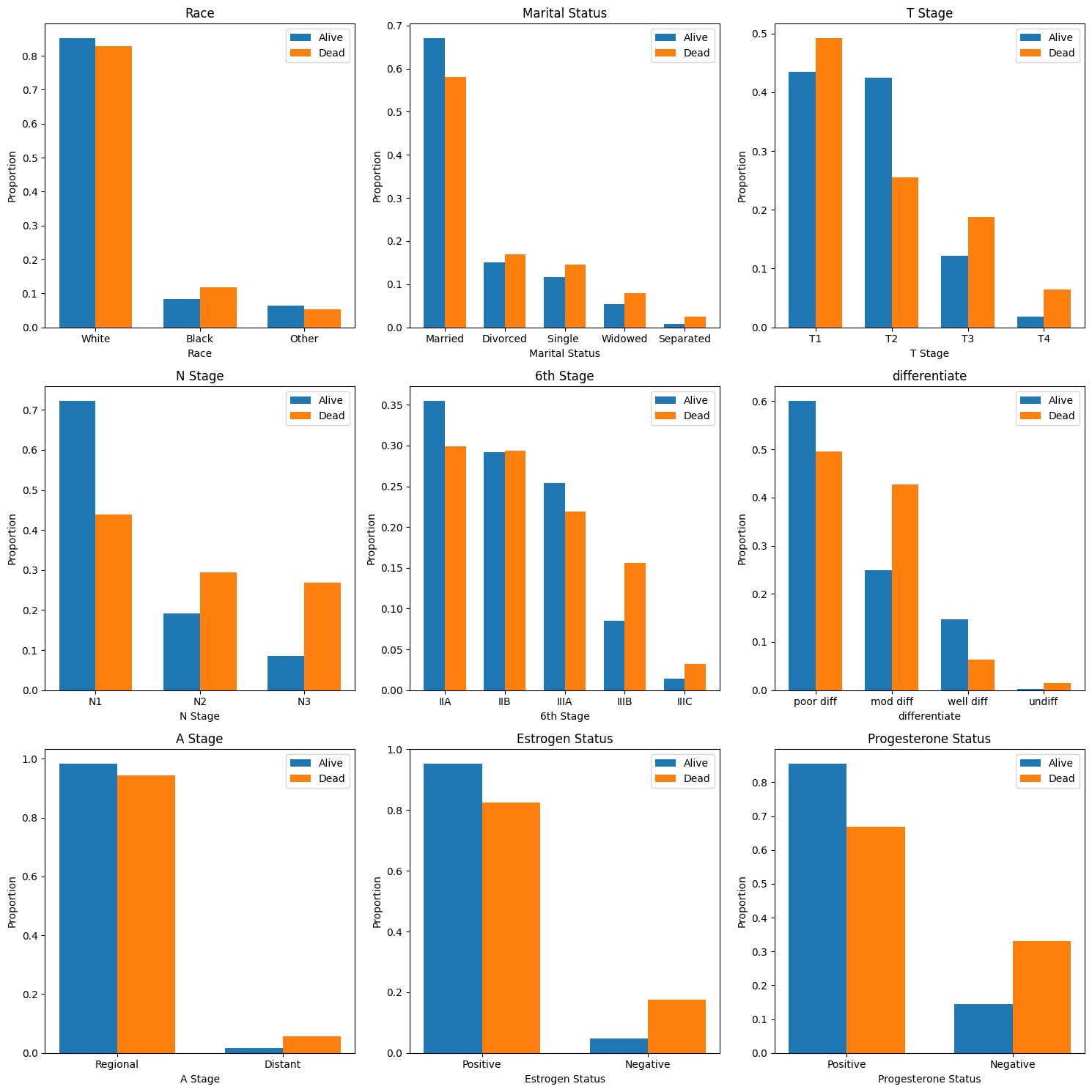
1. Univariate Outliers Analysis

| **Outlier Analysis for 'tumor\_size' Column**:  Q1 (25th percentile): 16.00  Q3 (75th percentile): 38.00  Interquartile Range (IQR): 22.00  Lower Bound for Outliers: -17.00  Upper Bound for Outliers: 71.00  Lower Bound for Extreme Outliers: -83.00  Upper Bound for Extreme Outliers: 137.00  Number of Outliers: 222  Number of Extreme Outliers: 5 | **Outlier Analysis for 'regional\_node\_examined' Column:**  Q1 (25th percentile): 9.00  Q3 (75th percentile): 19.00  Interquartile Range (IQR): 10.00  Lower Bound for Outliers: -6.00  Upper Bound for Outliers: 34.00  Lower Bound for Extreme Outliers: -36.00  Upper Bound for Extreme Outliers: 64.00  Number of Outliers: 72  Number of Extreme Outliers: 0 |
| --- | --- |
| **Outlier Analysis for 'regional\_node\_positive' Column:**  Q1 (25th percentile): 1.00  Q3 (75th percentile): 5.00  Interquartile Range (IQR): 4.00  Lower Bound for Outliers: -5.00  Upper Bound for Outliers: 11.00  Lower Bound for Extreme Outliers: -17.00  Upper Bound for Extreme Outliers: 23.00  Number of Outliers: 344  Number of Extreme Outliers: 54 | **Outlier Analysis for 'survival\_months' Column:**  Q1 (25th percentile): 56.00  Q3 (75th percentile): 90.00  Interquartile Range (IQR): 34.00  Lower Bound for Outliers: 5.00  Upper Bound for Outliers: 141.00  Lower Bound for Extreme Outliers: -97.00  Upper Bound for Extreme Outliers: 243.00  Number of Outliers: 18  Number of Extreme Outliers: 0 |

1. Bivariate numerical



1. Bivariate categorical

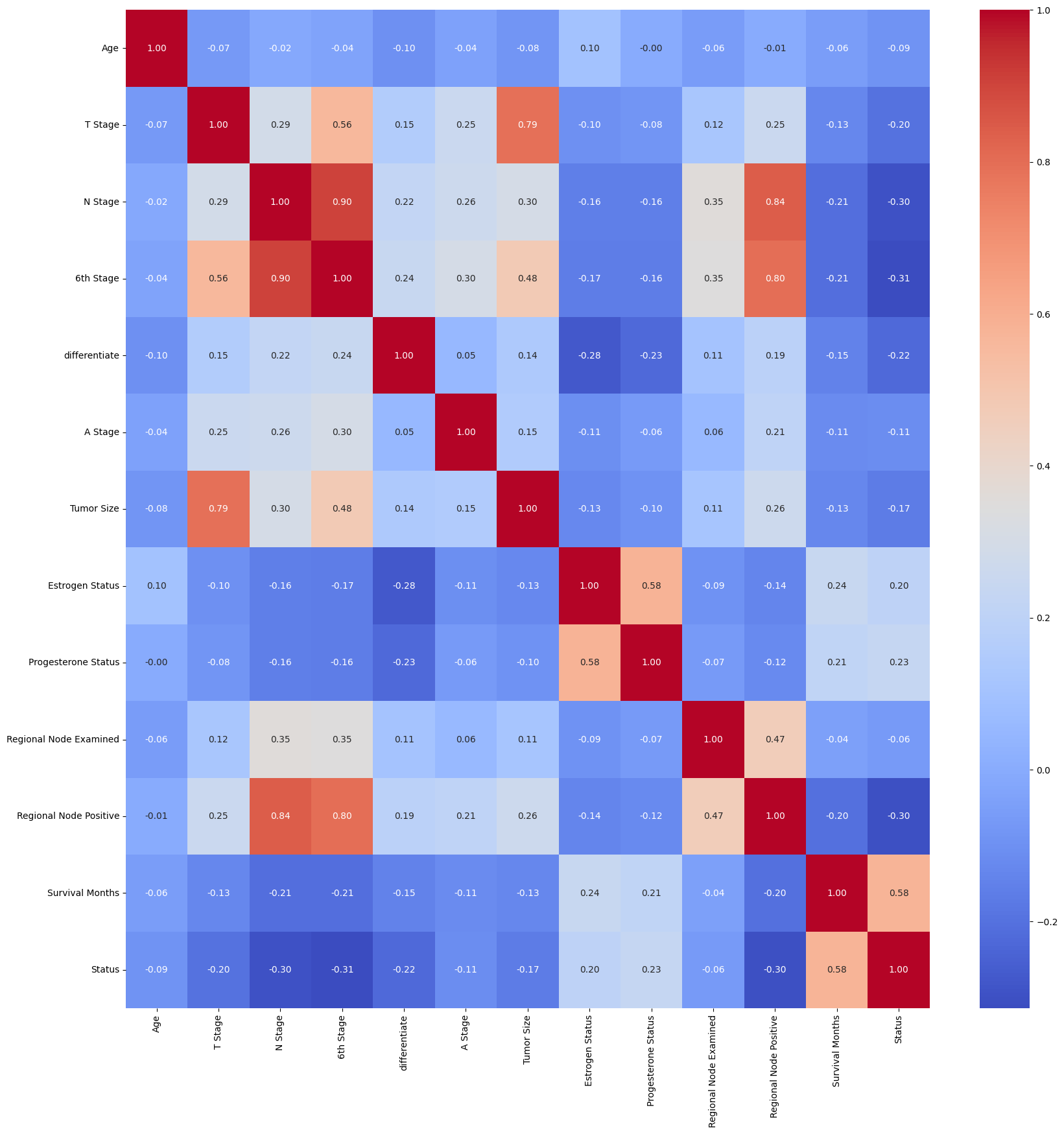


1. Multivariate Outlier Analysis

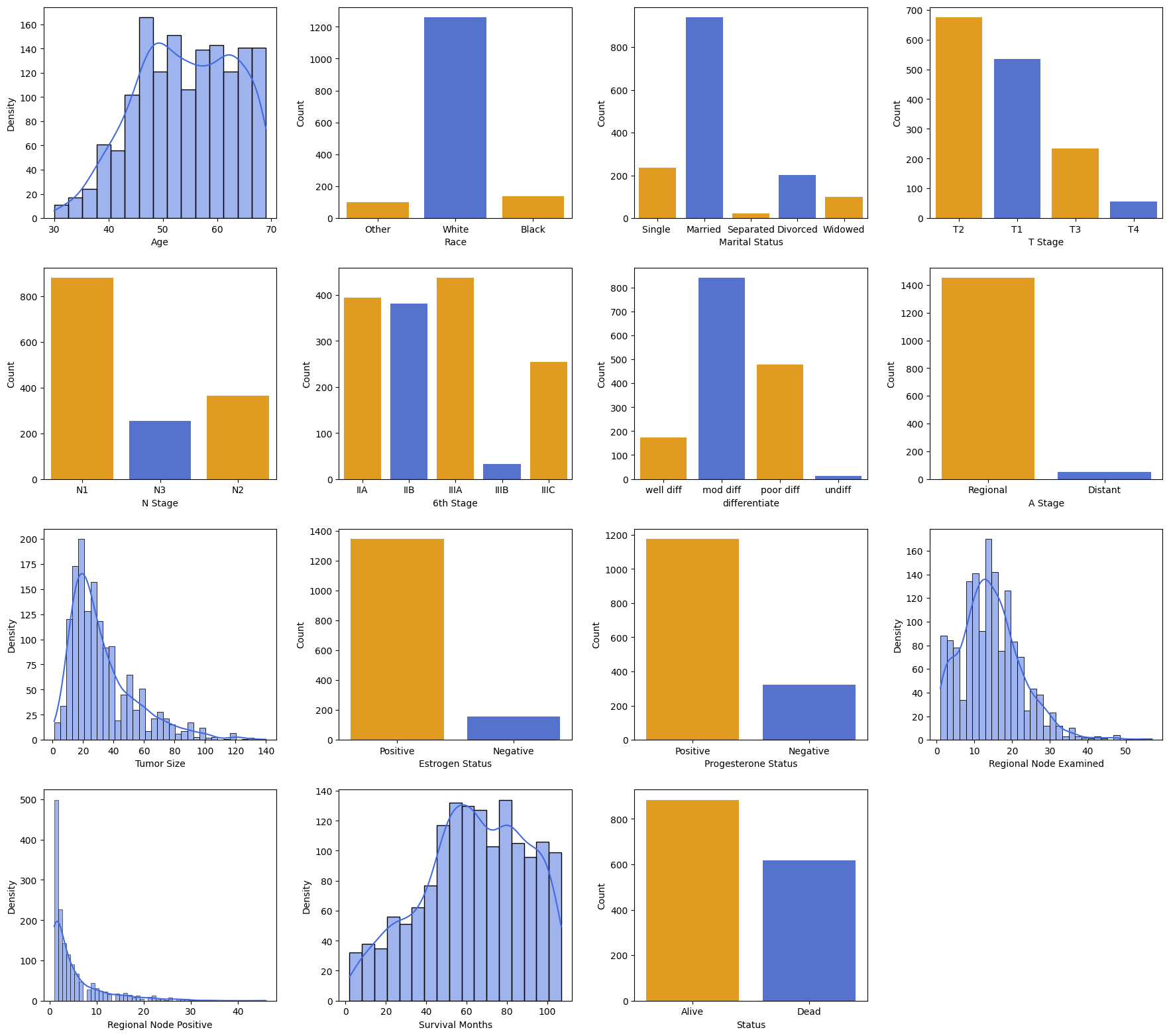
| **Variable\_1** | **Variable\_2** | **Chi2\_Statistic** | **p-value** |
| --- | --- | --- | --- |
| race | marital\_status | 137649912 | 7,31E-20 |
| race | t\_stage | 8309501 | 2,16E+05 |
| race | n\_stage | 6215014 | 1,84E+05 |
| race | 6th\_stage | 8923983 | 3,49E+05 |
| race | differentiate | 27855377 | 1,00E+02 |
| race | a\_stage | 0.303560 | 8,59E+05 |
| race | estrogen\_status | 13335971 | 1,27E+03 |
| race | progesterone\_status | 4982396 | 8,28E+04 |
| race | status | 27992525 | 8,35E-01 |
| marital\_status | t\_stage | 17496399 | 1,32E+05 |
| marital\_status | n\_stage | 22713668 | 3,75E+03 |
| marital\_status | 6th\_stage | 28533825 | 2,73E+04 |
| marital\_status | differentiate | 19112768 | 8,58E+04 |
| marital\_status | a\_stage | 7649645 | 1,05E+05 |
| marital\_status | estrogen\_status | 7647599 | 1,05E+05 |
| marital\_status | progesterone\_status | 10996051 | 2,66E+04 |
| marital\_status | status | 28472544 | 1,00E+01 |
| t\_stage | n\_stage | 321169353 | 2,37E-60 |
| t\_stage | 6th\_stage | 6758500577 | 0.000000e+00 |
| t\_stage | differentiate | 91921207 | 6,70E-10 |
| t\_stage | a\_stage | 588843504 | 2,64E-121 |
| t\_stage | estrogen\_status | 19987001 | 1,71E+02 |
| t\_stage | progesterone\_status | 14357154 | 2,46E+03 |
| t\_stage | status | 104268717 | 1,88E-16 |
| n\_stage | 6th\_stage | 6680736947 | 0.000000e+00 |
| n\_stage | differentiate | 115084469 | 1,75E-16 |
| n\_stage | a\_stage | 355992419 | 4,98E-72 |
| n\_stage | estrogen\_status | 42748359 | 5,22E-04 |
| n\_stage | progesterone\_status | 37162824 | 8,52E-03 |
| n\_stage | status | 267710959 | 7,37E-53 |
| 6th\_stage | differentiate | 157814744 | 1,46E-21 |
| 6th\_stage | a\_stage | 734944227 | 9,45E-152 |
| 6th\_stage | estrogen\_status | 52538339 | 1,06E-04 |
| 6th\_stage | progesterone\_status | 42945757 | 1,06E-02 |
| 6th\_stage | status | 279828664 | 2,43E-53 |
| differentiate | a\_stage | 10562507 | 1,43E+04 |
| differentiate | estrogen\_status | 217262787 | 7,84E-41 |
| differentiate | progesterone\_status | 147643156 | 8,49E-26 |
| differentiate | status | 111262966 | 5,87E-18 |
| a\_stage | estrogen\_status | 15546214 | 8,05E+01 |
| a\_stage | progesterone\_status | 2363510 | 1,24E+05 |
| a\_stage | status | 35795768 | 2,19E-03 |
| estrogen\_status | progesterone\_status | 1053273598 | 4,73E-225 |
| estrogen\_status | status | 135274057 | 2,88E-25 |
| progesterone\_status | status | 125062415 | 4,93E-23 |

BALANCED DATASET

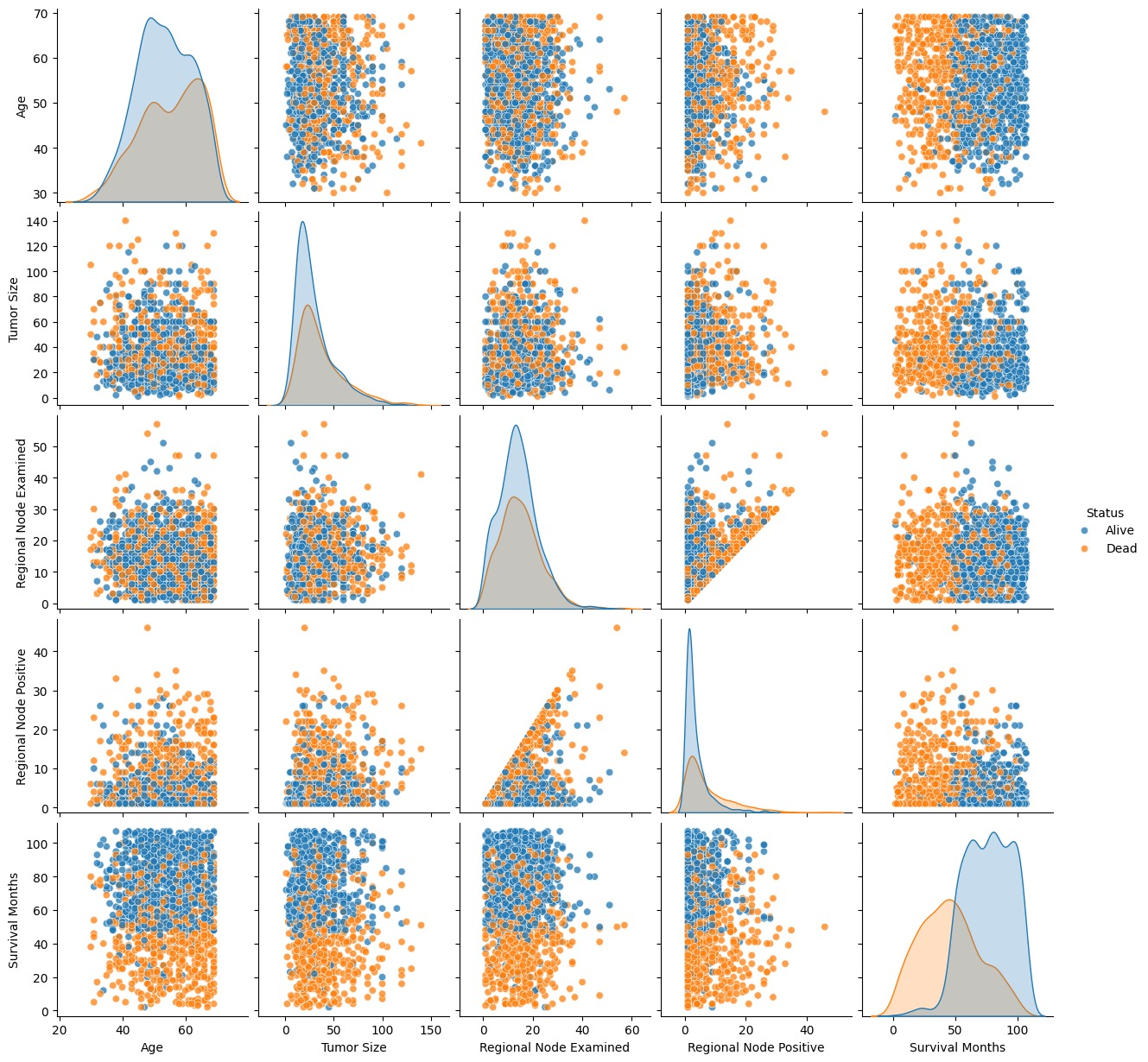
1. Correlation matrix - Balanced Dataset



1. Univariate Analysis



1. Bivariate numerical



1. Bivariate categorical

