

Dataset overview

Generated on: 2025-11-17 12:31:55 IST

Task type: classification

Target column: diagnosis

Rows: 569

Columns: 32

Numeric columns: 30

Categorical columns: 0

Overall missing: 0.00%

Model Performance (test split):

Accuracy: 0.9737

ROC AUC: 0.9944

Top Influencing Features:

perimeter_worst: 0.1431

area_worst: 0.1424

concave points_worst: 0.1130

concave points_mean: 0.0887

radius_worst: 0.0826

ID-like columns removed: 1

Top columns by missing values

id: 0 (0.0%)

diagnosis: 0 (0.0%)

radius_mean: 0 (0.0%)

texture_mean: 0 (0.0%)

perimeter_mean: 0 (0.0%)

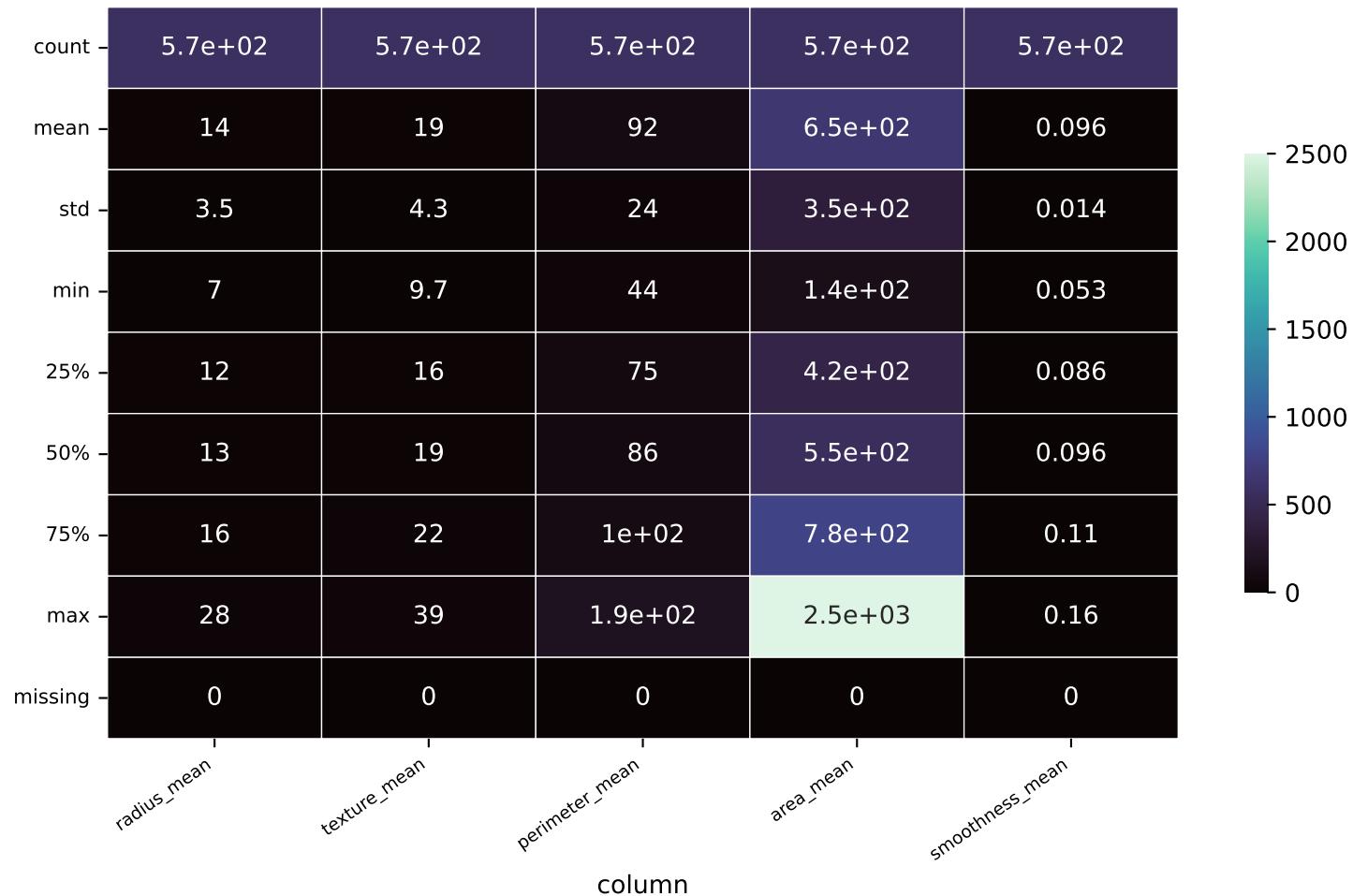
Top 5 Columns by Missing Values

column	dtype	unique	missing_count	missing_percent
id	int64	569	0	0.0
diagnosis	object	2	0	0.0
radius_mean	float64	456	0	0.0
texture_mean	float64	479	0	0.0
perimeter_mean	float64	522	0	0.0

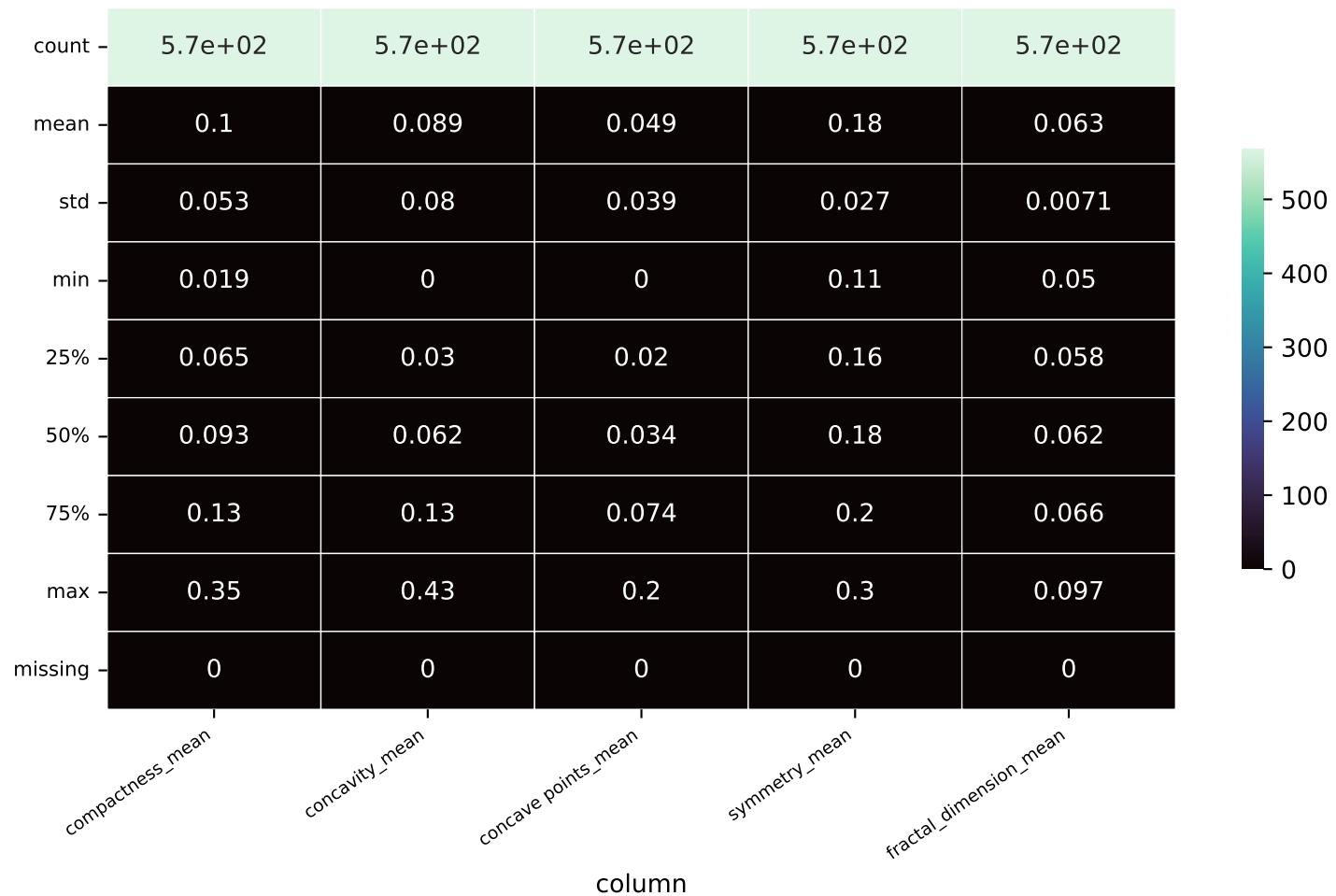
Numeric Columns

```
radius_mean | texture_mean | perimeter_mean  
area_mean | smoothness_mean | compactness_mean  
concavity_mean | concave points_mean | symmetry_mean  
fractal_dimension_mean | radius_se | texture_se  
perimeter_se | area_se | smoothness_se  
compactness_se | concavity_se | concave points_se  
symmetry_se | fractal_dimension_se | radius_worst  
texture_worst | perimeter_worst | area_worst  
smoothness_worst | compactness_worst | concavity_worst  
concave points_worst | symmetry_worst | fractal_dimension_worst
```

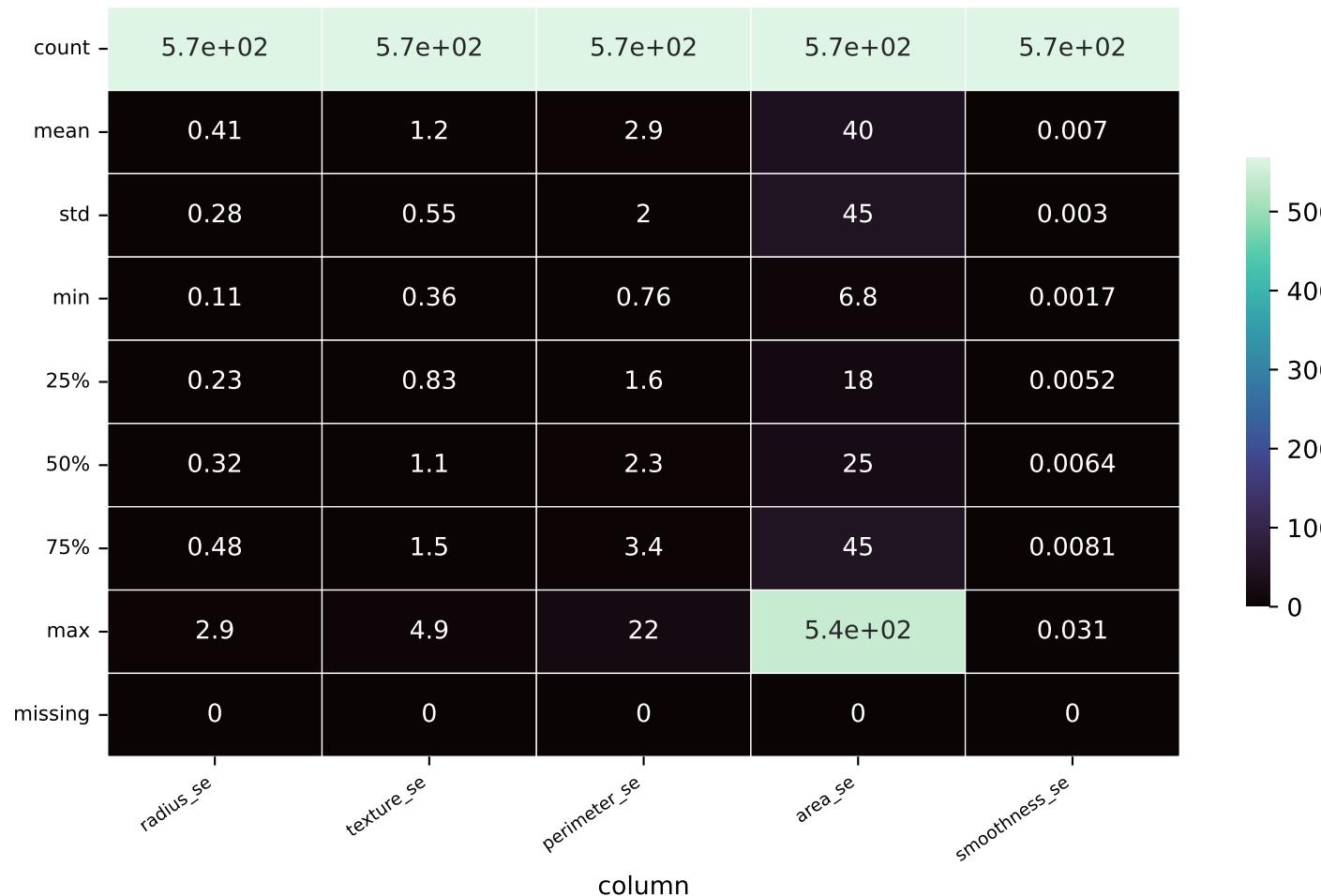
Numeric Summary (features 1-5)



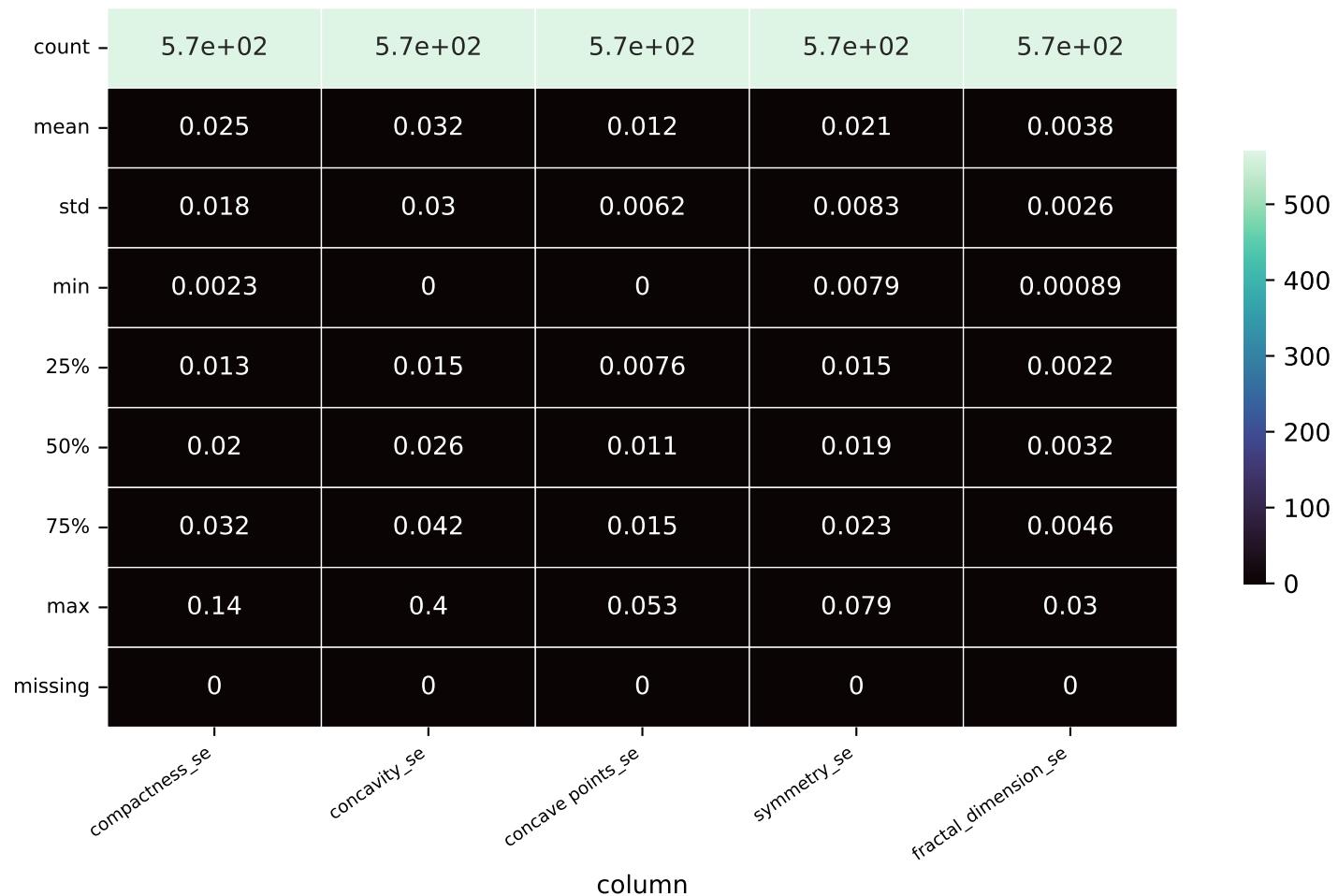
Numeric Summary (features 6-10)



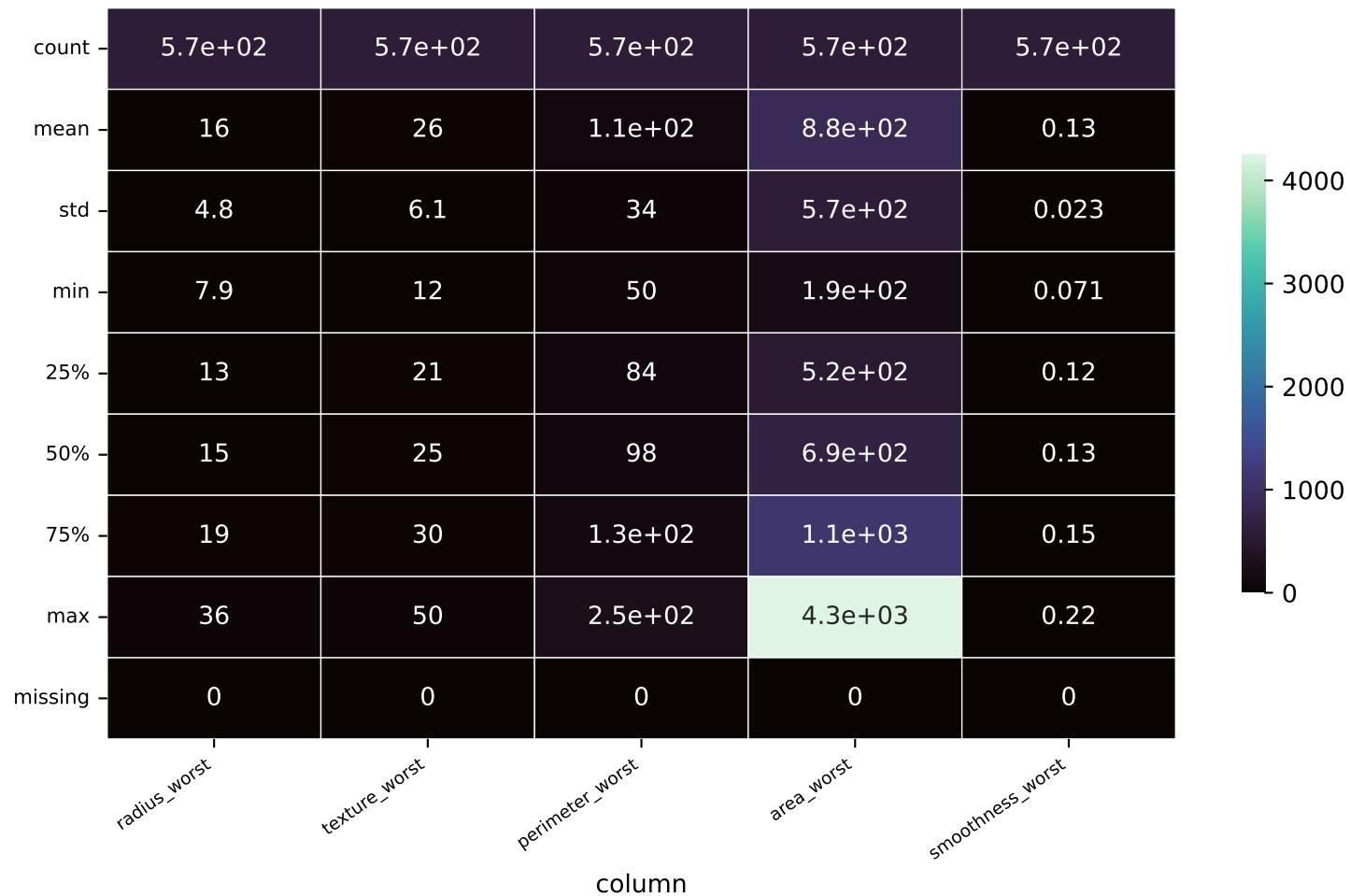
Numeric Summary (features 11-15)



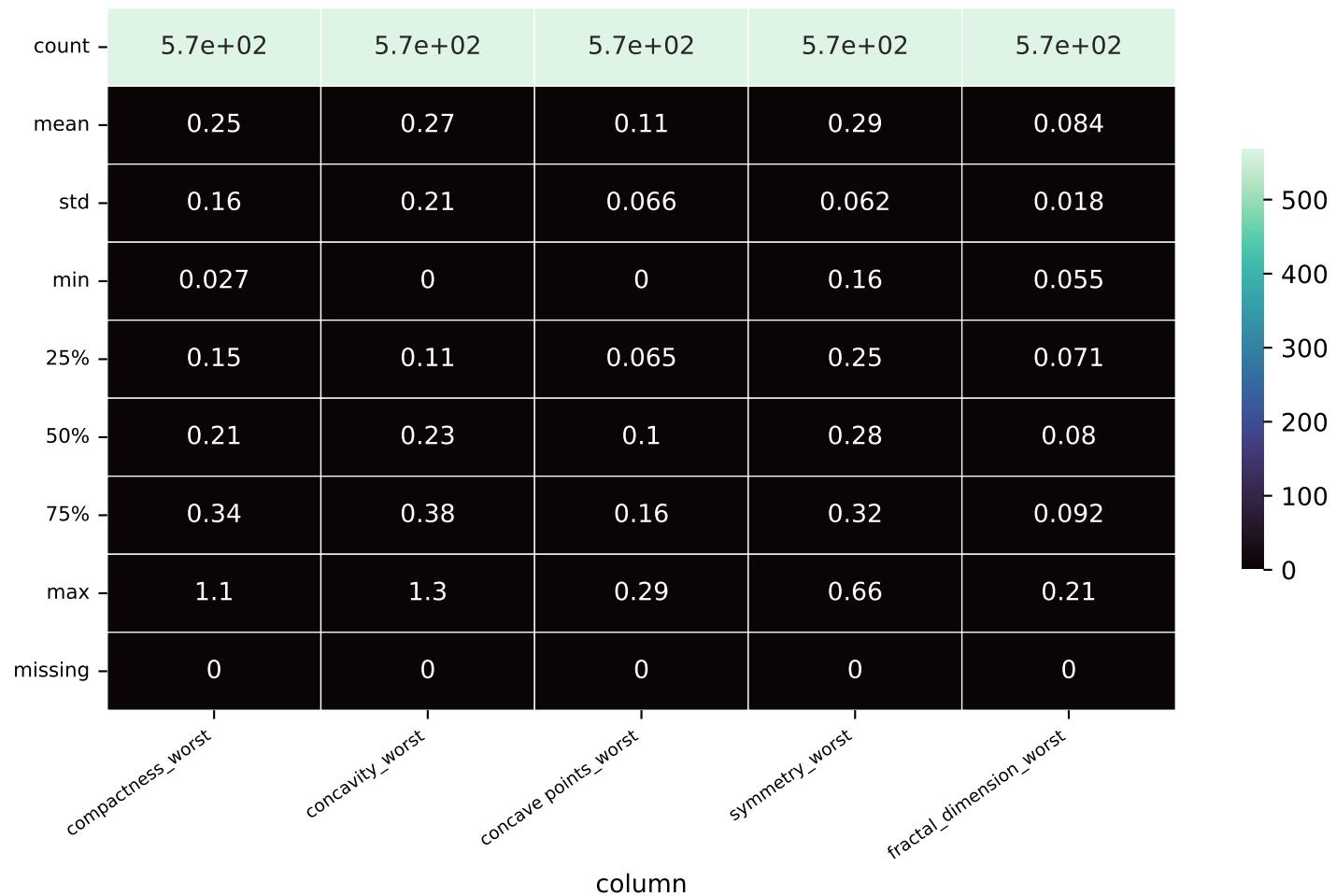
Numeric Summary (features 16-20)



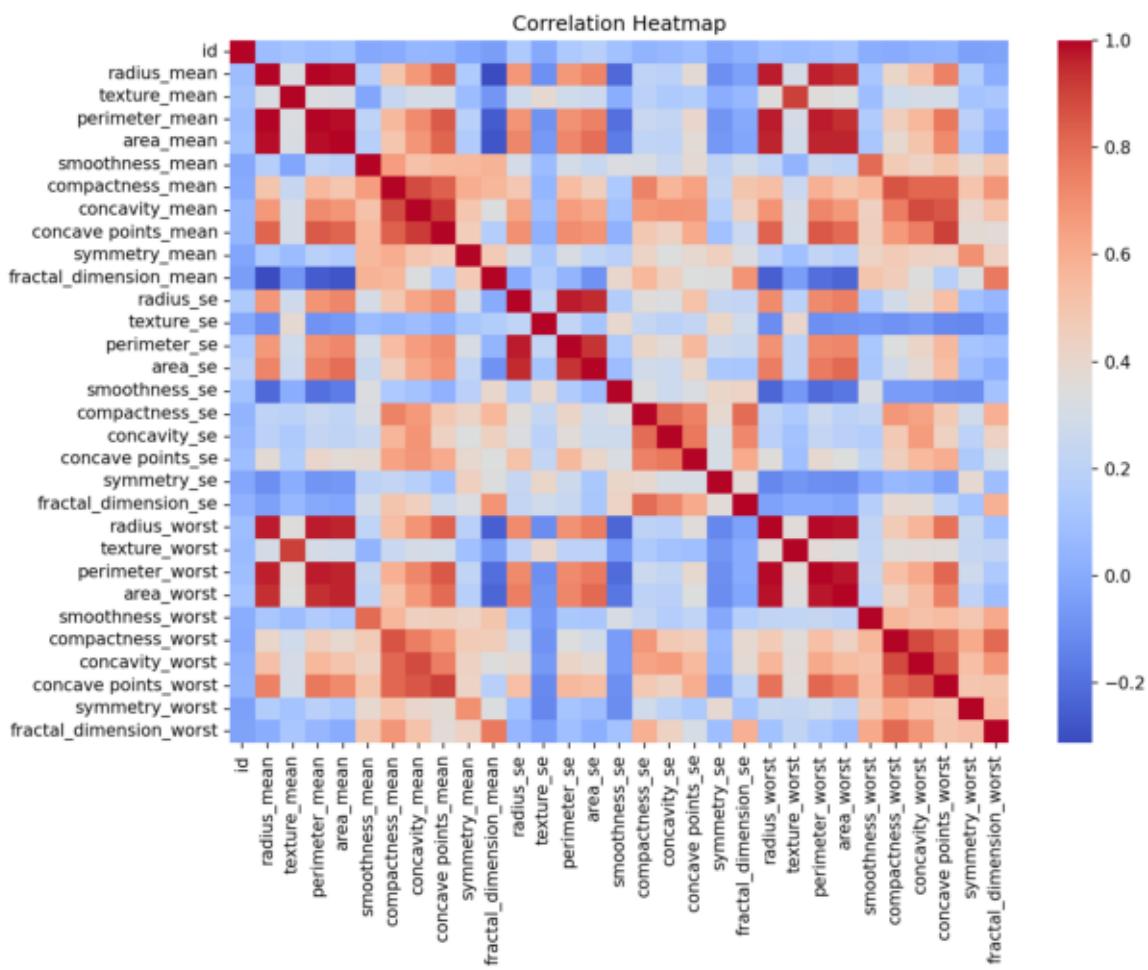
Numeric Summary (features 21-25)



Numeric Summary (features 26-30)



Correlation Heatmap



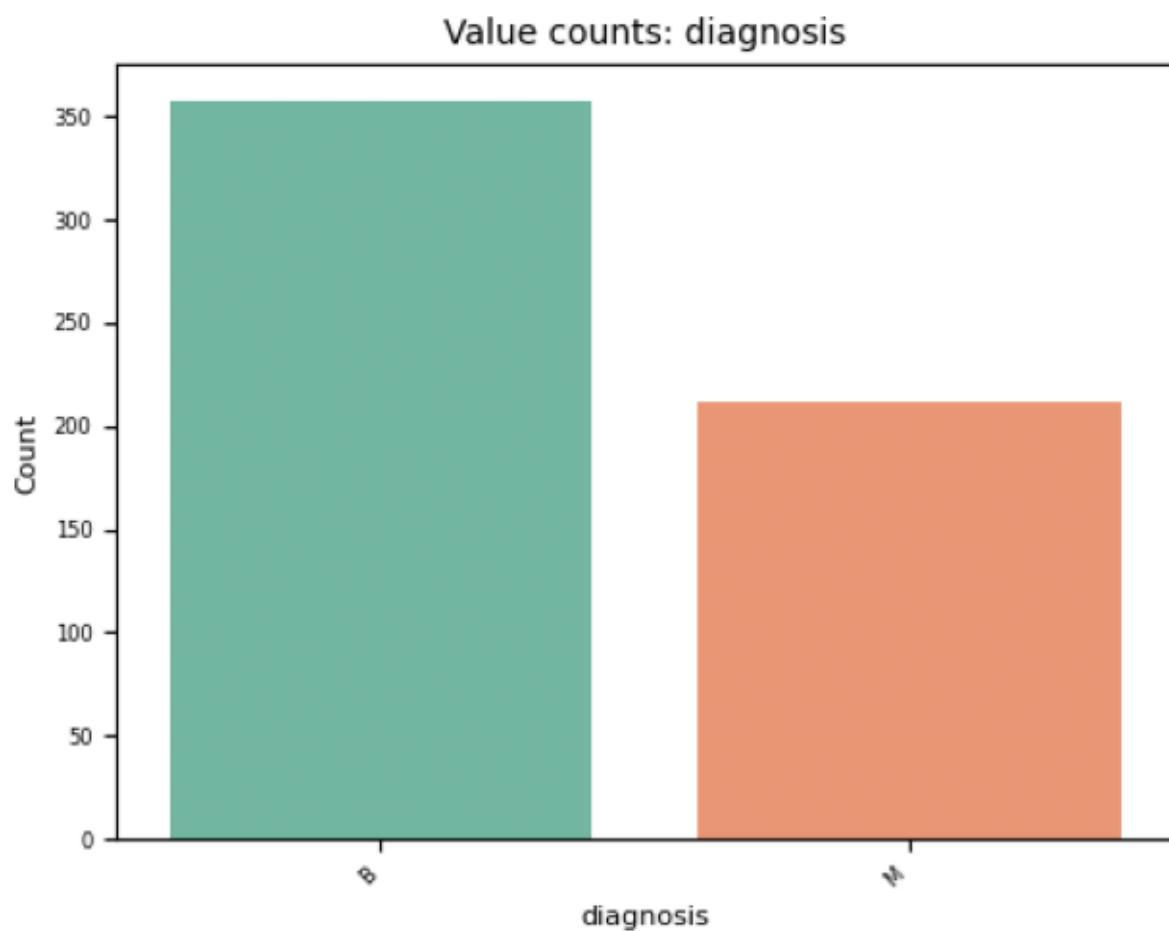
Categorical Columns

diagnosis

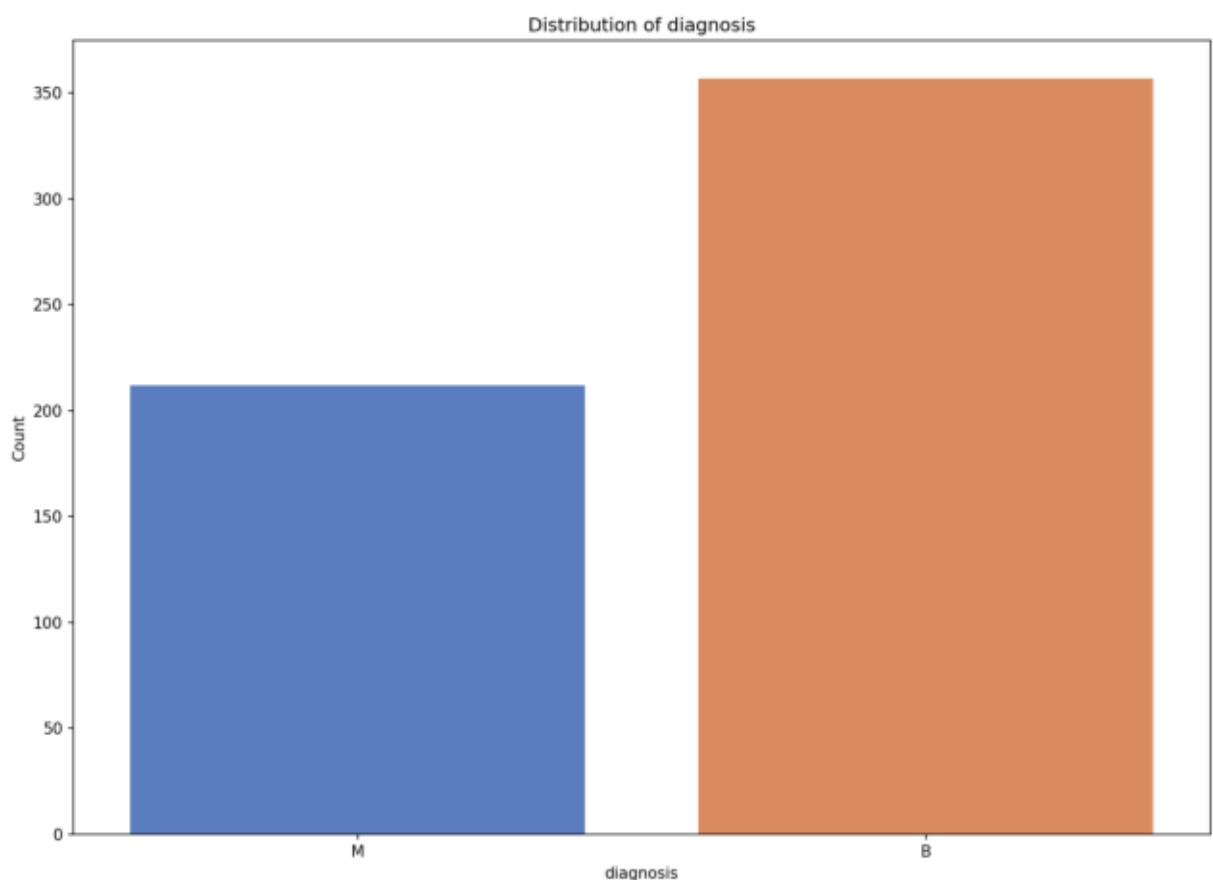
Categorical Columns Summary

column	count	unique	top	freq
diagnosis	569	2	B	357

Value counts: diagnosis

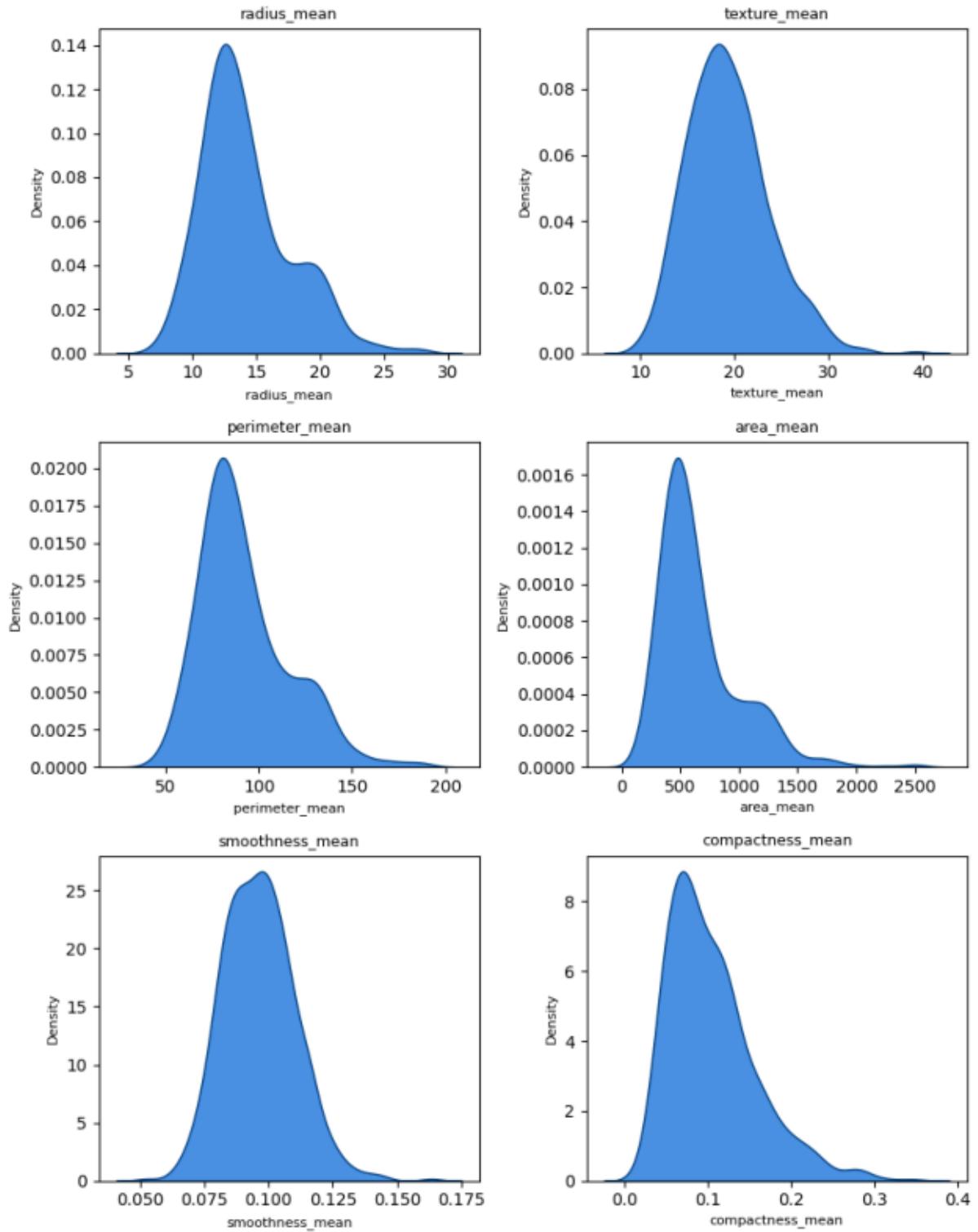


Distribution of diagnosis



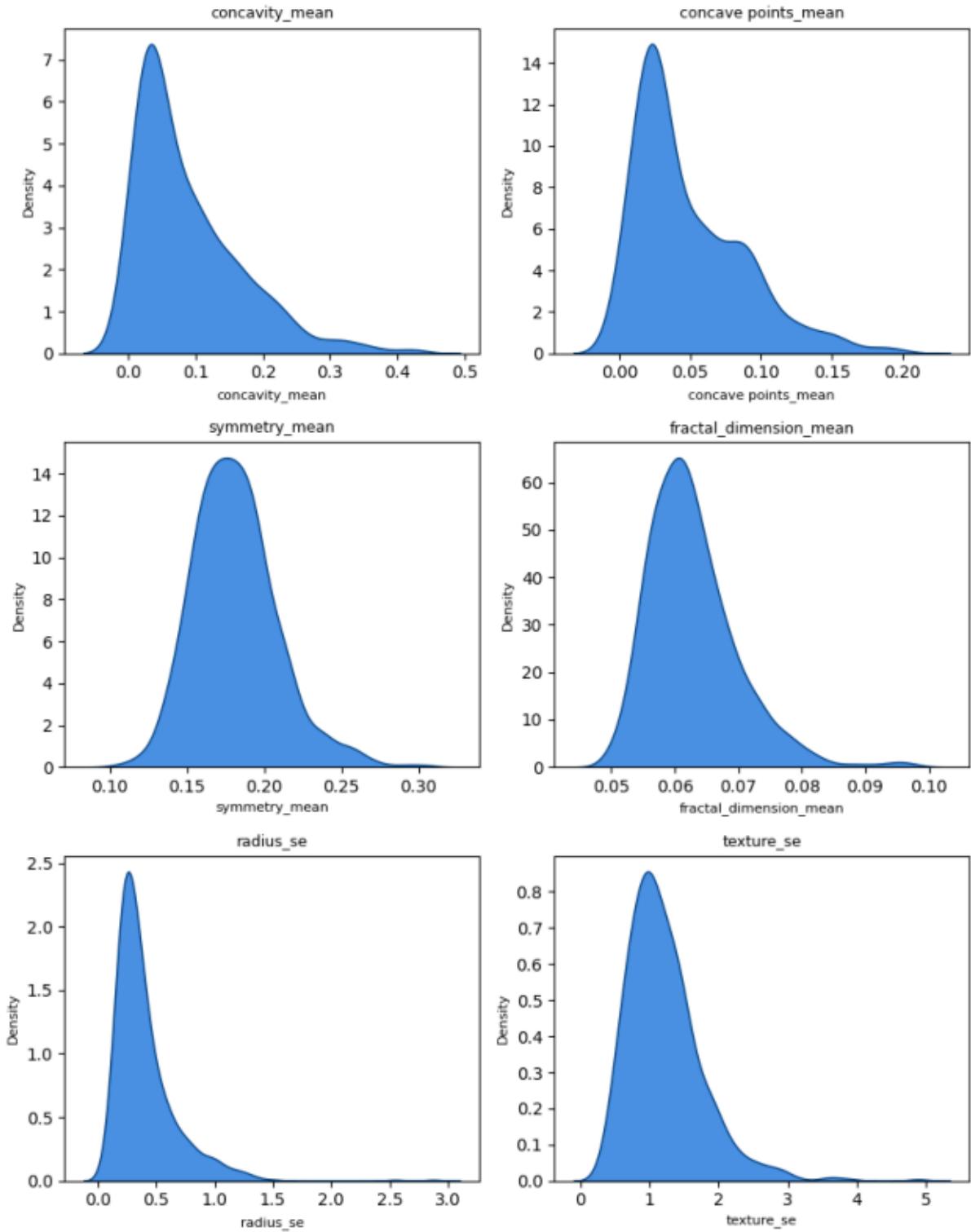
KDE Plots - Page 1

KDE/Scatter Plots



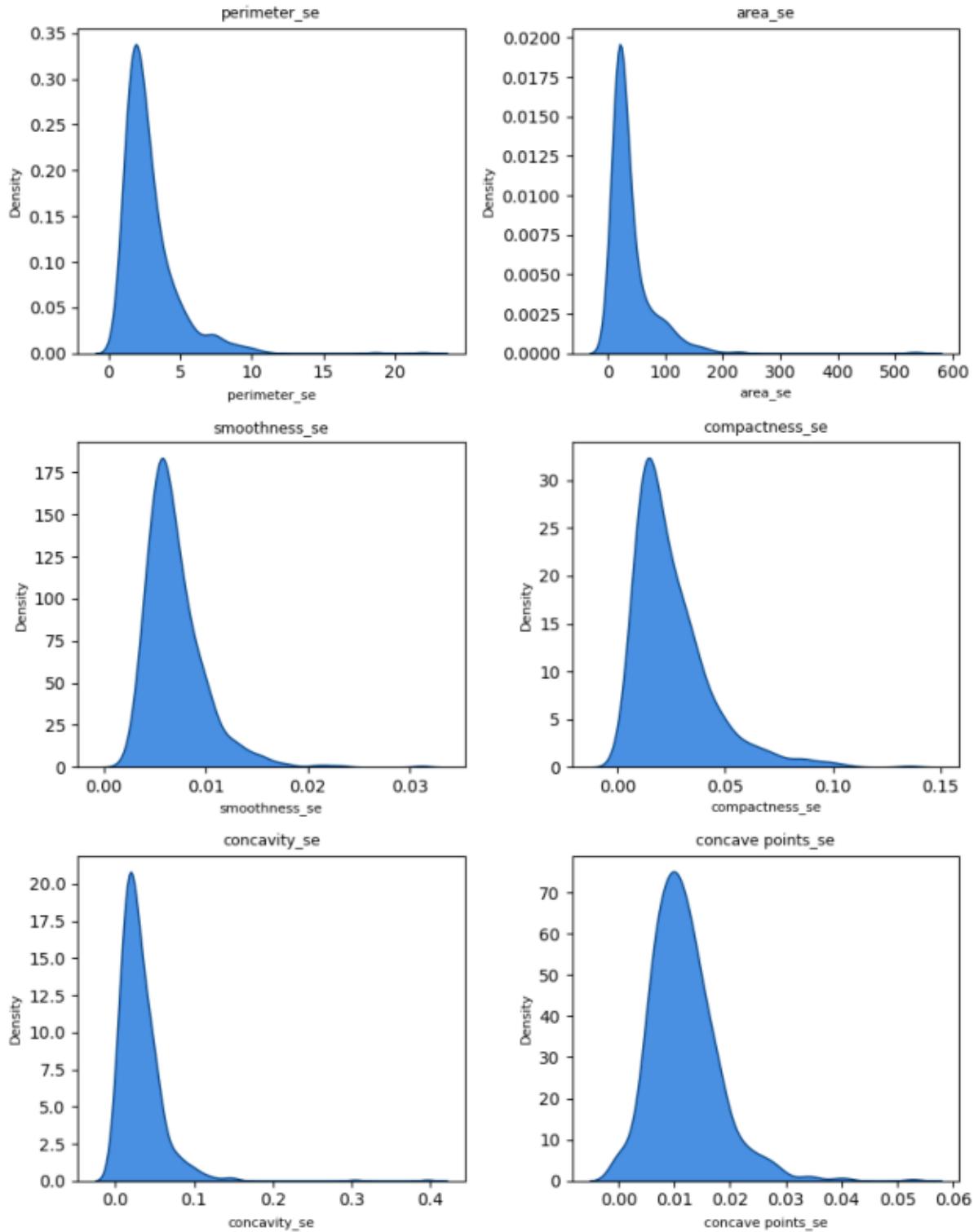
KDE Plots - Page 2

KDE/Scatter Plots



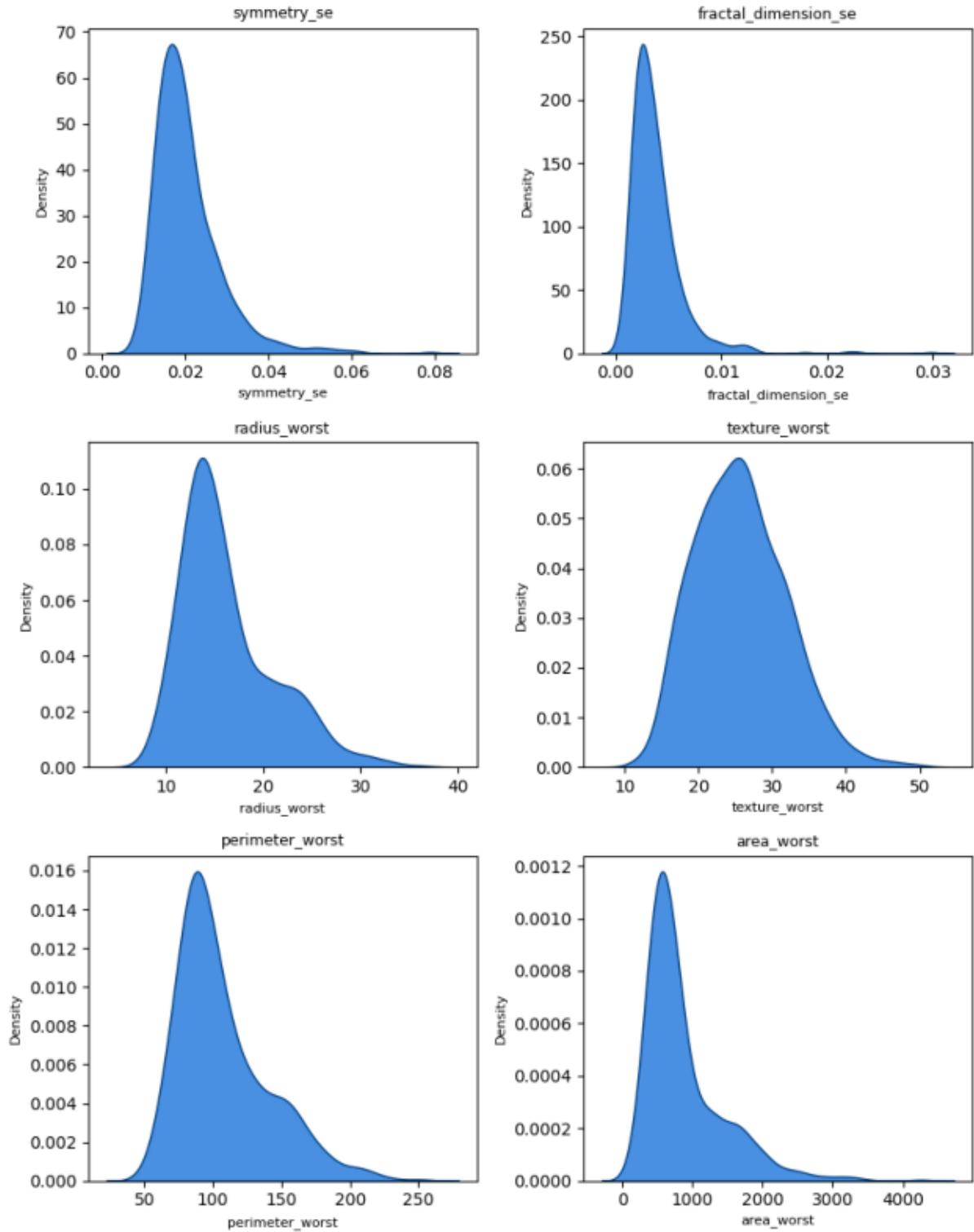
KDE Plots - Page 3

KDE/Scatter Plots



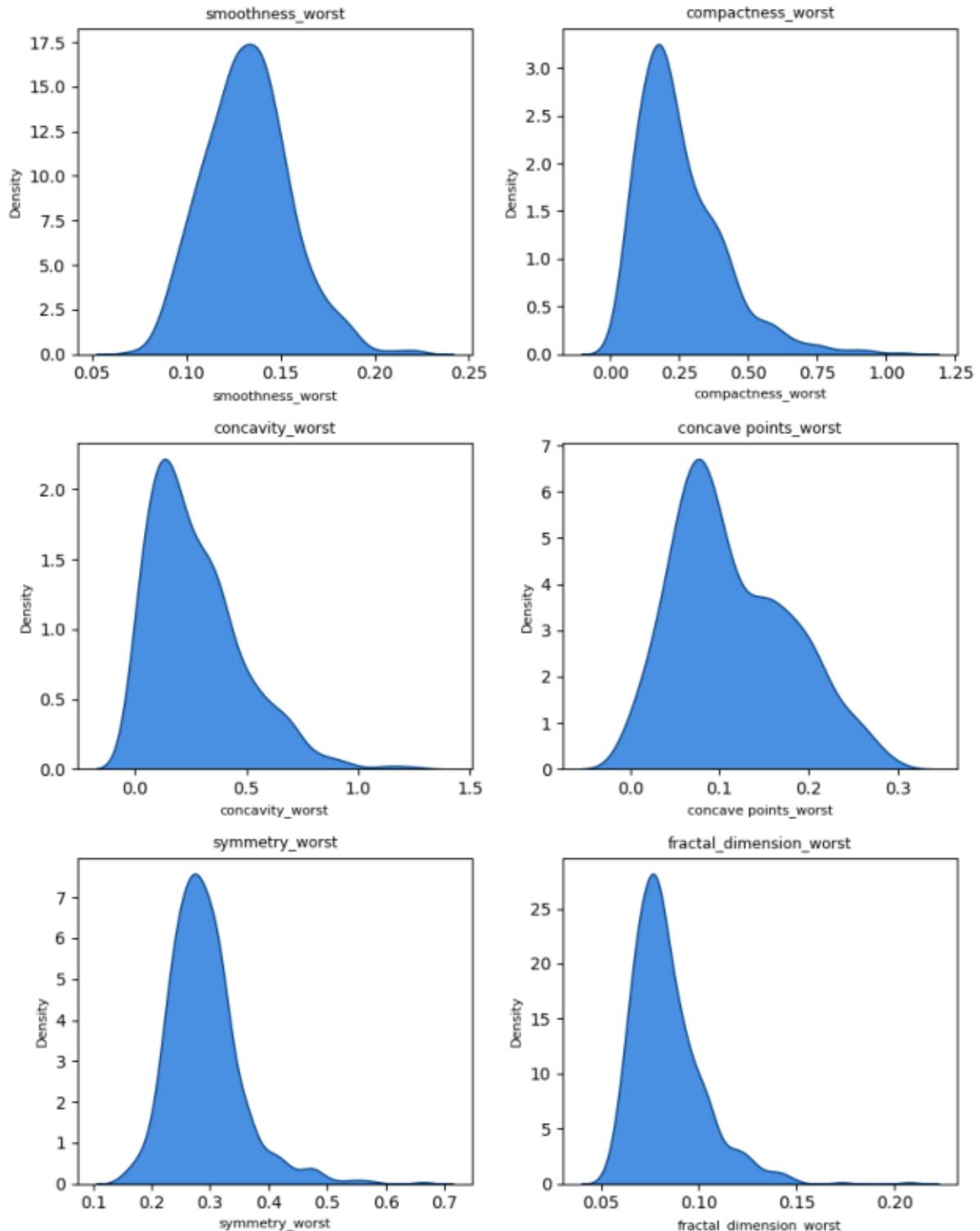
KDE Plots - Page 4

KDE/Scatter Plots

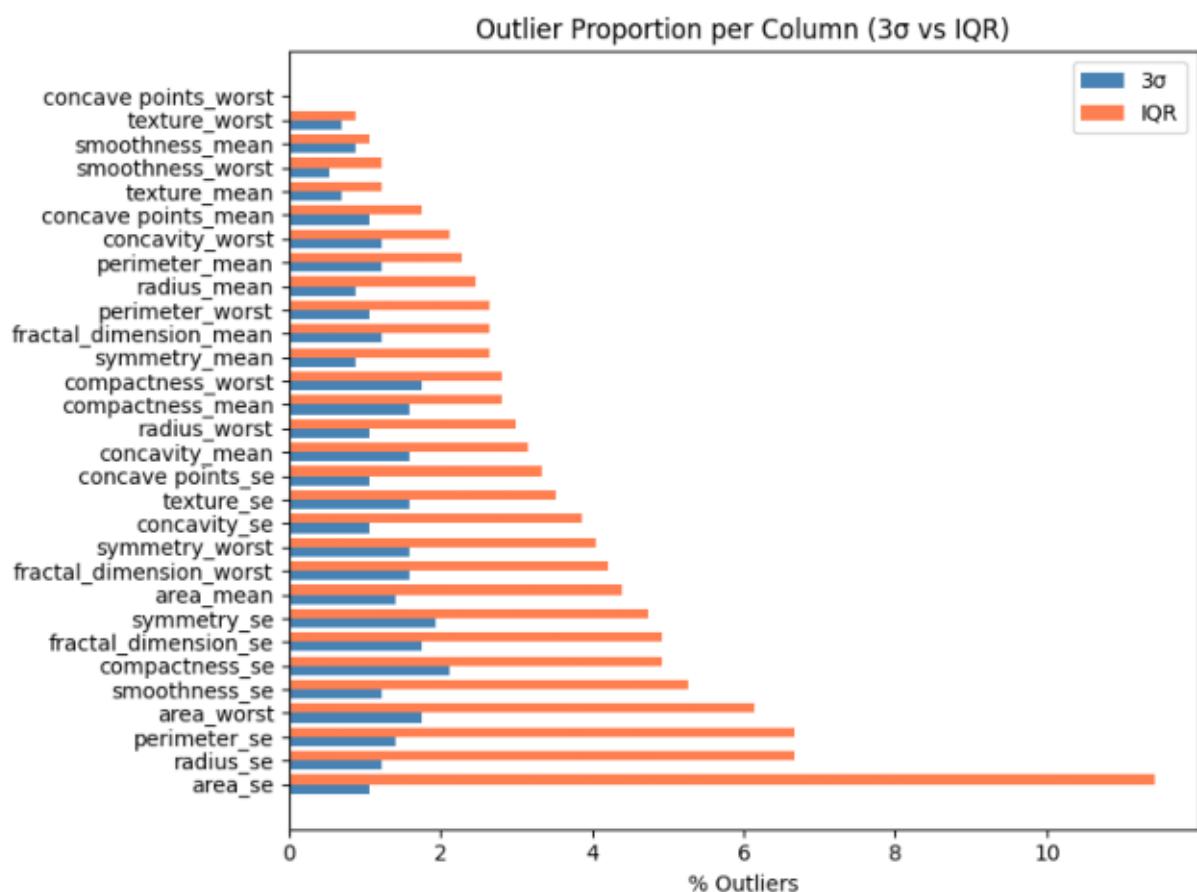


KDE Plots - Page 5

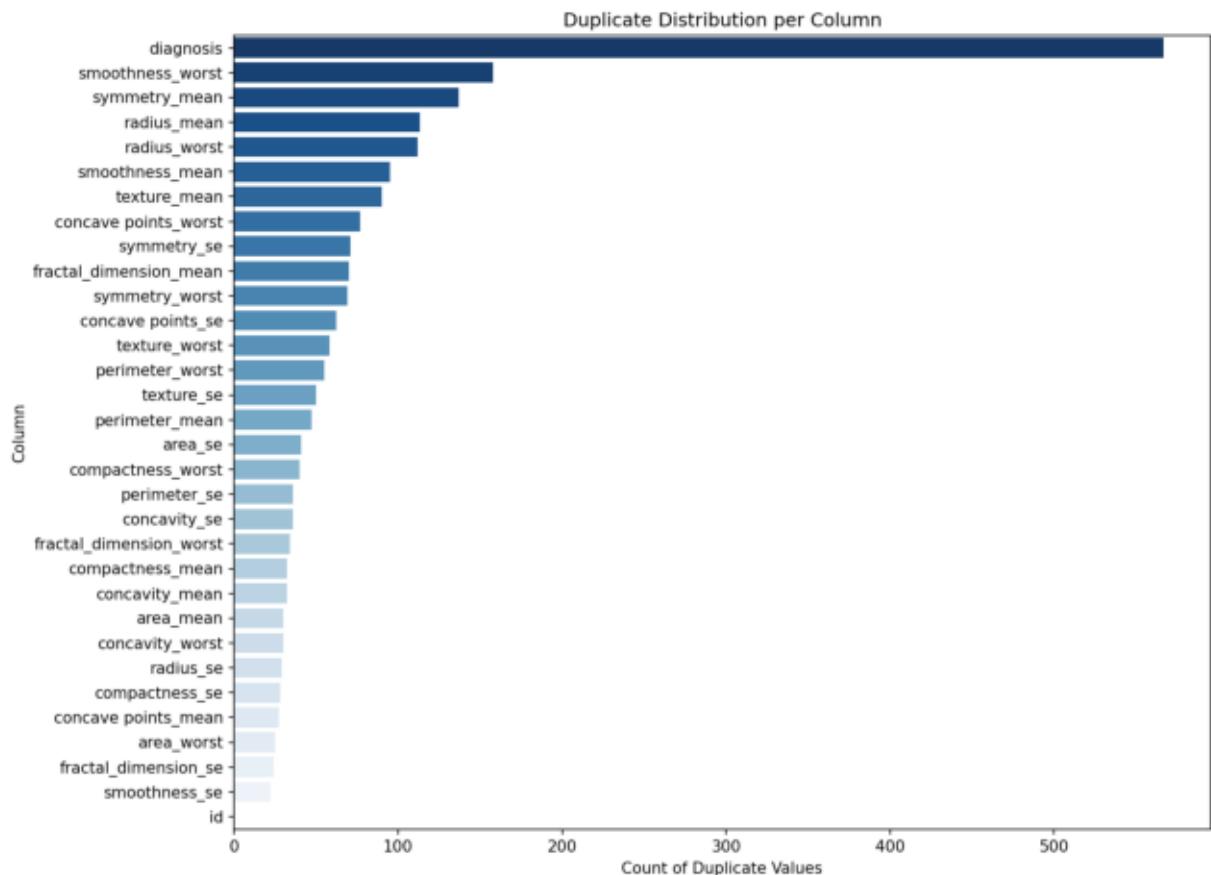
KDE/Scatter Plots



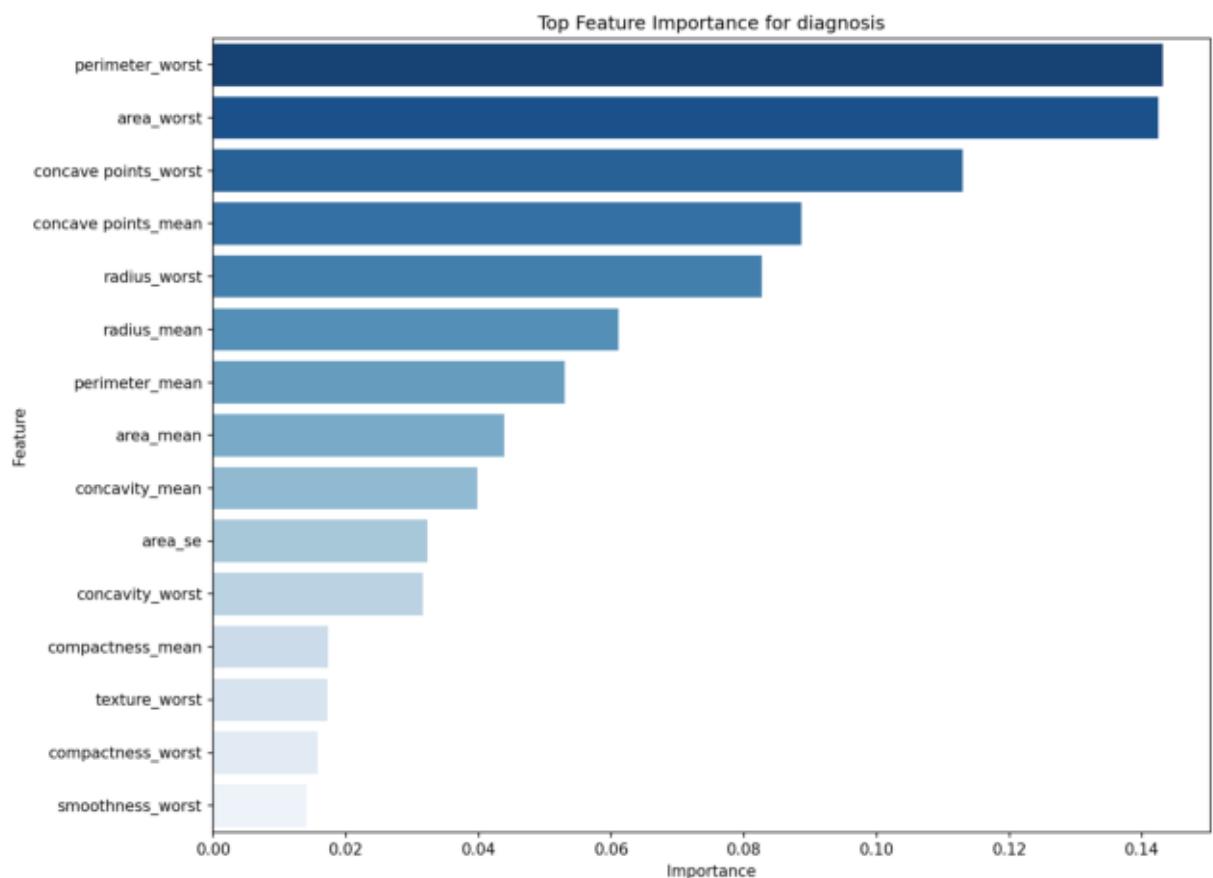
Outlier Proportion per Column



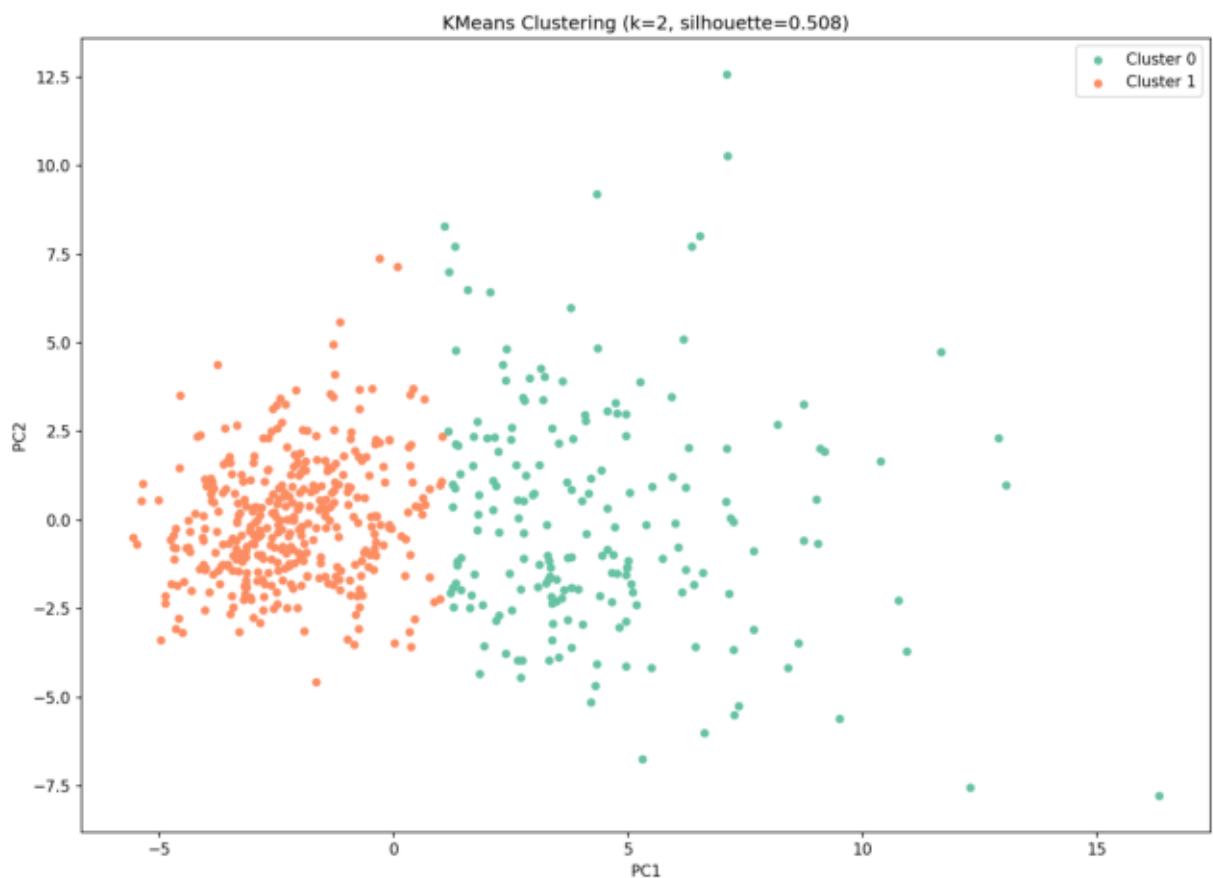
Duplicate Distribution per Column



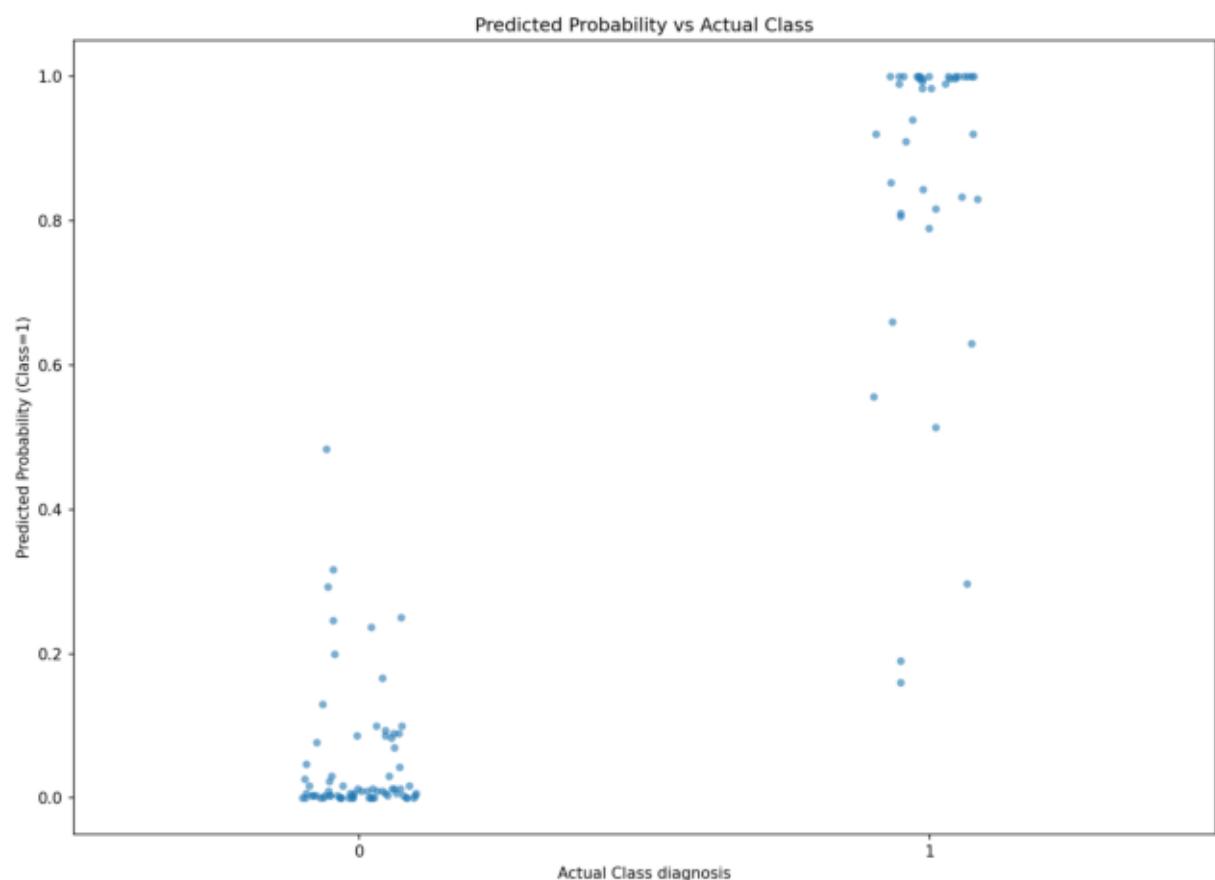
Feature Importance



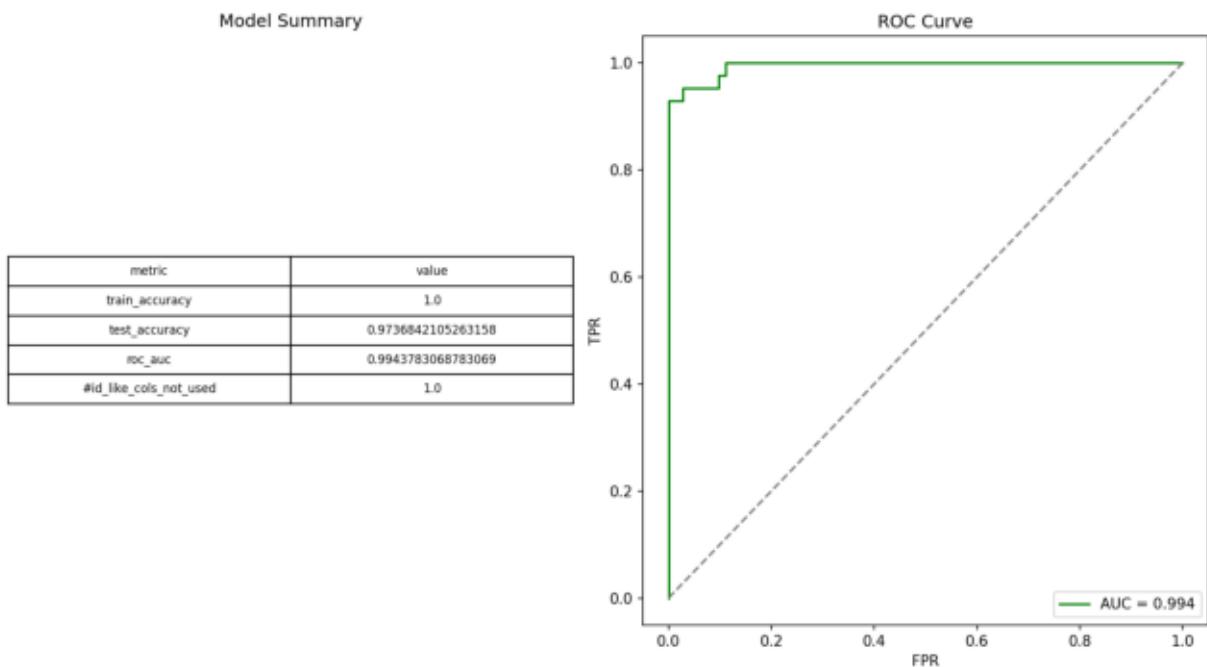
PCA Clusters (Silhouette = 0.508)



Predicted Probability vs Actual



Model Summary & ROC/Residuals

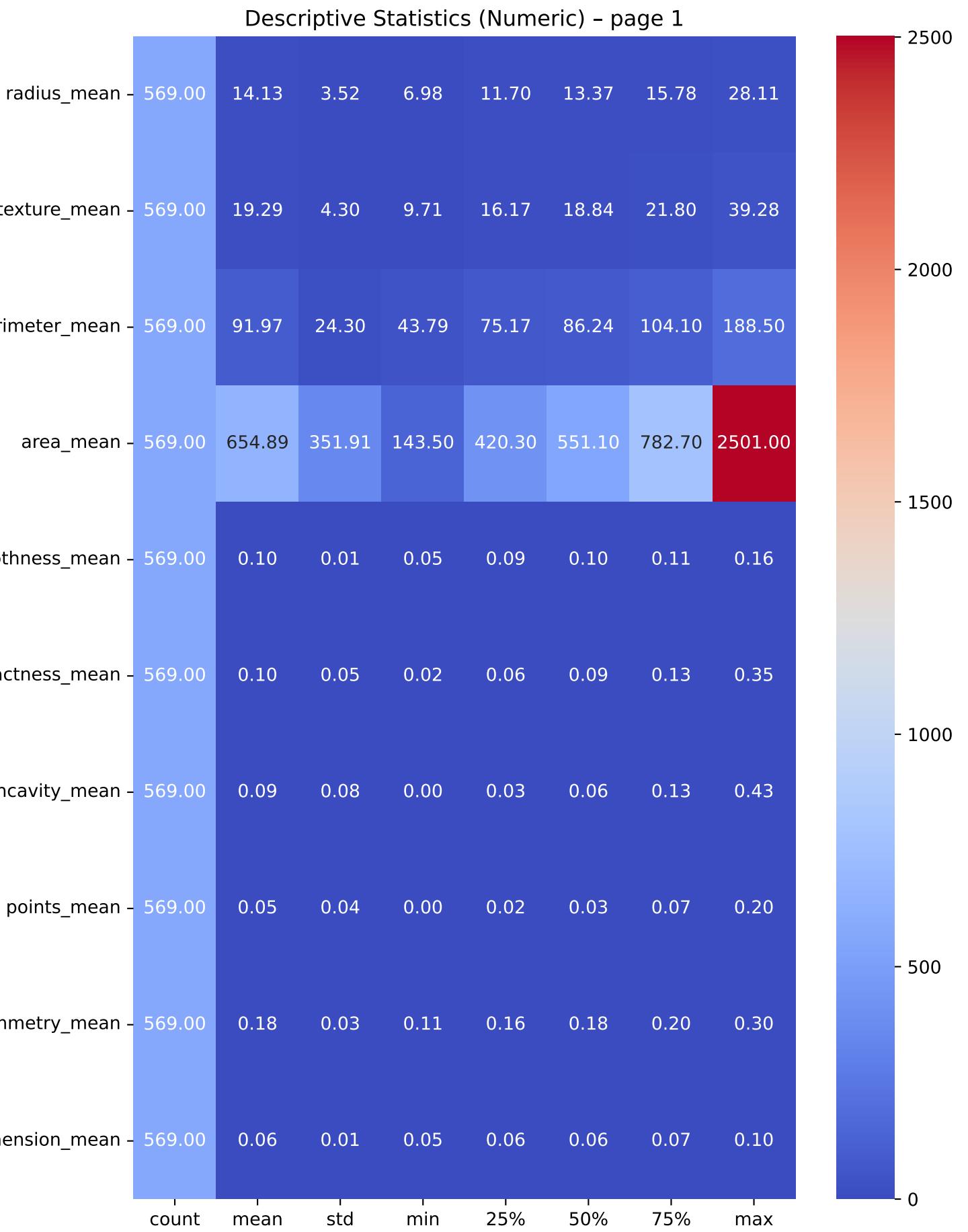


Categorical Columns

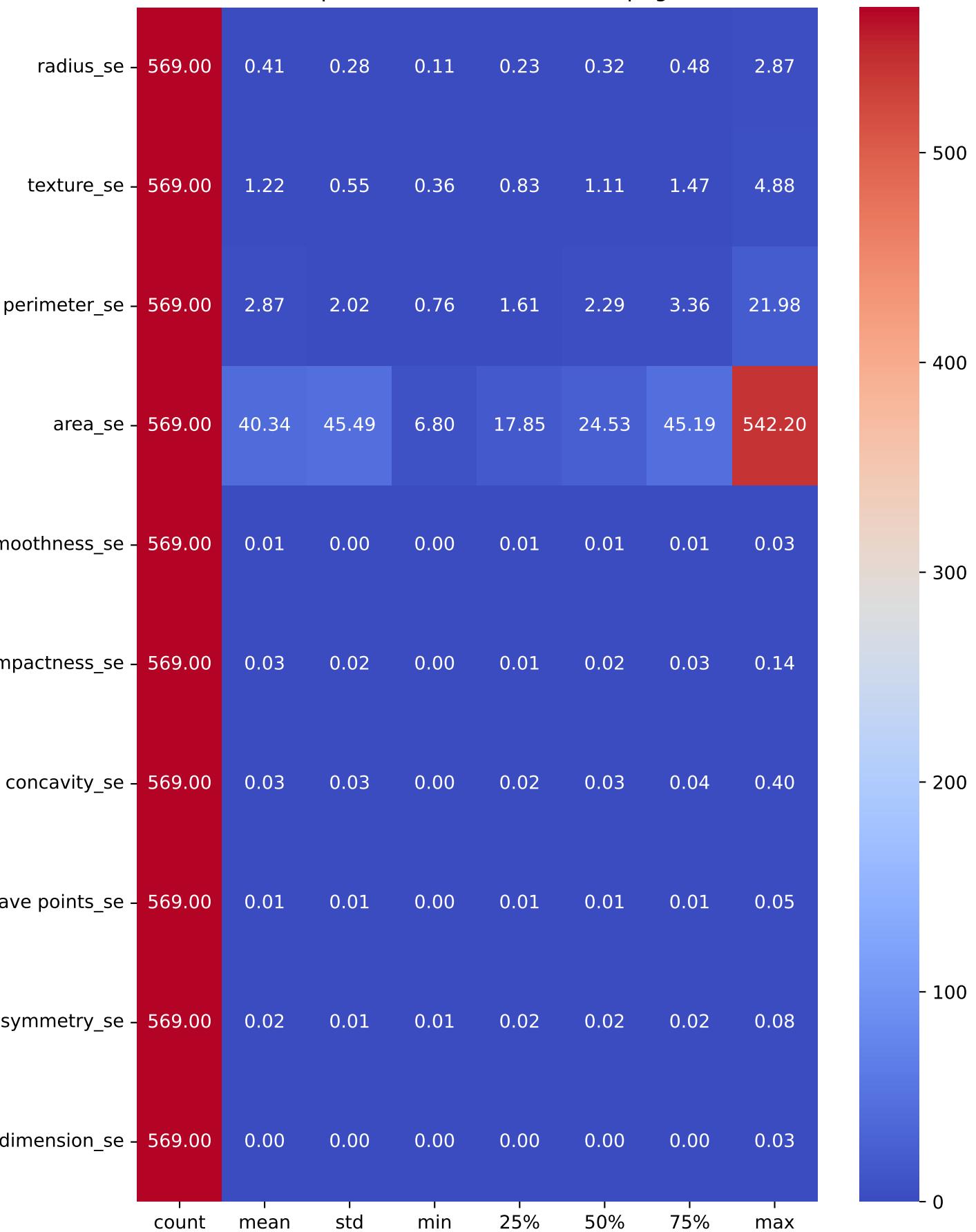
diagnosis

Categorical Summary

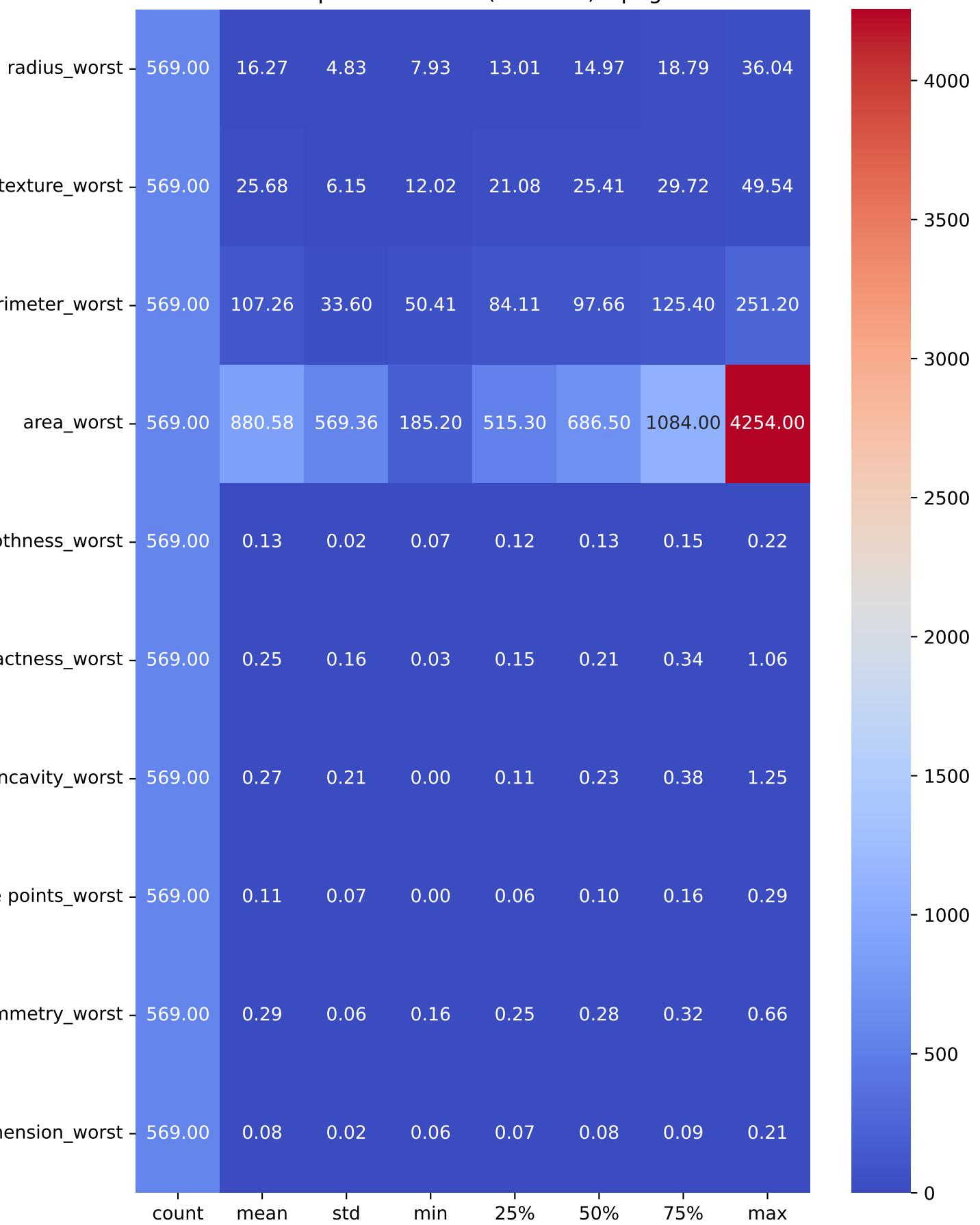
column	count	unique	top	freq
diagnosis	569	2	B	357



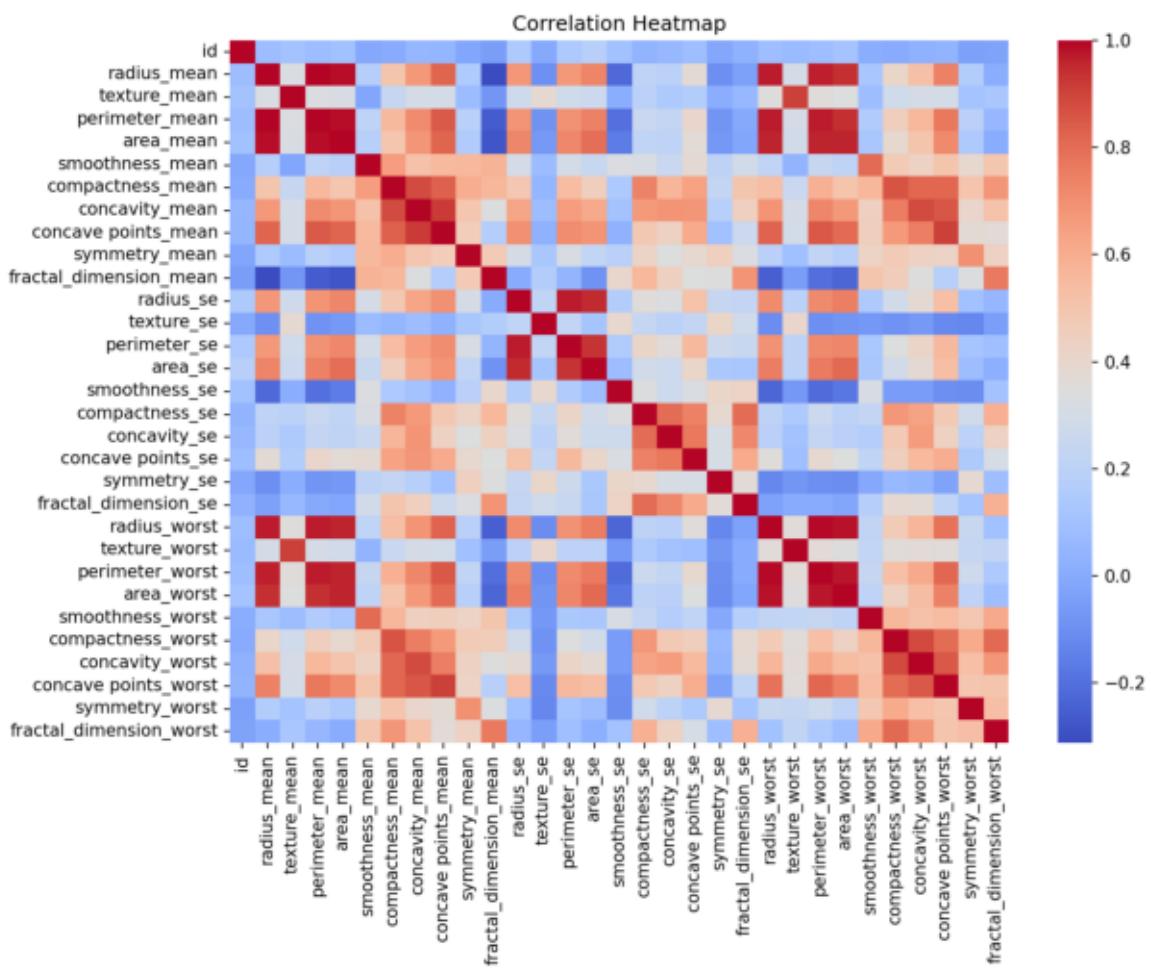
Descriptive Statistics (Numeric) - page 2



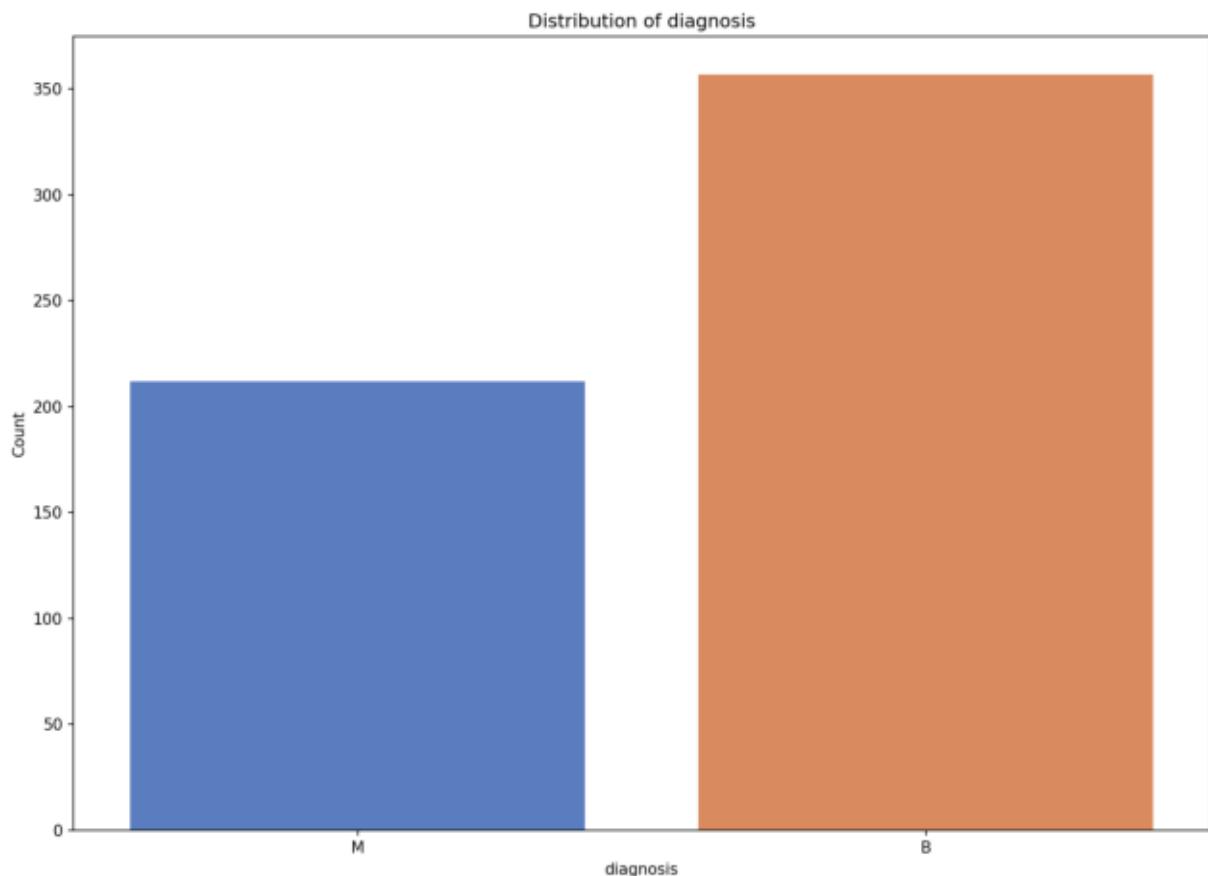
Descriptive Statistics (Numeric) – page 3



Correlation Heatmap



Target Distribution

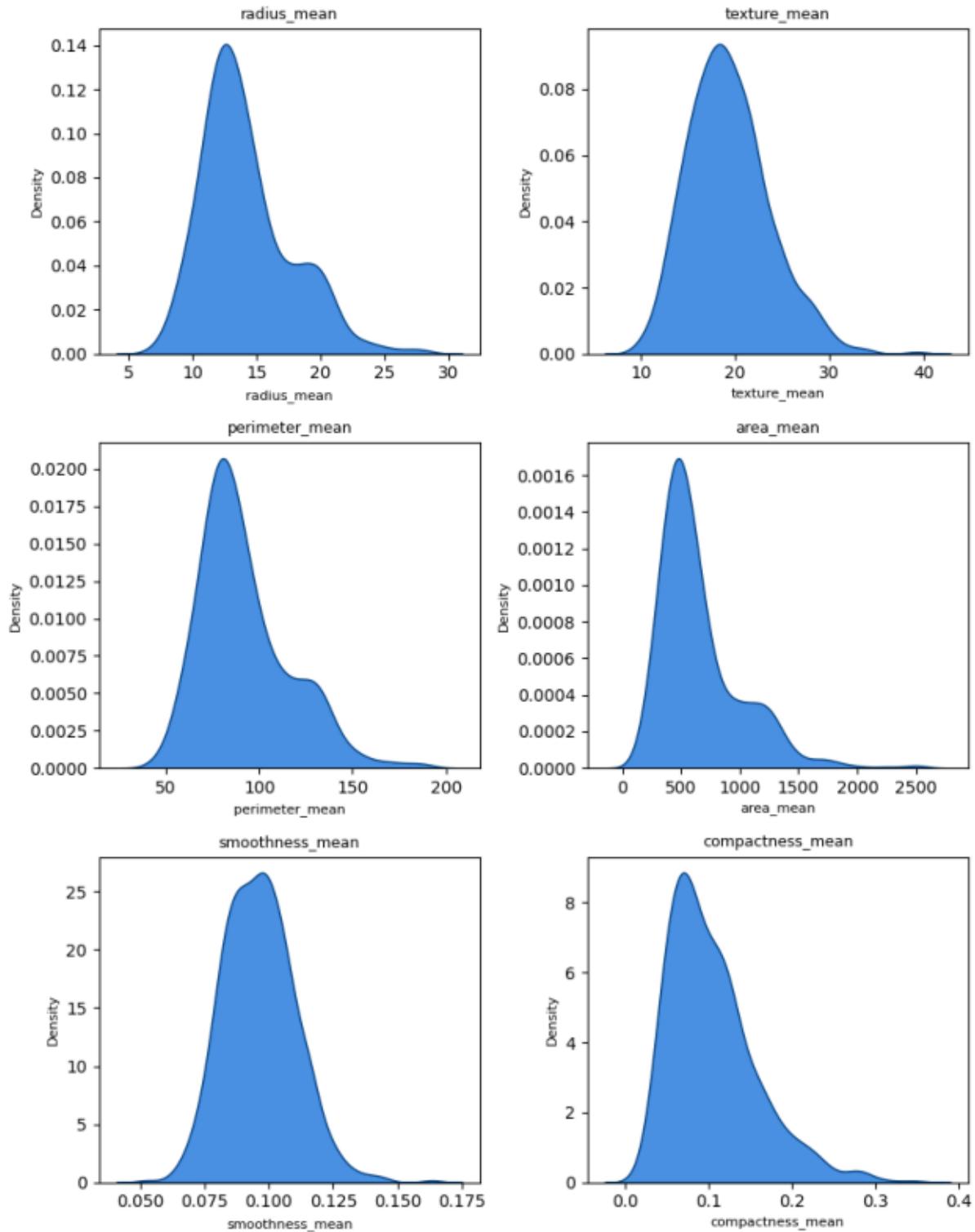


Target Summary

stat	value
count	569
unique	2
top	B
freq	357

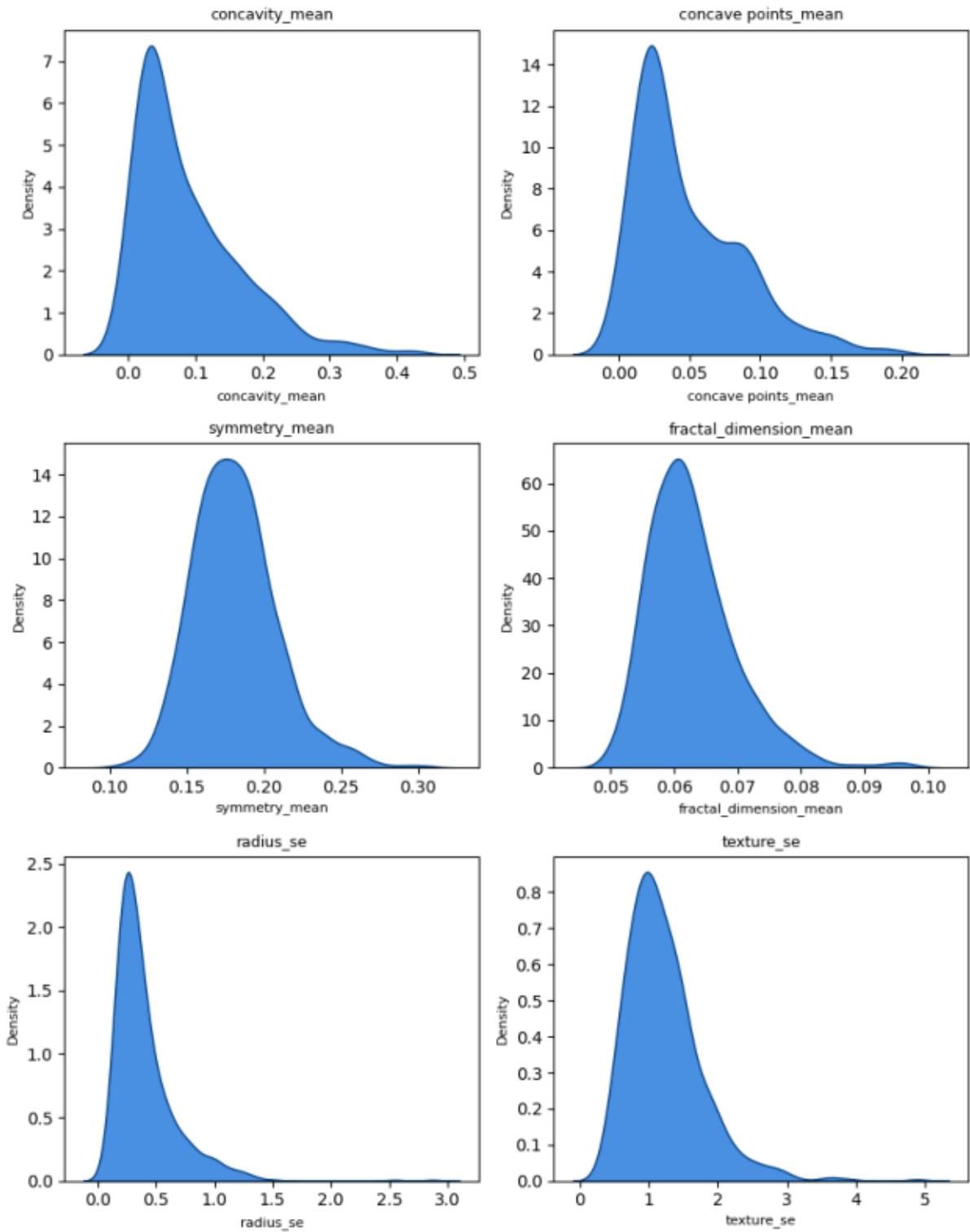
Feature vs Target (page 1)

KDE/Scatter Plots



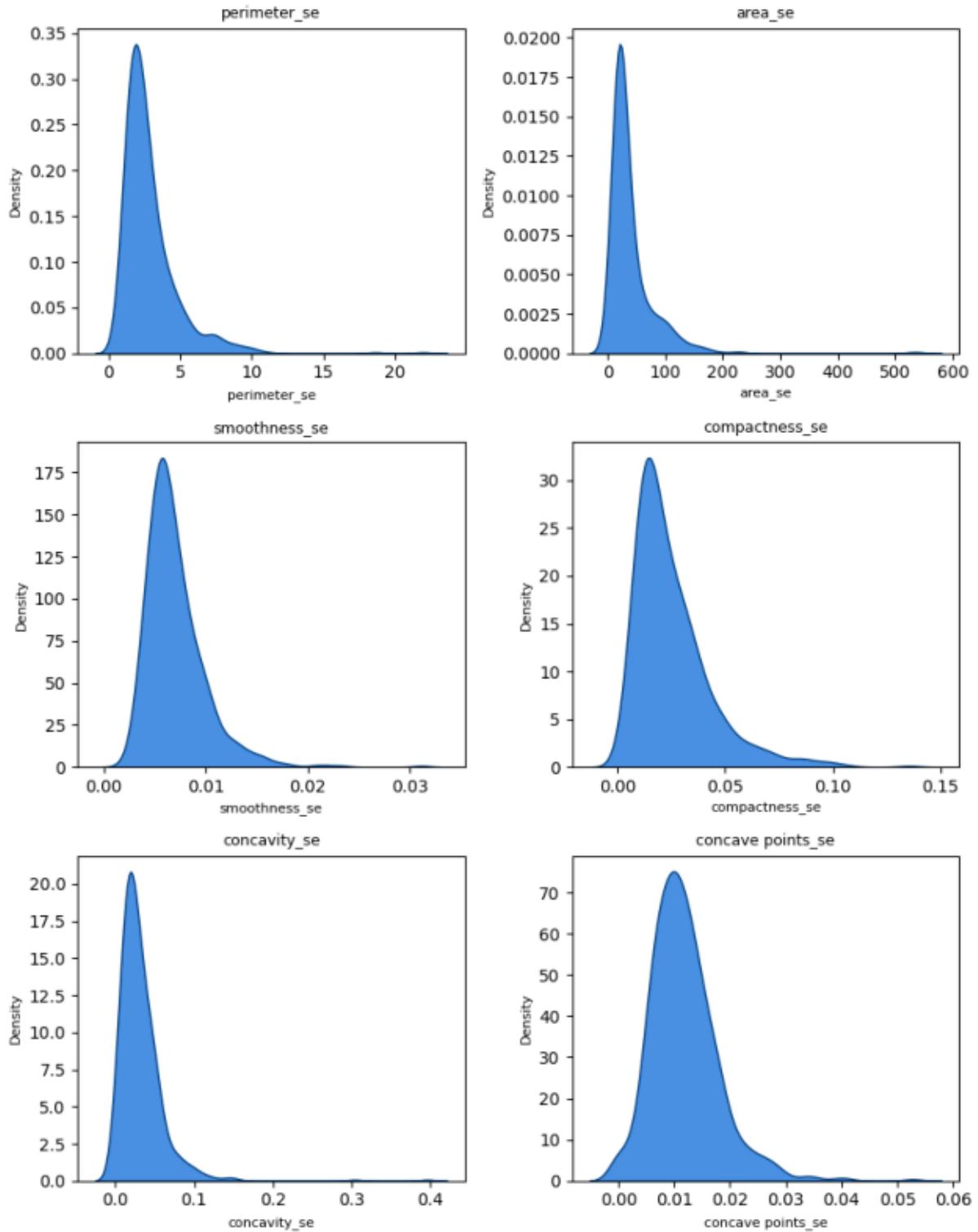
Feature vs Target (page 2)

KDE/Scatter Plots



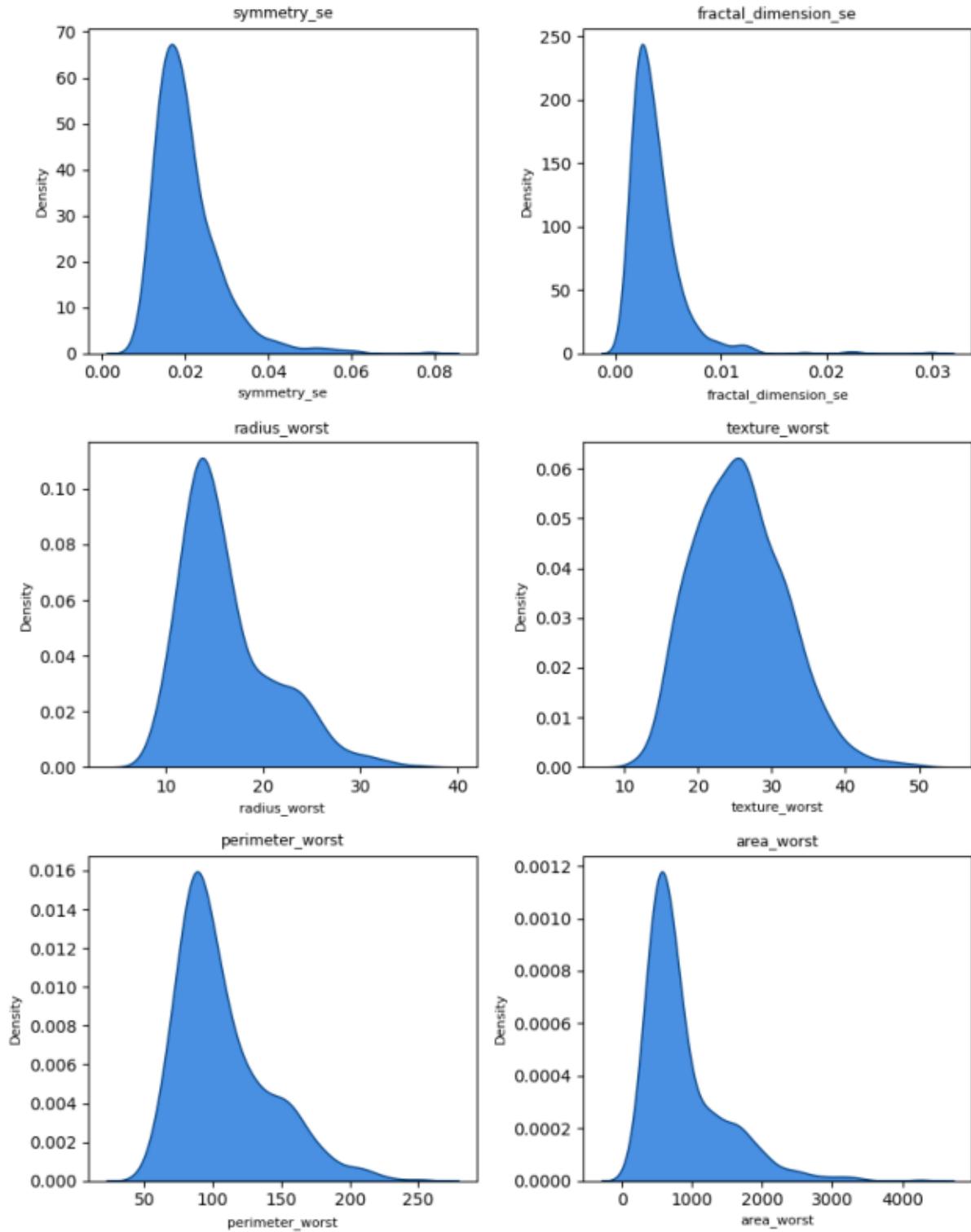
Feature vs Target (page 3)

KDE/Scatter Plots



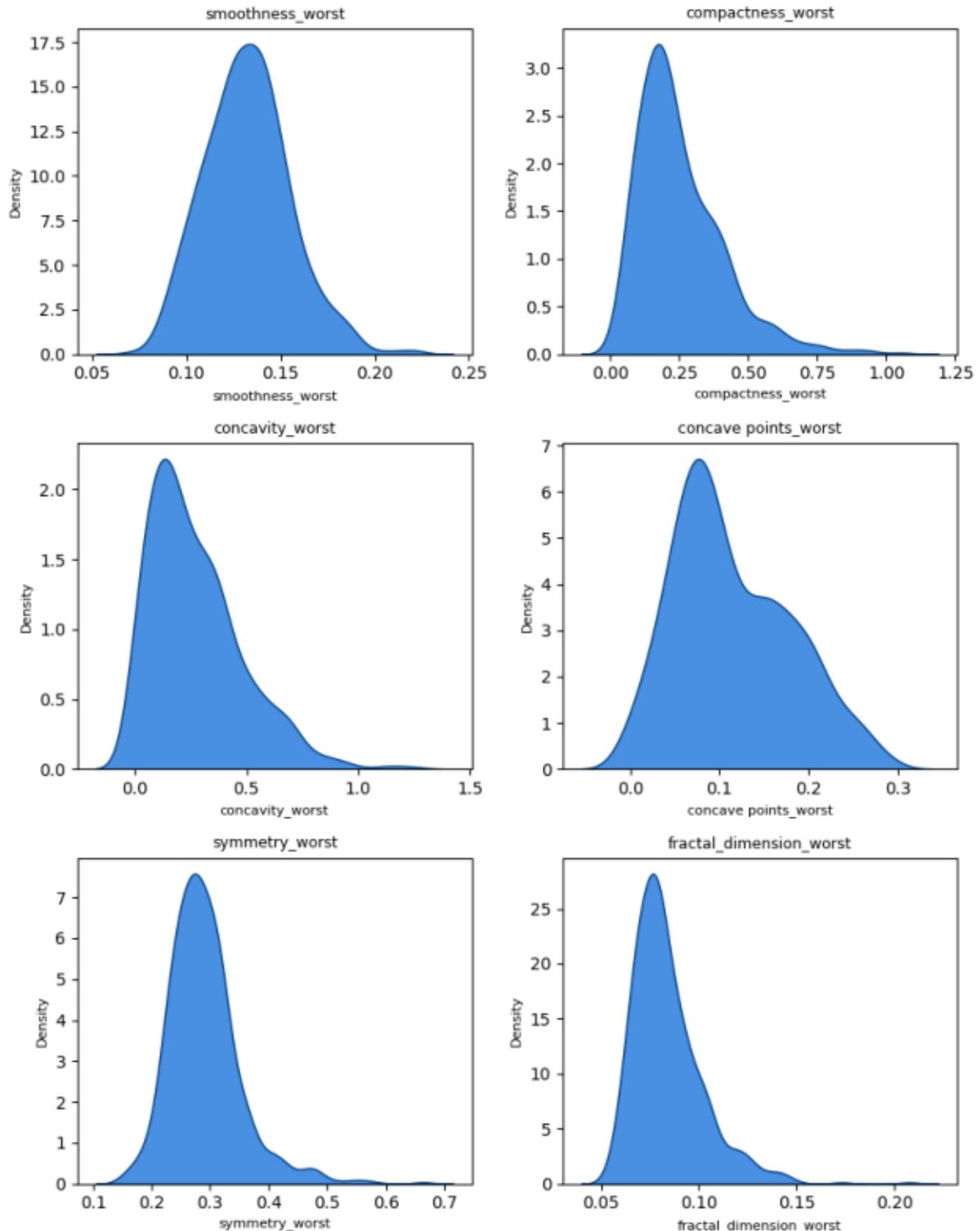
Feature vs Target (page 4)

KDE/Scatter Plots

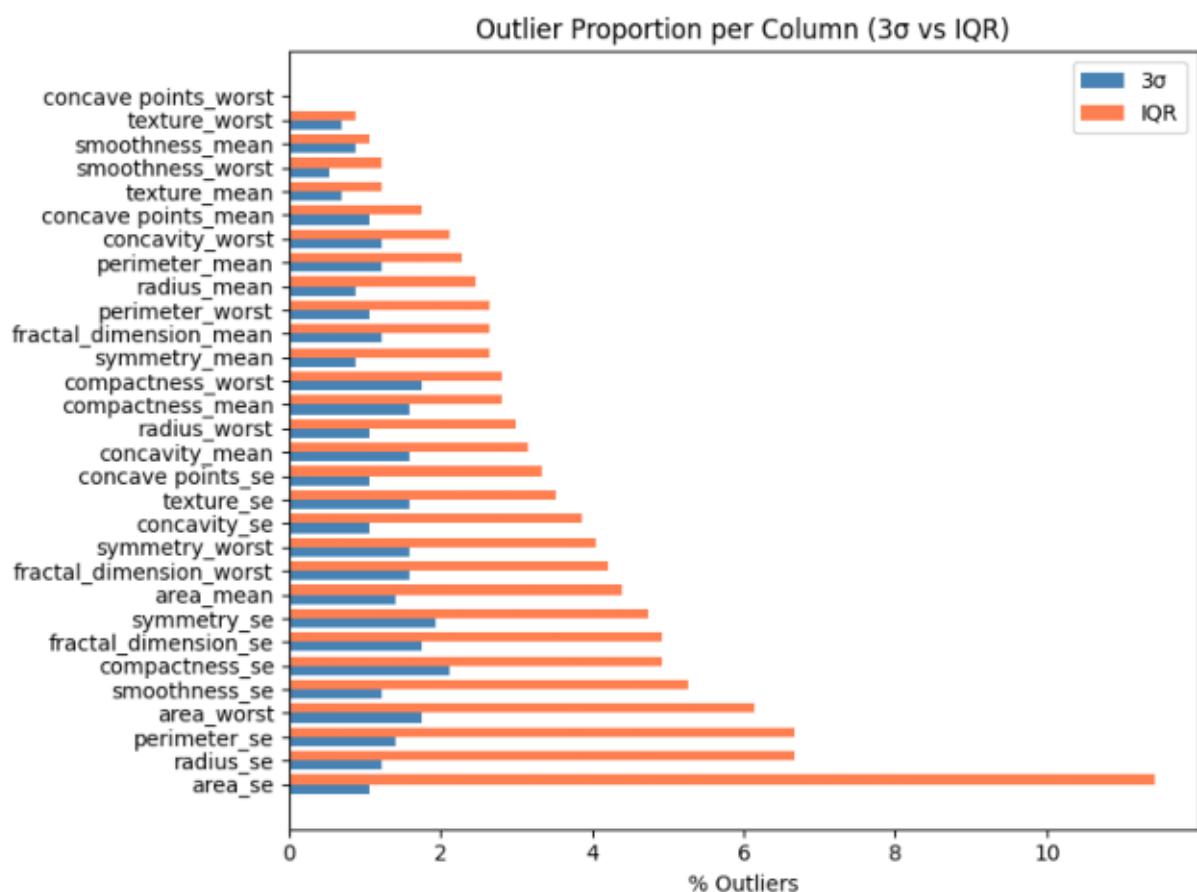


Feature vs Target (page 5)

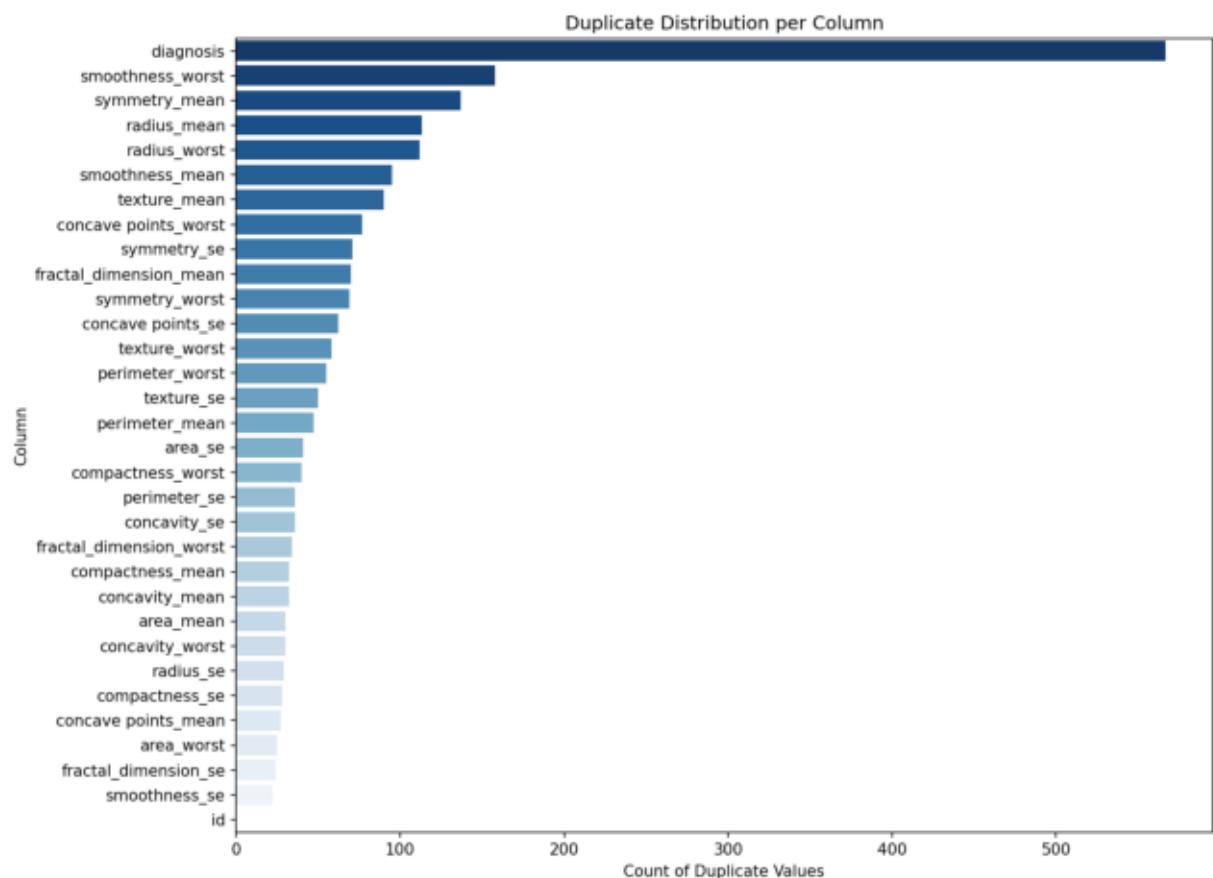
KDE/Scatter Plots



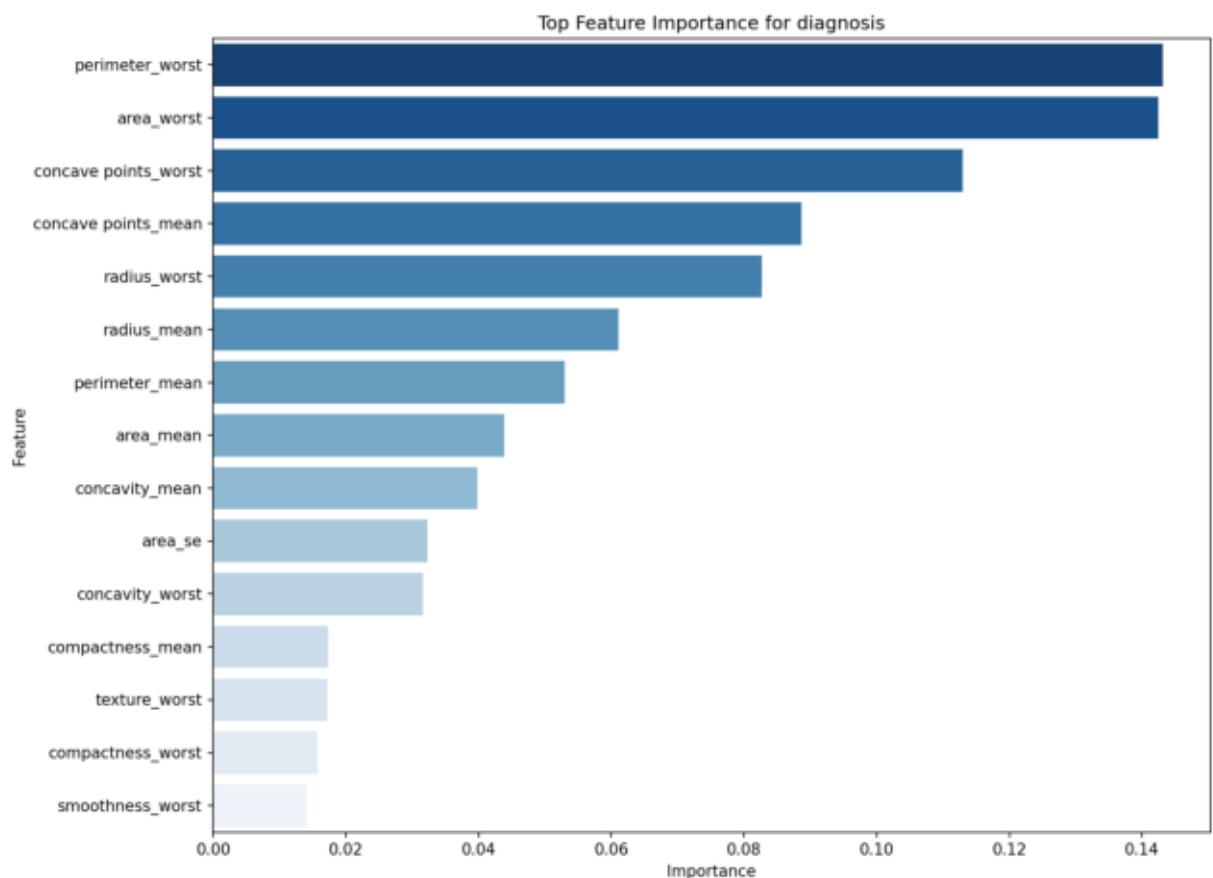
Outlier Proportion



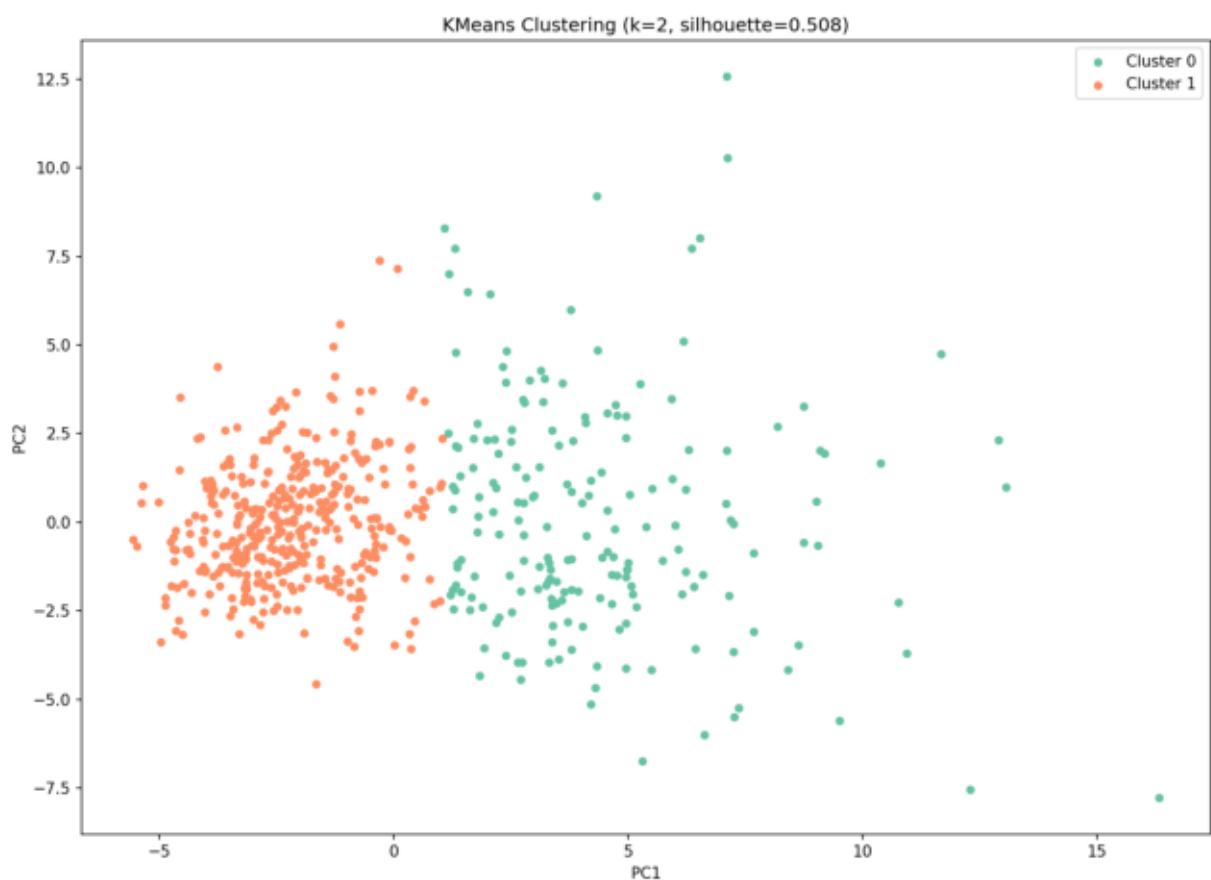
Duplicate Distribution



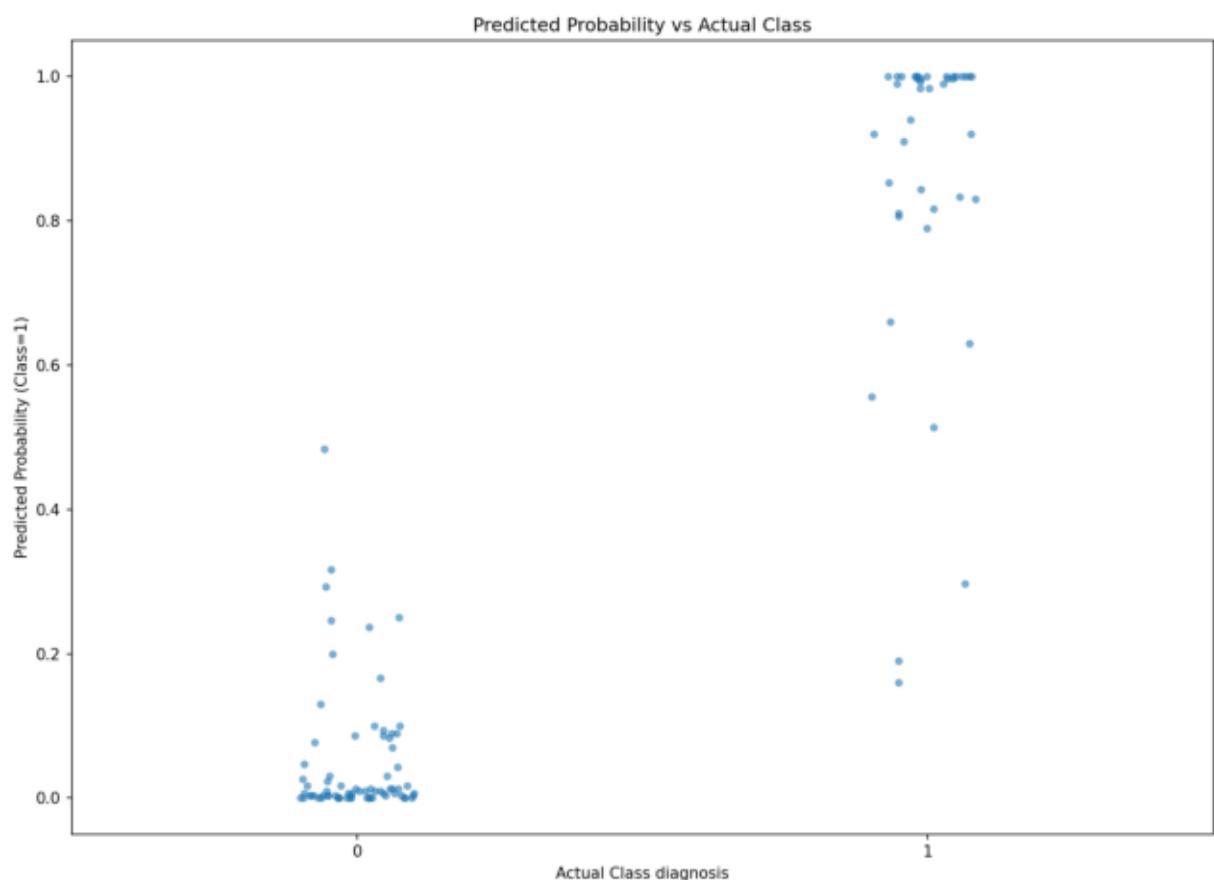
Feature Importance



PCA Clusters (silhouette=0.508)



Predicted Probability vs Actual



Model Summary & Performance

