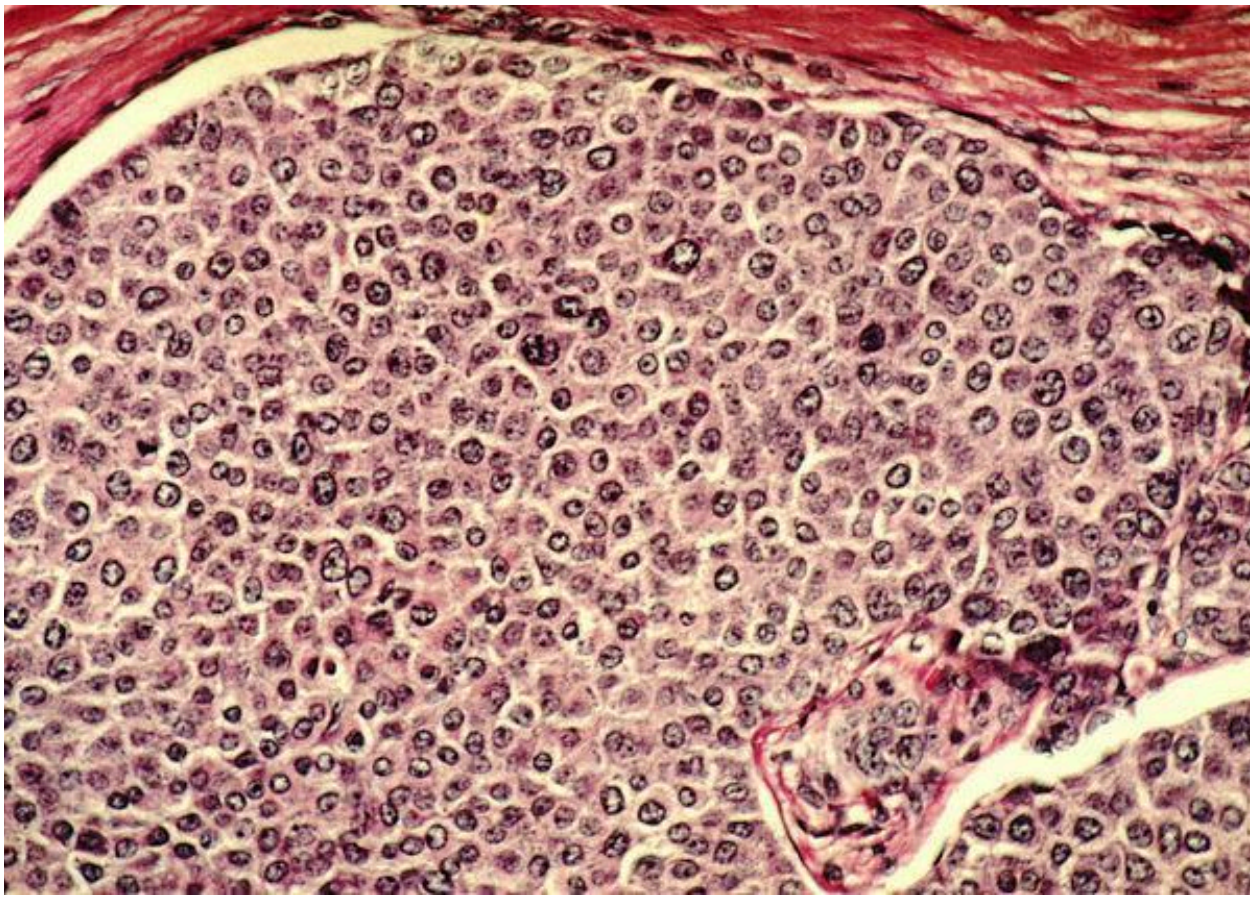
Breast Cancer Data Analysis Report



Introduction

Breast Cancer is a widely spread disease that affects many women all around the world.In this project we aim to get the most beneficial insights on what triggers this disease, what affects it the most and which women are most vulnerable to it.

Data Source:

The dataset used in this project is the Kaggle Breast Cancer Dataset

Link: [Kaggle Breast Cancer Dataset]

Data Explanation

This dataset has 31 features and the dependent variable(diagnosis) the features are:

Sure! Here are the rewritten names without quotes and numbered:

1. id
2. diagnosis
3. radius_mean
4. texture_mean
5. perimeter_mean
6. area_mean
7. smoothness_mean
8. compactness_mean
9. concavity_mean
10. concave_points_mean
11. symmetry_mean
12. fractal_dimension_mean
13. radius_se

14. texture_se

15. perimeter_se

16. area_se

17. smoothness_se

18. compactness_se

19. concavity_se

20. concave_points_se

21. symmetry_se

22. fractal_dimension_se

23. radius_worst

24. texture_worst

25. perimeter_worst

26. area_worst

27. smoothness_worst

28. compactness_worst

29. concavity_worst

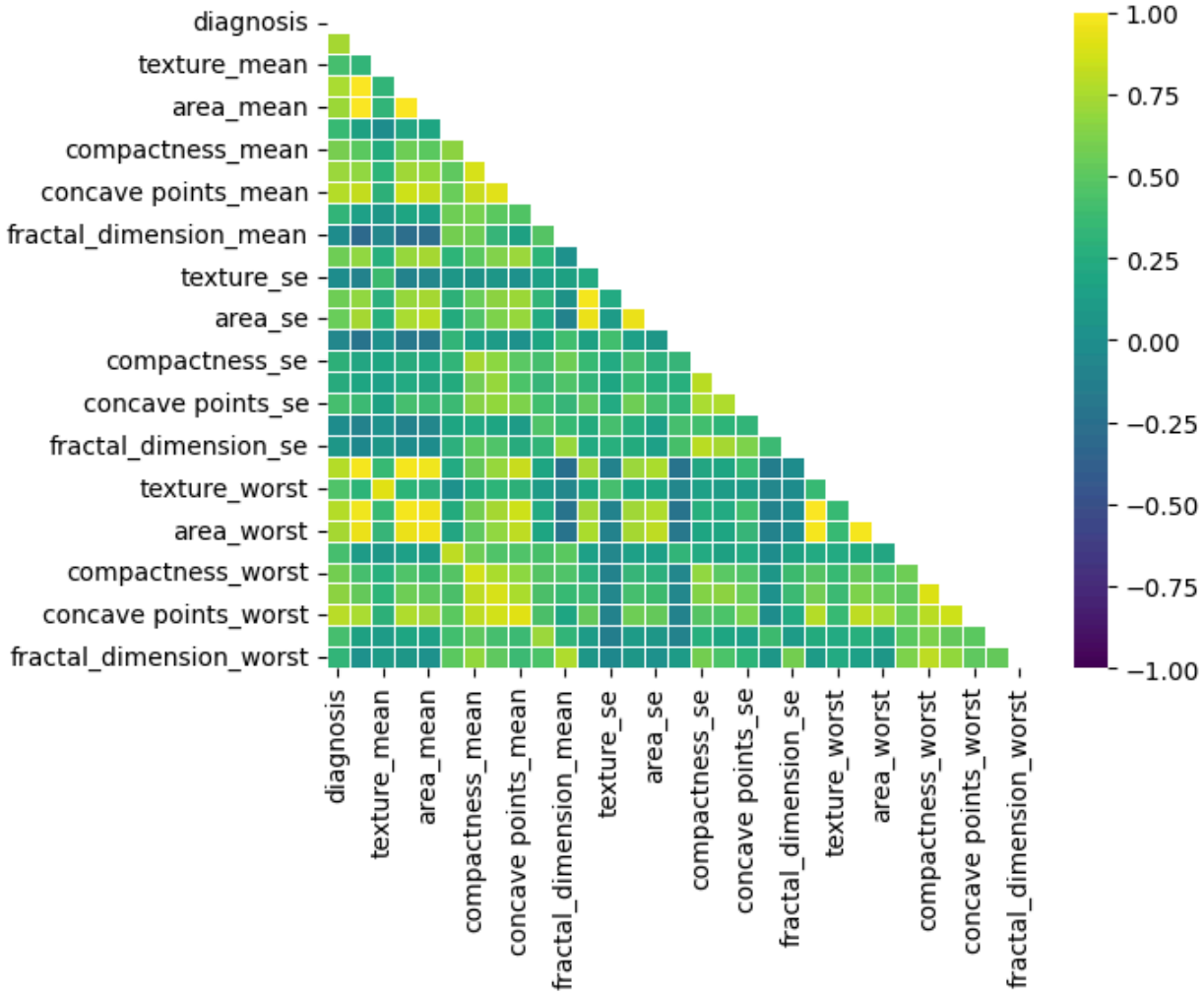
30. concave_points_worst

31. symmetry_worst

32. Fractal_dimension_worst

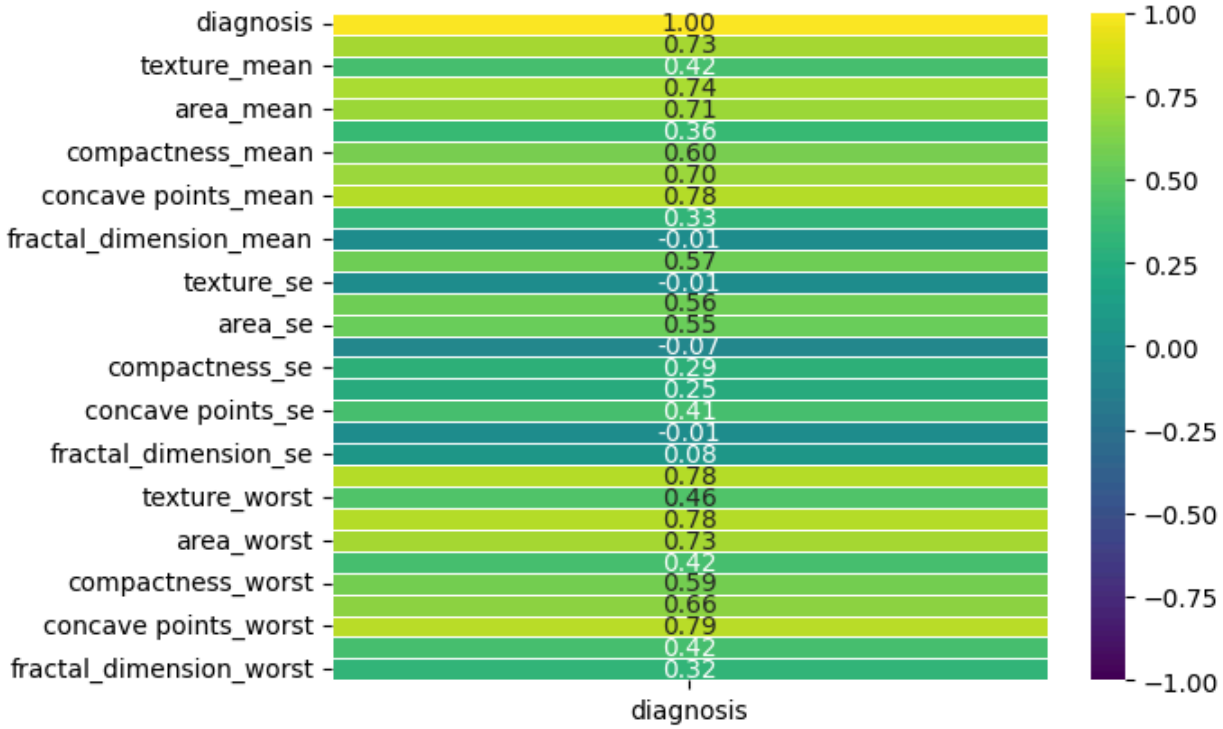
Analysis

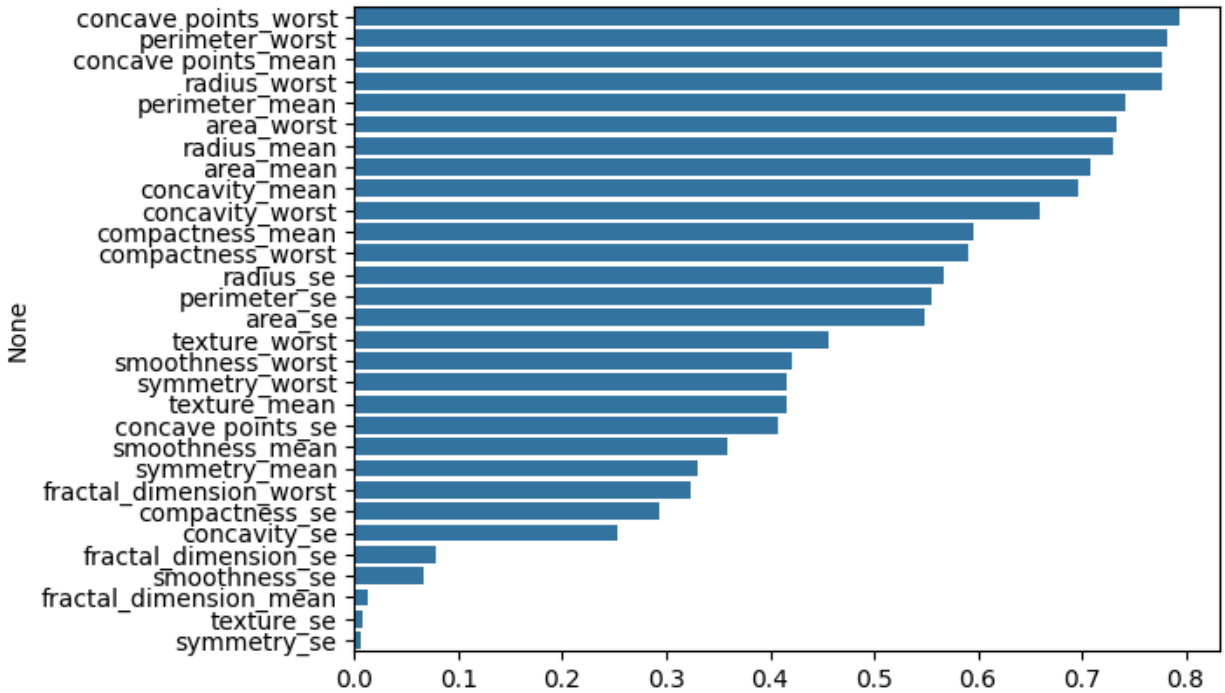
First we need to look at the correlation matrix



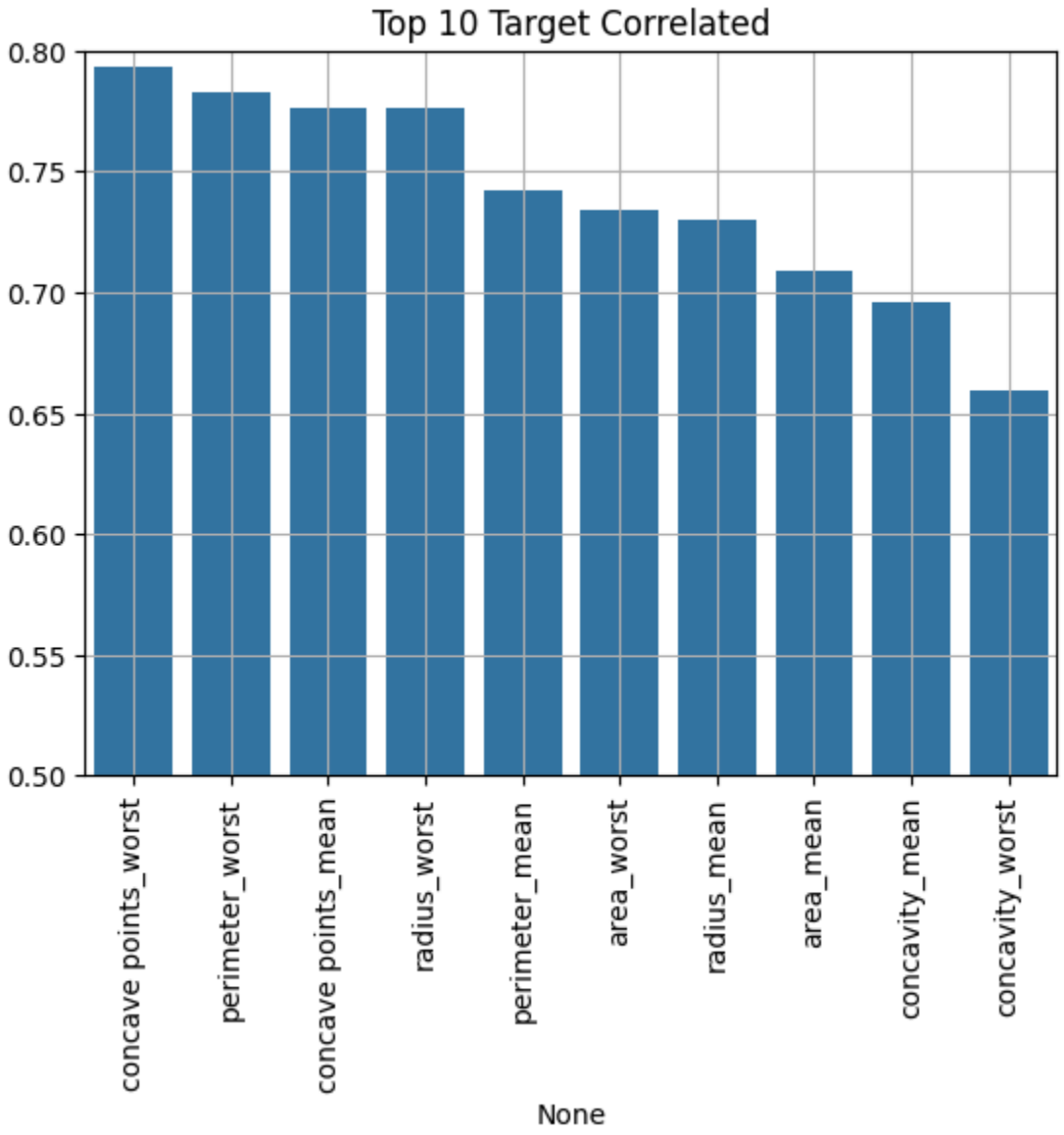
Note: Almost all features are positively correlated with each other

But we need to look further into the correlation between features and the dependet variable

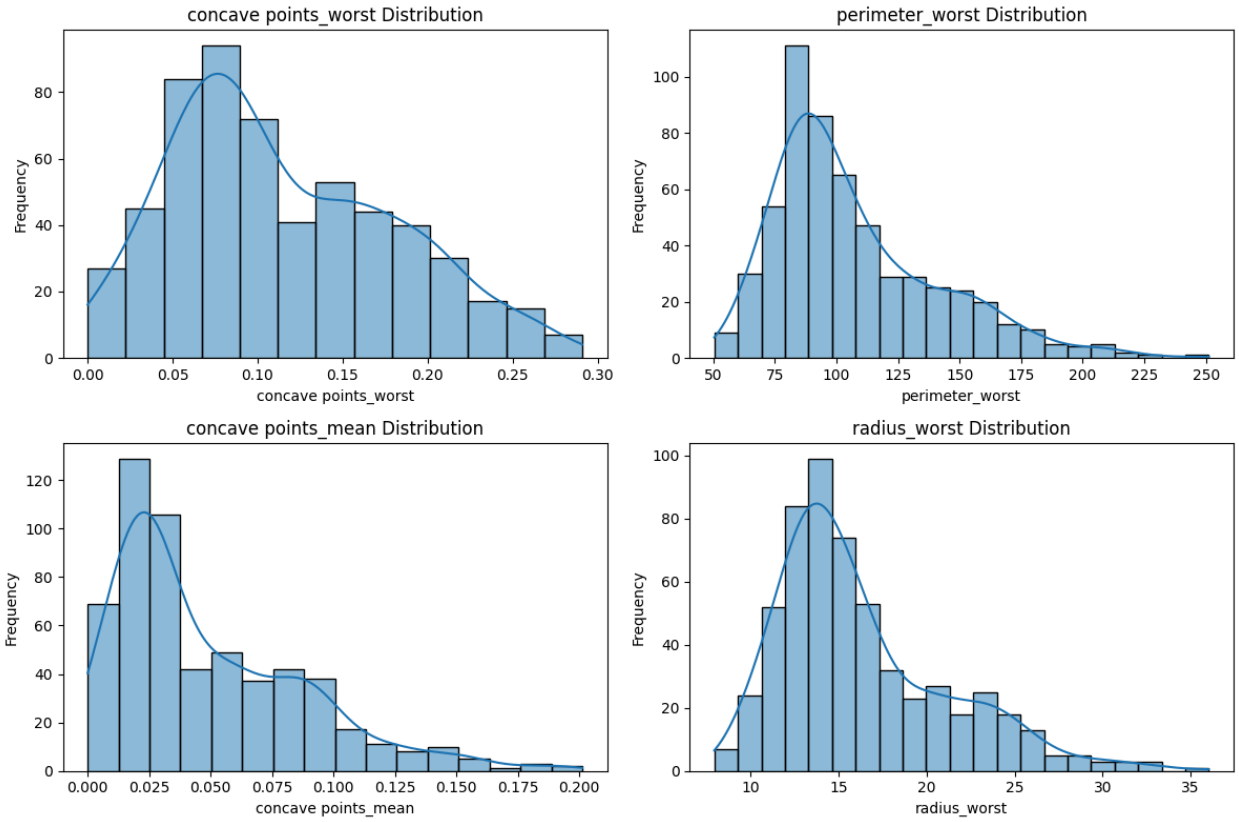




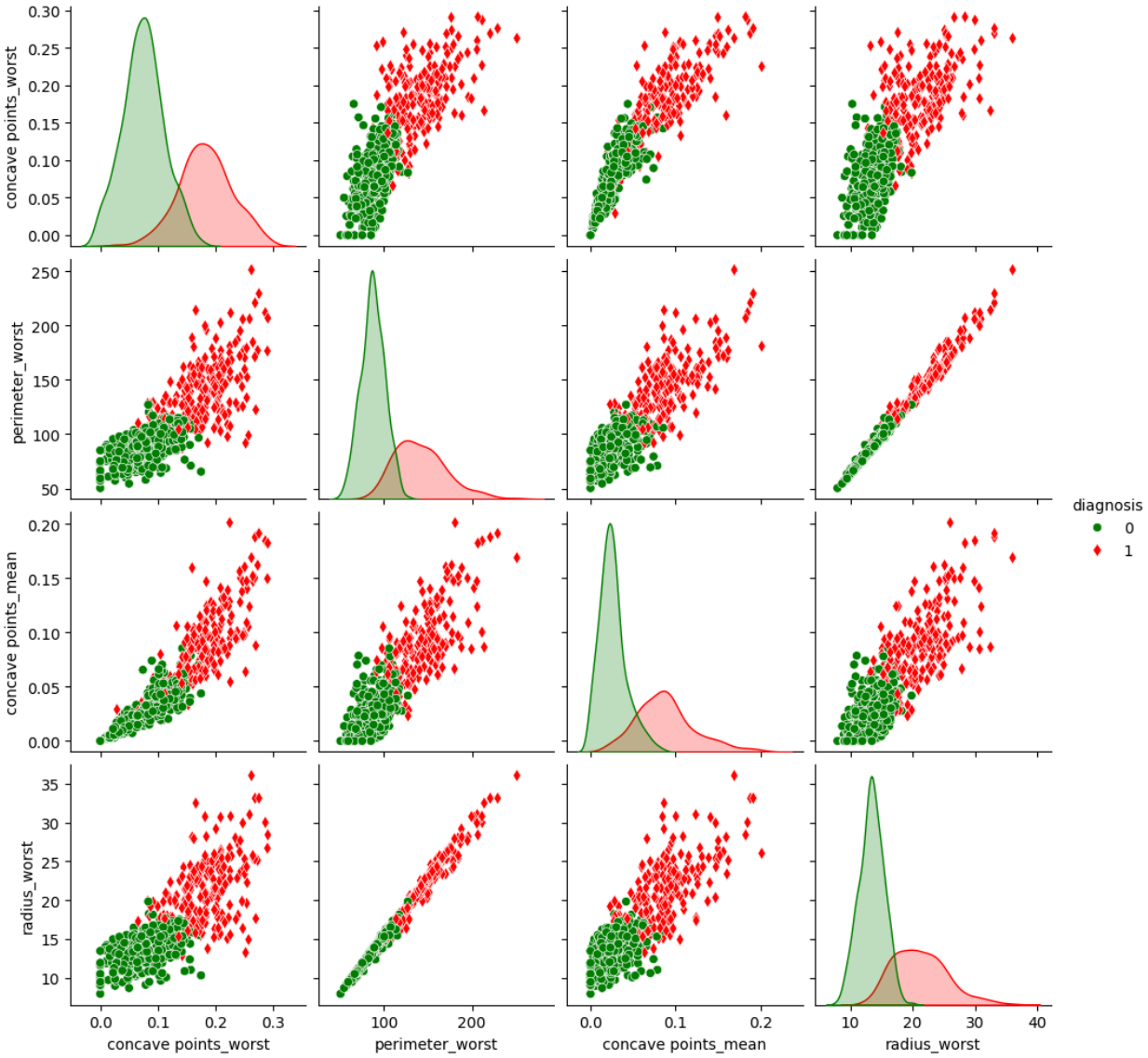
Lets see the top 10 correlated features with the target



Lets see the distribution of the top 4



Now lets see one of the most important visualtizations



Insights

From the statistical modelling and data analysis that we have done we can assure that:

- Increasing the worst perimeter from 80 to 100 results in a 38% higher chance of having a malignant tumor.
- Increasing concave points_worst from 0.1 to 0.2 results in a 28.5% higher chance of having a malignant tumor.
- Increasing radius_worst from 10 to 20 results in a 61.5% higher chance of having a malignant tumor.
- Increasing concave points_mean from 0.01 to 0.05 results in a 45.7% higher chance of having a malignant tumor.
- Patients with an area_mean greater than 700 have a 94% chance of having a malignant tumor.
- Patients with a compactness_mean greater than 0.13 have an 80% chance of having a malignant tumor.