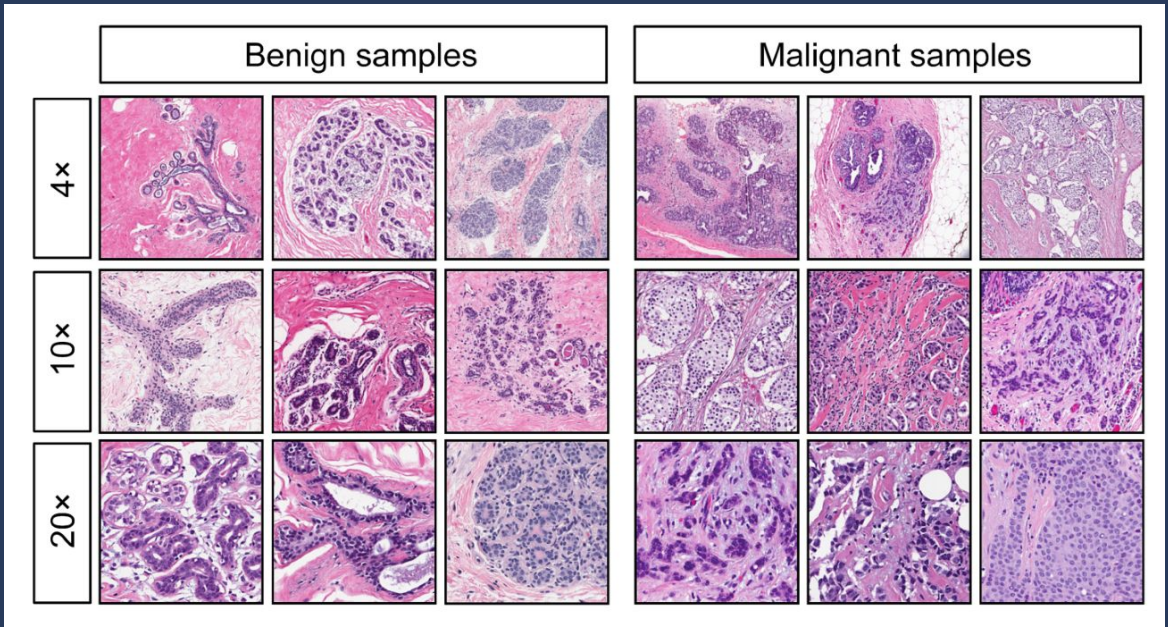
Applied Statistics Exercise

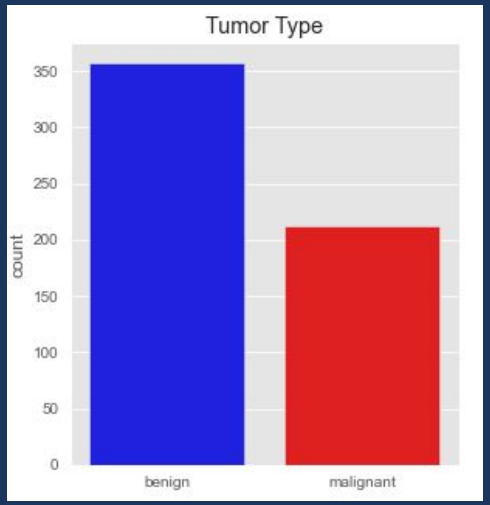
June 18, 2021

Case: Wisconsin Breast Cancer Dataset



Use Case: Wisconsin Breast Cancer Dataset

Problem: Explore the difference of benign and malignant tumors based on the diagnostic features



- **Benign tumors - they form only in one spot without spreading to surrounding tissue**
- **Malignant tumors - they are cancerous and can spread to nearby tissue**

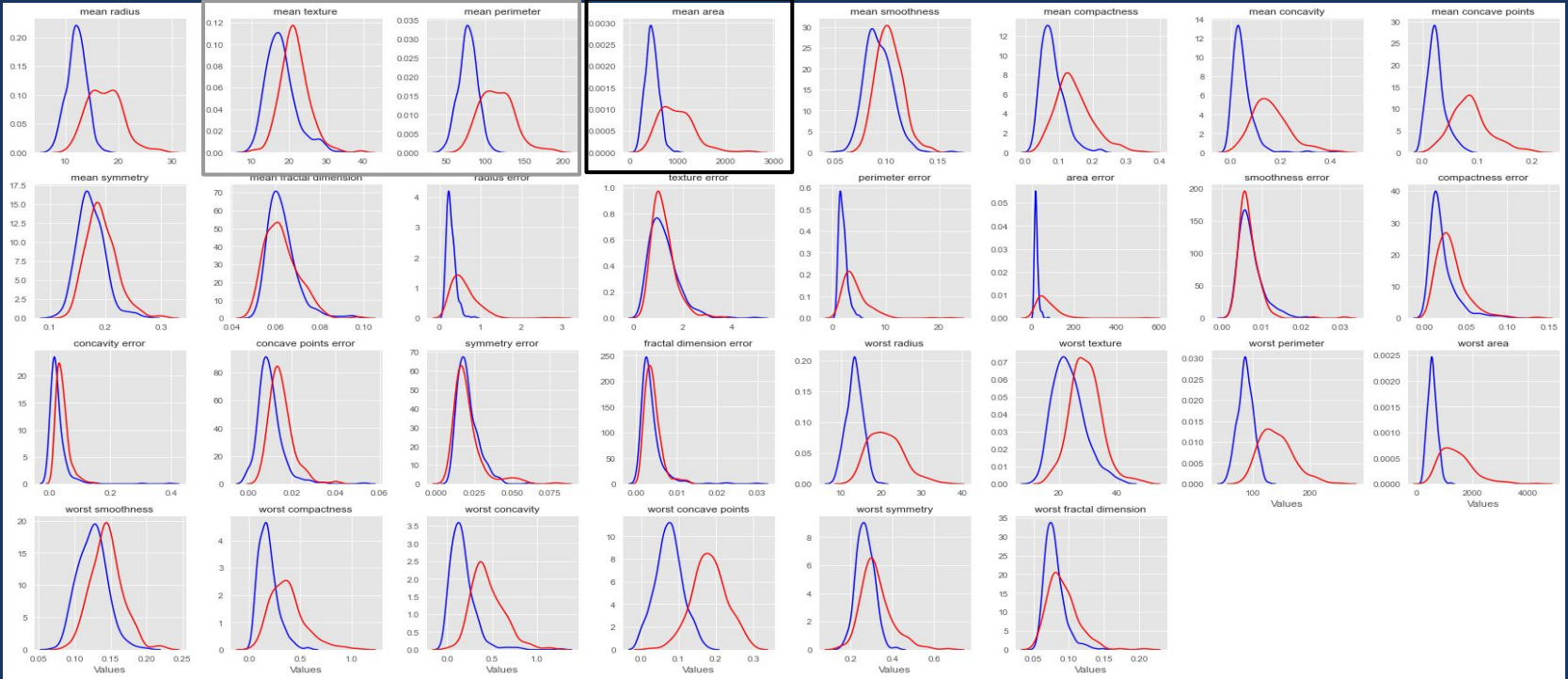
Use Case: Wisconsin Breast Cancer Dataset

This Task: Explore the difference of benign and malignant tumors

1. Perform descriptive statistics and the relevant exploratory data analysis methods to analyzed the dataset
2. Perform hypothesis testing by answering the following:
    - Do those with ‘malignant’ tumors have bigger ‘mean area’ on average than those with ‘benign’ tumors?
    - Find correlation between mean texture and mean perimeter for those with malignant tumors.

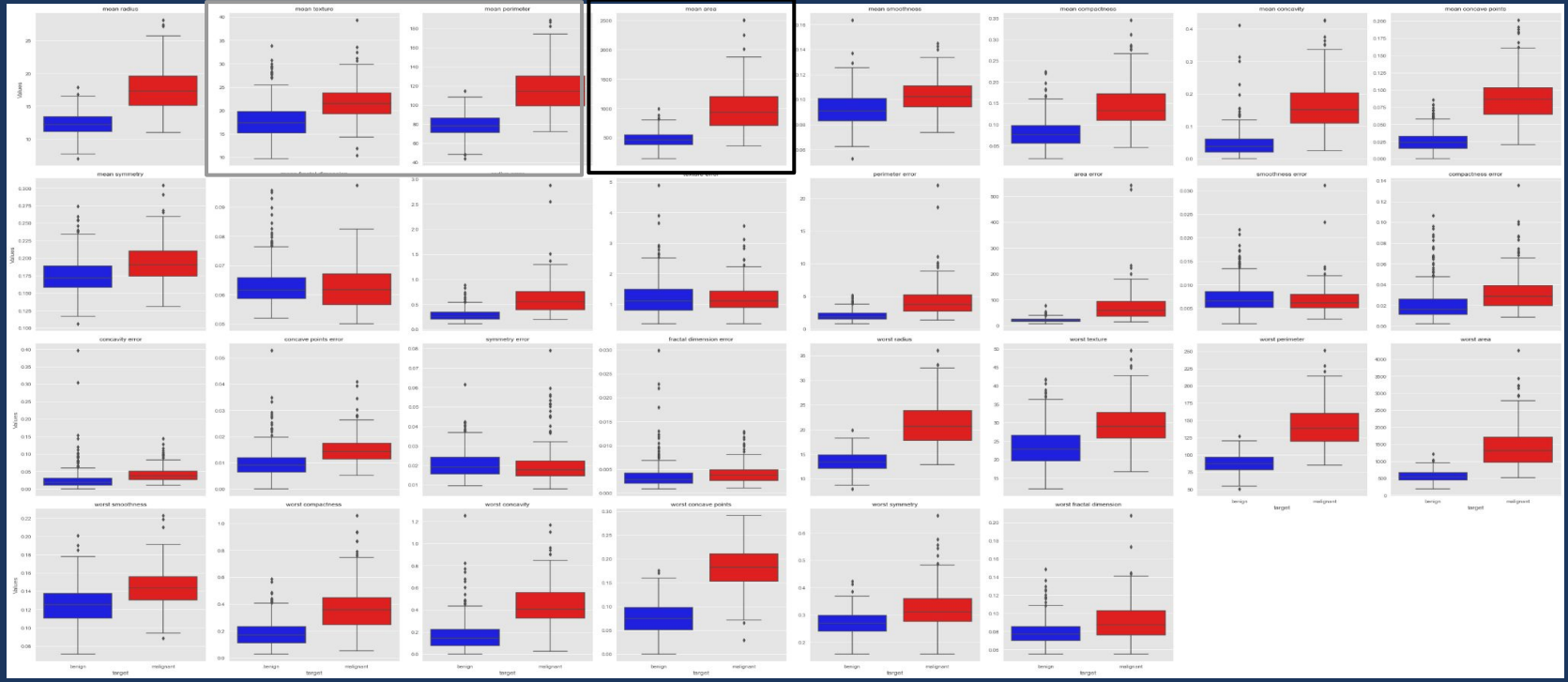
Use Case: Wisconsin Breast Cancer Dataset

Exploratory Data Analysis



Use Case: Wisconsin Breast Cancer Dataset

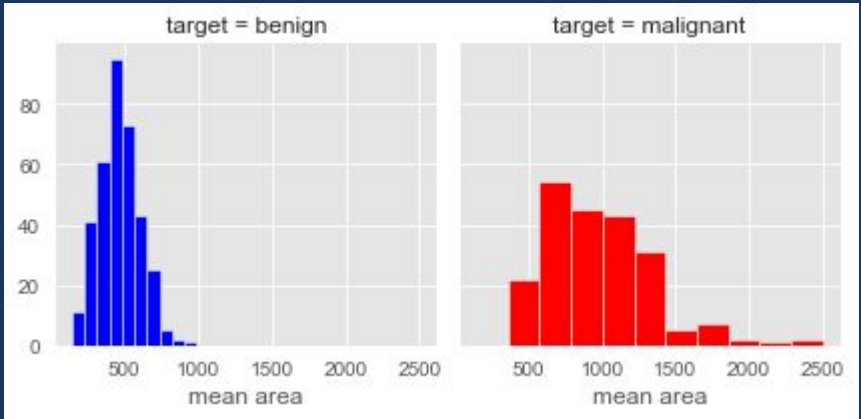
Exploratory Data Analysis



Use Case: Wisconsin Breast Cancer Dataset

Do those with ‘malignant’ tumors have bigger ‘mean area’ on average than those with ‘benign’ tumors?

Null Hypothesis: mean area of ‘malignant’ tumors is less than and equal to those with ‘benign’ tumors
Alternative Hypothesis: mean area of ‘malignant’ tumors is greater than those with ‘benign’ tumors



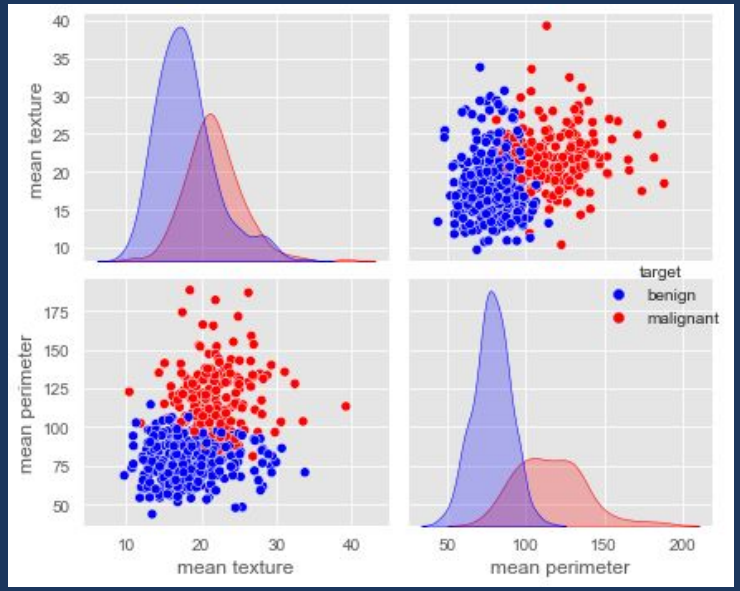
Results:

- Test Stat: 19.6861
- p-value (right-tailed): 1.3310e-52
- critical value of t: 1.6511
- Reject the null hypothesis (alpha = 0.05)

Have enough evidence that the ‘malignant’ tumors have bigger ‘mean area’ than those with ‘benign’ tumors

Use Case: Wisconsin Breast Cancer Dataset

Find correlation between mean texture and mean perimeter for those with malignant tumors.



Correlation = 0.1107
Low Positive Correlation