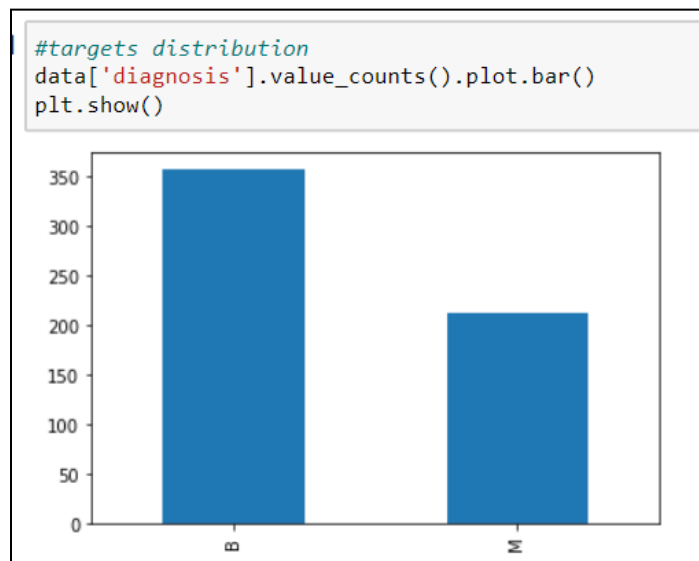


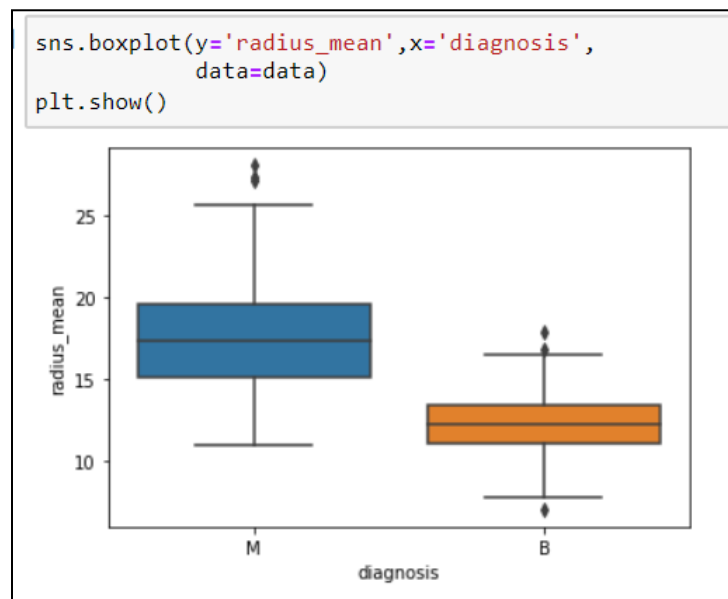
Breast Cancer prediction

Github link:

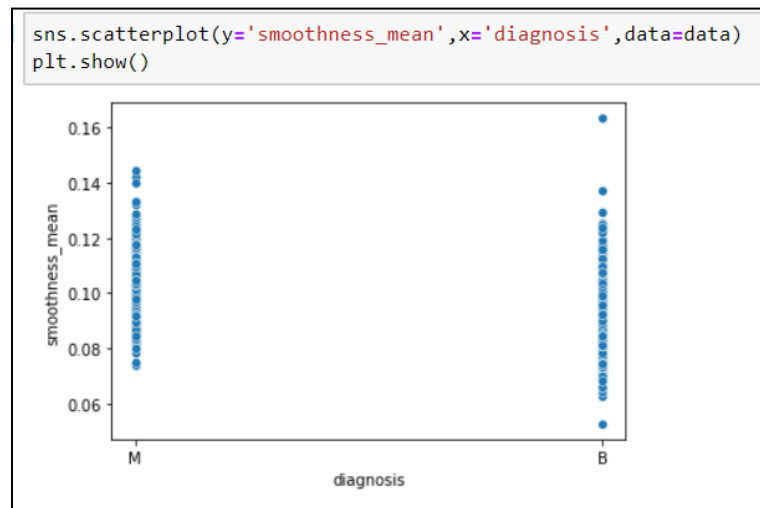
1) Target distribution of diagnostic features (benign and malignant)



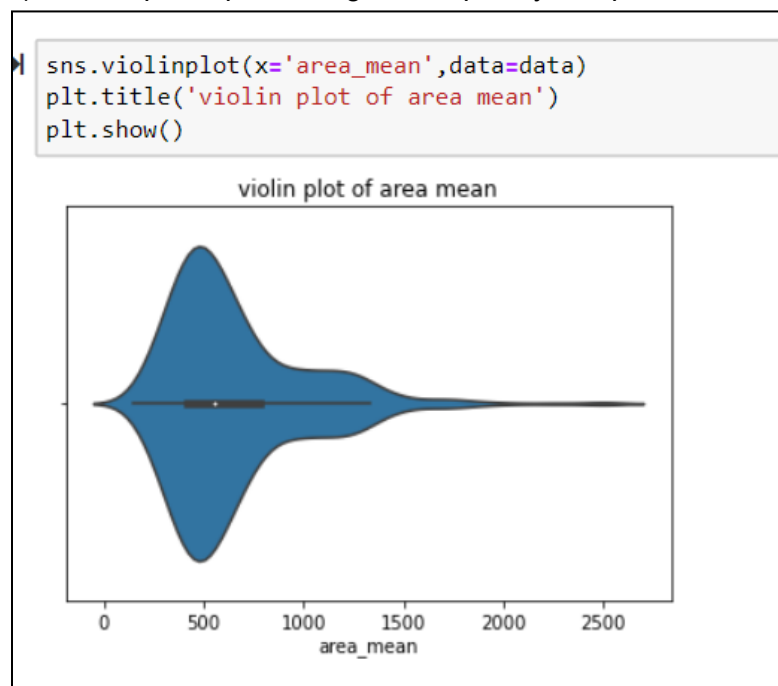
2) A boxplot representing the tissue radius of diagnostic features. The radius mean is high for the malignant category and low for the benign category



3) A scatterplot representing the tissue radius of diagnostic features. Smoothness mean is the same for both benign and malignant

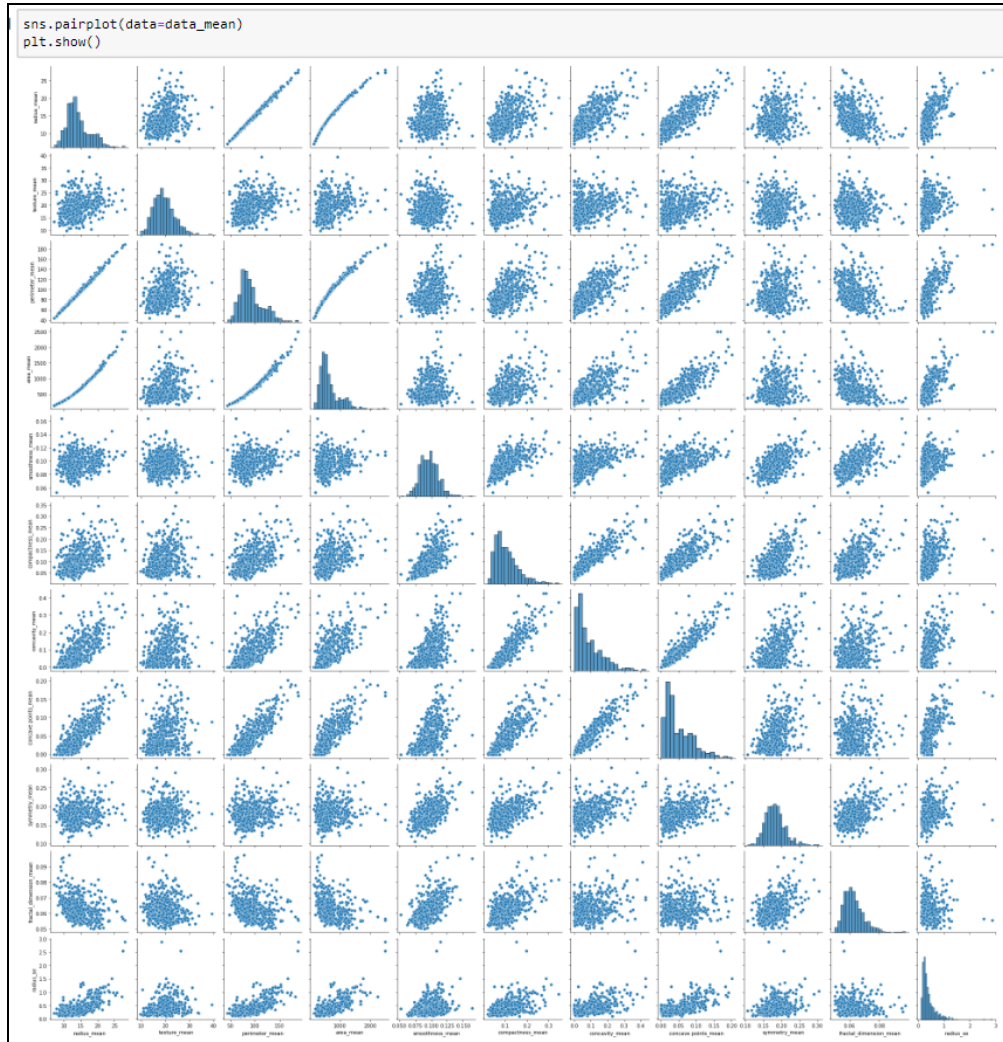


4) A violin plot representing the frequency components of the area



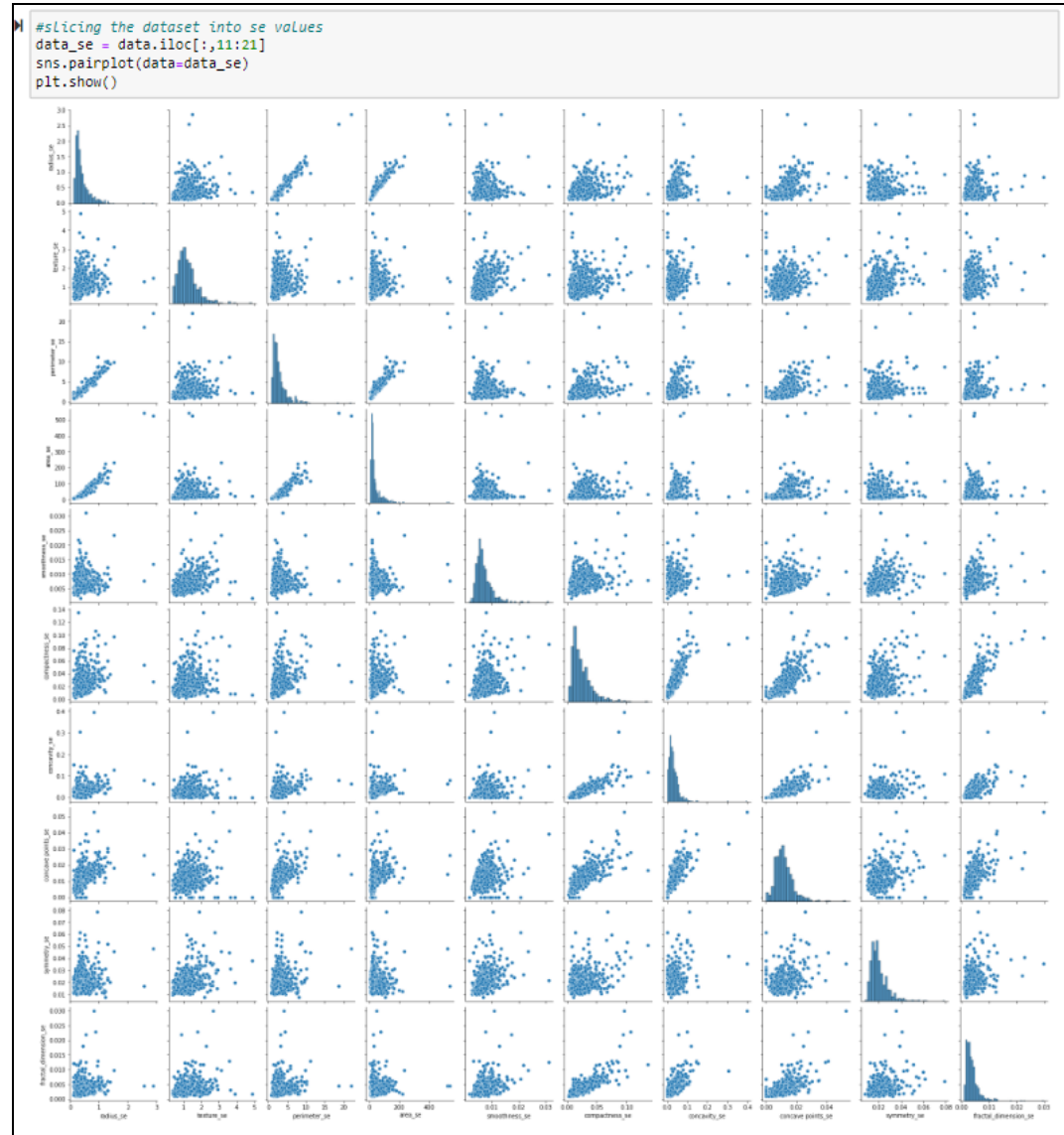
5) The comparison pair plot of the mean data for the feature characteristics

1. The area mean has exponential growth, and the perimeter mean has a linear correlation.
2. Most of the distributions are left skewed



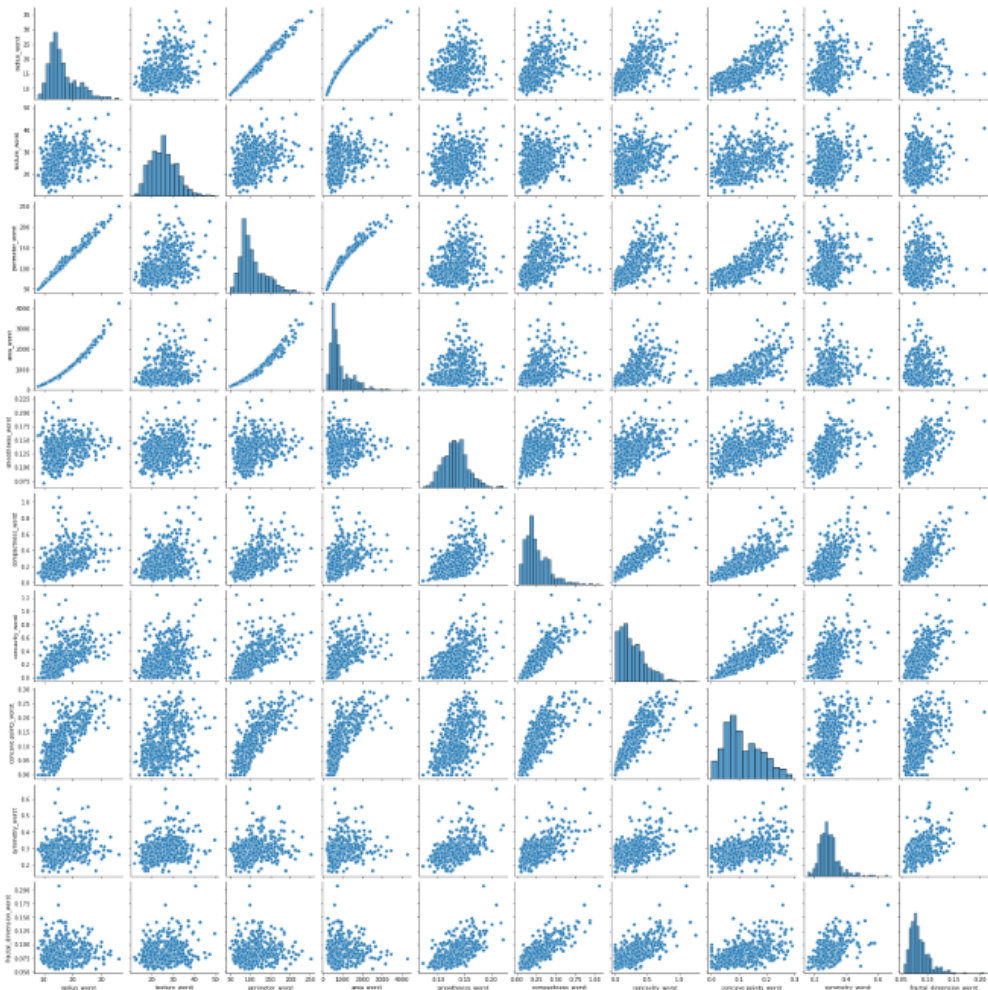
6) The comparison pair plot of the SE data for the feature characteristics

1. Most of the distributions are left skewed
2. We can observe the presence of outliers in the data



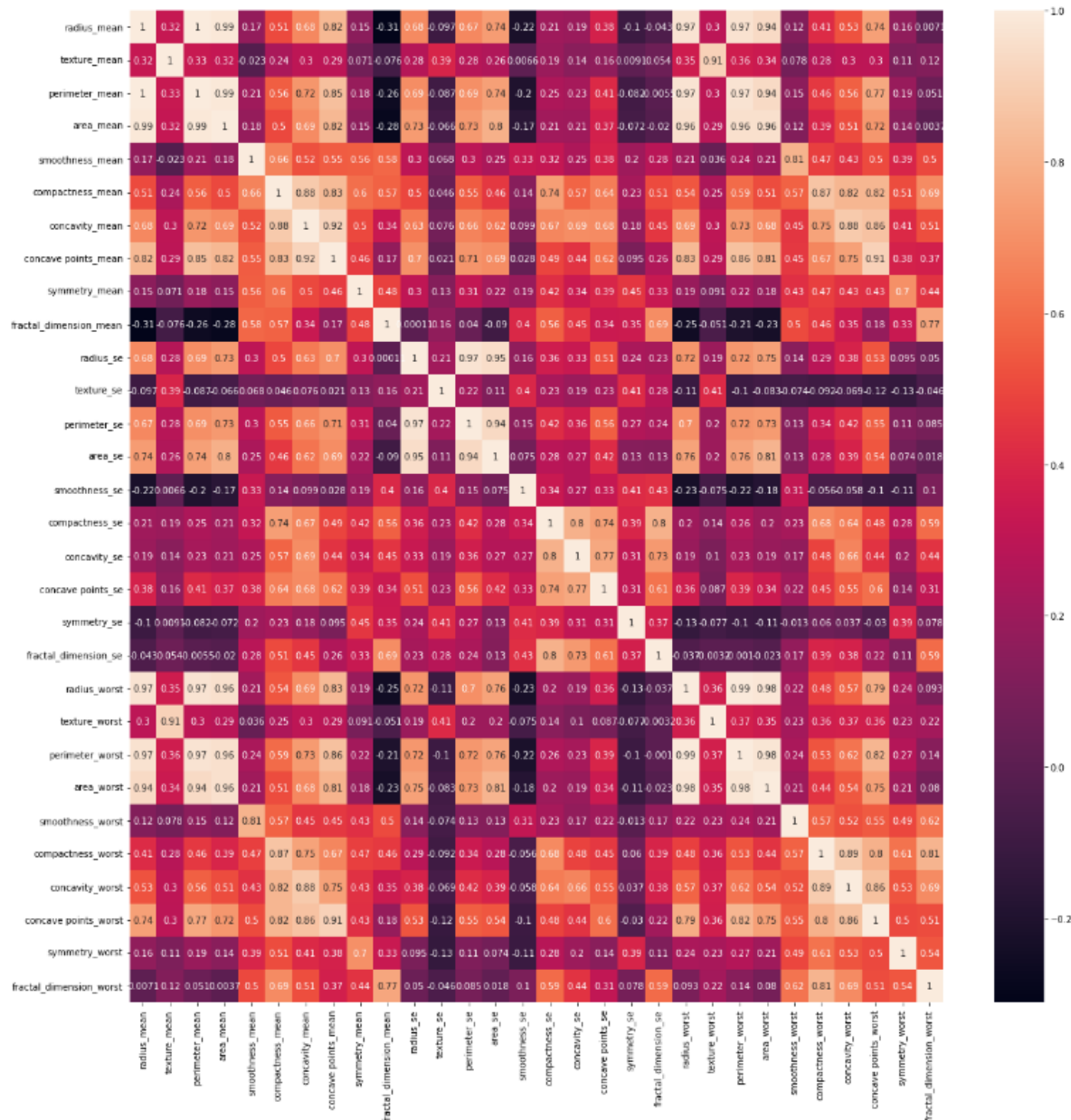
7) The comparison pair plot of the worst data for the feature characteristics

```
#slicing the data with worst parameters
data_worst = data.iloc[:,21:]
sns.pairplot(data=data_worst)
plt.show()
```



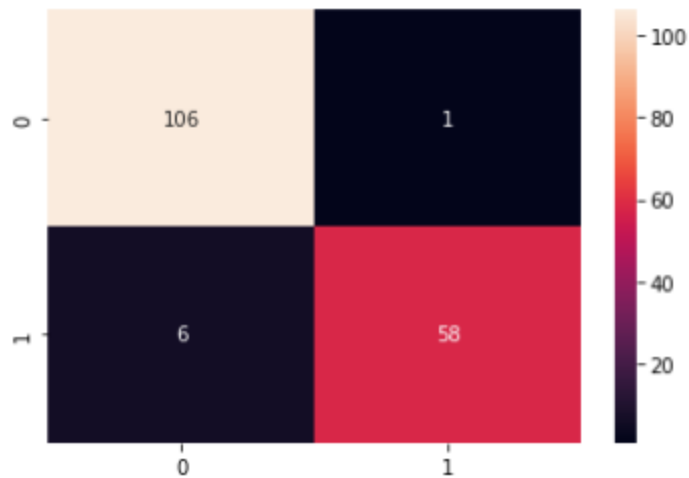
8) The correlation plot of all the data features

```
#correlation plot
plt.figure(figsize=(20,20))
sns.heatmap(data.corr(),annot=True)
plt.show()
```



9) Confusion matrix for LDA with an SVM classifier

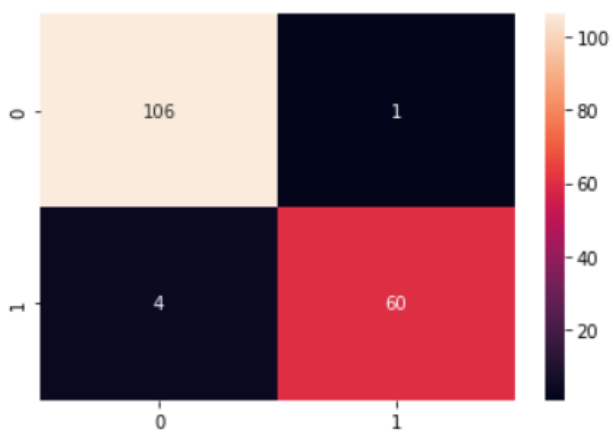
```
#confusion matrix
sns.heatmap(confusion_matrix(y_test,preds),annot=True,fmt='g')
plt.show()
```



10) Confusion matrix for QDA with an SVM classifier

```
confusion_matrix_qda = confusion_matrix(y_test, y_pred_qda)
false_positive_rate_qda, true_positive_rate_qda, thresholds = roc_curve(
    y_test, y_prob_qda
)
roc_auc_qda = auc(false_positive_rate_qda, true_positive_rate_qda)

#confusion matrix
sns.heatmap(confusion_matrix_qda,annot=True,fmt='g')
plt.show()
```



11) ROC curve QDA model

```
plot_roc(roc_auc_qda, false_positive_rate_qda, true_positive_rate_qda)
```

