**NAME: Suchetha Odapally**

**HALL TICKET:2203A51448**

**BATCH:12**

**AIML ASSIGNMENT-7**

Question 1:

<https://www.kaggle.com/datasets/kellistephenson/increasing-breast-cancer-awareness>

1. Read the data with pandas and find features and target variables
2. Find target variable
3. Train KNN
4. Find accuracy for different K values[2-20], plot a graph(X-axis=K, Y-axis=accuracy)

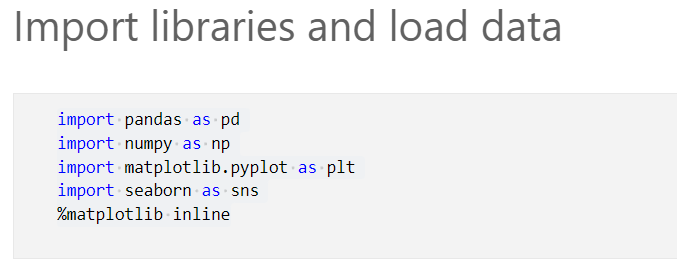
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

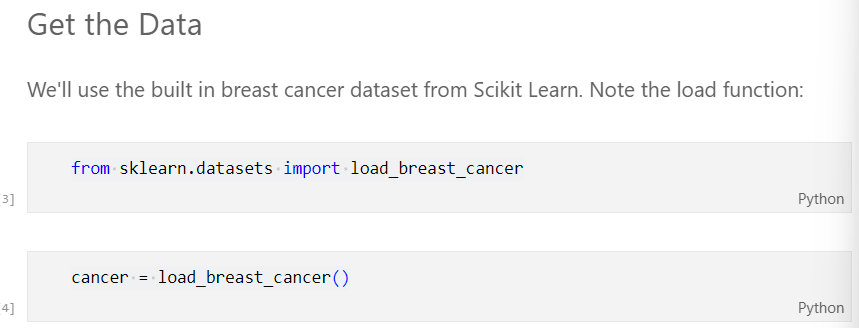
%matplotlib inline



**We will build in breast cancer dataset from scikit learn.**

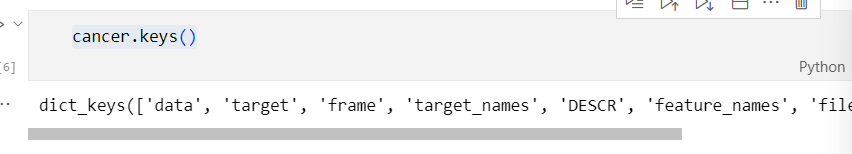
from sklearn.datasets import load\_breast\_cancer

cancer = load\_breast\_cancer()

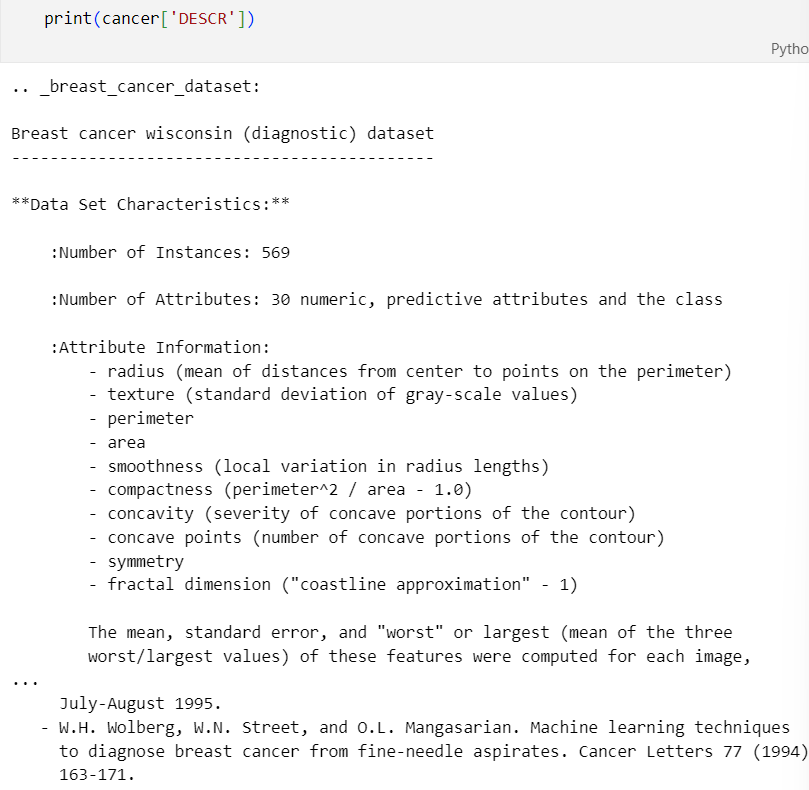


**The data set is presented in a dictionary form**

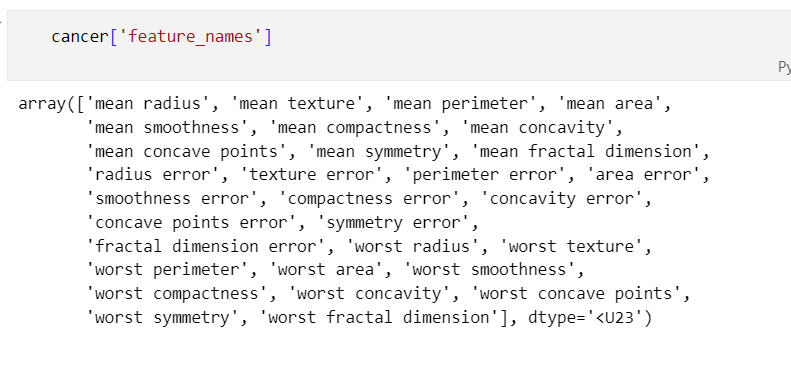
cancer.keys()



The description of features are as follows



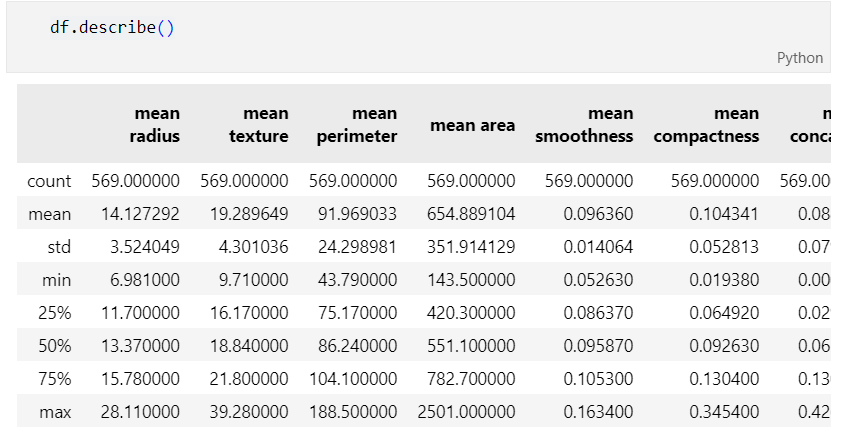
Show the feature names



Set up the data frame



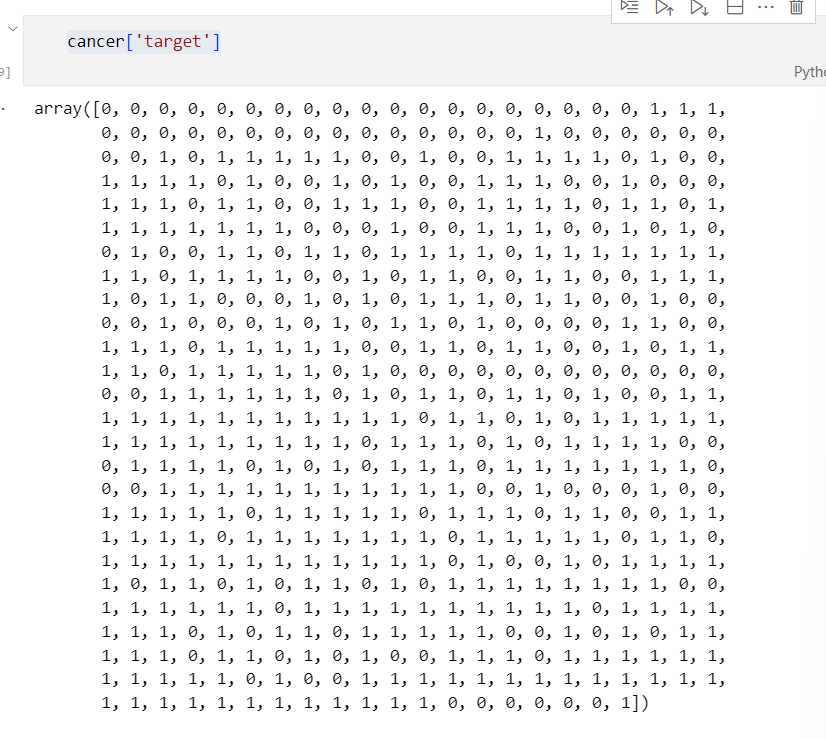
df.describe()

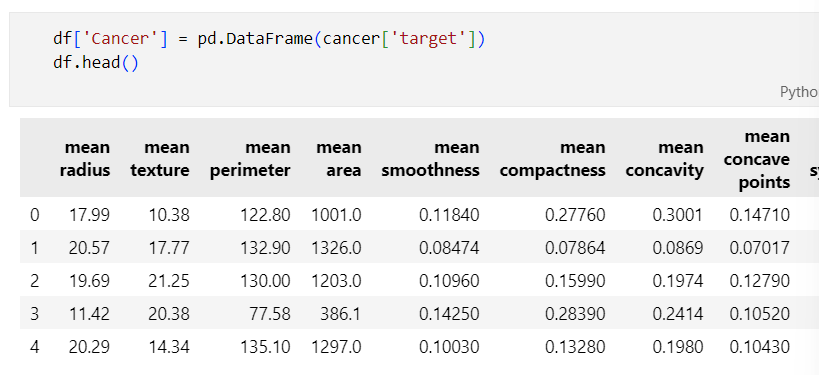


# Sum of the count of null objects in all columns of data frame

np.sum(pd.isnull(df).sum())

B. cancer['target']





C. sns.set\_style('whitegrid')

sns.countplot(x='Cancer',data=df,palette='RdBu\_r')

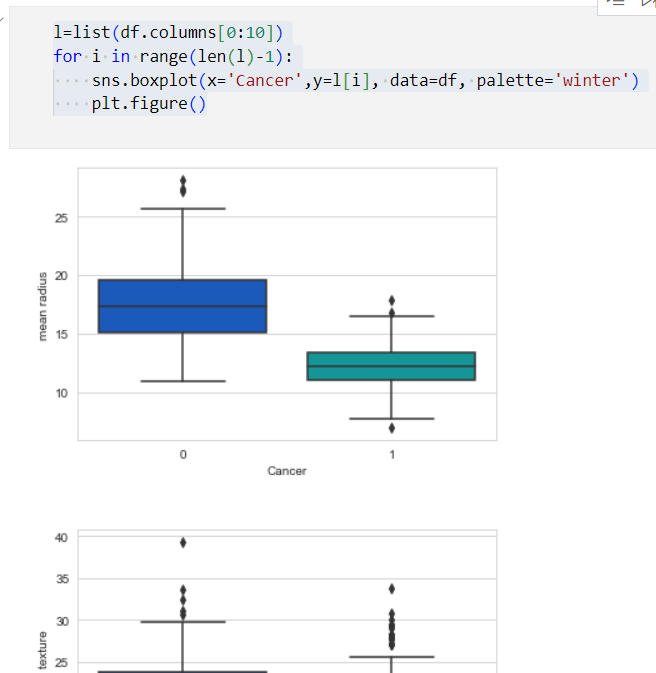


l=list(df.columns[0:10])

for i in range(len(l)-1):

    sns.boxplot(x='Cancer',y=l[i], data=df, palette='winter')

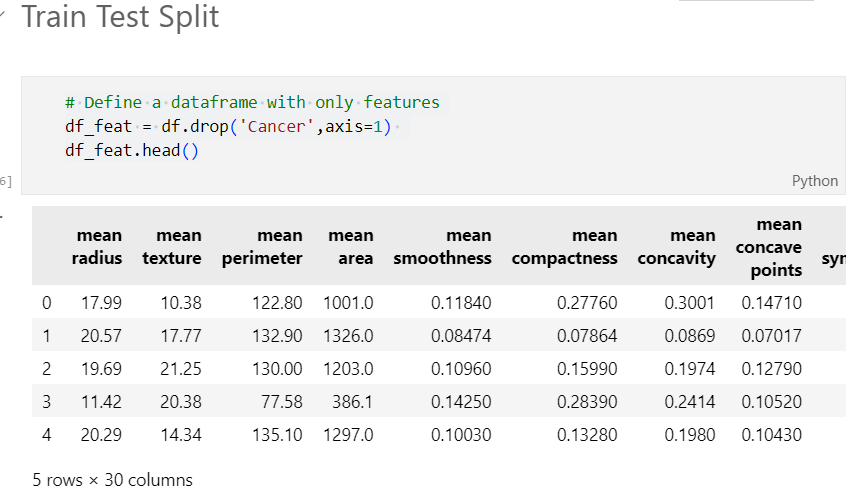
    plt.figure()



D. # Define a dataframe with only features

df\_feat = df.drop('Cancer',axis=1)

df\_feat.head()



# Define a dataframe with only target results for

#cancer detection

df\_target = df['Cancer']

df\_target.head()

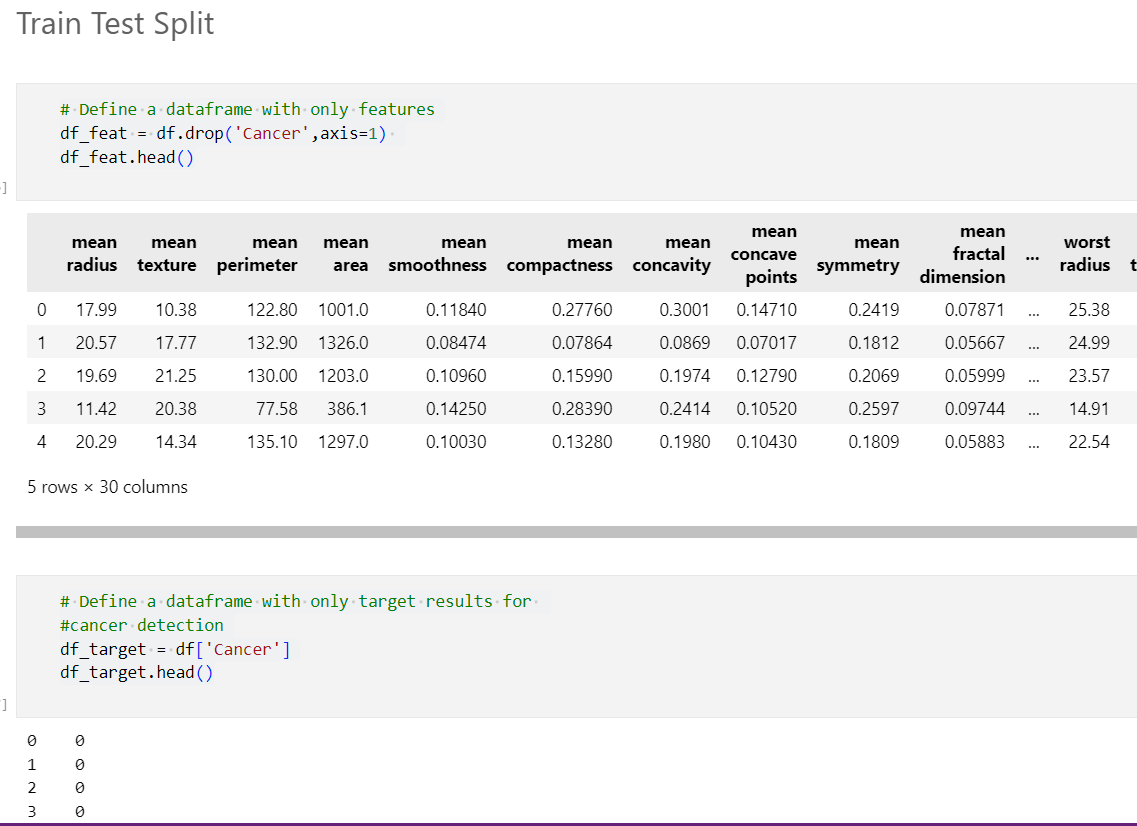
from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test =

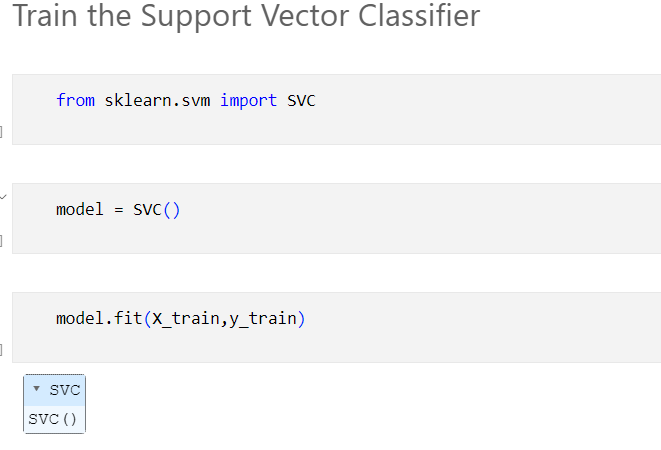
train\_test\_split(df\_feat, df\_target,test\_size=0.30,

                 random\_state=101)

X\_train.head()







**Question 2:  
By using above data set  
• Train Decision Tree  
And find accuracy, precision, recall, f1-score  
Compare both models**



