UNIT 2.4 GRADED ASSIGNMENT

Group members

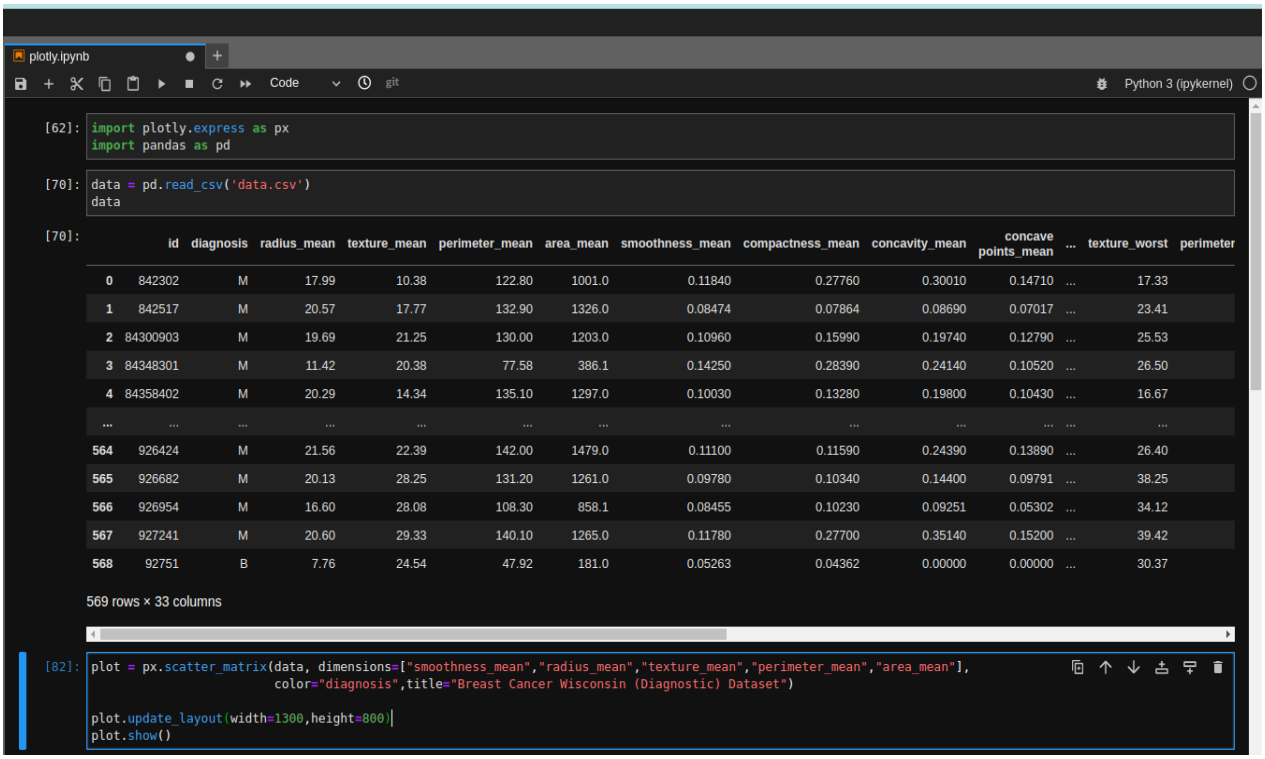
Ifra Saleem (2303.khi.deg.003)
Umaima Siddiqui (2023.KHI.DEG.033)

UNIT 2.4 GRADED ASSIGNMENT

Task:

Download the Breast Cancer Wisconsin dataset from https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data. After downloading, read about scatter matrix and implement it using plotly. Limit it to only a few (5-6) features of your choice. Try to make it as readable as possible (eg. use colors to represent target class)

Solution:

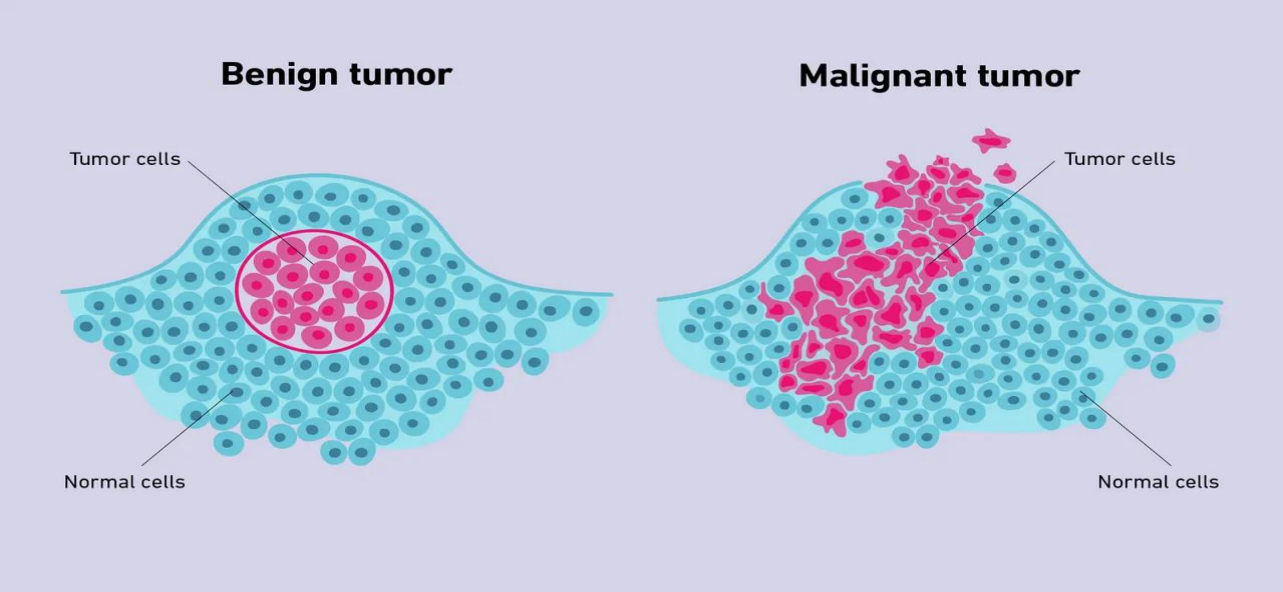


Explanation:

The dataset we used in this work is used to identify the type of breast cancer whether it is a benign tumor or a malignant tumor.

We took five features from the dataset (radius_mean, texture_mean, perimeter_mean, area_mean, smoothness_mean) and using the scatter matrix we plotted the five features from the dataset. And we set the color to the diagnosis (feature) so the features are distributed with respect to the diagnosis feature.

The distribution is according to the Benign tumor and Malignant tumor, so the scatter matrix plot will tell the type of tumor according to the values of the features.



Output: