

Unit 2.4 Graded Assignment: Scatter Matrix

Instructions:

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

Submitted by:

1. Ali Nasir (2303.KHI.DEG.012)
2. Saif ur Rehman (2303.KHI.DEG.007)

Solution:

1. First we have imported the dataset in to our cell.

Dataset

```
import pandas as pd
import numpy as np
import plotly.express as px

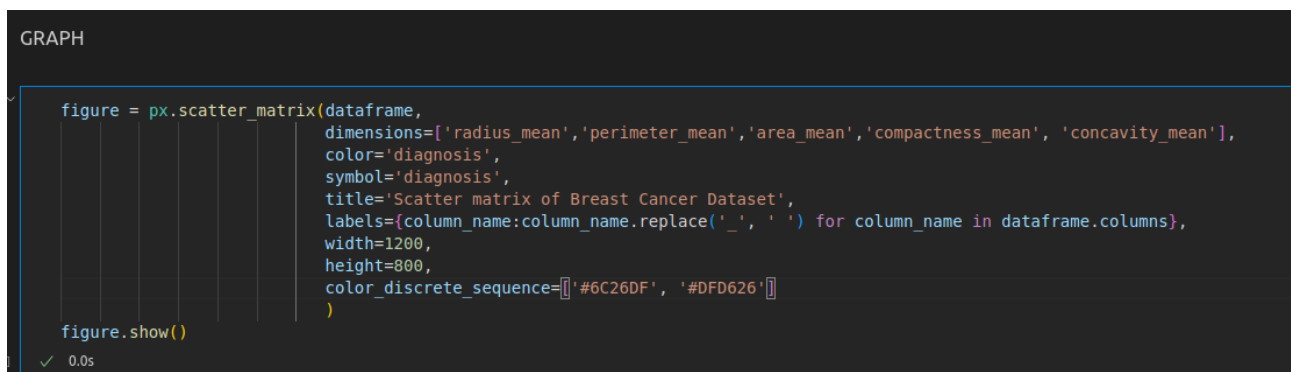
dataframe = pd.read_csv('Datasets/data.csv')
dataframe.sample(10)
```

0.0s Python

	Id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...	texture_worst	perimeter_worst	area_worst
455	9112085	B	13.38	30.72	86.34	557.2	0.09245	0.07426	0.02819	0.03264	...	41.61	96.69	705.
555	924934	B	10.29	27.61	65.67	321.4	0.09030	0.07658	0.05999	0.02738	...	34.91	69.57	357.
536	91979701	M	14.27	22.55	93.77	629.8	0.10380	0.11540	0.14630	0.06139	...	34.27	104.30	728.
328	8953902	M	16.27	20.71	106.90	813.7	0.11690	0.13190	0.14780	0.08488	...	30.38	129.80	1121.
142	869218	B	11.43	17.31	73.66	398.0	0.10920	0.09486	0.02031	0.01861	...	26.76	82.66	503.
264	889719	M	17.19	22.07	111.60	928.3	0.09726	0.08995	0.09061	0.06527	...	29.33	140.50	1436.
369	9012000	M	22.01	21.90	147.20	1482.0	0.10630	0.19540	0.24480	0.15010	...	25.80	195.00	2227.
506	91544001	B	12.22	20.04	79.47	453.1	0.10960	0.11520	0.08175	0.02166	...	24.17	85.13	515.
558	925277	B	14.59	22.68	96.39	657.1	0.08473	0.13300	0.10290	0.03736	...	27.27	105.90	733.
225	88143502	B	14.34	13.47	92.51	641.2	0.09906	0.07624	0.05724	0.04603	...	16.90	110.40	873.

10 rows × 33 columns

2. In the below code snippet, we can see a scatter matrix based on dimensions like radius mean, perimeter mean, area mean, compactness mean, concavity mean. As it is observed that the labels for the graph are in radius_mean format so we use list comprehension to rename the labels as radius mean.



3. Here is the final output of our code.

