Name: Osama Abdul Razzak(2303.KHI.DEG.029)
Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)

Assignment 2.4

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 few (5-6) features of your choice. Try to make it as readable
as possible (eg. use colors to represent target class).

First, I downloaded the .csv from link as per given instruction
After that I import the following required libraries

```python
import pandas as pd

import plotly.express as px
```

than, I fetch data from data.csv file using pandas built in function
pd.read_csv('data.csv') and stored it in a variable

```python
my_data = pd.read_csv('data.csv')

my_data
```

	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	smoothness_worst	compactness_worst	concav
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	...	17.33	184.60	2019.0	0.16220	0.66560	
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	...	23.41	158.80	1956.0	0.12380	0.18660	
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	...	25.53	152.50	1709.0	0.14440	0.42450	
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	...	26.50	98.87	567.7	0.20980	0.86630	
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	...	16.67	152.20	1575.0	0.13740	0.20500	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	...	26.40	166.10	2027.0	0.14100	0.21130	
565	926682	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	...	38.25	155.00	1731.0	0.11660	0.19220	
566	926954	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	...	34.12	126.70	1124.0	0.11390	0.30940	
567	927241	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	...	39.42	184.60	1821.0	0.16500	0.86810	
568	92751	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	...	30.37	59.16	268.6	0.08996	0.06444	

Name: Osama Abdul Razzak(2303.KHI.DEG.029)
*Peer Name: Rahima Siddiqui(2303.KHI.DEG.030)*

In addition, we also stored the columns name into the list name 'features'

```
features = ['radius_mean','texture_mean','perimeter_mean','compactness_mean','concavity_mean']
```

For visually observing the correlation of two columns, we used the scatter_matrix function which is the built-in function of plotly.express and passes it required argument

```
graph = px.scatter_matrix(my_data, dimensions=features, color='diagnosis',color_discrete_sequence=['#FF3333','#E633FF'],height=800)
```

and, here is final output
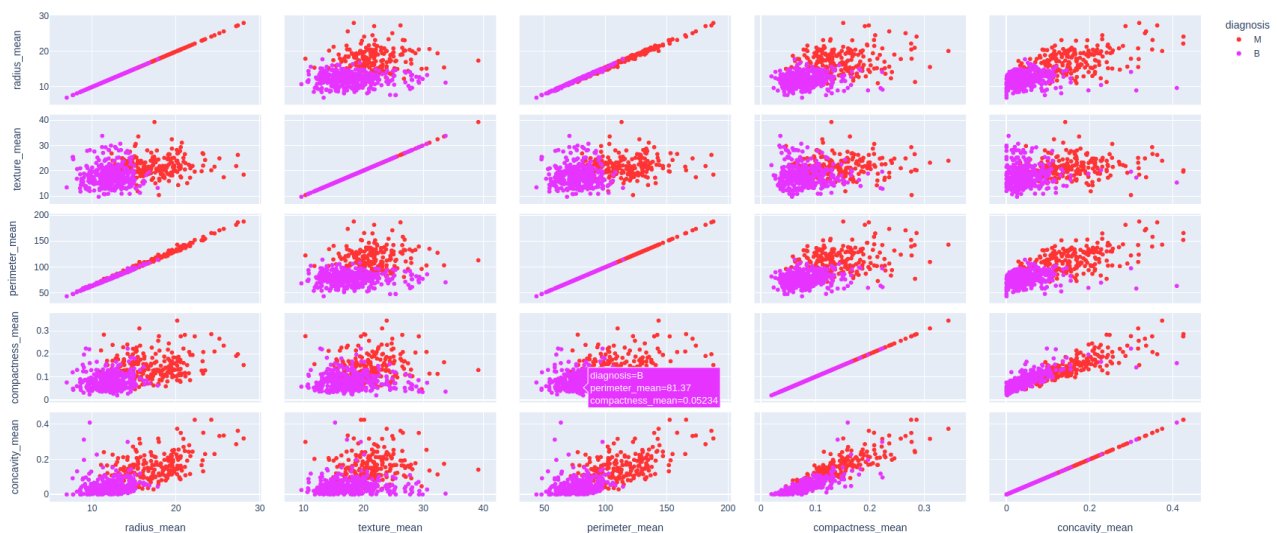
Basically, the graph shows the relationship of two column at a time with respect to diagnosis 'M' and 'B'