Saad+Hamza_Assignment_2.4

April 18, 2023

[ ]:
```
import pandas as pd
import plotly.express as px
```

[ ]:
```
df = pd.read_csv('cancer.csv')
df.head()
```

[ ]:
```
         id diagnosis  radius_mean  texture_mean  perimeter_mean  area_mean  \
0    842302         M        17.99         10.38          122.80     1001.0
1    842517         M        20.57         17.77          132.90     1326.0
2  84300903         M        19.69         21.25          130.00     1203.0
3  84348301         M        11.42         20.38           77.58      386.1
4  84358402         M        20.29         14.34          135.10     1297.0

   smoothness_mean  compactness_mean  concavity_mean  concave points_mean  \
0          0.11840           0.27760          0.3001              0.14710
1          0.08474           0.07864          0.0869              0.07017
2          0.10960           0.15990          0.1974              0.12790
3          0.14250           0.28390          0.2414              0.10520
4          0.10030           0.13280          0.1980              0.10430

   …  texture_worst  perimeter_worst  area_worst  smoothness_worst  \
0  …          17.33           184.60      2019.0            0.1622
1  …          23.41           158.80      1956.0            0.1238
2  …          25.53           152.50      1709.0            0.1444
3  …          26.50            98.87       567.7            0.2098
4  …          16.67           152.20      1575.0            0.1374

   compactness_worst  concavity_worst  concave points_worst  symmetry_worst  \
0             0.6656           0.7119                0.2654          0.4601
1             0.1866           0.2416                0.1860          0.2750
2             0.4245           0.4504                0.2430          0.3613
3             0.8663           0.6869                0.2575          0.6638
4             0.2050           0.4000                0.1625          0.2364

   fractal_dimension_worst  Unnamed: 32
0                  0.11890          NaN
1                  0.08902          NaN
```

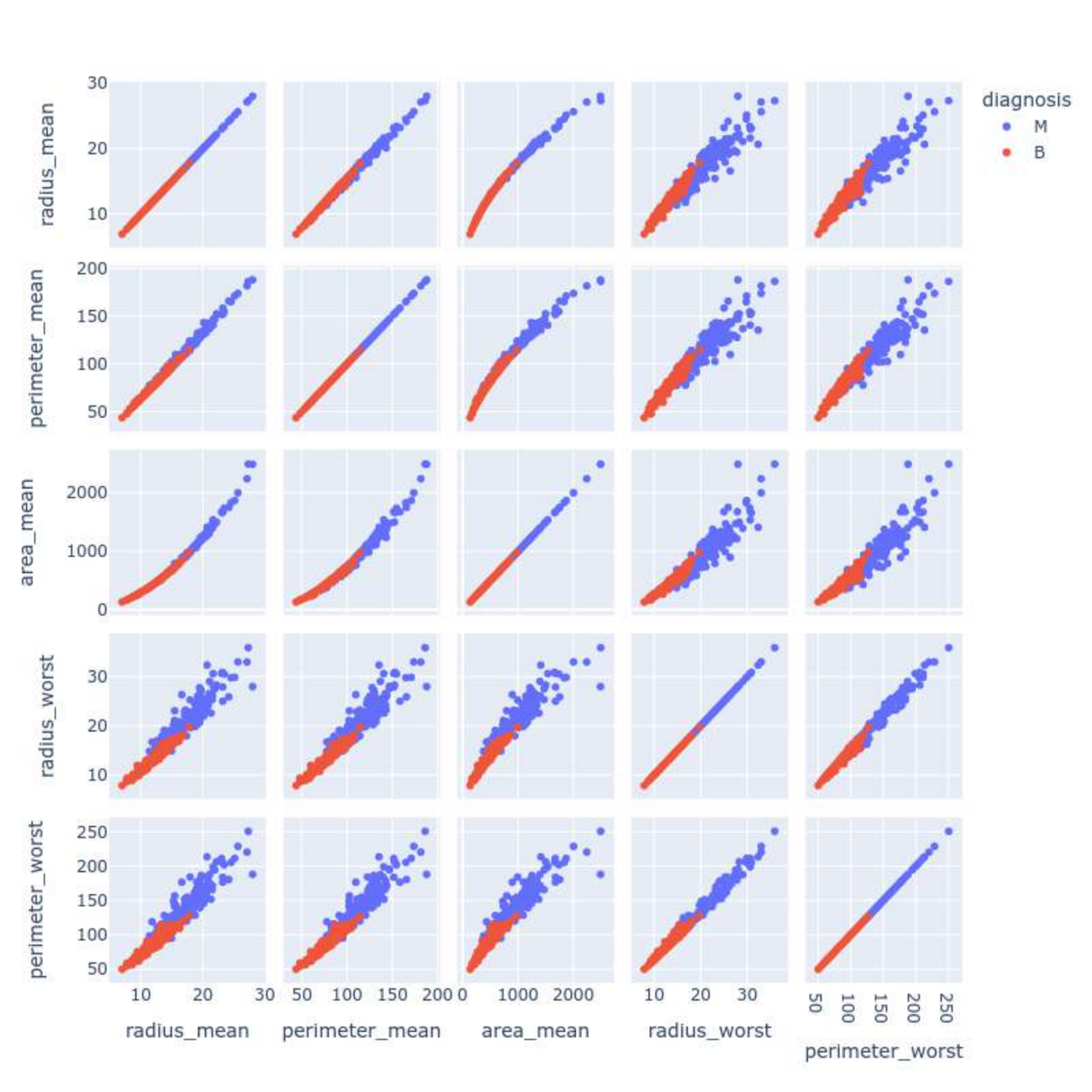
```
2                  0.08758          NaN
3                  0.17300          NaN
4                  0.07678          NaN

[5 rows x 33 columns]
```

[ ]:
```
#selecting the columns we want to use with '.loc' method
selected_df = df.loc[:, ['diagnosis', 'radius_mean',
                         'perimeter_mean', 'area_mean',
                         'radius_worst', 'perimeter_worst']]
selected_df
```

[ ]:
```
    diagnosis  radius_mean  perimeter_mean  area_mean  radius_worst  \
0           M        17.99          122.80     1001.0        25.380
1           M        20.57          132.90     1326.0        24.990
2           M        19.69          130.00     1203.0        23.570
3           M        11.42           77.58      386.1        14.910
4           M        20.29          135.10     1297.0        22.540
..        …           …               …          …            …
564         M        21.56          142.00     1479.0        25.450
565         M        20.13          131.20     1261.0        23.690
566         M        16.60          108.30      858.1        18.980
567         M        20.60          140.10     1265.0        25.740
568         B         7.76           47.92      181.0         9.456

     perimeter_worst
0             184.60
1             158.80
2             152.50
3              98.87
4             152.20
..               …
564           166.10
565           155.00
566           126.70
567           184.60
568            59.16

[569 rows x 6 columns]
```

[ ]:
```
# with color, we specify the column based on whose values we want to assign␣
 ↪colors in our plot
# with dimensions, we choose the columns we want to run the scatter plots␣
 ↪against
fig = px.scatter_matrix(selected_df, color='diagnosis',␣
 ↪dimensions=['radius_mean', 'perimeter_mean',
```

```
 ↪'area_mean', 'radius_worst',                                                ␣
 ↪'perimeter_worst'])                                                         ␣

#adjusting height and width for proper viewing
fig.update_layout(height=800, width=800)
fig.show()
```


diagnosis
M
B
radius_mean
perimeter_mean
area_mean
radius_worst
perimeter_worst