2.4 Assignment Data Visualization

Peer Members:

• Syed Muhammad Raqim Ali Shah (2303.KHI.DEG.008)

• Maaz Javaid Siddique (2303.KHI.DEG.004)

Question:

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).

Solution:

```
import pandas as pd
import plotly.express as explt


data = pd.read_csv("data.csv")



features = ['radius_mean', 'texture_mean', 'perimeter_mean', 'area_mean', 'texture_worst',
            'fractal_dimension_worst']

color_map = {'M': 'M is Blue', 'B': 'B is Red'}
data['color'] = data['diagnosis'].map(color_map)
```

```
df.head()
```

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	...
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	...
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	...
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	...
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	...
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	...

5 rows × 33 columns

First we import the necessary libraries , “pandas” for loading the dataset and “plotly.express” and “plotly.graph_objects” for creating the scatter matrix plot.

Then we loaded our dataset using “pd.raed_csv()” command from “pandas” and choose the 5 features randomly and save them into “features”.

The color parameter is used to specify the column to use for coloring the plots.

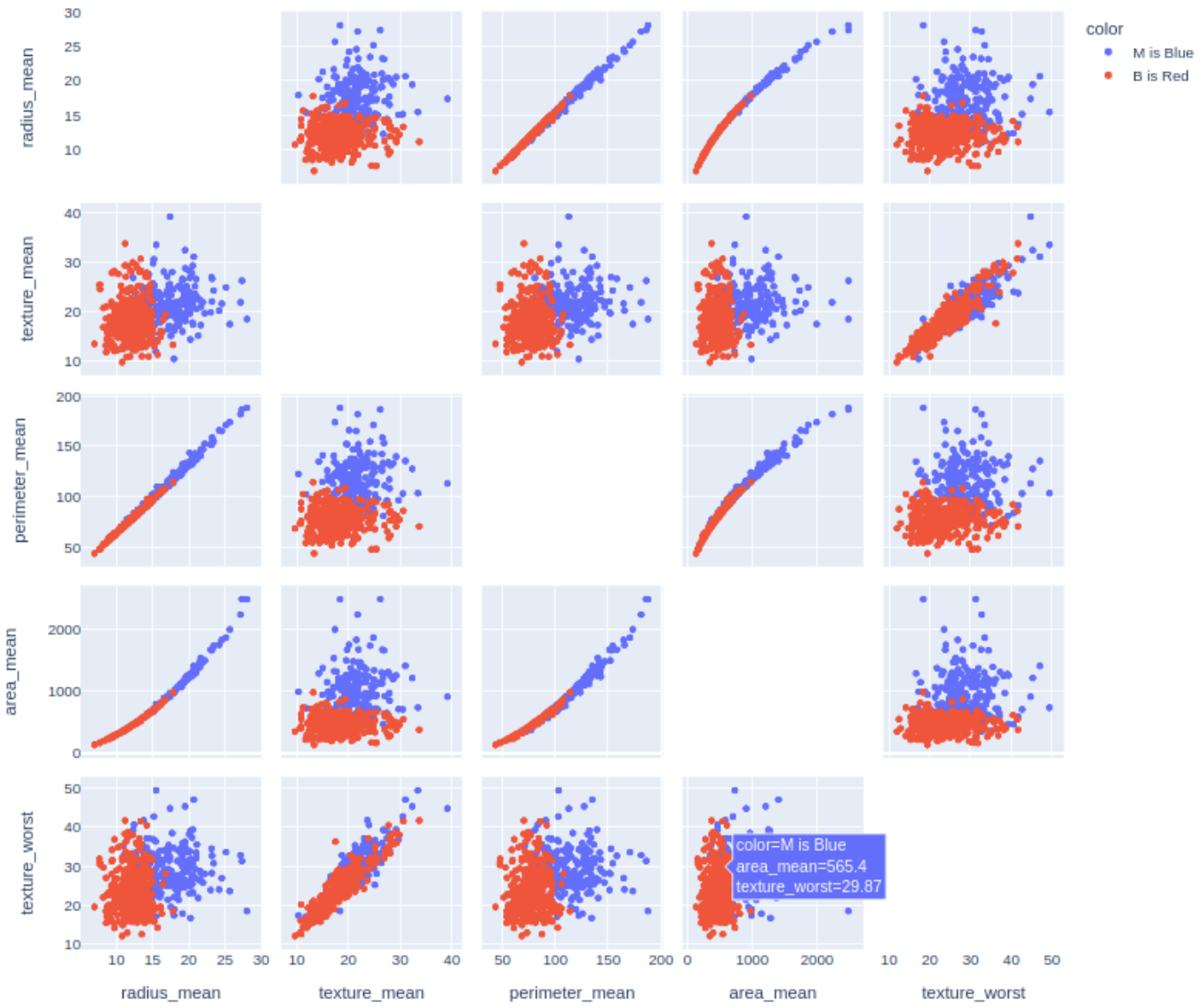
Using “.head()” command we printed 5 starting rows of our dataset.

```
fig = explt.scatter_matrix(data[features], dimensions=features[:-1], color=data['color'], height=900, width=1000)
fig.update_traces(diagonal_visible=False)
fig.show()
```

We used “scatter_matrix()” function from “plotly.express” to create the scatter matrix plot. We used dimensions parameter to specify the features to include in the scatter matrix, and the color parameter to specify the column to use for coloring the plots.

Finally, we used “update_traces()” function to set diagonal_visible=False to hide the diagonal plots. And we used “show()” function to display the plot in the notebook.

Output:


color
M is Blue
B is Red
radius_mean
texture_mean
perimeter_mean
area_mean
texture_worst
color=M is Blue
area_mean=565.4
texture_worst=29.87