In [1]:
```
import warnings, requests, zipfile, io
warnings.simplefilter('ignore')
import pandas as pd
from scipy.io import arff
```

In [2]:
```
f_zip = 'http://archive.ics.uci.edu/ml/machine-learning-databases/00212/vertebra
r = requests.get(f_zip, stream=True)
Vertebral_zip = zipfile.ZipFile(io.BytesIO(r.content))
Vertebral_zip.extractall()
```

In [3]:
```
data = arff.loadarff('column_2C_weka.arff')
df = pd.DataFrame(data[0])
```

In [4]:
```
df.shape
```

Out[4]:
```
(310, 7)
```

In [5]:
```
df.columns
```

Out[5]:
```
Index(['pelvic_incidence', 'pelvic_tilt', 'lumbar_lordosis_angle',
       'sacral_slope', 'pelvic_radius', 'degree_spondylolisthesis', 'class'],
      dtype='object')
```

In [6]:
```
df.dtypes
```

Out[6]:
```
pelvic_incidence           float64
pelvic_tilt                float64
lumbar_lordosis_angle      float64
sacral_slope               float64
pelvic_radius              float64
degree_spondylolisthesis   float64
class                       object
dtype: object
```

In [7]:
```
df['pelvic_incidence'].describe()
```

Out[7]:
```
count    310.000000
mean      60.496653
std       17.236520
min       26.147921
25%       46.430294
50%       58.691038
75%       72.877696
max      129.834041
Name: pelvic_incidence, dtype: float64
```

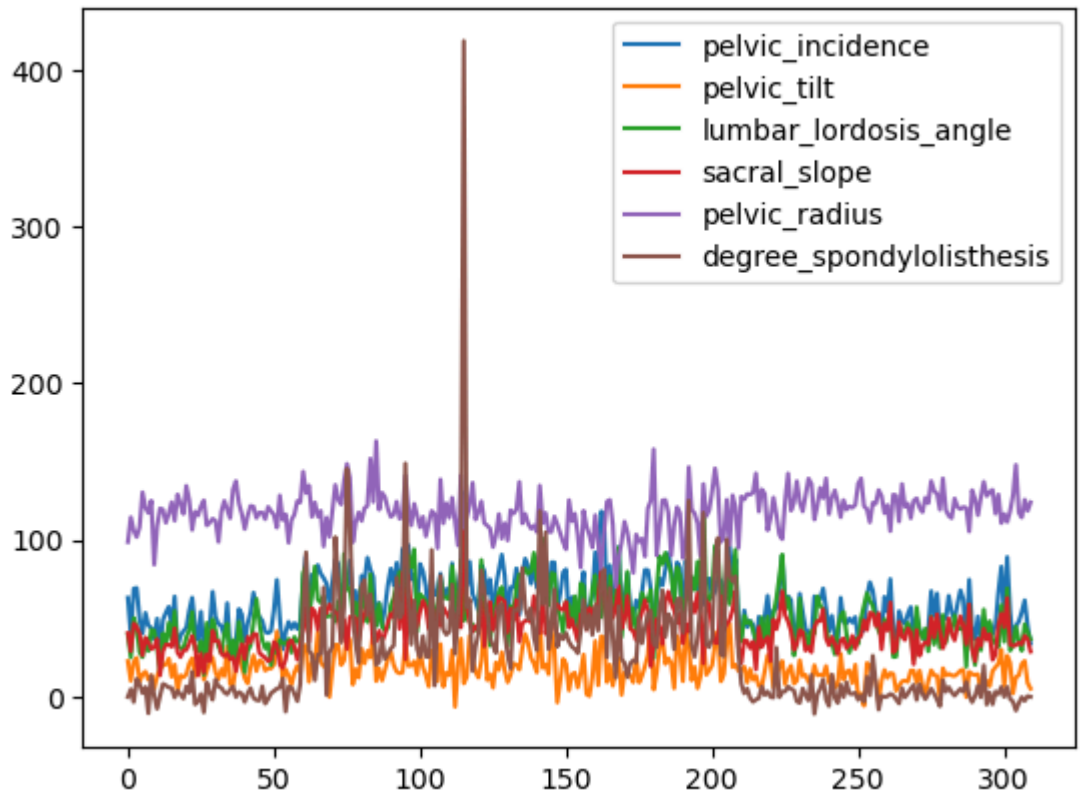
In [8]:
```
df.describe()
```

Out[8]:

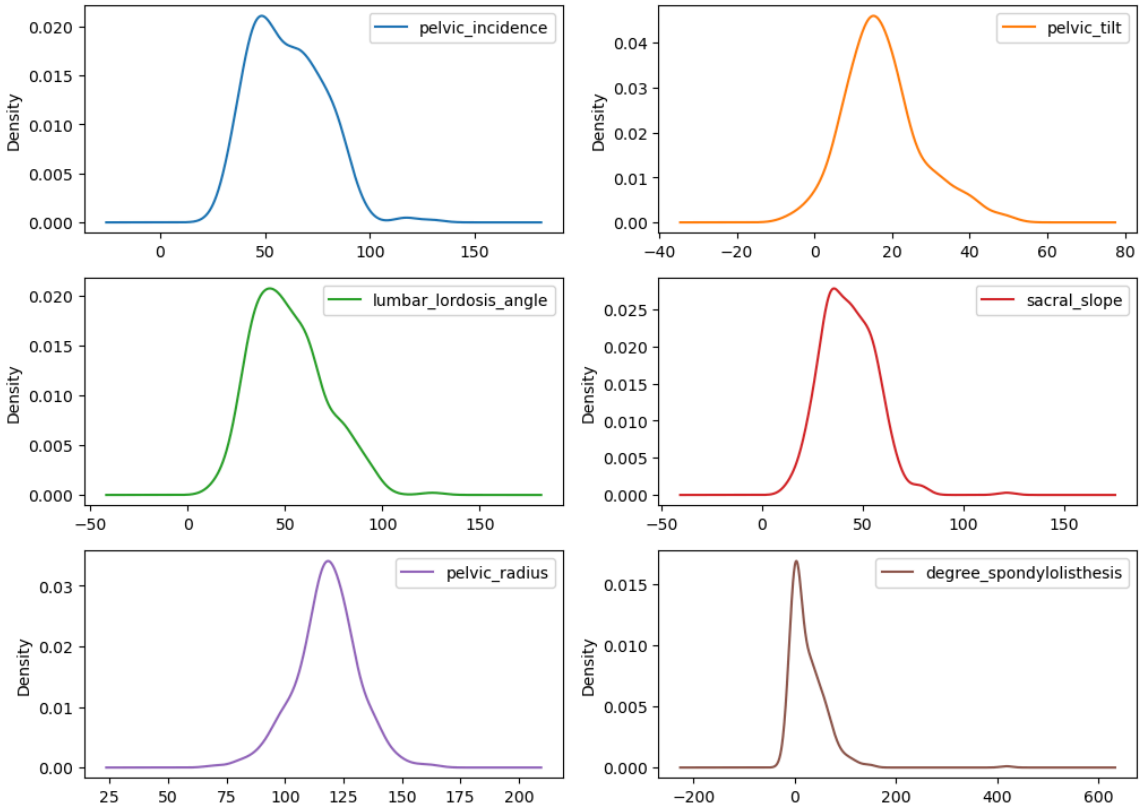
	pelvic_incidence	pelvic_tilt	lumbar_lordosis_angle	sacral_slope	pelvic_radius	degree_s
count	310.000000	310.000000	310.000000	310.000000	310.000000	
mean	60.496653	17.542822	51.930930	42.953831	117.920655	
std	17.236520	10.008330	18.554064	13.423102	13.317377	
min	26.147921	-6.554948	14.000000	13.366931	70.082575	
25%	46.430294	10.667069	37.000000	33.347122	110.709196	
50%	58.691038	16.357689	49.562398	42.404912	118.268178	
75%	72.877696	22.120395	63.000000	52.695888	125.467674	
max	129.834041	49.431864	125.742385	121.429566	163.071041	

In [9]:
```
import matplotlib.pyplot as plt
%matplotlib inline
df.plot()
```

Out[9]: <Axes: >

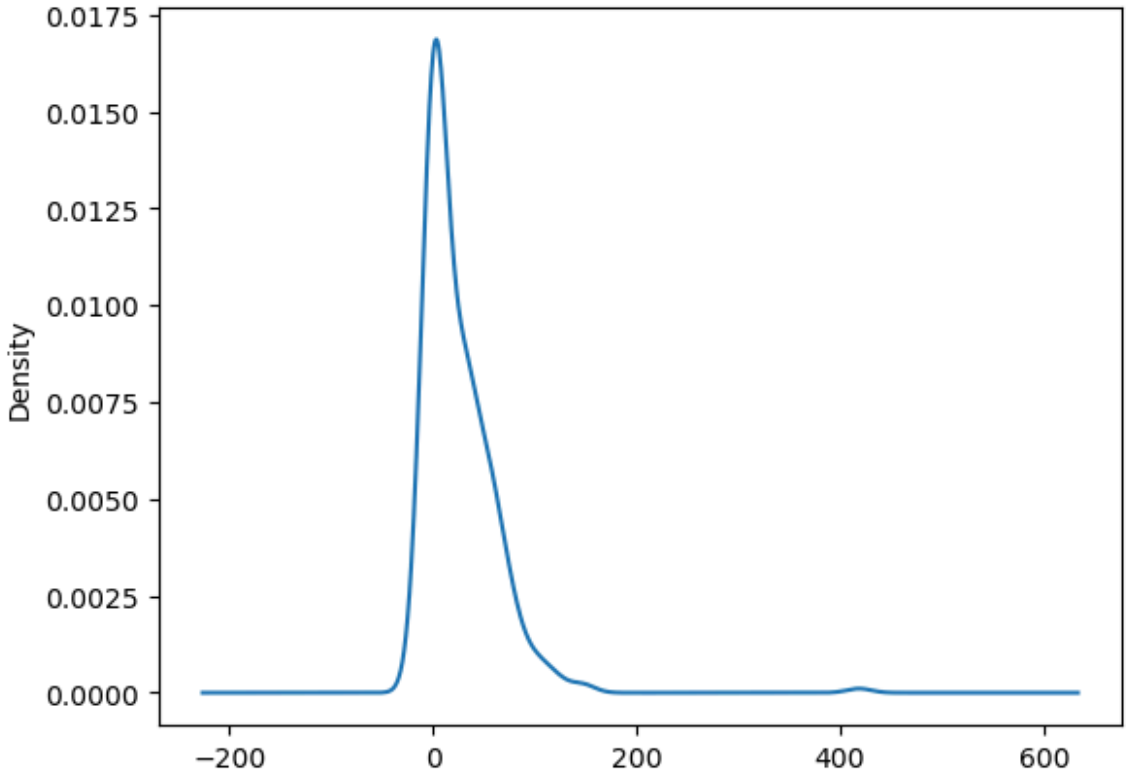


In [10]:
```
df.plot(kind='density',subplots=True,layout=(4,2),figsize=(12,12),sharex=False)
plt.show()
```



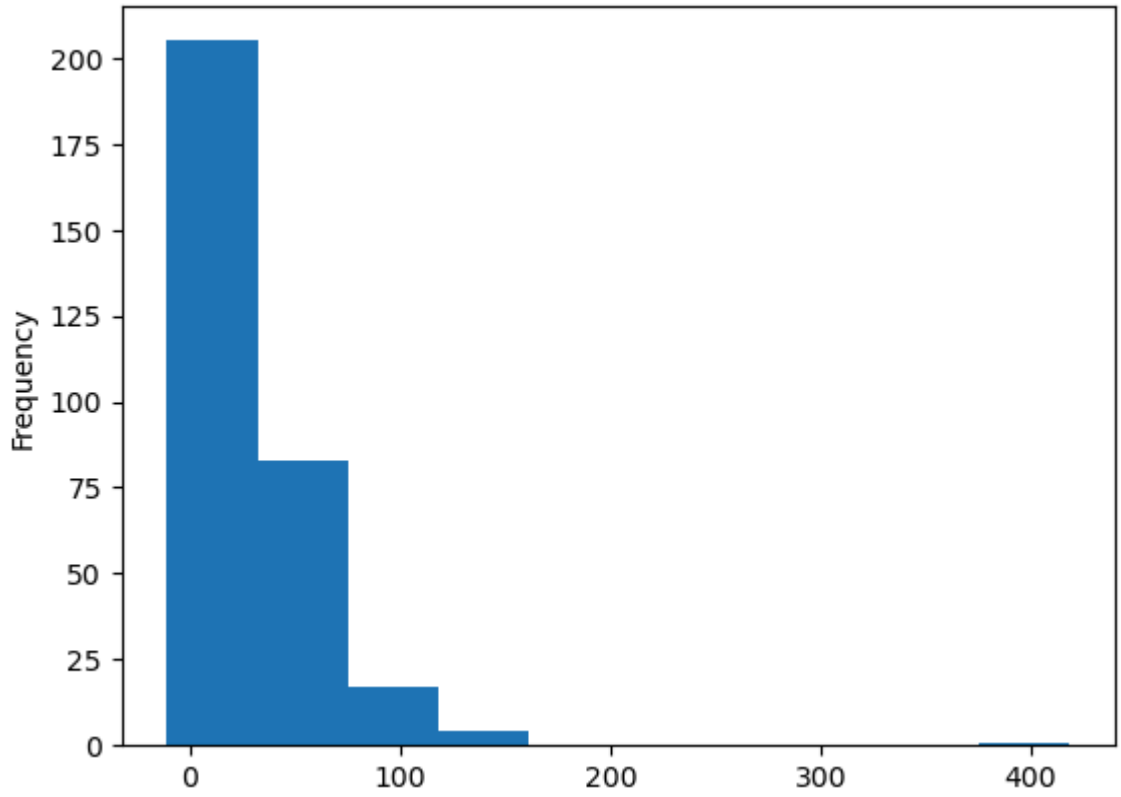
In [11]:
```
df['degree_spondylolisthesis'].plot.density()
```

Out[11]: <Axes: ylabel='Density'>



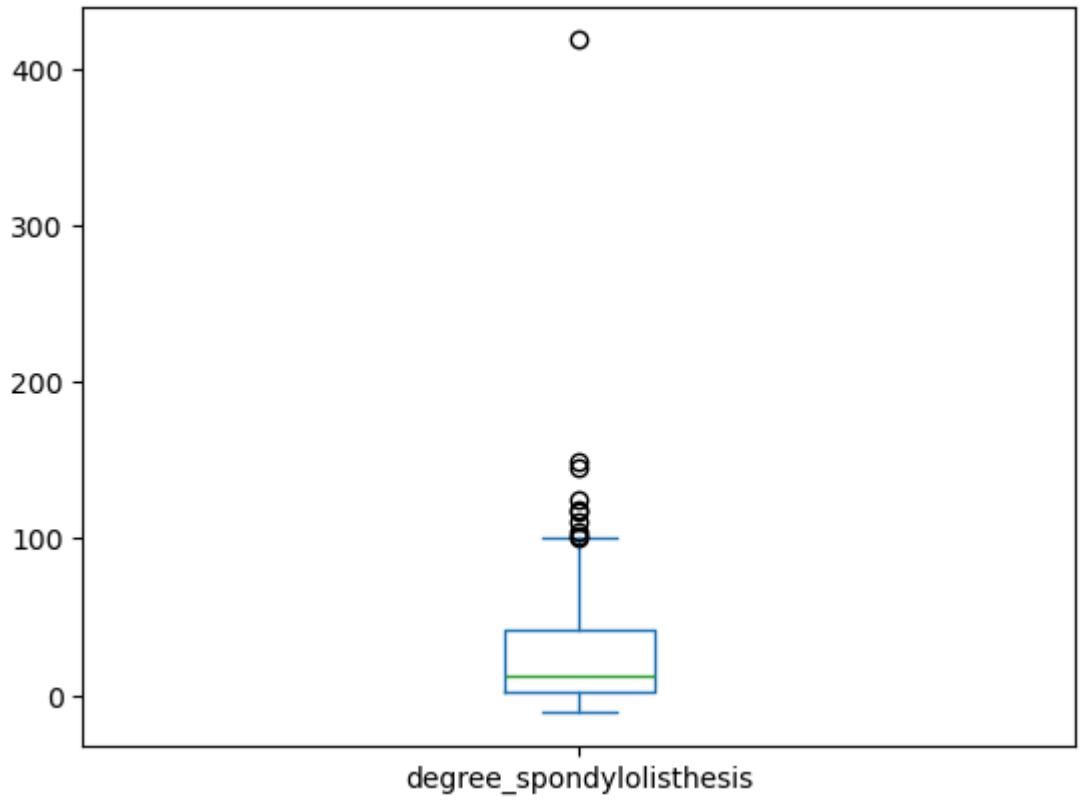
In [12]:
```
df['degree_spondylolisthesis'].plot.hist()
```

Out[12]: <Axes: ylabel='Frequency'>



In [13]:
```
df['degree_spondylolisthesis'].plot.box()
```

Out[13]: <Axes: >



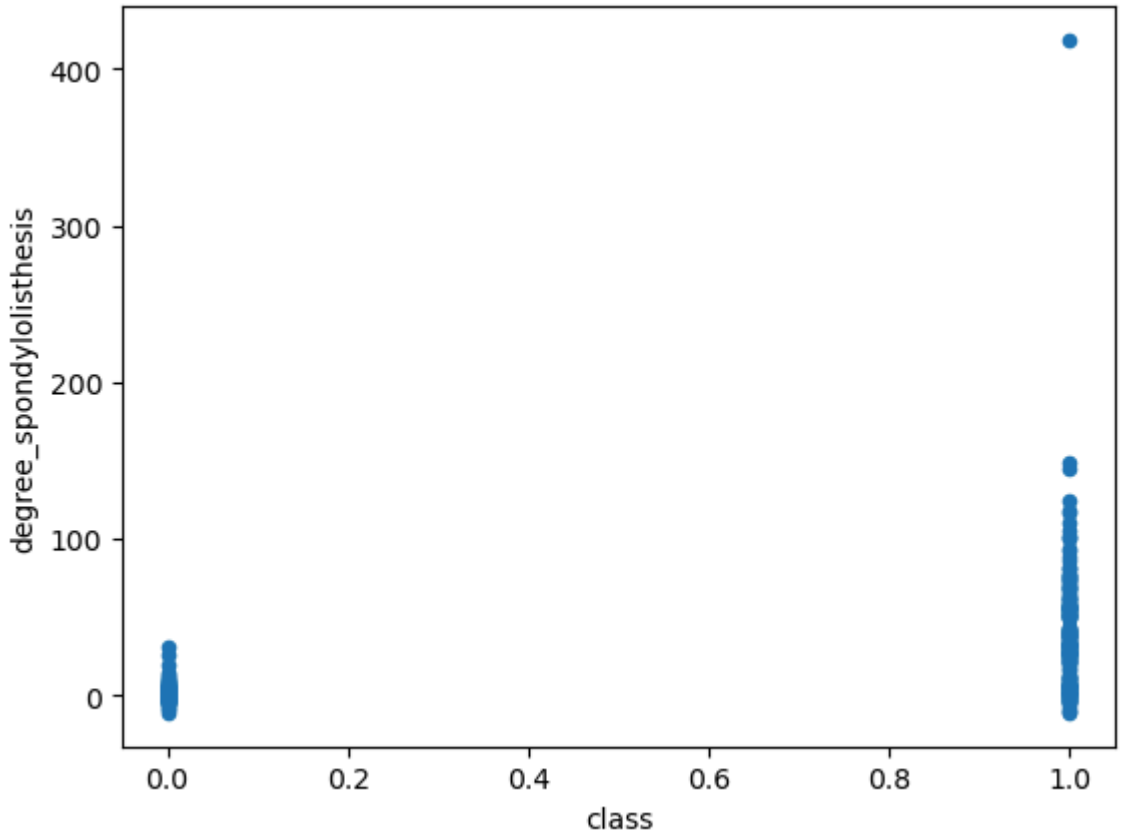
In [14]:
```
df['class'].value_counts()
```

Out[14]:
```
class
b'Abnormal'    210
b'Normal'      100
Name: count, dtype: int64
```

In [15]:
```
class_mapper = {b'Abnormal':1,b'Normal':0}
df['class']=df['class'].replace(class_mapper)
```

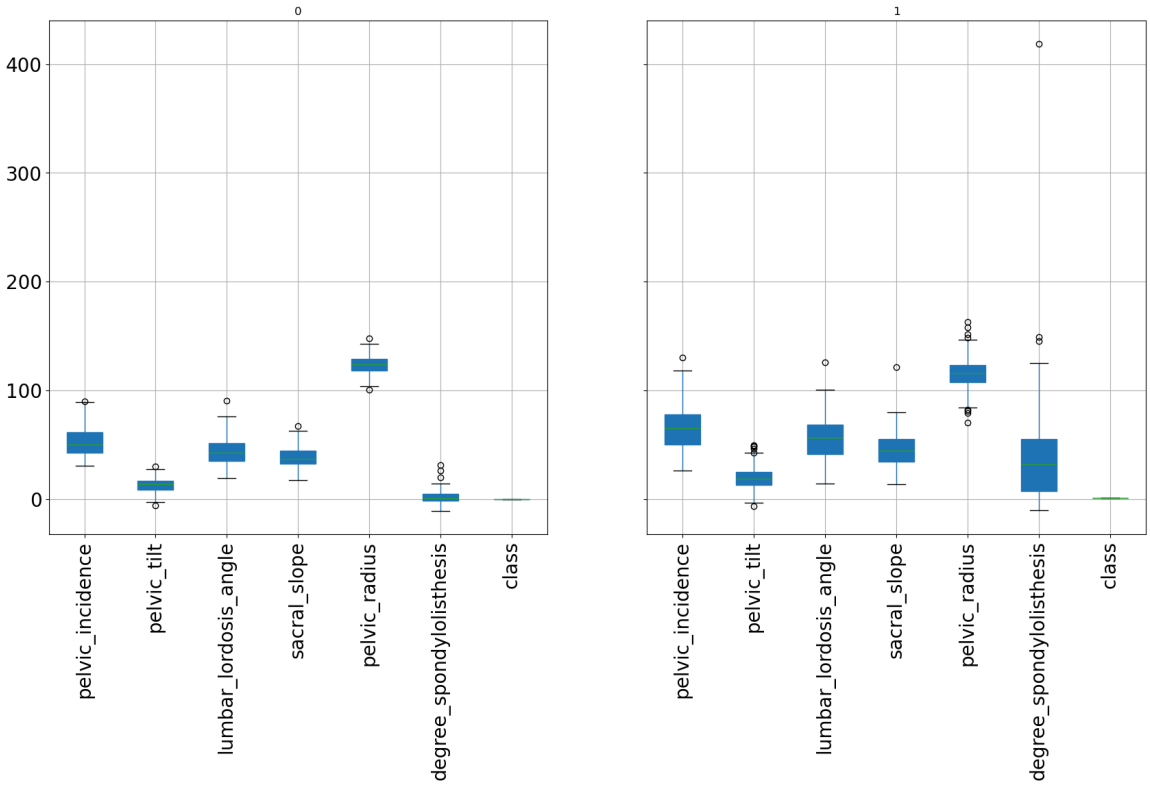
In [16]:
```
df.plot.scatter(y='degree_spondylolisthesis',x='class')
```

Out[16]:
```
<Axes: xlabel='class', ylabel='degree_spondylolisthesis'>
```



In [17]:
```
df.groupby('class').boxplot(fontsize=20,rot=90,figsize=(20,10),patch_artist=True
```

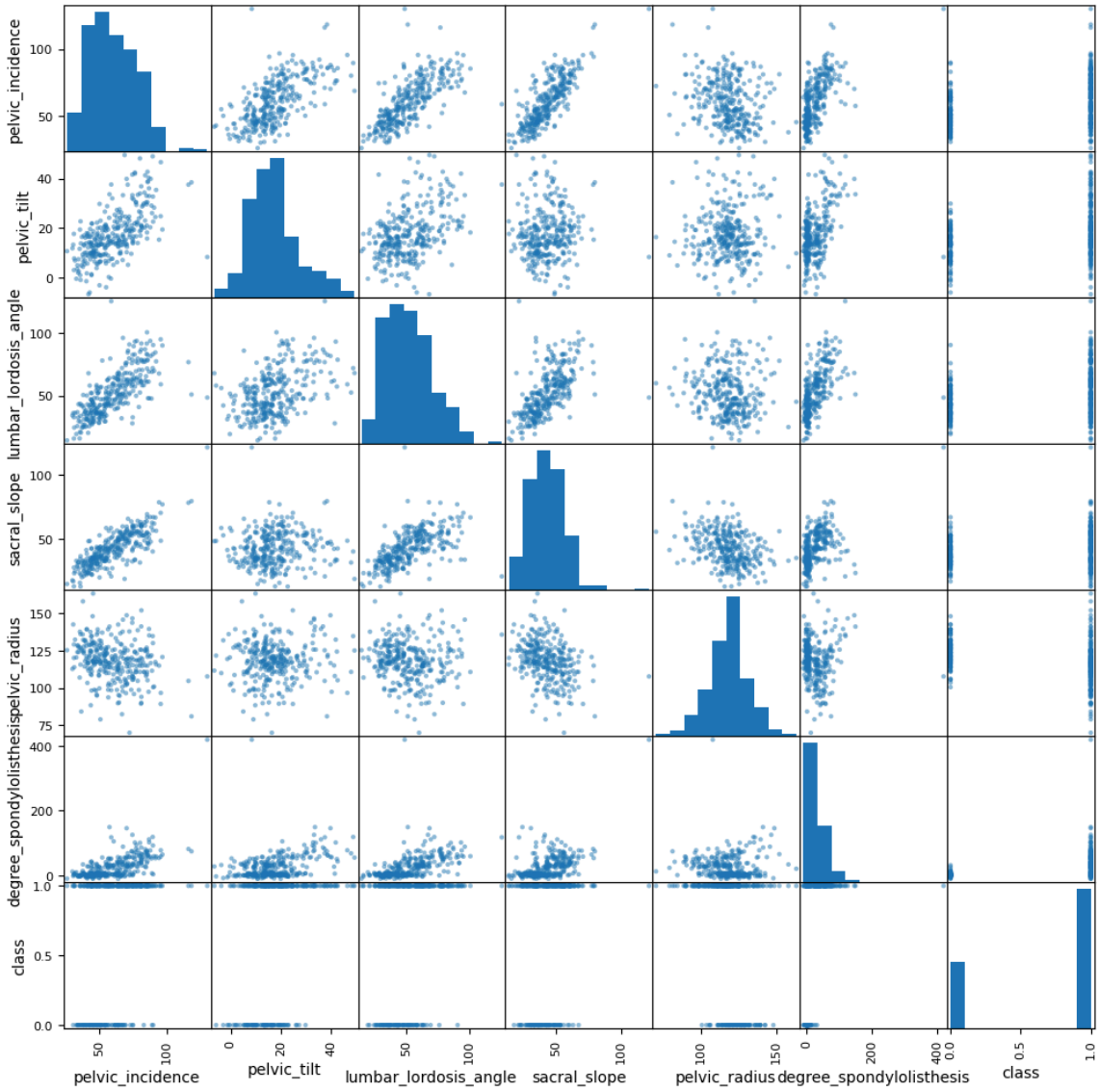
Out[17]:
```
0         Axes(0.1,0.15;0.363636x0.75)
1    Axes(0.536364,0.15;0.363636x0.75)
dtype: object
```



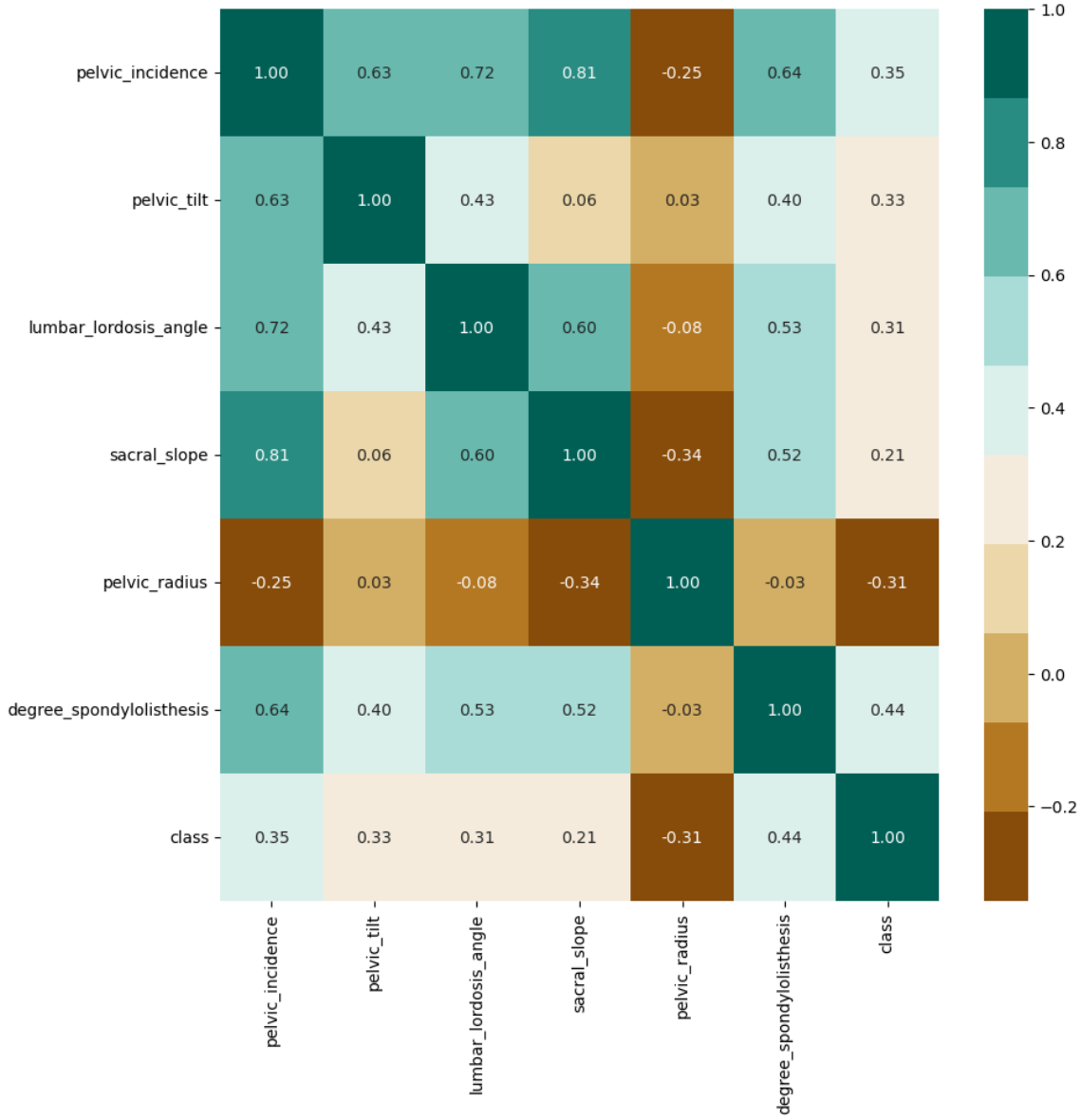
In [18]:
```
corr_matrix = df.corr()
corr_matrix["class"].sort_values(ascending=False)
```

Out[18]:
```
class                      1.000000
degree_spondylolisthesis   0.443687
pelvic_incidence           0.353336
pelvic_tilt                0.326063
lumbar_lordosis_angle      0.312484
sacral_slope               0.210602
pelvic_radius             -0.309857
Name: class, dtype: float64
```

In [19]:
```
pd.plotting.scatter_matrix(df,figsize=(12,12))
plt.show()
```



In [20]:
```
import seaborn as sns
# Plot figsize
fig, ax = plt.subplots(figsize=(10, 10))
# Generate Color Map
# colormap = sns.diverging_palette(220, 10, as_cmap=True)
colormap = sns.color_palette("BrBG", 10)
# Generate Heat Map, allow annotations and place floats in map
sns.heatmap(corr_matrix, cmap=colormap, annot=True, fmt=".2f")
#ax.set_yticklabels(column_names);
plt.show()
```


pelvic_incidence
pelvic_tilt
lumbar_lordosis_angle
sacral_slope
pelvic_radius
degree_spondylolisthesis
class
1.00 0.63 0.72 0.81 -0.25 0.64 0.35
0.63 1.00 0.43 0.06 0.03 0.40 0.33
0.72 0.43 1.00 0.60 -0.08 0.53 0.31
0.81 0.06 0.60 1.00 -0.34 0.52 0.21
-0.25 0.03 -0.08 -0.34 1.00 -0.03 -0.31
0.64 0.40 0.53 0.52 -0.03 1.00 0.44
0.35 0.33 0.31 0.21 -0.31 0.44 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

In [ ]: