breast-cancer-feature-analysis

December 24, 2023

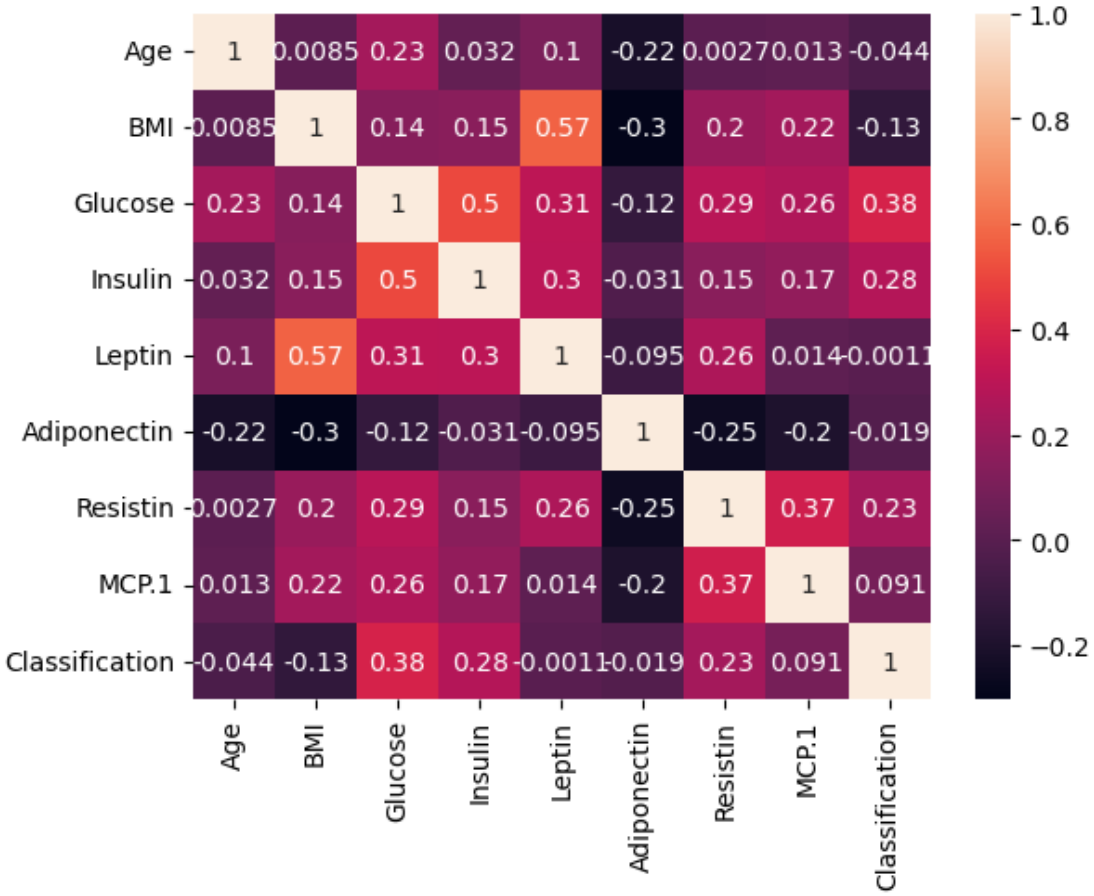
[9]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

[16]:
```
df = pd.read_csv('Datasets/Coimbra_breast_cancer_dataset.csv')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 116 entries, 0 to 115
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Age             116 non-null    int64  
 1   BMI             116 non-null    float64
 2   Glucose         116 non-null    int64  
 3   Insulin         116 non-null    float64
 4   HOMA            116 non-null    float64
 5   Leptin          116 non-null    float64
 6   Adiponectin     116 non-null    float64
 7   Resistin        116 non-null    float64
 8   MCP.1           116 non-null    float64
 9   Classification  116 non-null    int64  
dtypes: float64(7), int64(3)
memory usage: 9.2 KB
```

[18]:
```
sns.heatmap(df.corr(),annot=True)
```

[18]:
```
<Axes: >
```

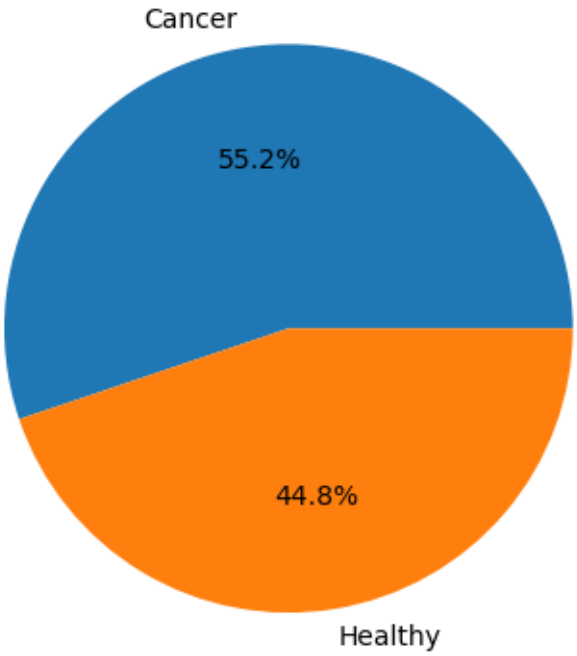


[20]:
```
df.Classification.value_counts()
```

[20]:
```
2    64
1    52
Name: Classification, dtype: int64
```

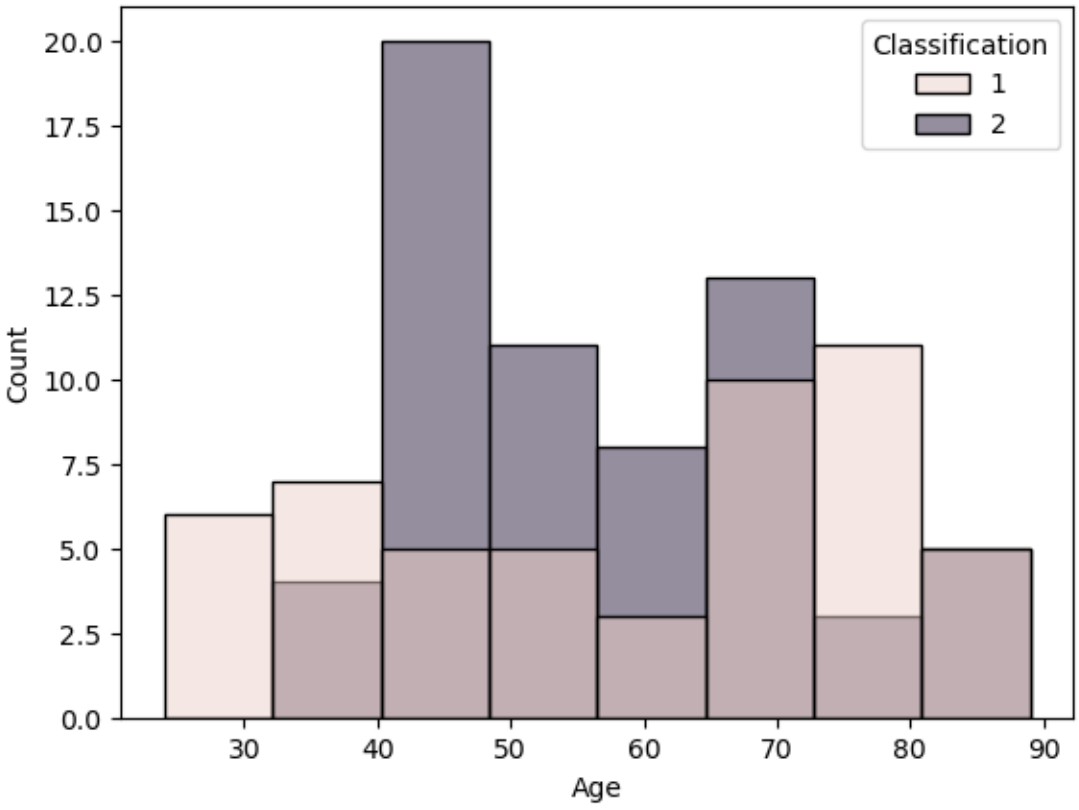
[22]:
```
plt.pie(df['Classification'].
  ↪value_counts(),labels=['Cancer','Healthy'],autopct='%1.1f%%')
```

[22]:
```
([<matplotlib.patches.Wedge at 0x132635c9330>,
  <matplotlib.patches.Wedge at 0x132635c9390>],
 [Text(-0.17796018919888915, 1.085509175944771, 'Cancer'),
  Text(0.17796018919888948, -1.0855091759447708, 'Healthy')],
 [Text(-0.09706919410848498, 0.5920959141516933, '55.2%'),
  Text(0.09706919410848516, -0.5920959141516932, '44.8%')])
```



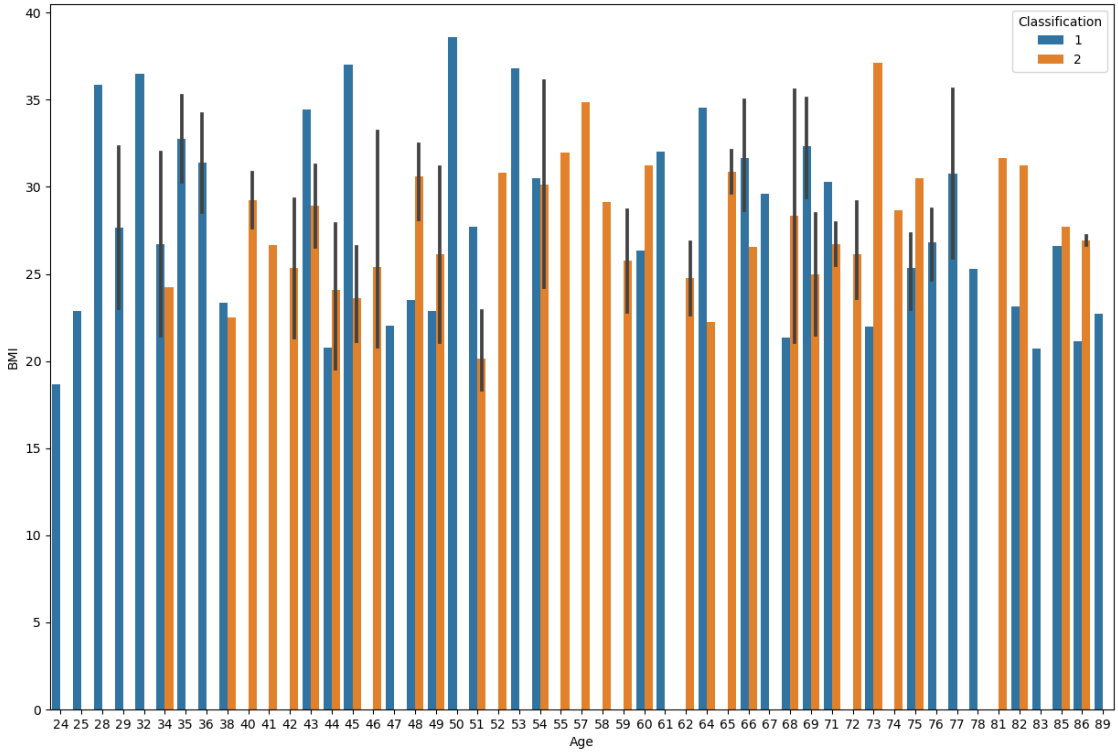
[28]:
```
sns.histplot(data = df, x ='Age' ,hue='Classification')
# Observation : There is a higher incidence of breast cancer among women aged␣
↪30 and 80 based on the available data.
```

[28]:
```
<Axes: xlabel='Age', ylabel='Count'>
```



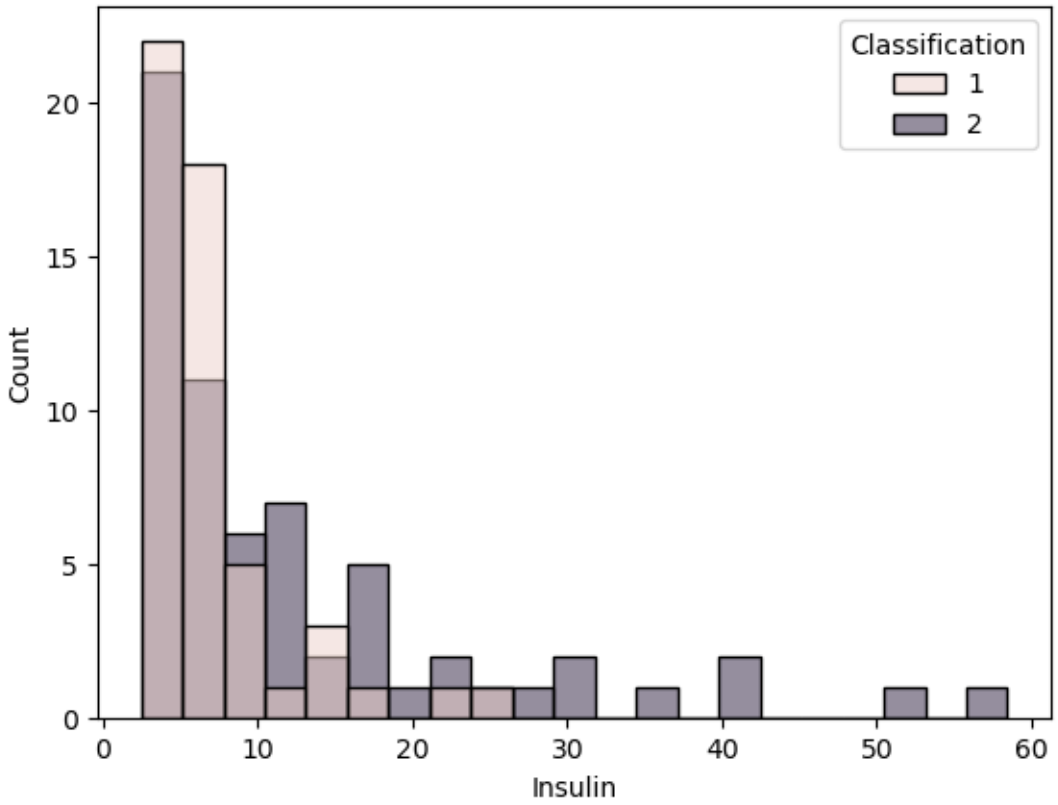
[34]:
```
plt.figure(figsize=(15,10))
sns.barplot(data = df,x = 'Age',y='BMI',hue='Classification')
```

[34]: <Axes: xlabel='Age', ylabel='BMI'>



[37]:
```
sns.histplot(data = df,x = 'Insulin',hue = 'Classification')
# Observation: The value of Insulin has no major effet on Cancer
```

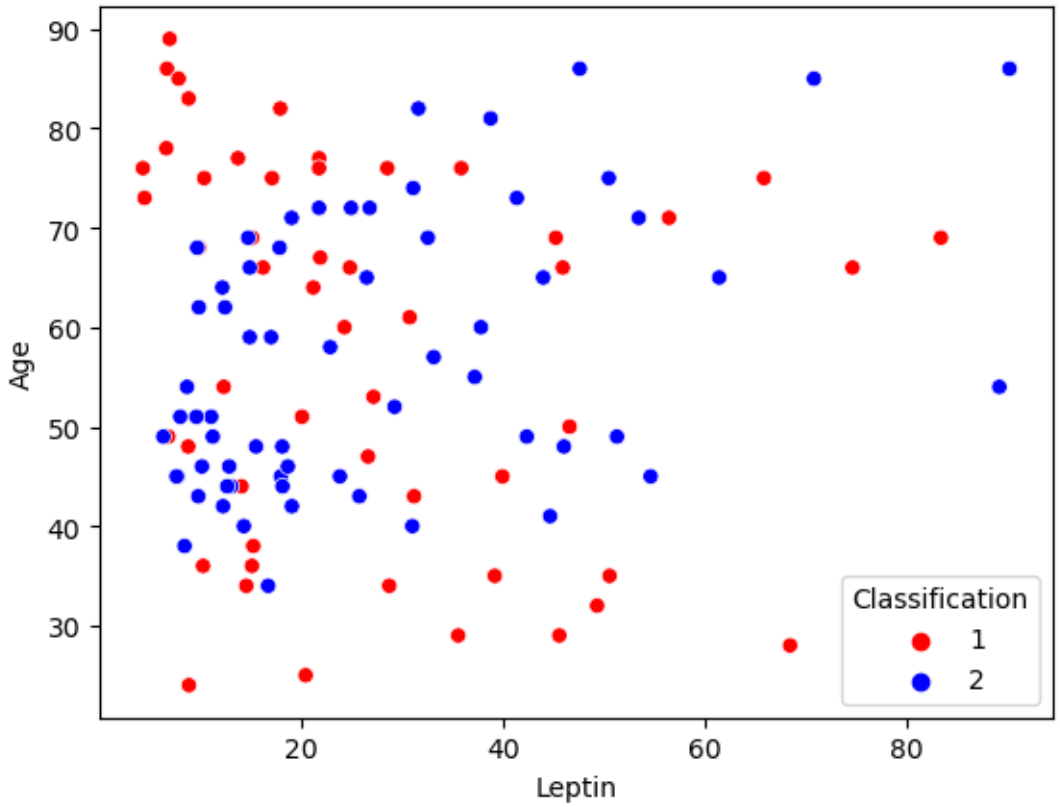
[37]: <Axes: xlabel='Insulin', ylabel='Count'>



Feature Info: Leptin : Leptin, a hormone primarily produced by fat cells, is known to play a role in regulating appetite and metabolism.

[45]:
```
# sns.histplot(data = df,x = 'Leptin',hue = 'Classification')
sns.scatterplot(data = df,x = 'Leptin',y =␣
 ↪'Age',hue='Classification',palette=['r','b'])
# Observation  : Elevated levels of leptin have been associated with an␣
 ↪increased risk of breast cancer,
```

[45]:
```
<Axes: xlabel='Leptin', ylabel='Age'>
```



[46]:
```
df
```

[46]:

	Age	BMI	Glucose	Insulin	Leptin	Adiponectin	Resistin	\
0	48	23.500000	70	2.707	8.8071	9.702400	7.99585	
1	83	20.690495	92	3.115	8.8438	5.429285	4.06405	
2	82	23.124670	91	4.498	17.9393	22.432040	9.27715	
3	68	21.367521	77	3.226	9.8827	7.169560	12.76600	
4	86	21.111111	92	3.549	6.6994	4.819240	10.57635	
..	…	…	…	…	…	…	…	
111	45	26.850000	92	3.330	54.6800	12.100000	10.96000	
112	62	26.840000	100	4.530	12.4500	21.420000	7.32000	
113	65	32.050000	97	5.730	61.4800	22.540000	10.33000	
114	72	25.590000	82	2.820	24.9600	33.750000	3.27000	
115	86	27.180000	138	19.910	90.2800	14.110000	4.35000	

	MCP.1	Classification
0	417.114	1
1	468.786	1
2	554.697	1
3	928.220	1
4	773.920	1

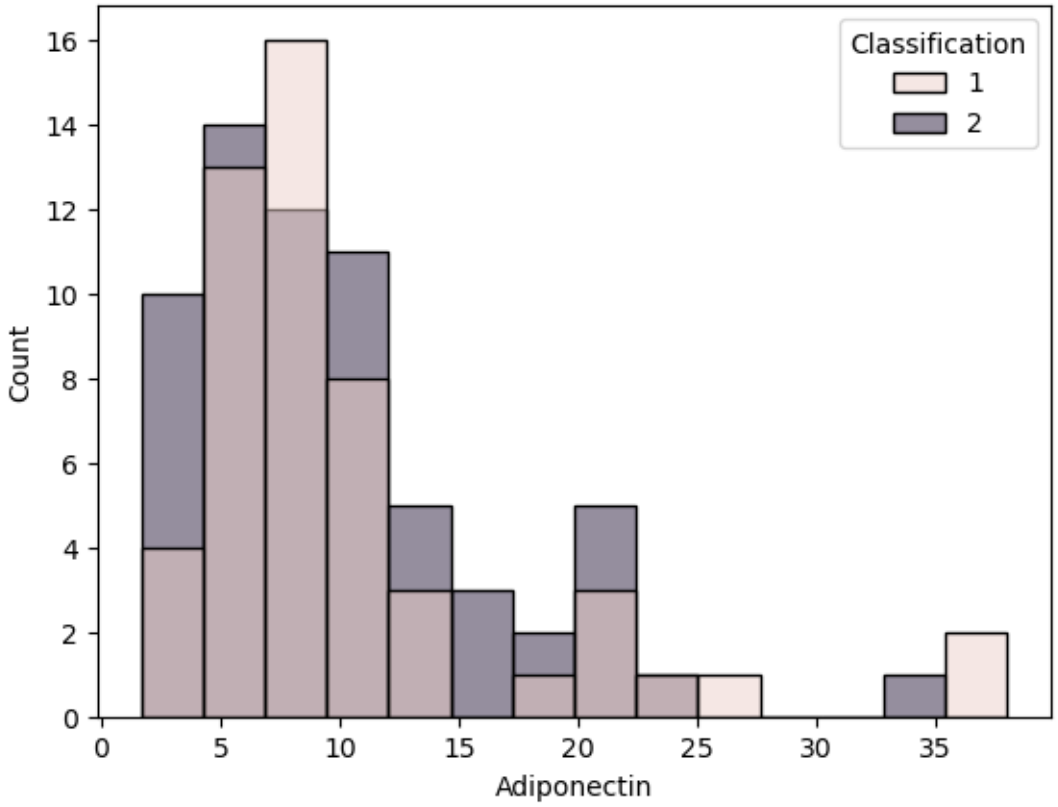
```
..       …              …
111  268.230              2
112  330.160              2
113  314.050              2
114  392.460              2
115   90.090              2

[116 rows x 9 columns]
```

Feature Info: Adiponectin : Leptin, a hormone primarily produced by fat cells, is known to play a role in regulating appetite and metabolism. Adiponectin is a hormone secreted by adipose (fat) tissue, and it plays a crucial role in regulating energy metabolism and insulin sensitivity.

[50]:
```
sns.histplot(data = df,x = 'Adiponectin',hue = 'Classification')
#  Observation - lower levels of adiponectin are associated with an increased␣
 ↪risk of breast cancer.
```

[50]:
```
<Axes: xlabel='Adiponectin', ylabel='Count'>
```



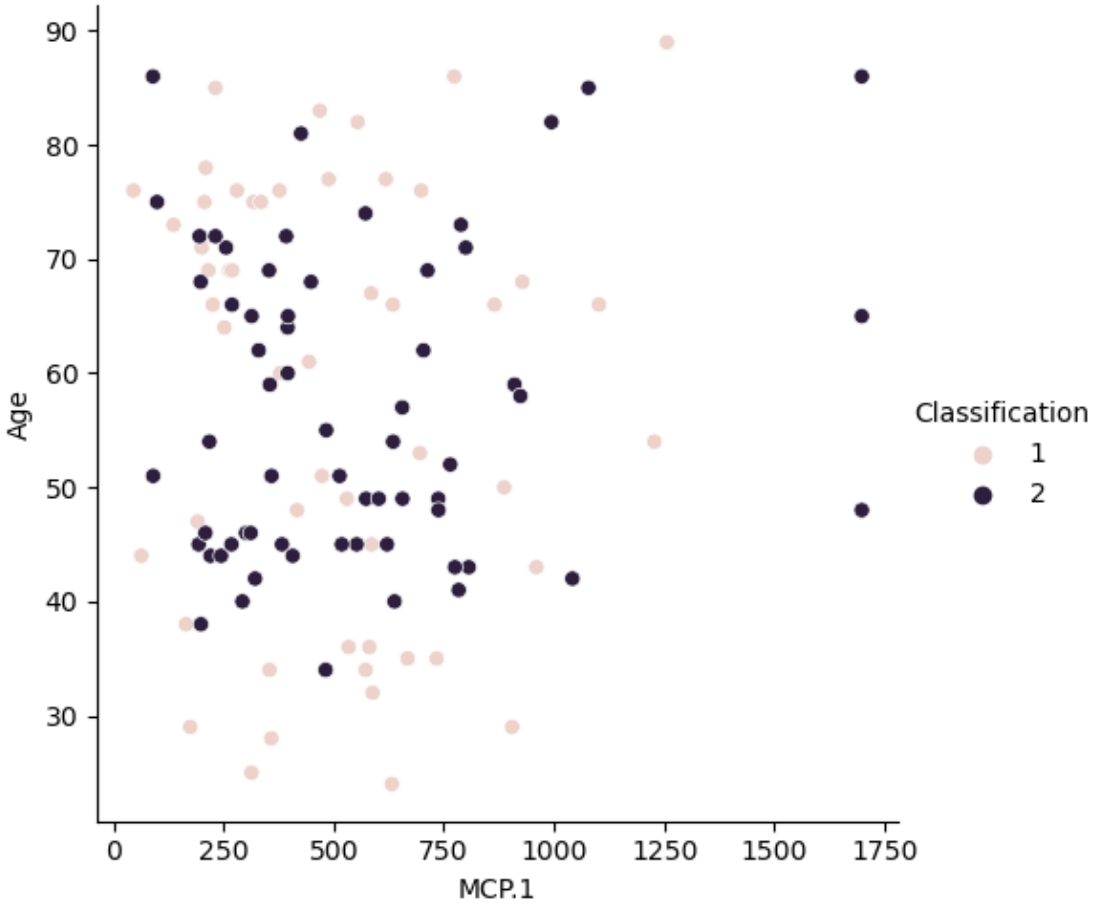
[57]:
```
df.columns
```

[57]:
```
Index(['Age', 'BMI', 'Glucose', 'Insulin', 'Leptin', 'Adiponectin', 'Resistin',
       'MCP.1', 'Classification'],
      dtype='object')
```

Feature Info: Leptin : Lesistin is a hormone that is primarily secreted by adipose tissue (fat cells), and it has been implicated in insulin resistance and inflammation. I

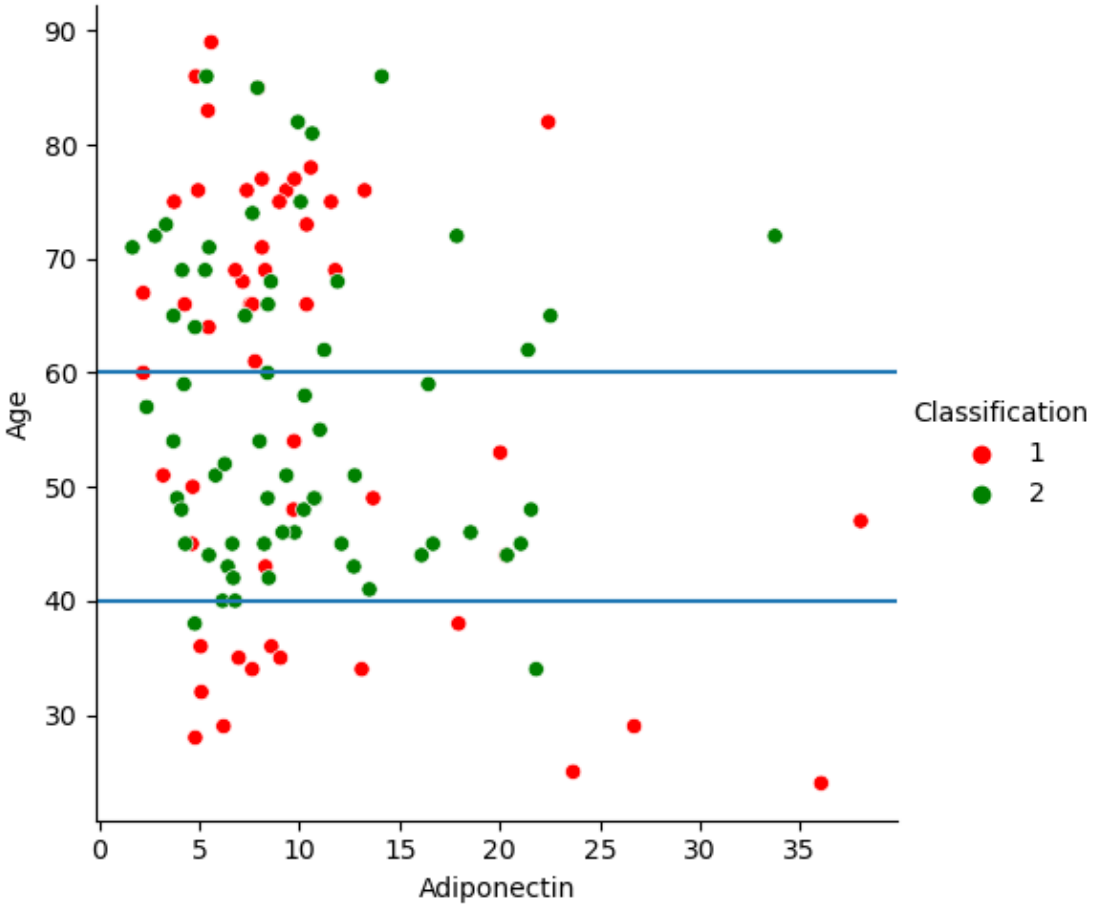
[59]:
```
sns.relplot(data = df,x ='MCP.1',y = 'Age',hue = 'Classification')
```

[59]:
```
<seaborn.axisgrid.FacetGrid at 0x1326c32add0>
```



[67]:
```
sns.relplot(data = df,x ='Adiponectin',y = 'Age',hue =␣
 ↪'Classification',palette=['r','g'])
plt.axhline(y=60)
plt.axhline(y=40)
# Observation between age 40-60 lower Adiponectin might be a concern
```

[67]:
```
<matplotlib.lines.Line2D at 0x1326ec8b8b0>
```


90
80
70
60
50
40
30
Age
0
5
10
15
20
25
30
35
Adiponectin
Classification
1
2

[ ]: