

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.metrics import confusion_matrix, classification_report
```

In [3]:

```
pdata= pd.read_csv('C:\\Users\\Jai mata di\\Downloads\\diabetes.csv')
pdata.head()
```

Out[3]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.62
1	1	85	66	29	0	26.6	0.35
2	8	183	64	0	0	23.3	0.67
3	1	89	66	23	94	28.1	0.16
4	0	137	40	35	168	43.1	2.28

In [4]:

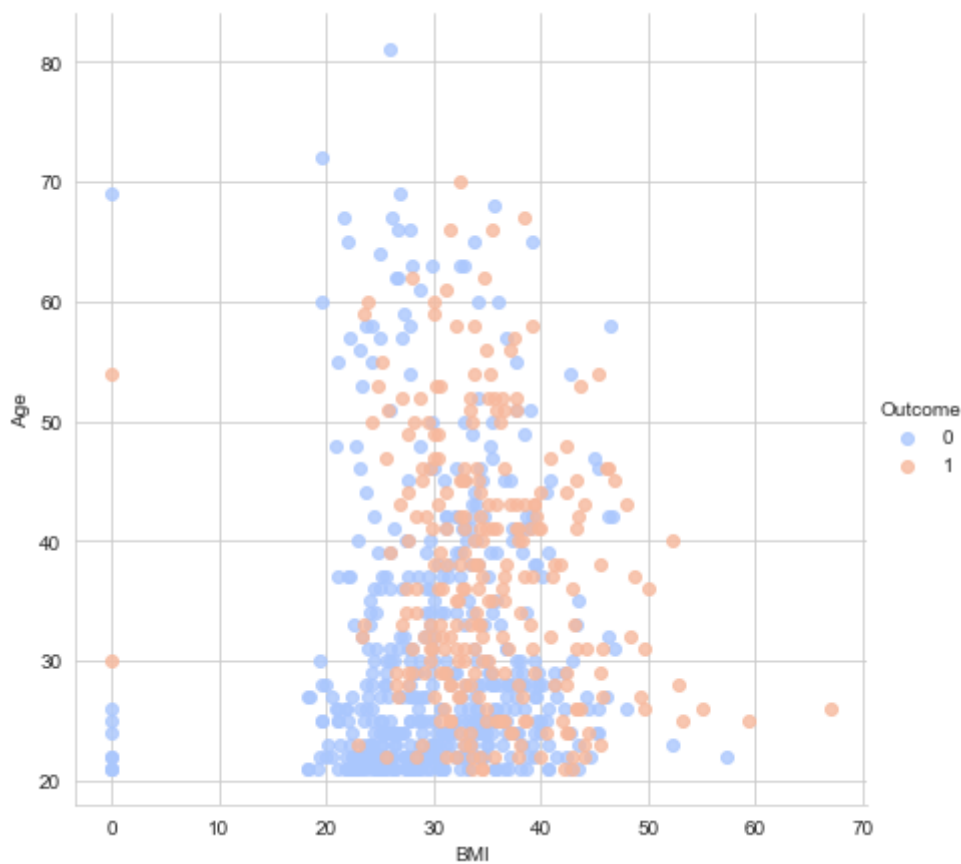
```
sns.set_style('whitegrid')
sns.lmplot('BMI', 'Age', data=pdata, hue='Outcome',
          palette='coolwarm', height=6, aspect=1, fit_reg=False)
```

C:\Users\Jai mata di\anaconda3\lib\site-packages\seaborn\\_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```

Out[4]:

<seaborn.axisgrid.FacetGrid at 0x1fb1e3fb970>



In [5]:

```
kmeans=KMeans(n_clusters=2)
kmeans.fit(pdata)
```

Out[5]:

KMeans(n\_clusters=2)

In [6]:

```
kmeans.cluster_centers_
```

Out[6]:

```
array([[ 3.88391376, 115.26699834, 68.09784411, 17.6185738 ,
        32.21227197, 31.17363184,  0.43757048, 33.11442786,
         0.30182421],
       [ 3.7030303 , 141.46060606, 72.78787879, 31.2       ,
        253.70909091, 34.98545455,  0.59724848, 33.7030303 ,
         0.52121212]])
```

In [7]:

```
def converter(cluster):
    if cluster=='Yes':
        return 1
    else:
        return 0
pdata['Cluster'] = pdata['Outcome'].apply(converter)
pdata.head()
```

Out[7]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0.62
1	1	85	66	29	0	26.6	0.35
2	8	183	64	0	0	23.3	0.67
3	1	89	66	23	94	28.1	0.16
4	0	137	40	35	168	43.1	2.28



In [8]:

```
print("Confusion Matrix: \n" ,confusion_matrix(pdata['Cluster'],kmeans.labels_))
print(classification_report(pdata['Cluster'],kmeans.labels_))
```

Confusion Matrix:

```
[[603 165]
```

```
[ 0  0]]
```

	precision	recall	f1-score	support
0	1.00	0.79	0.88	768
1	0.00	0.00	0.00	0
accuracy			0.79	768
macro avg	0.50	0.39	0.44	768
weighted avg	1.00	0.79	0.88	768

C:\Users\Jai mata di\anaconda3\lib\site-packages\sklearn\metrics\\_classification.py:1221: UndefinedMetricWarning: Recall and F-score are ill-defined and being set to 0.0 in labels with no true samples. Use `zero\_division` parameter to control this behavior.

```
_warn_prf(average, modifier, msg_start, len(result))
```

In [ ]: