In [27]:

```python
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

In [17]:

```python
data=pd.read_csv('dia.csv')
```

In [18]:

```python
data.head()
```

Out[18]:

	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 [19]:

```python
x=data.iloc[:,7]
x.head()
```

Out[19]:

```
0    50
1    31
2    32
3    21
4    33
Name: Age, dtype: int64
```

In [20]:

```python
y=data.iloc[:,6]
y.head()
```

Out[20]:

```
0    0.627
1    0.351
2    0.672
3    0.167
4    2.288
Name: DiabetesPedigreeFunction, dtype: float64
```

In [21]:

```python
x=np.array(x).reshape(-1,1)
```

In [22]:

```python
x
```

```
       [42],
       [45],
       [38],
       [25],
       [22],
       [22],
       [22],
       [34],
       [22],
       [24],
       [22],
       [53],
       [28],
       [21],
       [42],
       [21],
       [42],
       [48],
       [26],
       [22],
```

In [23]:

```python
y=np.array(y).reshape(-1,1)
```

In [24]:

```python
y
```

Out[24]:

```
array([[0.627],
       [0.351],
       [0.672],
       [0.167],
       [2.288],
       [0.201],
       [0.248],
       [0.134],
       [0.158],
       [0.232],
       [0.191],
       [0.537],
       [1.441],
       [0.398],
       [0.587],
       [0.484],
       [0.551],
       [0.254],
```

In [25]:

```python
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.20)
```

In [28]:

```python
classifier=LinearRegression()
```

In [29]:

```python
classifier.fit(x_train,y_train)
```

Out[29]:

```
LinearRegression()
```

In [30]:

```python
y_pred=classifier.predict(x_test)
```

In [31]:

```
y_pred
```

Out[31]:

```
array([[0.45921739],
       [0.47215969],
       [0.46784559],
       [0.48606067],
       [0.46065542],
       [0.45730001],
       [0.45730001],
       [0.4582587 ],
       [0.45777936],
       [0.46065542],
       [0.46544887],
       [0.47215969],
       [0.471201  ],
       [0.45730001],
       [0.46017608],
       [0.46640756],
       [0.46209346],
       [0.45777936],
```

In [34]:

```
r2_score(y_test,y_pred)
```

Out[34]:

```
-0.013647232579697022
```

In [36]:

```
mean_squared_error(y_test,y_pred)
```

Out[36]:

```
0.12172988403486278
```
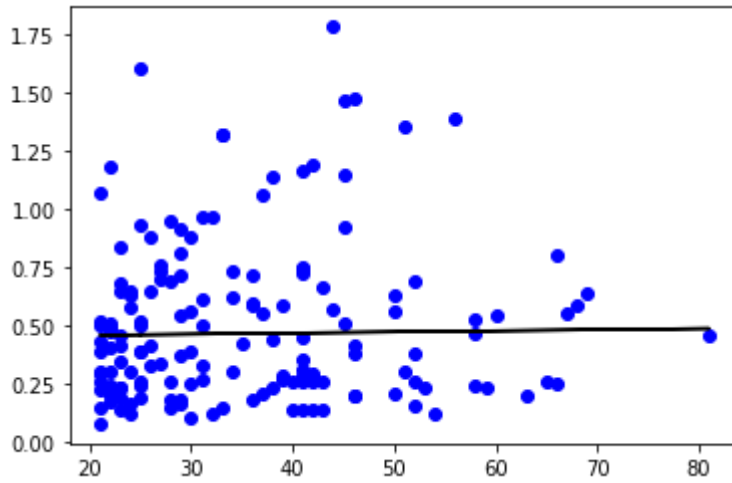
In [38]:

```
classifier.coef_
```

Out[38]:

```
array([[0.00047934]])
```

In [39]:

```python
plt.scatter(x_test,y_test,color='b')
plt.plot(x_test,y_pred,color='k')
```

Out[39]:

```
[<matplotlib.lines.Line2D at 0x180f950a560>]
```


In [ ]: