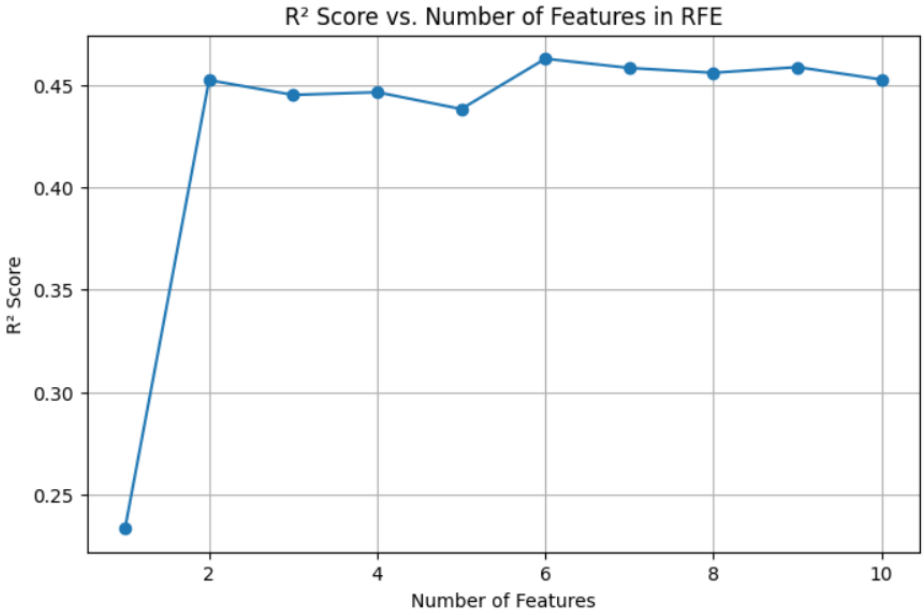
[19]
```
Dataset Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 442 entries, 0 to 441
Data columns (total 11 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   age     442 non-null    float64
 1   sex     442 non-null    float64
 2   bmi     442 non-null    float64
 3   bp      442 non-null    float64
 4   s1      442 non-null    float64
 5   s2      442 non-null    float64
 6   s3      442 non-null    float64
 7   s4      442 non-null    float64
 8   s5      442 non-null    float64
 9   s6      442 non-null    float64
 10  target  442 non-null    float64
dtypes: float64(11)
memory usage: 38.1 KB
None

First 5 rows of the dataset:
        age       sex       bmi        bp        s1        s2        s3  \
0  0.038076  0.050680  0.061696  0.021872 -0.044223 -0.034821 -0.043401
1 -0.001882 -0.044642 -0.051474 -0.026328 -0.008449 -0.019163  0.074412
2  0.085299  0.050680  0.044451 -0.005670 -0.045599 -0.034194 -0.032356
3 -0.089063 -0.044642 -0.011595 -0.036656  0.012191  0.024991 -0.036038
4  0.005383 -0.044642 -0.036385  0.021872  0.003935  0.015596  0.008142

         s4        s5        s6  target
0 -0.002592  0.019907 -0.017646   151.0
1 -0.039493 -0.068332 -0.092204    75.0
2 -0.002592  0.002861 -0.025930   141.0
3  0.034309  0.022688 -0.009362   206.0
4 -0.002592 -0.031988 -0.046641   135.0
```

```
.. _diabetes_dataset:

Diabetes dataset
----------------

Ten baseline variables, age, sex, body mass index, average blood
pressure, and six blood serum measurements were obtained for each of n =
442 diabetes patients, as well as the response of interest, a
quantitative measure of disease progression one year after baseline.

**Data Set Characteristics:**

:Number of Instances: 442

:Number of Attributes: First 10 columns are numeric predictive values

:Target: Column 11 is a quantitative measure of disease progression one year after baseline

:Attribute Information:
    - age     age in years
    - sex
    - bmi     body mass index
    - bp      average blood pressure
    - s1      tc, total serum cholesterol
    - s2      ldl, low-density lipoproteins
    - s3      hdl, high-density lipoproteins
    - s4      tch, total cholesterol / HDL
    - s5      ltg, possibly log of serum triglycerides level
    - s6      glu, blood sugar level

Note: Each of these 10 feature variables have been mean centered and scaled by the standard deviation times the square root of `n_samples` (i.e. the sum of squares of each column t
```


R² Score vs. Number of Features in RFE
R² Score
0.45
0.40
0.35
0.30
0.25
2
4
6
8
10
Number of Features

TASK 4

```
     10 features  9 features  8 features  7 features  6 features  5 features  \
age    37.904021         NaN         NaN         NaN         NaN         NaN
bmi   542.428759  542.799508  550.744365  551.866448  557.314167  597.892739
bp    347.703844  354.211438  363.791753  362.356114  350.178667  306.647913
s1   -931.488846 -936.350589 -947.823133 -660.643160 -851.515734 -655.560612
s2    518.062277  528.796592  541.585796  343.348089  591.093315  409.622184
s3    163.419983  167.800414  172.250588         NaN         NaN         NaN
s4    275.317902  270.396514  277.741072  185.140764         NaN         NaN
s5    736.198859  744.447429  761.921177  664.774591  803.121285  728.643647
s6     48.670657   53.350483         NaN         NaN         NaN         NaN
sex  -241.964362 -236.649588 -233.754686 -235.364224 -215.267423         NaN

     4 features  3 features  2 features  1 features
age         NaN         NaN         NaN         NaN
bmi  691.460102  737.685594  732.109021  998.577689
bp          NaN         NaN         NaN         NaN
s1  -592.977874 -228.339889         NaN         NaN
s2   362.950323         NaN         NaN         NaN
s3          NaN         NaN         NaN         NaN
s4          NaN         NaN         NaN         NaN
s5   783.168538  680.224653  562.226535         NaN
s6          NaN         NaN         NaN         NaN
sex         NaN         NaN         NaN         NaN
```

```
[30] #feature ranking
     rfe_final = RFE(model, n_features_to_select=3)
```

[31]
```
ranking_df = ranking_df.sort_values(by="Initial Rank")

print("Initial Feature Ranking:")
print(ranking_df)

print("\nFinal Selected Features:", selected_features)
```

```
Initial Feature Ranking:
  Feature  Initial Rank
0     age             1
1     sex             1
2     bmi             1
3      bp             1
4      s1             1
5      s2             1
6      s3             1
7      s4             1
8      s5             1
9      s6             1

Final Selected Features: ['bmi', 's1', 's5']
```