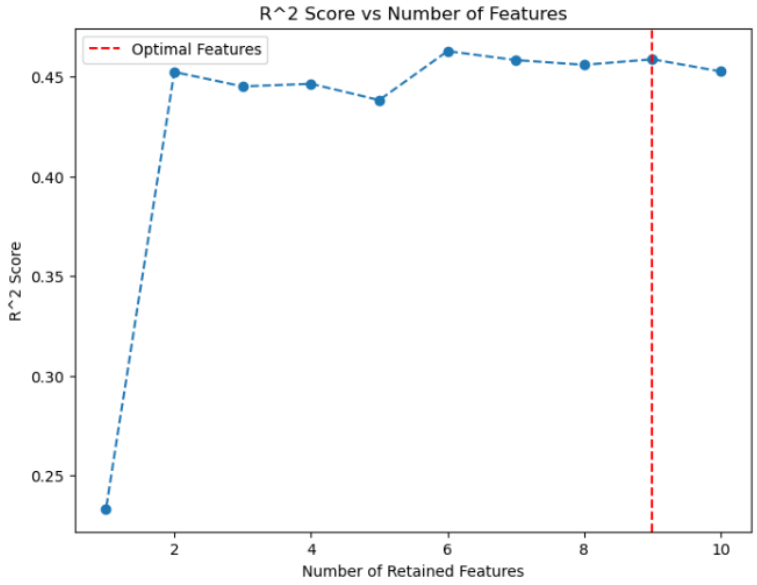
Objective:

This report presents an analysis of feature selection using Recursive Feature Elimination (RFE) with a linear regression model. The goal is to identify an optimal subset of features that maintains predictive performance while reducing complexity.

Dataset Overview:

- The dataset contains **10 features** and a target variable.
- The dataset was split into an **80-20** train-test ratio.
- The model's performance was evaluated using the **R^2 Score** on the test set.



Key Findings:

- The **R^2 Score** on the test set was **0.4526**, indicating moderate predictive power.
- Feature importance was assessed using Recursive Feature Elimination (RFE).
- The three most important features based on coefficient values are:
    1. **s1**: 931.49
    2. **s5**: 736.20
    3. **bmi**: 542.43

Feature Coefficients at Each Iteration:

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6
Iteration 1	37.904021	-241.964362	542.428759	347.703844	-931.488846	518.062277	163.419983	275.317902	736.198859	48.670657
Iteration 2	-236.649588	542.799508	354.211438	-936.350589	528.796592	167.800414	270.396514	744.447429	53.350483	NaN
Iteration 3	-233.754686	550.744365	363.791753	-947.823133	541.585796	172.250588	277.741072	761.921177	NaN	NaN
Iteration 4	-235.364224	551.866448	362.356114	-660.643160	343.348089	185.140764	664.774591	NaN	NaN	NaN
Iteration 5	-215.267423	557.314167	350.178667	-851.515734	591.093315	803.121285	NaN	NaN	NaN	NaN
Iteration 6	597.892739	306.647913	-655.560612	409.622184	728.643647	NaN	NaN	NaN	NaN	NaN
Iteration 7	691.460102	-592.977874	362.950323	783.168538	NaN	NaN	NaN	NaN	NaN	NaN
Iteration 8	737.685594	-228.339889	680.224653	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Iteration 9	732.109021	562.226535	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Iteration 10	998.577689	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Optimal Feature Selection:

- The optimal number of retained features was determined by analyzing R^2 score stabilization.
- The selected features that optimize performance while reducing dimensionality are:
    - **sex**
    - **bmi**
    - **bp**
    - **s1**
    - **s2**
- These features were found to contribute the most to the model's predictive capability.

Conclusion:

Feature selection using RFE successfully reduced the number of features while maintaining model performance. This process helps improve model interpretability and reduces computational costs without significantly impacting accuracy. The selected subset of features (sex, bmi, bp, s1, and s2) offers an optimal balance between performance and simplicity.