Introduction

This report presents the results of Recursive Feature Elimination (RFE) on the Diabetes dataset. RFE was applied to determine which features are most relevant for predicting diabetes progression.
We analyze feature rankings, their impact on the model, and how feature selection affects performance.

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.

Feature Coefficients at each RFE Iteration :

Feature Coefficients

	age	sex	bmi	...	s4	s5	s6
1	998.577689	NaN	NaN	...	NaN	NaN	NaN
2	732.109021	562.226535	NaN	...	NaN	NaN	NaN
3	737.685594	-228.339889	680.224653	...	NaN	NaN	NaN
4	691.460102	-592.977874	362.950323	...	NaN	NaN	NaN
5	597.892739	306.647913	-655.560612	...	NaN	NaN	NaN
6	-215.267423	557.314167	350.178667	...	NaN	NaN	NaN
7	-235.364224	551.866448	362.356114	...	NaN	NaN	NaN
8	-233.754686	550.744365	363.791753	...	761.921177	NaN	NaN
9	-236.649588	542.799508	354.211438	...	744.447429	53.350483	NaN
10	37.904021	-241.964362	542.428759	...	275.317902	736.198859	48.670657

Top 3 Most Important Features are s1 s5 and bmi

- s1 has a strong negative impact, meaning it may act as a protective factor.
- s5 and bmi both have strong positive effects, s1 and weight are critical factors in diabetes risk

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:

- **s1**: 931.49
- **s5**: 736.20

- **bmi**: 542.43

Conclusions :

RFE confirmed that all 10 features were needed for predicting diabetes progression.

- bmi, s5, and s1 were the strongest predictors, proving that body mass and blood glucose levels play a major role in diabetes severity.
- BP had moderate importance, while age and sex had lower impact but still contributed to the model.
- RFE ranked all features instead of eliminating them.

Since diabetes progression is influenced by **multiple factors working together**, keeping all features ensured the best model performance

This graph visualizes the **R^2 score** as a function of the number of retained features during Recursive Feature Elimination