1. Output of code:

Text

Description automatically generated

1. In my experience, it is much easier and quicker to use built-in functions in R compared to coding my own functions in C++. For functions like sum(), mean(), median(), and range() where most know how to calculate them, it is better to just use built-in functions. Built-in functions can also prevent human errors.
2. Mean: the average of a set of elements/data

Median: the middle of a (sorted) set of numbers

Range: the difference between the largest and smallest numbers in a dataset.

These measures are useful in gauging general information regarding the dataset and can be used in making simple hypotheses regarding the target. (For example, the mean/median can be initial good guesses for unknown target values of the same type.) They can be used to calculate other statistics which can give more information about the dataset regarding variance, correlation, and more.

1. Covariance: measures how changes in one variable are associated with changes in a second variable.

Correlation: covariance scaled to [-1,1]. This can be preferred as covariance can range wildly depending on the dataset. -1 means a perfect negative correlation, +1 means a perfect positive correlation, and 0 indicates little correlation.

These measures help us know how variables are correlated, which is essential information in machine learning. Using these, we can easily determine the predictors that our strongly correlated to our target and use them for prediction.