Making Quieter Cars

Neil DeBoer

Objective

Introduction

One of the most common sources of Noise Pollution are automobiles. I'm seeking to understand how we can engineer cars that are quieter.

Stakeholders

Car manufacturers, car buyers

The Dataset

- We are using the Kaggle dataset found here: https://www.kaggle.com/datasets/murtio/car-noise-specification
- This contains data on the noise levels of different cars alongside data on their physical characteristic.
- There is a lot of missing data in the characteristics of the car.
- In lieu of this we must restrict the number of features we examine in our analysis.

The Data

- I will be trying to see what impacts the change in intensity of noise at various speeds (as measured in decibels) from the noise made when the car is idling. The different speeds are:
  - 50 kilometers per hour (kmh)
  - 80 kilometers per hour (kmh)
  - 100 kilometers per hour (kmh)
  - 120 kilometers per hour (kmh)
  - 140 kilometers per hour (kmh)
- The features I will use to predict these changes are:
  - The noise when idling.
  - Car height.
  - Car width.
  - Car wheelbase.

  Limiting the analysis to these features will allow me to have enough data to preform a proper splitting of the data.

The Models

- I will be using 3 different regression models:
  - Linear Regression.
  - Decision Tree Regression.
  - Random Forest Regression.
- The Decision Tree and Random Forest Regression models both will allow me to look at which features have the most impact on the model

The Results

- Looking at the mean square error from a train test split with test size of 20% we got the following for our 50 kmh change and our 140 kmh change (Note that the maximum depth of the decision tree model was 4):
  - The mean squared error for 50 kmh with Linear Regression was 9.03
  - The mean squared error for 50 kmh with Decision Tree regression was 8.56
  - The mean squared error for 50 kmh with Random Forest regression was 7.4
  - The mean squared error for 140 kmh with Linear Regression was 6.76
  - The mean squared error for 140 kmh with Decision Tree regression was 6.15
  - The mean squared error for 140 kmh with Random Forest regression was 5.56

Graphs

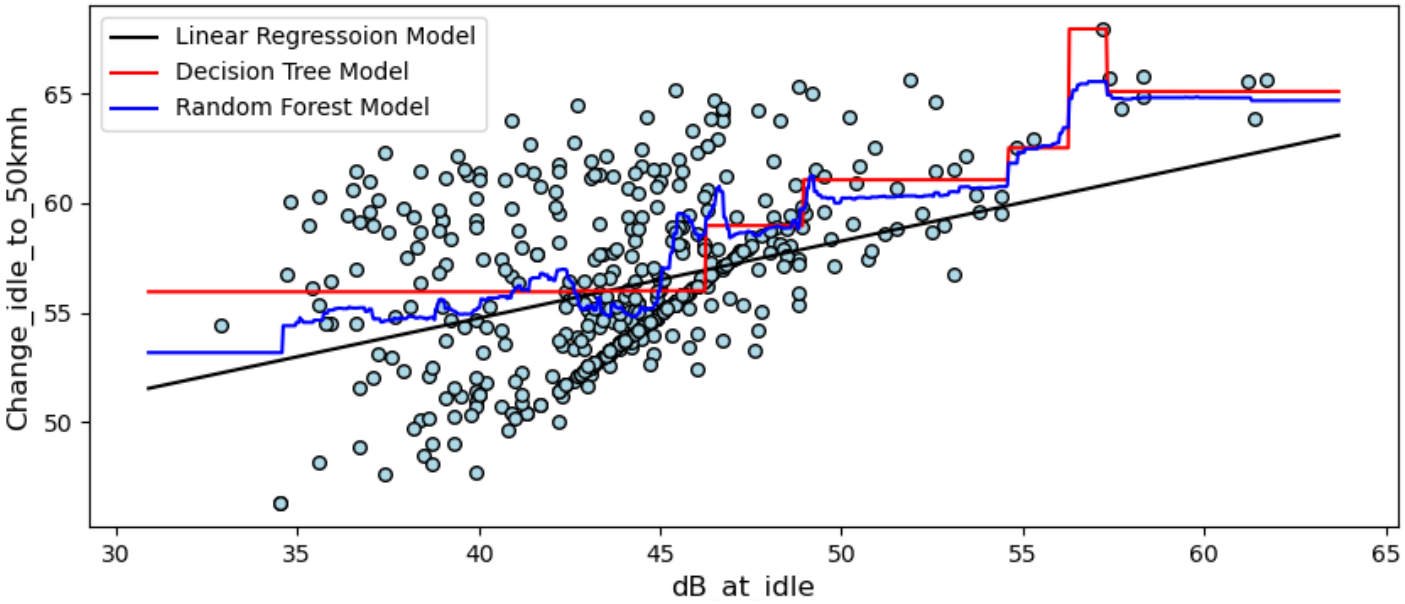

Figure: Idle Noise (decibels) vs Noise change at 50 kmh (decibels)

Graphs

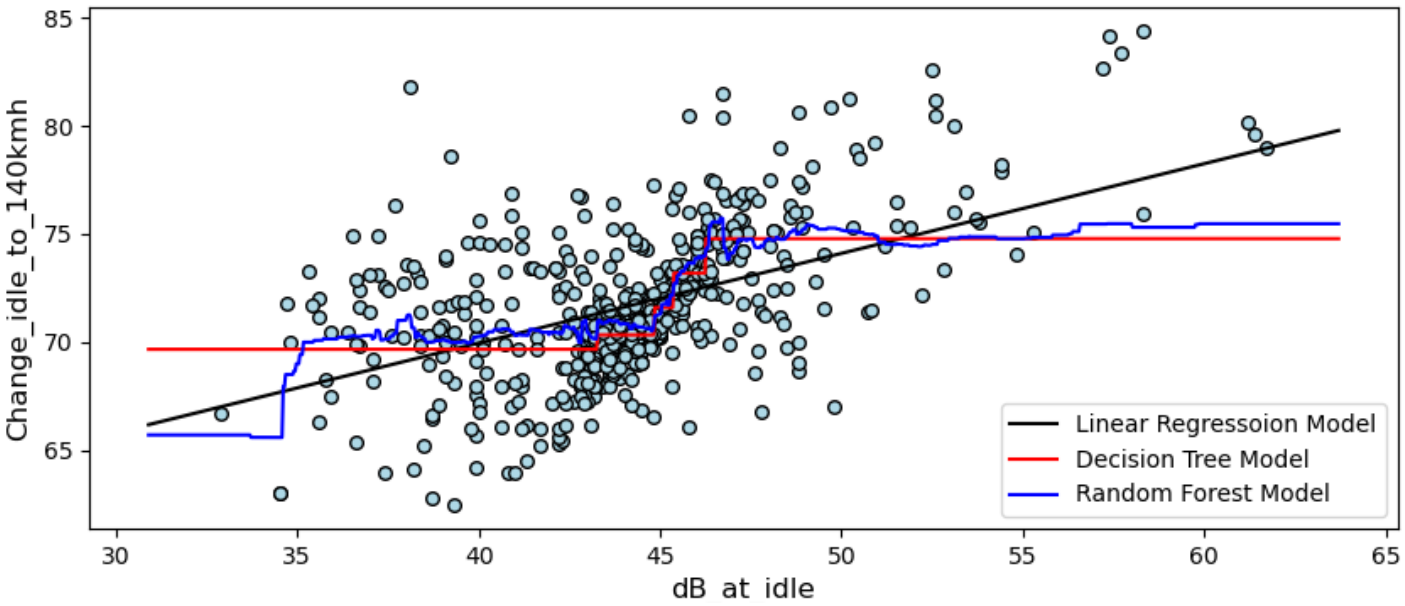

Figure: Idle Noise (decibels) vs Noise change at 140 kmh (decibels)

Graphs

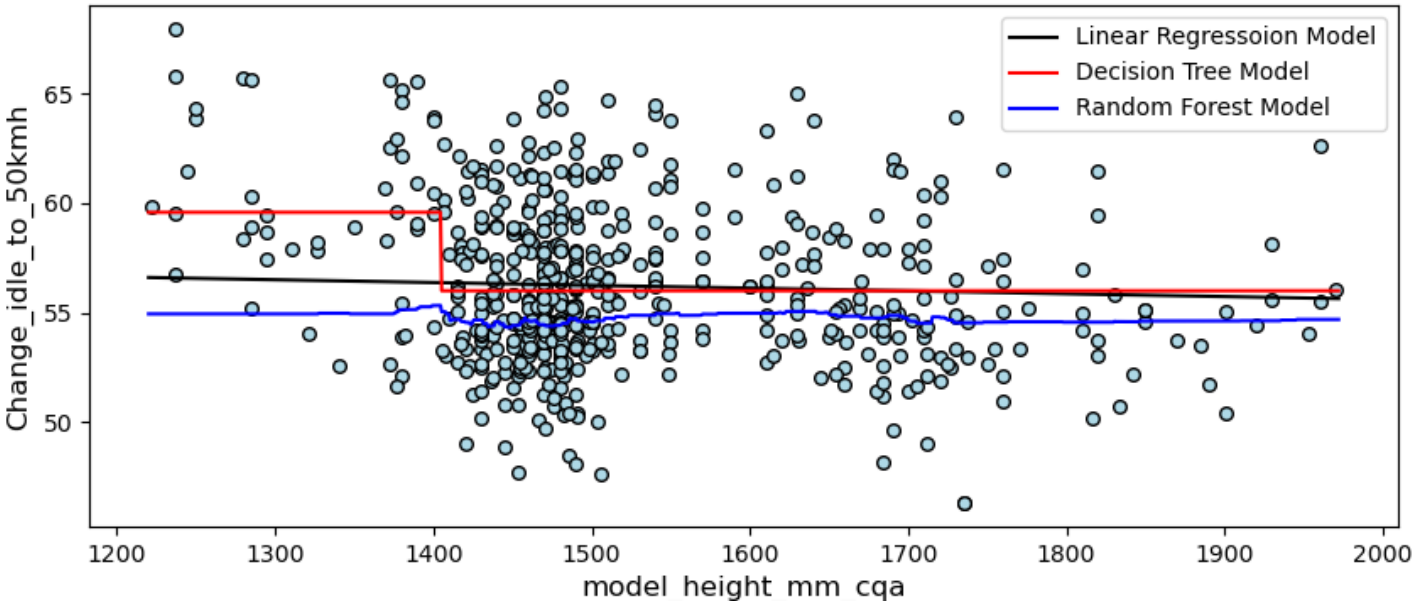

Figure: Car height (mm) vs Noise change at 50 kmh (decibels)

Graphs

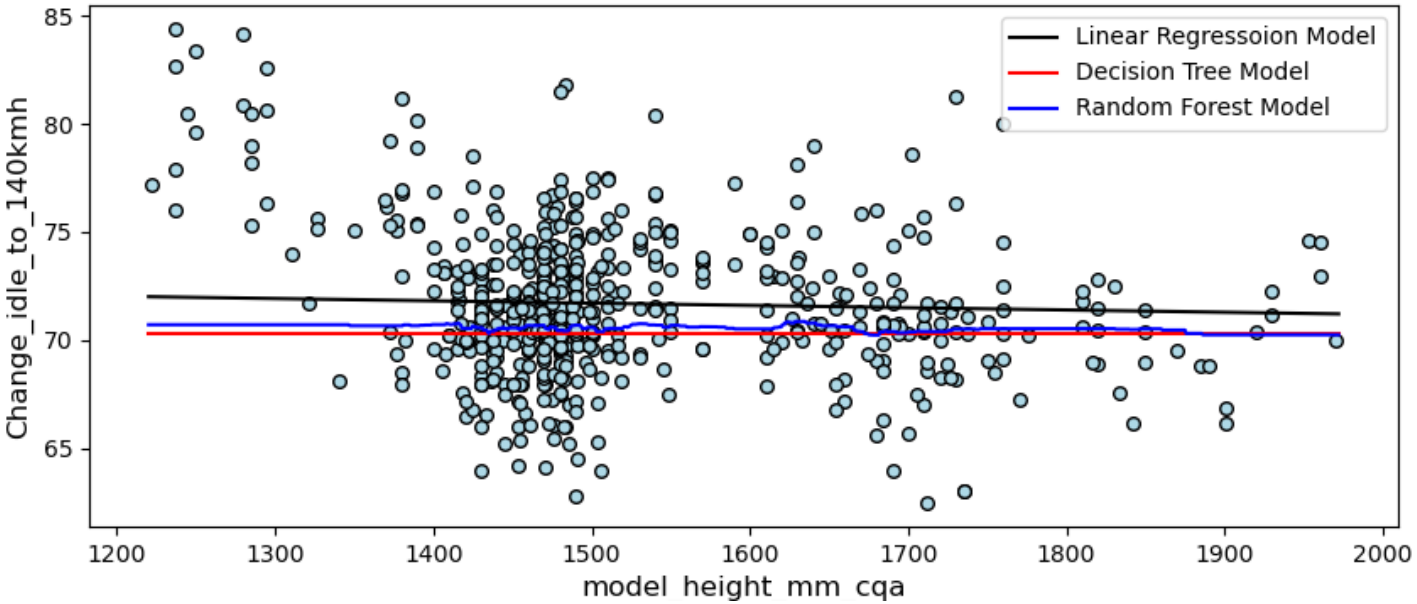

Figure: Car height (mm) vs Noise change at 140 kmh (decibels)

Graphs

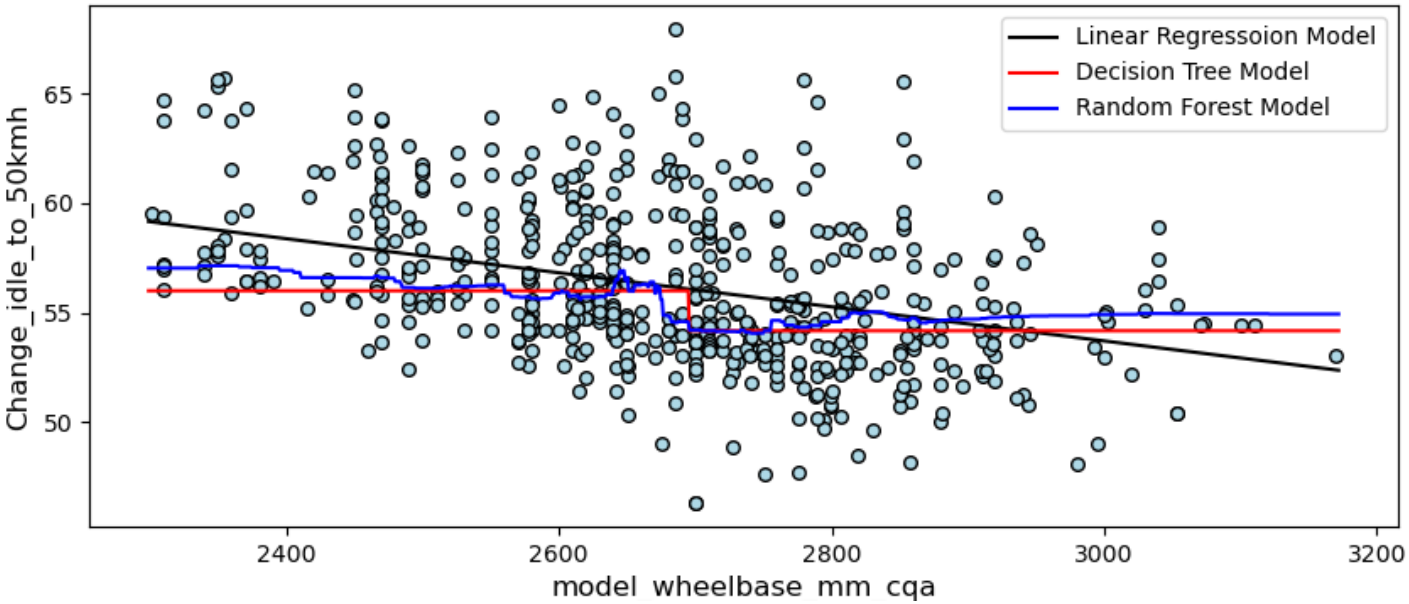

Figure: Car wheel base (mm) vs Noise change at 50 kmh (decibels)

Graphs

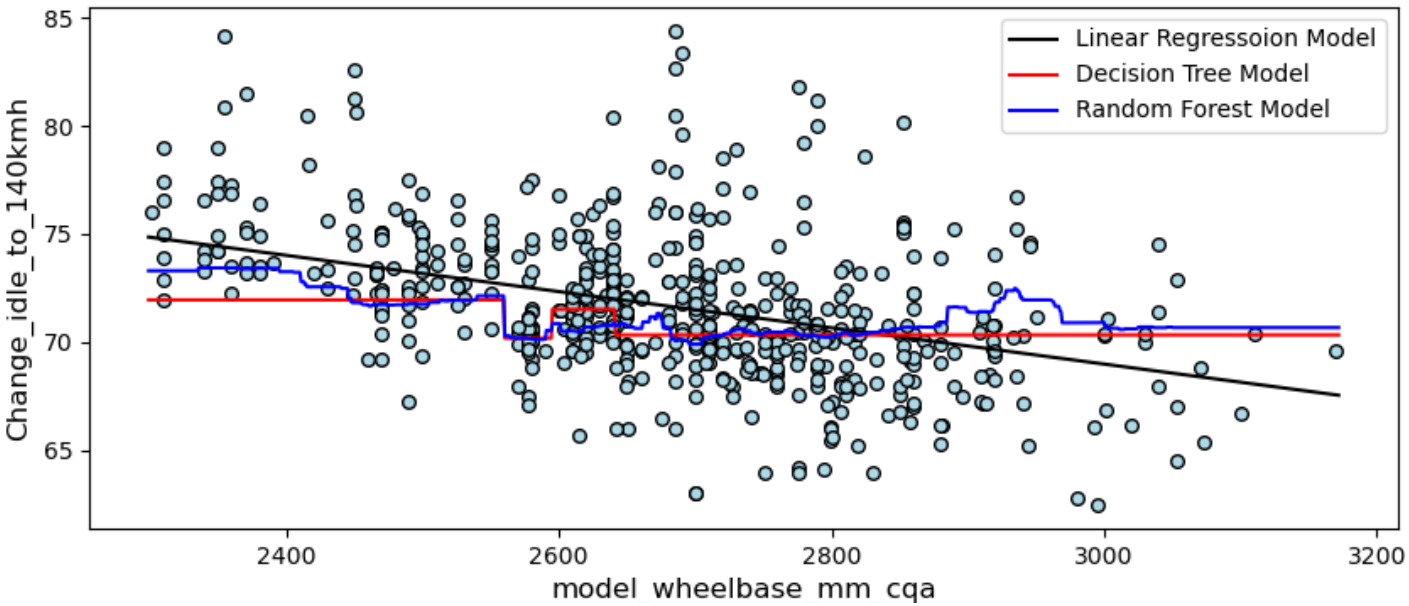

Figure: Car wheel base (mm) vs Noise change at 140 kmh (decibels)

Graphs

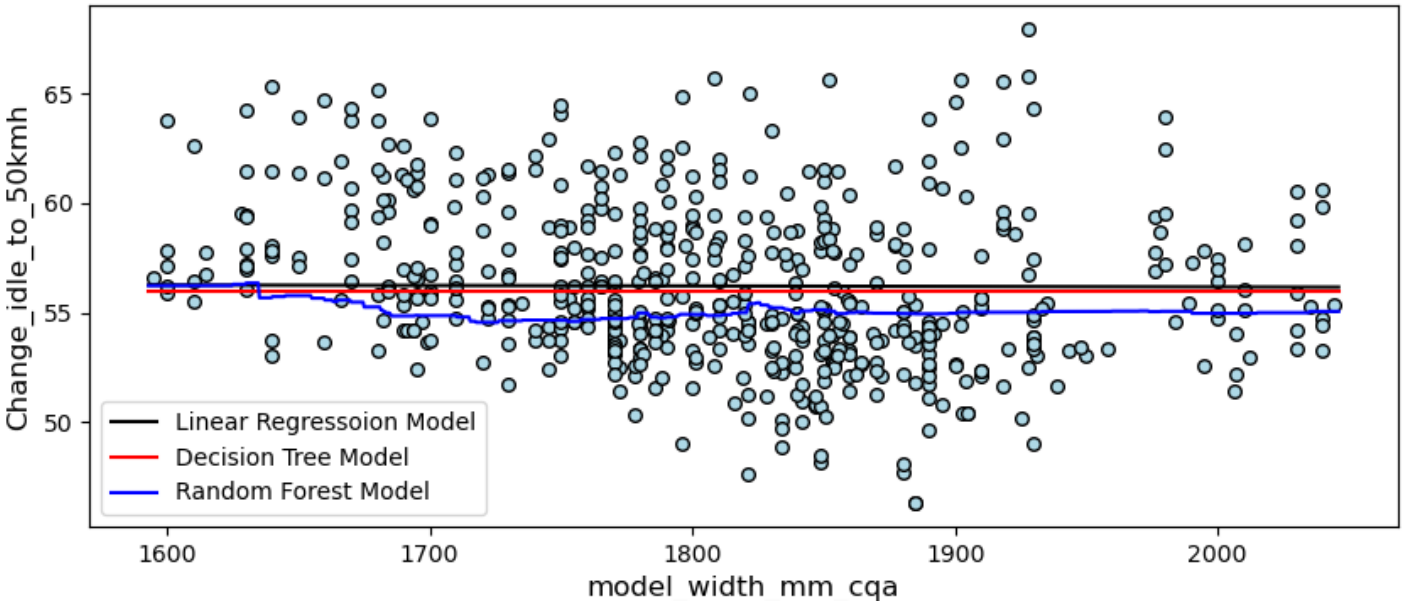

Figure: Car width (mm) vs Noise change at 50 kmh (decibels)

Graphs

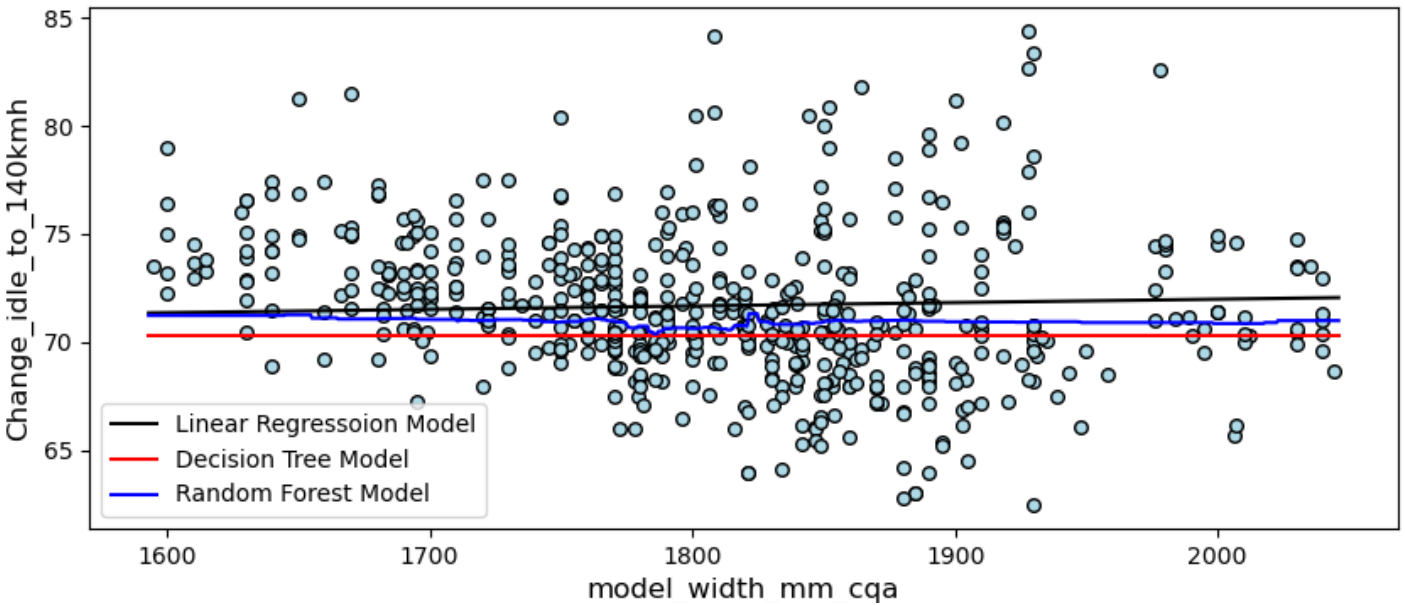

Figure: Car width (mm) vs Noise change at 140 kmh (decibels)

Feature Importance

According to the Random Forest model, the idling noise was the most important feature for determining the change in noise followed by the model wheel base. This held regardless of the speed we were looking at.