Introduction

The Ford F-150 is the most popular model in the Ford F-Series, and Texas is one of the states with the largest Ford F150 sales. This project will focus on the auto insurance premiums of the Ford F150 model in Texas, and explore what factors have a significant impact on insurance premiums given the fixed model and geographic location.

Dataset and Proposed Methods

By selecting four key independent variables:Driver Age, Accident Counts,Vehicle Year. I propose a multiple linear regression model to predict insurance premiums and analyze the most important factors affecting insurance premiums.

$$y_{ij} = \mu + \tau_j + \beta_{1j} x_{ij} + \epsilon_{ij}$$

$$\epsilon_{ij} \ (0, \sigma^2)$$

where y_{ij} is the predicted premium (Premium),

μ is overall mean, represents the average insurance cost when no other variables are considered.

τ_j represents the impact of the j age group (under 21 or over 21) on insurance costs.

affecting insurance premiums.

$\beta_{1j} x_{ij}$ Linear effect, represents the linear impact of independent variables (Accident Counts,Vehicle Year) on insurance costs.

Expected Results

Older drivers (typically 21+ or even older groups, such as 35+ or 50+) generally enjoy lower premiums. Younger drivers, especially those under 21, tend to have higher insurance premiums.

Drivers with a history of accidents, particularly when they were at fault, will experience significant increases in insurance premiums. The more accidents, the higher the premium increase.

Newer cars tend to have higher premiums, likely due to their higher replacement costs and advanced technology that might be more expensive to repair.