Problem statement

What type of vehicles shall Geely Auto select to produce and grow their revenue by 15% annually in the US for the first five years?

Context

A Chinese automobile company Geely Auto aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts. They have contracted an automobile consulting company to understand the factors on which the pricing of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market.

Criteria for success

Geely Auto grows revenue by 15% annually for the first five years once they enter the US market.

Scope of solution space

Apply data science method on provided file to clean and wrangle the data, explore the independence among predictors, the correlation between predictors and price, build and compare models with different parameters, and finalize a linear regression model to predict car price via selected predictors.

Constraints

Data is limited and unbalanced. There can be missing or corrupted data.

Stakeholders

Director of Marketing

Team of data scientists

Director of product R&D

Director of manufacturing

Data sources

CarPrice_Assignment.csv has data of 26 columns

Data Dictionary - carprices.xlsx shows the meaning and data type for each of the columns.