

A Case Study

In this case study we are going to use the automobile dataset, which plenty of car manufacturers with their specifications in in order to build a predictive model to find out the approximate car price. This dataset has 26 columns, including categorical and quantitative attributes.

In Table 1 some of the records from the above-mentioned dataset.

Table 1: Samples of the dataset records

| symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels |
|-----------|-------------------|-------------|-----------|------------|--------------|-------------|--------------|
| 3 | ? | alfa-romero | gas | std | two | convertible | rwd |
| 3 | ? | alfa-romero | gas | std | two | convertible | rwd |
| 1 | ? | alfa-romero | gas | std | two | hatchback | rwd |
| 2 | 164 | audi | gas | std | four | sedan | fwd |
| 2 | 164 | audi | gas | std | four | sedan | 4wd |
| 2 | ? | audi | gas | std | two | sedan | fwd |
| 1 | 158 | audi | gas | std | four | sedan | fwd |
| 1 | ? | audi | gas | std | four | wagon | fwd |
| 1 | 158 | audi | gas | turbo | four | sedan | fwd |
| 0 | ? | audi | gas | turbo | two | hatchback | 4wd |
| 2 | 192 | bmw | gas | std | two | sedan | rwd |
| 0 | 192 | bmw | gas | std | four | sedan | rwd |
| 0 | 188 | bmw | gas | std | two | sedan | rwd |
| 0 | 188 | bmw | gas | std | four | sedan | rwd |
| 1 | ? | bmw | gas | std | four | sedan | rwd |
| 0 | ? | bmw | gas | std | four | sedan | rwd |
| 0 | ? | bmw | gas | std | two | sedan | rwd |
| 0 | ? | bmw | gas | std | four | sedan | rwd |
| 2 | 121 | chevrolet | gas | std | two | hatchback | fwd |

You need to write descriptive answers to the questions under each task and also use a proper program written in Python and execute the code.

1. The missing values are presented as '?' in the dataset. Apply data wrangling techniques using Python programming language to solve missing values in all the attributes.
2. Check the data types of those columns with the missing values, and convert the data type if needed.
3. Find all the correlated features to the 'Price'.
4. Build a predictive model to predict the car price based on using one of the independent correlated variables.
5. Continue with the same built model in No.4, but choose different independent variables and discuss the result.