Business analytics
Analysis of car advertisement data

Alina Ivanova, Moritz Hangen, Simon Gosch

2024/03/24

Contents

List of Figures

List of Tables

1 Introduction

[1]

2 Theoretical background

3 Methodology

4 Findings and discussion

4.1 Model usage

Given the regression model, it can be applied to example data to examine its behavior.

4.1.1 Assessing expectations

Intuitive assumptions draw you to the conclusion, that the vehicle's mileage has an inverse relationship to the predicted advertisement price. To evaluate if the model also follows this behavior, it was applied to manually created data differentiating only by mileage.

Registration year	**Mileage (mi)**	Horse-power	Width (mm)	Length (mm)	Average mpg	Top speed (mph)	**Predicted price (£)**
2017	**60000**	135	2027	4284	49	116	**16157**
2017	**130000**	135	2027	4284	49	116	**13828**

Table 4.1: Influence of mileage on recently registered cars

As it is evident in Table 4.1, a higher mileage in fact reduces the predicted price. However, for an over 2-fold increase in miles, the depreciation is with 14.4 % not as high as initially expected.

While comparing two otherwise identical cars, it should be noted that the other values still contribute significantly to the result.
In particular the registration year, which, as shown before, strongly correlates with the target variable, has an effect on the limited influence of the mileage here. Given the data set's sampling of data up to 2017, both of the Volkswagen (VW) Golfs in Table 4.1 have been first registered very recently, so the base price is measurably higher. If you apply the same example to cars registered in 2010, which therefore have been running for seven years, the influence of the mileage grows.

Registration year	**Mileage (mi)**	Horse-power	Width (mm)	Length (mm)	Average mpg	Top speed (mph)	**Predicted price (£)**
2010	**60000**	135	2027	4284	49	116	**11122**
2010	**130000**	135	2027	4284	49	116	**8793**

Table 4.2: Influence of mileage on older cars

As shown in Table 4.2, the gap between the two otherwise identical vehicles has widened to 20.9 %, an increase by 45.3 %. Potential reasons for the strong correlation between year and target value will be investigated further in section 4.2.2. Nevertheless, for that small subset of data, the efficacy of the model is evident.

4.1.2 Usage of example data

However, this small example is not cohesive enough to demonstrate the ability to solve the aforementioned business problem. To apply the model to a day-to-day use case as it regularly appears in a dealership, it was used to predict the advertisement price of four automobiles as they could be in its yard.

Registration year	Mileage (mi)	Horse-power	Width (mm)	Length (mm)	Average mpg	Top speed (mph)	**Predicted price (£)**
2014	180000	110	1799	4204	45	110	**5601**
2016	150000	120	2027	4255	48	112	**12130**
2018	80000	130	2027	4255	50	115	**16136**
2015	190000	115	1799	4204	44	108	**5819**

Table 4.3: Assessing model for business use cases

Table 4.3 illustrates that the model is able to predict an appropriate advertisement price for a VW Golf. As the model's performance with $R^2 = 0.927$ is very good, the dealer can rely on it, receiving a data driven estimate for a competitive price, thus not having to rely solely on human estimate.

4.2 Findings

In section 4.1 the application of the model to exemplary vehicles has been demonstrated. Here, some trends that could have already been anticipated after the correlation analysis, have emerged more clearly.
These trends that have been found regarding the model and the underlying data can be divided into two groups, the predicted and the unexpected.

4.2.1 In line with presumptions

In this first part, the findings which can be considered logical or self-evident will be discussed.

Mileage ↔ Price

Horsepower ↔ Price

4.2.2 Outliers

Engine size ↔ Price

Width ↔ Price

Year ↔ Price

TODO include example of higher influence of mileage the older the car is

5 Conclusion

5.1 Limitations of the analysis

5.1.1 Recency of the data

5.1.2 Lack of sales data regarding advertisement

Firstly, while the given dataset includes conclusive data regarding the advertisement price of cars in the used vehicle market, there is no indicator given, whether the car was actually sold at the price that it has been advertised for.
Nevertheless, while one may anticipate a few dealerships to over- / undershoot their prices, considering the scale of the dataset, that effect evens out for the overall market. However, evaluating whether there is a trend that pre-owned cars are systematically under- / overpriced in commercials is only possible if you compare the given data set to real sales information.

5.2 Further research topics

Given the scale of the available data, more potential research questions may be examined.

5.2.1 Inter-model comparison of findings

Additionally, the analysis is currently only valid for the small subset of the data including VW Golf. To levitate generalizability to the whole used vehicle market and to assess potential disparities as well as similarities in between car models, expanding the scope to the entirety of available data is obligatory.

5.2.2 Assess value depreciation

Having inter-model differences evaluated, a possible further research topic is the comparison of each car's new price to its future advertisement prices. For each model in the advertisement dataset, there is a corresponding data point in the basic information table that contains, among others, the manufacturer's suggested retail price (MSRP).
Given that information, the following questions could be analyzed:

- Which model retains the most value compared to its MSRP?
- Given five years of use, which car's price decreased the most?
- Is there a correlation between value depreciation and the manufacturer of the car?

This information can be useful for customers considering the purchase of a new car in order to assess its potential resale value in the future.

5.3 Résumé

The conducted analysis provides a data driven solution to the business problem of a competitive advertisement price of a given VW Golf. After preparing the raw data, key influences have been separated using Pearson-correlation and used to train a linear regression model to predict a suitable advertisement price.

Considering the parameters

- Registration year
- Mileage (mi)
- Horsepower
- Width (mm)
- Length (mm)
- Average mpg
- Top speed (mph)

the model accurately predicts a suitable advertisement price with $R^2 = 0.927$, solving the business problem therewith. The end results align with logical assumptions, for instance that newer cars sell for higher prices. Nonetheless, they provide further insight by revealing each factor's contribution to the total amount.

Provided that outcome, further potential research areas are examined and left open for future analysis.

Bibliography

[1] *Industrie-Roboter - Hirata Engineering Europe GmbH.* URL: https://www.hirata.de/de/produkte/scara-roboter (visited on 02/02/2024).