Evaluation of Vehicles data Set

Eric Riutort

Business Understanding¶

1.1 Determine business objectives:¶

1.1.0 Business Objectives and Background¶

The group of used car dealerships is interested in understanding what attracts customers to used cars, and what their preferences are for used cars.

There are many used car dealerships in the United States, and they have a vast number of vehicle inventory from which to choose.

The dealership group in question would like to determine what type of used cars to acquire and then resell to their customers. In otherwords, what types of used cars are in the highest demand for their customer base, so they dealership can focus on specific target markets and determine what inventory to purchase and sell, at the highest margin, but also at the largest volume.

The dealership group plans to fine-tune their inventory based on the analysis recommendations.

1.1.1 Business success criteria¶

The business success is measured by the total net profit of all the dealers together. This net profit is driven by the ability of each individual dealership to purchase vehicole inventory and then sell the inventory for the highest margin with the lowerst number of days of the inventory being on the lot.

For this reason, understanding what features of a vehicle drive demand is important, as the faster a vehicle is sold, and at a higher price the vehicle is sold, the assumed margin is higher.

It is noted, that this assumption is missing the cost aspect, as the end gross margin is a result of the final sales price less the cost of the good sold, sales commission, other promotions to clients, losses or profits from client vehicle trades ins, but may also include profits resulting from vehicle financing, service agreements, or other ancilliary services and products bundled into sales transaction.

1.2 Assess Situation¶

For the purposes of this analysis, the used car dealership group is assumed to be in a regional location, and focuses on newer automobiles in fair to excellent shape, which are available to purchase from regional vehicle auction venues. This group does not trade in older vintage, historical, or antique vehicles, for which NADA bluebook prices are not available, or in salvage automobiles.

For this reason, the focus of this analysis will be on newer vehicles fair or better condition.

1.2.0 Inventory of resources¶

Data Resources: The data resource available is the Kaggle derived dataset of ~426K cars.

Technical Resources: Anaconda, Jupyter notebooks, Google Colab, Python.

Manpower Resources: One invidual

1.2.1 Requirements, assumptions, constraints¶

Requirements:

- Identify possible features which are attractive to customers
- Validate features which are attrative to customers

Assumptions

- Determinations and recommendations will be made on complete data
- The scope of the analysis will fit the region the dealership is in
- The group's target market is for newer vehicles, in fair or better condition
- Older, vintage vehicles AKA any vehicle which is not currently valued by the NADA bluebook for the year of the last vehicle year listed in the data set, will not be in scope for the analysis
- Salvage vehicles, or other vehicles whose condition is not determinable, will not be in the analysis
- Vehicles whose brand or make is not within the NADA bluebook will not be within the analysis
- The group is interested in vehicles within a certain region specific to their business sales territory

Constraints

- As an exercise, it is not possible to verify business objectives or requirements
- The business objectives, requirements, assumptions and constraints can only be assumed
- Only one data set is available for analysis

1.2.2 Risk and Contingencies¶

The primary risk inherent in this analysis is the data set itself, which may lack some primary data important to the conduct of an effective analyis to meet the business objectives.

Additionally, terminology and Costs and Benefits are beyond the scope of this assessment.

.3 Determine the data mining goals¶

1.3.0 Data Mining Goals¶

Identify the features of a vehicle, which drive price and thus demand, for vehicles. This may include derived features, such as the vehicle's sales location.

1.3.1 Data Mining Success Criteria¶

One or more features are identified and validated to be correlated to the price AKA demand for a vehicle.

1.4 Project Plan¶

1.4.0¶

The data set will be accesssed, understood, prepared, modeled and evaluated in order to meet the data mining success criteria, and, thusly, the business objectives

1.4.1 Project plan steps¶

1. Import data
2. Understand data
3. Create Visualizations
4. Prepare Data
5. Create Visualizations
6. Modeling with Linear Regression or other techniques
7. Modeling - use K-fold or Hold-Out Cross-Validation, use Grid-search
8. Evaluate the models using coefficients and error metrics. For error metrics evaluate their coefficients.
9. Identify the influence of the factors on the price of the car
10. Use SFS, Permutation , Transform Target Regression
11. Define next steps and recommendations

For Sections 2 - 4, see the Notebook.

5.0 Deployment¶

The results of this analysis of the Kaggle data set,show the group of car dealerships, in order to trim their inventory, can look at this analysis and implement a set of recommendations about how to more accuratley predict automobile prices for their target market.

Clean vehicle data is hard to come by, and if analysis is conducted and then business decisions made on dirty or incomplete data, that analysis will be inaccurate or misleading and the business decisions could be potentially less than ideal or even lead to negative outcomes, such as overpaying for a vehicle that a dealer intends to resell.

For this reason, used car dealers should focus on their specific target market and make purchase acquisition decisions for vehicles where the complete history of the vehicle and information of the vehicle is available. Purchasing vehicles where this information is not available means the vehicle cannot be accurately priced.

In order to do so, they should not acquire inventory where the automobile's characteristcs are not well defined (which actually happens), such as inventory where title is missing, VIN number is missing, odometer reading is unavailable, etc.

Additionally, they should trim or not acquire inventory which does not match their target market, where the models or manufacturers are infrequently traded (sold), as these types are difficult to price_violin_plot.png

Also, they should trim or not acquire inventory that is in a salvage, or parts only condition, or where the title is for a lien, salvage or rebuilt vehicle.

And, they should trim or not acquire inventory where the type is unusually, or unique, such as buses, motor-cycles, certain fuel types (electric), vehicles with unusual cylinder counts.

Finally, they should not acquire very low or high mileage vehicles or vehicles that are vintage or antique (which for our purposes are vehicles made before 1992).

By limiting or trimming their inventory in accordance with the above guidelines, dealers can then leverage the above model to accurately predict (R-squared: 0.783 and MAE: 3442) what the selling price of a vehicle will be. They can do this by leveraging this model and inputting the features of the vehicle, and then from this predicted sales price, determine a offer price for a vehicle acquisition, which would be based on the desired profit margin each dealership would expect to make on its sales.