2025年8月10日

18:01

List of Variables：

**PurchaseID:**

Purchase Identification Number

**PurchaseTimestamp:**

Purchase Timestamp

**PurchaseDate**:

Purchase Date

**Auction:**

Auction company

**VehYear:**

Year Vehicle is made

**Make:**

Vehicle’s make

**Color:**

Color of the car

**Transmission:**

Auto or manual Transmission

WheelTypeID:

Wheel type Identification Number

WheelType:

Wheel type

VehOdo:

Vehicle’s Odometer reading

Nationality:

Nationality of the vehicle

Size:

Size of the vehicle

TopThreeAmericanName:

If the vehicle is from one of the top three American manufacturers.

MMRAcquisitionAuctionAveragePrice:

Acquisition price for this vehicle in average condition at the time of purchase

MMRAcquisitionAuctionCleanPrice:

Acquisition price for this vehicle in the above Average condition at the time of purchase

MMRAcquisitionRetailAveragePrice:

Acquisition price for this vehicle in the retail market in average condition at the time of purchase

MMRAcquisitonRetailCleanPrice:

Acquisition price for this vehicle in the retail market in above average condition at the time of purchase

MMRCurrentAuctionAveragePrice:

Acquisition price for this vehicle in average condition as of current day

MMRCurrentAuctionCleanPrice:

Acquisition price for this vehicle in the above condition as of the current day

MMRCurrentRetailAveragePrice:

Acquisition price for this vehicle on the retail market in average condition as of the current day

MMRCurrentRetailCleanPrice:

Acquisition price for this vehicle on the retail market in above average condition as of current day

MMRCurrentRetailRatio:

Ratio of MMRCurrentRetailAveragePrice and MMRCurrentRetailCleanPrice

PRIMEUNIT:

Level of demand with respect to a standard purchase

AUCGUART:

The risk that can be run with the vehicle, meaning how much guarantee the seller is willing to give

VNST:

Geographic region

VehBCost:

Acquisition cost paid for the vehicle at time of purchase

IsOnlineSale:

1 = Sale done online, 0 = No

Warranty cost:

Warranty price (term = 36month and millage = 36K)

ForSale:

Whether is car is available for sale

IsBadBuy:

1 = Yes, 0 = No

----------

What is the difference between PurchaseTimestamp and PurchaseDate?

Are Make and Nationality redundant?

Is **VNST** – Geographic region code redundant with Nationality?

There is TopThreeAmericanName, does it mean that the car is from American manufacturer? If so, why is there still a nationality variable,

Are **WheelTypeID** / **WheelType redundant?**

**Redundancy**

To reduce the complexity of the modelling tasks, I'm going to reduce the dimensionality of the data.

I'm going to implement redundancy analysis to check which columns are redundant.

In this dataset, there are many columns related to prices. It is likely that some of them might be redundant. So,

WheelTypeID might be redundant with WheelType. Nationality could also be concluded using Make. TopThreeAmericanName could also be concluded using Make. PurchaseDate and PurchaseTimestamp likely convey the same information.

|  |  |  |
| --- | --- | --- |
| **Status** | **Feature** | **Reason** |
| **Kept** | MMRAcquisitionAuctionAveragePrice | Representative for auction prices at acquisition time (average condition) |
| **Dropped** | MMRAcquisitionAuctionCleanPrice | r = 0.990 with MMRAcquisitionAuctionAveragePrice; redundant |
| **Kept** | MMRAcquisitionRetailAveragePrice | Representative for retail prices at acquisition time (average condition) |
| **Dropped** | MMRAcquisitonRetailCleanPrice | r = 0.990 with MMRAcquisitionRetailAveragePrice; redundant |
| **Kept** | MMRCurrentAuctionCleanPrice | No highly correlated twin; provides unique “current auction clean” info |
| **Kept** | MMRCurrentRetailAveragePrice | Representative for retail prices at current time (average condition) |
| **Dropped** | MMRCurrentRetailCleanPrice | r = 0.990 with MMRCurrentRetailAveragePrice; redundant |
| **Kept** | MMRCurrentRetailRatio | Low-to-moderate correlations (r < 0.65); provides distinct “condition premium” info |

From <[*https://chatgpt.com/g/g-p-686f470df750819183989f2d4d6e36dc-machine-learning/c/6893222f-44ec-8328-8539-013a7b0d10b5*](https://chatgpt.com/g/g-p-686f470df750819183989f2d4d6e36dc-machine-learning/c/6893222f-44ec-8328-8539-013a7b0d10b5)>