

## Data Enrichment Recommendation

### Data Dictionary

Observation from the given dataset used for the analysis shows that most of the column names although self explanatory, their values have arbitrary meanings making it difficult to draw concrete conclusions.

For example, the state column holds values such as `ca`, `fl`, `tx`, `ny`, `hi`, etc which are an abbreviation of state names, however we are not so sure what the actual state names are. Same thing applies to other such columns. Also the `price` column just holds arbitrary integers without specific mentions whether the values are in dollars or cents. A currency type column would also aid interpretation and analysis. Another observation is that the `removal_date` column is quite ambiguous hence conclusive statements can't be made whether the vehicle was finally sold on that date or it was merely removed from the listings. Finally the odometer measuring unit was not specified whether it is in kilometres or miles.

For this problem, I would suggest that a data dictionary be provided so that the analyst can have an in depth understanding of the underlying analysis when drawing conclusions and making inference especially for columns such as price, state and odometer

### Show a sold flag

One important column I considered missing is a column that flags whether the listed vehicle was sold within the period it was listed.

The available data only shows the `posting_date` and `removal_date` which is not very intuitive because it will be erroneous to say the vehicle was sold on the removal date when in fact it was simply removed from the listings for some other reasons.

For this, I recommend that column indicating sold out vehicles be added to the dataset for proper intuition and inference