**Car Price Report and Analysis: 02/2025**

### **Business Understanding**

Buying and selling used cars is a difficult problem because the market participants do not have enough information to make an informed decision.

This problem is called "problem of asymmetric information in markets", and it's particularly relevant in the context of buying and selling used cars.

Here's a summary of the business problem it addresses:

Business Problem: Asymmetric Information in the Used Car Market

* In the used car market, sellers typically have more information about the quality of the car than buyers.
* This imbalance of information creates a business problem:
* Adverse Selection: Buyers cannot easily distinguish between high-quality used cars and low-quality ones (the "lemons").
* As a result, they are only willing to pay an average price that reflects the expected quality of all cars in the market.
* This average price is lower than what sellers of high-quality cars would demand, leading those sellers to exit the market.
* Over time, the market becomes dominated by low-quality cars, as sellers of good cars are unwilling to sell at the lower prices.

Market Breakdown:

* As high-quality cars disappear from the market, buyers further lower their price expectations, exacerbating the problem.
* This cycle can lead to a market collapse, where mostly the low-quality cars are traded, or the market may fail.

Implications for Business:

* Trust and Reputation: Businesses must find ways to signal the quality of their products to overcome buyer skepticism.
* Information Transparency: Markets function better when information is better distributed accross buyers and sellers.
* Platforms or intermediaries that provide reliable information about product quality can help mitigate the problem.
* Price and Quality Mismatch: Sellers of high-quality goods may struggle to obtain fair prices.

The Opportunity for Business:

* Knowing what drives car prices may shed light into this problem to help businesses understand what drive sales in used cars.

Reference: "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". George A. Akerlof, The Quarterly Journal of Economics, Vol. 84, No. 3 (Aug., 1970), pp. 488-500 (13 pages). JStor link: https://www.jstor.org/stable/1879431

### **Data Understanding and Data Preparation**

* The data is coming from the “vehicles.csv” provided by the client.
* The data contains 426,880 rows with car data, including price and it’s features.
* The data contains many errors that would avoid modeling.
* The data cleansing process led to 200,677 cars with useful data for modeling purposes.
* Cars with prices below $1,000 and year before 1970 were also removed, as they were biasing the data.
* The actual price paid for a car as used as the target variable.
* The features analyzed were:
  + year
  + odometer
  + cylinders
  + manufacturer
  + condition
  + fuel
  + title status
  + transmission
  + drive
  + size
  + type
  + paint color

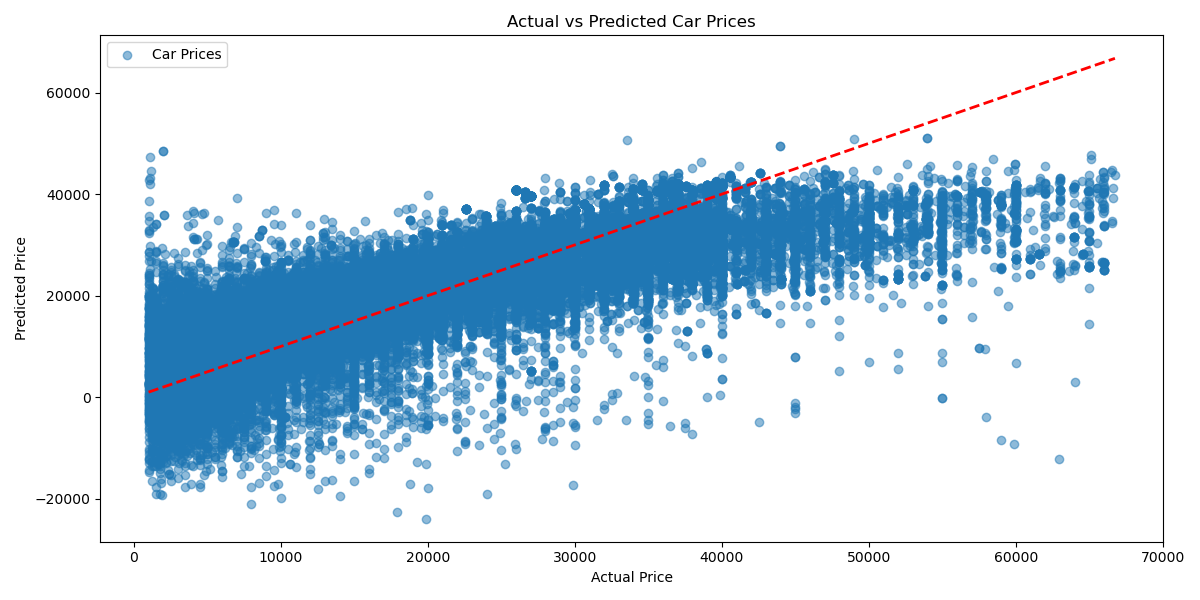
### **Modeling and Evaluation**

* We used various linear regression techniques to evaluate and predict the car prices:
  + Simple linear regression
  + Multiple linear regression
  + Ridge regression
  + Lasso regression
* The multiple linear regression - Ridge performed better than the simple linear regression.
* Ridge and Lasso performed similarly. We opted for Ridge as it was slightly better than Lasso.
* We used techniques to process and select the most relevant features:
  + SVD
  + PCA - principal component analysis
  + k-means clustering
  + SFS - sequential feature selection

### **Deployment and Recommendations**

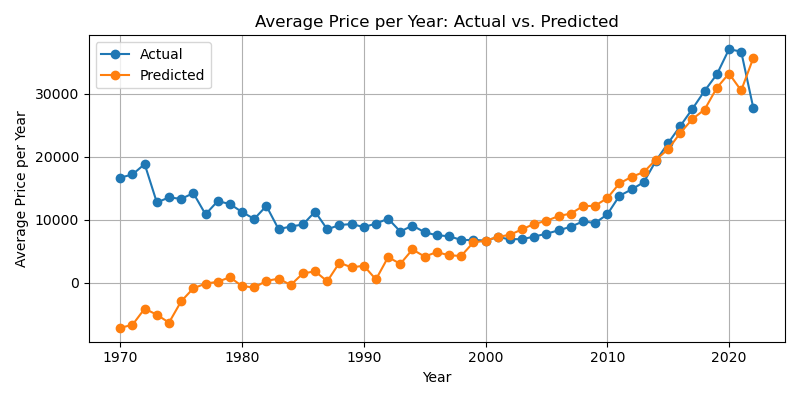
Please look at the images folder for more detailed data. We only added the most relevant charts in this report.

Our model predicted the car prices reasonably well for prices below $40,000 dollars.



Approximately between the years 2000 and 2014, the predicted prices were ABOVE the actual prices sold to customers. This indicates that there is an opportunity to RAISE prices for the types of cars sold in that timeframe. We recommend breaking down the data of this period to design marketing strategies for similar situations in the future.

The opposite occurred between 2014 and 2020. The predicted prices were below the actual prices sold. We recommend another study to understand what changed, and apply this knowledge to a version 2 of this model.



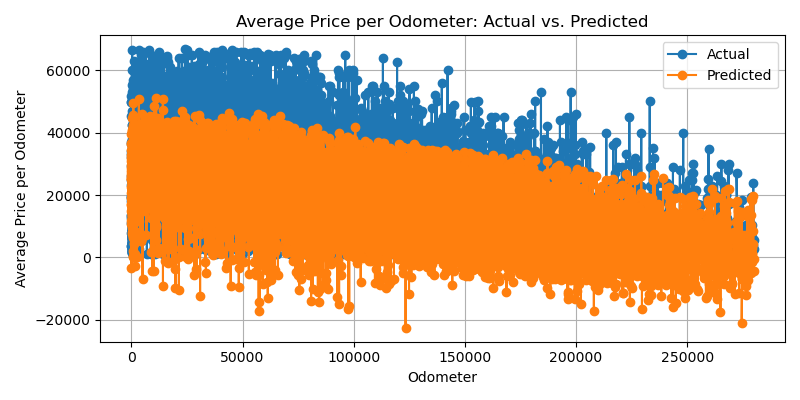
The two most important features affecting prices were:

* Year
* Odometer

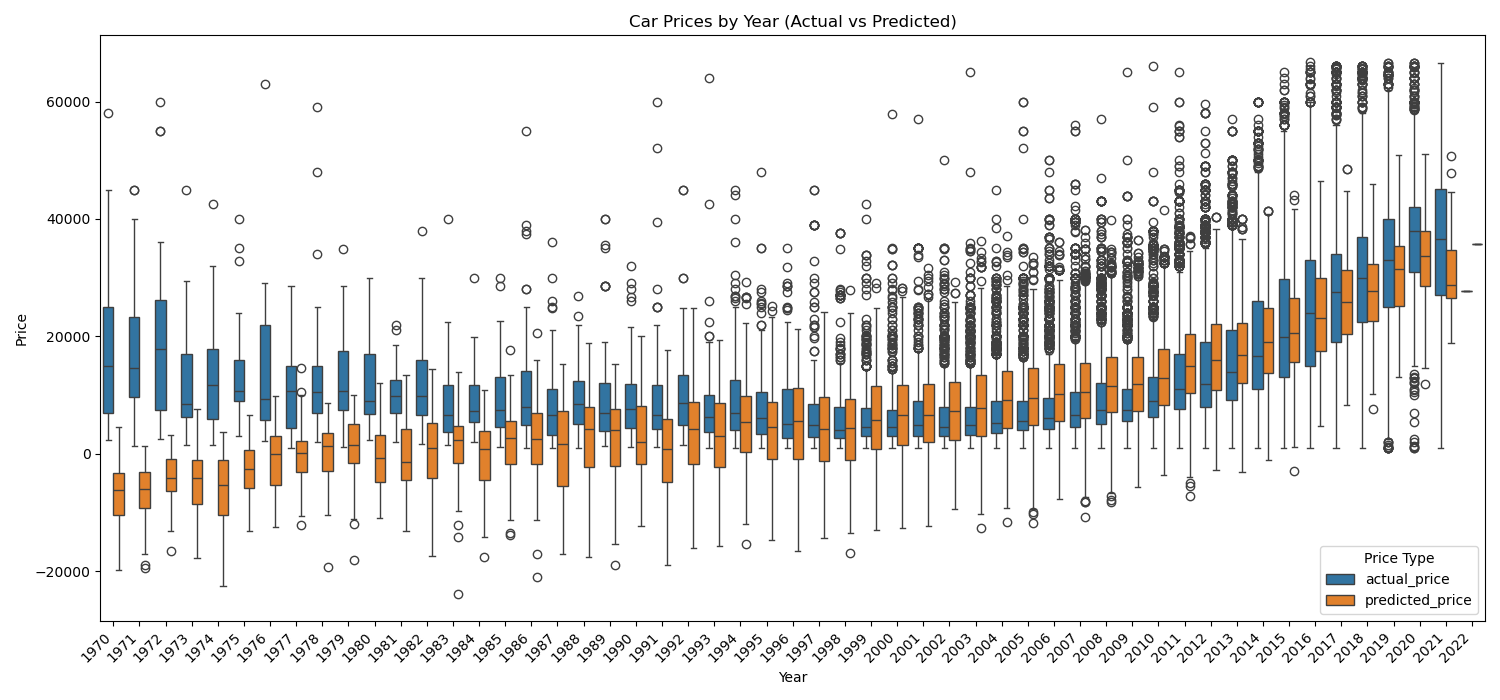
Odometer was considered more important than Year. That means customers were willing to buy older cars if the odometer was low. This strategy can be explored in the marketing campaigns and sales efforts.

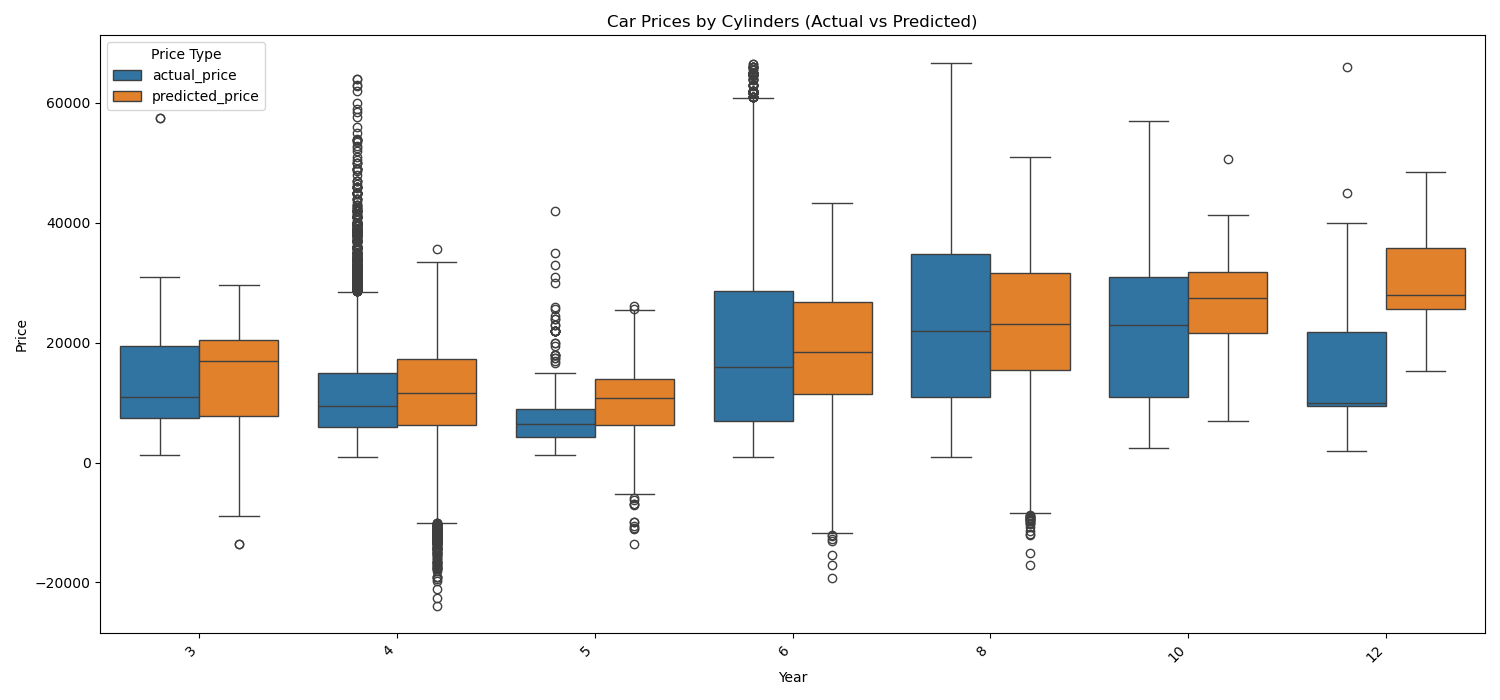
Emphasizing odometer and year combined appears to be the best strategy when the two features are ‘good’ (that is, reference a ‘good’ car). If only one of the two features is ‘good’, the odometer seems to be a better option for customers.

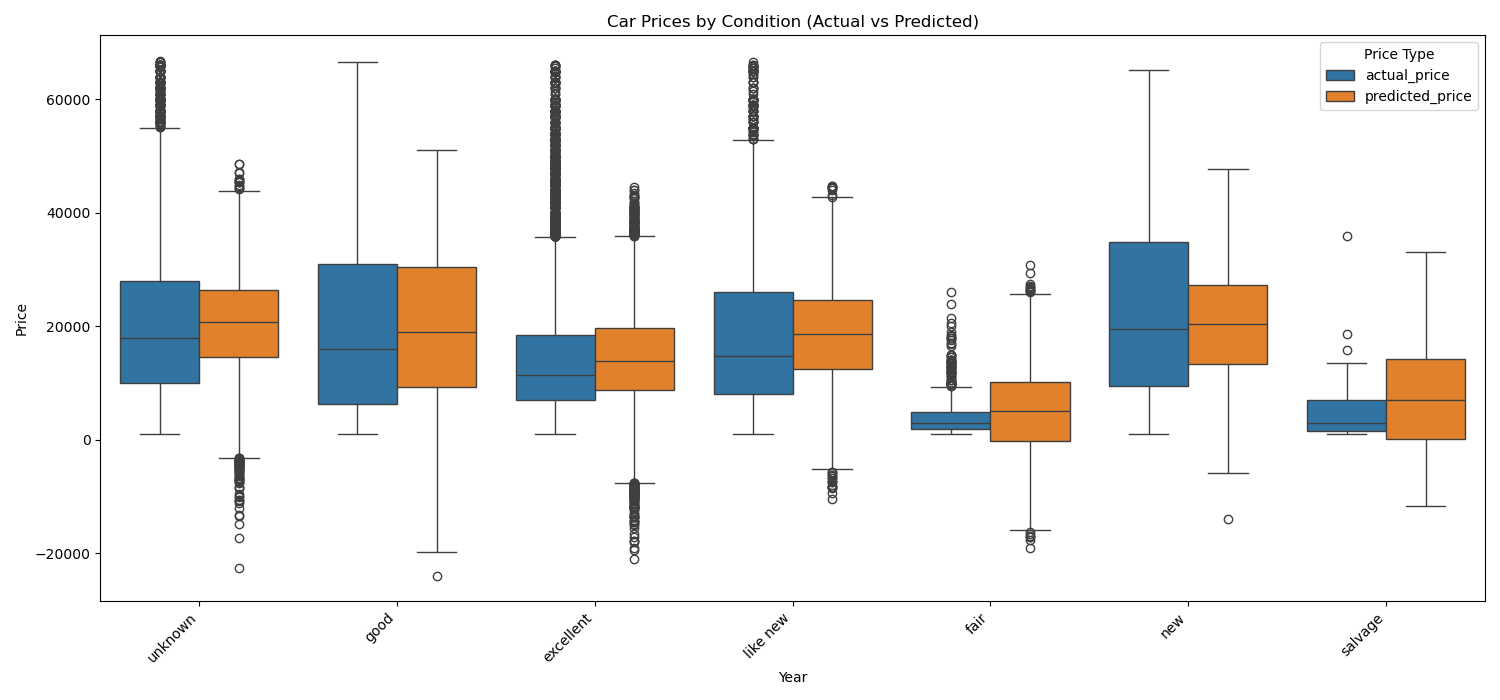
Our model also predicted prices per odometer values that are consistent with this recommendation.

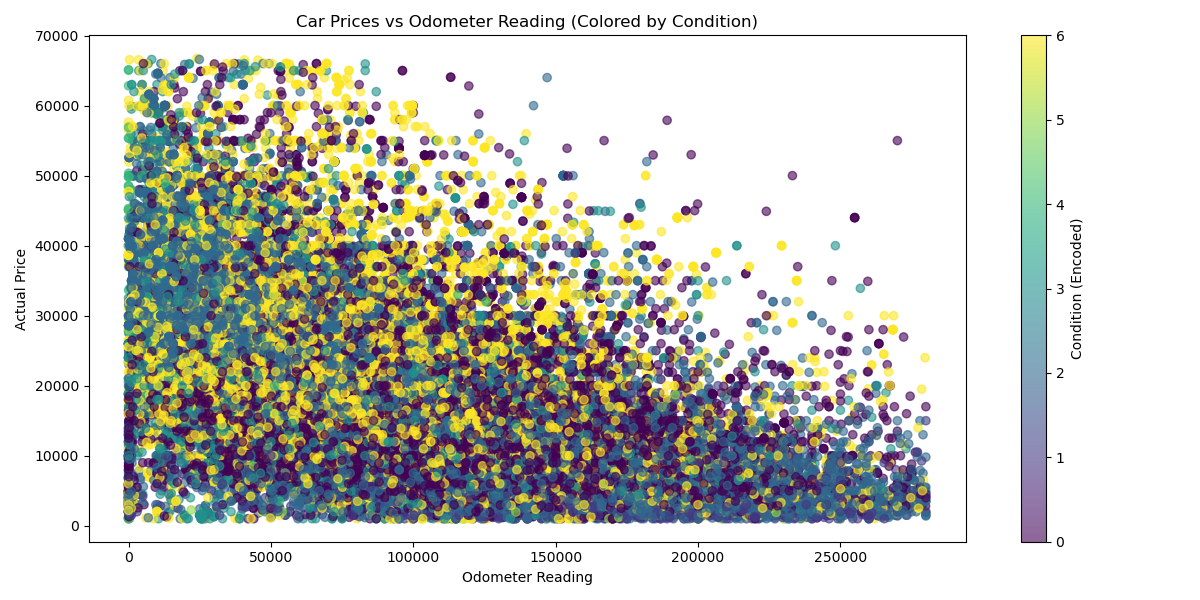


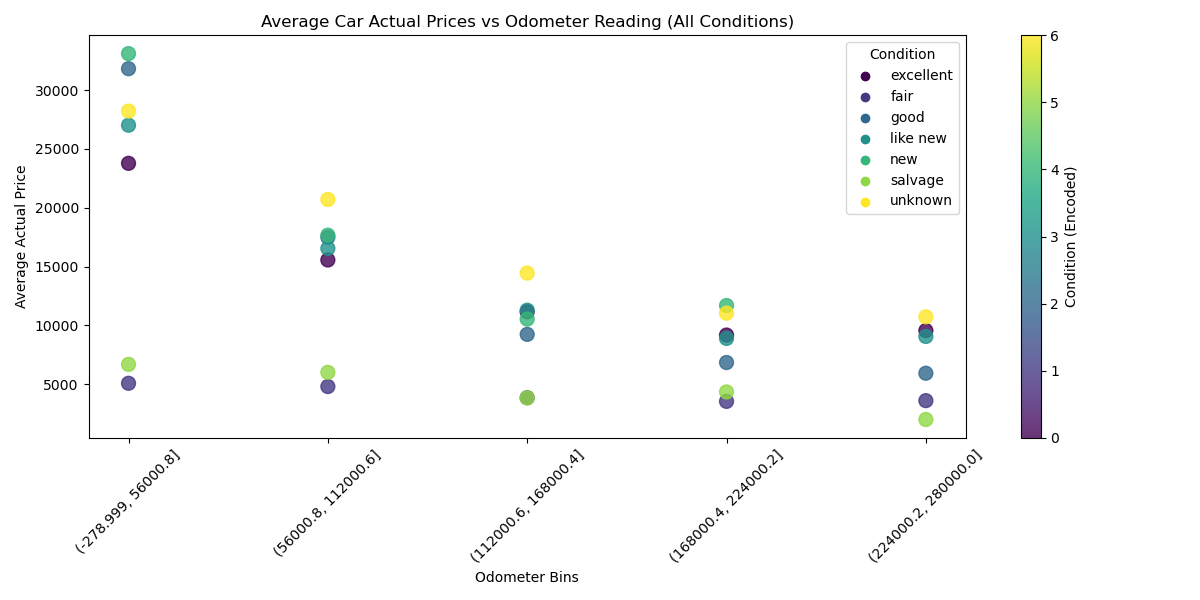
The report below show the variation of car prices per year, odometer and condition of the car (good, bad):

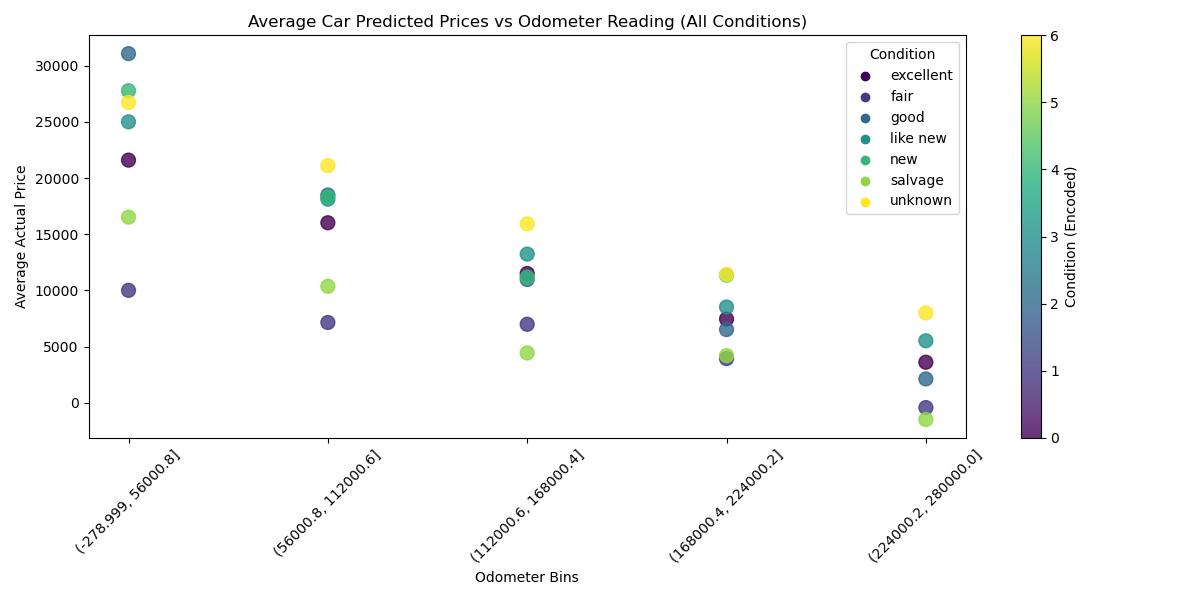












The analysis below show the histogram for odometer, actual car prices and year:

