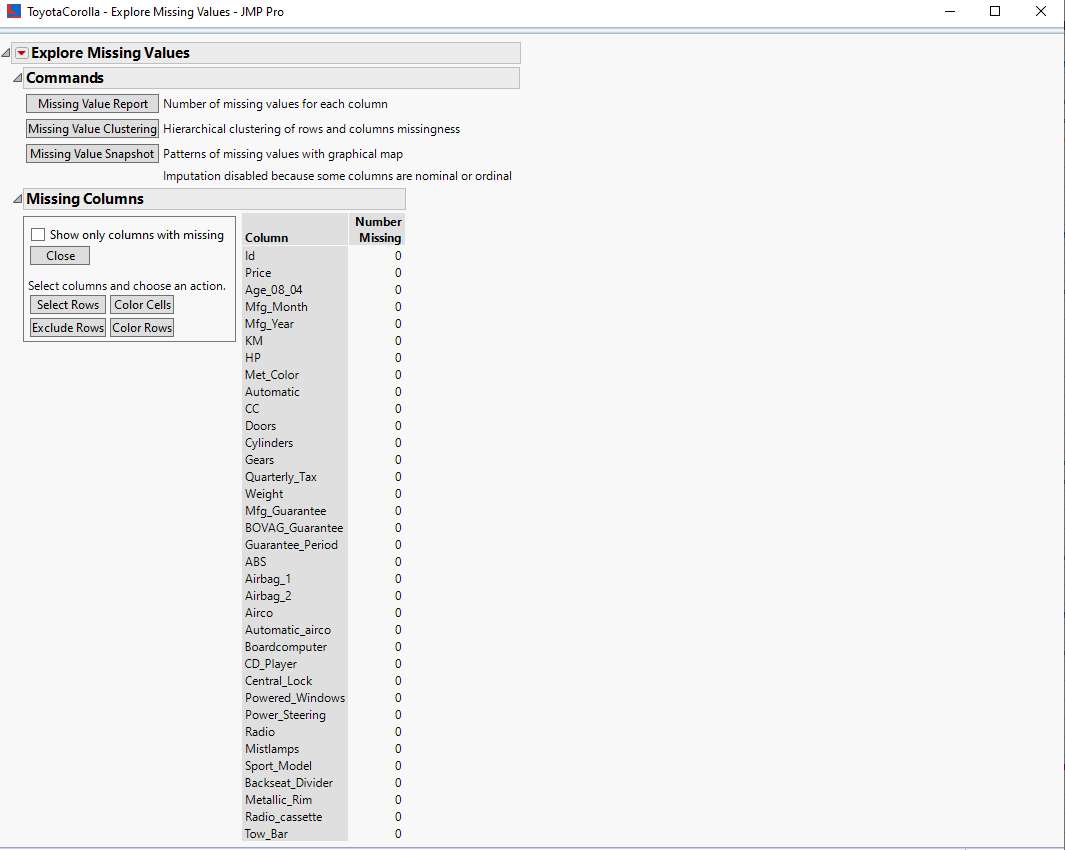
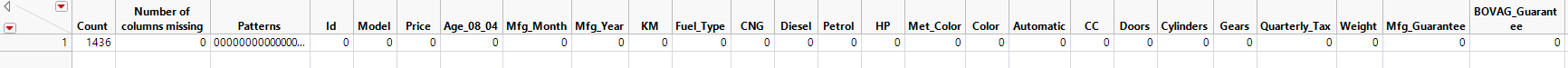
1. **I checked for missing values in the data using two methods:**
2. **By Screening and exploring missing values**

Analyze -> Screening -> Explore missing values



1. **By Analyzing the missing data pattern**

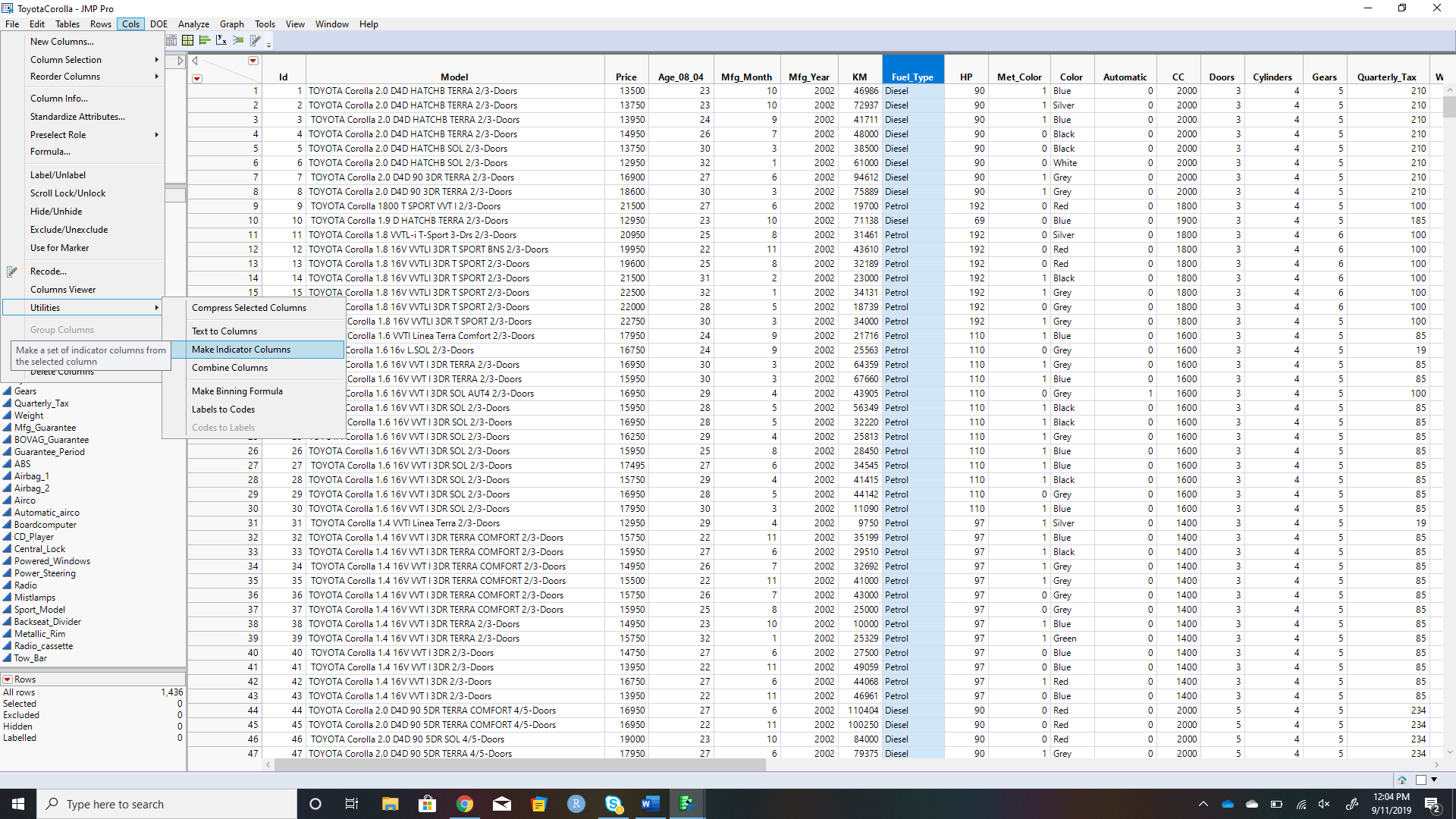
Tables -> Missing Data Pattern

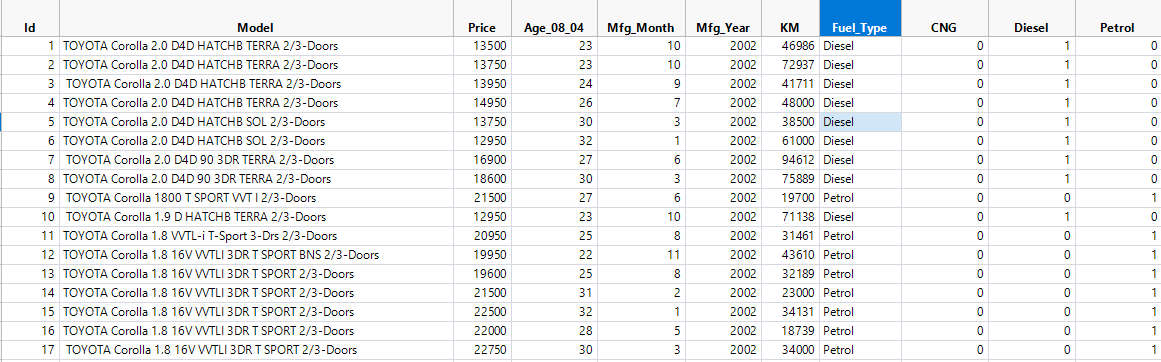


**There was no missing value found.**

1. Category variables are created by selecting

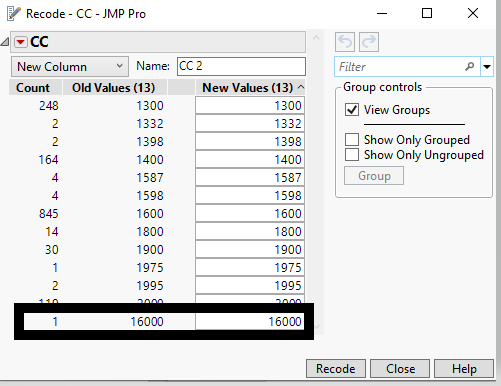
Cols -> Utilities -> Make indicator Columns

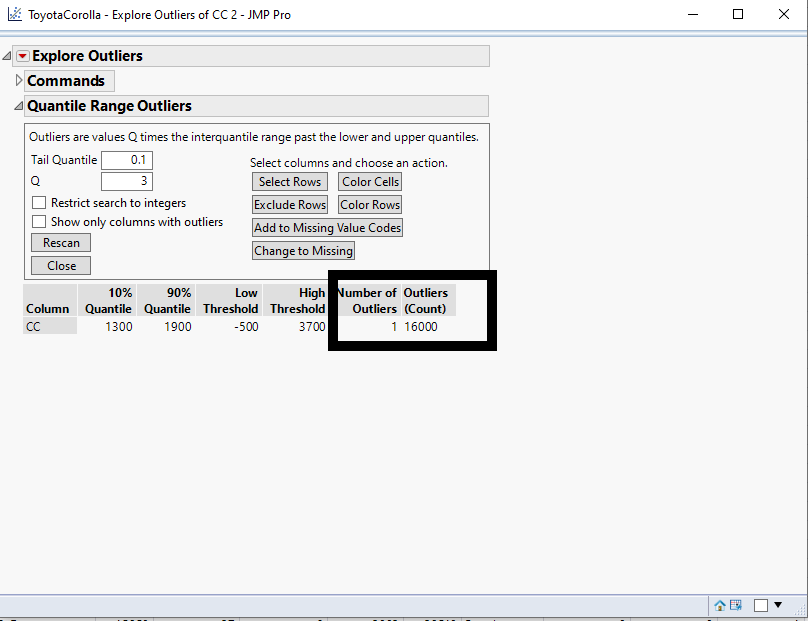




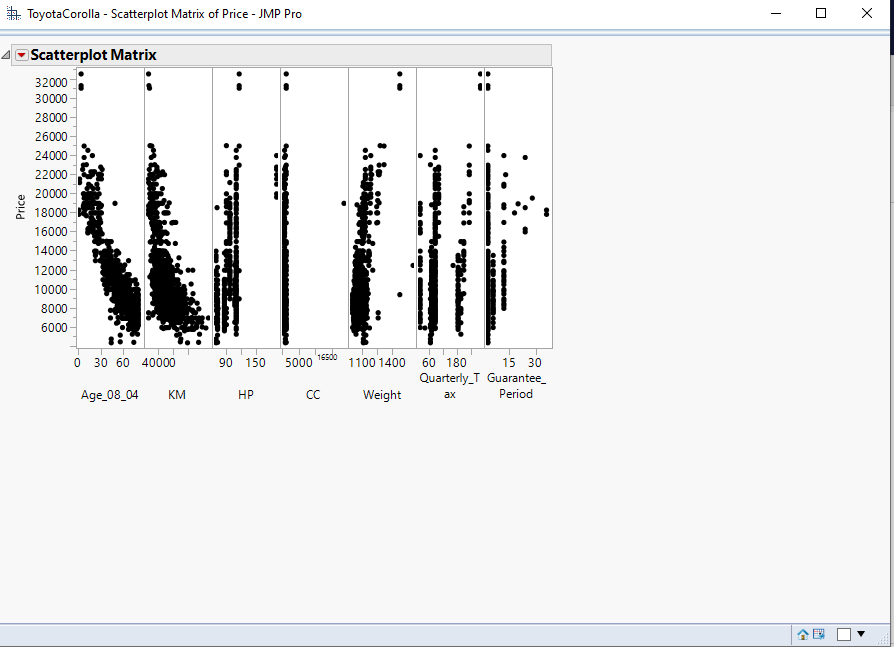
“Fuel Type” predictor has three categories -Petrol, Diesel, and CNG. Therefore, we have two dummy variables in the model (Petrol (0/1) and Diesel (0/1); the third CNG (0/1) is redundant given the information can be derived from the first two dummy variables.

1. There is one outlier in this column – 16000. The Quantile Range Outliers method of outlier detection uses the quantile distribution of the values in a column to locate the extreme values. Quantiles are useful for detecting outliers because there is no distributional assumption associated with them.





**4.**



The following observations have been made from the matrix,

**AGE** : The newer the car, the more expensive it is.

**KM** : The more miles the car has run, the cheaper it is.

**HP** : The more horsepower, the more expensive the car is. Though not always the case(There are some outliers).

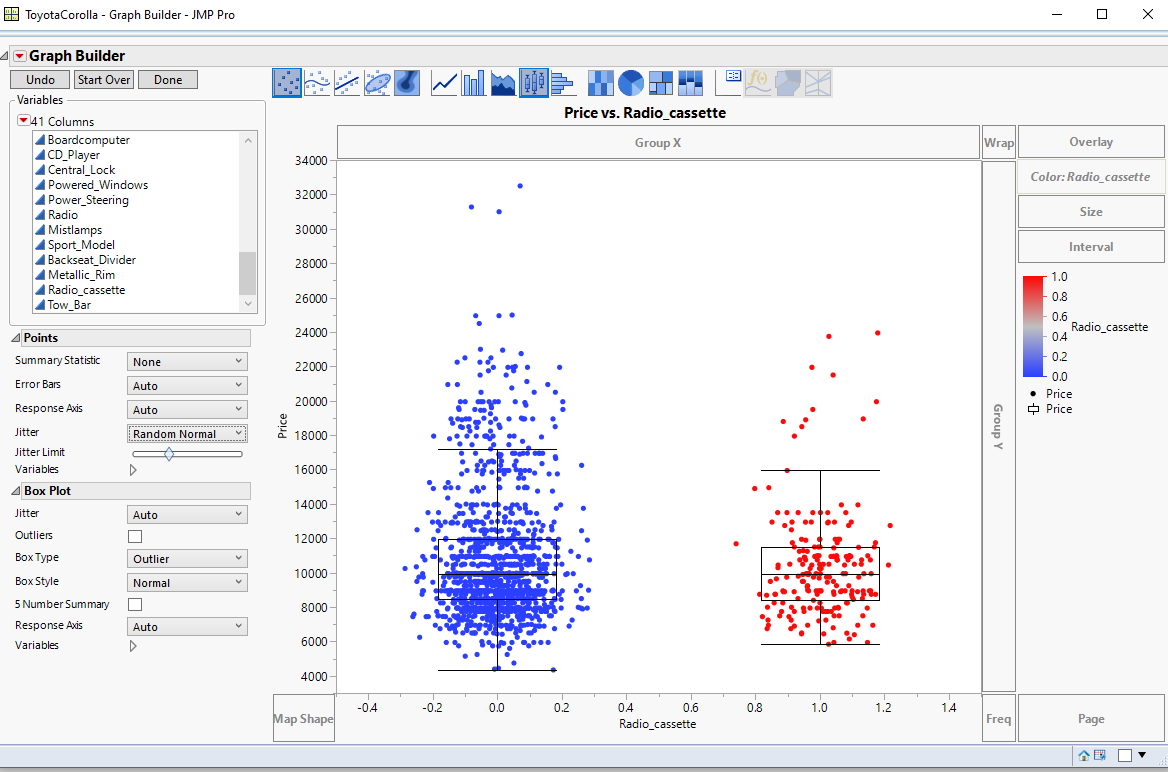
**CC :** The more cylinder, the more expensive the car is. Though not always the case(There are some outliers).

**Weight :** Bigger cars cost more. Though not always the case(There are some outliers).

**Quarterly\_Tax :** It depends. (There are some outliers).

**Guarantee\_Period :** It depends. (There are some outliers).

**5.**

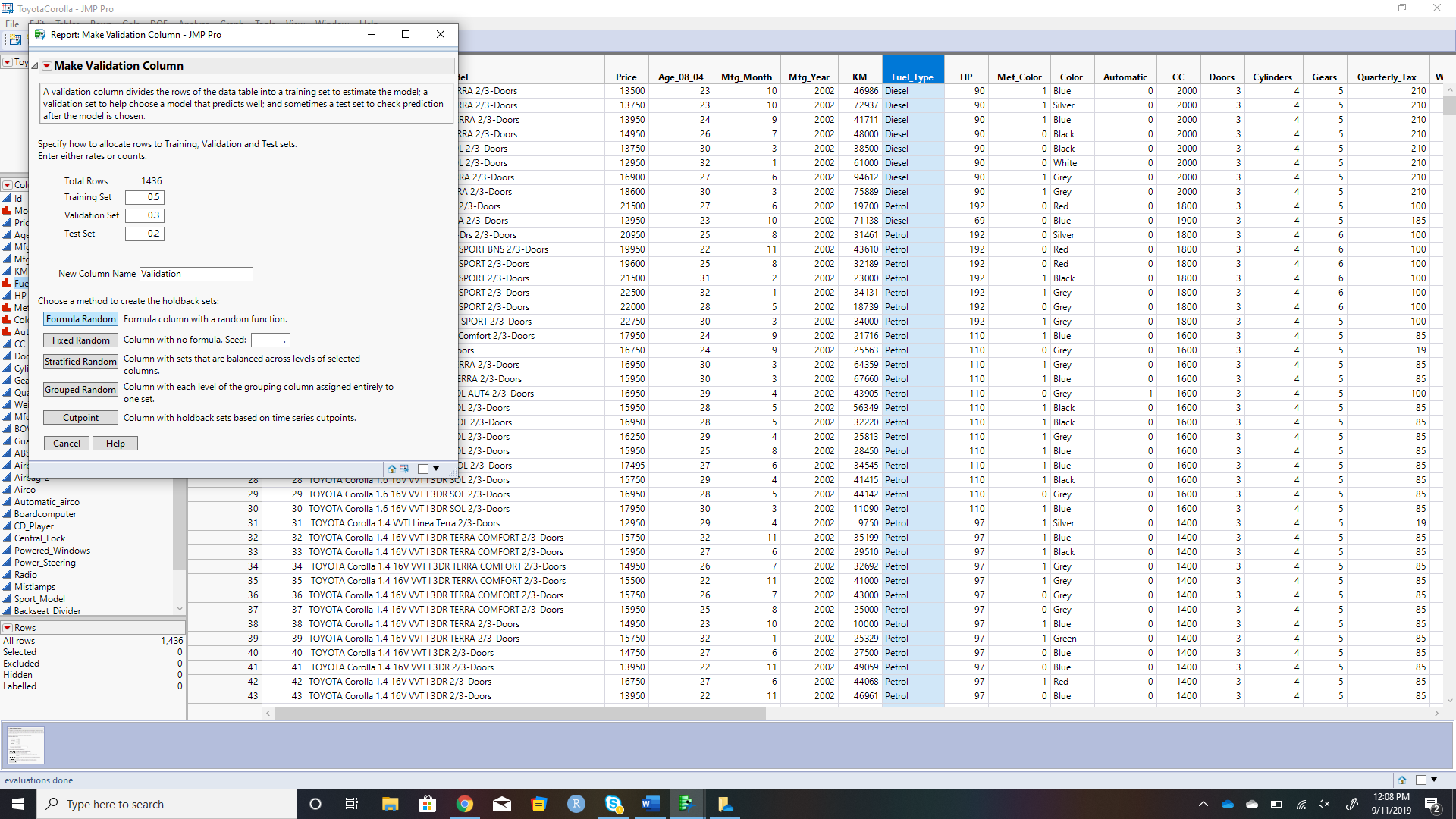


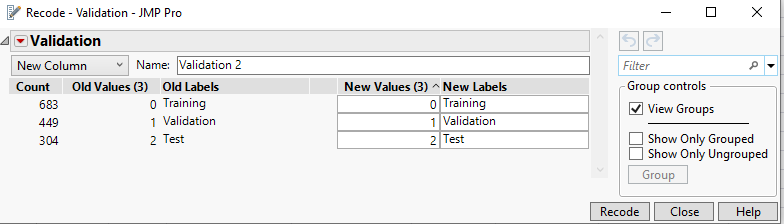


As per my interpretation, CD\_Player is more predictive of vehicle Price. We can see that the car having CD\_Player has average price greater than the car without one from the box plot. And with respect to Radio\_Cassette, the average price doesn’t change with or without this feature. However, there are some outliers (High Priced cars) which indicates that there are high end cars not using the obsolete radio cassettes. There are a greater number of cars without radio cassettes.

**6.** Set the partition by selecting

**Analyze -> Predictive Modeling -> Make Validation Column and setting the values to 50% training, 30% validation and 20% testing.**





**The row count of the sets are listed in the above screenshot captured via recoding the Validation field.**