How Different Factor Influence The Price Of The Donated Cars

**Introduction**

Car donations are becoming much more popular around the world today. The main reason can be summarized in terms of eight main points. The first reason is charitable contributions, since donating cars can allow individuals to support organizations and initiatives they care about without a direct monetary donation. The second reason is the tax benefits. In some countries, individuals will be able to receive tax benefits by donating the vehicles to qualifying charitable organizations. The third reason is environmental awareness. As more environmental problems occur around the world, people tend to donate their cars as a way to promote recycling and reduce the environmental impact associated with manufacturing and disposing of vehicles. The fourth reason is convenience. Car donation programs often make the process easy and convenient for the donors. The fifth reason is that social media and online platforms have made it easier for charitable organizations to promote their car donation programs and for individuals to learn about these opportunities. The increased visibility has contributed to the growing popularity of car donations. The sixth reason is marketing and outreach. Charities and nonprofit organizations have been more adept at marketing and outreach, using various channels to highlight the benefits of car donations. This has helped raise awareness and attract more donors. The seventh reason is financial support for charities. Car donations provide charities with an additional source of funding. The proceeds from selling donated vehicles can support various charitable initiatives, ranging from education and healthcare to environmental conservation. The eighth reason is an alternative to selling and trading. Since people find it more appealing to donate their cars rather than go through the process of selling or trading them in, donating is the quicker and simpler way to part with a vehicle while supporting the cause. All eight reasons have contributed to the increasing popularity of car donations worldwide. It’s a trend that aligns with a growing desire among individuals to make a positive impact through their actions, including the disposal of unwanted assets like old vehicles.

Analyzing how different factors affect the price of donated cars has become a crucial process for companies that have car donation programs. Not only for maintaining transparency, optimizing fundraising efforts, and building donor trust, but also for ensuring legal compliance and being able to continuously improve the effectiveness of the car donation program. There are nine key major benefits. The first benefit is transparency and accountability: understanding the factors that affect cars contributes to transparency in the donation process. Donars, in public, have the right to know how the value of their contribution is determined. The second benefit is optimizing fundraising. By analyzing factors affecting prices, organizations can optimize their fundraising strategies. This includes identifying which types of vehicles tend to fetch higher prices and tailoring marketing efforts accordingly. The third benefit is donor confidence. Let the donors feel confident that their contributions are being used effectively. Knowing that a charity has a well-thought-out and transparent process for valuing and selling donated cats can enhance donor trust and confidence. The fourth benefit is the legal and ethical compilance. Different factors can influence the tax implications of car donations. Ensuring the valuation process complies with legal and ethical standards is crucial for both charities and donors. The fifth benefit is fairness and consistency. A clear understanding of the factors affecting donated car prices helps ensure fairness and consistency in the valuation process. Consistency is important for building trust among donars and avoiding perceptions of bias or inconsistency. The sixth benefit is educating Donars. By analyzing and communicating the factors influencing donated car prices, the company can educate donors about the complexities of the valuation process. This education can help manage donor expectations and dispel misconceptions. The seventh benefit is improving processes. Regular analysis allows charities to identify areas for improvement in their car donation programs. Involve refining valuation methodologies, enhancing marketing strategies, and streamlining administrative processes. The eight benefits armarketet Trends and changes The automotive market is dynamic, and factors influencing car prices can change over time. Regular analysis enables companies to stay informed about market trends, allowing them to adapt their strategies accordingly. The ninth benefit is maximizing impact. The company aims to maximize the impact of their fundraising efforts. Analyzing factors affecting donated car prices helps organizations make informed decisions to ensure they are getting the most value from each donation.

**Data Exploration**

The Sound of Hope car donation dataset is filled with records from 2020 to 2022. With the basic eight information groups, The first group is general information, including operation ID and status, dates related to the sale, and time-related details. The second group is the vehicle's details. Which include the name of the donar and the car details. The third group is the contact information, which includes the phone, address, and email of the donor. The fourth group is additional information, which includes remarks, information sources, carfax information, and registration fees. The fourth group is Sales and Pricing, which includes sale-related details and buyer information. The sixth group is documentation and recordkeeping, such as leave messages, post-bill dates, and ticket information. The seventh group is inventory and eferral, such as stick work logs, referrer information, and the buyer’s phone. The eighth group is miscellaneous details, such as information on obtaining a new plate from a dealer or using tax. After having a basic idea of what fields the dataset included, the next step is data exploration.

Data type is always important when doing any kind of exploration. The first step is to check what types of data are in the dataset. After checking, most of the data have the type of floar and object. The next step is to check our target variable price. After checking, find out if the price variables have the type of object. By doing data type conversion, transform the price into a float type. The next step is to separate the data into numerical and categorical categories. For numerical data, you need to use a histogram to check the distribution; for categorical data, use a bar plot for distributions. For numerical data, the main focus is the sold\_year, sold\_month, car\_year, mileage, cost, and price. The sales year is mainly distributed from 2019 to 2022. For the sold\_month variable, the distribution is uniformly distributed. As for the car\_year, the distribution is little bit skewed to the left, spread around 1981 to 2014. For mileage, the distribution is skewed to the right, with most values spread around from 50k to 350k. For cost the distribution is also skewed to the right, with most of the values spread from 100 to 400. Lastly, we need to check the most important variable, the price, with most of the value spread from 1100 to 6100.

**Question Exploration**

**Question: Does the time spent donating the car affect the car's price? Season or quarter**

|  |  |
| --- | --- |

|  |  |
| --- | --- |

After looking at the dataset and doing some basic exploration, the next step is to raise a question to do further investigation. The first question that comes to mind is does the time factor play an important role in terms of price prediction? In order to answer this question, the first step is to separate the month into different quarters and seasons. After creating a dictionary for quarter and season, we are mapping it back to the car donation dataset by adding two new columns, Quarter and Seasonality. The second step is to do a groupby for A using quarter and seasonality columns and also aggregate both of them using the mean price and mean cost. By printing out the results, we can observe that prices vary across quarters, with the highest average price in quarter 3 and the lowest average price in quarter 2. The cost also shows variability, with quarter 3 having the highest average cost. After checking the quarter, we next need to check the seasonality. By checking the result, we can observe that there is a clear seasonal pattern observed in both price and cost. The highest average price is in summer, while the lowest is in fall; the highest average cost is in winter, while the lowest is in fall. In conclusion, quarter 3 has the highest average price and cost, and quarter 2 has the lowest average cost. The concussion for seasonality is that summer has the highest average price, fall has the lowest, winter has the highest average cost, and fall has the lowest. Then, by creating scatterplots with trendline, we cn observe the trends for seasons and quarters more clearly

The reason behind the observation is that quarter 3 may experience higher demand, leading to an increase in both price and cost. Increased demand can often drive prices higher, and meeting this demand might require additional resources, contributing to higher costs. For the seasonal trend, certain industries or products may experience a natural peak in sales or activity during Q3, impacting both prices and costs. Quarter 2 has the lowest average cost; the reason behind this might be that operational efficiency may improve in Q2, leading to a lower production or operational cost. This could be due to factors such as improved processes, economies of scale, or cost-saving measures. Next, the conclusion about seasonality is that summer has the highest average price. The reason behind this might be seasonal demand, since some products or services might experience higher demand during the summer months, leading to an increase in price. This can be due to seasonal preferences, holidays, or specific events associated with the summer. The second reason might be a supply constraint. Limited availability of certain goods or services during the summer season may drive prices higher. Fall has the lowest average price. The reason behind this might be reduced demand. Fall might be a period of lower demand for certain products and services, leading to a lower price. Consumers may also be less active in the market during this season. Winter has the highest average cost. The reason behind this might be operational challenges. Winter conditions can pose operational challenges, leading to increased costs. This could include higher energy costs, maintenance expenses, or logistical challenges associated with winter weather. The second reason behind this might be seasonal labor costs. Industries relying on seasonal labor might face higher costs during the winter months, impacting overall production costs.

**Question: Is there a chance that different people who repair the car tend to have a different price for selling that car?**

|  |  |
| --- | --- |

|  |  |
| --- | --- |

|  |
| --- |

The second question I am curious about is do different people who repair the car tend to have a different price for selling that car? Their might be some workers who have the best skills and are able to fix cars quickly. Which might lead to the fact that the donated car’s price tends to be higher. By grouping the dataset using repair and price columns aggregated by means, according to the bar pot, we can observe that the repair person with the name Bai or Khai tends to have the highest average price. Next, by creating a pie chart, we can also observe that Khai has done the most of the car fixes, since the percentage of cat fixes done by Khai is 69.2%, which is almost 70%. There is also another important factor that determines a car’s price. We can observe Khai mainly taking care of Ford and Toyota; prices for both makes are high. Also, create a box plot to check for outliers and compare the distribution between different people's average car donation prices. Next, in order to have a deeper and clearer understanding of the effect of fixed people on car prices, we need to perform an ANOVA test. Before using Anova, you first need to check the dot plot. Both people with the names Bai and Khai have the highest average price for donated car’s. After checking the dot plot, we are ready to do the ANOVA analysis. After performing the analysis, the p-value for the repair variable is 0.537932, which is greater than the typical significance level of 0.05. This suggests that there is not enough evidence to reject the null hypothesis. The repairer variable has no significant effect on the dependent variable. Based on the current analyses, there is no statistically significant difference in the dependent variable across different levels of the regression variable Conclusion: The repairer variable does not seem to be a statistically significant predictor of dof theendent variable based on the given p value. The observed variability in the dependent variable does not appear to be significantly associated with different Repairers

**Question: How do different economic events happening around the world impact the frequency of car donations?**

The economic event will also play an important role in terms of price influence. As an car donation company, they all very curiosis and worried how economical event such as inflation of financial crisis or even pandemic might influence their company’s sales and revenues. The First factor wish to explore is how federal interest Rate iinfluencesthe price of the donated car’s. First, get the outside federal interest rate data from website and combine with the car donation dataset. Then mainly focus on the range from 2020 to 2022, since this is the time for car donation program. Next, we are matching the axis by setting a double axis in order to compare.

|  |
| --- |

By observing the plot, it was revealed that there is only a small portion of the time range where the car donation dataset aligns with the federal interest rate. In order to examine this more deeply, the next step is to filter the data to focus only on the time range from 2020 to 2022.

|  |
| --- |

After focusing on the specific range, the conclusion is that although the federal interest rate is quite low, around 0.25, the price of donated car’s tends to increase overall. The relationship between the federal interest rate and the price of donated cars can be influenced by various economic factors.

Some of the reasons behind this might be economic stimulus and a low-interest rate. During periods of economic stimulus or an effort to spur economic growth, central banks, such as the Federal Reserve, may lower interest rates. Low-Interest rates can make borrowing more affordable, leading to increased spending and investment. This can contribute to economic expansion, potentially boosting consumer confidence and purchasing power, including in the automotive market. The second reason might be consumer spending and confidence. Low-Interest rates can encourage consumers to make big-ticket purchases, such as cars, by reducing the cost of financing. When interest rates are low, individuals may be more willing to take out loans to buy cars, leading to increased demand and potentially higher prices. The third reason might be credit availability. Low-Interest rates are often associated with easier access to credit. When credit is available and interest rates are low, consumers may find it easier to finance to purchase the cars. This increased accessibility ro credit can contribute to higher demand and prices in the car market. The fourth reason might be investor Behavior. Investors seeking higher returns may shift their investments away from traditional interest-bearing assets(like bonds) when interest rates are low. This shift in investment behavior can lead to increased demand for alternative assets. Including tangible assets like cars, potentially contributing to higher prices. The fifth reason is inflation Expectations. Low-Interest rates may be accompanied by expectations of low inflation. In such an enviornment, individuals may view certain assets, including cars, as a relatively stable store of value. This perception could contribute to increase demand and higher price. The sixth reason is Supply and Demand Dynamics. Economic conditions, including interest rates, can influence both the supply and demand for cars. If demand increases while supply remains relatively constant, prices are likely to rise. Factors such as changes in manufacturing costs, consumer preferences, and overall economic outlook can also play a role.

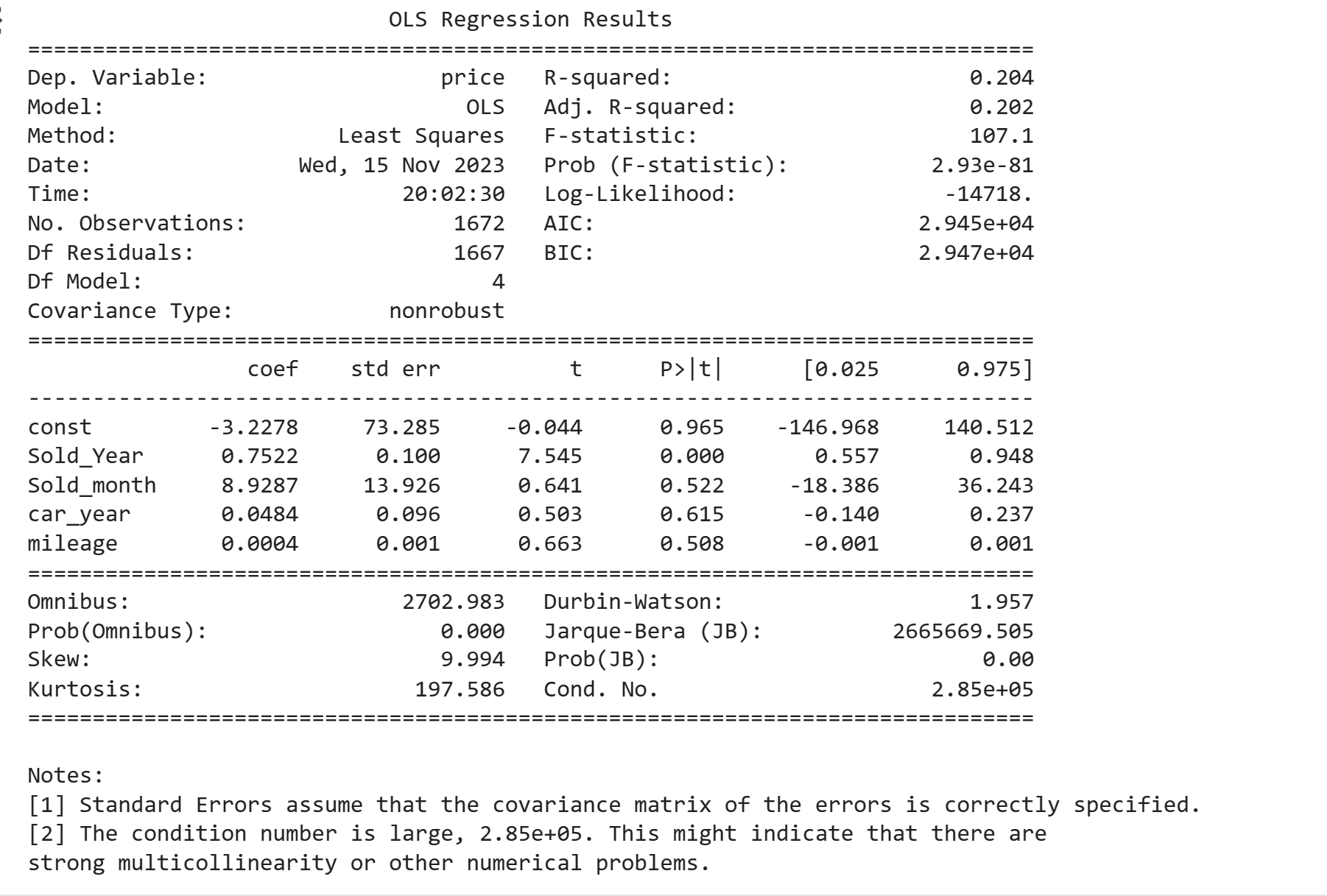
Next, need to check another important factor: Federal Funds. By getting data from the government website, and compare with the car donation dataset by setting into same time range. Here is the plot

|  |
| --- |

According to the plot, we can observe the the funds rate is really low amost close to 0 and the price of donated car is increasing. The reason are kind of simler as federal interest rate such as Economic Stimulus. Cost of Fiancing, Investor Behavior, Supply and Demand Dynamics, Consumer Confidence. The different part is there might also be in inflation Pressure. The Central banks may keep interest rate low to counteract low inflation. Inflation is a general increase in prices across the economy. When inflation is low, central banks may use low interest rates to encourage spending and investment to boost inflation.

**Question: Price Prediction Model**

**Model1 : Simple linear Regression**

****

During the data exploration phase, we explore the correlations between the numerical variable and the target variable price, then draw conclusions about the variables sold\_year, sold\_month,car\_year, and ,mileage. We fit an simple linear regression model using those four variables and the target variable price. According to the final output, the R-square measures the proportion of the variance in the dependent variable that is explained by the independent variables. In this case, R-squared is 0.204, indicating that approximately 20.4% of the variance in the dependent variable is explained by the independent variables. Adjusted R-squares adjusts the R-squared value for the number of predictors in the model. Here, it is 0.202. Next, need to check the F-statistics. The F-statistic tests the overall significance of the regression model. The value of 107.1 is associated with a p-value(Prob (F-stastic)) of 2.93e-81, indicating that the overall model is statistically significant. The fourth step is to check the coefficients, Each independent variable has an associated coefficient, standard error, t-statistic and p -vale. First check the intercept, -3.2278 with a large standard error (73.285), is not statistically significant (P>|t|=0.965). Next, we check the different predictor variables. First check the Sold\_Year. 0.7522 with a small p-value(P>|t|<0.001), suggesting a statistical significance. Next, check the Sold\_month, 8.9287 with a p-vale of 0.522, suggesting no statistical significance. Then check the car\_year, 0.0484 with a p-value of 0.615, suggesting no statistical significance. Lastly check mileage, 0.0004 with a p-value of 0.508, suggesting no statistical significance. In conclusion, the model is a whole, is statistically significant based on the F-stastic. Sold\_year is statistically significant and positively associated with price. Sold\_Month, car\_year, and mileage are not statistically significant in predicting the price. The overall model explains around 20.4% of the variance in price.

|  |
| --- |

After fitting a simple linear regression model, the next step is to check the residual plots between the residuals and the fitted values. The observation shows that most of the points are clustered around one side, in this case, the right side. It suggests there are some potential issues with the model. Some problems that need to be considered are the homoscedasticity issues. If the spread of residuals increases or decreases systematically as the predicted values increase, this indicatproblemblems of homoscedasticity. In this case, the residuals are clustered on one siwhich, it might suggest that the model’s variance is not constant across all levels of the predicted values. The second problem might be non-linearity; if the residuals are not randomly scattered around the horizontal axis and exhibit patterns (like being clustered to one side), it may suggest non-linearity in the data. The model may not adequately capture the underlying relationship between the predictor variables and the response variable. The third problem might be outliers or influential points; the cluster of residuals on one side may indicate the presence of outliers or influential points. These data points have a significant impact on the regression model. Identifying and addressing outliers may improve model performance. The fourth problem might be model pecification; a cluster of residuals on one side could also be a sign of model misspecification. The chosen model might not be the best fit for the data, or important predictor variables might be missing.

|  |
| --- |

The next step is using the qq plot to check for normality. Throughout the trend line, it is clearly linear; however, there are still some outliers that influence the normality. The observation Most points are falling. Along a line, since the majority of points fall along a straight line, it suggests that the central position of your data conforms well to the theoretical distribution (normal distribution). However, there is deviation at the ends. Deviations, particularly at the ends of the QQ plot, suggest that the tails of the data may not follow the expected distribution. This can indicate heavier or lighter tails than the theoretical distribution. Points deviating at the ends may be indicative of outliers or skewness in the data. Outliers can disproportionately affect the tails of the distribution and result in deviations from the expected linear pattern. If the deviations are systematic and consistent, it might suggest that the underlying distribution of the data differs from the assumed theoretical distribution. This could prompt a reevaluation of the distributional assumption.

Next, we wish to normalized the dataset and start making prediction for the testing set. Below is the results.

|  |
| --- |

By examining the results, we first need to check the model, checking the i is 1144.2697. The expected value of price when all predictor variables are 0 Next, check the sales-year coefficients. For each unit increase in sold year, the price is expected to increase by 721.9316 units. Then check the sold-month coefficient. For each unit increase in sold\_month, the price is expected to increase by 36.8038 units. Next, check the car\_year coefficient. For each unit increase in car year, the price is expected to increase by 43.9402 units. Lastly, check the mileage coefficient. For each unit increase in mileage, the price is expected to increase by 38.4716 units. Next, checking the model fit, an R-squared of 0.204 indicates approximately 20.4% of the variance in price is explained by the model.

Then we need to check the residual plot and qqplot again for the testing dataset. Nothing really special here; it is still very similar to the training datasets.

|  |  |
| --- | --- |

**Model2 : Random Forest**

After fitting an simple linear regression model, the next model to give a try is the Random Forest Model. By Fitting an Random Forest Model the result and check for feature importance. The result are below.

|  |
| --- |

By checking the feature importance. The observation are the Sold\_year accounts for approximately 16.93% of the importance in predicting the target variable. The sold\_month contributes around 20.68% to the model’s predictive power. Car\_years has an importances of about 19.66%. Mileage is the most influential feature, contributing approximately 42.73% to the model’s predictive power. Feature importances show the relative contribution of each feature to the model’s prediction, mileage appear to be the most infleunetial feature, followed by 'Sold\_month', 'car\_year', and 'Sold\_Year, suggest mileage have the most impact on predicting the target variable. Below is an chart to show feature importance.

|  |
| --- |

By checking the residual plot, the observation that the is flat and evenly spread across the range of predicted values, indicates that the variance of the residuals is consistent. This suggest that the model’s predictions have a constant level of uncertainty. The second observation are Correct Model Specification, in an ideal scenario, the residuals should not exhibit an clear pattern or trend. A perfectly horizontal line would implay that the model is correctly specified and that the relationship between the predictors and the target variable is adequately captured.

|  |
| --- |

**Model 3 : classification using clustering**

By doing regression and random forest, there are still some problems for price predictions. Since different model or make car’s tend to have different price