

Second-hand car analysis in the United State

Ruihong Sun*

Oregon State University United States

*Corresponding author: sunruihong6666@gmail.com

Keywords: Optimization; Second hand car; Suitable; Customer; Price.

Abstract: In the United States, a car is a necessity for most families. Buying a second hand car is a good choice, but not all buyers have enough knowledge to know about second hand cars. It is necessary to show them an overview of the used-car market in the US and to help them build expectations of where they stand in the used-car market based on their budget. In this paper, the heat map and 3D map and line chart are used respectively to analyze the relationship between the condition of the price of equipment and the status quo of second-hand cars, so that customers can quickly find their favorite second-hand cars in a short time. Through the display of these charts, the turnover of second-hand cars and the popular styles, types and colors are clearly displayed in front of us. The white car sells the most among all the cars. Petrol cars sell best, followed by diesels. Six-cylinder used cars are most popular with buyers. All the results in this paper May saves customers time, helps them address their immediate needs and meets their needs.

1. Introduction

In the United States, with a car, a family can travel, buy groceries, commute to work, and do other activities. Cars allow people to get to their destinations quickly and directly, saving time and comfort. Since then, people's lifestyle, life concept and quality of life have changed because of cars. Owning a car allows you to travel at any time to expand your life and broaden your horizons. Owning a car also allows you to enjoy air conditioning in a warm car instead of being exposed to rain and wind on cold winter trips. In the society, not all families can afford a new car with high financing costs, or they may be reluctant to pay too much for it [1]. Buying a second hand car is a bargain compared to buying a new one, which has led to the growth of the second hand car market in the United States to be huge. Not all buyers have enough knowledge about the show, the advantages, and disadvantages of different brands of cars [2]. In today's booming world of big data, finding an item seems to be simple, and second hand cars are no exception. With the help of big data, it is convenience to help customers select their preferred car simply and effectively, which will not only improve the efficiency of the dealership's service, but also create peace of mind and trust in the service and make the customer a potential beneficiary. [3] Therefore, it is very important to help customers choose products that are suitable for them. It has been proposed that one way to compete effectively in the 21st century is to better meet customers' needs over time than competitors by offering a diverse product line. [4] But at present, no one has paid attention to this problem, which is basically a blank field. Consumers and salesmen have not conducted conditioning and in-depth research, they focus on only one-sided demand.

How to spend the money properly and maximize the comfort of life without exceeding their affordability is a problem that every potential car owner must pay attention to. However, where they come from may be the first problem they need to think. The epidemic has greatly affected the used-car market in the US, where second hand cars outsell new ones by a ratio of nearly 2 to 1, while prices of second hand cars have also soared. According to existing report (<https://www.goodcarbadcar.net/>), 39.3 million second hand cars were sold in 2020, about 2.8 times as many as new cars. According to the U.S. Bureau of Labor Statistics' Consumer Price Index, used-car prices in the U.S. have risen a staggering 39.8% since March 2020. Over the same period, new car prices in the U.S. rose 8.9%, while the overall U.S. inflation rate was 6.3%. The main reasons may be as follows: 1. During the epidemic period, due to the shortage of automotive chips and the implementation of epidemic prevention

measures, the global automotive supply chain had serious problems, leading to the shortage of chips and other key components, resulting in the decrease of new car production and tight supply, and thus the price of second-hand car market rose. 2. Due to the strong infectivity of the virus, the public are worried about taking public transportation, so they buy their own cars one after another, and the first choice is second-hand cars. As a result, the supply of second-hand cars is in short supply, and the car price increases accordingly. Interest rates on car loans in the United States also fell to record lows during the pandemic, which also boosted people's desire to buy vehicles.

At present, the price of second-hand cars in the United States is still rising, and the United States government has adopted a series of policies to alleviate the situation. Such as opening the port of Los Angeles around the clock to address slow order deliveries and tightening ultra-loose monetary policy to slow inflation. However, judging from the current development, it will take a long time to recover the price of second-hand cars in the United States.

Most enterprises only focus on the product and fail to realize that satisfying the customer's needs and service is the basis of sales. And our research is based on the needs of customers, to realize the value exchange with customers in a real sense, to meet the needs of customers, so that customers can choose their own truly satisfied second-hand cars. We classify and sort second hand cars from sources, performance, price and other aspects, and build a search engine to help customers from the perspective of big data. As we can see from the graphs after the survey, white gas ford sedan in good condition with 6 Cylinders and clean title is the most popular type in the second hand car market. Those buyers whose budgets are less than 10000 occupies the most parts of second hand car market. However, the cars they can choose are less than the other two types of buyers. The price is usually higher for newer car with less odometer but the price rule may be different in states because the inflation rate and the type people prefer are different.

The first chapter mainly introduces the source and content of the data, and the second chapter introduces the sorting data conclusion. The third chapter introduces the current market, the fourth chapter introduces customers' needs and budgets for second-hand cars, the fifth chapter introduces how to use our engines, and the sixth chapter summarizes the above content.

2. Literature review

At present, the used-car market in the United States is very complicated and the price is very strange, which makes the market somewhat out of balance. Gavazza [6] makes a comprehensive analysis of the second-hand car market in the United States from different aspects, such as consumer welfare, purchase quality, configuration and so on. Berkovec [1] analyzes the second hand car market by a discrete choice model of consumer demand for cars and car production in combination with second hand car scrapping the simple model of development for the automobile market's short-term general equilibrium model, gave the analysis of the used-car market new point of view, give us more new inspiration.

Odometer fraud is also a problem in the second hand car market, with 3.5 percent of all vehicles having their odometers returned, which translates to 450,000 cases of odometer fraud in the United States each year. Direct costs to U.S. customers are estimated at \$1 billion. In the investigation of second-hand car market, data falsification, false propaganda and other chaotic phenomena occur frequently, which are the problems that consumers who buy second-hand cars need to face. (Montag, 2022) Many dealers have come to believe that new cars are worth more, and second hand cars are less profitable, mainly in the transaction process. From the point of view of cost, the main source of profit still comes from the difference between sales and imports. Manufacturers are also starting to consider the key role second hand car management plays in boosting salvage value, new car sales and manufacturer brand equity.(Duvan,2022) The life trend and scrappage behavior of vehicles that the second-hand car market is facing are also very important issues. Emerging technologies such as the Internet of Things and big data can be actively introduced to build an information interaction system for the whole life cycle of automobiles. Establish sales ledger of scrapped auto parts, so that the source can be checked and the destination can be traced. (Bento, 2022)

For the analysis of second-hand car market, linear regression model can also be used to fit the data through brand, age, title state, primary color, state, remaining time and mileage and other factors, so as to predict the price according to the selected factors, using simple linear model, stepwise function and standardization technology, random forest and neural network. (Lin, 2022) The structural empirical model can also be used to estimate the seller's profit under published price and bargaining, and to study how market conditions affect the seller's optimal pricing format. This model integrates a simple bargaining mechanism into the standard random coefficient discrete selection model. By extension, the product-level demand system uses only list price data for estimation, and also restores unobserved price discounts in the estimation. (Huang, 2022)

However, these studies cannot show the situation of second-hand car market intuitively, and are not applicable to ordinary consumers, who need a simple search engine to help them quickly locate the car with the highest cost performance. In this regard, my research results help consumers to find the car they need in the shortest time, and achieve this through detailed and credible data.

3. Date and analysis

In this paper, we collect our data on the second hand car from Kaggle (<https://www.kaggle.com/>) in which data about second hand cars in every state in the United States can be found. Within this data, some of the key data was necessary for us to conduct research and presentations while some was irrelevant.

The data was taken from 1920 to 2020, the original file have 14 columns and 426853 rows. After cleaning useless columns and empty rows, there are 12 columns including eight factors that are condition, Paint color, Cylinders, Fuel, Manufacturer, Title status, Type, and State. Otherwise, there are 124946 rows (cars) left in the clean data. We focus on the authenticity and validity of the data in the tables in the selection data, and some problematic data will not be considered.

Based on these bar plots and pie charts (Figure 1-4), it can be clearly seen that cars with high quality sell better. Otherwise, those cars with white, black and silver colors sell best among all cars. Secondly, buyers also consider the number of cylinders when buying a car, 6 cylinders is the best since. Cars that run on petrol sell best, followed by diesel. Hybrid and electric cars only occupy a little percentage. As for the brand, Japanese cars and American cars divide the market. Ford, Chevrolet, Toyota are the top three brands for buyers. Otherwise, California buys the most cars, followed by Florida, which is also a state with a large population, so car sales are also high. In the meanwhile, buyers prefer SUV, Sedan and pickup as their goal since these types are more useful for families.

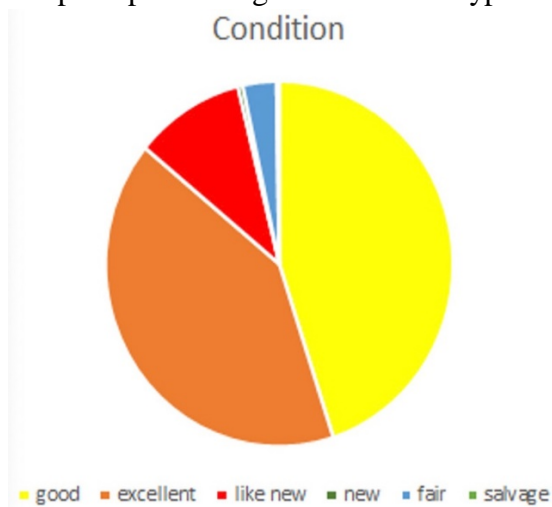


Figure 1. Different condition levels For second hand cars.

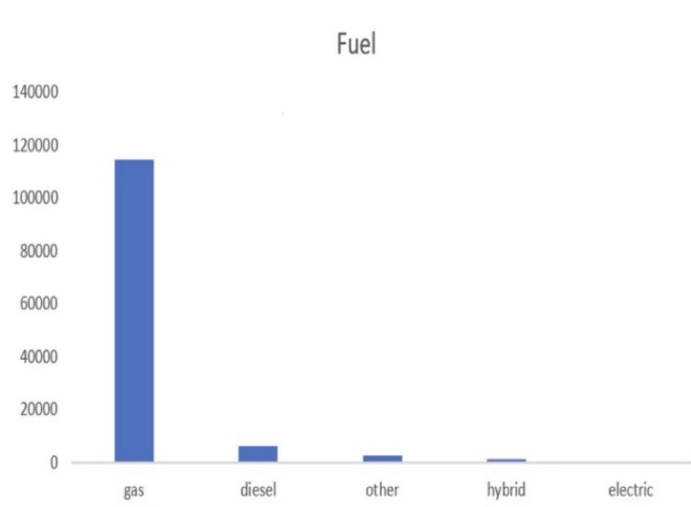


Figure 2. Energy use of second hand cars.

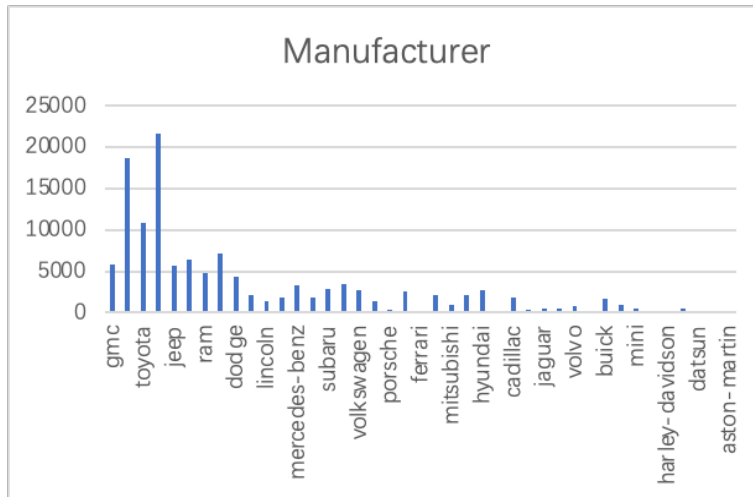


Figure 3. Manufacturer use of second hand cars.

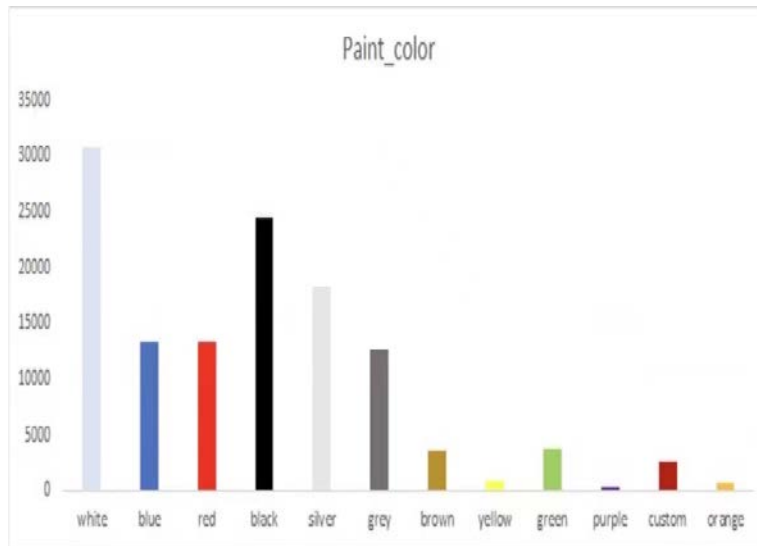


Figure 4. Colors used in second hand cars.

The construction of the car finder is based on these eight factors, assuming that the factors can largely affect the preferences of consumers when they are choosing cars. Therefore, through the car finder engine, car buyers will be able to obtain the information of vehicles based on different states they come from and different budgets they have. By doing so, car buyers can quickly narrow down the database and pay attention to the information of vehicles which potentially fits in to their appetite.

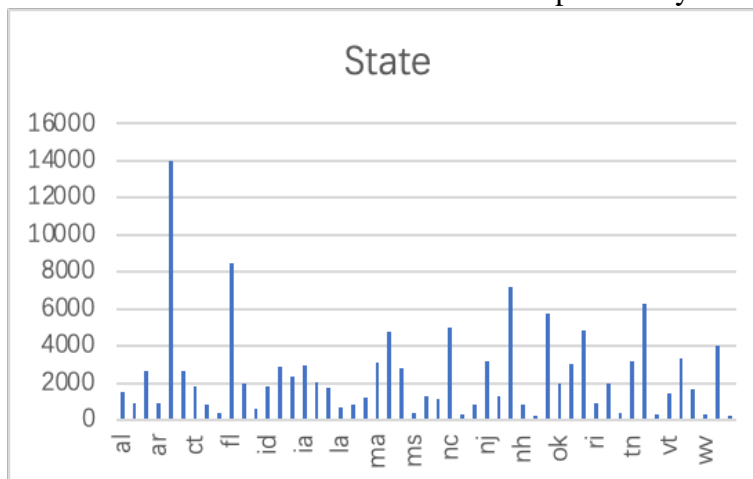


Figure 5. State distribution of second hand cars.

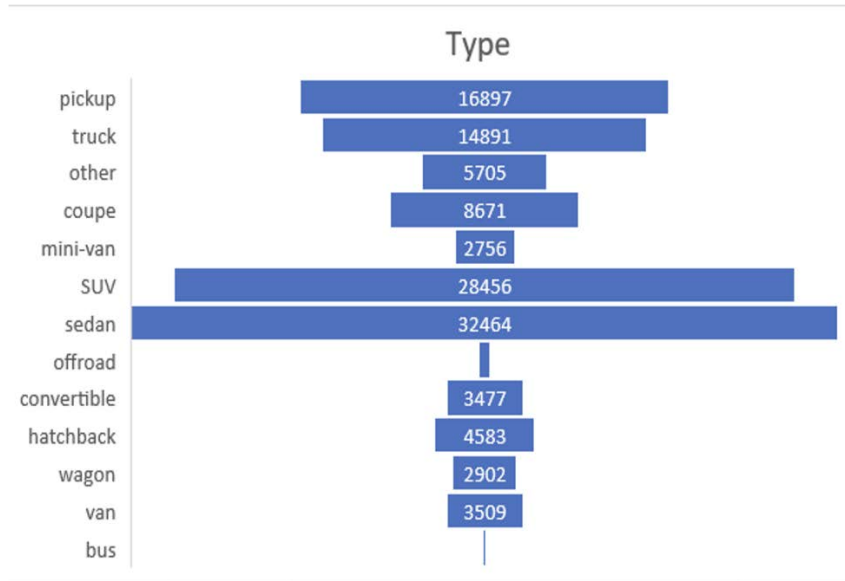


Figure 6. Types of second hand cars.

Data can be viewed from top, top, and side views to avoid overstretching. According to this green bar chart, there are three types of buyers. A rational car buyer will think twice before buying a car. It's a question many consumers will consider. Buying a car is not a small expense for most people. How to spend money reasonably and maximize the comfort of life under the premise of not exceeding their own affordability is a problem that every potential car owner must pay attention to. Therefore, our data analysis is particularly meaningful and well suited to the concerns of every car buyer. By discovering the data and car finder, also can be seen in the green-bar chart, the used-car buyers can be categorized into three types depending on their budget. The first type of buyer is people with sufficient budgets, namely higher than \$30000 expense. Thus, they can get whatever kinds of models they like. The second type of buyers are people with relatively sufficient budgets, namely between \$10000 to \$30000 expense. Thus, they can choose to buy some of the premium models but with some limited. The last type of buyers are people with relatively insufficient budgets, namely under \$10000 expense. Thus, they would have been strictly limited when choosing cars



Figure 7. The price and budget of a second hand car buyer

After processing and refining the data, import the data and create a Heat Map and 3D Plot (Figure 8-11) which allowed people to visualize and analyze our data. Heat map is a data visualization

technique that shows the magnitude of a phenomenon as color in two dimensions. Use New York and California as two examples for the analysis. Because they are representative states with large populations. According to the heat maps of two different states, the price in New York is positively correlated with the year, but the data in California is not, because the inflation in California leads to a non-positive correlation between car price and year. Moreover, from the 3D Plot, it is more visually expressive to see the situation of the car that fits the buyer. Readers can observe the data from the angle of face up, overlooking and side view to avoid overstretching.

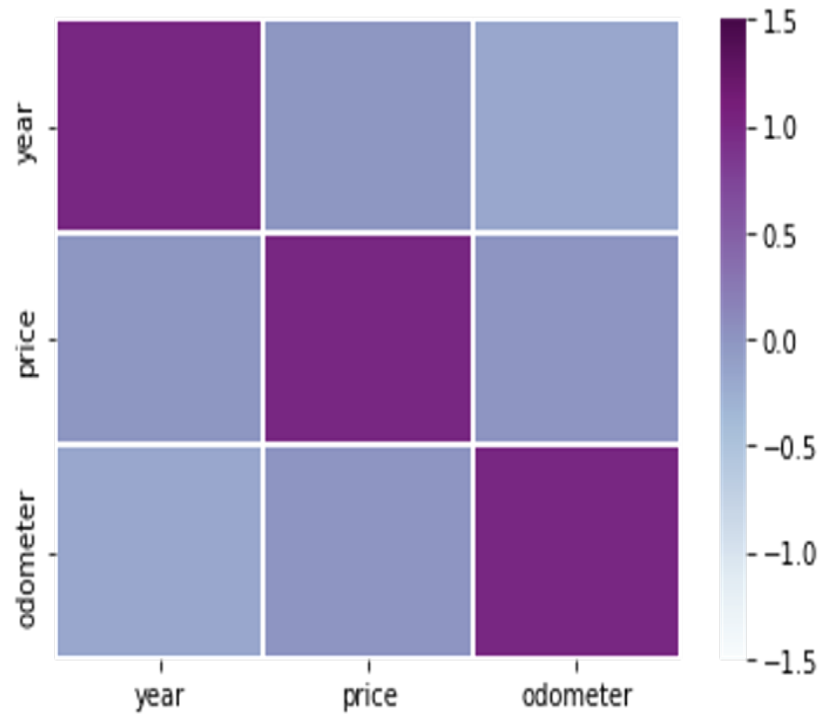


Figure 8. Heat map distribution of second hand cars in CA.

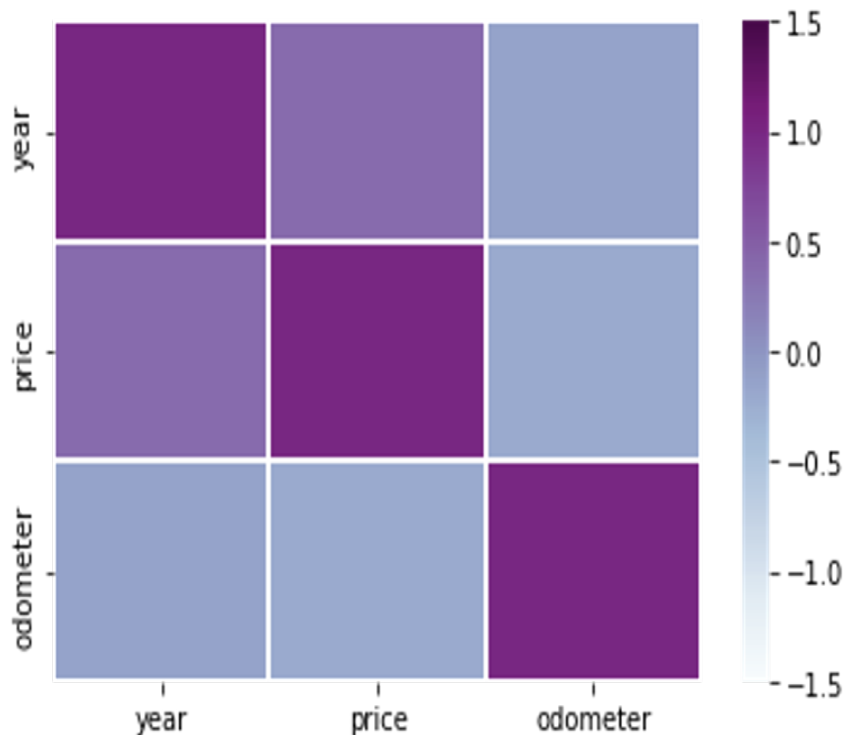


Figure 9. Heat map distribution of second hand cars in NY.

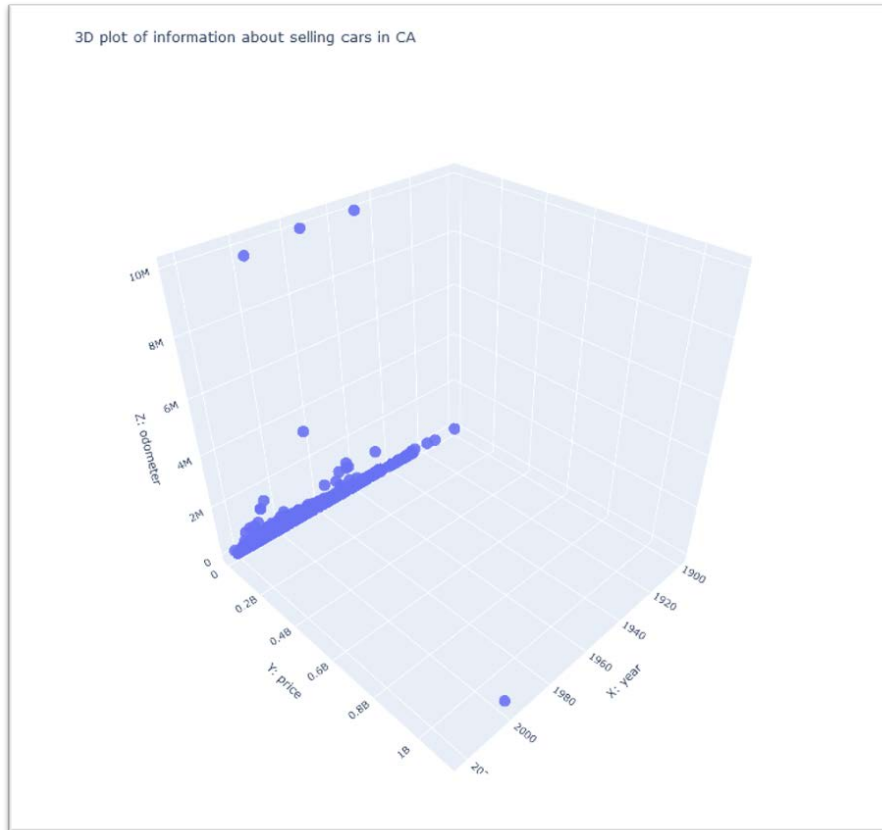


Figure 10. 3D plot of information about cars in CA.

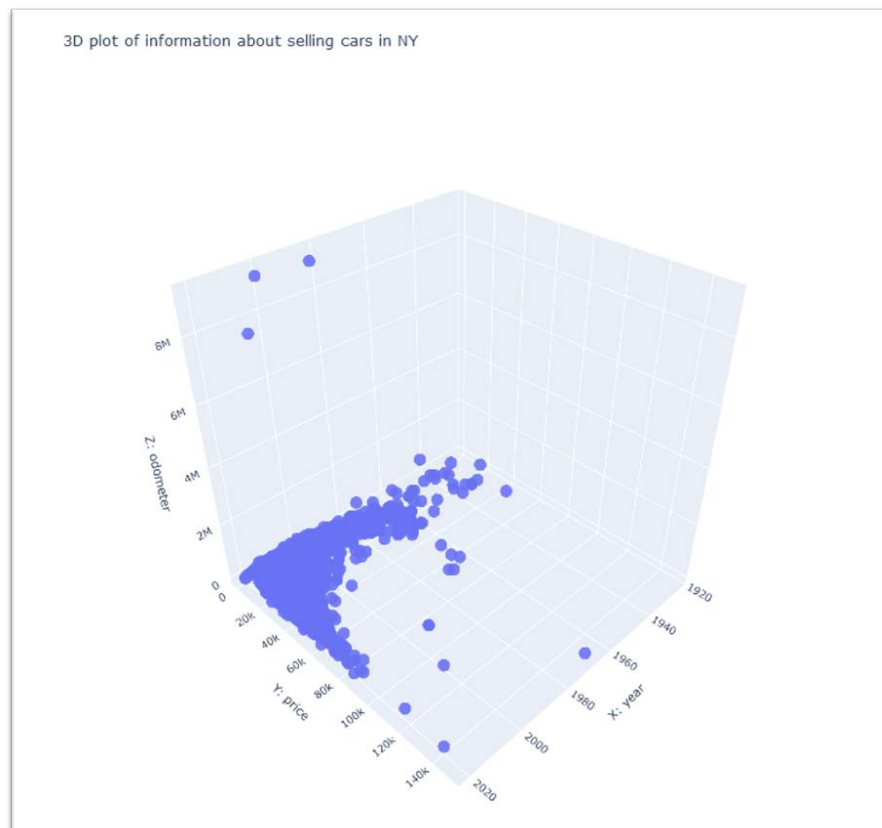


Figure 11. 3D plot of information about cars in NY.

Once all the raw data was collected, the rest of the work was to visualize the data in Python in order to vividly represent what the data told us. First, import the collected and finely filtered data into Colab,

which contains the year, make, model, color and mileage of vehicles in each state of the United States. After importing, we generated a Heat Map by code, we engendered a 3D plot by the code, which was show the position of the three sub-data in the 3D plot to analyze the connection and correlation among them. Then, as long as our consumers input their own state or the name of the state they want to buy a car in the search engine, they can display all the eligible second-hand cars, so as to make a better choice, saving consumers' time cost and energy, but also provide a high cost performance of second-hand cars.

4. Conclusion

Through the analysis of second-hand cars, this paper starts to conduct research in a new field with limited research and makes a research beneficial to consumers and people. The analysis results in this paper can be summarized as follows. All aspects of used car performance are concerned by consumers.

From the chart, it can be clearly seen that cars with high quality sell better, and white cars sell best among all cars. Buyers also consider the number of cylinders when buying a car, 6 cylinders is the best since. Cars that run on petrol sell best, followed by diesel. California buys the most cars, followed by Florida, which is also a state with a large population, so car sales are also high. According to the heat maps of the two states, it can be clearly seen that the price in New York is positively correlated with the year, but the data in California is not, because the inflation in California leads to a non-positive correlation between car price and year. Choosing a car in a different state is also part of the car purchase process. From the 3D Plot, we can also feel more visually expressive and clearly see the situation of the car that fits the buyer.

Our data is based on a sample of vehicles sold by Kaggle to introduce the domestic used-car market in the US. Relationship between odometer, production date and price based on heat and 3D map showing in specific states (such as CA and NY). Analyze prices and divide buyers according to their budget. In addition, through search, we can tell buyers how many cars they can choose from our sample and prove the analysis results to help buyers find their favorite car in different situations.

References

- [1] Berkovec, J. New Car Sales and Second hand car Stocks: A Model of the, Automobile Market. *The RAND Journal of Economics*, 16(2), 195–214. <https://doi.org/10.2307/2555410>
- [2] Does mass customization pay? An economic approach to evaluate customer integration. Taylor & Francis. (n.d.). Retrieved December 18, 2021, from <https://www.tandfonline.com/doi/abs/10.1080/0953728042000238773>
- [3] Kahn, B. E. Dynamic relationships with customers: High-variety strategies - Barbara E. Kahn, 1998. *SAGE Journals*. Retrieved November 27, 2021, from <https://journals.sagepub.com/doi/abs/10.1177/0092070398261005>.
- [4] U.S. Bureau of Labor Statistics. CPI Home. U.S. Bureau of Labor Statistics. Retrieved November 27, 2021, from <https://www.bls.gov/cpi/>.
- [5] Turrentine, T. S., & Kurani, K. S. Car buyers and fuel economy? *Energy Policy*. Retrieved December 18, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S0301421506001200?casa_token=1z1ZjQCAMUEAAAAA%3AxaHYFevAdgBf4xxN3V-vveeBE7hDrkFxFgILPQU8EVn6uIu_AEI2LrTnvqhn1Go34xkrQo
- [6] Gavazza, A., Lizzeri, A., & Roketskiy, N. A quantitative analysis of the used-car market . *American Economic Review*. Retrieved December 18, 2021, from <https://www.aeaweb.org/articles?id=10.1257%2Faer.104.11.3668>

- [7] Montag, J. Identifying odometer fraud in second hand car market data. *Transport Policy*. Retrieved January 1, 2022, from https://www.sciencedirect.com/science/article/pii/S0967070X16308083?casa_token=WFHJ_izrn_IAAAAA%3Ausd2E3zOJ5uii9NnNxAttPLSnir0pNb0XIYhBUXWn7mCRz1PK34YmfumVQTmmo659MG7NU4
- [8] Duvan, B. S., & Ozturkcan, S. Second hand car remarketing. *SSRN*. Retrieved January 1, 2022, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1435184
- [9] Bento, A., Roth, K., & Zuo, Y. Vehicle lifetime trends and scrappage behavior in the U.S. second hand car market. *The Energy Journal*. Retrieved January 1, 2022, from <https://www.iaee.org/energyjournal/article/3032>
- [10] Lin, A., Marlow, B., & Dixon, A. USA used car price prediction using machine learning. *Murray State's Digital Commons*. Retrieved January 15, 2022, from <https://digitalcommons.murraystate.edu/postersatthecapitol/2022/EKU/10/>
- [11] Huang, G. When to haggle, when to hold firm? Lessons from the used-car retail market. *Wiley Online Library*. Retrieved January 15, 2022, from <https://onlinelibrary.wiley.com/doi/abs/10.1111/jems.12385>