

|  |
| --- |
| Carvana: Is Bad Buy? |
|  |
| JANUARY 2022  SCIT | ITBM | DIV D  Data Visualization and Modelling  Authored by: Suraj Kumar Shrivastava |

..

|  |
| --- |
| Carvana: To Buy or Not to Buy?**1. Discuss the role of visual analytics in data analytics?** The use of sophisticated tools and procedures to evaluate datasets using visual representations of the data is known as visual analytics. Users may find patterns and produce actionable insights by visualizing data in graphs, charts, and maps. These insights assist businesses in making better data-driven decisions.  Visual Analytics creates host of advantage in field of data analytics by:   * Improving data exploration and analysis while lowering total costs. * Improving data understanding and speed for better decision-making. * Helping consumption of larger amounts of data in less time, resulting in increased operational efficiency * Early detection of previously unnoticed trends, anomalies, and correlations between data sets, which could lead to a competitive advantage * Real-time updates and instant feedback keep data current and correct. * It assists organizations in resolving relevant challenges thanks to significant advancements in computing and data storage. * It has the ability to solve huge and difficult problems, resulting in more accurate findings and more profitable business decisions; * It has a variety of visualization trends, ensuring that data is presented in a comprehensible manner.  **2. What do you think about Carvana’s mission, business model and achievements?** "To alter the way people buy vehicles," is Carvana's business objective. Carvana was created in Phoenix, Arizona, in 2012. Carvana successfully launched its service and sold its first vehicle in January 2013. The American automobile industry was massive and extremely segregated. The used-car market in the United States boomed in the mid-2010s, earning over $710 billion in yearly sales, with 38 million automobiles sold at an average price of $18,552 in 2015. Instead of franchises, private auto owners sold roughly 80% of used cars in 2016. In North America, the top 100 used-car dealerships accounted for only about 7% of the market. Carvana's mission addressed all of the issues that most dealers had when selling cars. The majority of these dealerships were brick-and-mortar operations, with customers coming in to look at restricted automobile inventories.  These car dealerships lacked the capacity and expertise to frequently purchase high-quality vehicles. According to a Gallup poll from 2015, over 80% of North Americans who bought used automobiles disliked the transaction, and they had little faith in car salespeople. Carvana was able to target these major pain points thanks to its well-defined vision. This vision also helped Carvana disrupt the $1 trillion per year US vehicle market, as customers were excited about the prospect of buying a car with Carvana's online business model's simplified sales, financing, and other processes.  In comparison to prior used automobile retailers, Carvana's business approach was unique. Carvana was a used automobile web marketplace that allowed users to research and select vehicles, view 360-degree photos of them, apply for finance and warranty coverage, and finish the transaction. Vehicles could thereafter be delivered or collected from one of a dozen automobile vending machines. The habit of buying online was only beginning to take off at the time. Because of the ease, the majority of Americans choose to shop online. Carvana's business model addressed the trend of online shopping, which accounted for 8.4% of all US retail sales in the third quarter of 2016. This business model was designed to appeal to the three-quarters of US vehicle buyers who claimed they would prefer to buy a car primarily online if given the option. Carvana's 100% online process business model helped the customer acquire a lower price because it eliminated the need for middlemen. Because there were no commissions, no pressure, and no hidden fees, this internet business model made the process more open.  The aim and business model of Carvana aided in achieving greater success. Carvana didn't just want to be a vehicle salesman; instead, he wanted to improve the client experience. On all automobiles, Carvana offered a seven-day money-back guarantee. Carvana provided one-day delivery for the convenience of its consumers. Consumers can improve their car-buying experience by using Carvana's fully automated vending machines. Customers who lived more than 100 miles away from a vending machine were eligible for a $200 travel subsidy, as well as transportation from Carvana. Carvana's customer-centric approach became the primary differentiator between it and other web-based used car vendors later on. **3. Would you invest in Carvana?** Yes, we would invest in Carvana because it is disruptive and ahead of its time; they are using machine learning-based algorithms to minimize faults and enhance profits, indicating that they are more focused on innovation and optimization. This exemplifies why it's critical for businesses to use data analytics to keep up with data science-related technologies like AI and ML for a more efficient manner of increasing earnings.  The following are some of the highlights of the Carvana stocks:   * According to case study data, the company had targeted 60 (and increasing) major regions across the United States by 2017, and had a nationwide pooled inventory of almost 10,000 used vehicles. * Carvana's goal was to disrupt the $1 trillion-a-year US automobile business. Carvana fought in a market that was both huge and fragmented. The automotive business in the United States accounted for over 20% of the US retail economy in 2015, with sales of $1.1 trillion. * The industry's sales increased by roughly 5% each year from 2003 to 2017, suggesting rising disposable income and greater credit access for consumers. * A sizable addressable market - Used car transactions in the United States totaled almost 40 million last year, with revenues exceeding $700 billion. That's the largest single retail sector in the United States, so the company has a lot of room to grow in the long run if it can keep expanding its reach. * The used automobile industry is highly fragmented, owing to the lack of economies of scale and the fact that no single company can dominate it. Carvana stock may appear pricey based on standard valuation metrics (shares sell at nine times trailing 12-month sales with no profits to report on the bottom line), but given its tremendous potential runway, it should fetch a premium. * Carvana's ability to upend this almost $1 trillion industry and continue to beat the market should not be underestimated by investors. * Furthermore, if we look at the present stock price, we can see that it is a good lucrative price. The price was 11 dollars when the IPO began, and it is now 261 dollars.  **4. Have Carvana’s algorithms been effective in predicting which of its cars are and are not “kicks”?** Carvana held a "Don't Get Kicked" competition on Kaggle to learn more about how effective their business is at obtaining used cars at auctions and then selling the same used cars to customers with higher customer satisfaction. The dataset for the competition consisted of 72983 records with 34 attributes. We created numerous models in order to generate the best/optimize algorithm for reliably predicting good/kick cars.  Carvana implemented a classification algorithm. It's the process of recognizing, analyzing, and organizing concepts and objects into predetermined "sub-populations." Simply said, machine learning classification algorithms use input training data to predict whether following data will fall into one of the established categories. Carvana utilized a classification algorithm to determine whether a car was defective based on the data available, and cars were then divided into two groups: Good Buy and Bad Buy.  The steps we followed were as follows:  a) Data Collection  b) Data Preparation  c) Choosing a model  d) Training  e) Evaluation and parameter tuning  f) Prediction  According to the data analysis, the dependent variable ‘IsBadBuy’ has significantly imbalanced values, with 88% good buys and only 12% bad buys. The dataset has features that are categorical, numeric and date type.  We created seven different types of algorithms to find the best prediction model. The following are the outcomes of each algorithm:    We don't have access to Carvana's algorithm and can't verify its exact accuracy to determine efficiency, but we do have an algorithm that works in a similar vein and uses the same database as Carvana's, and it has an efficiency of 89.4 %. As a result of the foregoing, we can conclude that boosting algorithms (i.e. Decision Trees and Gradient Boosted Trees) are the best for predicting risky cars purchased at auctions.  Following the above processes, we can conclude that 88 % of Carvana's cars are good buys, and that predicting the same with different models yielded an accuracy of around 88 %, implying that Carvana's algorithms are effective in predicting which of its cars are ‘kicks’ and which are ‘not kicks’. **5. What is visual analytics and how can it be applied to Carvana’s data set in Kaggle? Explore the Hans Rosling video, “Let my dataset change your mindset.” TED@State, 19:41, June 2009, Hans Rosling: Let my dataset change your mindset | TED Talk. What makes his visualization so effective?** The science of analytical reasoning supplemented by interactive visual interfaces is known as visual analytics. Data is being generated at an astonishing rate today, and the ability to collect and store it is growing faster than the ability to analyze it. A wide number of automatic data analysis methods have been developed over the previous few decades. The complexity of many situations, on the other hand, necessitates the inclusion of human intelligence early in the data analysis process. Visual analytics methods enable decision makers to acquire insight into complicated situations by combining their human flexibility, creativity, and background knowledge with the massive storage and processing capacities of today's computers. Humans can directly interact with today's computer's data analysis skills via enhanced visual interfaces, allowing them to make well-informed judgments in complex situations.  Visual analysis is a holistic technique that combines visualization, human factors, and data analysis. The domains of visual analytics research are depicted in the diagram. Human elements, including cognition and perception, play a significant role in human-computer communication and decision-making, in addition to perception and data analysis. Visual analysis encompasses the domains of information visualization and computer graphics, as well as data analysis approaches in the fields of data recovery, information management and representation, and information management and knowledge recognition, as well as Data Mining.  Hans Rosling's video "Let my dataset alter your thinking" explains research done over the last 100 years on the relationship between health and prosperity in many countries and locations. Rosling discusses how other parts of the world are misunderstood. He illustrates his point by displaying data sets on countries and GDP. The majority of graphs have a higher magnitude in locations such as Europe and North America. Africa was one of the poorest continents in the world.  The Western World has been described by Rosling's students as a lengthy existence inside a small family and a developing economy. This means that people with large families are less likely to survive owing to a lack of resources, and vice versa. They have received these assumptions and beliefs about the world's worldview from their professors and the year they were born.  This is a significant difference when Rosling separates Regional Bubbles into smaller, country-like bubbles on his charts, because regional bubbles are higher in children's survival and Africa is the lowest. When the bubbles are divided into countries, Africa's percentage will drop even lower than the previous average of 80%. The African countries' actual baseline has dropped to 70%. (That being Sierra Leone).  In global health, Rosling is a difficulty since affluent countries have greater rates of health and survival. The Rosling data set demonstrates that poorer regions/countries eventually become rich nations as the world develops. Rosling refers to this as "development" convergence because it is a convergence of equal health and prosperity, rather than a world rejuvenation.  As Rosling demonstrates, the more relevant research we conduct, the more we learn about our generalizations. The more we learn about it, the better. Rosling's speech is upbeat because he explains how the rate of convergence has changed dramatically over the last 100 years. We shall be able to attain total integration in health and survival into a precisely identical globe if the 'developed' countries assist the 'developing' nations. I agree with Roslings' findings since we are not establishing, but rather converging in the wrong direction, particularly as Americans. **6. Use visual analytics tools to find 5 different hidden information in the dataset which will change business perspective of the company?** Following are the 5 different hidden information in the dataset which might change business prospective of the company, that we obtained by analysing the dataset in Tableau.  1. Vehicles from American manufacturers have the highest bad buys.    2. Bad buys are not restricted to a single vehicle size.    3. Older vehicles have a higher probability of bad buys.    4. Ford manufactured vehicles have the highest bad buys.    5. A Steady increase in bad buys till the year 2005 after that there was a gradual decline.   **7. The Story** We analyzed Good buys vs Bad buys based on “*Auction Provider at which the vehicle was purchased*”. We found that Manheim has highest count i.e. 4715 of bad buys and other two categories i.e. “*Adesa and others*” are less than 50% of Manheim. We narrowed our analysis to Bad Buys based on the “*Nationality of Car Manufacture”* and “*Auction Provider at which the vehicle was purchased”.* We observed that the maximum count of Bad Buys Vehicles are American.  We have then analyzed those Good Buys vs Bad Buys based on the vehicle size. Here we observed the distribution of Bad Buys across all size of vehicle. In addition, bad buy is not restricted single segment.  We analyzed the Bad Buys vs Good Buys based on the Vehicle Age. Here we observed that higher the age of vehicle, the more is the probability of Bad Buys.  We analyzed the Bad Buys trend across all the years, here we observed that till 2005 the count of Bad Buys are increasing and right from 2005 it started decreasing. The highest count is 418 in the year 2005.  Here we divided the Bad Buys in three segment: (i) Before 2005 (ii) 2005, and (iii) After 2005. We also observed the Bad Buys is happening in both Online and Offline mode and we have also observed that the Bad Buys after 2005 is 48% of total Bad Buys.    We have already analyzed that Chevrolet, Dodge, Ford and Chrysler are the brand which have count of bad buys greater than 100. Now when we segregated the count in 3 segments (i) Before 2005 (ii) 2005, and (iii) After 2005 and we have found that the count of Bad Buys had comparatively decreased for all the brands ‘After 2005’ as compared to ‘Before 2005’ and ‘During 2005’.    When we analyzed the “Acquisition [Auction vs Retail] Price analysis” graph we observed that ‘Acquisition Retail Average Price’ and ‘Acquisition Auction Average Price’ are much higher for the category ‘After 2005’. This is the reason why we believe that the count for the bad buys has decreased.  When we analyzed the “Current [Auction vs Retail] Price analysis” graph we observed that ‘Current Retail Average Price’ and ‘Current Auction Average Price’ are much higher for the category ‘After 2005’. This is the reason why we believe that the count for the bad buys has decreased.    Now we did ‘Vehicle Acquisition Cost Analysis’. It means that Acquisition cost paid for the vehicle at time of purchase for all the categories i.e. (i) Before 2005 (ii) After 2005. The unit price for vehicles before 2005 was approx. $5666 but the unit price for the vehicles after 2005 was approx. $7024. Now this also supports why the count for the bad buys has decreased after 2005.    We did the analysis of the bad buys according to the state wise distribution and found that bad buy vehicle count was highest for the state of Texas and percentage of vehicle acquisition cost was around 20%, which means that 20% of the total vehicles sold were bad buys for the state of Texas for instance.  Drive folder for the “TABLEAU report”, click to download the files: <https://drive.google.com/drive/folders/1oeiscyJIhQ9AUqpnqR-KRfFp2pyXEFQQ?usp=sharing> |