# IST 652 - Scripting for Data Analysis

Kelsey Kirby, Sean Deery, Mark Stiles

May 23, 2023

Final Project: Honda Data

# General

The dataset we selected was about Honda car sales from Cars.com. This is a large consumer website containing a lot of specific information about many Honda models. It offers a rich source of information to analyze what similarities or differences there are between models as well as compare features, pricing, and fuel efficiency.

## Questions

The six questions that this report will attempt to answer are:

1. What car components are the key drivers of price points?
2. What states have the highest or lowest cost?
3. How has MPG changed over the years per model?
4. Which models have the best ratings?
5. What are the best deals for different budgets ($10,000, $20,000, ~ $60,000)?
6. Which models were sold for the most years?

# Data Analysis

The data used in this analysis was sourced from [Kaggle](https://www.kaggle.com/datasets/omartorres25/honda-data) for this report. The main file contained 4,999 rows and 25 columns.

## Cleanup

The dataset had several columns that weren’t lined up properly and needed to be shifted over. This was done through scripting. We also had missing transmission data. There was a significant number, but there were so many configurations that scripting was prohibitive. The data was cleaned manually in Excel by finding similar configurations that did have the transmission data available.

The data needed some fine-tuning the following changes were also made:

* set years column as int
* convert all words to null in mileage and convert the column to int
* removed the text and symbols from price and converted column to int
* fix misnamed states or set them to null
* align the drivetrain options
* removing dashes from transmission values

## New Variables

There were several variables created to help create better subsets and to extract information from dense textual values like the engine information. Here’s what was created:

* Model Category - cars grouped by model, ignoring specific trim and body styles
* Average MPG - the MPG was a range that was split and then averaged
* Exterior Color Category - colors were a dazzling array of names converted to a basic color set
* Transmission Category - the transmissions were trimmed to a
* Cylinder Size - used regex to extract the cylinder volume from the engine details
* Cylinder Count - used regex to extract the cylinder count from the engine details
* Cylinder Valves - used regex to extract the valve count from the engine details
* Cams - used regex to extract DOHC or SOHC from the engine details and converted to 0/1
* Injection Type - used regex to extract GDI or MPFI from the engine details and converted to 0/1
* Hybrid - used regex to extract Hybrid from the engine details and converted to 0/1
* Turbo - used regex to extract Turbo from the engine details and converted to 0/1

## Missing or Null Data

There was missing data throughout the dataset. For most variables, we consider that the data wasn’t missing significant volumes of data, and we compensated by filtering out the null values. There was enough remaining data to produce quality comparisons.

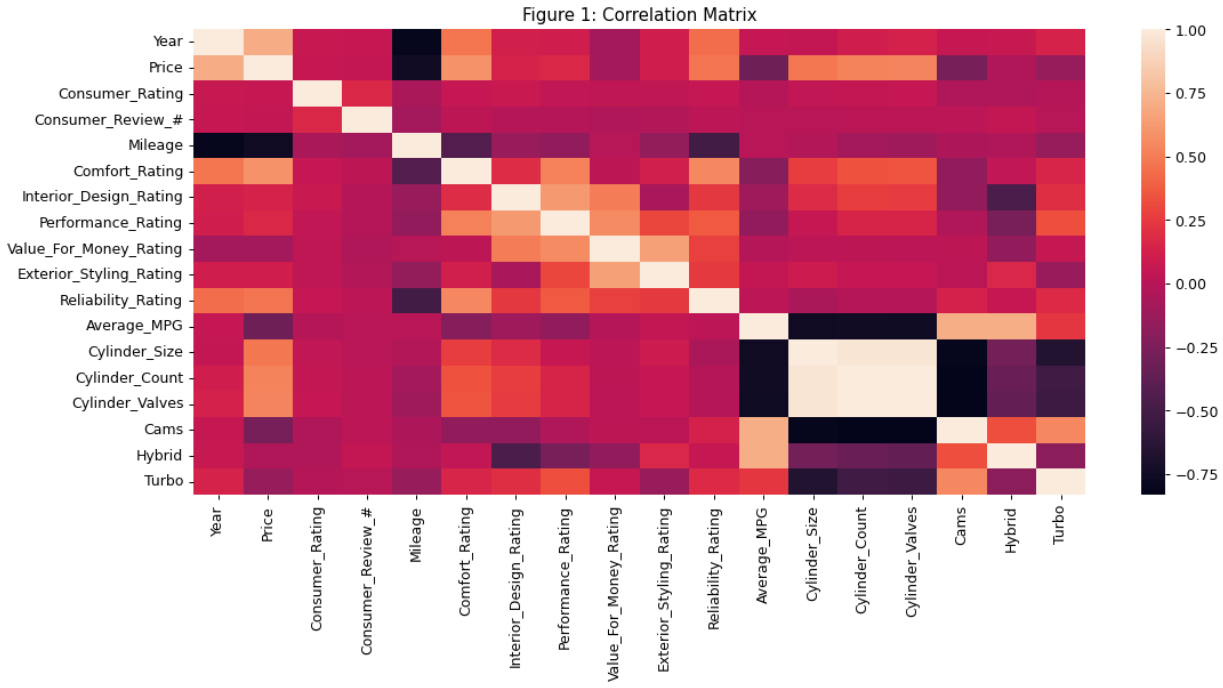
## Program

The program begins by loading the CSV data files, defining new functions, cleaning up column data, and then creating or extracting new variables. Then it begins each question with distribution information and further plots to extract additional information. It covers all six questions posed in the opening.

# Results

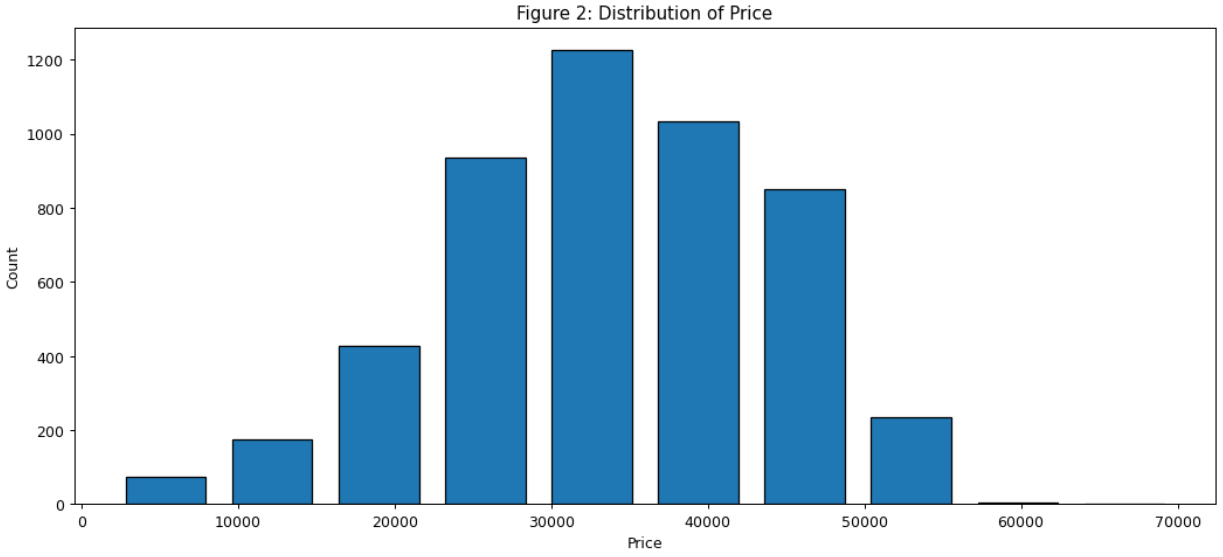
## Exploratory Data Analysis

The initial exploratory analysis showed some areas of interest in the numerical fields. Using a correlation matrix, Figure 1 shows there were correlations between mileage, year, and price, which is not surprising. There’s also a negative relationship between miles per gallon and larger engine features, as well as a positive one between miles per gallon and hybrid. One of the most interesting features was between value for the money rating and exterior styling rating or the negative relationship between mileage and reliability rating. They followed reasonable logic but were interesting.



## Question 1: What car components are the key drivers of price points?

To start exploring what drives the price point, the analysis started with the distribution of the price itself. Figure 2 shows a mean near $35,000 with a low in the single thousands up to just over $60,000. It’s normally distributed and gives some shape to the range of new and used Honda car values.



The next step in the analysis explored the engine components to see how well the parts are represented and their effect on price. Figure 3.1 shows a varied distribution of sizes with a notable absence around 3.0. This must be where the compromise between efficiency and power meets. For smaller engines, efficiency is a priority. For larger engines, power becomes more important.

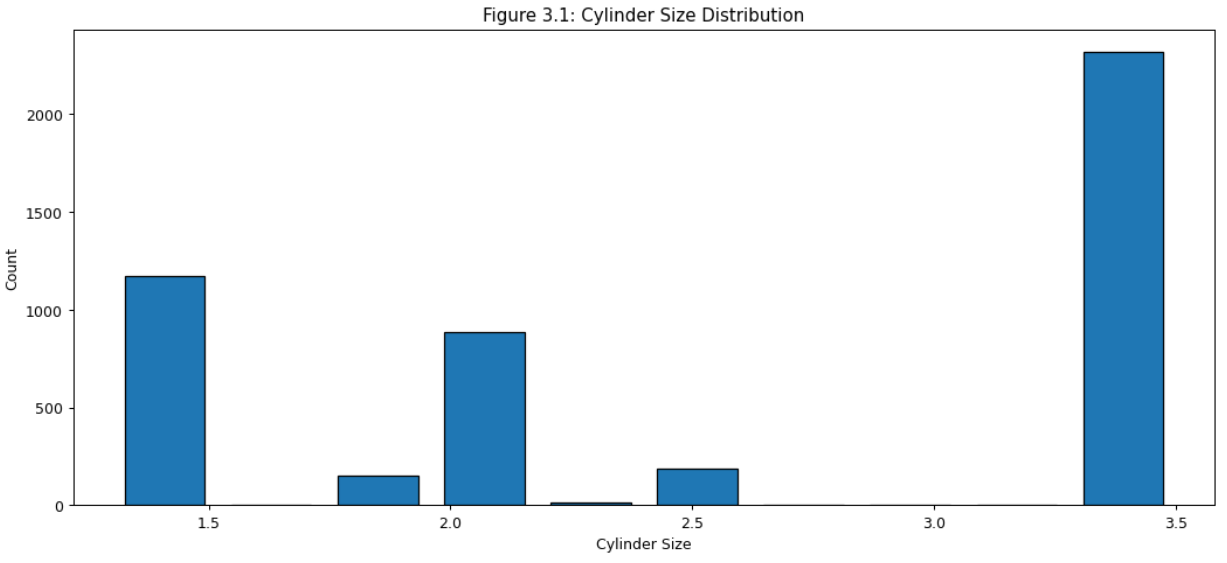


Figure 3.2 shows that the value is highest for 2.0 or 3.5-liter engines. This most likely reflects the engine meeting the demands of a consumer willing to pay for that specific benchmark.



Figure 3.3 shows an even split among the data between the 4 and 6-cylinder engines. Both are common engine configurations.

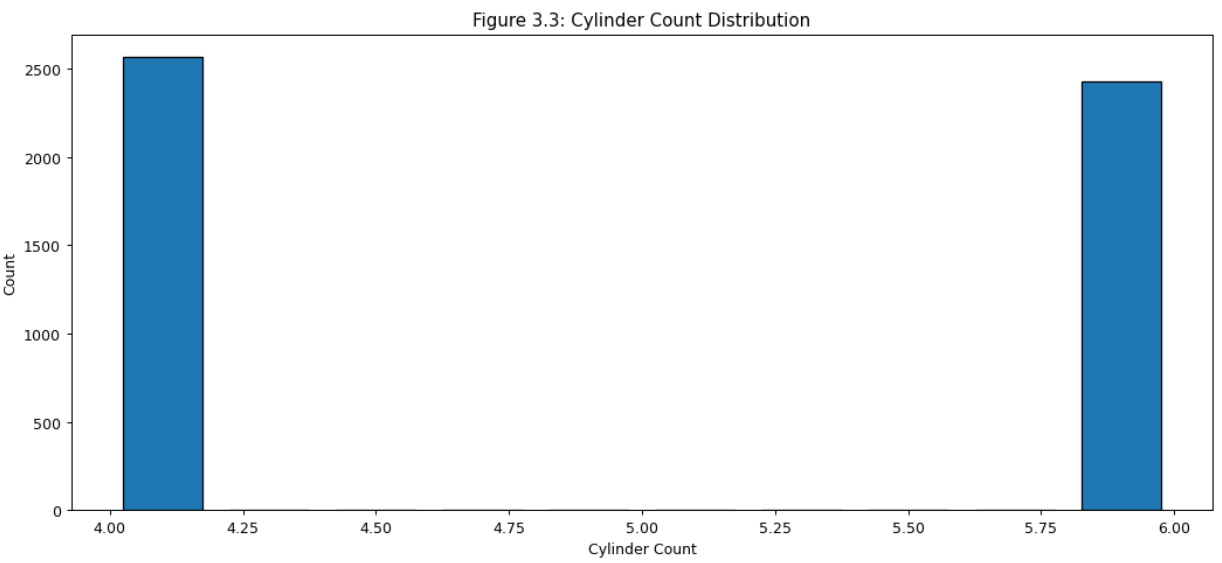


Figure 3.4 shows that the 4-cylinder car prices are slightly less than cars with 6-cylinders ,but there are a few 4-cylinder cars drawing in higher prices.

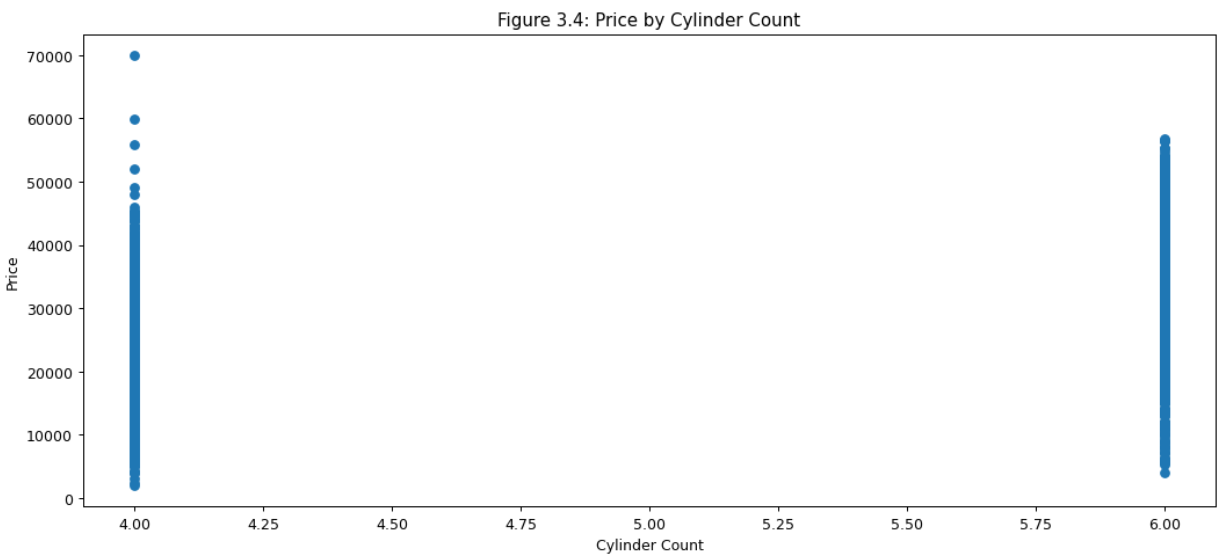
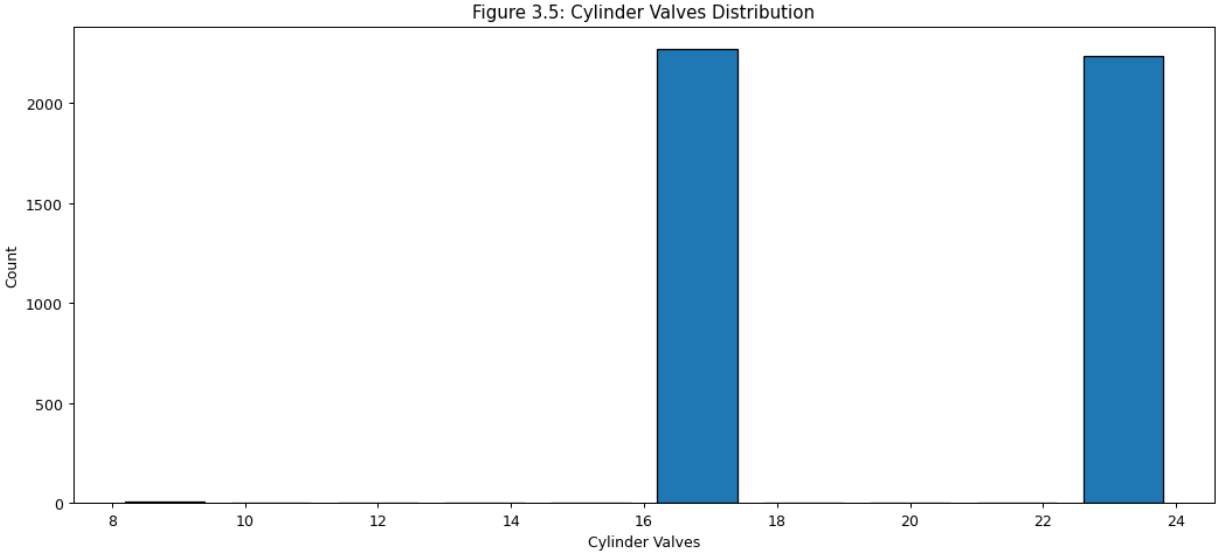


Figure 3.5 shows there’s a relatively equal split between 16 and 32-valve engines. There are almost no 8-cylinder engines represented in the dataset.



The 16-valve engines draw a predominantly lower price with a few outliers higher than the 24-cylinder engines but appear to be a correlation between higher price and more valves by around $10,000.

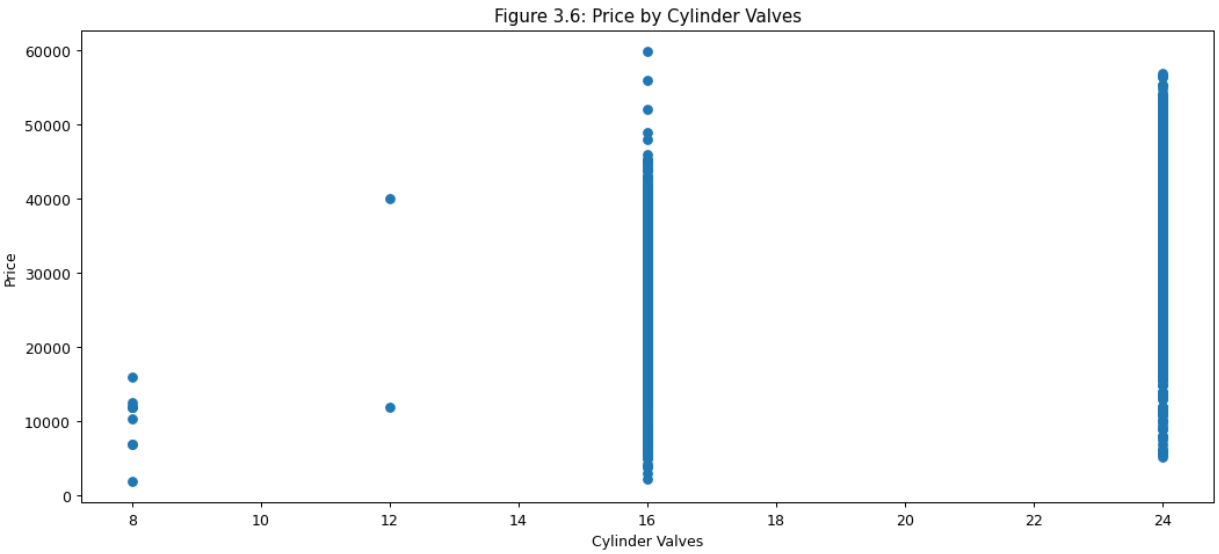


Figure 3.7 shows there is an even split between an engine with 1 or 2 cams.

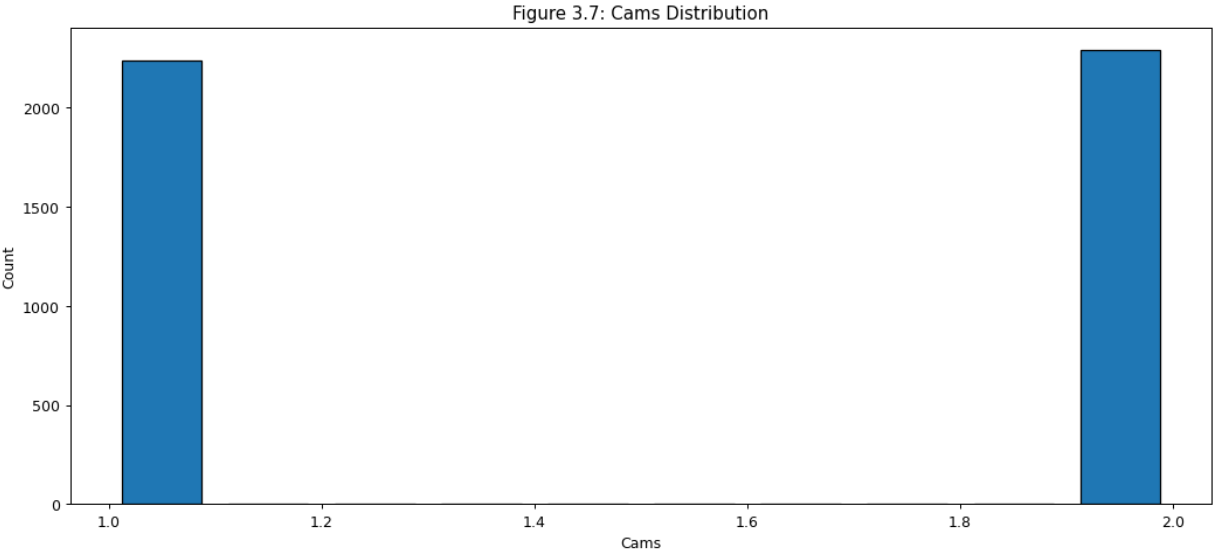


Figure 3.8 shows not much difference between the number of cams and the price point.

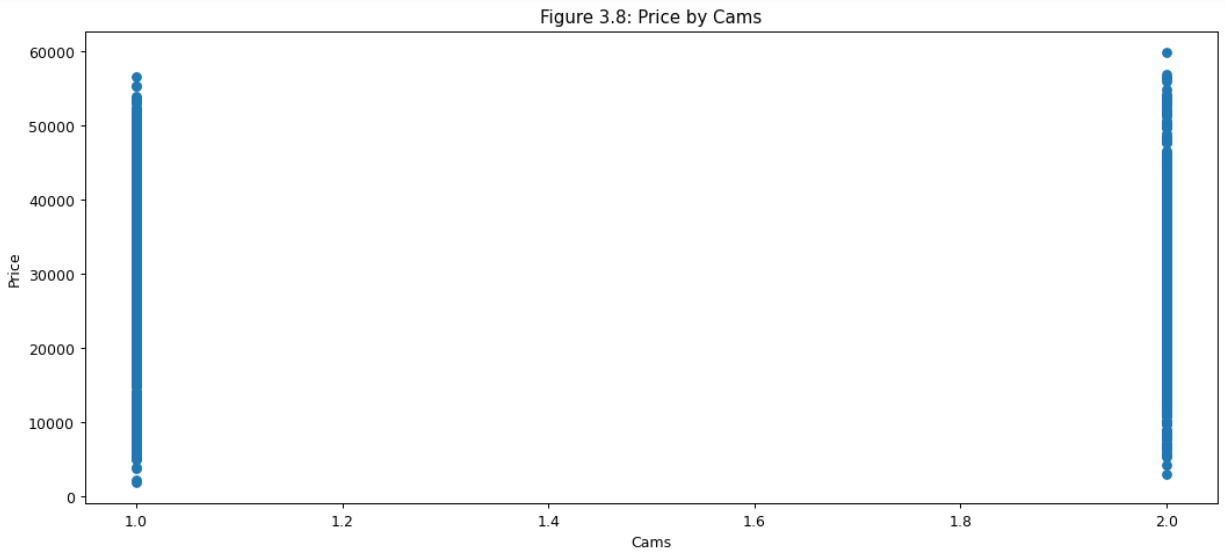
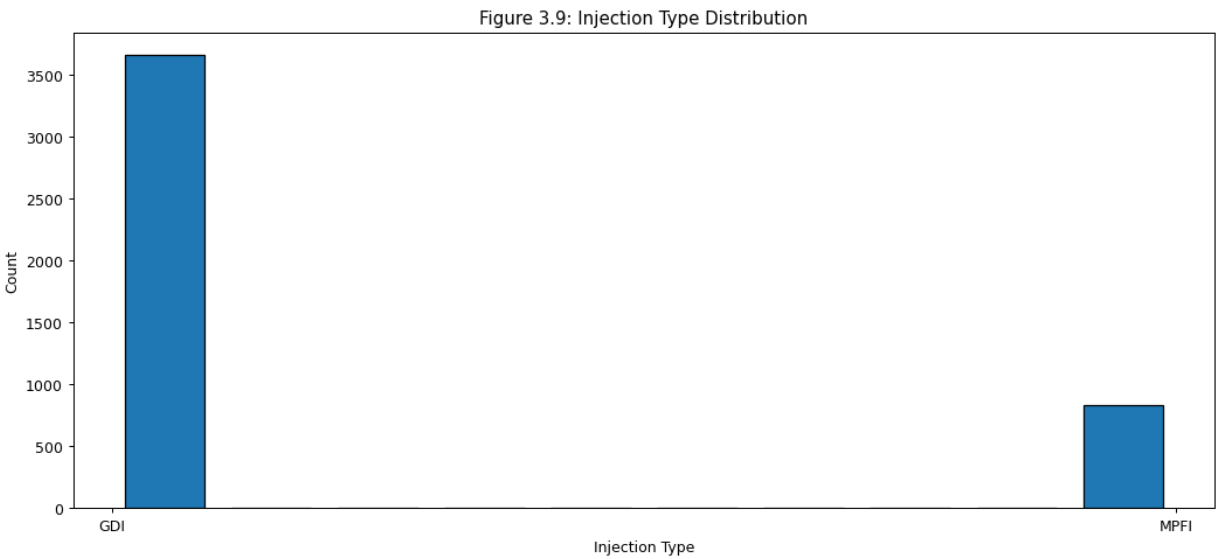


Figure 3.9 shows that the injection type leans much higher for the Gas Direct Injection (GDI). This means any comparison should consider their relative ratio.



The GDI engines do have a price premium. The Multi-Point Fuel Injection (MPFI) engines don’t have any outlier values as high as the GDI engines. The price premium is around $15,000. The GDI system is newer than MPFI and injects directly into the cylinder with higher precision. Consumer values may be evident in the driving experience.

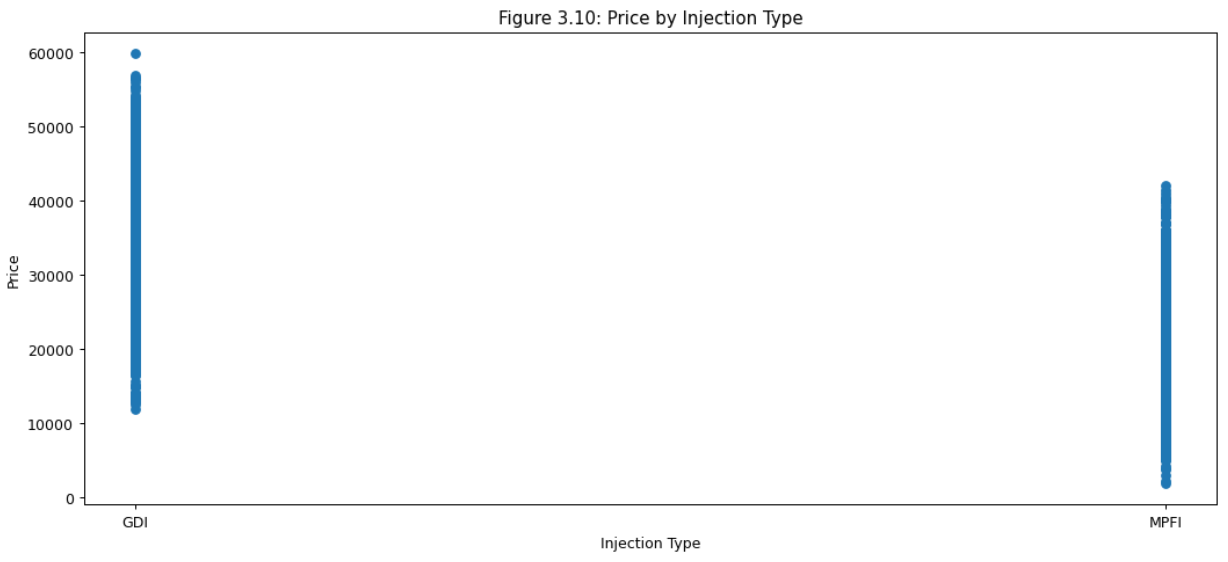
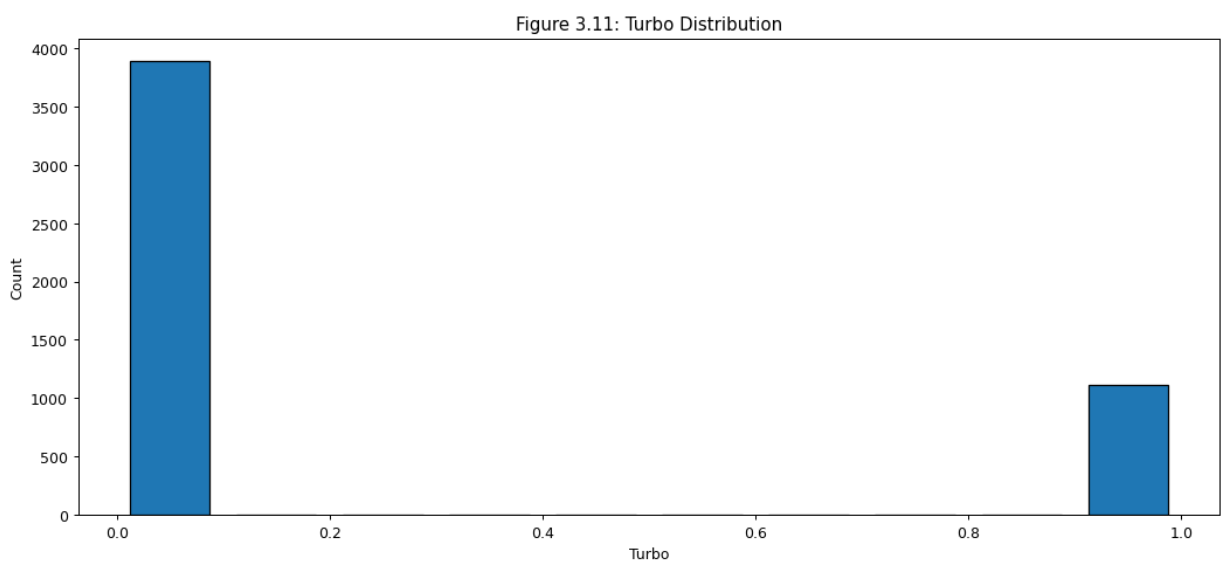


Figure 3.11 shows that turbo engines are almost a quarter of the number of non-turbo engines.



The turbo engines compress air and provide more bang for the buck, but what’s interesting is that there is no indication that the price point is any higher with or without it.

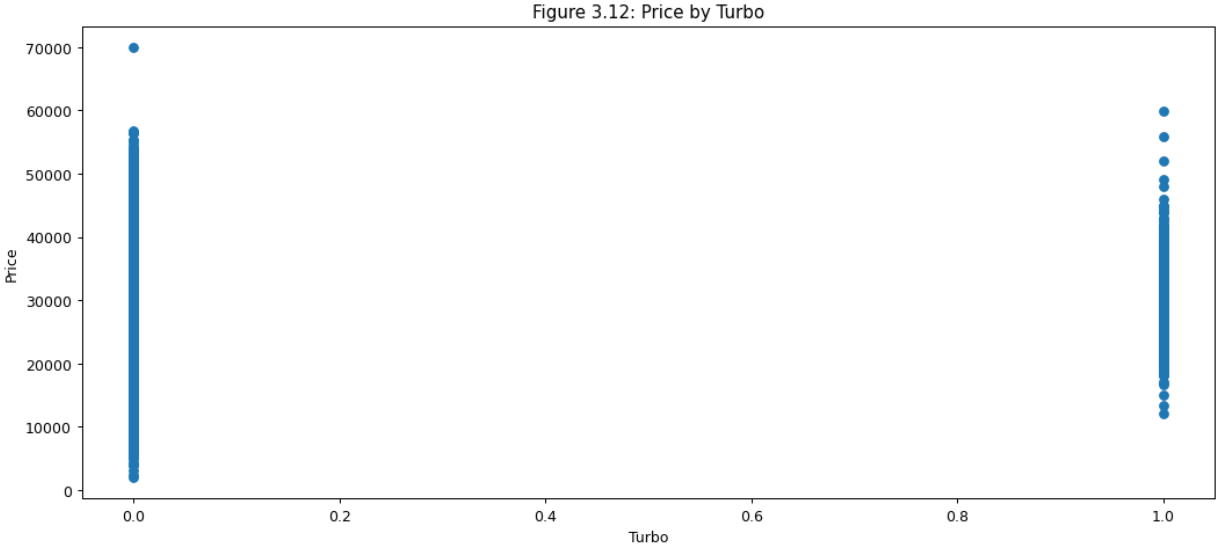


Figure 4 begins by showing which brands command more value from customers. It’s not surprising to see larger vehicles with a higher price point and smaller vehicles with a lower price. The Element being the lowest cost is somewhat surprising in that it’s not a compact car but costs less than a CR-Z or a Del Sol which was last produced in 1998.

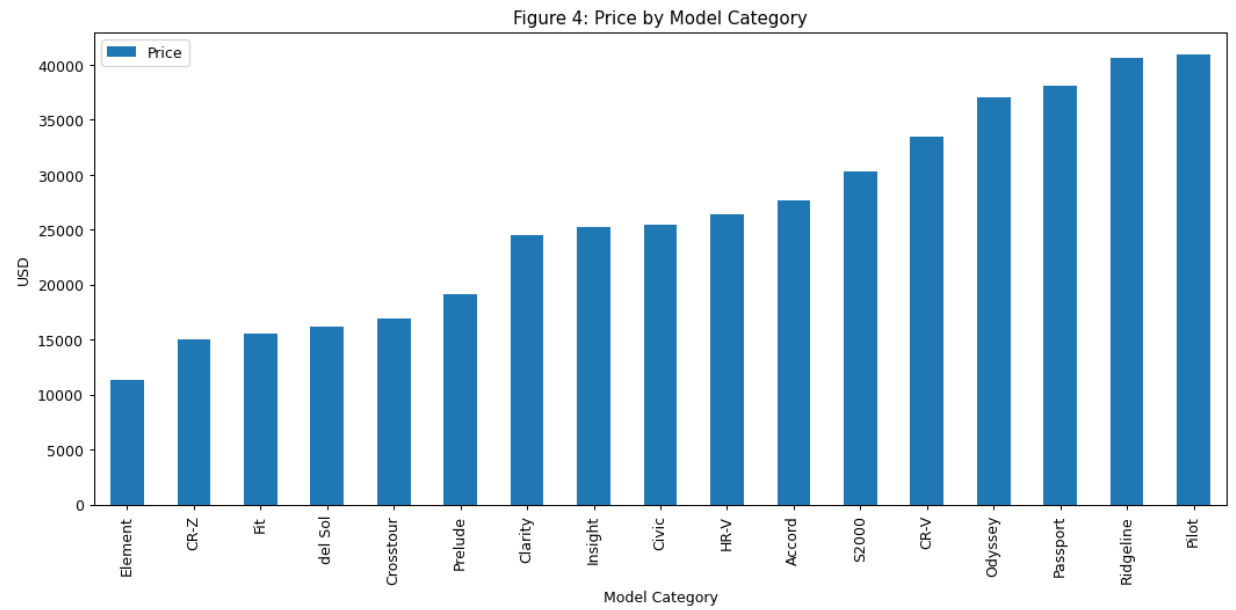
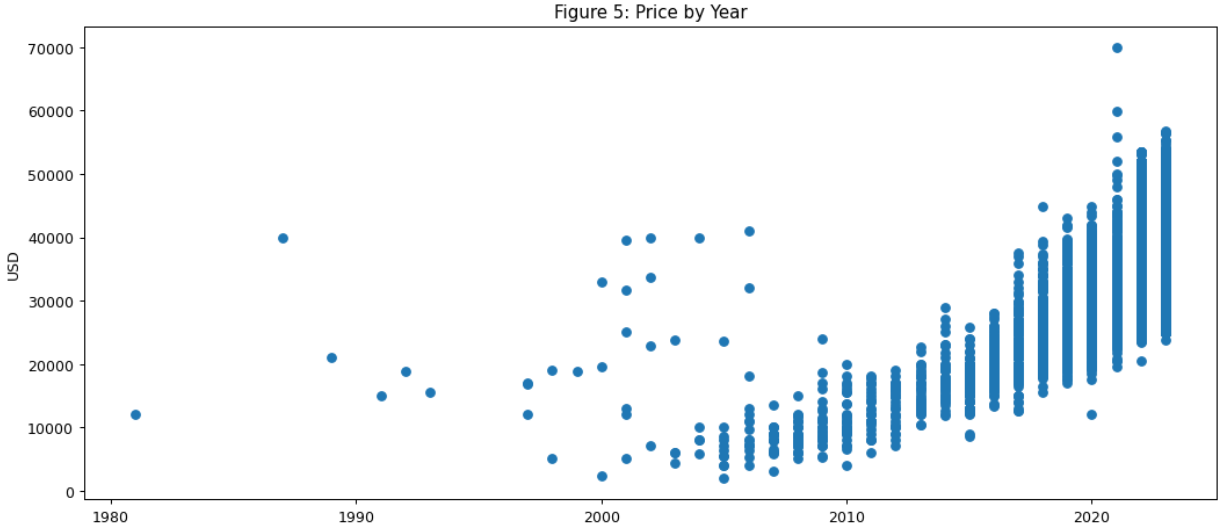


Figure 5 shows that prices have a positive correlation with newer models. This isn’t surprising at all. What is interesting are the oldest cars. There are six built prior to 1995 whose price seems to stay high despite their age. Four of them are Preludes. The other two are a Del Sol and a Civic.



Figures 6.1 through 6.17 show the average price for each model per year. A positive relationship would be expected, but a different outcome may indicate something about the quality of the model.

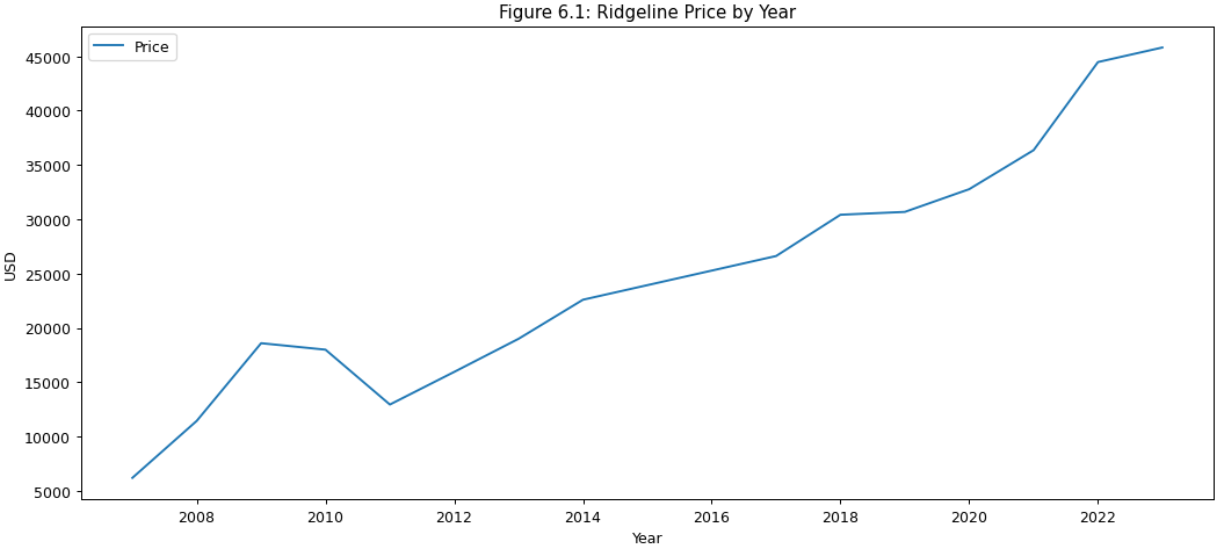


Figure 6.1 shows a strong positive relationship by year for the price of a Ridgeline.

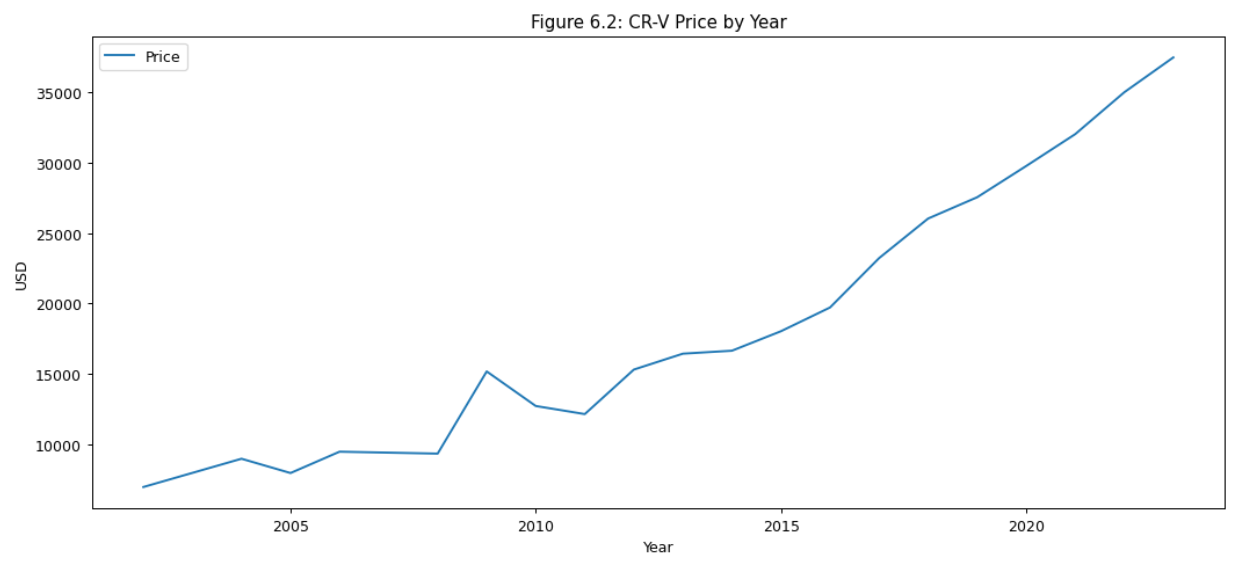


Figure 6.2 shows a positive relationship with a somewhat slow start for the price of a CR-V.

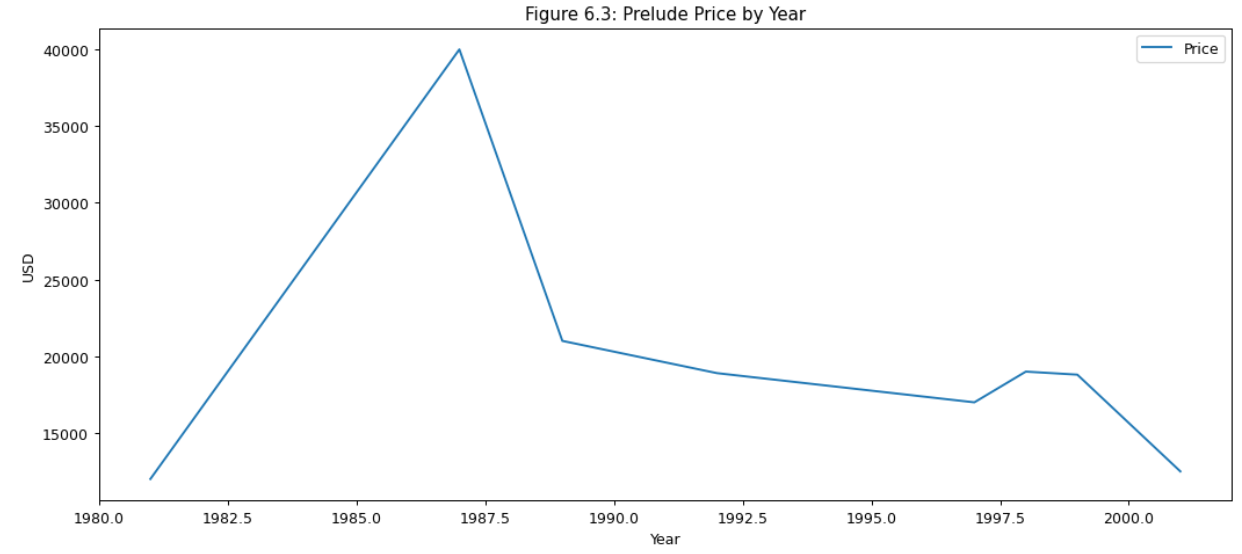


Figure 6.3 shows a complicated relationship with a somewhat slow start for the price of a CR-V. The high values in the 1980s are from the few high-valued car sale records that came up earlier. They are outliers with too few records to determine if there was a reason it was highly valued, but none of the models after that seem to indicate the value held.

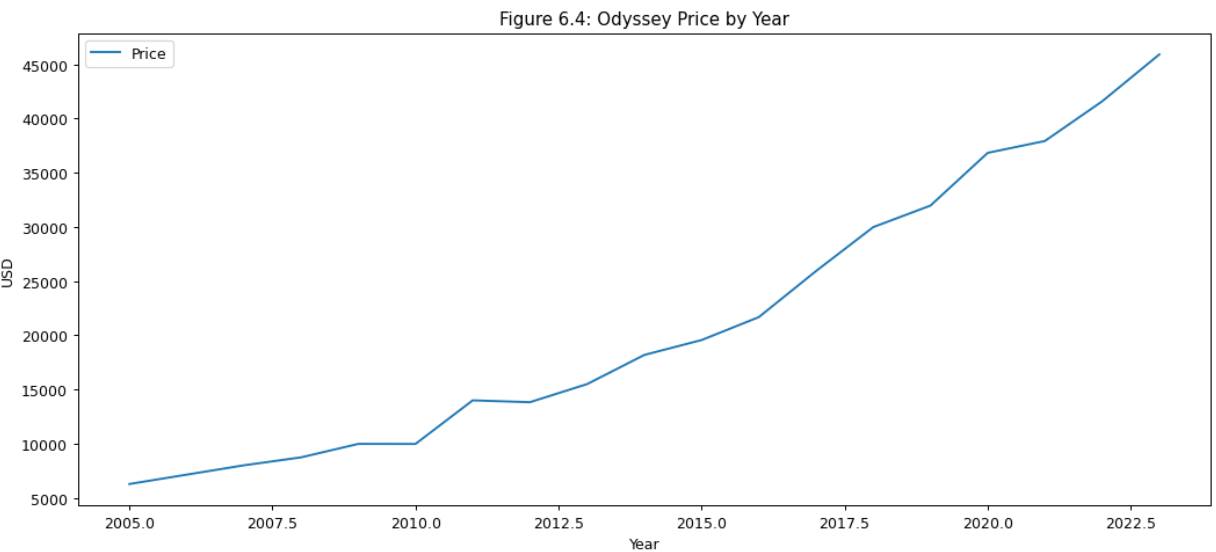


Figure 6.4 shows a positive relationship with a somewhat slow start for the price of an Odyssey. Its numbers are higher overall than the CR-V

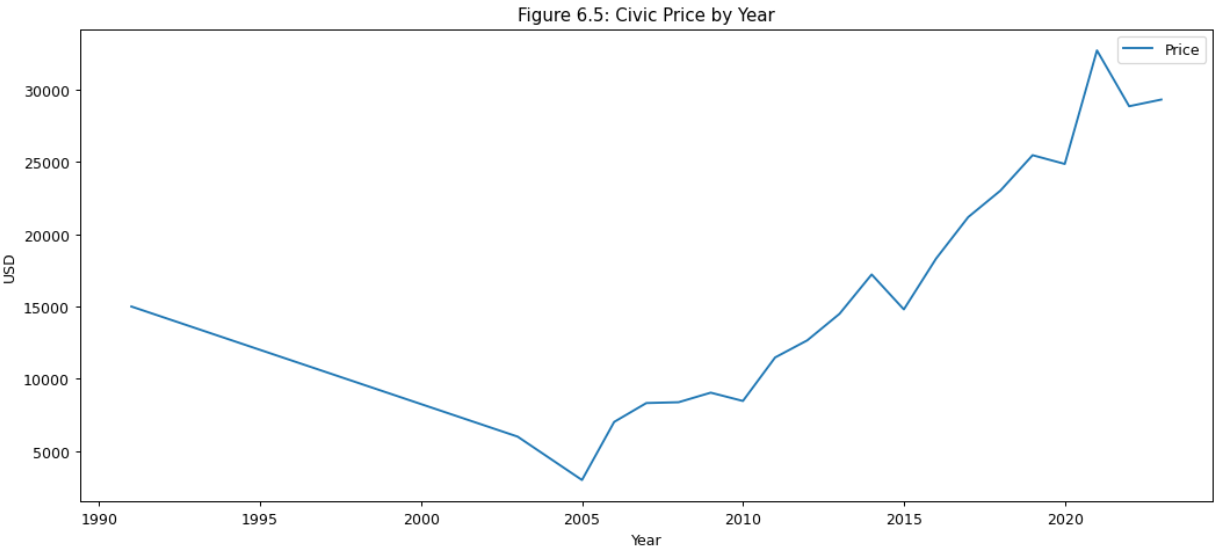


Figure 6.5 shows a crash and then a climb for the Civic which, bottomed out in 2005. It may indicate an interesting turnaround story for that brand.

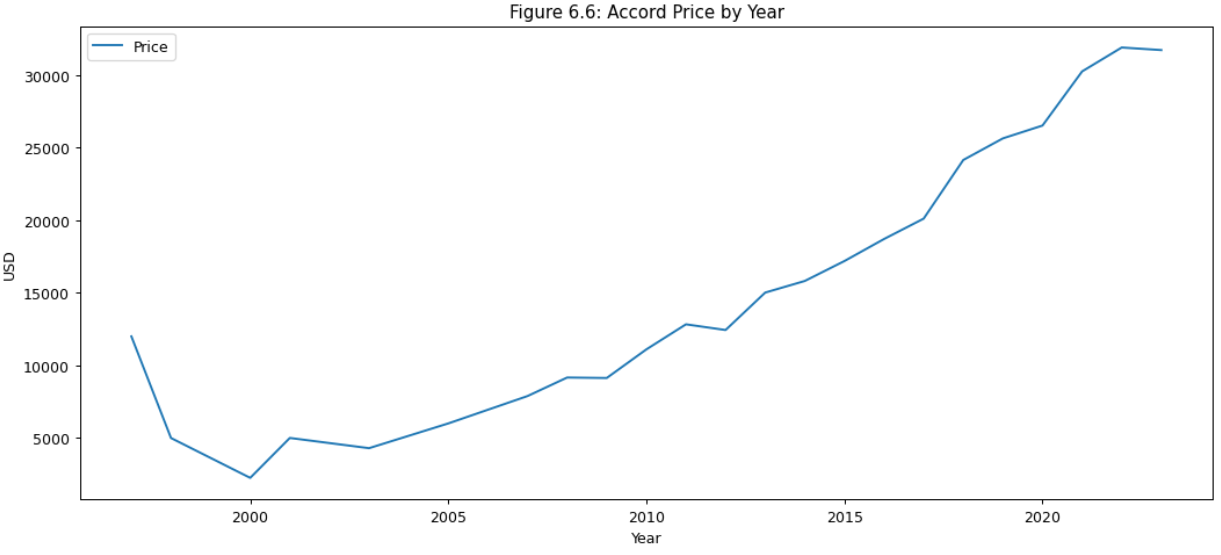


Figure 6.6 shows a mostly positive relationship for the price of an Accord with the initial dip in the years just after the first models.

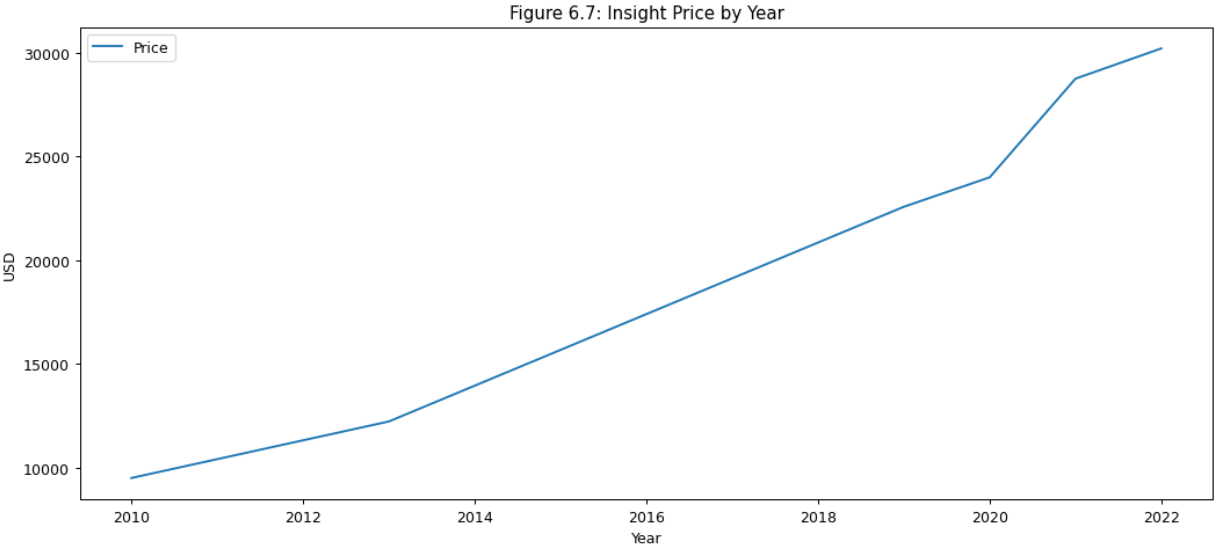


Figure 6.7 shows a steady positive relationship with the price of an Insight.

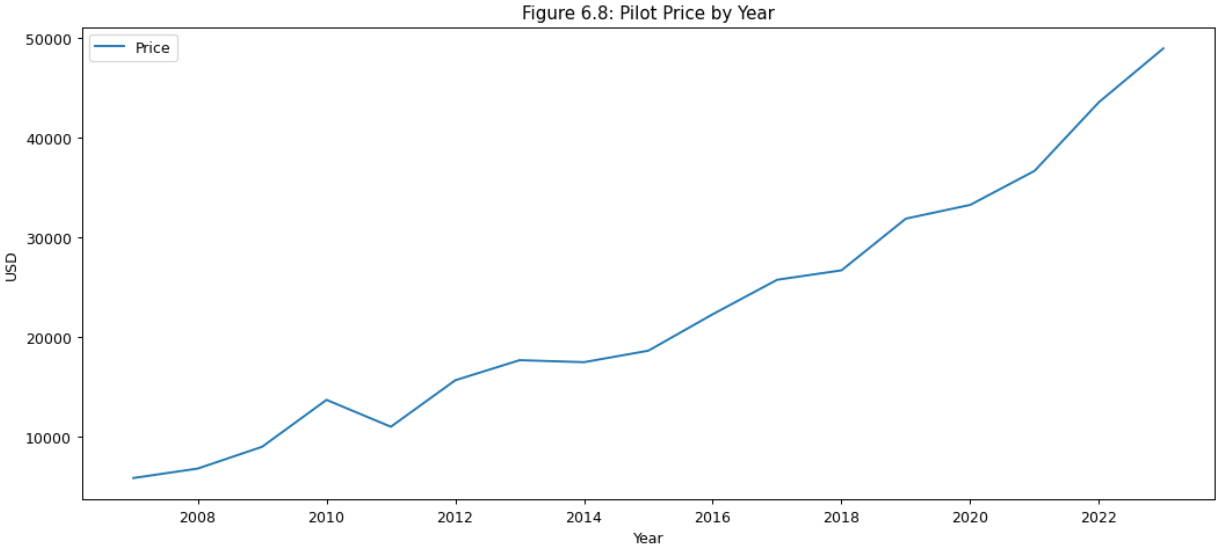


Figure 6.8 shows a modest positive relationship with the price for a Pilot.

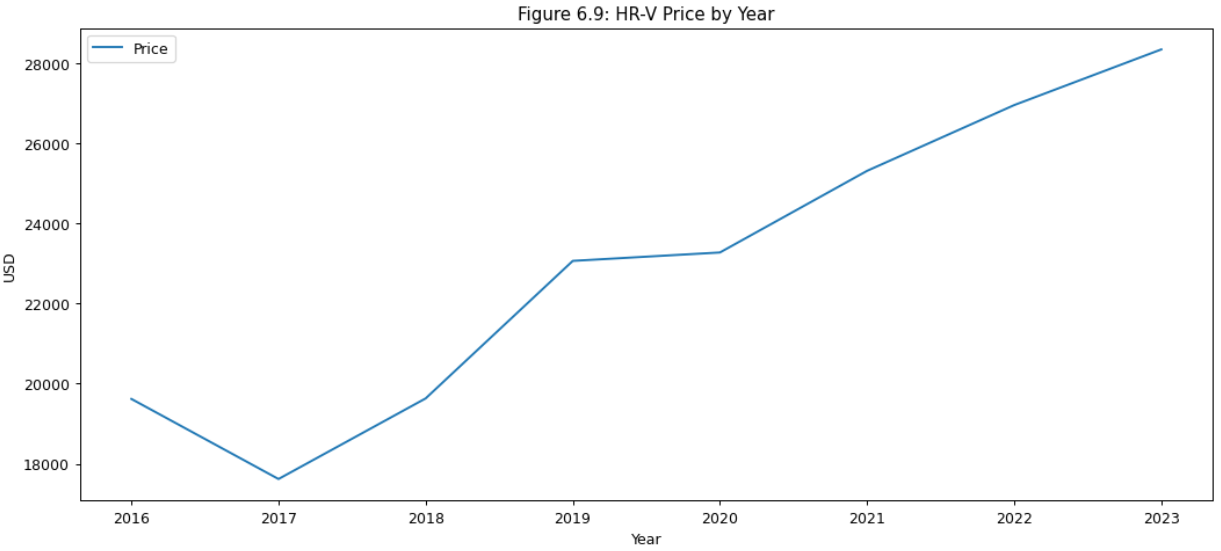


Figure 6.8 shows a strong positive relationship with the price for an HR-V with only a small bit of variance in the first years of the US-released model.

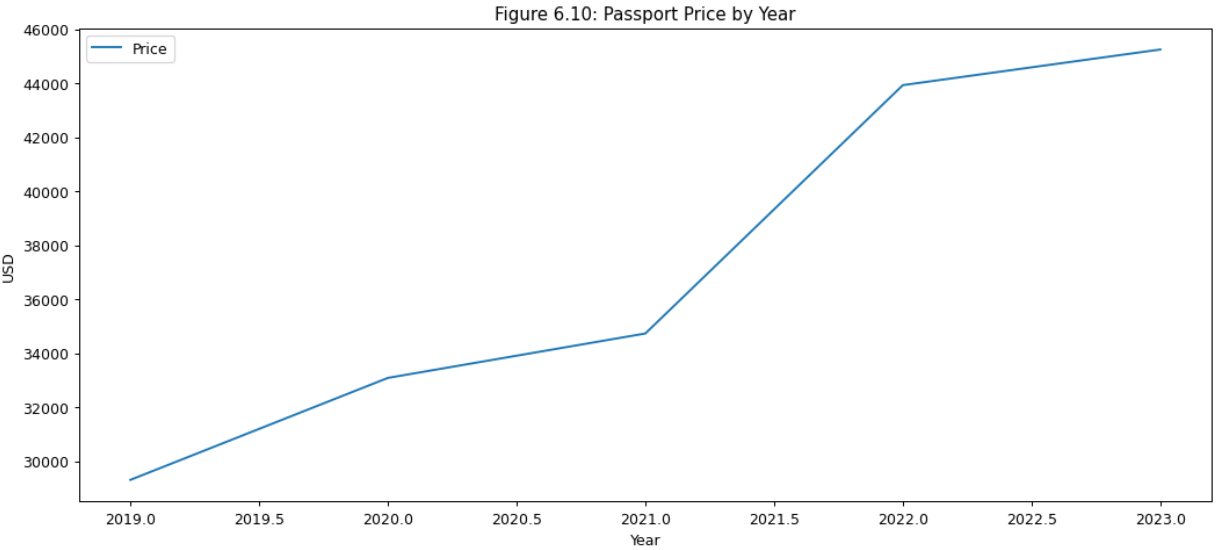


Figure 6.10 shows a strong positive relationship with the price of the Passport.

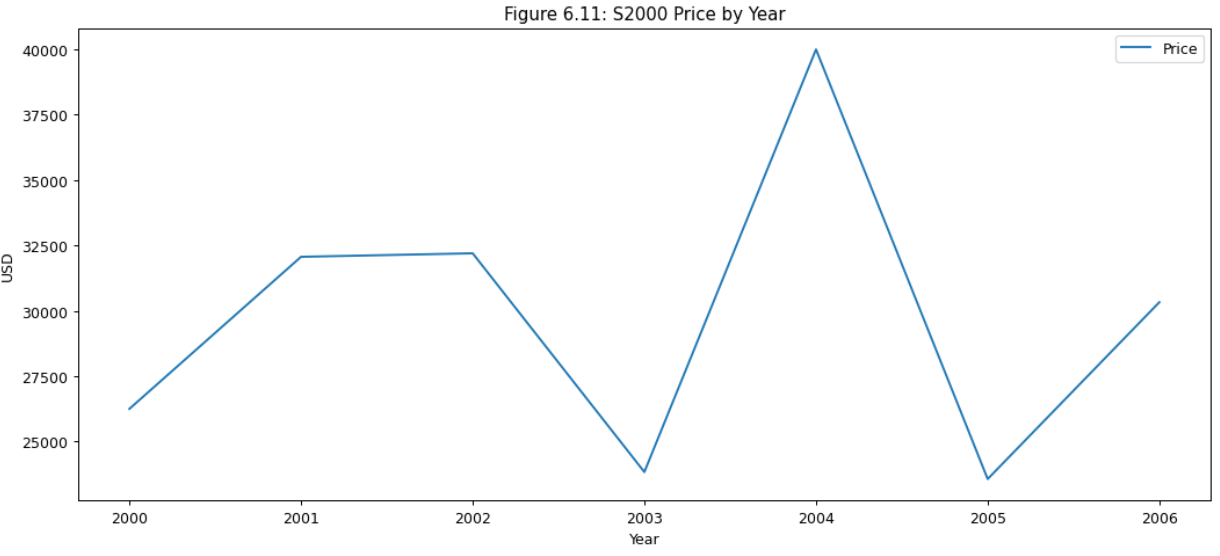


Figure 6.11 shows a very interesting price point for the S2000. It ranges below $25,000 up to $40,000 but with no discernable pattern. The mean is roughly $28,000. The model lifespan data spans six years. Wikipedia indicates the car was a roadster that was quite powerful and popular. The cars seem to hold their value which is why the price seems relatively flat.

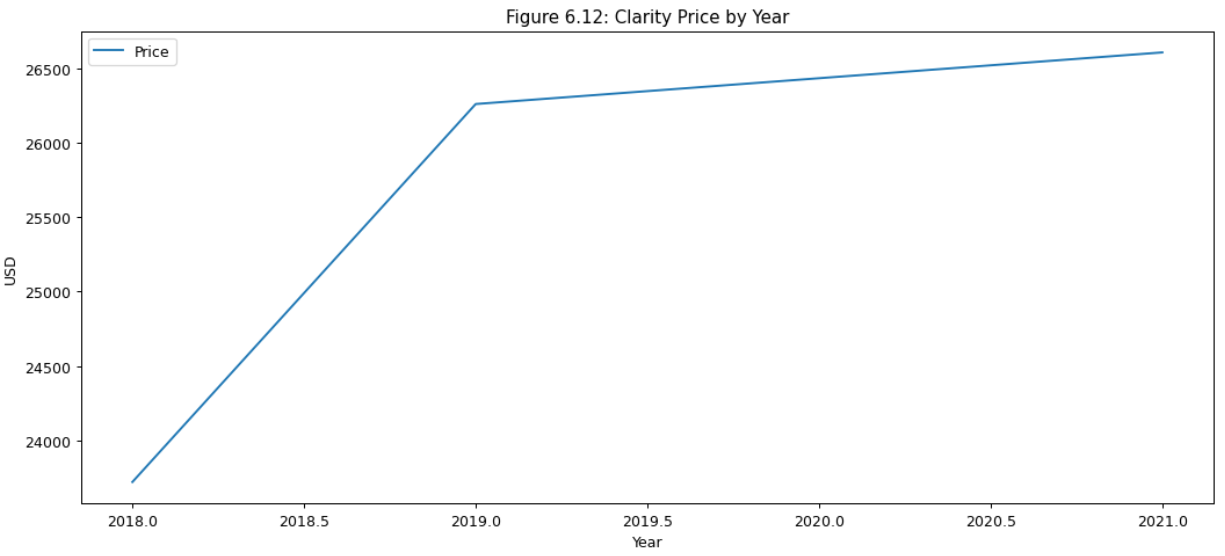


Figure 6.12 shows an interesting increase in price and then relatively flat. Some of this is due to the zoomed-in nature of the data, with only 3 years of data and a range in price of $3,000.

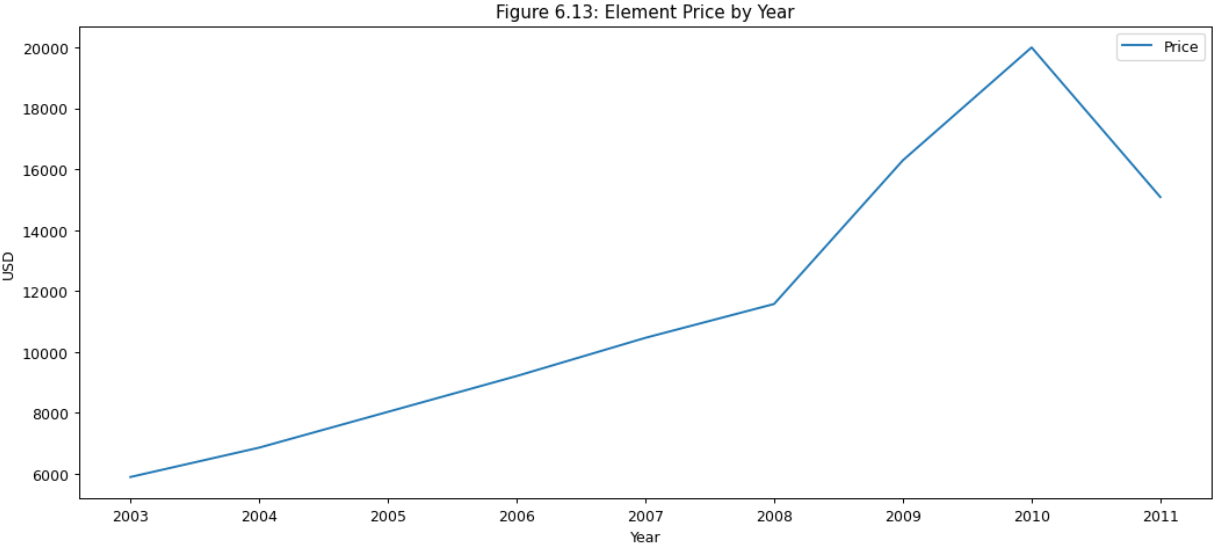


Figure 6.13 shows a slow climb, then a steep increase, then a fall in the final year. The peak price is certainly lower than most but also started from a very low price.

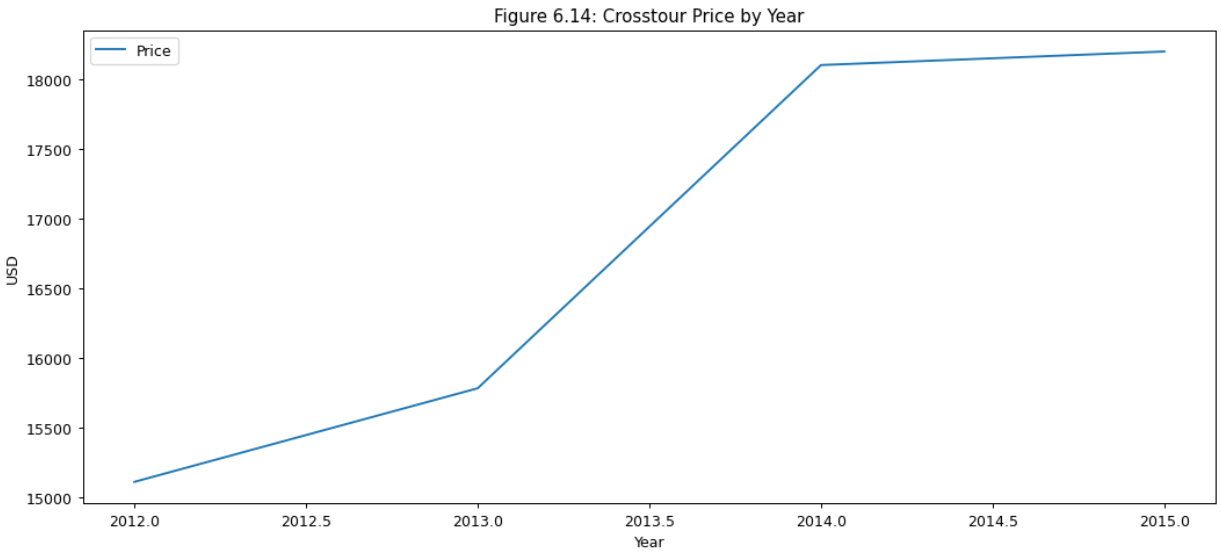


Figure 6.14 shows a short range of $3,000 in price over 3 years but steady growth in that time.

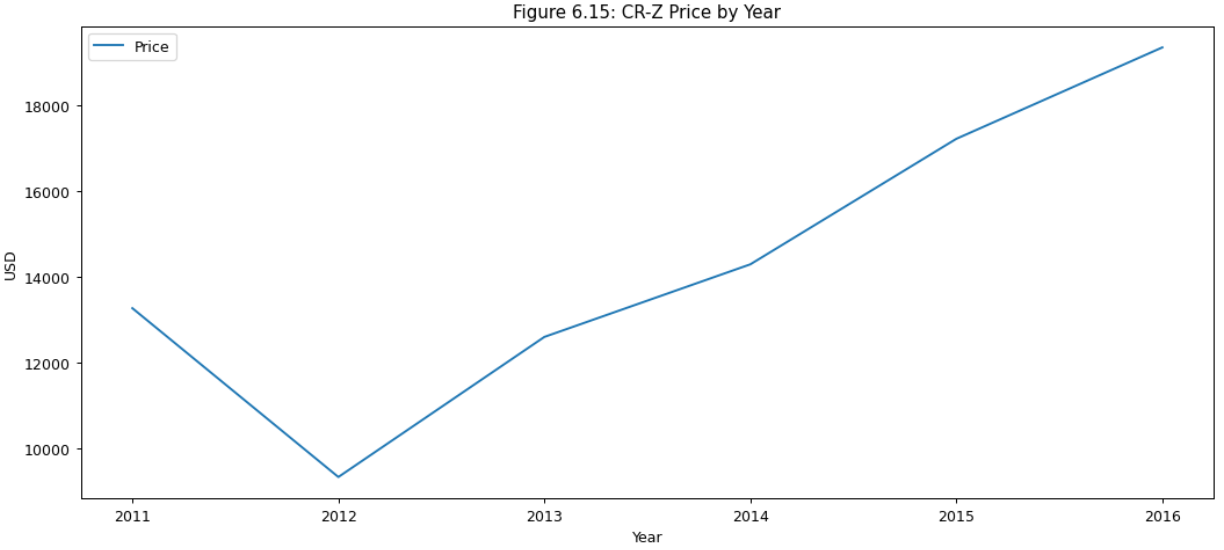


Figure 6.15 shows a relatively positive trend for the price of a CR-Z, with a notable exception in the drop of $4,000 in price in the first year of model records.

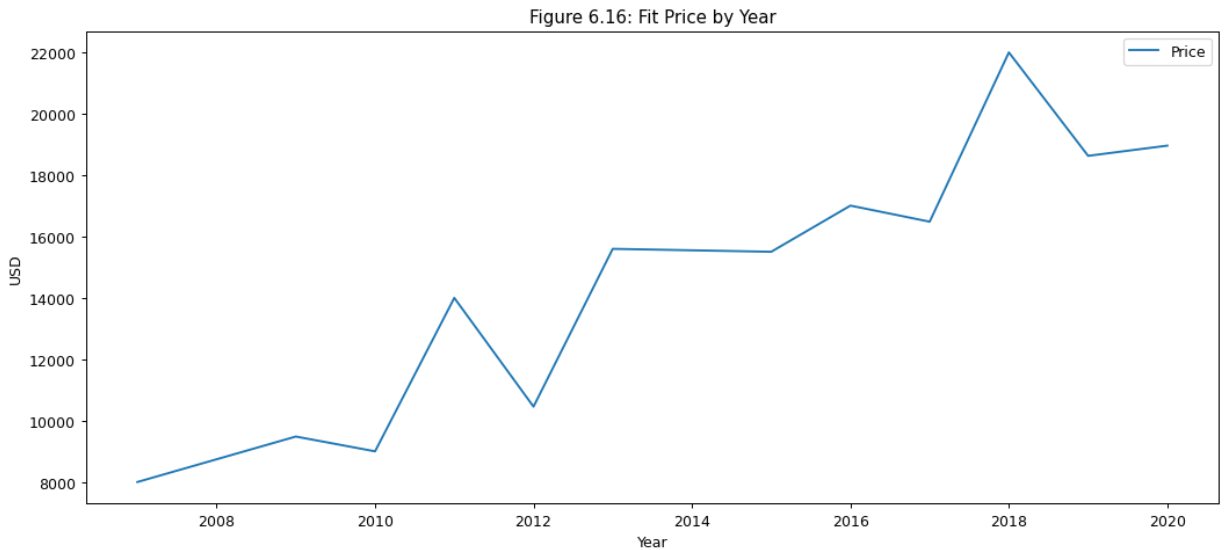


Figure 6.16 shows a very jagged but positive climb in price over 13 years to almost triple the price. A strong performer with a somewhat inconsistent value.

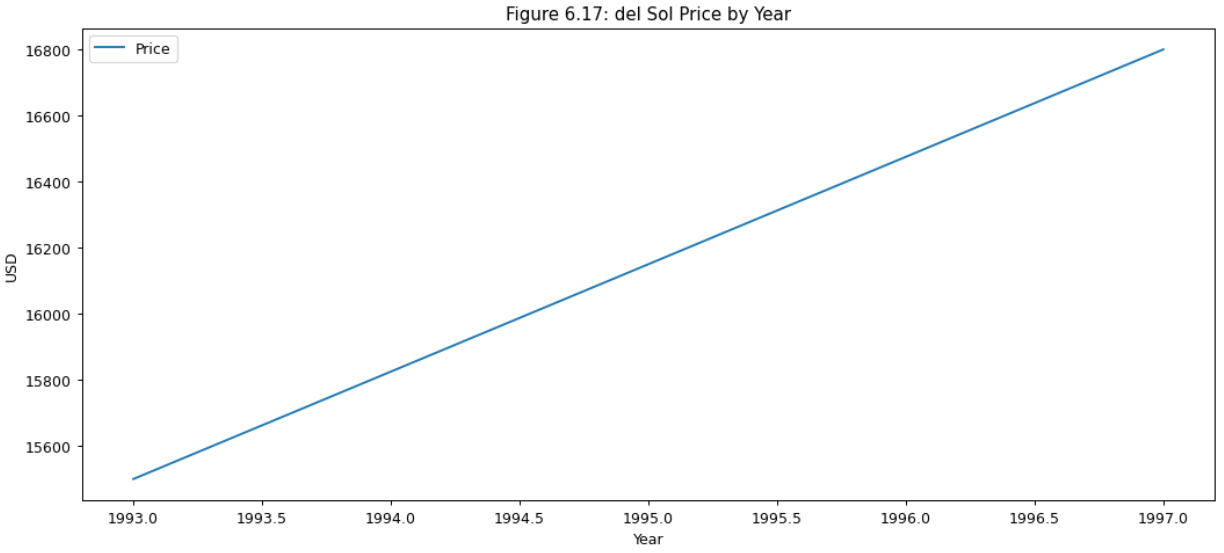
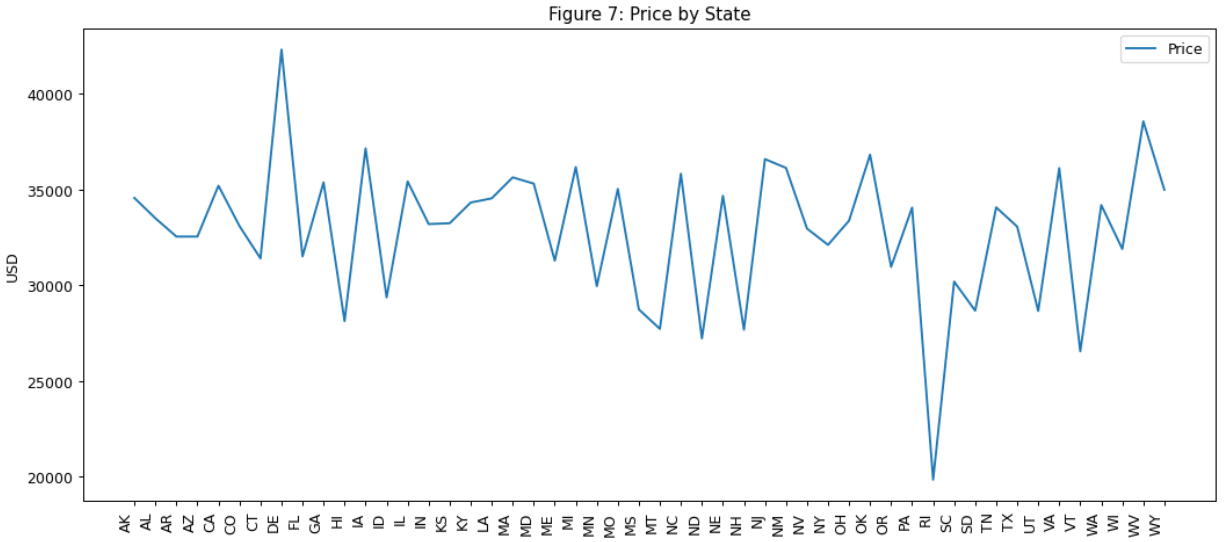


Figure 6.17 shows a 4-year window where the price of a Del Sol increased by $1,000. There must not have been a large market for this model, and it doesn’t appear to be highly valued.

Figure 7 now focuses on the average price by state. Most states show a trend of roughly $32,000 for a vehicle, with two notable exceptions. Rhode Island has the lowest average selling price at $20,000, while Delaware has the highest average price above $40,000.



Figures 8.1 through 8.5 show additional car features by price.

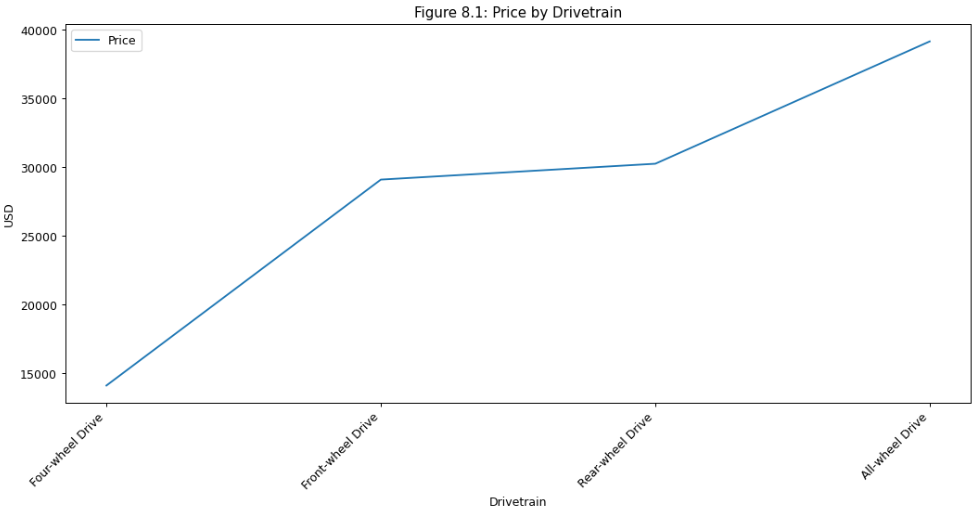


Figure 8.1 shows the price premium for the engine drivetrain. The all-wheel drive vehicles demand the highest price. Front-wheel drive isn’t as costly as rear-wheel drive, but both are remarkably less expensive than front-wheel drive while four wheel drive is by far the lowest cost. This may be due to the specific models providing higher value to consumers for managing hard driving conditions like snow.

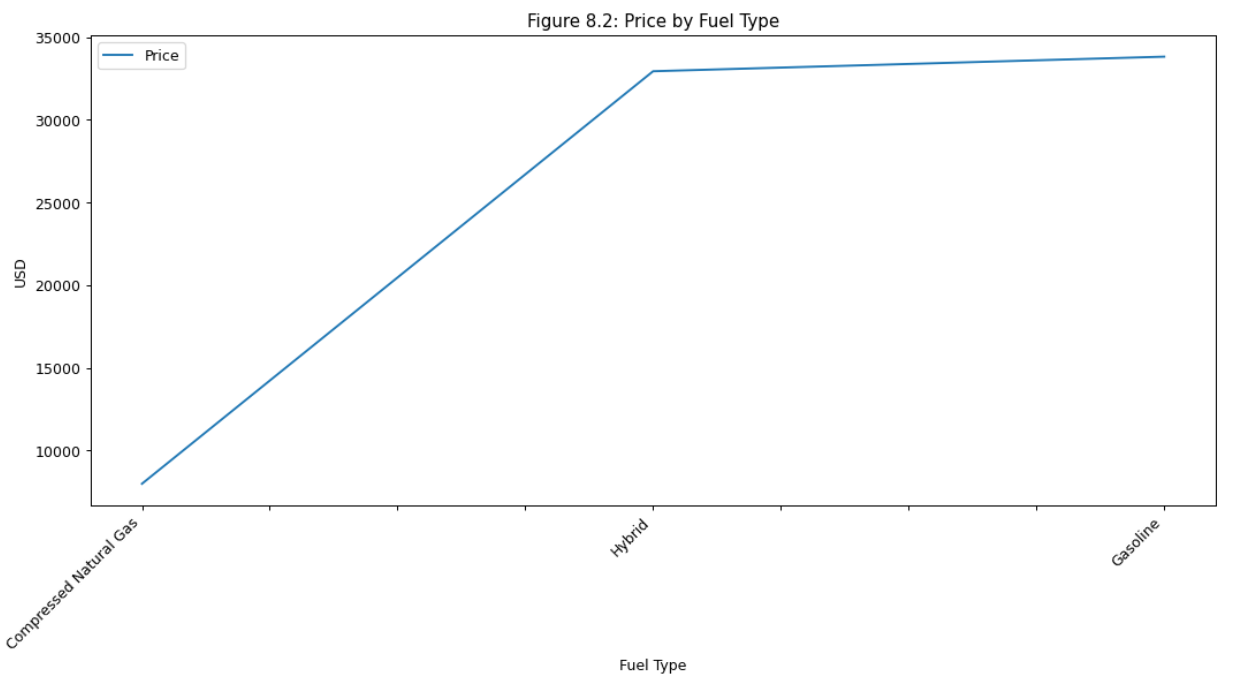


Figure 8.2 shows that Compressed Natural Gas (CNG) engines do not hold a price premium. Gas and Hybrid engines hold a similar value with, Gas being slightly more expensive.

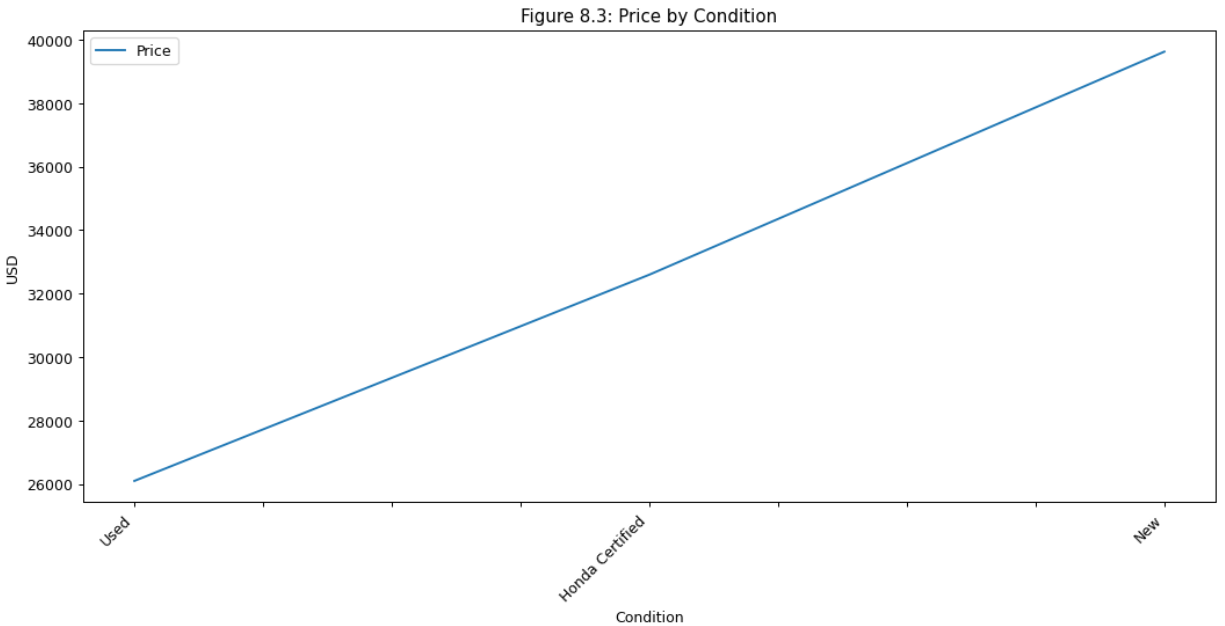


Figure 8.3 shows the price range for car conditions. It’s not a surprise that the used cars are the least valued, with Honda Certification above it but below a new car value.

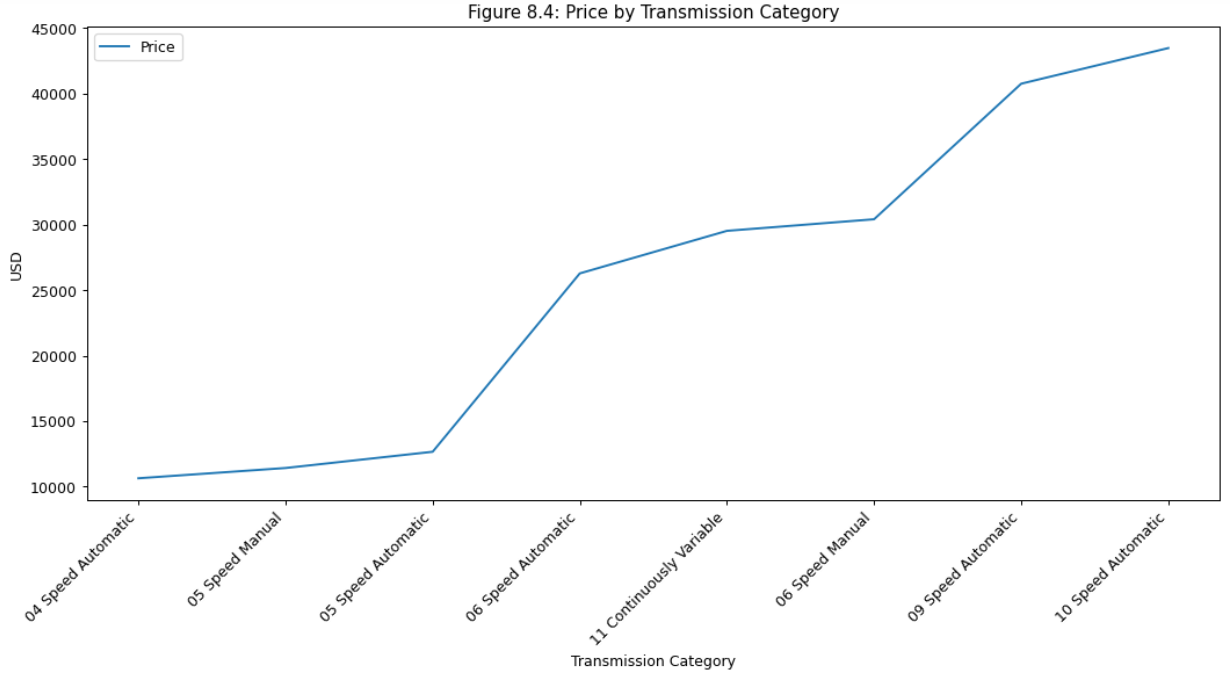


Figure 8.4 shows a trend of increasing price with more gears, with a notable exception of the Continuously Variable Transmission (CVT,) which is a high-performing all-in-one gear with a mid-range price.

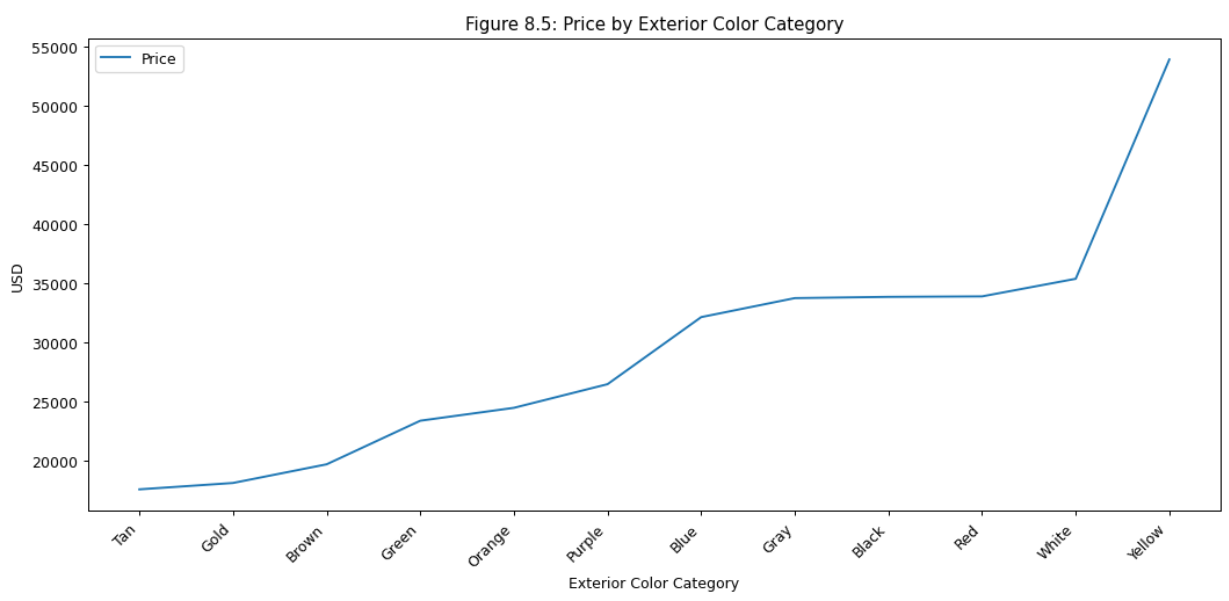


Figure 8.5 shows that Tan, Gold, and Brown at the lowest end of the pricing range, while Gray, Black, Red, and White are at the top. Yellow is an anomaly with a price point premium of $20,000.

Figures 9.1 to 9.3 show distributions of different car features.

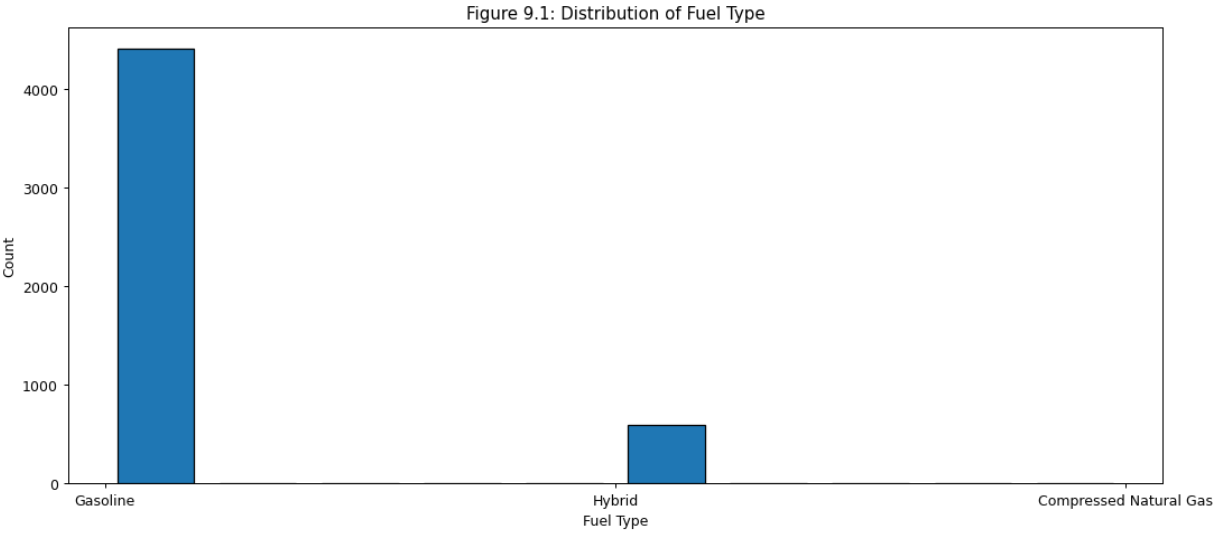


Figure 9.1 shows the distribution of fuel types. Gas engines are almost 9 times more represented in the dataset. CNG has almost no records.

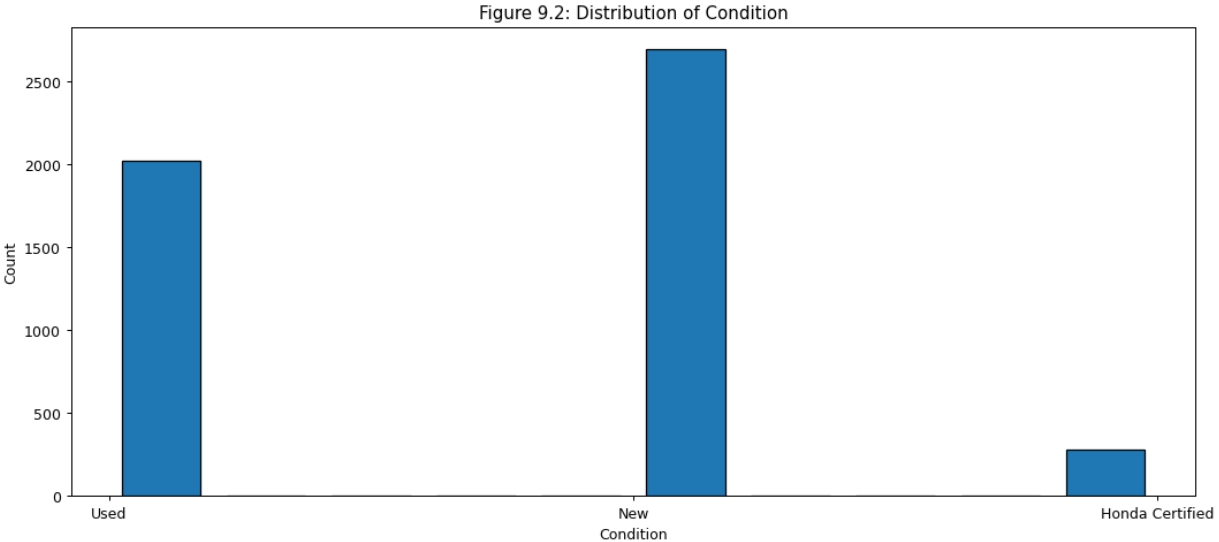


Figure 9.2 shows that Honda Certified cars don’t make up much of the market, while new cars are slightly more represented in the dataset than used cars.

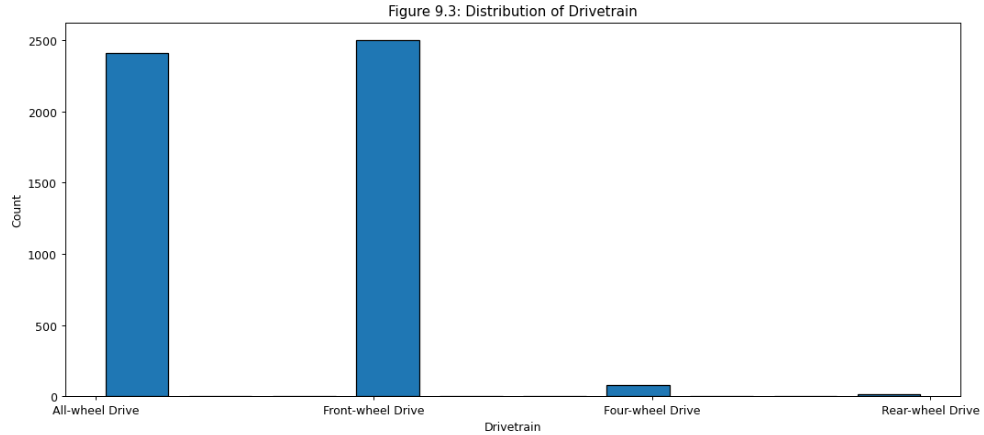


Figure 9.3 shows an almost equal distribution of all-wheel drive and front-wheel drive vehicles, with very few records of rear-wheel drive vehicles.

Figure 10 shows the average price against the average miles per gallon. I would expect to see a positive correlation between price and mpg, but there doesn’t appear to be. There may even be a somewhat negative correlation which is unexpected. This could be accounted for by larger gas vehicles commanding a higher price and smaller, more efficient vehicles having better gas mileage at an affordable price.

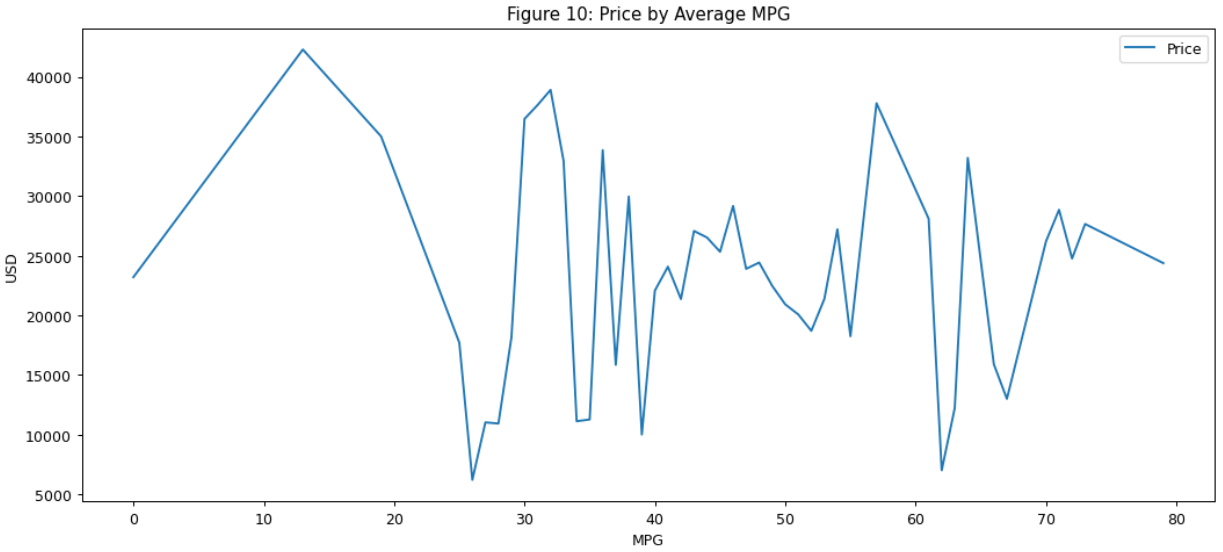
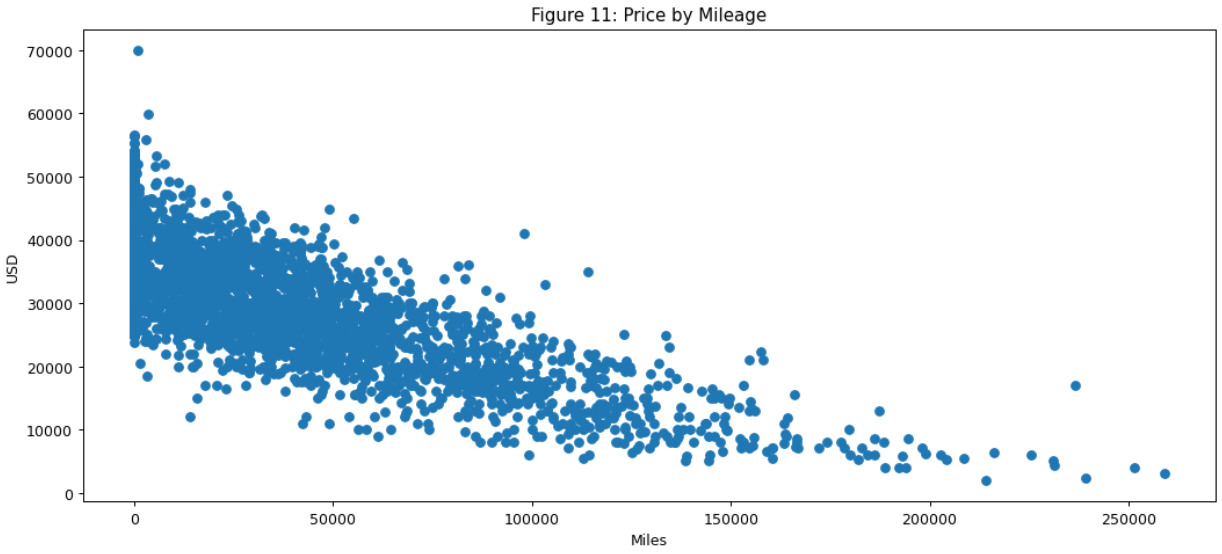


Figure 11 shows a very strong correlation between miles driven and price. Price favors the newer cars.



## Question 2: What states have the highest or lowest cost?

Question 2 focuses on the cost for states, so Figure 12 shows the distribution of vehicle sale records by state. There are roughly three groups: 250-300 cars, 50-150 cars, and less than 50. Most states fall into the second category.

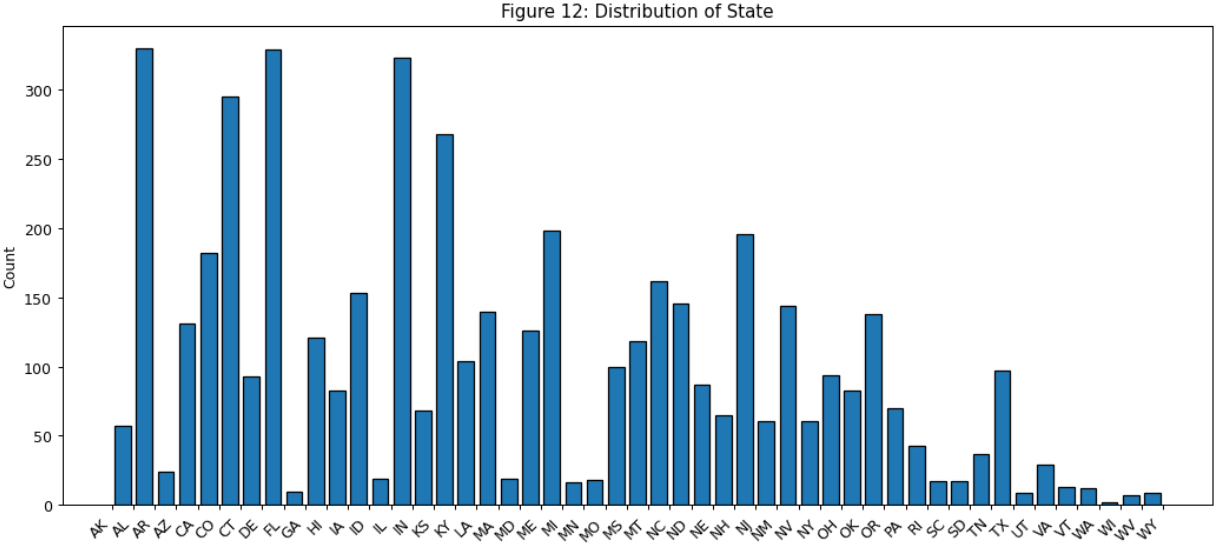


Figure 13 shows another distribution view that includes prices for each state. California has the highest number of cars in the $42,000 range. There are a few other states with high-value car sales but not many.

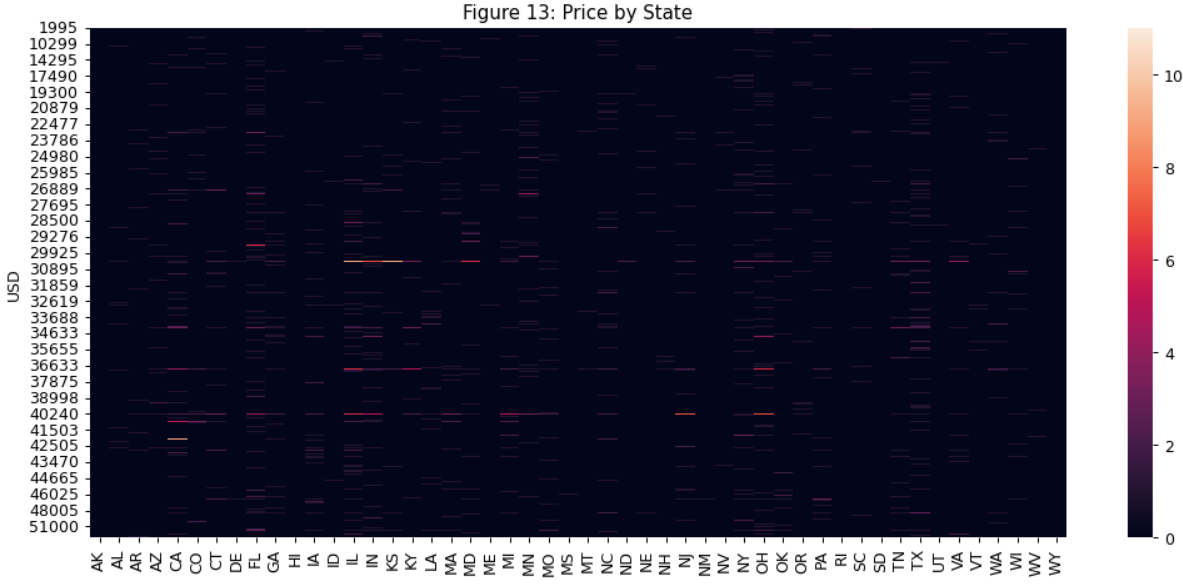


Figure 14 shows the average miles per gallon (MPG) by state. California has reasonably moderate MPG indicating that those prices may afford better overall car economy. Delaware has the lowest MPG. Combine that with the highest average car prices, and Delaware seems to be quite expensive to own a Honda.

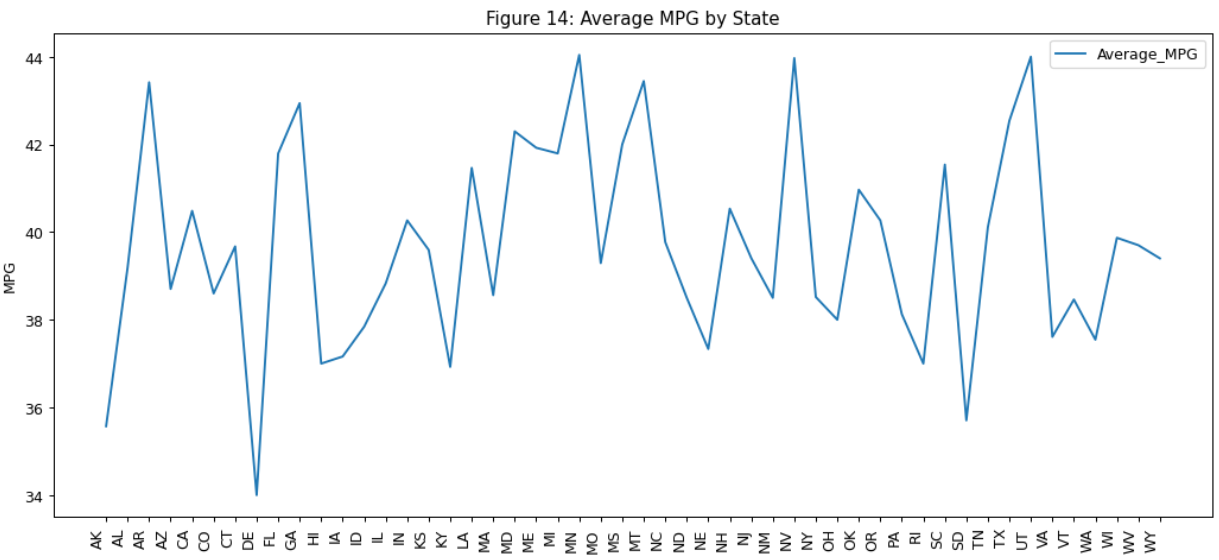


Figure 15 shows a different view of the MPG per state with a clear band around the 30-34 MPG range as well as a second band around 44-46 MPG. There are a few states with substantial numbers of high-range vehicles like California, Florida, Georgia, Illinois, Minnesota, and North Carolina.

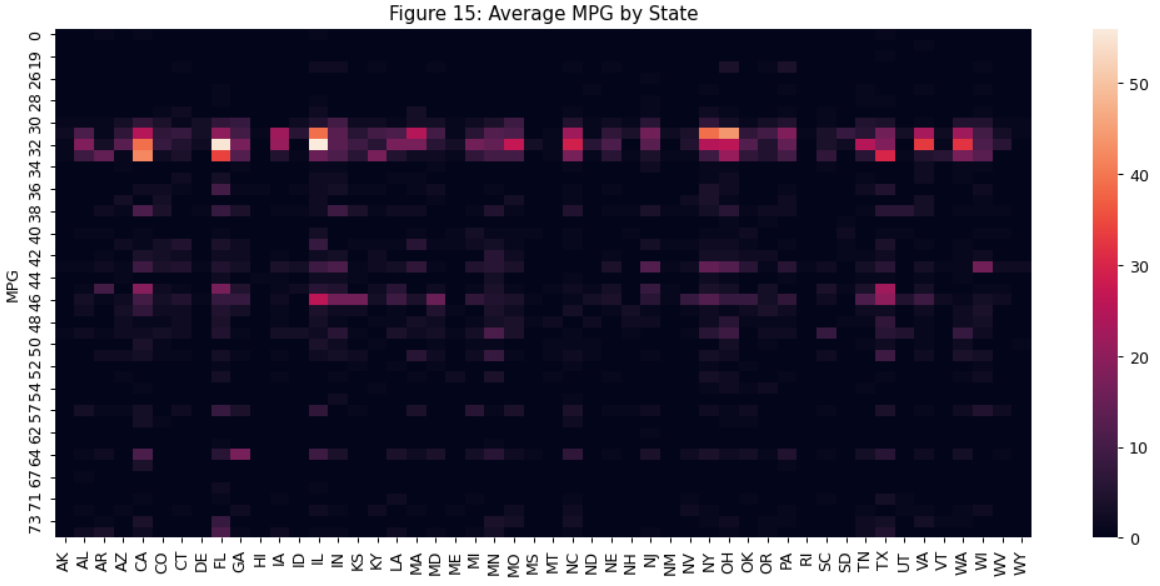


Figure 16 shows what condition the Honda cars are being sold in each state. California and Illinois are selling the most new Hondas. Texas and Florida are next behind them. Florida, on the other hand, sells the most used Honda cars.

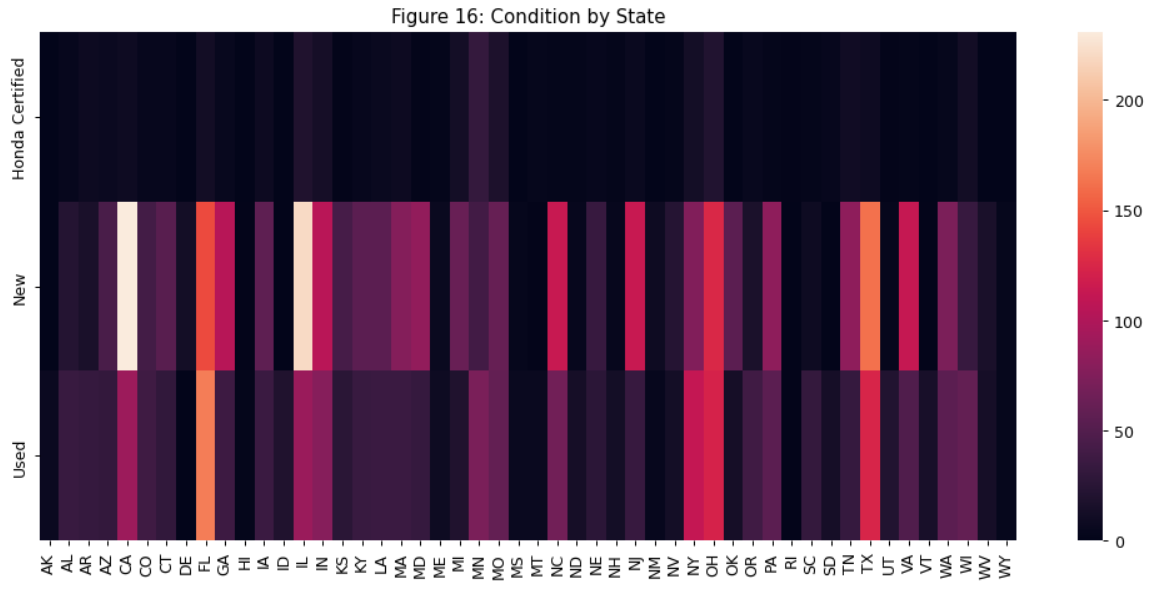


Figure 17 shows what drive trains in Hondas are being sold by state. Florida leads with front-wheel drive, with California and Texas close behind. Illinois leads with all-wheel drive Hondas with Ohio not far behind.

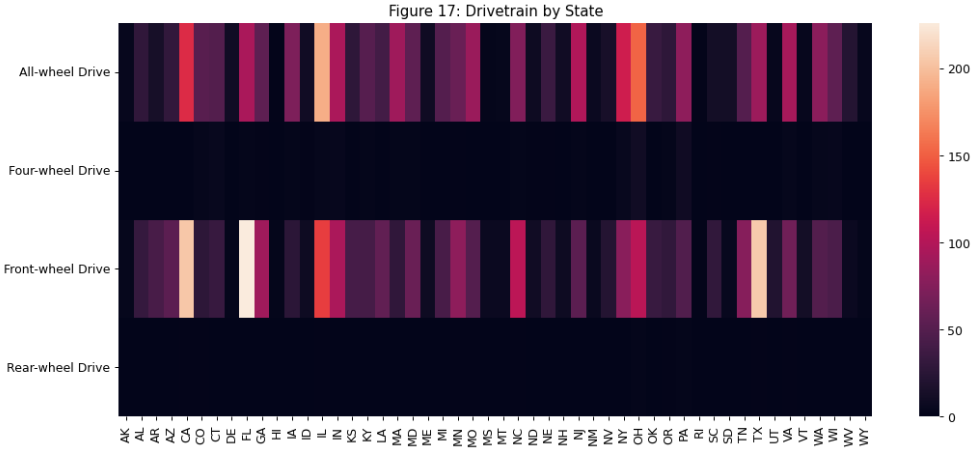


Figure 18 shows fuel types in Hondas sold by state. Illinois, California, Florida, Texas, and Ohio lead with gas vehicles. California, Florida, Georgia, Illinois, Indiana, North Carolina, and Texas are leaders in Hybrid sales.

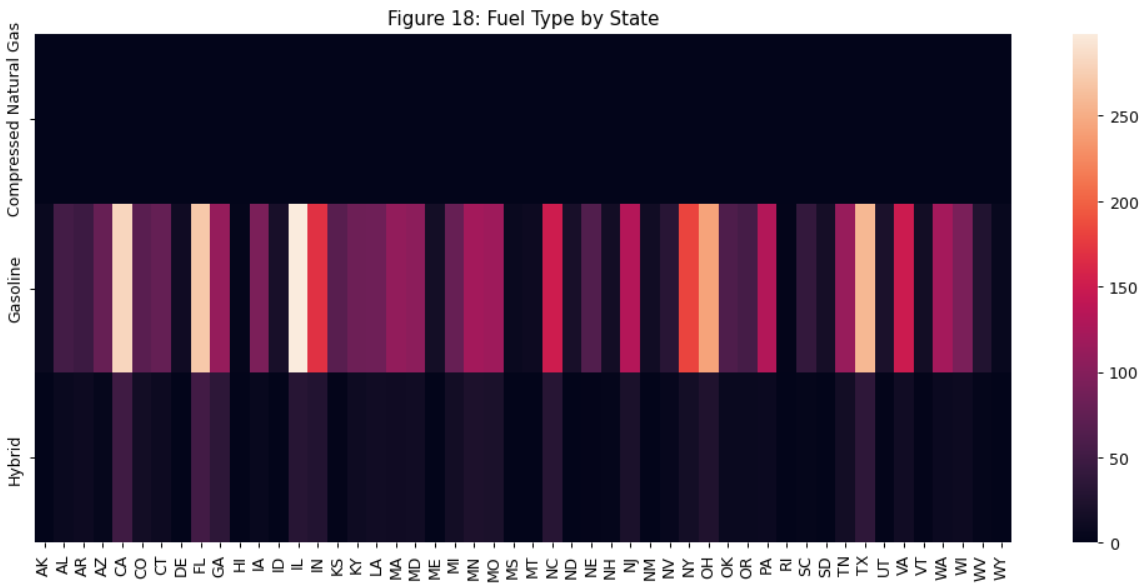
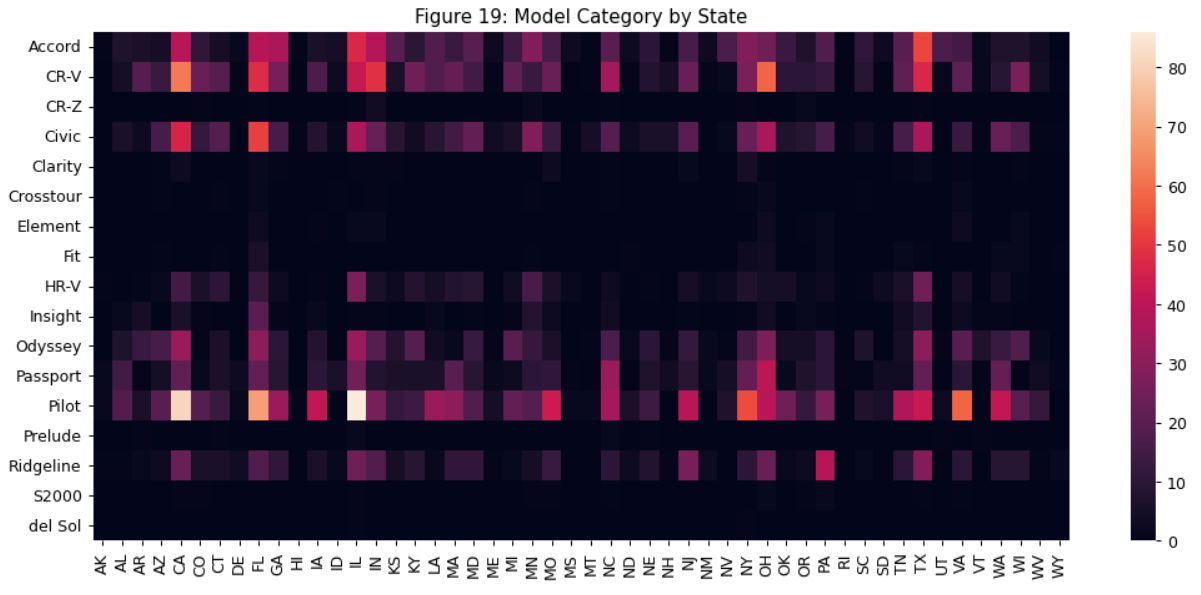


Figure 19 shows what model categories each state is selling. Illinois, California, and Florida have a large volume of Pilot sales. California and Ohio also lead with CR-V sales.



## Question 3: How has MPG changed over the years per model?

Question three focuses on miles per gallon (MPG) and the change over time. Figures 20.1 to 20.3 show distributions of the years, MPG, and model category.

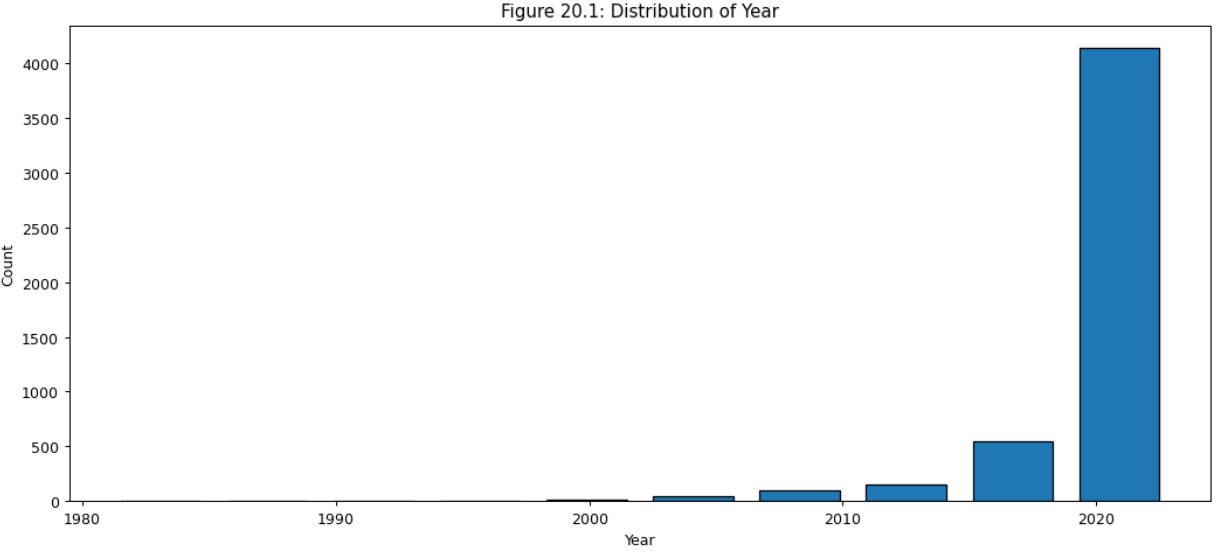


Figure 20.1 shows that most of the records are for newer sales. This could indicate that data on the low end may suffer from higher outlier effects.

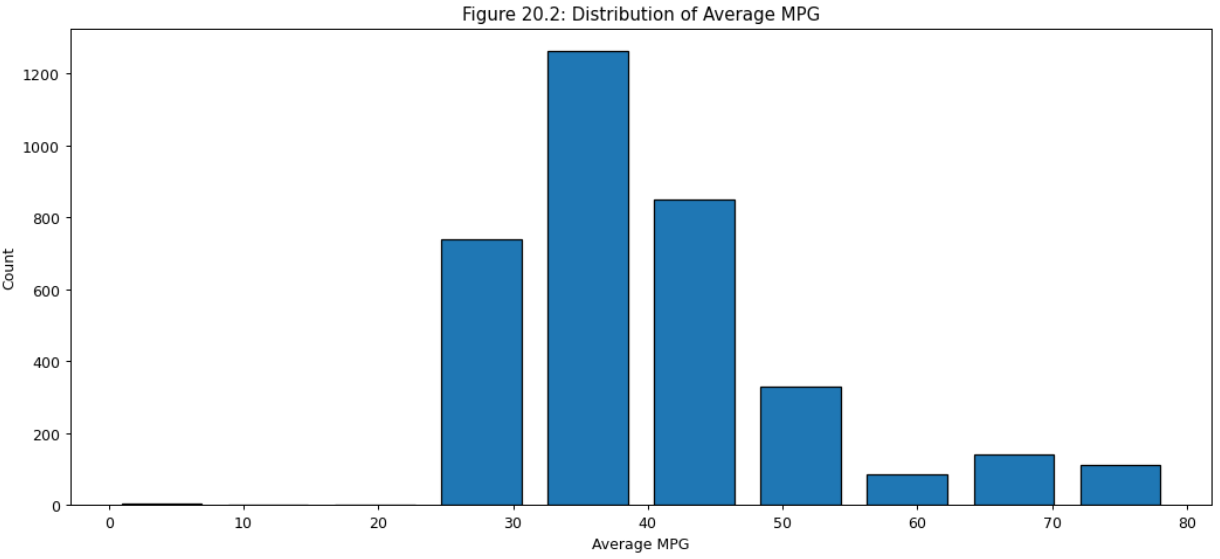


Figure 20.2 shows the distribution of MPG. There are very few models with less than 25 MPG, and they top out just below 80, but the mean is around 35 MPG.



Figure 20.3 shows about half the models have a reasonable number of records, with the rest having a very small number.

Figures 21.1 to 21.11 show how MPG stacks up against engine features.

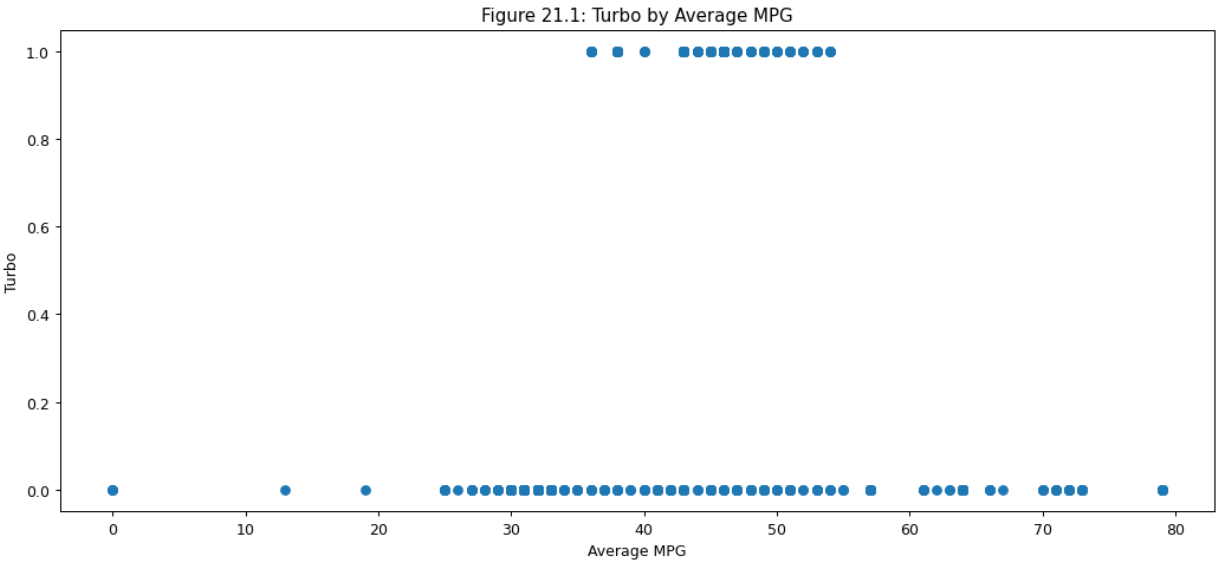


Figure 21.1 shows that turbo engines have a range between 35-55 MPG, and non-turbo engines range between 25-75 MPG. There’s no clear indication that turbo engines provide a genuine advantage over regular engines, but that may also be due to the limited number of uses for it. If all models had some samples using turbo engines, it may prove out different.

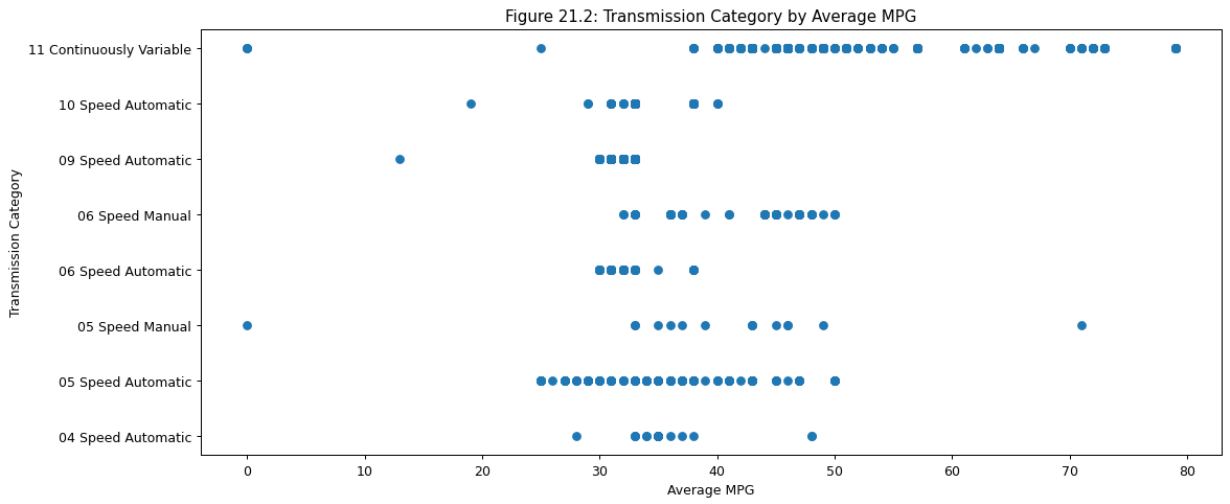


Figure 21.2 shows how the transmission gearing affects MPG. There is a clear indication that CVT engines provide much higher MPG than traditional geared transmissions.

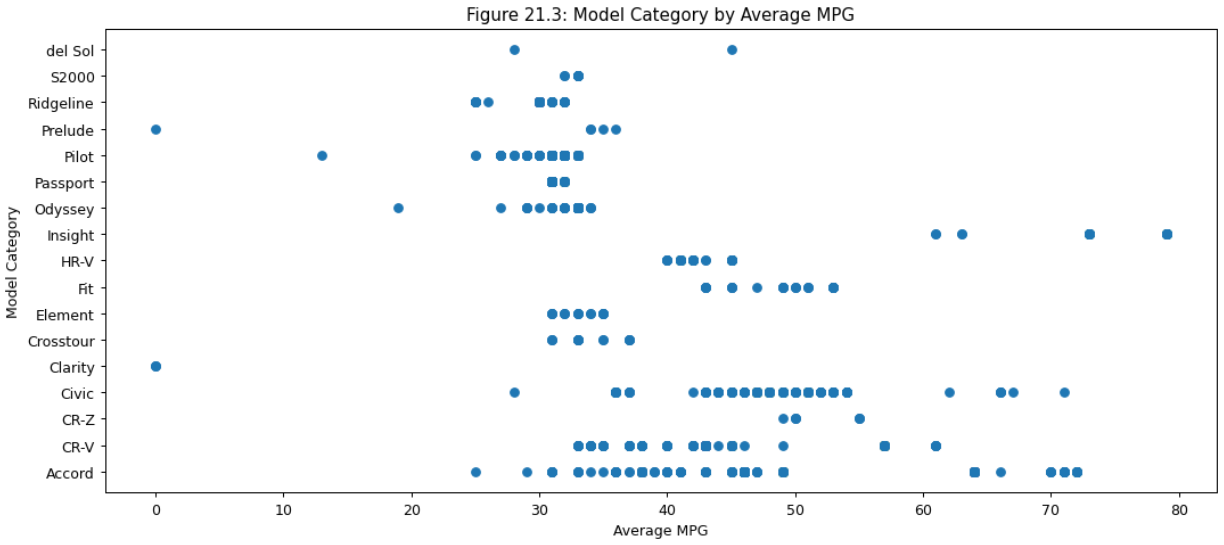


Figure 21.3 shows that there are a few models that have higher MPG. They are the Insight, Civic, and CR-Z. All other models have a mean between 30 and 50 MPG.

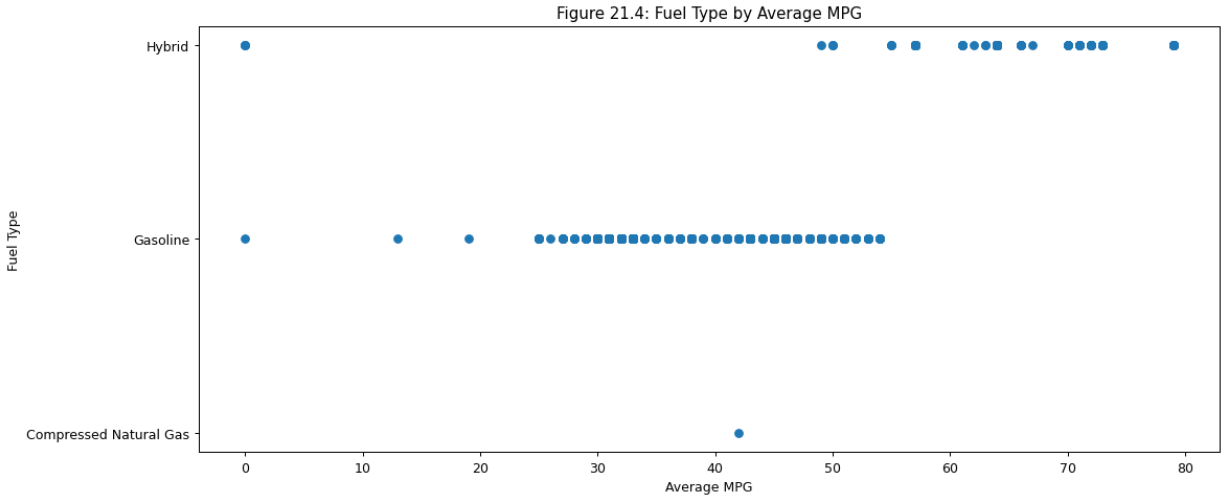


Figure 21.4 shows that the Hybrid engines have a significantly different range for MPG than standard gas models. Interestingly the single CNG engine has a relatively good MPG rating that’s about average compared to regular gas models.

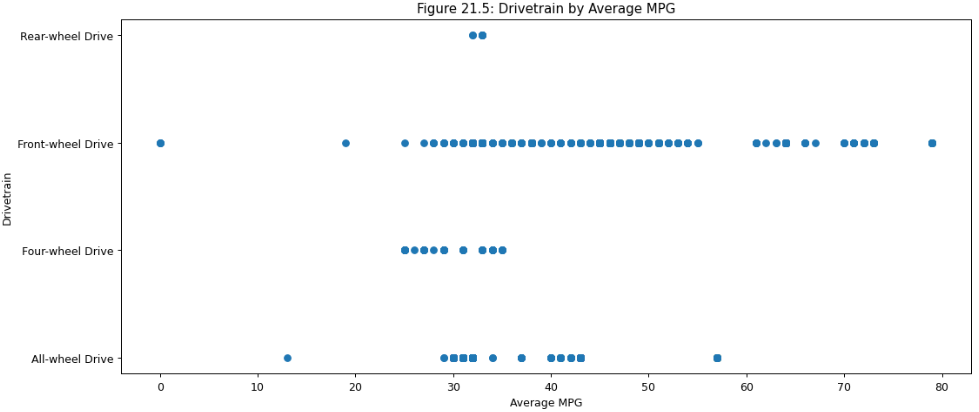


Figure 21.5 shows the front-wheel drive having a higher MPG band than all-wheel drive or rear-wheel drive.

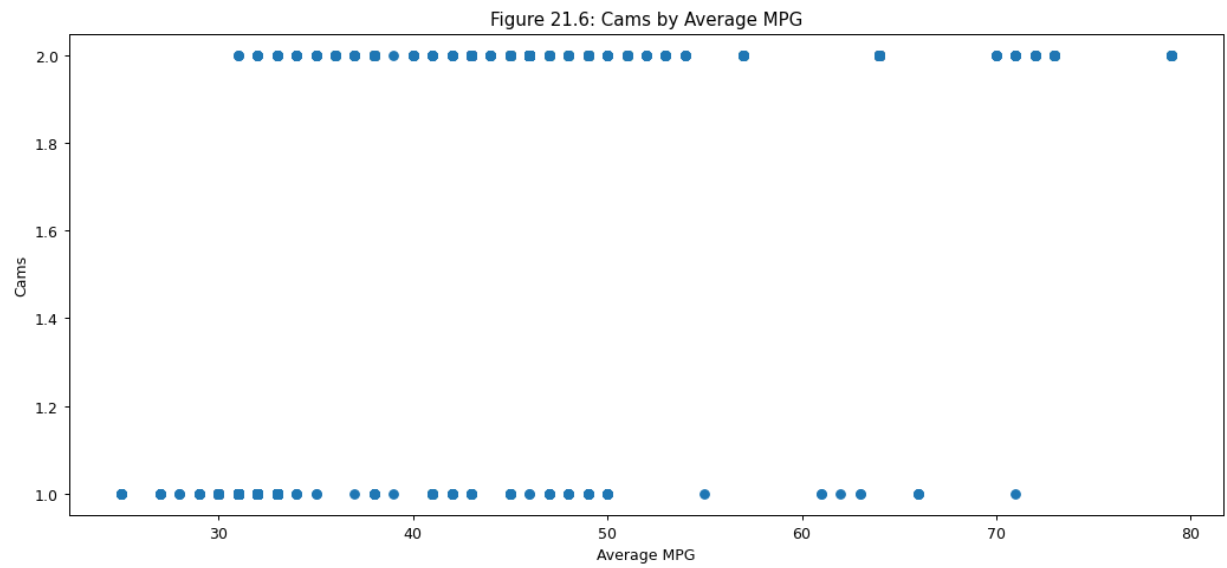


Figure 21.6 shows a slight but significant advantage with two-cams over one-cam engines.

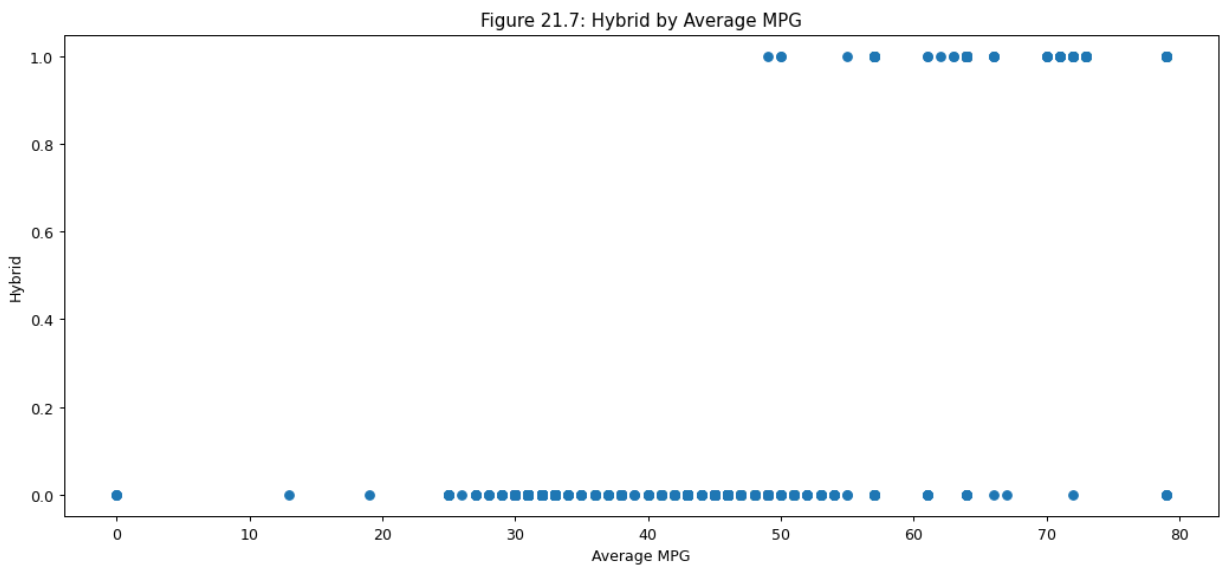


Figure 21.7 shows a higher affinity for higher MPG for hybrid engine models than all other engine types, further confirming Figure 21.4.

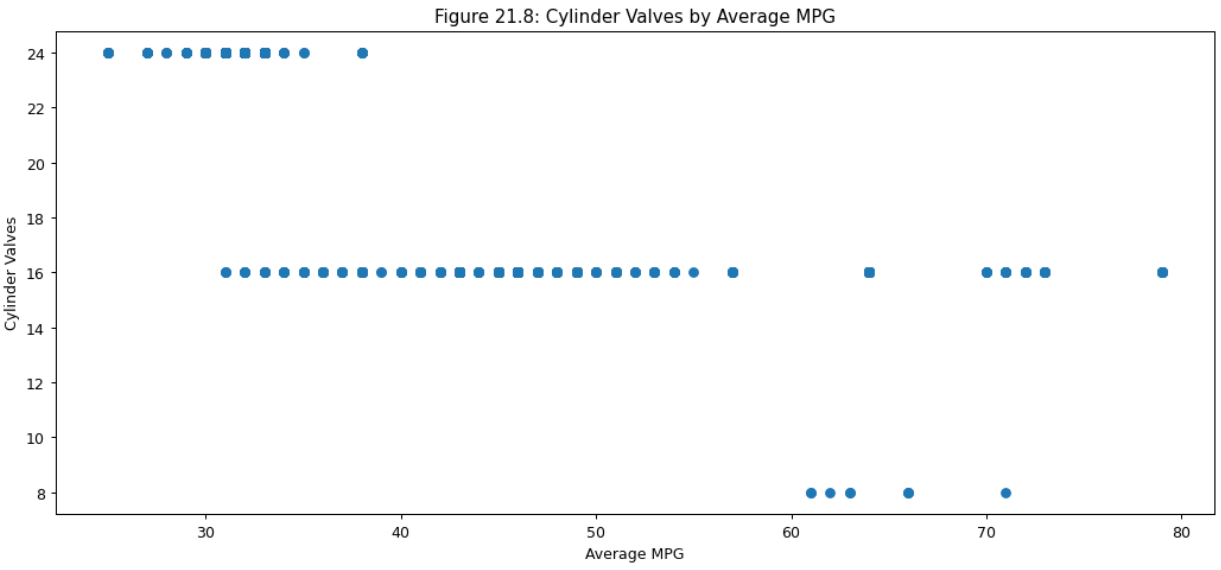


Figure 21.8 shows a relatively strong relationship between fewer valves and higher MPG. This is intuitive since the valves usually are associated with larger engines and more gas being injected.

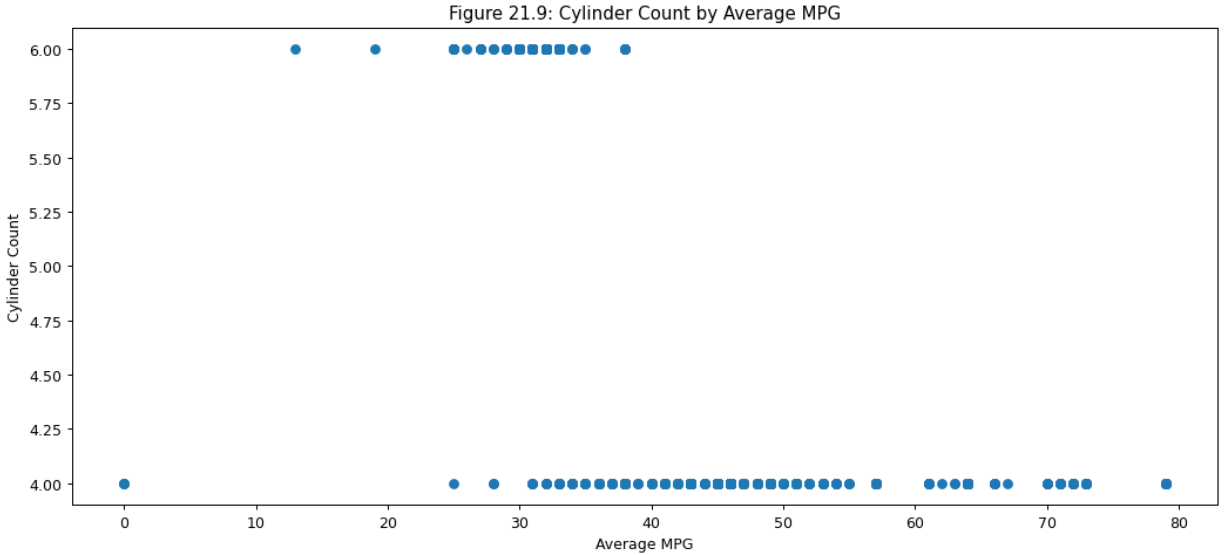


Figure 21.9 shows a much higher MPG for 4-cylinders than for 6-cylinders. This is also intuitive since more cylinders use more gas in exchange for providing more power.

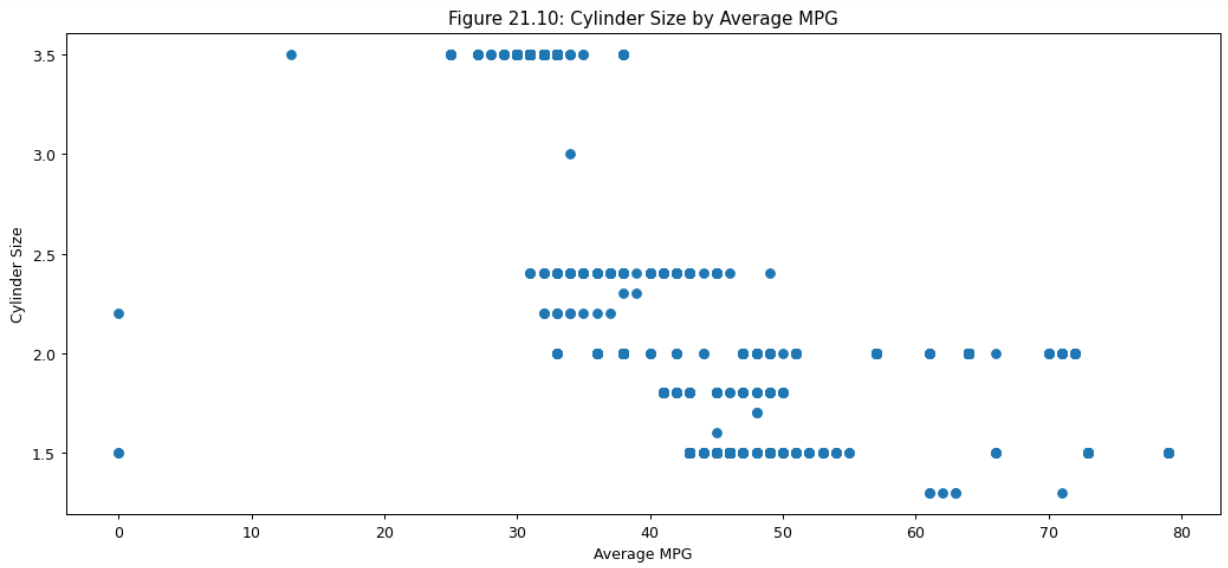


Figure 21.10 shows a clear positive relationship that lower cylinder volumes will produce higher fuel efficiency. This is also intuitive, given that a larger cylinder volume will require more gas to burn.

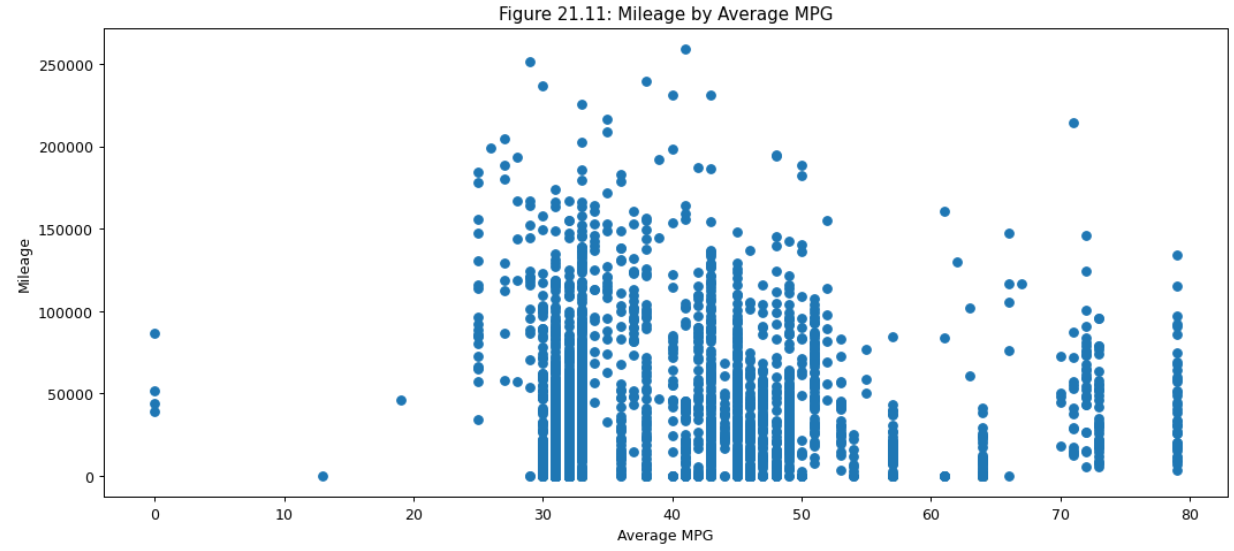


Figure 21.11 shows a somewhat negative relationship between MPG and mileage. This could be due to the relative newness of the Hybrid / CVT engines, which have higher mileage but have not been on the market as long.

Figure 22 shows that most all-wheel drive vehicles have a fuel efficiency between 31 and 32 MPG, with a few in the 46 MPG range and even less in the 61 MPG range. Front-wheel drive vehicles have a wider spread of fuel efficiency ratings between 32 and 64 MPG.

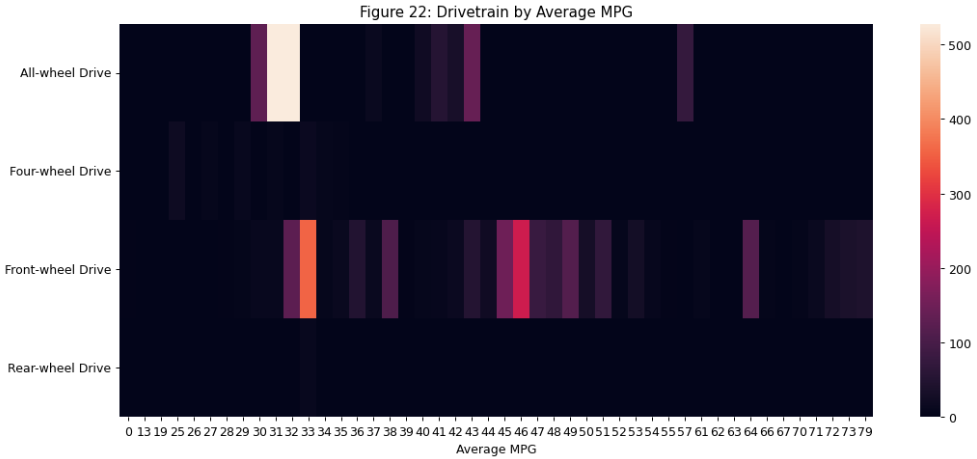


Figure 23 shows that gas engines predominantly have a range of 30 to 34 MPG but also a significant band in the 43 to 49 MPG range. Hybrid engines tend to center on the 57 to 64 MPG range.

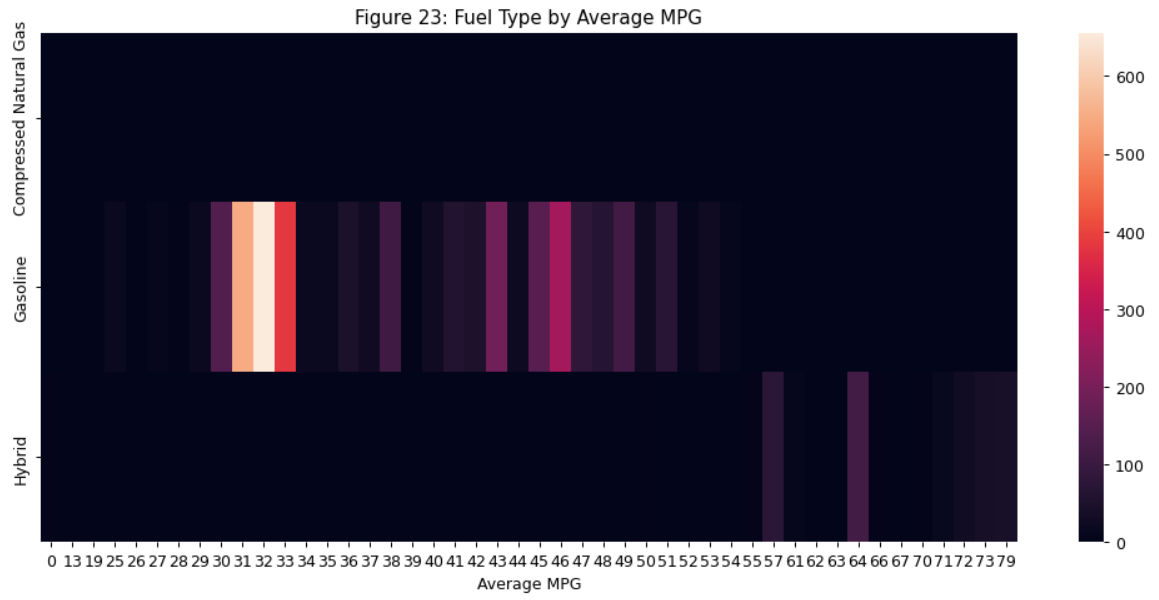


Figure 24 shows a cluster of models including the Pilot, Passport, Odyssey, and Ridgeline, in the 30 to 33 MPG range. There’s another cluster from the 37 to 64 MPG range for the Civic, CR-V, and Accord models.

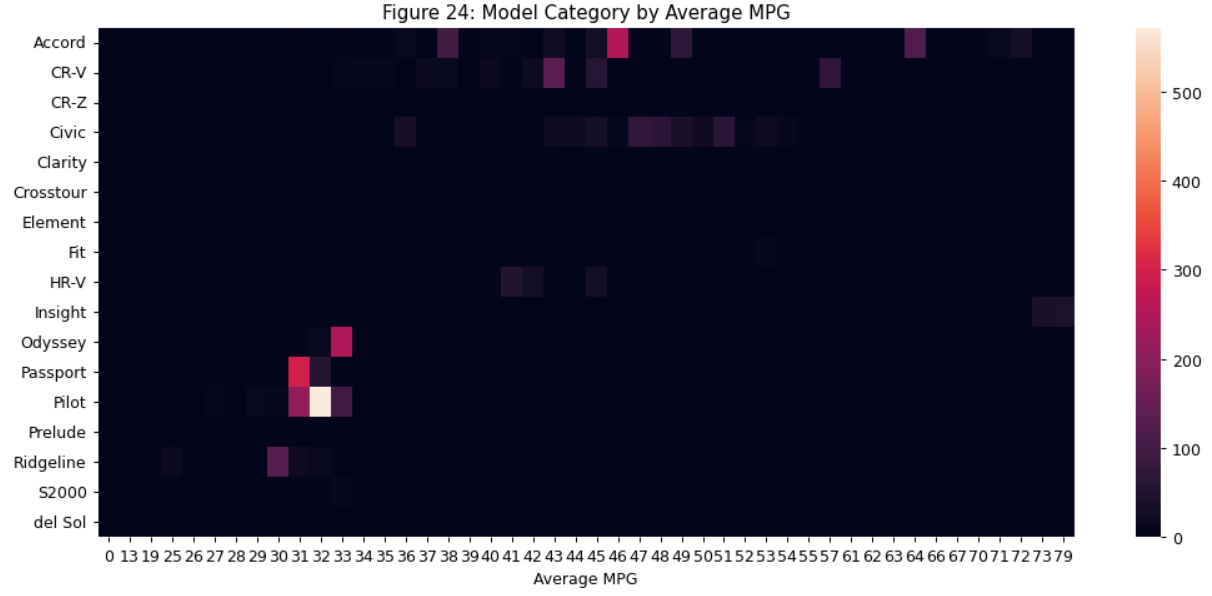
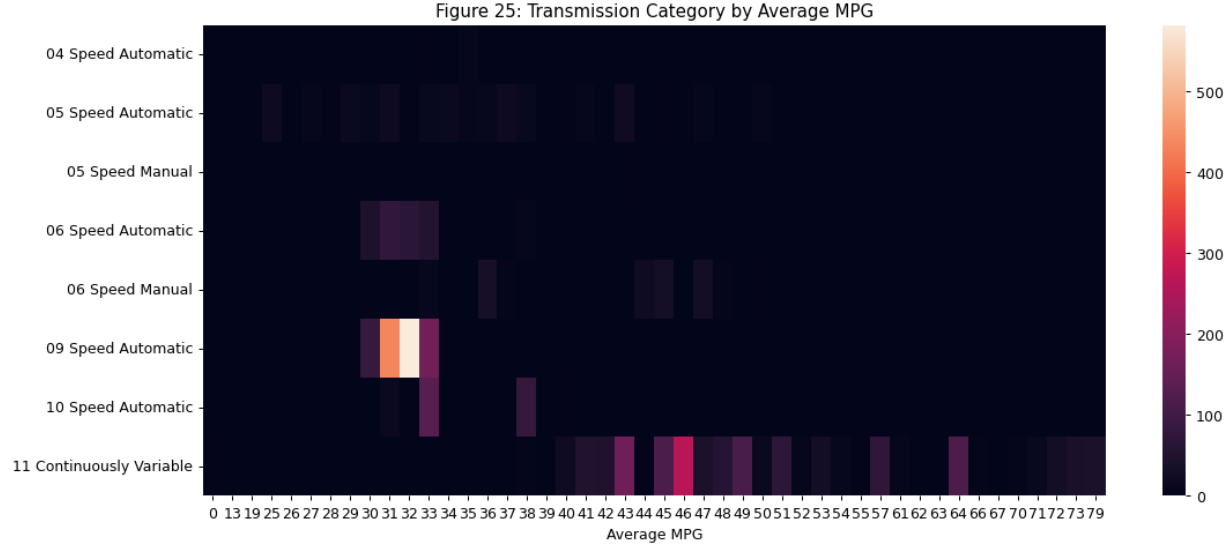


Figure 25 shows a dense cluster of 9-speed automatics around the 31 to 33 MPG range. There’s another less dense cluster between 43 to 64 MPG for the CVT engines.



Figures 26.1 to 26.17 shows the changes in MPG per year model.

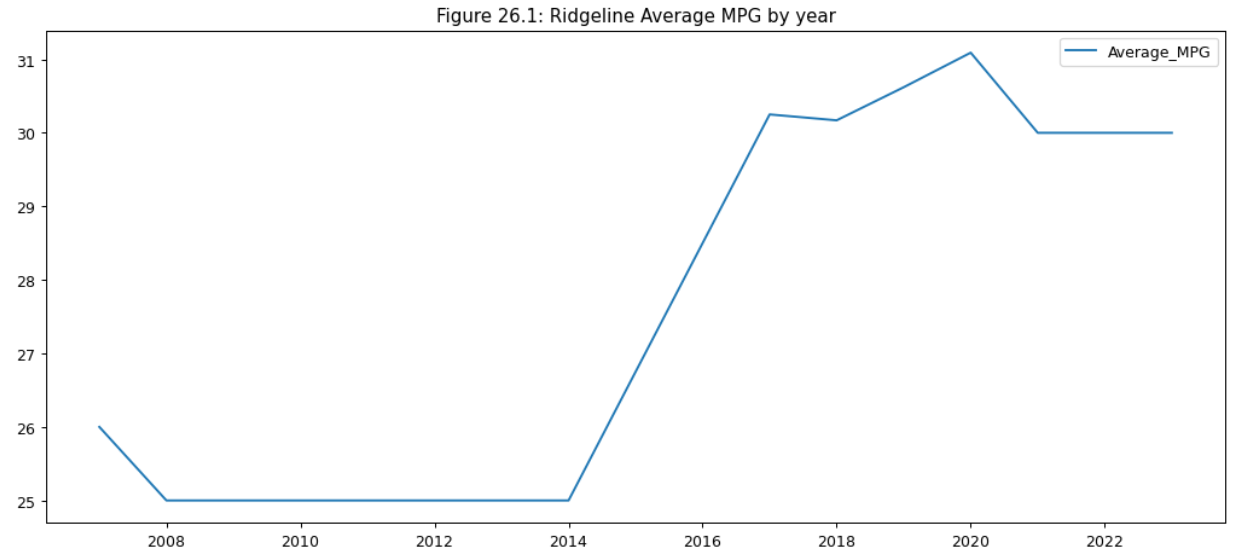


Figure 26.1 shows a significant increase in fuel efficiency for the Ridgeline in the most recent few years.



Figure 26.2 shows a dip in fuel efficiency for the Odyssey from 2008 to 2009 but generally a steady 33 MPG overall.

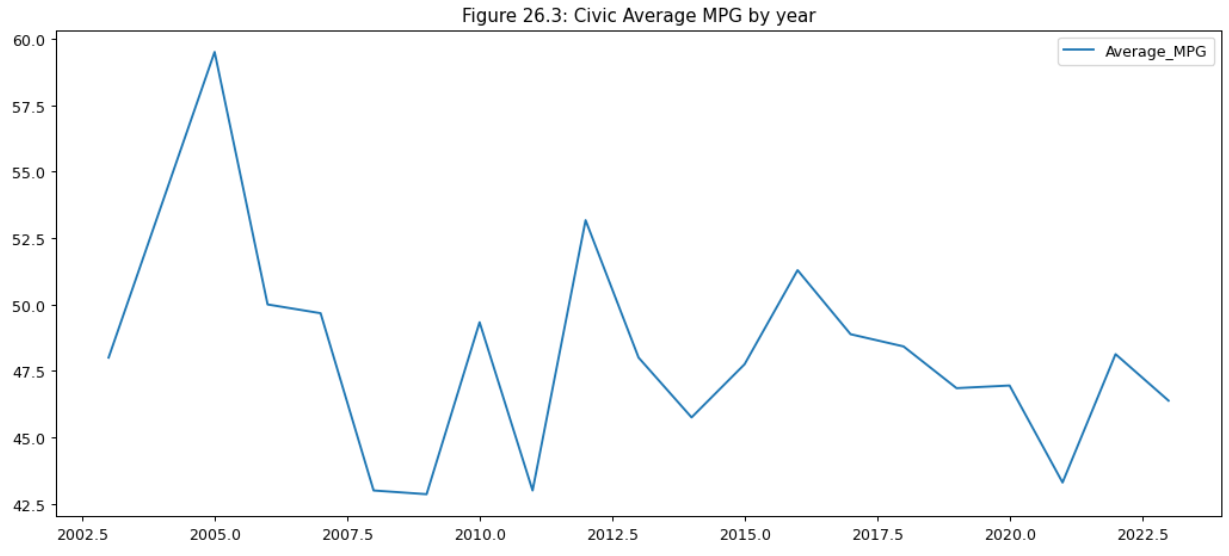


Figure 26.3 shows a relatively high fuel efficiency over the life of the Civic, with an average of around 46 MPG.

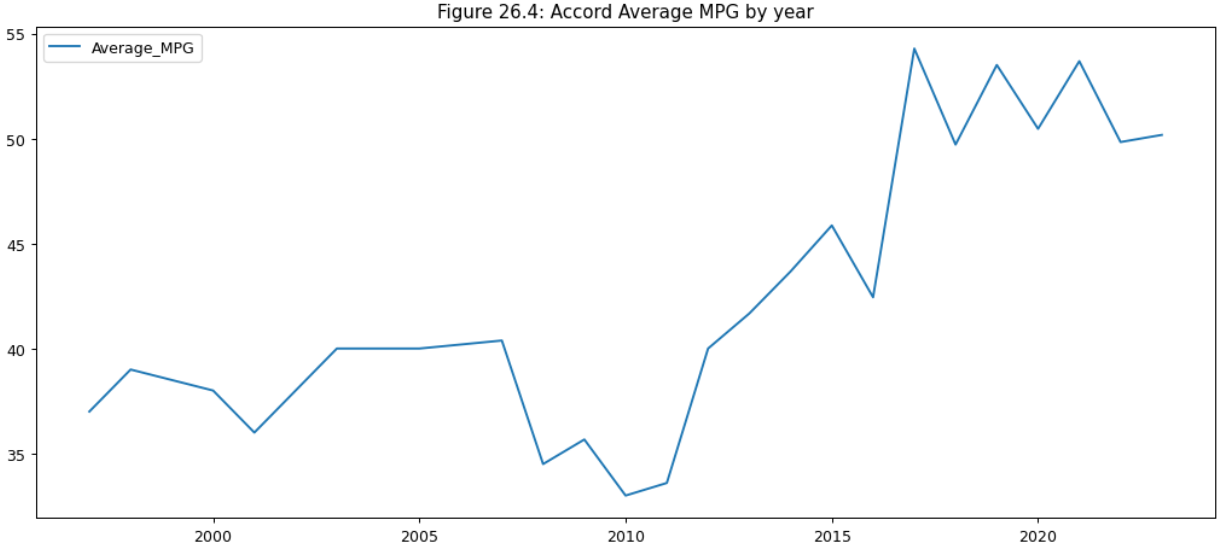


Figure 26.4 shows a significant increase in fuel efficiency for the Accord around 2015 to a high of 54 MPG.

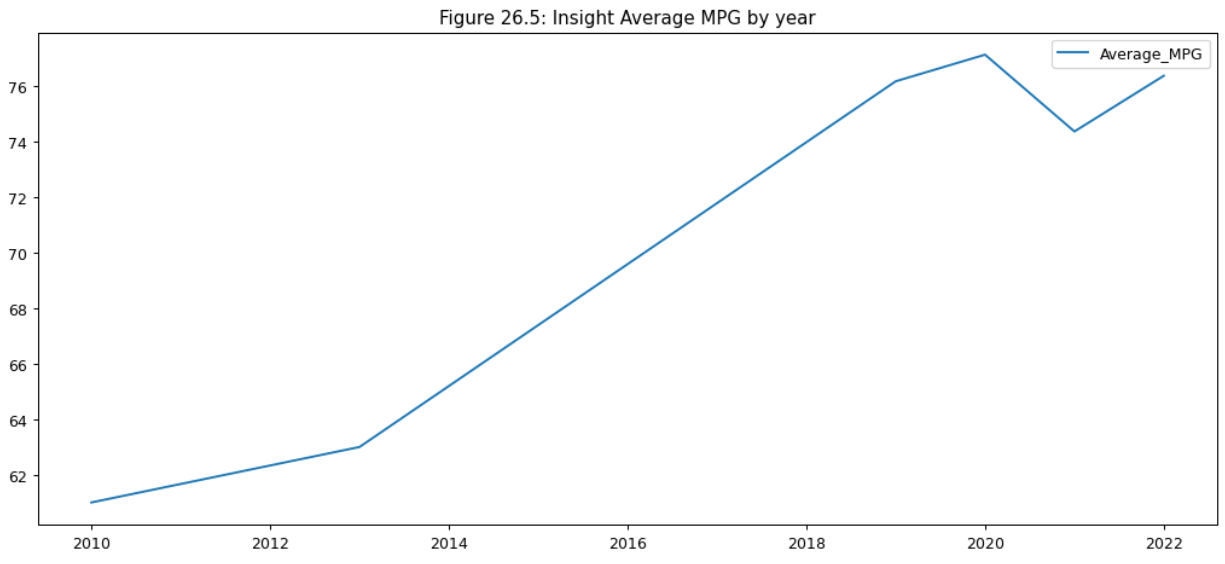


Figure 26.5 shows a steady increase in fuel efficiency for the Insight from an initial high of 61 up to 77 MPG.

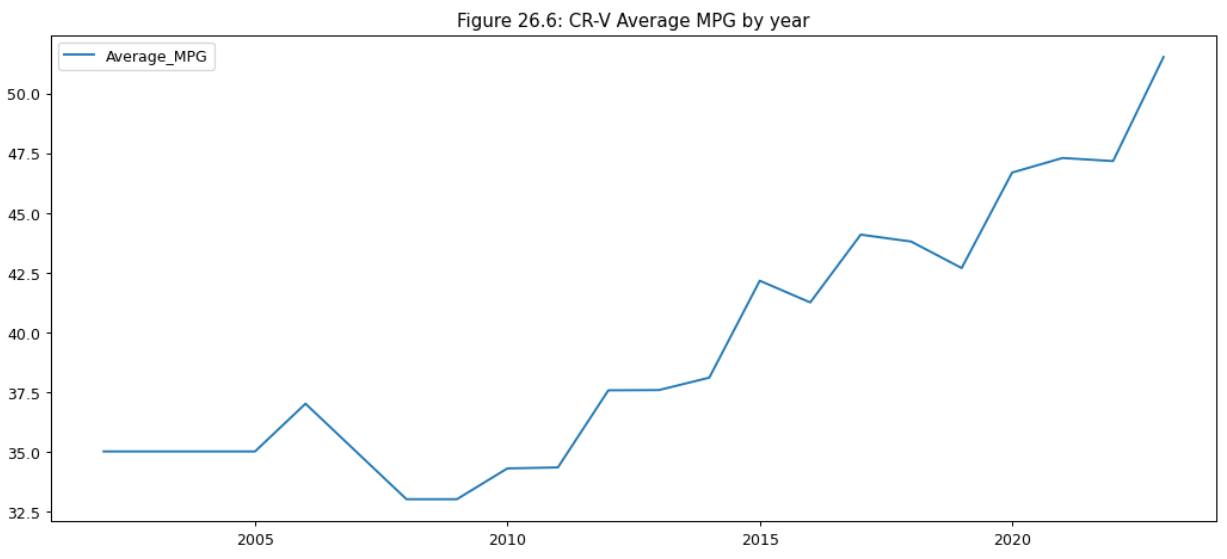


Figure 26.6 shows a slow start, then a steady increase in fuel efficiency for the CR-V starting around 2010 up to 51 MPG.

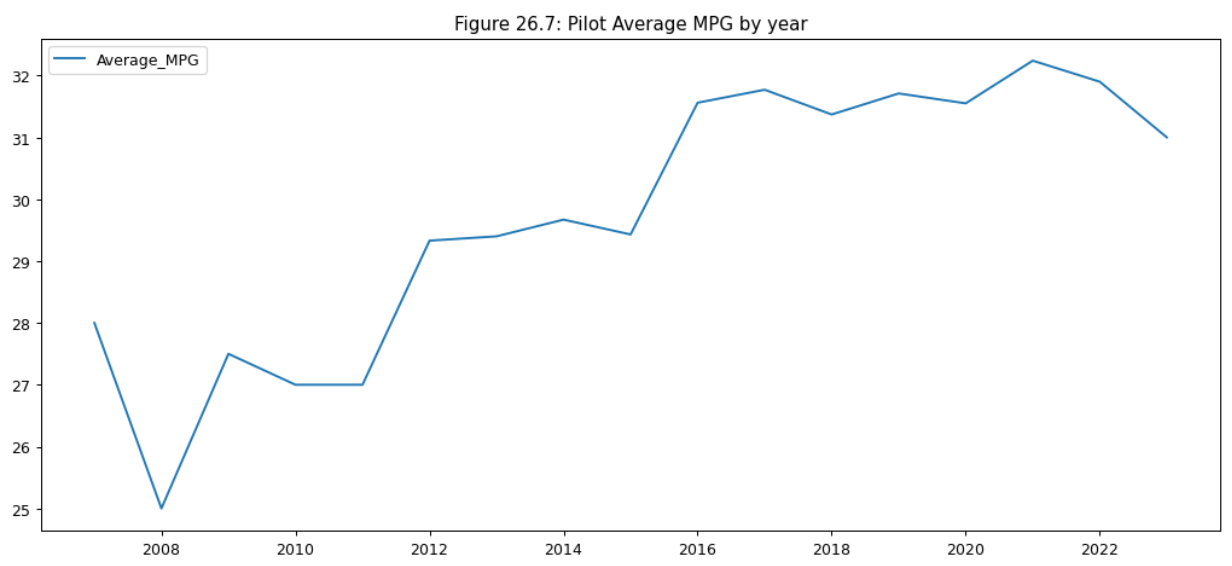


Figure 26.7 shows a steady increase in fuel efficiency for the Pilot leveling off at 32 MPG.

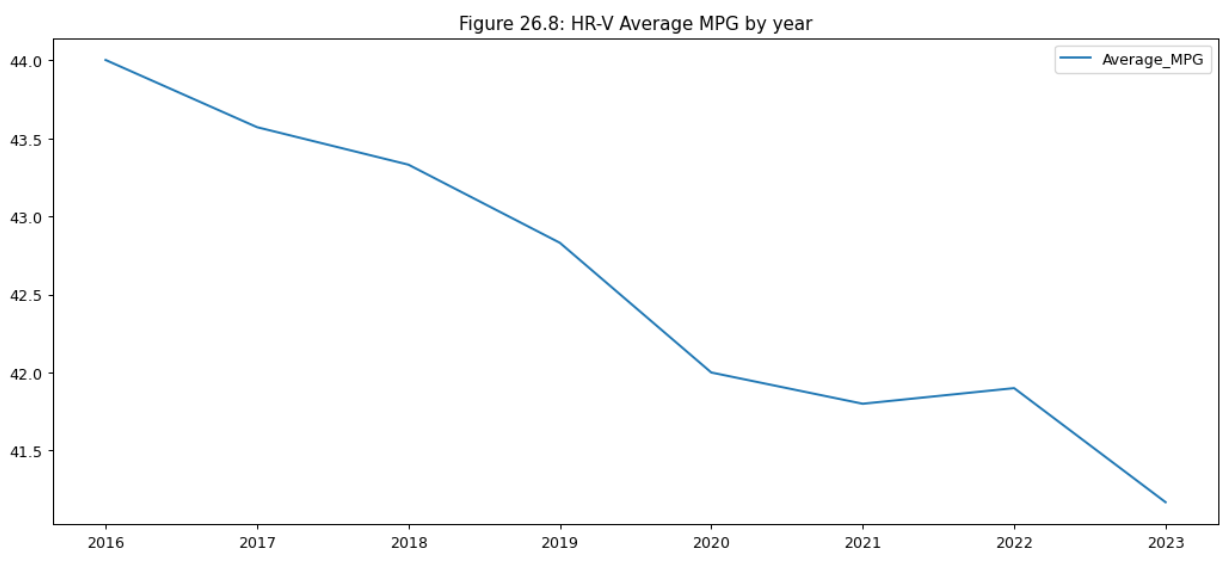


Figure 26.8 shows a steady decrease in fuel efficiency for the HR-V from a high of 44 MPG to a low of 41 MPG.

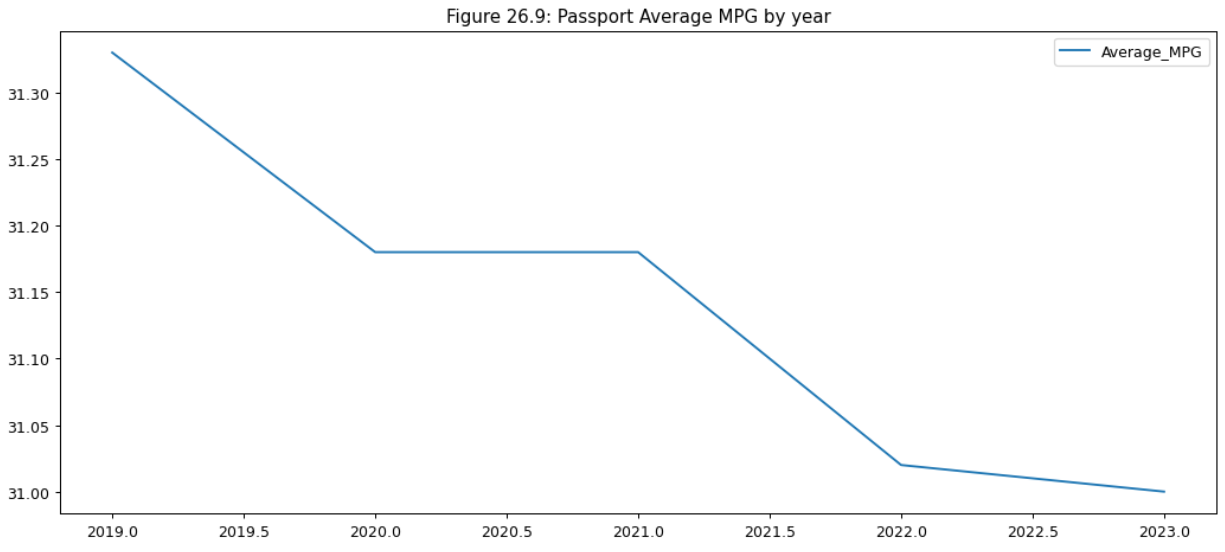


Figure 26.9 shows a steady decrease in fuel efficiency for the Passport, but the drop is a very small range from 31.3 to 31 MPG. The decrease can be attributed to random noise.

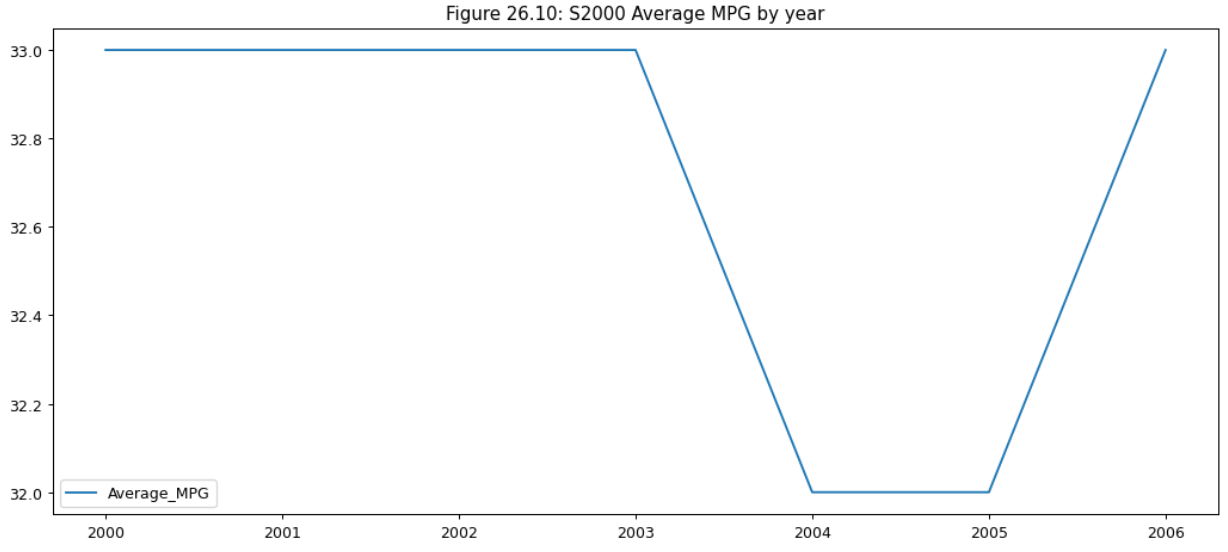


Figure 26.10 shows a relatively flat level of fuel efficiency for the S2000 for the short period of time it was available, with a mean of around 32.5 MPG.

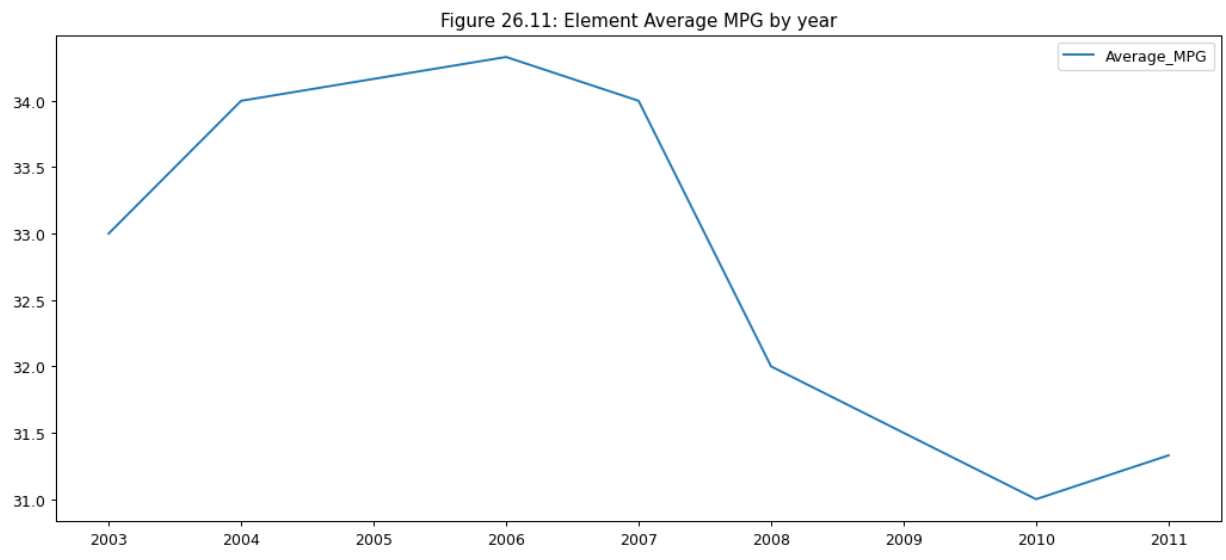


Figure 26.11 shows a bit of a drop in fuel efficiency for the Element around 2007-2008, leveling off at around 31 MPG.



Figure 26.12 shows a marked increase in fuel efficiency followed by a slow decline for the Crosstour, leveling off at around 34 MPG.

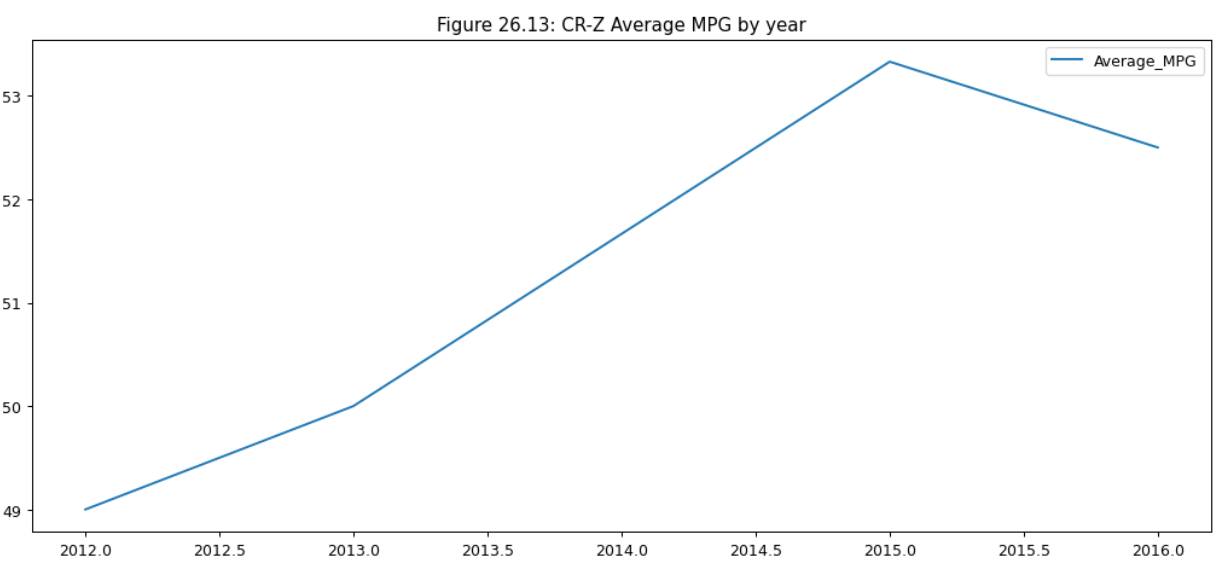


Figure 26.13 shows a steady increase in fuel efficiency for the CR-Z over a short period of time, followed by a small dip leveling off at around 52 MPG.

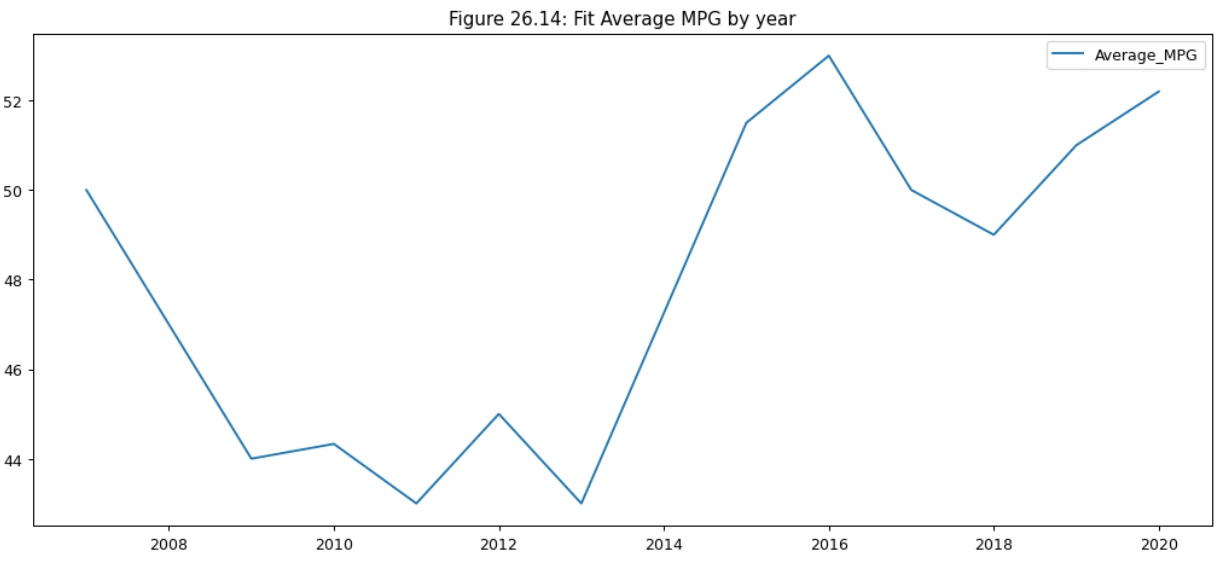


Figure 26.14 shows a rocky ebb and flow in fuel efficiency for the Fit. It started around 50 MPG and then dropped to 44 MPG, then back up to around 52 MPG.

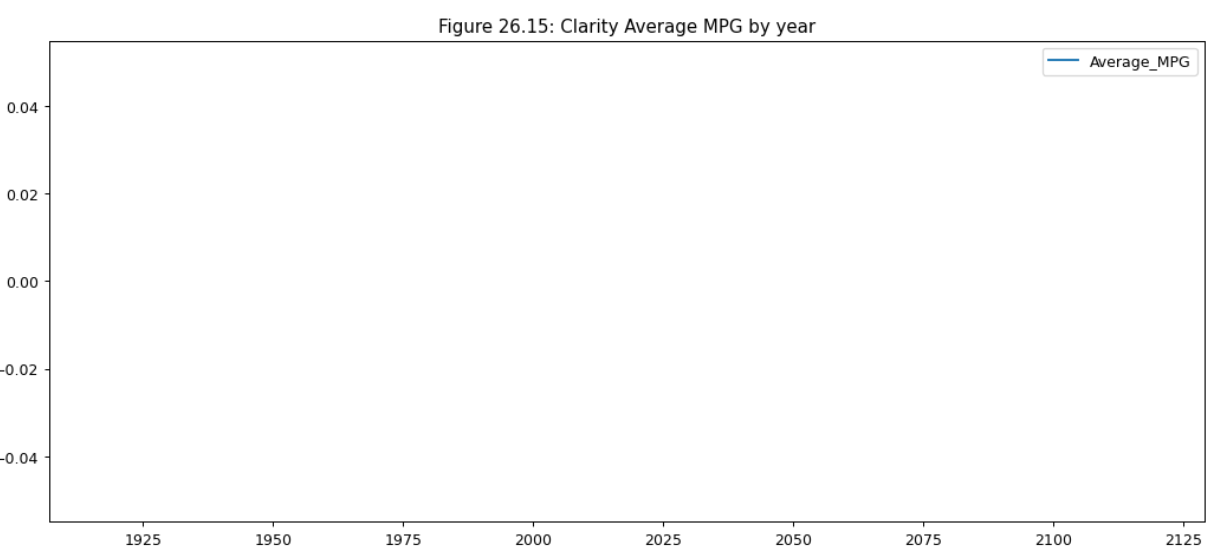


Figure 26.15 doesn’t have any records containing MPG data for the Clarity.

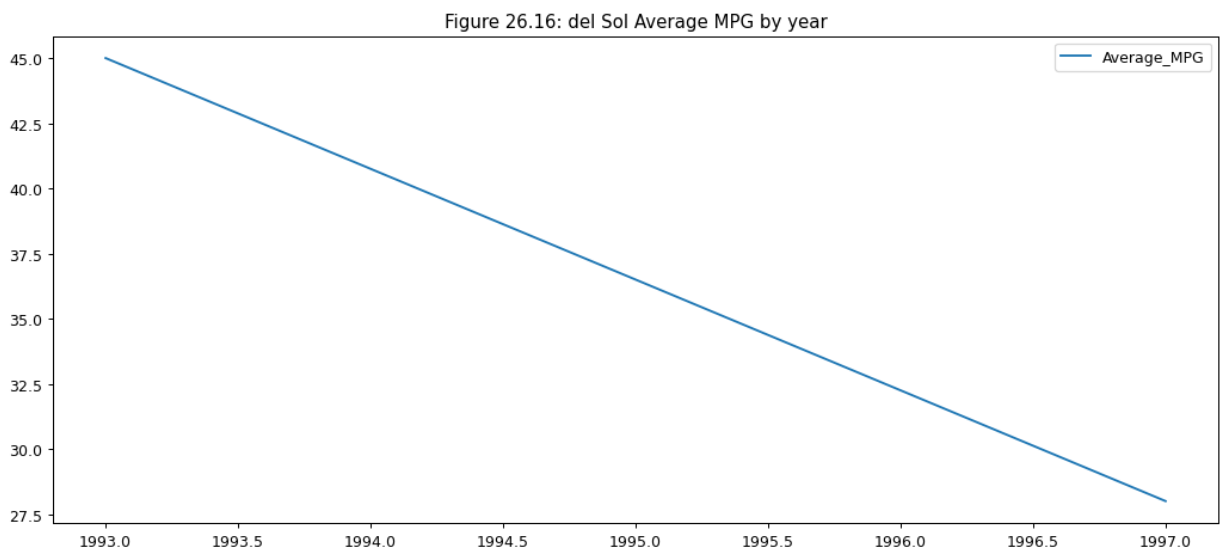


Figure 26.16 shows a steady decline in fuel efficiency for the Del Sol, falling from a peak of 45 MPG to a low of 28 MPG.

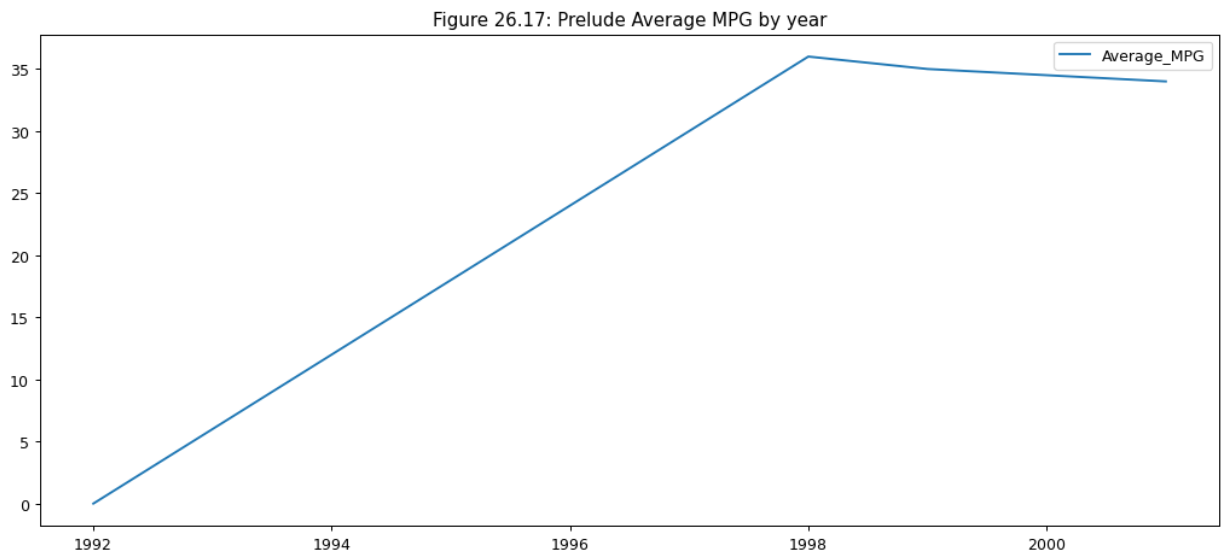


Figure 26.17 shows a steady increase in fuel efficiency for the Prelude, leveling off at around 35 MPG.

## Question 4: Which models have the best ratings?

In the analysis of the ratings, we start by looking at the distribution of the ratings themselves. Figures 27.1 to 27.8 break down each of the ratings and rating numbers.



Figure 27.1 shows that most consumer ratings were 4 or above. Few were low-rated, but there is a long tail.

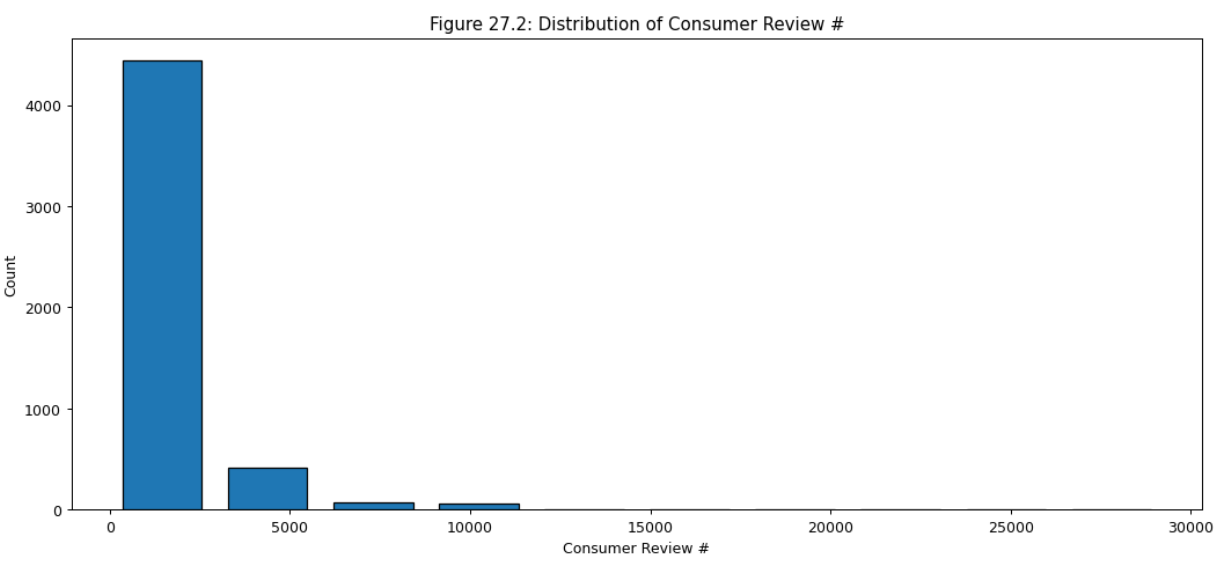


Figure 27.2 shows that most models had less than 5000 votes.

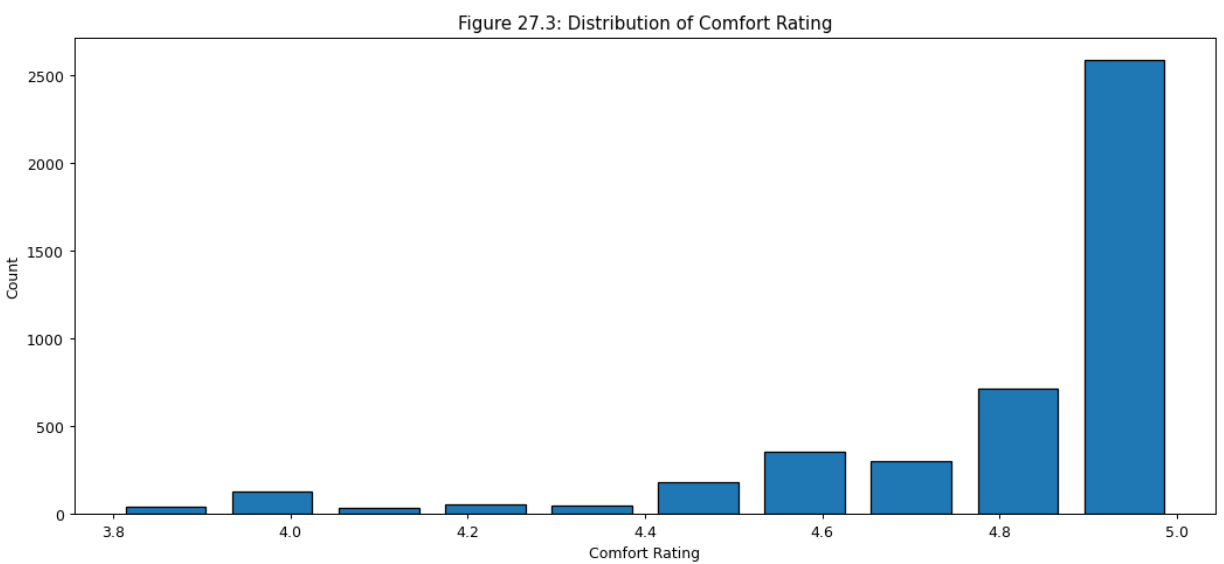


Figure 27.3 shows that most comfort ratings were above 4.4. There were a few lower ratings but never reached below 3.8.

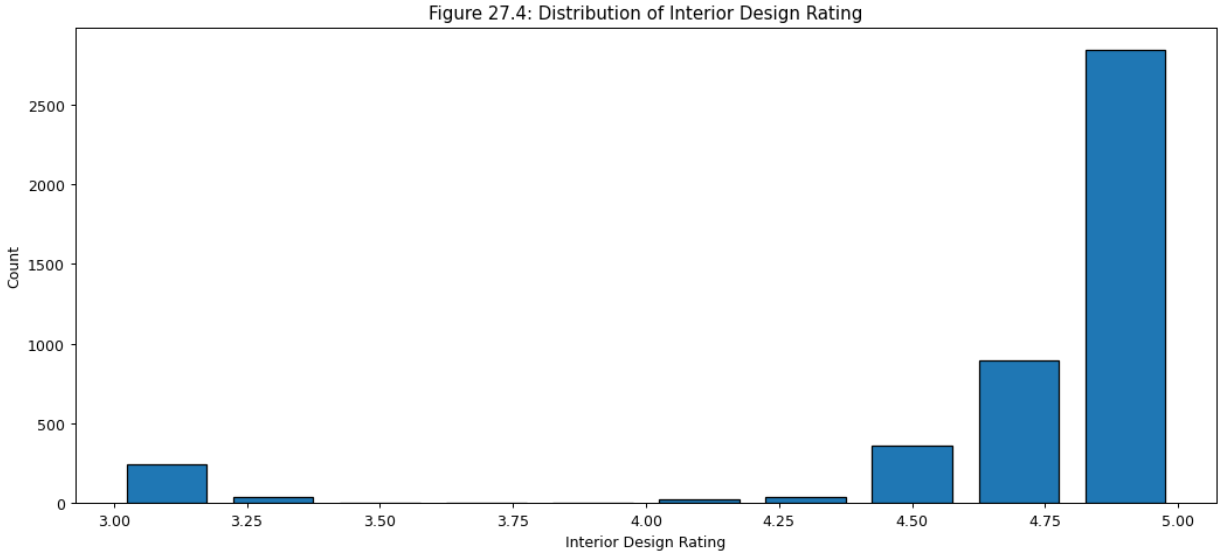


Figure 27.4 shows that most interior design ratings were above 4.5. There were a few lower ratings but never reached below a 3.

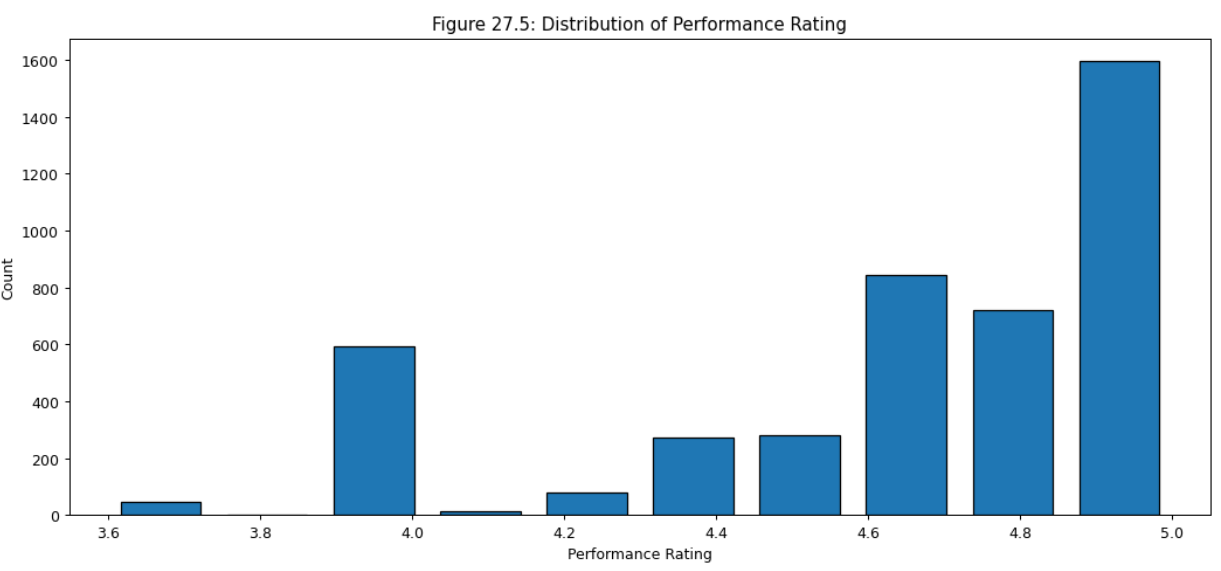


Figure 27.5 shows that most performance ratings were above 4. There were very few lower ratings but never reached below 3.6.

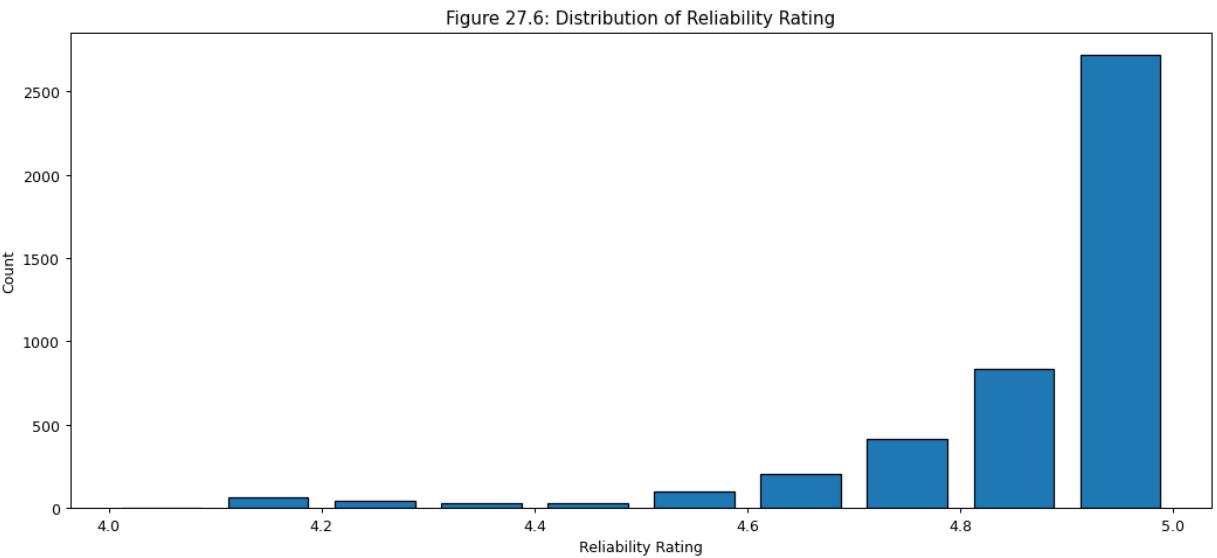


Figure 27.6 shows that most reliability ratings were above 4.6. There were a few lower ratings but never reached below 4.

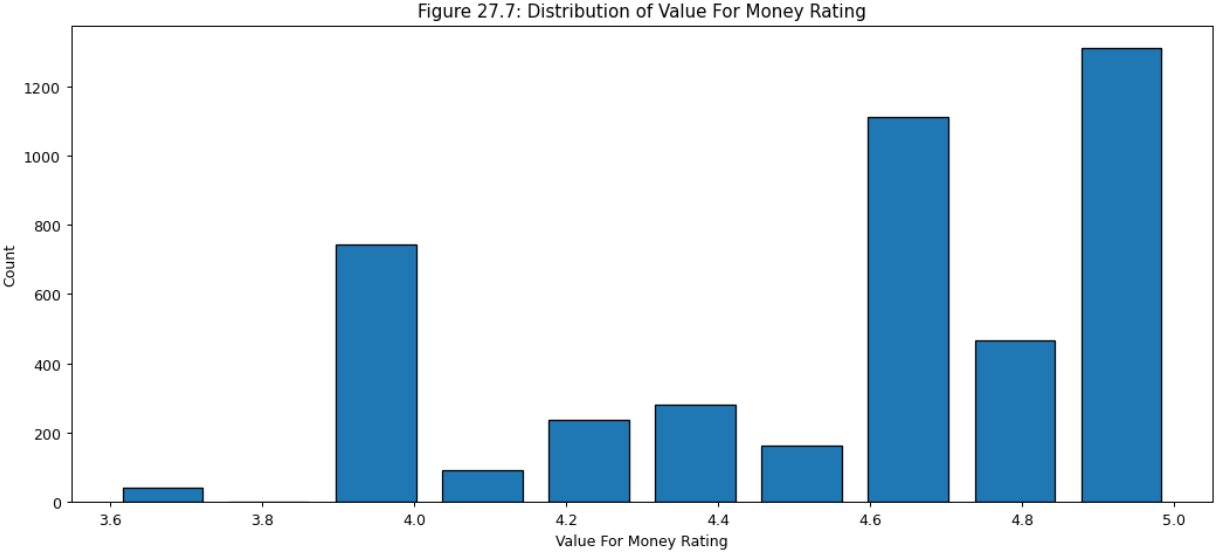


Figure 27.7 shows that most value-for-money ratings were above 4. There were a few lower ratings but never reached below 3.6.

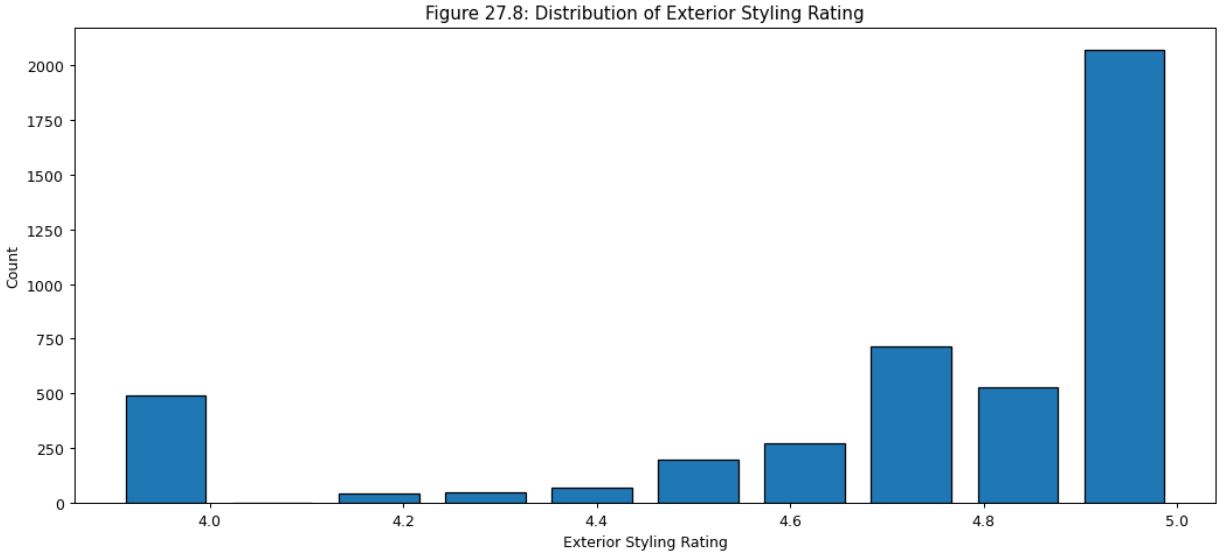


Figure 27.8 shows that most exterior style ratings were above 4.5. There were a few lower ratings but never reached below 4.

Figure 28 shows that consumer ratings have largely stayed high, with a few dips in 1990. These were inspected closer and shown to be due to two outlier values. More data points within this range would help to clarify if this was caused by actual issues or just statistical noise.

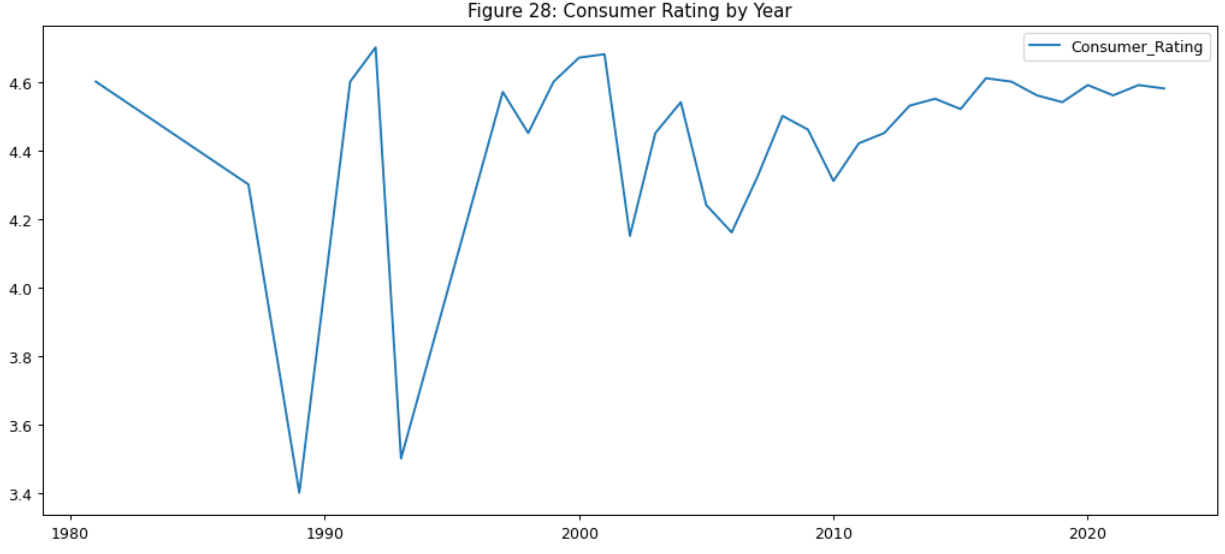
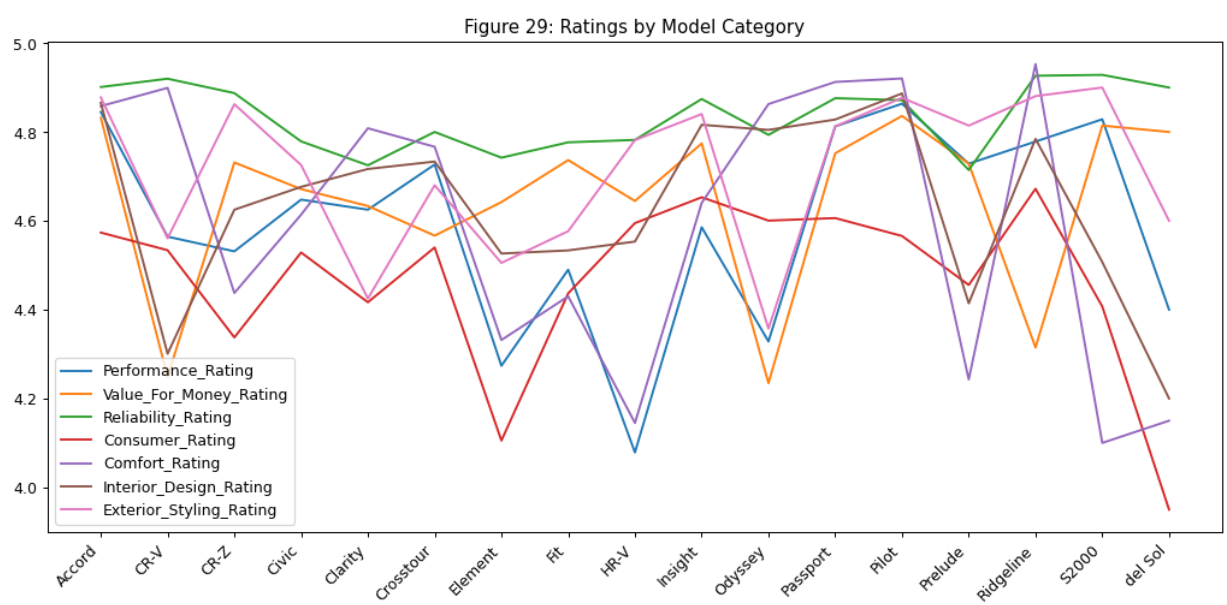


Figure 29 shows a combination of ratings for each of the models. There is a generally high rating for reliability across the board. There are significant dips for the CR-V, Element, HR-V, Odyssey, Prelude, and Del Sol. The Ridgeline seems to have mostly high ratings with a dip only in the value for the money rating. The Pilot and Passport have the highest overall ratings.



## Question 5: What are the best deals for different budgets ($10,000, $20,000, ~ $60,000)?

In the analysis for question 5, a closer look was taken at the models for different price ranges, comparing the MPG with the alpha values (darker = higher) indicating the mileage. The higher on the graph with a lighter point indicates a better value. Figures 30.1 to 30.6 show which models offer good values for each price point.

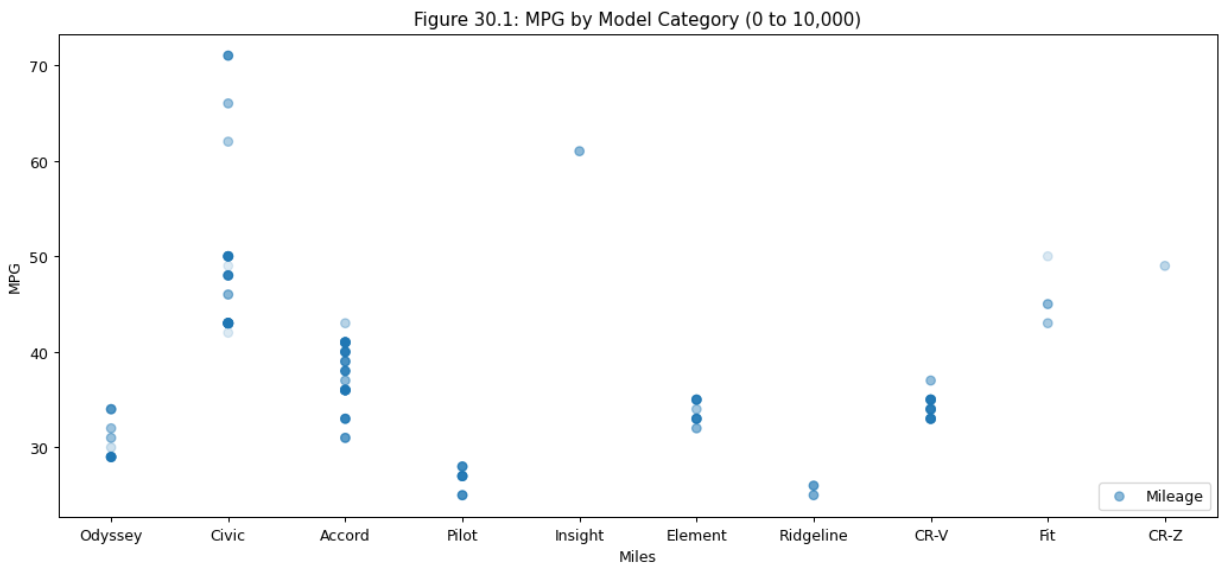


Figure 30.1 shows that the Civic, Insight, Fit, and CR-Z models would be a good deal in the 0 to $10,000 price range.

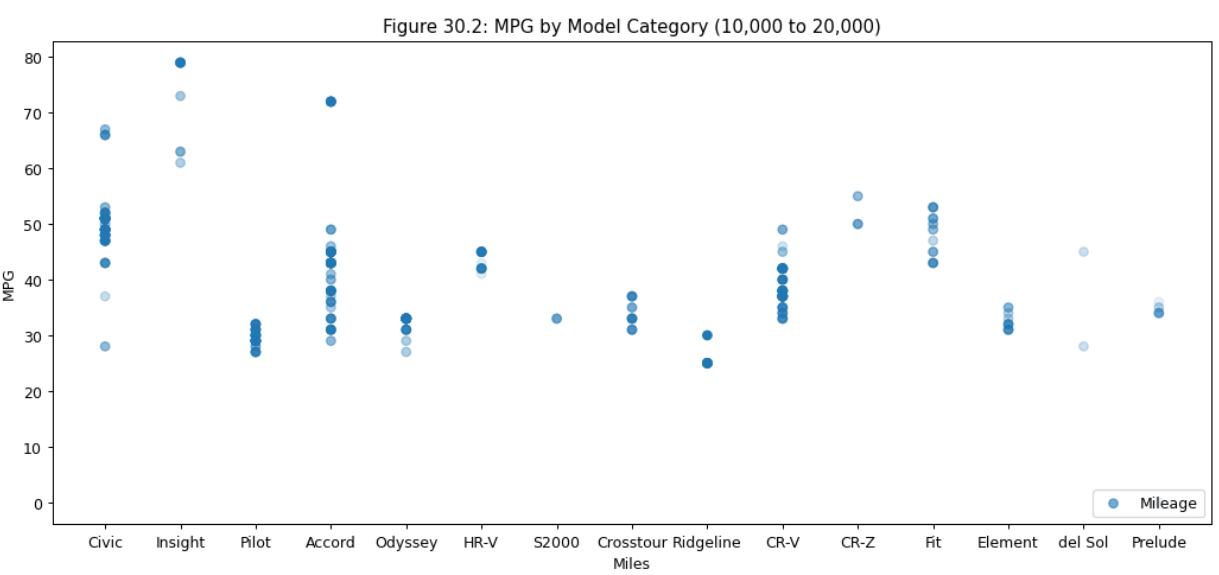


Figure 30.2 shows that the Civic, Insight, and CR-Z models would be a good deal in the $10,000 to $20,000 price range.

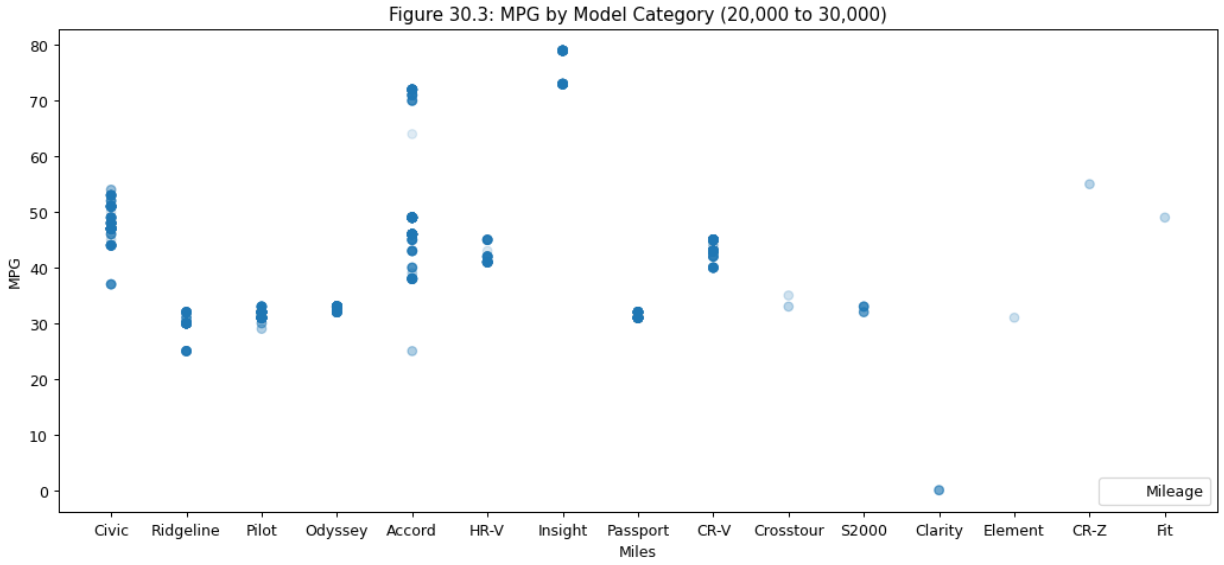


Figure 30.3 shows that a few Civics, Accord, Fit, and CR-Z models would be a good deal in the $20,000 to $30,000 price range.

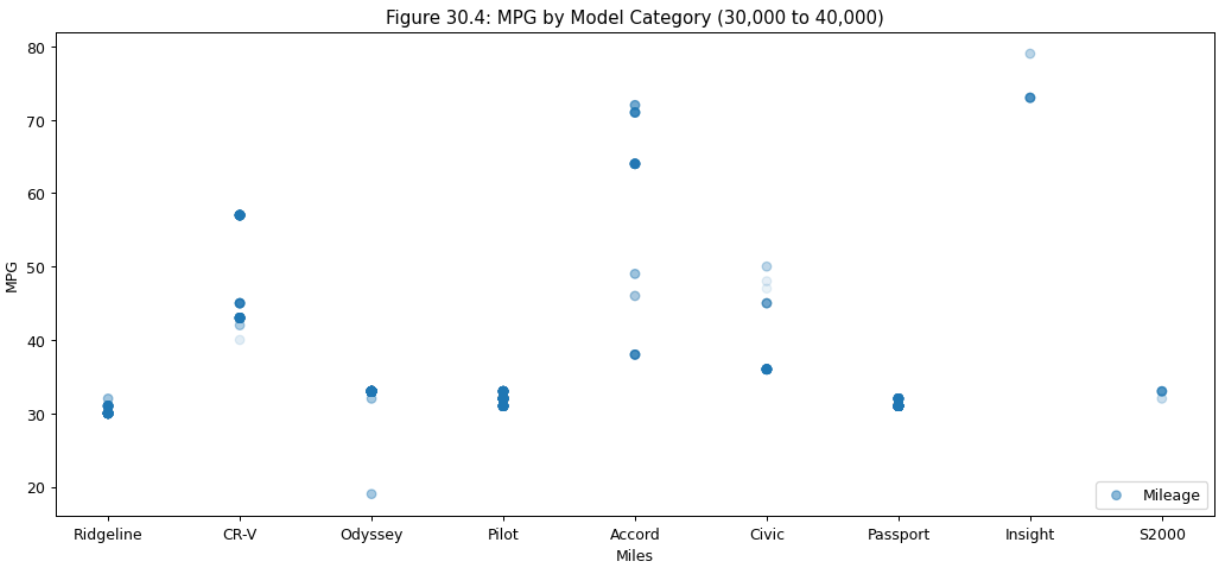


Figure 30.4 shows that Accord and Insight models would be a good deal in the $30,000 to $40,000 price range.

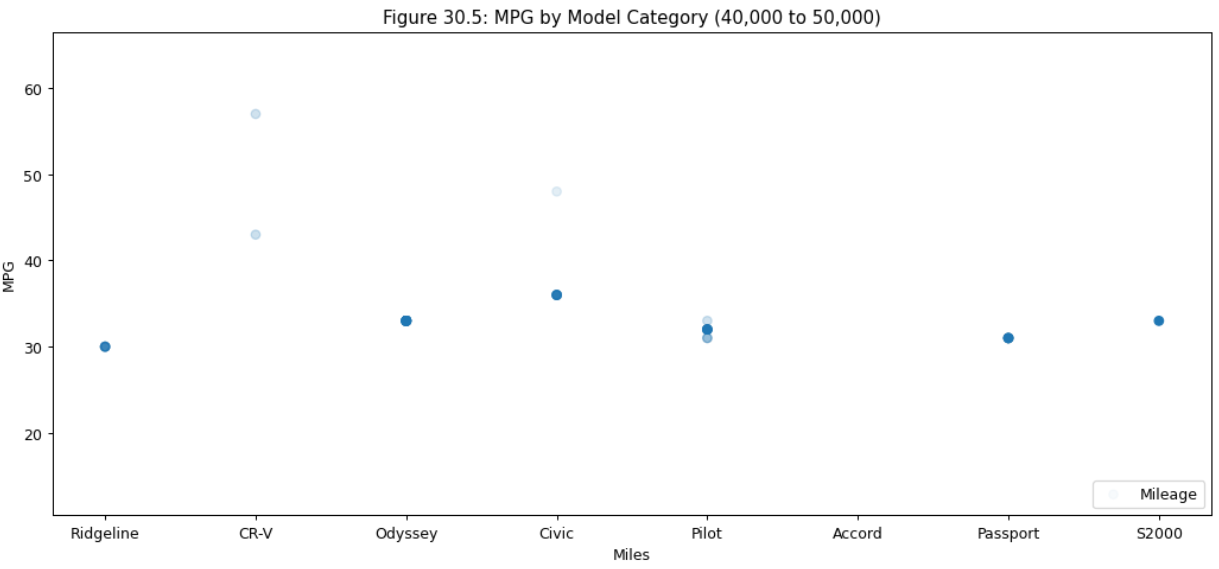


Figure 30.5 shows that CR-V and Civic models would be a good deal in the $40,000 to $50,000 price range.

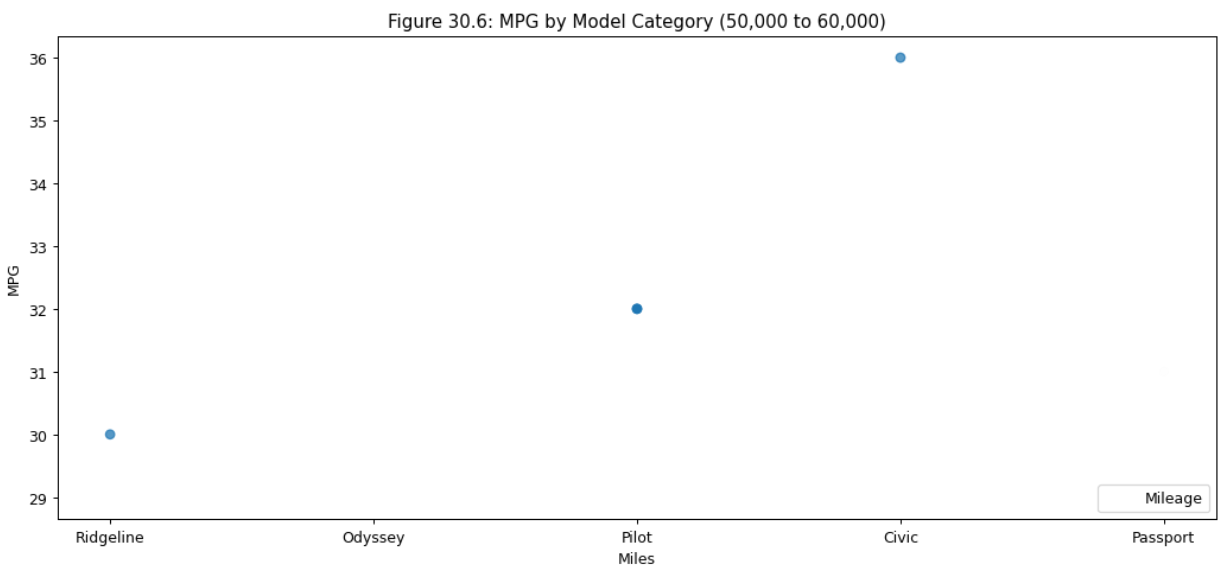
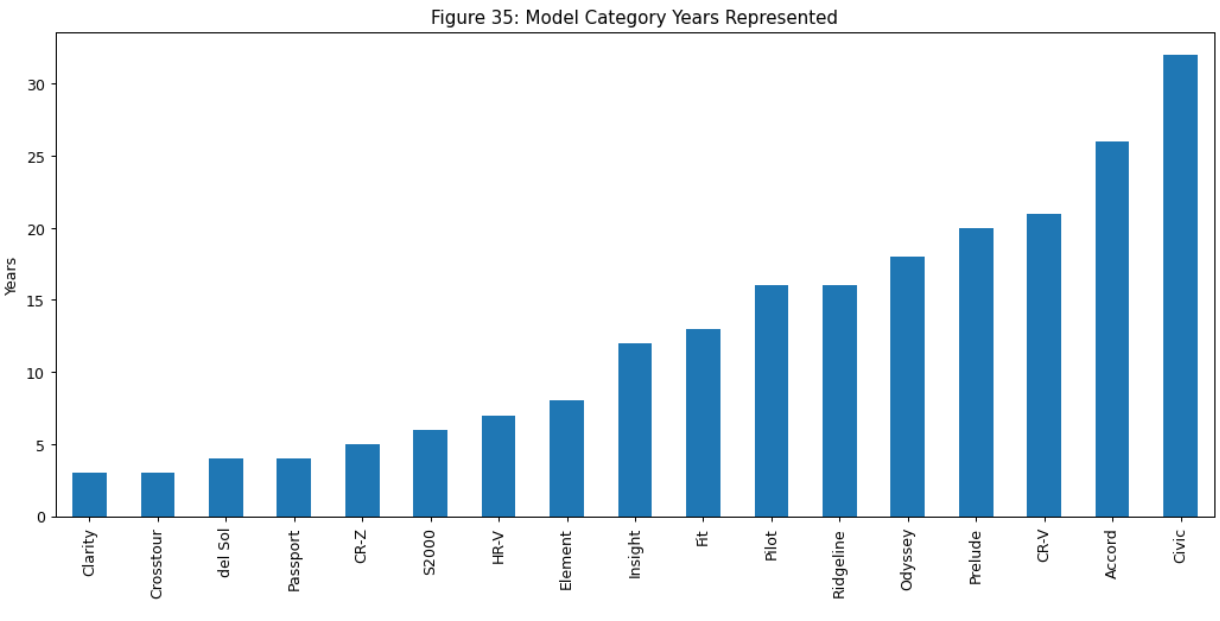


Figure 30.6 shows that the Civic model would be a good deal in the $50,000 to $60,000 price range.

## Question 6: Which models were sold for the most years?

For the analysis of the model duration, the calculation was straightforward. Figure 35 shows the range of years per model. The Civic is the longest-running model at 30 years, and the Clarity is the shortest at 3 years. These are not the actual model lifespan but just the ranges within the dataset of Honda cars being sold currently. About half the models are sold for in than 15 years.



# Conclusion

The data analysis focused on several main questions around car price, location, fuel efficiency, ratings, and market duration. This dataset isn’t a full spread of all Honda vehicles and only represents vehicles newly for sale or chosen models for resale. There are some gaps in the data that make true comparisons between engine or fuel types challenging, but there was sufficient data to make some valuable conclusions.

The car features that correlate to a higher price are 2 and 3.5 Liter engines, 6 cylinders, higher valve counts, a GDI engine, larger models like SUVs, lower mileage, lower age, all-wheel drive, higher gears, and interestingly yellow exterior paint job. There were also a few unintuitive results. The MPG was not a good predictor of price, which is interesting because that could potentially save a consumer money over the life of the car. Also, there was one location, Delaware, where the cost of a car was consistently much higher than everywhere else.

The cost for Honda cars per state showed a lot of similarities. There were a few states that stood out. Delaware had the highest average cost combined with a low average MPG would indicate that ownership in that state could be the worst. On the other hand, California had an average cost with many newer cars, front-wheel drive, and a higher than average MPG due to most likely the large number of hybrid vehicles there. This may indicate a lower overall cost of ownership. Rhode Island was the state with the lowest average cost per car by a wide margin.

When exploring how fuel efficiency has changed over the years, a few things stood out. There were several models that have improved over the years, like the Ridgeline, Accord, Insight, CR-V, CR-Z, Pilot, Fit, and Prelude. There were a few that declined over the years, like the HR-V, Element, and Del Sol. The highest correlation between fuel efficiency was with non-turbo engines, CVT engines, Hybrids, front-wheel drive, Dual Overhead Cams (DOHC), 4-cylinder engines and smaller volume engines. The models with the highest MPG were the Civic, Accord, Insight, and CR-V.

Car ratings showed mixed results over all the models, but there are significant dips for the CR-V, Element, HR-V, Odyssey, Prelude, and Del Sol. The Ridgeline seems to have the highest ratings, with the lowest rating in the value for the money rating. The Pilot and Passport have the highest overall ratings.

When looking at price ranges, the best deals by model for under $10,000 were Civic, Insight, Fit and CR-Z. The best deals by model between $10,000 to $20,000 were Civic, Insight, and CR-Z. The best deals by model between $20,000 to $30,000 were Civic, Accord, Fit, and CR-Z. The best deals by model between $30,000 to $40,000 were Accord and Insight. The best deals by model between $40,000 to $50,000 were Civic and CR-V. The best deal by model between $50,000 to $60,000 was the Civic. The Civic appeared to be a good deal at almost any range.

The last area of research was the lifetime of each model. The Civic was the model with the widest range of years. It had a 30-year span. Clarity is the model with only 3 years represented. About half the models had more than 15 years represented.

Overall, the analysis of the dataset showed that most of the cars being sold were with dealers selling new or reselling recent models. There are fewer records for older models, which is probably a natural artifact of fewer older cars surviving for a longer period. What we can’t determine from large quantities of numbers, we can infer based on what types of cars are still available. Older models that continue to hold their value, like the Prelude, show a type of quality not measured in the dataset.