On the Used Car Market: Why Do American Cars Fail to Hold Resale Value?

Emerson Fleming

9/10/2022

2000 Porsche 911 GT3

Intro

(1-2 Pages)

The question I address in my paper is car value retention. I investigate using the Car Sales data set, why some cars hold their value better than others. Throughout my research, I found that high-end German automobile manufacturers like Porsche, Mercedes-Benz and BMW hold their value better than American manufacturers. Additionally, I found that Japanese manufacturers like Toyota and Lexus also hold their value better than American cars.

Indeed, it can be argued brand reputation and status determine how well vehicles resell. Japanese manufacturers are known to be reliable. Although they are less expensive overall, many buyers are drawn to them because of their reputation for reliability. German automotive manufacturers are known for their sophistication, performance and status. Though German cars are more expensive and less reliable, they have a solid

amount of demand because of their speed and sophistication. Both Japanese and German cars have a niche. What they do, they do very well.

According to my findings, American cars struggle to maintain value because they are in between the ideal thresholds for horsepower and initial price. Affordable Japanese manufacturers do not drive up unnecessary cost on horsepower for affordable vehicles. Instead, affordable Japanese vehicles tend to offer only the horsepower necessary to maximize affordability. German manufacturers offer exceptional horsepower at an optimal price for expensive vehicles. More premier Japanese manufacturers like Lexus and Infinity offer expensive vehicles with great performance much like German manufacturers. As a result, both Japanese and German automotive manufacturers have great value retention; Japanese and German manufacturers resell for more because they produce vehicles in ideal thresholds for the price and horsepower.

I hypothesized that American manufacturers struggle because they are outside the ideal thresholds for horsepower and initial price. As my findings demonstrate, there is a niche for affordable cars and a niche for expensive cars. American cars fetch lower resale prices because they fail to offer optimal affordability and performance. Instead, they exist somewhere in between.

Data

```
myDataPath %>%
  select(Manufacturer, Model, Sales_in_thousands, X__year_resale_value) %>%
  arrange(desc(Sales_in_thousands, X__year_resale_value)) %>%
  slice(1:10)
```

```
##    Manufacturer      Model Sales_in_thousands X__year_resale_value
## 1          Ford   F-Series            540.561                15.075
## 2          Ford   Explorer            276.747                16.640
## 3        Toyota      Camry            247.994                13.245
## 4          Ford     Taurus            245.815                10.055
## 5         Honda     Accord            230.902                13.210
## 6         Dodge Ram Pickup            227.061                15.060
## 7          Ford     Ranger            220.650                 7.850
## 8         Honda      Civic            199.685                 9.850
## 9         Dodge    Caravan            181.749                12.025
## 10         Ford      Focus            175.670                    NA
```

In this data set, we are given 157 different car models across 30 manufacturers with 16 different variables (columns). Variables include the number of sales, resale value, price, fuel efficiency, and more. My research indicates this data set was compiled using data from year 2000.

While attempting to input market price information for the BMW 3-Series, I found the year used to generate the data set. To find the exact year, I began by using obscure models in my data set. The last year of production for the BMW 323i occurred in year 2000 (Cars). Additionally, the last year of production for the Chrysler Cirrus also occurred in year 2000 (Cars). At this juncture, I searched MSRP values for year 2000 across various car databases and finally found that "Price_in_thousands" was generated using the lower end estimate from cars.com. At this point, I was certain **the price for each car was calculated using data from year 2000**.

Overall, while it is true the data is over 2 decades old, this data set is still viable and applicable to contemporary sales. For instance, of all auto manufacturers, Porsche still has the most exceptional residual value (Business Insider). Of all vehicles on this data set, the Porsche 911-Series also had the highest value retention of any car.

The Porsche 911 Coupe has the highest resale value of any car for the 2019 year (Business Insider). Additionally, 70% of the top 10 vehicles with the highest resell value for year 2019 are German or Japanese (Business Insider). Therefore, even now, German and Japanese automobiles continue to have higher resell prices overall than American cars.

Ultimately, I wish to assess why certain vehicles resell for the most and to prove numerically why American models fail to hold their value as well as their contemporaries. To construct my argument, I will use the following variables.

Variables:

Manufacturers: Indicates the make of each car in the data set

Model: Indicates the model of the vehicle listed

Sales__in__thousands: Lists the amount of cars sold for a specific model for a given year. In the future, I will research which year this is exactly(as the data does not specify).

___year_resale_value: Tells us the resale value for the car based on a given year; I am not sure yet which year this is.

Price__in__thousands: Indicates how much the vehicle costs for the year the data was collected during

Horsepower: Indicates the power output of the vehicle

Before I begin implementing visualizations for my argument, I will clean and tidy my data accordingly.

```
myDataPath <- rename(myDataPath, resale__value = X__year_resale_value)
myDataPath$Vehicle_type <- NULL #deleted the vehicle_type column, was redundant. Only had the option of
myDataPath <- myDataPath[-9,] #deleted the 328i row in order to combine in places and create a 3-Series
myDataPath[myDataPath$Sales_in_thousands==19.747,3] <- 28.978 #added the sales from the 328i to the new
myDataPath[myDataPath$Sales_in_thousands==28.978,4] <- 25.744 #used the mean of the resell value of the
myDataPath[myDataPath$Sales_in_thousands==28.978,2] <- "3-Series" #replaced the model from 328i to 3-Se
myDataPath[myDataPath$Sales_in_thousands==28.978,5] <- 30.195 #used the mean of the price of the 328i a

myDataPath <- myDataPath[-97,] #deleted the SLK230 row in order to combine in places and create an SLK-
myDataPath[myDataPath$Sales_in_thousands==7.998,3] <- 9.524 #added sales from SLK230 to SLK-Series
myDataPath[myDataPath$Sales_in_thousands==9.524,4] <- 25.581 #used the mean of the resell value of the
myDataPath[myDataPath$Sales_in_thousands==9.524,2] <- "SLK-Series" #created a new SLK Series row
myDataPath[myDataPath$Sales_in_thousands==9.524,5] <- 39.950 #used the mean of the price of the SLK and

myDataPath <- myDataPath[-125,] #deleted the Porsche Carrera Cabriolet row to combine in places and cre
myDataPath[myDataPath$Sales_in_thousands==1.280,3] <- 3.146 #added sales from the Porsche former Cabrio
myDataPath[myDataPath$Sales_in_thousands==3.146,4] <- 64.088 #used the mean of the resell value of the
myDataPath[myDataPath$Sales_in_thousands==3.146,2] <- "911 Series" #created a new 911 series row
myDataPath[myDataPath$Sales_in_thousands==3.146,5] <- 72.995 #used the mean of the price of the Porsche
```

```
expensive_cars <- filter(myDataPath, `Price_in_thousands` >= 35) #created a data set that only shows ex
  affordable_cars <- filter(myDataPath, `Price_in_thousands` < 35) #created a data set that only shows
```

```
myDataPath[myDataPath$Sales_in_thousands==14.785,4] <- 30.858
#changed resell value of Cadillac Escalade

myDataPath[myDataPath$Sales_in_thousands==14.114,5] <- 23.100
#added price for the Acura CL, used year 1999 because the Acura Cl was not made in year 2000 (hence why
```

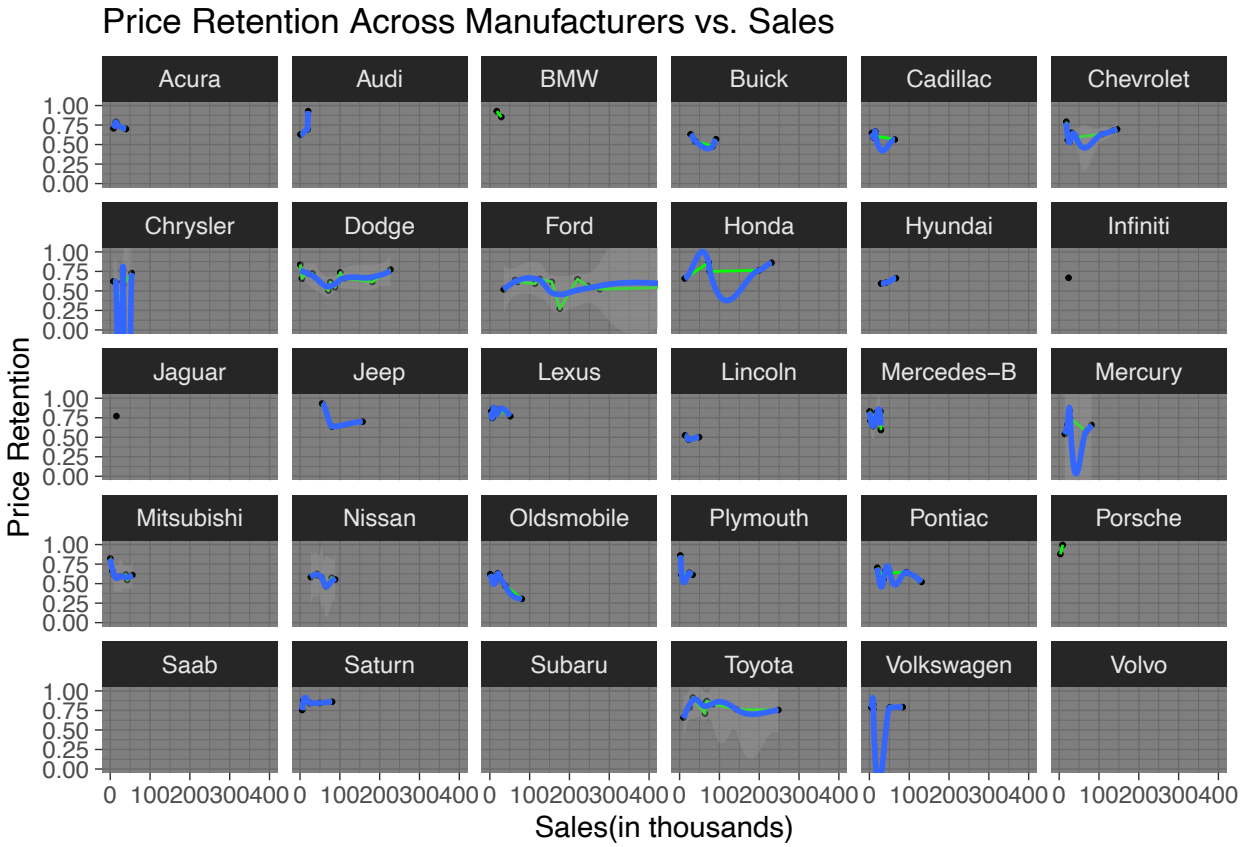
```
myDataPath[myDataPath$Sales_in_thousands==107.995,4] <- 11.956
#added resell value of Chevrolet Impala

myDataPath[myDataPath$Sales_in_thousands==30.696,4] <- 8.072
#added resell value of Chrysler 300M

myDataPath[myDataPath$Sales_in_thousands==101.323,4] <- 19.349
#added resell value of Dodge Durango

myDataPath[myDataPath$Sales_in_thousands==175.670,4] <- 3.334
#added resell value of Ford Focus

myDataPath[myDataPath$Sales_in_thousands==15.467,4] <- 33.000
#Estimated Jaguar with a rather low resale keeping in mind that Jaguar models have some of the worst re

myDataPath[myDataPath$Sales_in_thousands==15.467,2] <- "S-Type"
#Re-added the model name for the Jaguar because I accidentally deleted it.

myDataPath[myDataPath$Sales_in_thousands==3.334,4] <- 38.342
#added resell value of Lexus GS400
myDataPath[myDataPath$Sales_in_thousands==9.126,4] <- 52.142
#added resell value of Lexus LS470
myDataPath[myDataPath$Sales_in_thousands==51.238,4] <- 26.642
#added resell value of Lexus RX300

myDataPath[myDataPath$Sales_in_thousands==22.925,4] <- 20.080
#added resale value of Lincoln Navigator

myDataPath[myDataPath$Sales_in_thousands==11.592,4] <- 27.231
#added resale value of Mercedes CLK Coupe
myDataPath[myDataPath$Sales_in_thousands==0.954,4] <- 71.131
#added resale value of Mercedes CL500
myDataPath[myDataPath$Sales_in_thousands==28.976,4] <- 20.931
#added resale of Mercedes M-Class

myDataPath[myDataPath$Sales_in_thousands==54.158,4] <- 13.244
#added resale of Nissan Xterra
myDataPath[myDataPath$Sales_in_thousands==65.005,4] <- 8.335
#added resale of Nissan Frontier

myDataPath[myDataPath$Sales_in_thousands==38.554,4] <- 11.444
#added resale of Oldsmobile Intrigue
myDataPath[myDataPath$Sales_in_thousands==80.255,4] <- 5.564
#added resale of Oldsmobile Alero

myDataPath[myDataPath$Sales_in_thousands==1.872,4] <- 37.002
#added resale of Plymouth Prowler

myDataPath[myDataPath$Sales_in_thousands==39.572,4] <- 17.046
#added resale of Pontiac Montana

myDataPath[myDataPath$Sales_in_thousands==8.472,4] <- 16.525
#added resale of Saturn LW
```

```
myDataPath[myDataPath$Sales_in_thousands==49.989,4] <- 12.700
#added resale of Saturn LS

myDataPath[myDataPath$Sales_in_thousands==65.119,4] <- 18.858
#added resale of Toyota Sienna

myDataPath[myDataPath$Sales_in_thousands==53.480,5] <- 26.800
#added price for the Chrysler Town & Country via Cars.com
myDataPath[myDataPath$Sales_in_thousands==53.480,6] <- 3.3
#added the engine size for the Chrysler Town & Country via Edmunds.com.
myDataPath[myDataPath$Sales_in_thousands==53.480,7] <- 158
#added the horse power for the Chrysler Town & Country via Edmunds.com.
myDataPath[myDataPath$Sales_in_thousands==53.480,8] <- 119.3
#added the wheelbase for the Chrysler Town & Country via Edmunds.com
myDataPath[myDataPath$Sales_in_thousands==53.480,9] <- 76.8
#added the width for the Chrysler Town & Country via Edmunds.com
myDataPath[myDataPath$Sales_in_thousands==53.480,10] <- 199.7
#added the length for the Chrysler Town & Country via Edmunds.com
myDataPath[myDataPath$Sales_in_thousands==53.480,11] <- 4.045
#added the curb weight for the Chrysler Town & Country via Edmunds.com
myDataPath[myDataPath$Sales_in_thousands==53.480,12] <- 20.0
#added the fuel capacity for the Chrysler Town & Country via Edmunds.com
myDataPath[myDataPath$Sales_in_thousands==53.480,13] <- 18.0
#added the fuel efficiency for the Chrysler Town & Country via Edmunds.com

myDataPath[myDataPath$Model=="Seville",11] <- 3.970
#added the curb weight for the Cadillac Seville via Edmunds.com

myDataPath[myDataPath$Price_in_thousands==22.505,13] <- 28.0
#added the fuel efficiency for the Dodge Intrepid via Edmunds.com

myDataPath[myDataPath$Curb_weight==3.455 & myDataPath$Length==195.9,13] <- 20.0
#added the fuel efficiency for the Oldsmobile Intrigue via Edmunds.com
```

Visualizations

```
ggplot(myDataPath,
  aes(x = Sales_in_thousands, y = resale__value/Price_in_thousands)) +
  coord_cartesian(xlim = c(0, 400), ylim = c(0, 1)) +
  geom_point(size = 0.5) +
  geom_line(colour = "green") +
  geom_smooth() +
  facet_wrap(~Manufacturer) +
  labs(title = "Price Retention Across Manufacturers vs. Sales",
       x = "Sales(in thousands)",
       y = "Price Retention")+
  theme_dark()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```



For the first visualization to construct my argument, I thought observed what I call the "Price Retention"(meaning how much a each car is able to retain its original value) across all manufacturers, versus the total sales in thousands.

I felt the line graph paired with the facet wrap function most accurately represented the variation in price retention across all manufacturers in the data set. I utilize line graphs in my visualizations because I compared numeric variables as opposed to categorical data. Therefore, the line graph seemed the most intuitive. I first attempted to try and build histograms to represent my data due to their impactful and aesthetically pleasing nature. However, I did not believe histograms were the best choice for my final visualizations. Overall, histograms are best used to visualize occurrences of a single categorical variable.

Despite lower unit sales, BMW, Porsche, and Mercedes-Benz have a great deal of price retention. Additionally, Toyota and Lexus exhibit considerable price retention as well. Conversely, American brands like Oldsmobile, Lincoln, Mercury, Dodge, and even Ford have significantly lower price retention ratios–**but why is this?**

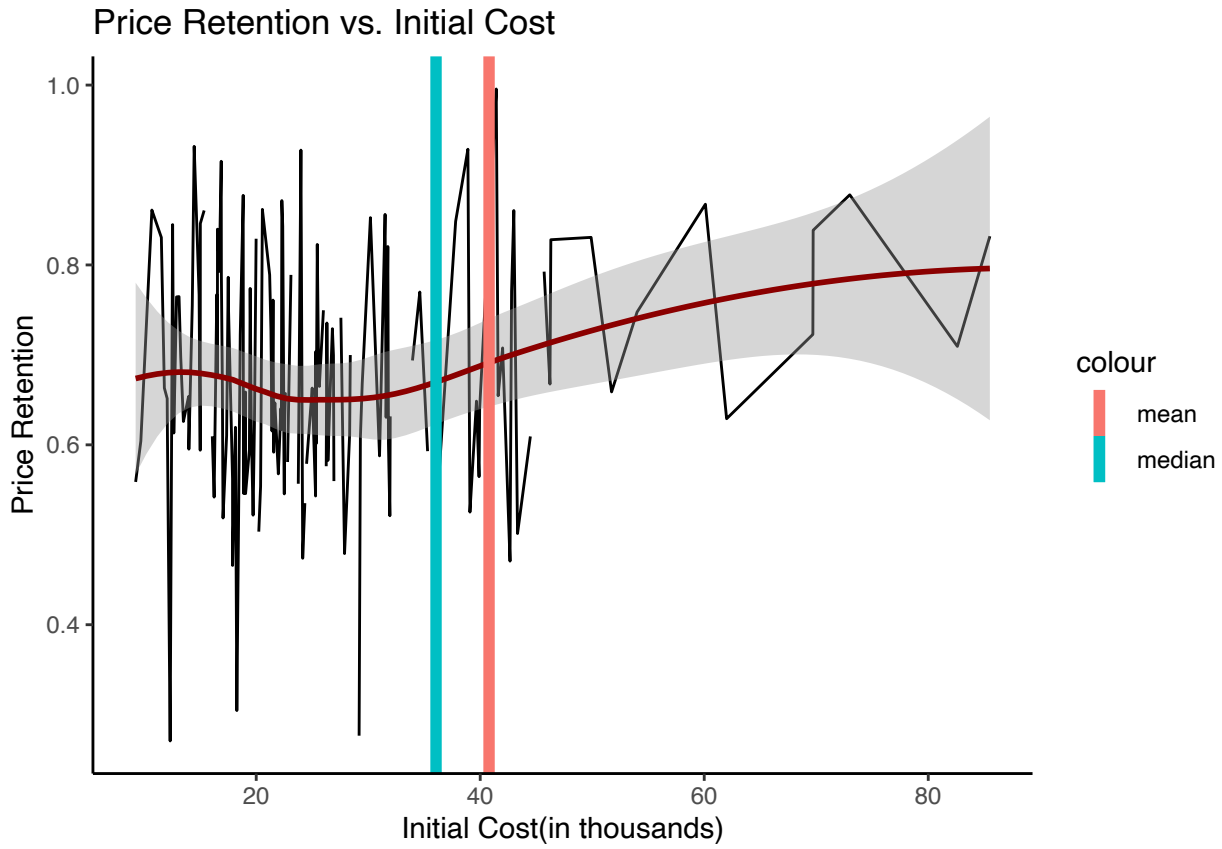
I will attempt to explain this phenomenon using only information from this data set. My hypothesis is that most American manufacturers fail to have high price retention values because they fail to create models that fit optimal price and horsepower thresholds. My research indicates when people are in the market to purchase a car, they purchase either an affordable or an expensive vehicle. American models suffer from being somewhere in between. **The problem with most American cars is that they exist somewhere in between the ideal thresholds for price and horsepower.** Let us first visualize price retention versus initial price.

```
ggplot(myDataPath,
  aes(x = Price_in_thousands, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "dark red") +
```

```
  geom_vline(aes(xintercept = mean(Price_in_thousands/Price_Retention, na.rm = TRUE), color = 'mean'), s
  geom_vline(aes(xintercept = median(Price_in_thousands/Price_Retention, na.rm = TRUE), color = 'median
  labs(title = "Price Retention vs. Initial Cost",
       x = "Initial Cost(in thousands)",
       y = "Price Retention") +
  theme_classic()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```

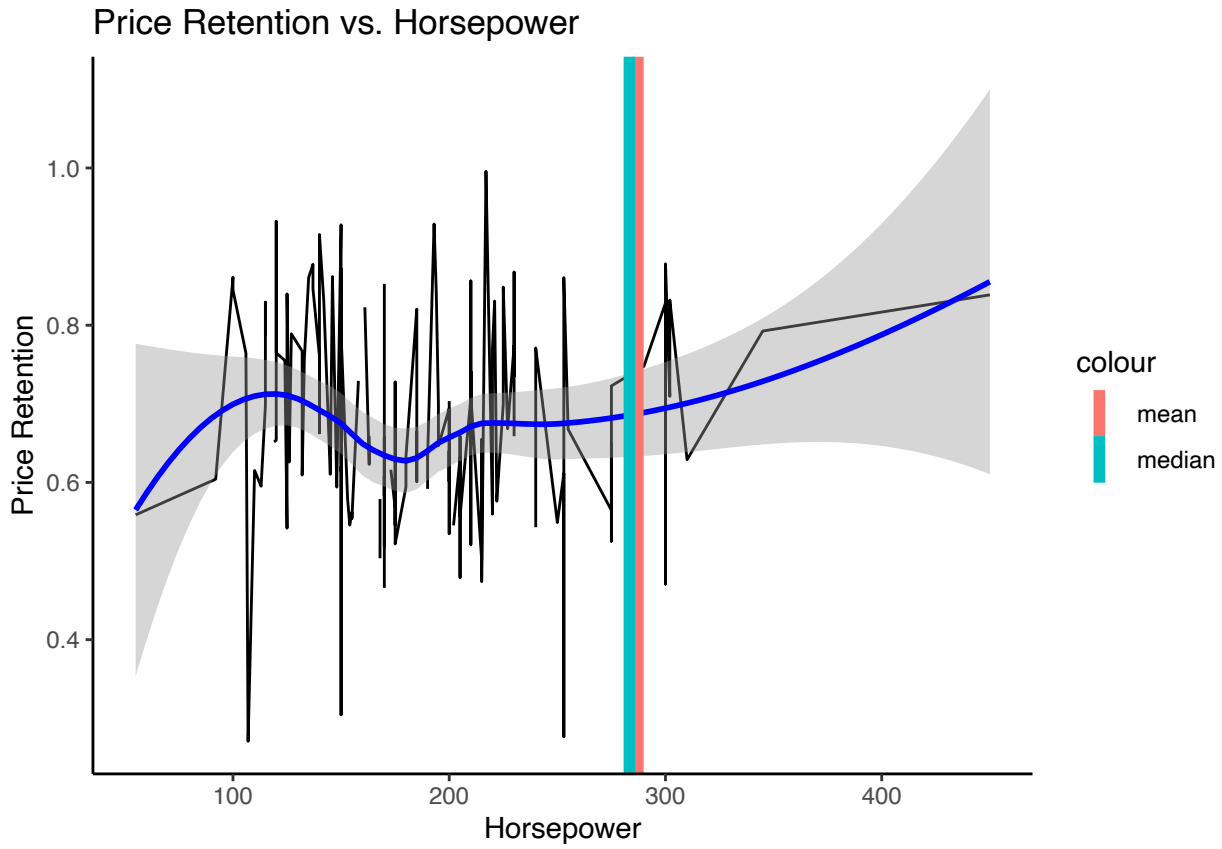


Cars with an initial cost less than about 20k exist in the ideal range for price retention in the inexpensive car market. Cars over around 37k exist in the ideal threshold for the expensive car market. After 37k, the higher the initial cost of the car, the more it will resell for.

```
ggplot(myDataPath,
  aes(x = Horsepower, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "blue") +
  geom_vline(aes(xintercept = mean(Horsepower/Price_Retention, na.rm = TRUE), color = 'mean'), show.leg
  geom_vline(aes(xintercept = median(Horsepower/Price_Retention, na.rm = TRUE), color = 'median'), show
  labs(title = "Price Retention vs. Horsepower",
       x = "Horsepower",
       y = "Price Retention") +
  theme_classic()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```



As the following representation indicates, two ideal bands exist for expensive and affordable cars in horsepower as well as initial price. For affordable cars, price retention is particularly high for cars under about 165 horsepower. Similarly, for expensive cars, the price retention increases as the horsepower increases beginning at around 265 horsepower. Models that fall outside these two ideal thresholds suffer from having the lowest Price Retention overall.

At this point, we will investigate cars with the lowest value retention in the data set to see if existing outside the optimal thresholds presented causes low resell value overall.

```
myDataPath %>%
  select(Manufacturer, Model, Price_in_thousands, Horsepower, Price_Retention) %>%
  arrange(Price_Retention) %>%
  slice(1:20)
```

```
##    Manufacturer       Model Price_in_thousands Horsepower Price_Retention
## 1          Ford       Focus              12.315        107       0.2707268
## 2      Chrysler        300M              29.185        253       0.2765804
## 3    Oldsmobile       Alero              18.270        150       0.3045430
## 4        Nissan    Frontier              17.890        170       0.4659027
## 5       Lincoln   Navigator              42.660        300       0.4706985
## 6    Oldsmobile    Intrigue              24.150        215       0.4738716
## 7         Buick     LeSabre              27.885        205       0.4791106
```

```
## 8       Lincoln    Town car             43.330        215       0.5013847
## 9         Dodge     Stratus             20.230        168       0.5034602
## 10         Ford     Contour             17.035        170       0.5186381
## 11         Ford    Explorer             31.930        210       0.5211400
## 12      Pontiac    Grand Am             19.720        175       0.5218053
## 13      Lincoln Continental             39.080        275       0.5252047
## 14    Chevrolet      Camaro             24.340        200       0.5351274
## 15      Mercury    Mystique             16.240        125       0.5418719
## 16        Buick       Regal             25.300        240       0.5430830
## 17        Dodge    Intrepid             22.505        202       0.5454343
## 18   Mitsubishi     Eclipse             19.047        154       0.5457552
## 19    Chevrolet      Lumina             18.890        175       0.5457914
## 20   Oldsmobile      Aurora             36.229        250       0.5490077
```

Firstly, here are the top 20 vehicles in the entire data set with the lowest resell values. As we can see all but one are American. Next, we will try and see how many of these vehicles fall within the optimal threshold for affordable models.

```
myDataPath %>%
  select(Manufacturer, Model, Price_in_thousands, Horsepower, Price_Retention) %>%
  arrange(Price_Retention) %>%
  filter(Price_Retention < 0.5490077) %>%
  filter(Horsepower <= 165, Price_in_thousands < 20)
```

```
##   Manufacturer    Model Price_in_thousands Horsepower Price_Retention
## 1         Ford    Focus             12.315        107       0.2707268
## 2   Oldsmobile    Alero             18.270        150       0.3045430
## 3      Mercury Mystique             16.240        125       0.5418719
## 4   Mitsubishi  Eclipse             19.047        154       0.5457552
```

As we can see, only 4 cars fall within the optimal thresholds for price and horsepower for the affordable car market. Now we will display cars that fall into the higher threshold for price and horsepower and add both the results together.

```
myDataPath %>%
  select(Manufacturer, Model, Price_in_thousands, Horsepower, Price_Retention) %>%
  arrange(Price_Retention) %>%
  filter(Price_Retention < 0.5490077) %>%
  filter(Horsepower >= 265, Price_in_thousands >= 37)
```

```
##   Manufacturer       Model Price_in_thousands Horsepower Price_Retention
## 1      Lincoln   Navigator             42.66        300       0.4706985
## 2      Lincoln Continental             39.08        275       0.5252047
```

As we can see, only 2 cars fall within the ideal thresholds for price and horsepower in the expensive car market.

Therefore, **at least 70% of the cars with the lowest price retention suffer this issue because their price and/or horsepower are not ideal**. As the research indicates, there are "sweet spots" in horsepower and initial price that vehicles with high price retention tend to fall between. The sweet spots exist for inexpensive vehicles and expensive vehicles.

19 out of the top 20 cars with the lowest price retention are American cars. This demonstrates most American cars have low price retention because they exist outside of the ideal thresholds for horsepower and price. American cars fail to retain as much value as their German and Japanese contemporaries because they exist outside of the ideal thresholds for initial price and horsepower.

Reflection

In order to reflect on the accuracy of my results, I will build models to understand my data further.

```
lm(Price_in_thousands~Price_Retention, data = Expensive_Car_Sweetspot) %>%
  summary()
```

```
##
## Call:
## lm(formula = Price_in_thousands ~ Price_Retention, data = Expensive_Car_Sweetspot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.4977  -6.4257  -0.9336   9.0823  25.1021
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.832     21.216   0.086   0.9326
## Price_Retention   78.464     29.858   2.628   0.0221 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.8 on 12 degrees of freedom
## Multiple R-squared:  0.3653, Adjusted R-squared:  0.3124
## F-statistic: 6.906 on 1 and 12 DF,  p-value: 0.02206
```

Based on the R-squared variable, we can see the initial price is about 37% correlated to price retention for expensive cars. Now let us look at the model as a graph.

```
Expensive_Car_Price_Sim <- lm(Price_in_thousands~Price_Retention, data = Expensive_Car_Sweetspot)
ggplot(Expensive_Car_Price_Sim,
  aes(x = Price_in_thousands, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "dark red") +
labs(title = "Price Retention vs. Price in Thousands[Expensive Cars]",
       x = "Initial Price[in thousands]",
       y = "Price Retention") +
  theme_classic()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Interestingly, our model indicates a dip in the 50 to 70k price range for price retention versus initial price for expensive cars. This is something our data does not tell us initially. Furthermore, the model predicts a decline after the 80k mark regarding price retention. This tells us the relationship between Price and Price Retention is not entirely linear at all for expensive cars. In fact, it is much less linear that it would have seemed originally.

Now, let us look at the correlation between price retention and horsepower for expensive cars.

```
lm(Horsepower~Price_Retention, data = Expensive_Car_Sweetspot) %>%
  summary()
```

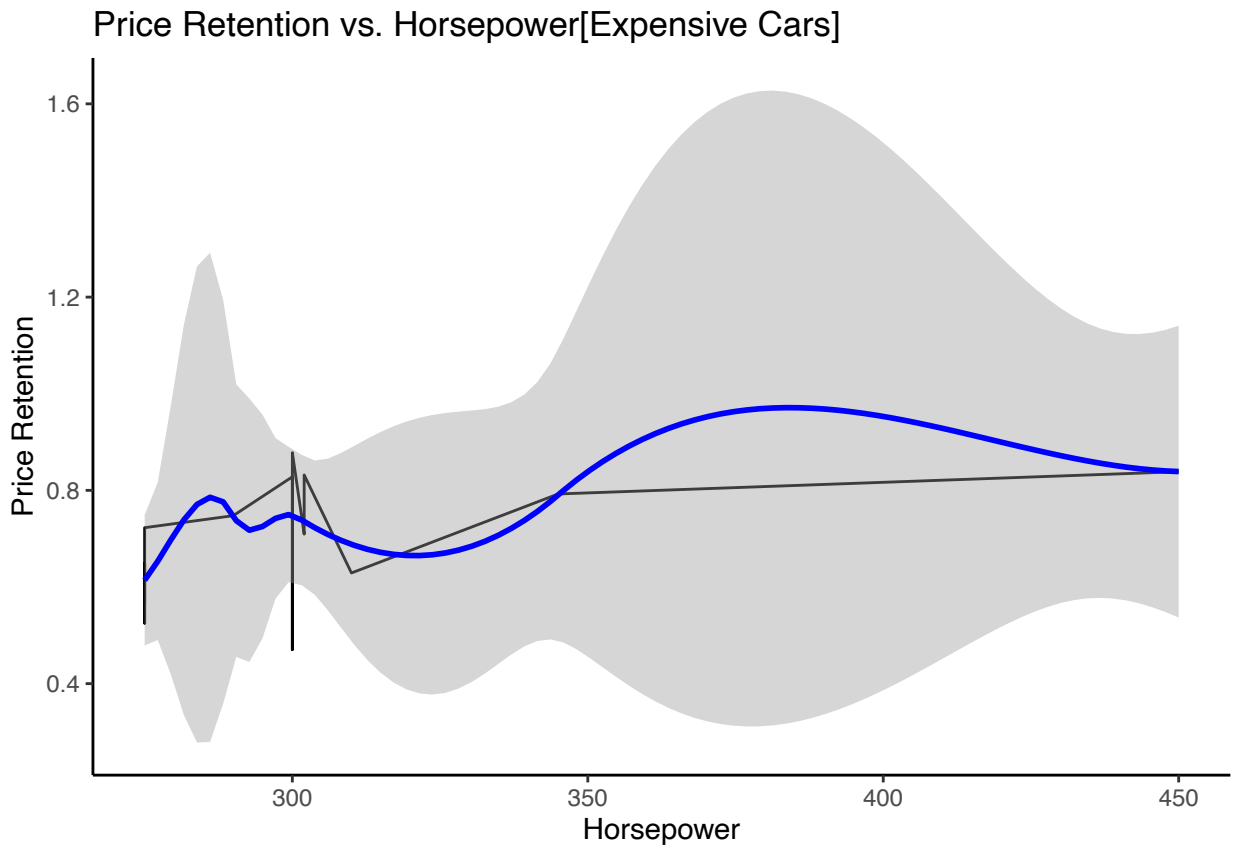
```
##
## Call:
## lm(formula = Horsepower ~ Price_Retention, data = Expensive_Car_Sweetspot)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -34.35 -24.40 -11.90  11.72 122.08
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       191.20      65.53   2.918   0.0129 *
## Price_Retention   163.04      92.23   1.768   0.1025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.61 on 12 degrees of freedom
```

```
## Multiple R-squared:  0.2066, Adjusted R-squared:  0.1405
## F-statistic: 3.125 on 1 and 12 DF,  p-value: 0.1025
```

Based on our R-squared variable, we can see horsepower is about 21% correlated to price retention for expensive cars.

```
Expensive_Car_HP_Sim <- lm(Horsepower~Price_Retention, data = Expensive_Car_Sweetspot)
ggplot(Expensive_Car_HP_Sim,
  aes(x = Horsepower, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "blue") +
labs(title = "Price Retention vs. Horsepower[Expensive Cars]",
       x = "Horsepower",
       y = "Price Retention") +
  theme_classic()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The model predicts a dip in the data for initial price versus horsepower that my initial visualizations did not demonstrate. As we can see, a dip exists in the price retention for cars between 300 and 340 horsepower. Furthermore, there is a peak where price retention is the very highest at about 370 horsepower. After this point, the model indicates price retention relative to initial price will decline.

Now, let's assess the same models for affordable cars.

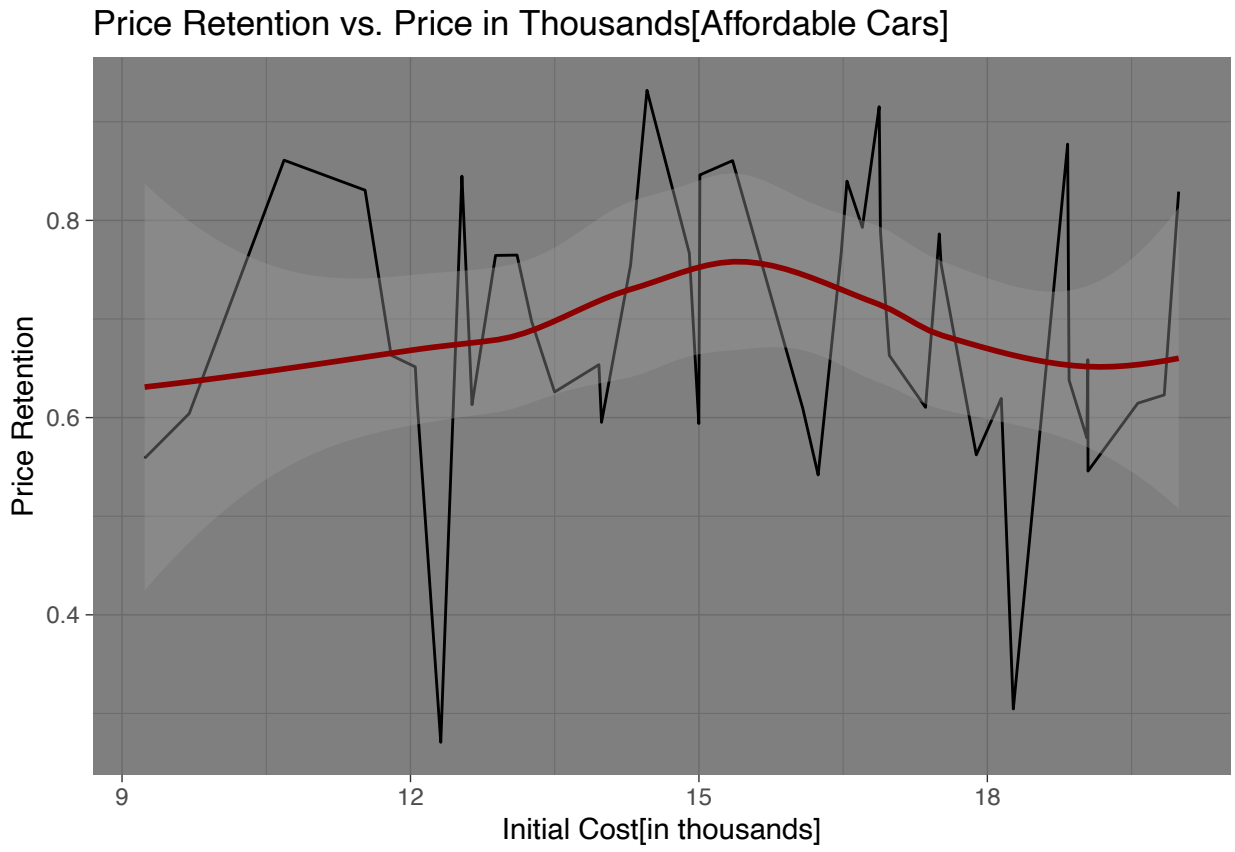
```
lm(Price_in_thousands~Price_Retention, data = Affordable_Car_Sweetspot) %>%
  summary()
```

```
##
## Call:
## lm(formula = Price_in_thousands ~ Price_Retention, data = Affordable_Car_Sweetspot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1232 -2.4535  0.0207  2.1787  4.6577
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     15.41173    2.20716   6.983 1.36e-08 ***
## Price_Retention -0.09573    3.14987  -0.030    0.976
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.937 on 43 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  2.148e-05,  Adjusted R-squared:  -0.02323
## F-statistic: 0.0009237 on 1 and 43 DF,  p-value: 0.9759
```

As we can see, for affordable cars, the R-squared value indicates initial price and price retention are highly correlated. Price is incredibly important for price retention in affordable cars.

```
Affordable_Car_Price_Sim <- lm(Price_in_thousands~Price_Retention, data = Affordable_Car_Sweetspot)
ggplot(Affordable_Car_Price_Sim,
  aes(x = Price_in_thousands, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "dark red") +
labs(title = "Price Retention vs. Price in Thousands[Affordable Cars]",
       x = "Initial Cost[in thousands]",
       y = "Price Retention") +
  theme_dark()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



As we can see in the graph, our model demonstrates a more specific threshold in price retention for affordable cars than indicated by prior visualizations. At about the 16k initial price point, the model predicts affordable cars will retain their value the best.

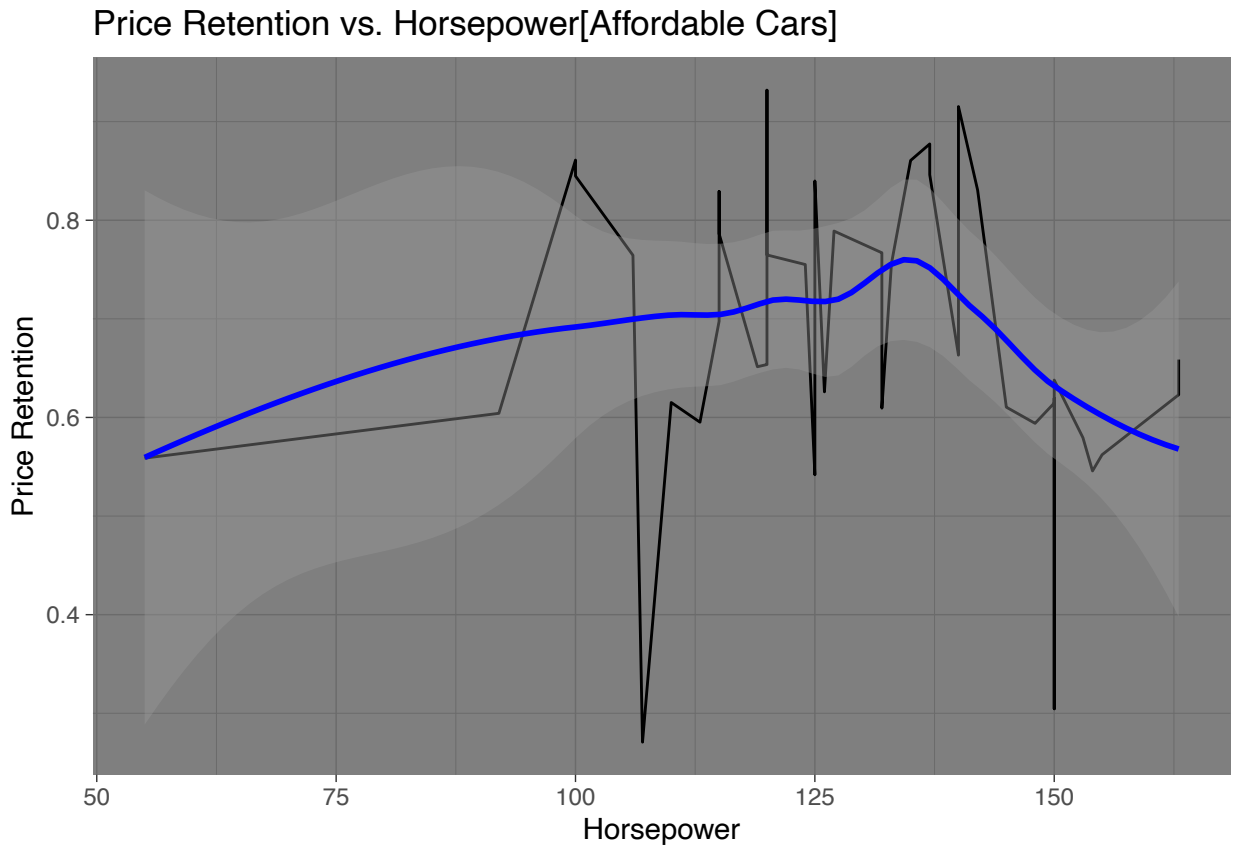
```
lm(Horsepower~Price_Retention, data = Affordable_Car_Sweetspot) %>%
  summary()
```

```
##
## Call:
## lm(formula = Horsepower ~ Price_Retention, data = Affordable_Car_Sweetspot)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -75.160 -11.090   0.844  15.739  34.578
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       139.88      15.59   8.969  2.1e-11 ***
## Price_Retention   -17.39      22.26  -0.781    0.439
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.75 on 43 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.014,  Adjusted R-squared:  -0.008933
## F-statistic: 0.6104 on 1 and 43 DF,  p-value: 0.4389
```

Lastly, I will interpret the model for affordable cars versus horsepower. According to the R-squared value, we can deduce the correlation between price retention and horsepower for affordable cars is almost negligible at just 1.4%. In other words, horsepower has almost no effect on price retention for affordable cars.

```
Affordable_Car_HP_Sim <- lm(Horsepower~Price_Retention, data = Affordable_Car_Sweetspot)
ggplot(Affordable_Car_HP_Sim,
  aes(x = Horsepower, y = Price_Retention)) +
  geom_line(color = "black") +
  geom_smooth(color = "blue") +
labs(title = "Price Retention vs. Horsepower[Affordable Cars]",
       x = "Horsepower",
       y = "Price Retention") +
  theme_dark()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The model predicts a small optimum at about the 130 horsepower mark. Additionally, the model suggests any more than 130 horsepower becomes gradually less ideal for affordable vehicles.

Reflection (cont'd)

Overall, I am pleased with my research. However, I wish the data set I used to generate my findings was more recent. Indeed, my findings are still worthwhile as many of the trends described continue to occur more

than two decades later. However, I realize my findings would be more meaningful had the data been more recent.

Additionally, I wish I was able to use multiple imputation to generate missing resale values in the initial data set. I was informed of the possibility of using multiple imputation but failed after many attempts to get the mice library to load on my computer. Therefore, I used a formula I created to generate the resell values myself:

NA (for resale value) = Sales_in_thousands - (price_in_thousands - resale___price)/(number of other models) ^^^ **repeat this step for all models made by the manufacturer and add the sum together. This sum is then divided by however many other models there are that do not include the NA value. Subtract this value from the original market price of the car missing the resell price.

Thirdly, I wish I was able to generate resale values for Saab, Subaru and Volvo. I would have been able to draw a more comprehensive and valid conclusion with these manufacturers. However, I did not believe it was useful to input resale values for manufacturers without at least one resale value to start with.

I also wish I had known which variables I would be interested in at an earlier juncture. Instead, I only knew the variables I would use to build my argument at a later stage in my research process. This would have saved the time used to input each NA value for every variable. It also would have demonstrated to me earlier that price and horsepower are correlated to price retention but other variables exist in my data that also contribute to price retention. If I had to continue, I would start by attempting to quantify the other variable(s) for price retention.

Conclusion

For expensive cars, the initial price is 21% correlated to price retention and horsepower is 37% correlated to price retention. This demonstrates at least one other significant factor exists that affects price retention for expensive cars.

For affordable cars, my models demonstrate horsepower was almost negligible for affordable cars and price was more important than for expensive cars.

```
Affordable_Car_Sweetspot %>%
  slice(1:10)
```

```
##    Manufacturer         Model Price_in_thousands Horsepower Price_Retention
## 1     Chevrolet      Cavalier             13.260        115       0.6975867
## 2     Chevrolet         Prizm             13.960        120       0.6536533
## 3     Chevrolet         Metro              9.235         55       0.5587439
## 4      Chrysler Sebring Coupe             19.840        163       0.6229839
## 5      Chrysler        Cirrus             16.480        132       0.7669903
## 6         Dodge          Neon             12.640        132       0.6131329
## 7         Dodge       Avenger             19.045        163       0.6587031
## 8         Dodge        Dakota             16.980        120       0.6631331
## 9         Dodge       Caravan             19.565        150       0.6146179
## 10         Ford        Escort             12.070        110       0.6151616
```

American cars that do fall within the optimal thresholds of horsepower and initial price tend to have higher price retentions than contemporaries that do not. This proves even American cars can maintain their value well, so long as they fit within the optimal thresholds of horsepower and price.

```
Expensive_Car_Sweetspot %>%
  arrange(desc(Price_Retention)) %>%
  slice(1:10)
```

```
##    Manufacturer      Model Price_in_thousands Horsepower Price_Retention
## 1       Porsche 911 Series             72.995        300       0.8779779
## 2         Dodge      Viper             69.725        450       0.8385801
## 3    Mercedes-B      CL500             85.500        302       0.8319415
## 4         Lexus      GS400             46.305        300       0.8280315
## 5     Chevrolet   Corvette             45.705        345       0.7925829
## 6         Lexus      LS400             54.005        290       0.7476160
## 7    Mercedes-B    S-Class             69.700        275       0.7227403
## 8    Mercedes-B   SL-Class             82.600        302       0.7094431
## 9      Cadillac   Eldorado             39.665        275       0.6485567
## 10         Audi         A8             62.000        310       0.6290323
```

Here, we see that American cars can maintain their value quite well, so long as they they fit within the optimal thresholds for horsepower and initial price. In fact, of all the cars in this entire data set, **the car that has the second best price retention of any is American.**

Why does it resell so well? Based on the results, cars that fit within the ideal thresholds for price and horsepower tend to hold their value the best overall. However, because price and horsepower together are only 58% correlated to the Price Retention of expensive cars, there is clearly at least one other considerable variable in price retention for expensive cars that I was unable to quantify.

Annotated Bibliography

Bhatia, Gagan. *Car Sales*. (Kaggle, 2017). The following source is the data set this entire paper is based around. The data set was originally derived from Analytix Labs and was created to predict car unit sales.

Cars. "BMW 323." (Cars.com, 2021). The following source gives the MSRP Range for every year the BMW 323i model was made. This particular site helped the most to find which year the Cars Data set was derived from.

Car Sales Base. (Caralesbase.com, 2021). The following source contains a multitude of data sets for readily-available vehicles. This website was used in the early stages of attempting to determine which year was used to create the Car Sales data set.

Chang, Brittany. "These are the 10 sports cars that have the best resale value 5 years after purchase." (Business Insider, 2019). The following source demonstrates the modern applicability of my research findings generated from the Cars Data Set. The article names the top 10 cars that have the highest value retention for the 2019 year. Two of which are Porsche 911 models exactly like in the Cars Data Set. Additionally, 70% of all models listed are German or Japanese, and only 30% are American.

Dataslice. "Make Beautiful Graphs in R: 5 Quick Ways to Improve ggplot2 Graphs." (YouTube, 2020). The following source demonstrates 5 easy ways to aesthetically improve ggplot2 graphs within R.

Edmunds. "Specs & Features." (Edmunds.com, 2021). The following source gives the user access to an enormous data base containing the exact specifications and features of nearly every car ever made for almost every year. This data set was used to derive the missing specification values of several vehicles.

Meaden, Richard. "History of the Porsche 911 GT3 - driving the original 20 years on." (Evo.co.uk, 2020). The following source is where the title image for my final paper was derived from.

Motor 1. "10 Cars with the Worst Resale Value After One Year." (Motor1.com, 2021). The following web site lists all current car models with the lowest resale value after one year. This source was used to input the resale price for Jaguar. However, I later decided to not use Jaguar in my research.

Statistics for Sustainable Development (Stats4SD). "Statistical Modeling in R." (YouTube,2021). The following video demonstrates the basics of generating effective models within R. This video was particularly helpful in the modeling portion of my research.