Visualization Project

Arka Roy

2023-09-24

Data Set- Mercedes Used Car Listing

Introduction

This is a data set on used mercedes cars models depicting the price,mileage and various other aspects of these cars.Various factors determine cost of an used car given a certain model like how many miles it has run, and what year the car was manufactured. Additionally, there are various cheap and expensive models for each brand, especially for a Mercedes. Thus, we will be analyzing these data to see what the prices would be.

Data Description

- The data set contains 9 columns and 13119 rows.
- It has no missing data.
- Column 'model' tells the car model.
- Column 'Year' gives the year of manufacturing of the models which ranges from 1970 to 2020 with the majority models being manufactured in 2018.
- Column 'price' gives the selling prices of these used cars which ranges from 650 pounds to 159999 pounds with the majority models being sold at 24699 pounds.
- Column 'Transmission' tells us the type of transmission like automatic,manual or semi-transmission.
- Column 'mileage' tells us the no of miles the car has run before being sold. It ranges from as low as 1 mile to as high as 259000 miles with the most frequent being 15189 miles.
- Column 'Fueltype' have three data types- 'Petrol', 'Diesel','Hybrid'.
- Column 'tax' tells the government taxation rate applicable on these cars when they were sold.
- Column 'mpg' tells us the mileage per gallon for these models when they were sold.
- Column 'engineSize' gives information about volume of the engine.

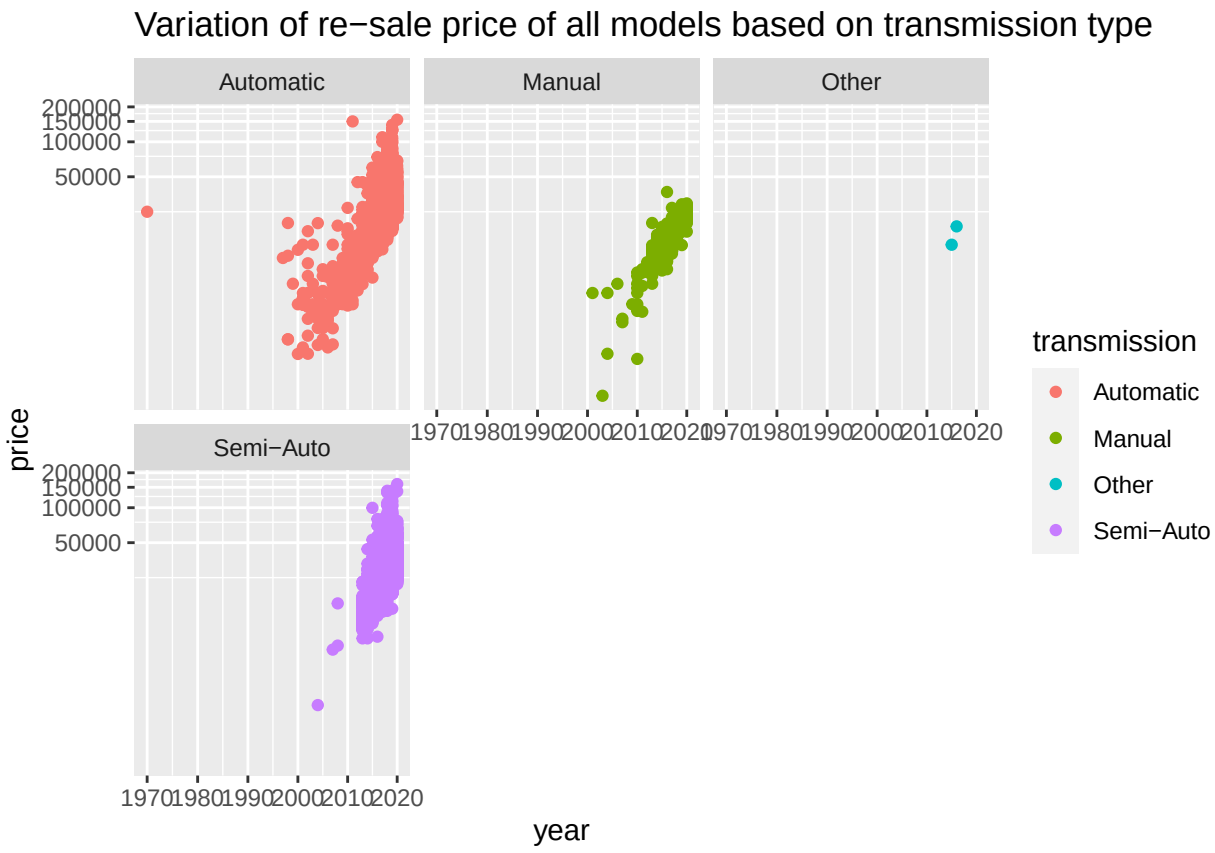
```
##       model year price transmission mileage fuelType tax  mpg engineSize
## 1       SLK 2005  5200    Automatic   63000   Petrol 325 32.1        1.8
## 2   S Class 2017 34948    Automatic   27000   Hybrid  20 61.4        2.1
## 3  SL CLASS 2016 49948    Automatic    6200   Petrol 555 28.0        5.5
## 4   G Class 2016 61948    Automatic   16000   Petrol 325 30.4        4.0
## 5   G Class 2016 73948    Automatic    4000   Petrol 325 30.1        4.0
```

Exploratory Data Analysis

```
##
## Automatic    Manual     Other Semi-Auto
##      4825      1444         2      6848
```

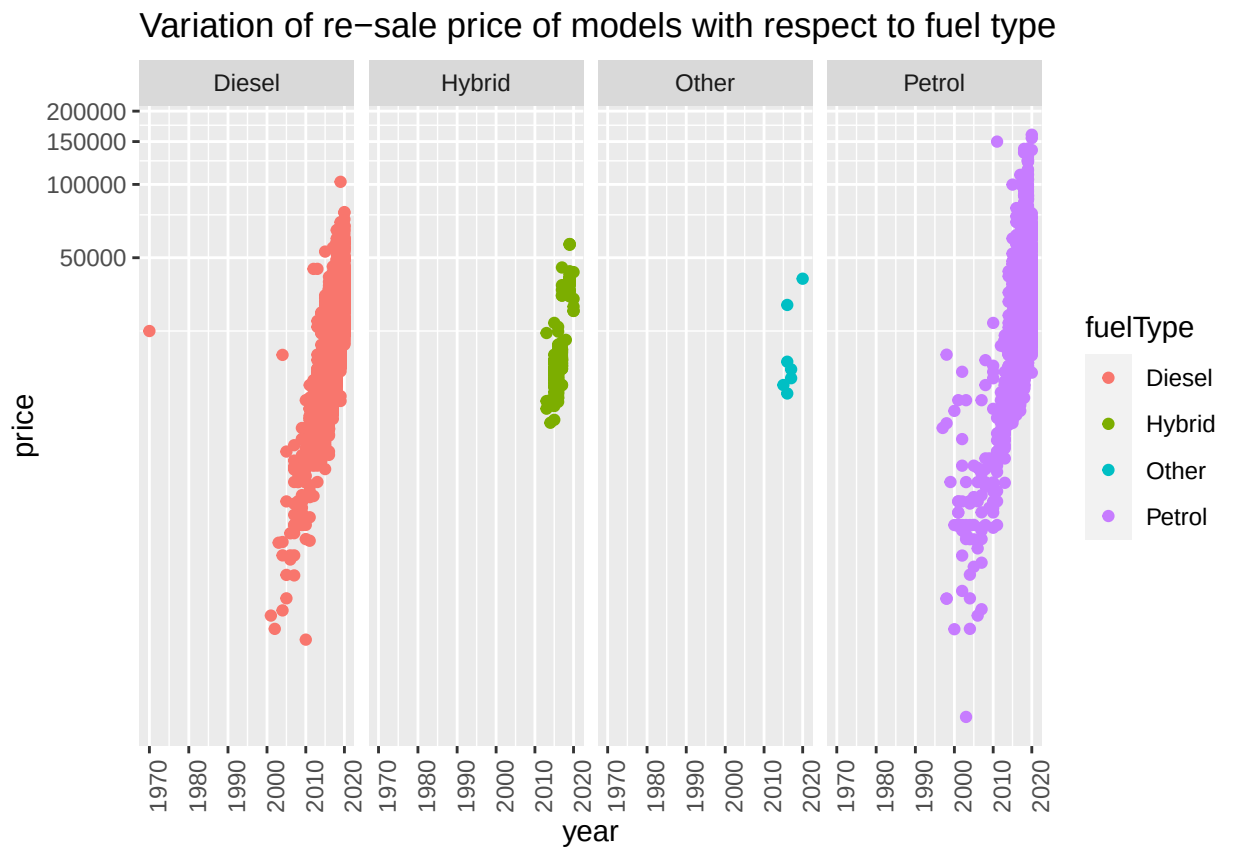
- Initial Exploration shows that among the total models sold, Highest were of Semi-Auto Transmission followed by Automatic and Manual.
- If we explore the registration year with re-sale pricing for these transmission types, we can easily observe that apart from certain out liers, mostly semi-Auto type transmission with latest registration year is more concentrated at higher re-sale price zone. The price for automatic type is more distributed. One can also infer that semi-automatic transmission type is a recent product from Mercedes since the majority registrations were only 2 decades old.

```
library(ggplot2)
ggplot(mydata)+geom_point(aes(x = year, y = price,colour=transmission))+ggtitle("Variation of re-sale pr
```



- Similarly upon observation,we can infer that petrol variants have re-sale values more distributed with certain high re-sale price models than diesel.

```
library(ggplot2)
  ggplot(mydata)+geom_point(aes(x = year, y = price,colour=fuelType))+ggtitle("Variation of re-sale pri
  coord_trans(x="log",y="log")+facet_wrap(~fuelType,ncol=5) + theme(axis.text.x = element_text(angle = 9
```



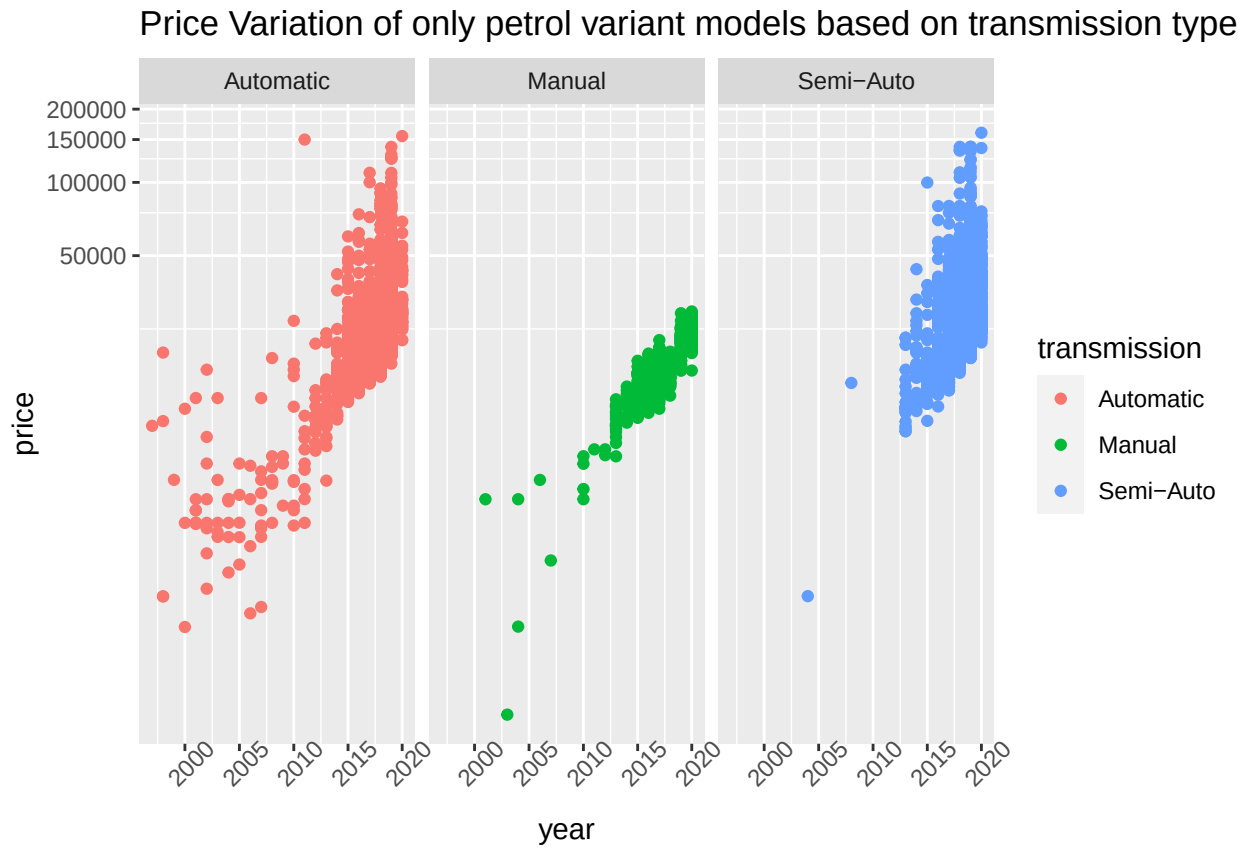
- Now we will try to explore and find out the reason behind the high price of certain petrol variant models. we will look into a filtered data set with only petrol variants as fueltype.
- So we will first filter the fuelType to "Petrol" variants only.

```
mydata_petrol <- dplyr::filter(mydata, fuelType=="Petrol")
head(mydata_petrol)
```

```
##       model year  price transmission mileage fuelType tax  mpg engineSize
## 1       SLK 2005   5200    Automatic   63000   Petrol 325 32.1        1.8
## 2  SL CLASS 2016  49948    Automatic    6200   Petrol 555 28.0        5.5
## 3   G Class 2016  61948    Automatic   16000   Petrol 325 30.4        4.0
## 4   G Class 2016  73948    Automatic    4000   Petrol 325 30.1        4.0
## 5  SL CLASS 2011 149948    Automatic    3000   Petrol 570 21.4        6.2
## 6   S Class 2012  10948    Automatic  107000   Petrol 265 36.7        3.5
```

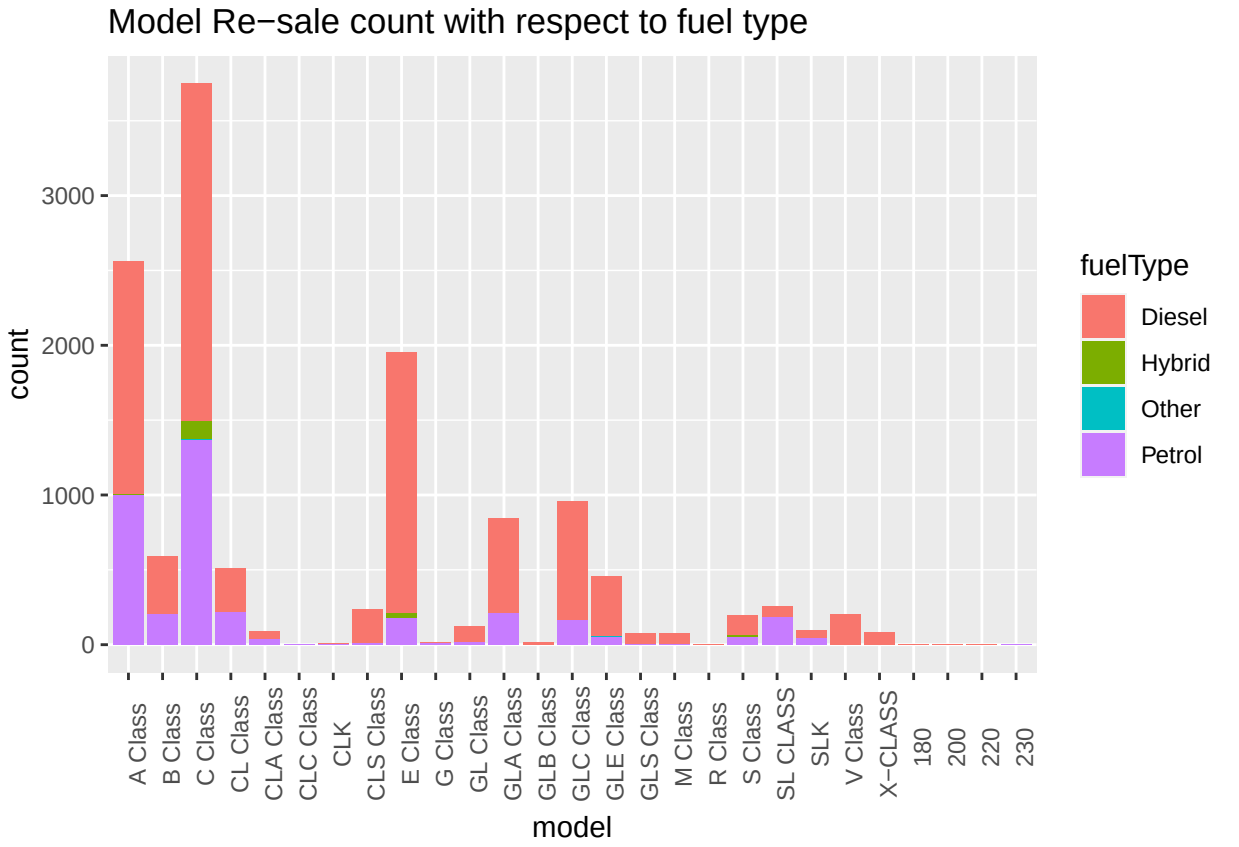
- Now we have a filtered data set where we will be exploring the reason behind high prices of certain petrol variants.
- We will try to explore whether this high re-sale price of petrol variants is connected to the transmission type.

```
ggplot(mydata_petrol)+geom_point(aes(x = year, y = price,colour=transmission))+ggtitle("Price Variation
```



- Clearly, for petrol as fuel type, transmission type is not the factor here as both automatic and semi-auto are falling in the same price range.
- Let's explore the model types to find out whether it is dependent on the models. first we will see what are the different model types and how their sale depended on fuel type.

```
ggplot(data = mydata,
       aes(x = model, fill = fuelType)) +
  geom_bar() +  ggtitle("Model Re-sale count with respect to fuel type")+
   theme(axis.text.x = element_text(angle = 90))
```



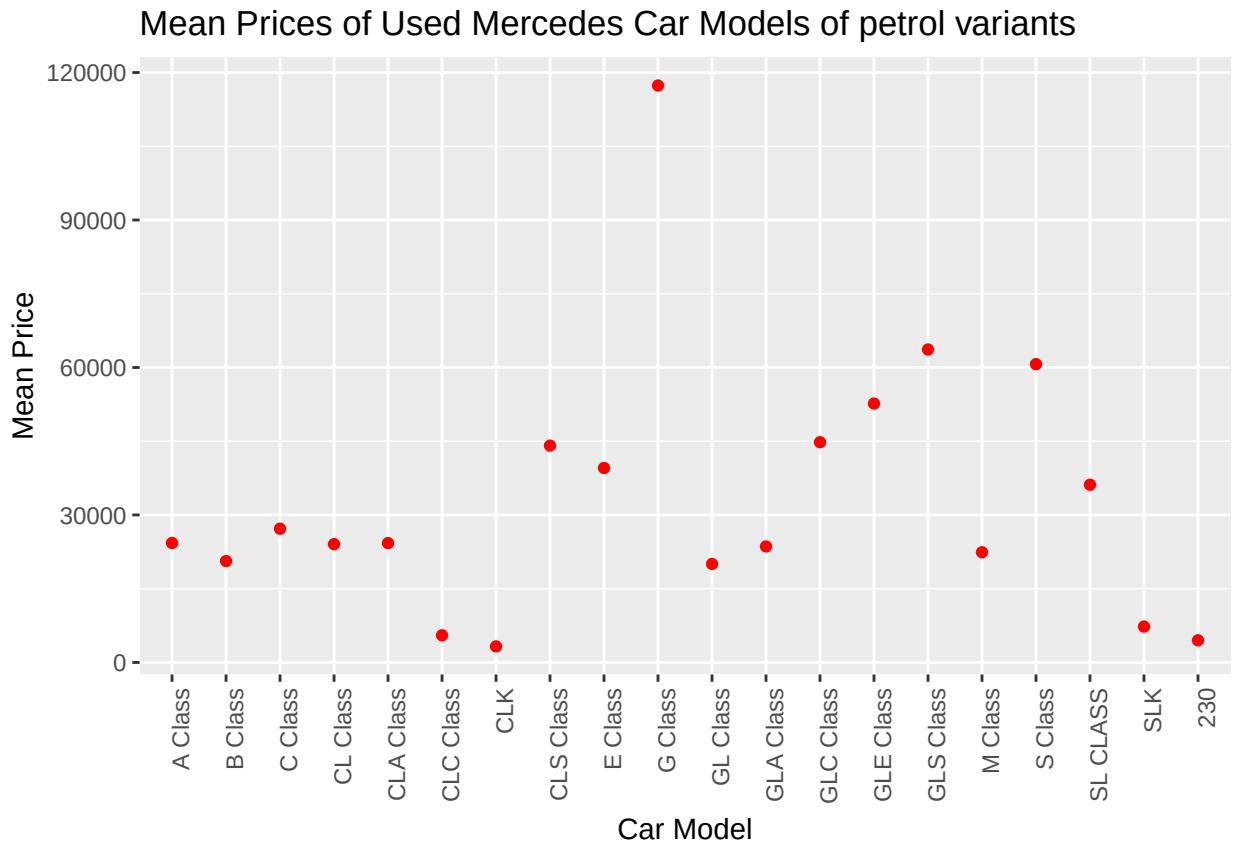
- So there are total 27 different models of different fuel types. Among all of them C class model has been re-sold the highest which is a compact sedan within an affordable price range followed by A-Class which is an hatch back.
- Clearly C-Class model with diesel as fuel type surpasses the sale count when compared to its immediate competitor-petrol fuel type.
- Also one can observe that in almost every model type, diesel variety re-sale count exceeds that of petrol.
- Now lets find out how prices of different model types varies in petrol segment. We will try to analyze the mean price of the different models to get an idea.
- First let calculate the mean price for each model type.

```
mean_prices <- aggregate(mydata_petrol$price, by = list(mydata_petrol$model), FUN = mean)
colnames(mean_prices) <- c("Model", "Mean_Price")
head(mean_prices)
```

```
##        Model Mean_Price
## 1    A Class  24301.885
## 2    B Class  20626.256
## 3    C Class  27204.950
## 4   CL Class  24056.642
## 5  CLA Class  24281.054
## 6  CLC Class   5516.667
```

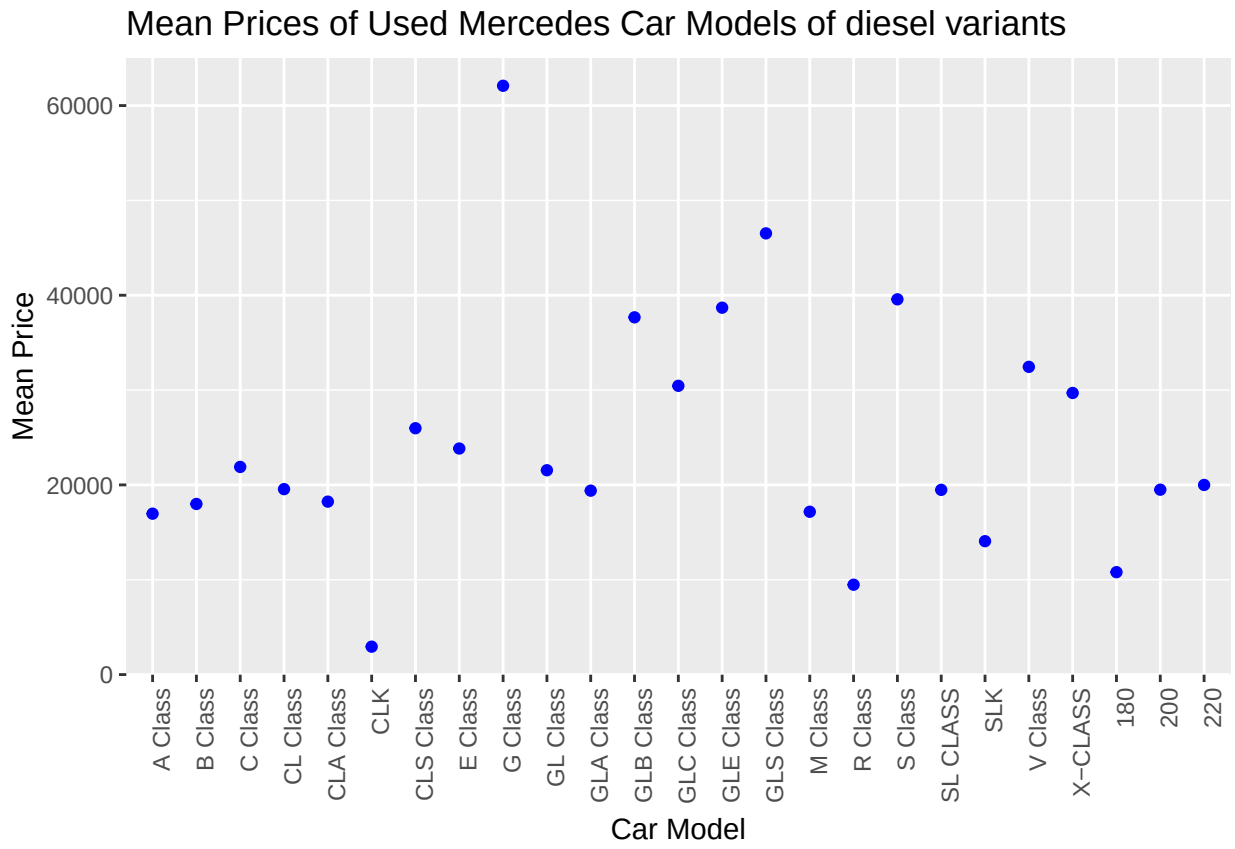
- Now let's find out which model types among petrol variant has higher prices.

```
library(ggplot2)
ggplot(mean_prices)+geom_point(aes(x = Model, y = Mean_Price),color="red") +
  labs(x = "Car Model", y = "Mean Price") +
  ggtitle("Mean Prices of Used Mercedes Car Models of petrol variants") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



- So clearly the mean re-sale price of G-class with petrol variant is very high when compared to other models, which is because it is an SUV wagon from the house of Mercedes. The lowest price point is that of CLK followed by CLC class and 230.
- Lets find out what is the scenario for models with diesel as fuel type.

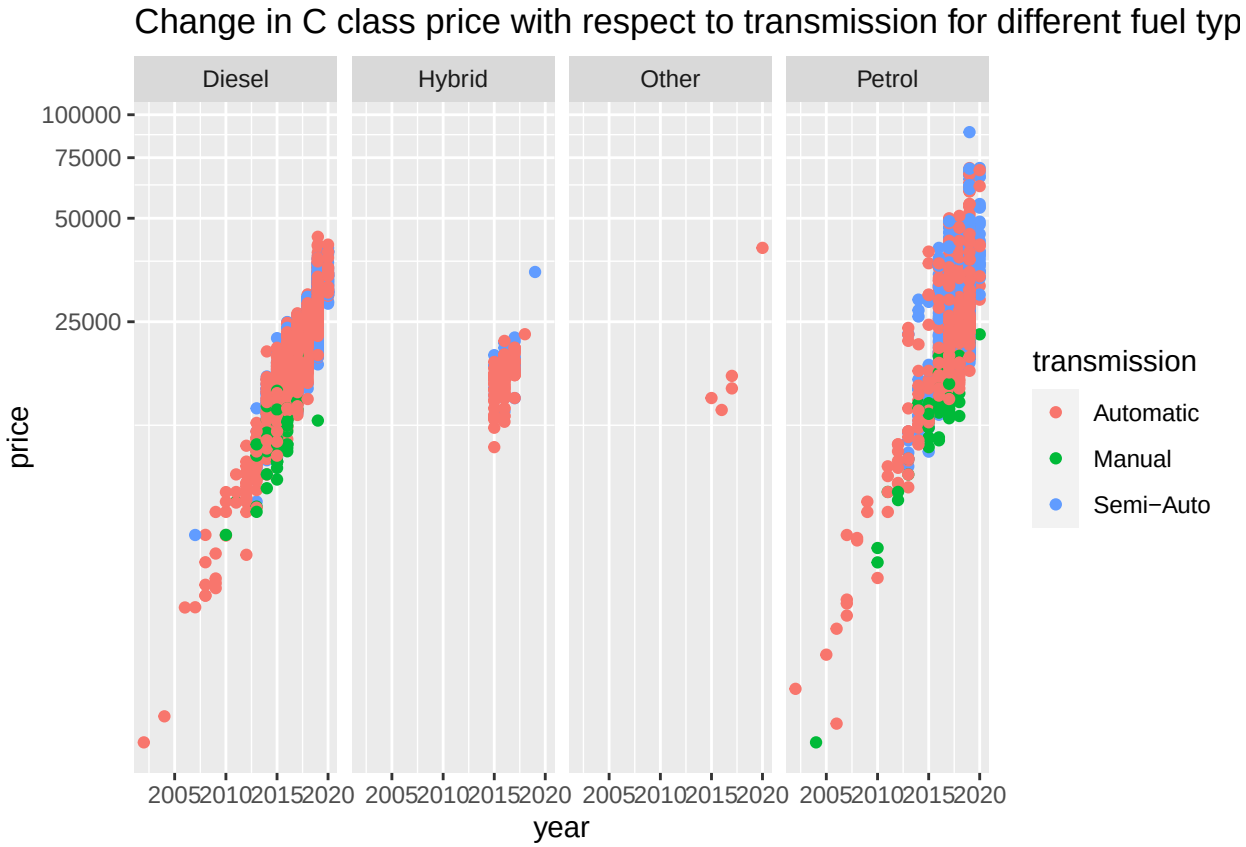
```
mydata_diesel <- dplyr::filter(mydata, fuelType=="Diesel")
mean_pricesDiesel <- aggregate(mydata_diesel$price, by = list(mydata_diesel$model), FUN = mean)
colnames(mean_pricesDiesel) <- c("Model", "Mean_Price")
library(ggplot2)
ggplot(mean_pricesDiesel)+geom_point(aes(x = Model, y = Mean_Price),color="blue") +
  labs(x = "Car Model", y = "Mean Price") +
  ggtitle("Mean Prices of Used Mercedes Car Models of diesel variants") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



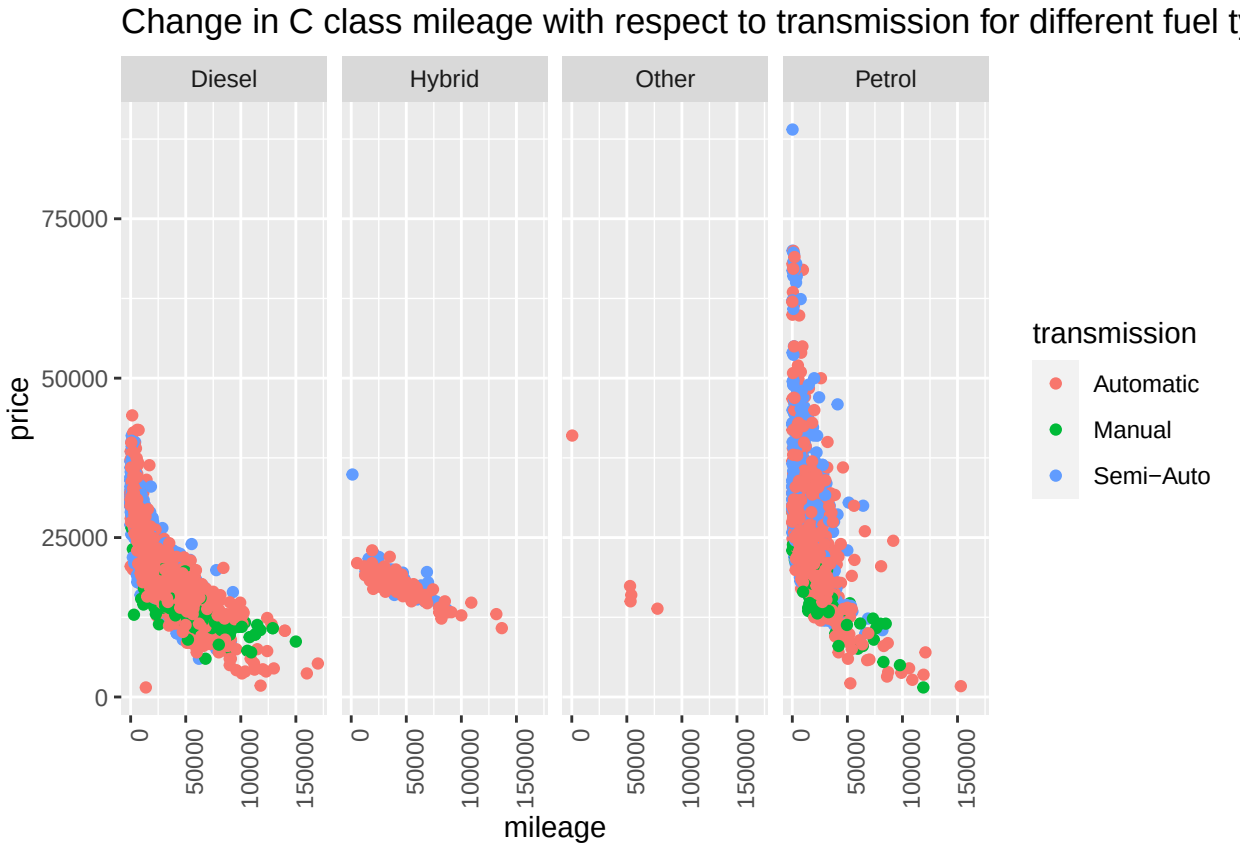
- Here also we can see that price point of G-CLASS dominates the other models which was expected. But the price of Diesel variant of G class is lower significantly than that of the petrol variant which answers our question as to which petrol models have higher price point. So the reason is solely dependent on the model type.
- We can also see that other models like A CLASS and C class of Diesel variants also have lower price points when compared to Petrol variants. .
- Now we will try to understand how price of C Class which is the most re-sold car varies with different factors

```
Cclass_data <- subset(mydata, model == " C Class")

library(ggplot2)
ggplot(Cclass_data)+geom_point(aes(x = year, y = price,colour=transmission))+ggtitle("Change in C class
```



- Clearly , we can observe that in case of C-class , the manual mode of transmission has slowly phased out from the re-sale market and is being replaced by the Semi-auto type transmission
- The re-sale price point for both Diesel and Petrol variants of C class slowly has increased manifolds over last 15 years.
- The semi-automatic transmission type is slowly dominating the automatic type in the re-sale market both in Petrol and Diesel,whereas for hybrid fuel-type,it is still dominated by automatic type.
- Another important observation for both A-class and C -class is that both entered the re-sale market well after 2005.
- Now lets see how the prices of C class varied with respect to mileage



- As expected we can see that re-sale price point dropped drastically when the mileage increased.
- The striking outcome is that in the re-sale data, majority petrol variant C -class models have a mileage in the range of 0 to 50000 mileas whereas that of Diesel is 25000 to 75000 miles.
- So the reason behind higher price points of petrol variants over diesel variants is clearly because overall mileage of the petrol variant models is lower than that of C-class models.

Result

1. The Mercedes resale car market really flourished after 2005 where although initially the automatic transmission type dominated alongwith manual, they are slowly being overtaken by semi-auto type transmission.
2. A class and C class are the most resold models of Mercedes among which C-Class is the highest.
3. For every model except SL Class, Diesel fueltype variants have dominated the re-sale market over Petrol type.
4. G-Class models have the highest price. In general, for every model, petrol variants in the re-sale market has a higher mean price than diesel variants.
5. Manual mode of transmission has slowly phased out from the re-sale market.
6. Re-sale prices of A-Class and C-Class have increased manifolds in the recent year and generally price points are higher when the mileage is lower and registration year is not old.

Conclusion

Mercedes re-sale car market in the Europe has really flourished over the years since 2005 attracting young executives and working professionals to get hold of the prestigious Mercedes Benz cars. Most preferred model types are A class and C class which are kind of entry level models for Mercedes with an affordable price of around 50,000 pounds.Although the electric era has boomed, for Mercedes, the preferred fuel type is still Diesel followed by Petrol and the preferred transmission has become semi-automatic. Manual transmission has lost its charm in the market and is slowly getting phased out. Inspite of the advent of covid in 2020, the market was steadfast and in fact prices increased during this period.