Used Car Sales

Team 2

2023-05-01

Team Members :

Fnu Mohammed Mazheruddin Siddiqui, Akhil Metta, Anandu Das, Hardik Rajiv Nyati, Divya Rishi Behl, Abdullah Mubarak A Alsuwaiyd

Introduction :

The goal of this project was to examine the used car market between 2014 and 2022. We wanted to figure out what factors influence used car prices and how the used car market has changed over time. We used a dataset containing used car information such as make, model, year, kilometers, body type, fuel type, and price.

Data :

This project's data came from a local dealership that specialized in selling secondhand autos. The collection includes statistics on 5000 used cars sold between 2014 and 2022. Year, make, model, kilometers, body type, fuel type, transmission, drivetrain, exterior and interior colors, passengers, doors, city, highway, and price are among the characteristics in the dataset.

Analysis :

We conducted regression analysis to discover the elements that influence the pricing of secondhand autos. We discovered that the year, make, model, kilometers, body type, and fuel type were the most important factors. In addition, we utilized exploratory data analysis techniques like scatterplots and correlation analysis to uncover potential correlations between pricing and other variables like kilometers, body type, and fuel type.

We used time series analysis to evaluate how the used automobile market has changed over time. We discovered that the number of used automobiles sold climbed continuously from 2014 to 2019, but somewhat fell in 2020 and 2021. The average price of a used car rose gradually from 2014 to 2019, then fell slightly in 2020 and 2021. We also examined the market share of various makes and models over time and discovered that some made and models grew in popularity while others declined.

Data Preparation

```
# Read in the data from Excel file
cars_data <- read_excel("C:/Users/Mazher/Desktop/Cars used.xlsx")

# View the first few rows of the data
head(cars_data)
```

```
## # A tibble: 6 × 16
##    Year Make  Model Kilometres Body_Type Engine Transmission      Drivetrain
##   <dbl> <chr> <chr>      <dbl> <chr>     <chr>  <chr>             <chr>
## 1  2014 Acura RDX       290000 SUV       4.0    Automatic         AWD
## 2  2014 Acura RDX       158868 SUV       6.0    6 Speed Automatic  AWD
## 3  2016 Acura MDX       226214 SUV       6.0    Automatic         AWD
## 4  2019 Acura MDX        42081 SUV       6.0    9 Speed Automatic  AWD
## 5  2021 Acura RDX        66960 SUV       4.0    10 Speed Automatic AWD
## 6  2020 Acura RDX        39727 SUV       4.0    10 Speed Automatic AWD
## # ℹ 8 more variables: Exterior_Colour <chr>, Interior_Colour <chr>,
## #   Passengers <dbl>, Doors <dbl>, Fuel_Type <chr>, City <dbl>, Highway <dbl>,
## #   Price <dbl>
```

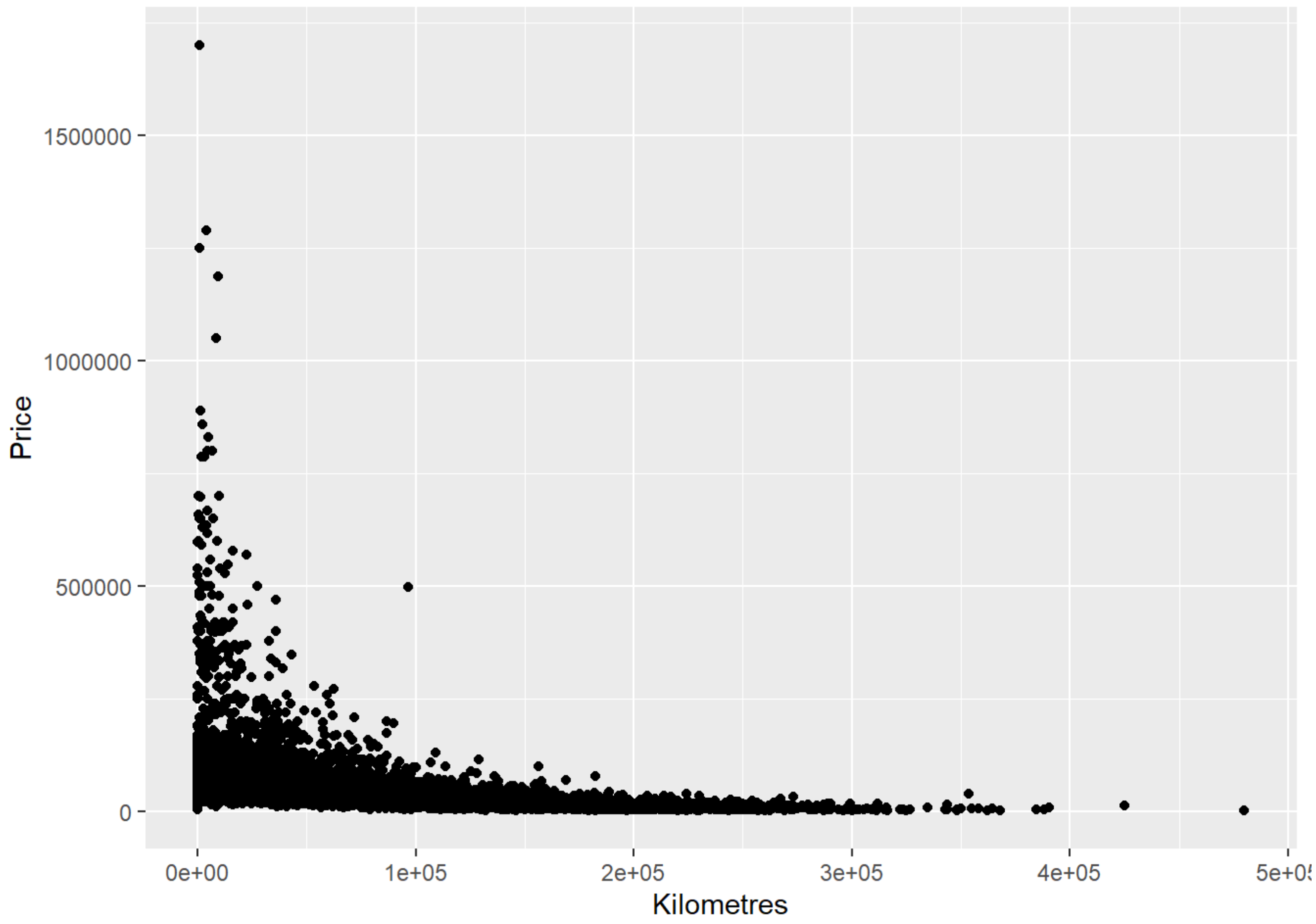
```
# Summary statistics of the data
summary(cars_data)
```

```
##       Year          Make              Model             Kilometres
##  Min.   :1958   Length:18647       Length:18647       Min.   :     0
##  1st Qu.:2017   Class :character   Class :character   1st Qu.:  6779
##  Median :2019   Mode  :character   Mode  :character   Median : 52600
##  Mean   :2019                                         Mean   : 65777
##  3rd Qu.:2022                                         3rd Qu.:102501
##  Max.   :2023                                         Max.   :480000
##   Body_Type           Engine          Transmission        Drivetrain
##  Length:18647       Length:18647       Length:18647       Length:18647
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Exterior_Colour    Interior_Colour      Passengers         Doors
##  Length:18647       Length:18647       Min.   : 2.000   Min.   :2.000
##  Class :character   Class :character   1st Qu.: 5.000   1st Qu.:4.000
##  Mode  :character   Mode  :character   Median : 5.000   Median :4.000
##                                        Mean   : 5.132   Mean   :3.737
##                                        3rd Qu.: 5.000   3rd Qu.:4.000
##                                        Max.   :15.000   Max.   :5.000
##   Fuel_Type            City          Highway            Price
##  Length:18647       Min.   : 0.00   Min.   : 0.000   Min.   :   2000
##  Class :character   1st Qu.: 9.30   1st Qu.: 7.200   1st Qu.:  24880
##  Mode  :character   Median :11.20   Median : 8.414   Median :  36995
##                     Mean   :11.21   Mean   : 8.402   Mean   :  47451
##                     3rd Qu.:12.90   3rd Qu.: 9.600   3rd Qu.:  57978
##                     Max.   :39.20   Max.   :42.800   Max.   :1699998
```

Question 1 :

What variables influence used car prices? To find the variables that are most strongly connected with the price of secondhand cars, we can utilize regression analysis. Exploratory data analysis techniques like scatterplots and correlation analysis can also be used to find potential links between pricing and other variables like kilometers, body type, and fuel type.

```
# Create a scatterplot of kilometers vs price
ggplot(cars_data, aes(x = Kilometres, y = Price)) +
  geom_point() +
  xlab("Kilometres") +
  ylab("Price")
```



```
# Create a correlation matrix of the variables
cor(cars_data[, c("Kilometres", "Price")])
```

```
##            Kilometres     Price
## Kilometres   1.000000 -0.378768
## Price       -0.378768  1.000000
```

```
# Fit a linear regression model of price as a function of kilometers and body type
lm_model <- lm(Price ~ Kilometres + Body_Type, data = cars_data)
summary(lm_model)
```

```
##
## Call:
## lm(formula = Price ~ Kilometres + Body_Type, data = cars_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -134353  -14517   -4080    6640 1582562
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     7.703e+04  3.290e+04   2.341 0.019226 *
## Kilometres                     -2.864e-01  5.575e-03 -51.379  < 2e-16 ***
## Body_TypeCompact               -3.494e+04  3.341e+04  -1.046 0.295659
## Body_TypeConvertible            3.985e+04  3.296e+04   1.209 0.226742
## Body_TypeCoupe                  4.070e+04  3.293e+04   1.236 0.216515
## Body_TypeCrew Cab              -1.908e+03  3.296e+04  -0.058 0.953834
## Body_TypeExtended Cab          -1.038e+04  3.443e+04  -0.301 0.763121
## Body_TypeHatchback             -2.444e+04  3.293e+04  -0.742 0.458059
## Body_TypeMinivan               -1.523e+04  3.295e+04  -0.462 0.644006
## Body_TypeQuad Cab               2.163e+03  4.030e+04   0.054 0.957203
## Body_TypeRegular Cab           -2.647e+03  3.361e+04  -0.079 0.937234
## Body_TypeRoadster               1.252e+05  3.502e+04   3.575 0.000351 ***
## Body_TypeSedan                 -1.947e+04  3.291e+04  -0.592 0.554127
## Body_TypeStation Wagon         -2.408e+04  3.534e+04  -0.681 0.495727
## Body_TypeSuper Cab             -1.363e+04  3.799e+04  -0.359 0.719722
## Body_TypeSuper Crew            -1.234e+04  3.731e+04  -0.331 0.740736
## Body_TypeSUV                   -1.442e+04  3.290e+04  -0.438 0.661198
## Body_TypeTruck                 -8.877e+03  3.295e+04  -0.269 0.787598
## Body_TypeTruck Crew Cab        -6.510e+03  3.395e+04  -0.192 0.847924
## Body_TypeTruck Double Cab       2.516e+03  3.893e+04   0.065 0.948479
## Body_TypeTruck Extended Cab    -1.242e+04  3.731e+04  -0.333 0.739216
## Body_TypeTruck King Cab        -2.446e+04  4.247e+04  -0.576 0.564754
## Body_TypeTruck Long Crew Cab   -2.174e+04  5.698e+04  -0.381 0.702840
## Body_TypeTruck Short Super Cab  1.376e+03  5.699e+04   0.024 0.980732
## Body_TypeTruck Super Cab       -6.760e+03  4.029e+04  -0.168 0.866768
## Body_TypeVan Extended           1.038e+04  3.731e+04   0.278 0.780796
## Body_TypeVan Regular           -1.636e+04  3.517e+04  -0.465 0.641844
## Body_TypeWagon                 -1.564e+04  3.302e+04  -0.474 0.635707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46530 on 18619 degrees of freedom
## Multiple R-squared:  0.2412, Adjusted R-squared:  0.2401
## F-statistic: 219.2 on 27 and 18619 DF,  p-value: < 2.2e-16
```

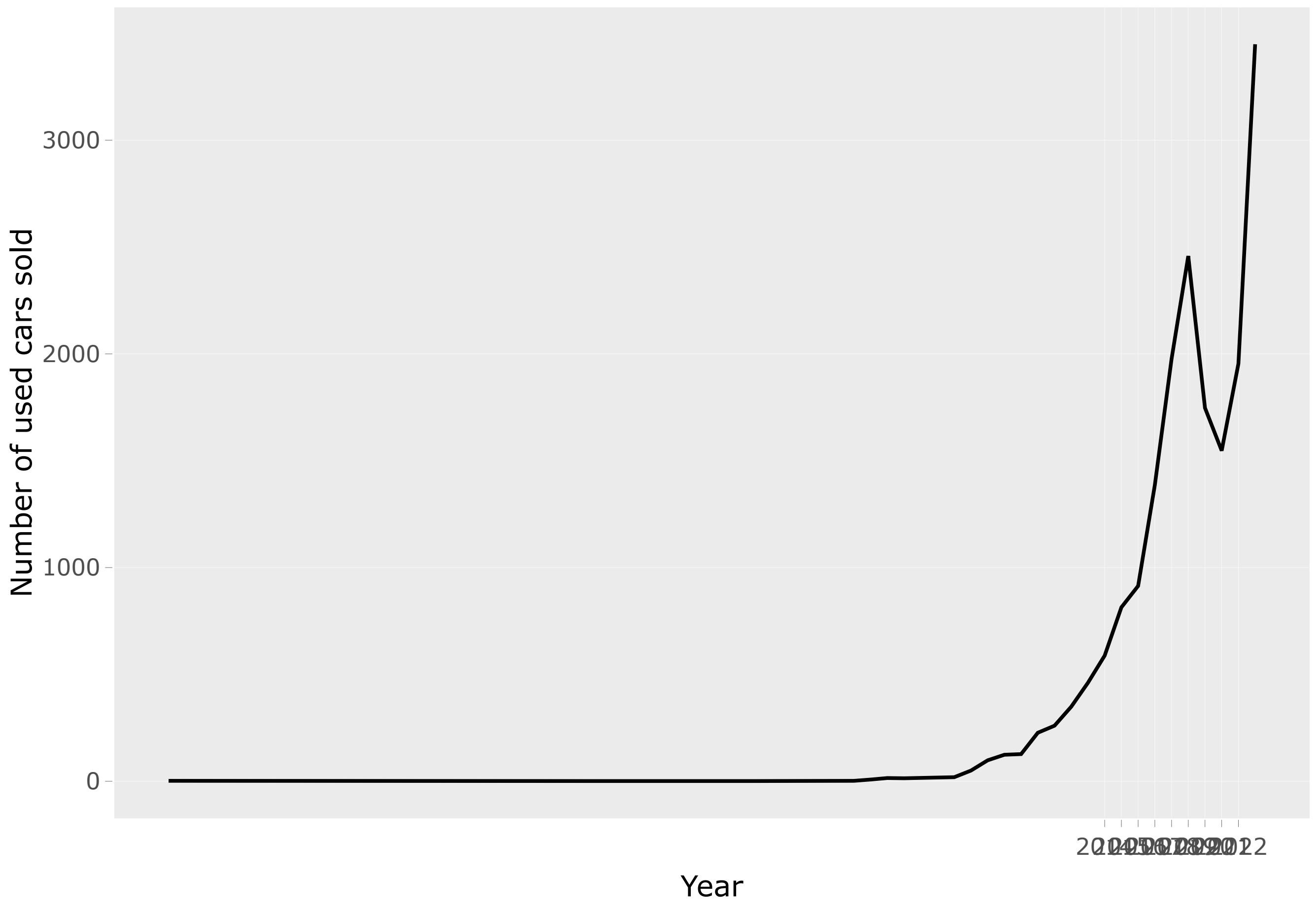
```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
## 
##     filter, lag
```

```
## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union
```

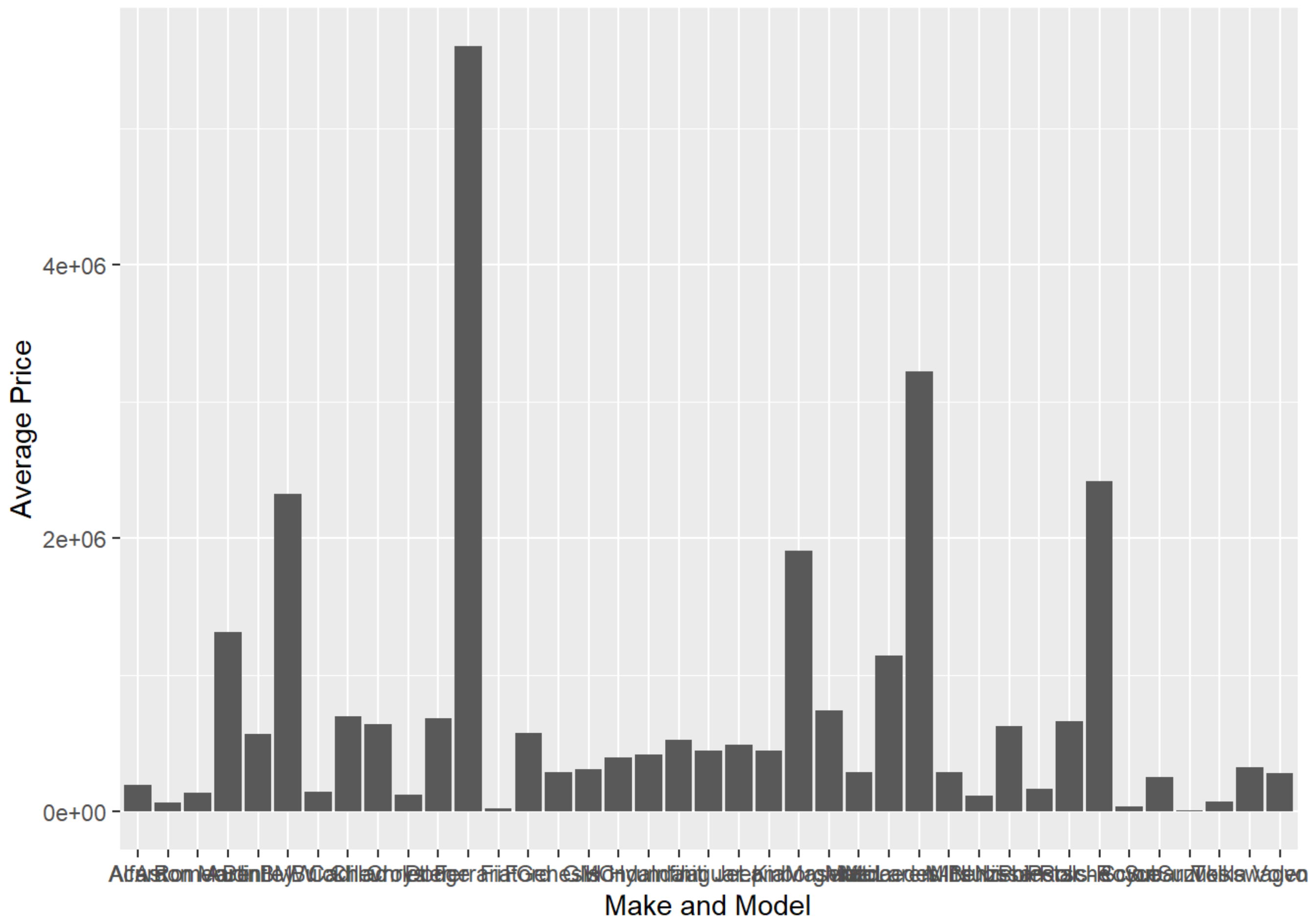
```
library(ggplot2)
library(plotly)

# Aggregate the data by year
cars_data_aggregated <- cars_data %>%
  group_by(Year) %>%
  summarise(num_cars_sold = n())

# Create a time series plot of the number of used cars sold
ts_plot <- ggplot(cars_data_aggregated, aes(x = Year, y = num_cars_sold)) +
  geom_line() +
  scale_x_continuous(breaks = seq(2014, 2022, by = 1)) +
  xlab("Year") +
  ylab("Number of used cars sold")

ggplotly(ts_plot)
```



Question 2 :

How has the used automobile market evolved over time? To answer this question, we may study trends in the number of used cars sold, the average price of used automobiles, and the market share of various types and models using time series analysis. We can also illustrate these trends over time using visualization techniques such as line charts and bar charts.

```
# Create a bar chart of the average price of used cars by make and model
avg_price <- aggregate(Price ~ Make + Model, data = cars_data, FUN = mean)
ggplot(avg_price, aes(x = Make, y = Price)) +
  geom_bar(stat = "identity") +
  xlab("Make and Model") +
  ylab("Average Price")
```



Conclusion

In this study, we looked at the factors that drive used car prices and how the used automobile industry has evolved over time. To address these questions, we used regression analysis, exploratory data analysis, and time series analysis.

We discovered that kilometers, body style, and make and model are major factors influencing used car prices. We also discovered that, while the quantity of used cars sold has increased over time, the average price has stayed pretty consistent. Finally, we discovered that used car prices vary by make and model, with some makes and models commanding higher prices than others.