

# Toyota Used Car Analysis

Natatkit

2025-03-31

## Introduction

Toyota is one of the most popular car brands in the used car market. This report analyzes the factors affecting car prices, such as mileage, year, fuel type, and transmission. The goal is to provide insights that help dealerships optimize pricing and inventory.

## 1. Load and Explore Data

```
df <- read_csv("toyota.csv")

## Rows: 6738 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (3): model, transmission, fuelType
## dbl (6): year, price, mileage, tax, mpg, engineSize
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
# Display first few rows
head(df)

## # A tibble: 6 x 9
##   model  year price transmission mileage fuelType   tax   mpg engineSize
##   <chr> <dbl> <dbl> <chr>         <dbl> <chr>   <dbl> <dbl>      <dbl>
## 1 GT86   2016 16000 Manual         24089 Petrol    265  36.2         2
## 2 GT86   2017 15995 Manual         18615 Petrol    145  36.2         2
## 3 GT86   2015 13998 Manual         27469 Petrol    265  36.2         2
## 4 GT86   2017 18998 Manual         14736 Petrol    150  36.2         2
## 5 GT86   2017 17498 Manual         36284 Petrol    145  36.2         2
## 6 GT86   2017 15998 Manual         26919 Petrol    260  36.2         2

glimpse(df)

## Rows: 6,738
## Columns: 9
## $ model      <chr> "GT86", "GT86", "GT86", "GT86", "GT86", "GT86", "GT86", "~
## $ year       <dbl> 2016, 2017, 2015, 2017, 2017, 2017, 2017, 2017, 2020, 201~
## $ price      <dbl> 16000, 15995, 13998, 18998, 17498, 15998, 18522, 18995, 2~
## $ transmission <chr> "Manual", "Manual", "Manual", "Manual", "Manual", "Manual~
## $ mileage     <dbl> 24089, 18615, 27469, 14736, 36284, 26919, 10456, 12340, 5~
## $ fuelType    <chr> "Petrol", "Petrol", "Petrol", "Petrol", "Petrol", "Petrol~
## $ tax        <dbl> 265, 145, 265, 150, 145, 260, 145, 145, 150, 265, 265, 14~
```

```
## $ mpg          <dbl> 36.2, 36.2, 36.2, 36.2, 36.2, 36.2, 36.2, 36.2, 33.2, 36.~
## $ engineSize   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
```

```
summary(df)
```

```
##      model          year      price      transmission
## Length:6738      Min.   :1998      Min.    : 850      Length:6738
## Class :character  1st Qu.:2016      1st Qu.: 8290      Class :character
## Mode  :character  Median :2017      Median :10795     Mode  :character
##                      Mean   :2017      Mean   :12522
##                      3rd Qu.:2018      3rd Qu.:14995
##                      Max.    :2020      Max.    :59995
##      mileage      fuelType      tax      mpg
## Min.   :      2      Length:6738      Min.    : 0.0      Min.    : 2.80
## 1st Qu.: 9446      Class :character  1st Qu.: 0.0      1st Qu.: 55.40
## Median :18513      Mode  :character  Median :135.0     Median : 62.80
## Mean   :22857                      Mean   : 94.7      Mean   : 63.04
## 3rd Qu.:31064                      3rd Qu.:145.0     3rd Qu.: 69.00
## Max.   :174419                      Max.    :565.0     Max.    :235.00
##      engineSize
## Min.    :0.000
## 1st Qu.:1.000
## Median :1.500
## Mean    :1.471
## 3rd Qu.:1.800
## Max.    :4.500
```

## 2. Data Cleaning & Preparation

```
# Remove missing values
df <- df %>% drop_na()

# Remove duplicates
df <- df %>% distinct()

# Convert categorical variables
df$fuelType <- as.factor(df$fuelType)
df$transmission <- as.factor(df$transmission)

# Create new variable: Car Age
df <- df %>% mutate(Car_Age = 2024 - year)
head(df)
```

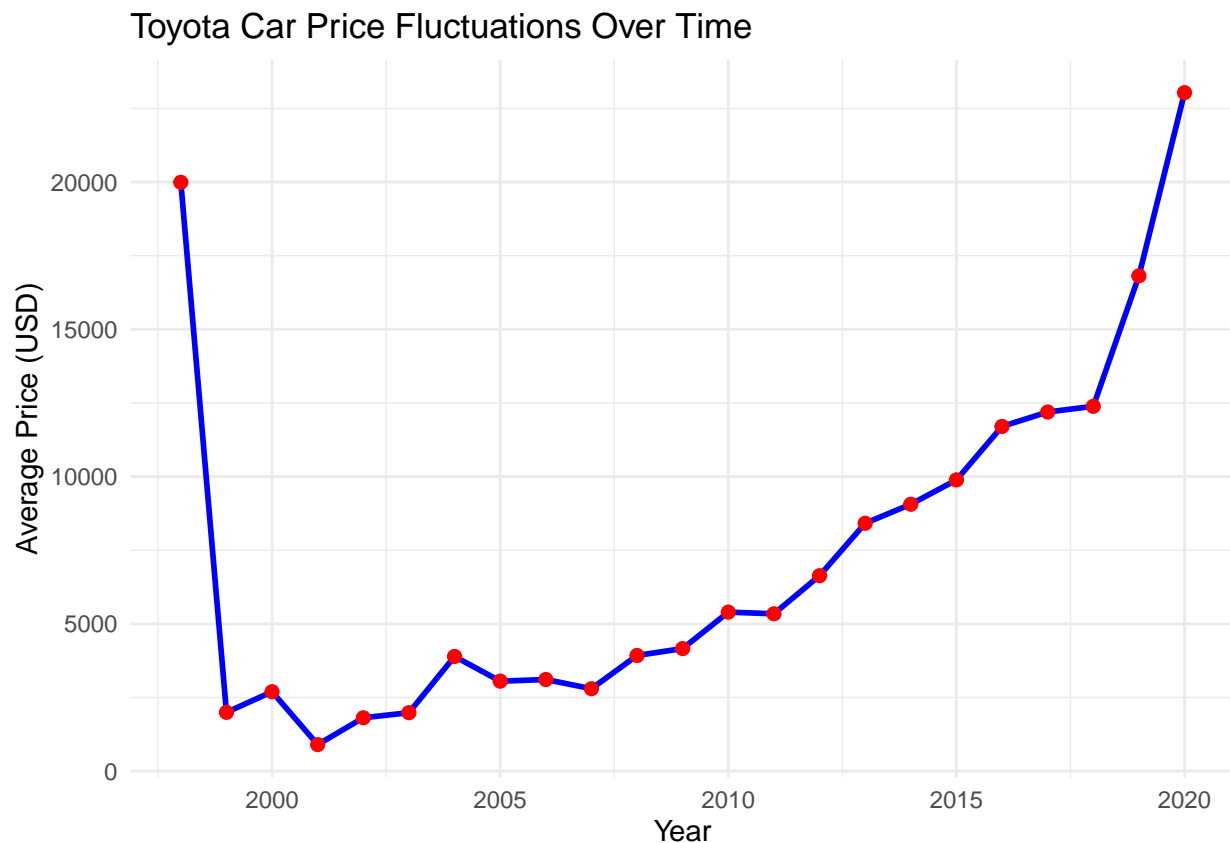
```
## # A tibble: 6 x 10
##   model year price transmission mileage fuelType tax mpg engineSize Car_Age
##   <chr> <dbl> <dbl> <fct>         <dbl> <fct>   <dbl> <dbl>      <dbl>   <dbl>
## 1 GT86  2016 16000 Manual         24089 Petrol   265  36.2         2       8
## 2 GT86  2017 15995 Manual         18615 Petrol   145  36.2         2       7
## 3 GT86  2015 13998 Manual         27469 Petrol   265  36.2         2       9
## 4 GT86  2017 18998 Manual         14736 Petrol   150  36.2         2       7
## 5 GT86  2017 17498 Manual         36284 Petrol   145  36.2         2       7
## 6 GT86  2017 15998 Manual         26919 Petrol   260  36.2         2       7
```

### 3. Exploratory Data Analysis (EDA)

#### 3.1 Plot price trends over time

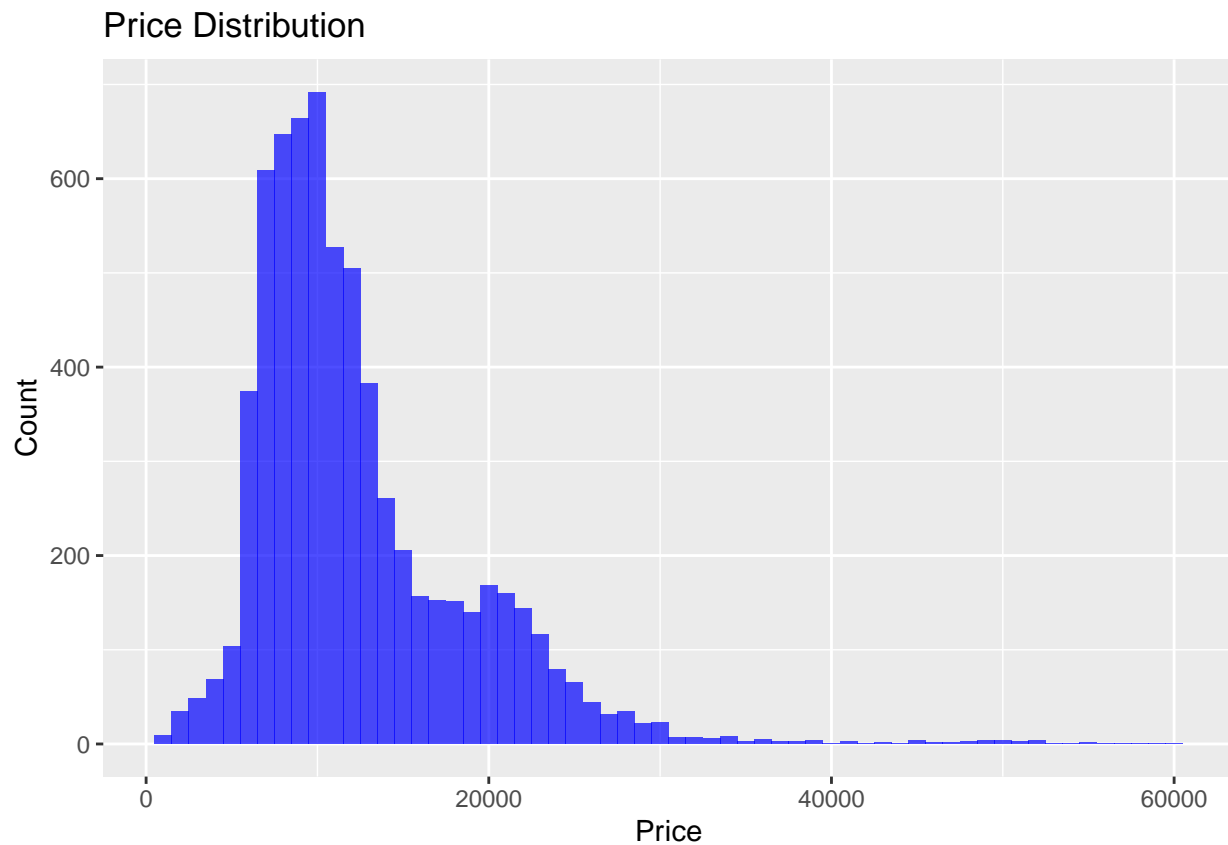
```
# Calculate average price per year
price_trends <- df %>%
  group_by(year) %>%
  summarise(Average_Price = mean(price, na.rm = TRUE))

ggplot(price_trends, aes(x = year, y = Average_Price)) +
  geom_line(color = "blue", linewidth = 1) +
  geom_point(color = "red", size = 2) +
  labs(title = "Toyota Car Price Fluctuations Over Time",
       x = "Year",
       y = "Average Price (USD)") +
  theme_minimal()
```



#### 3.2 Price Distribution

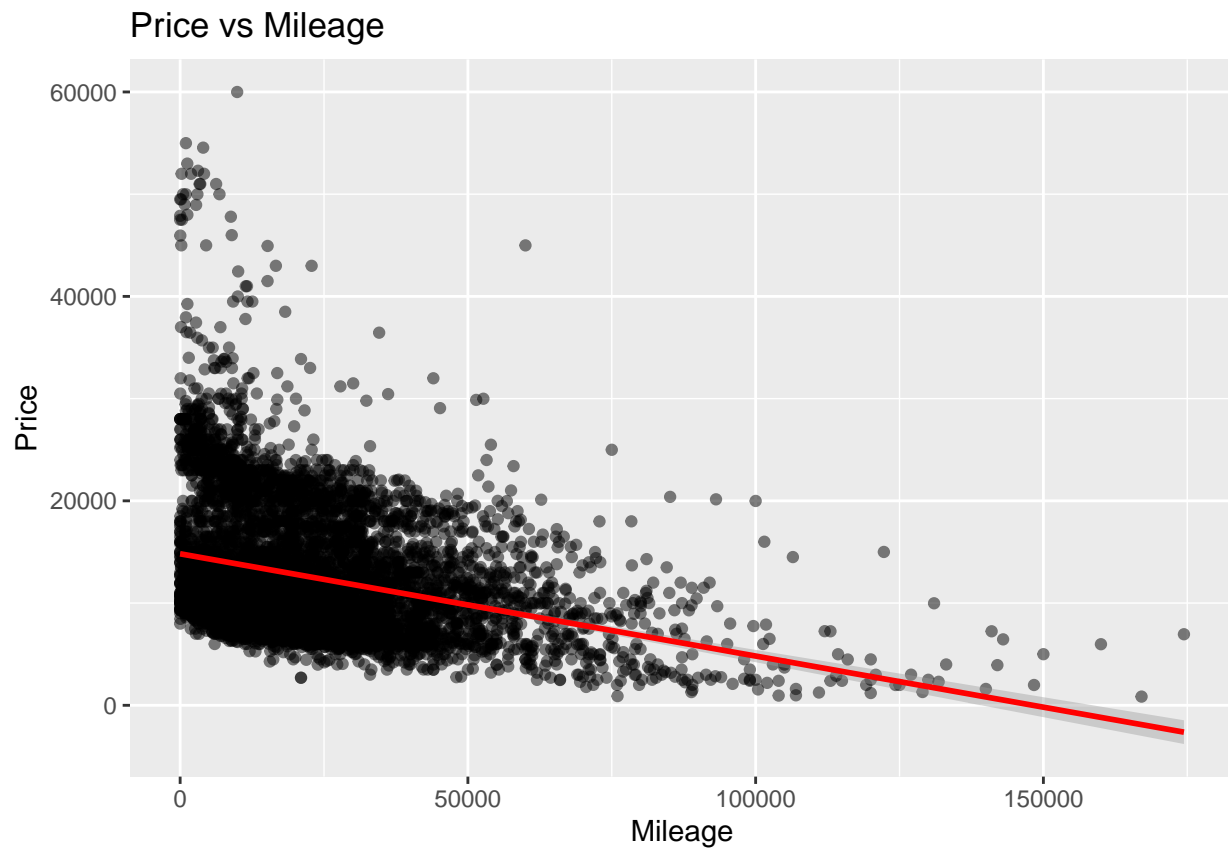
```
ggplot(df, aes(x = price)) +
  geom_histogram(binwidth = 1000, fill = "blue", alpha = 0.7) +
  labs(title = "Price Distribution", x = "Price", y = "Count")
```



### 3.3 Relationship Between Mileage and Price

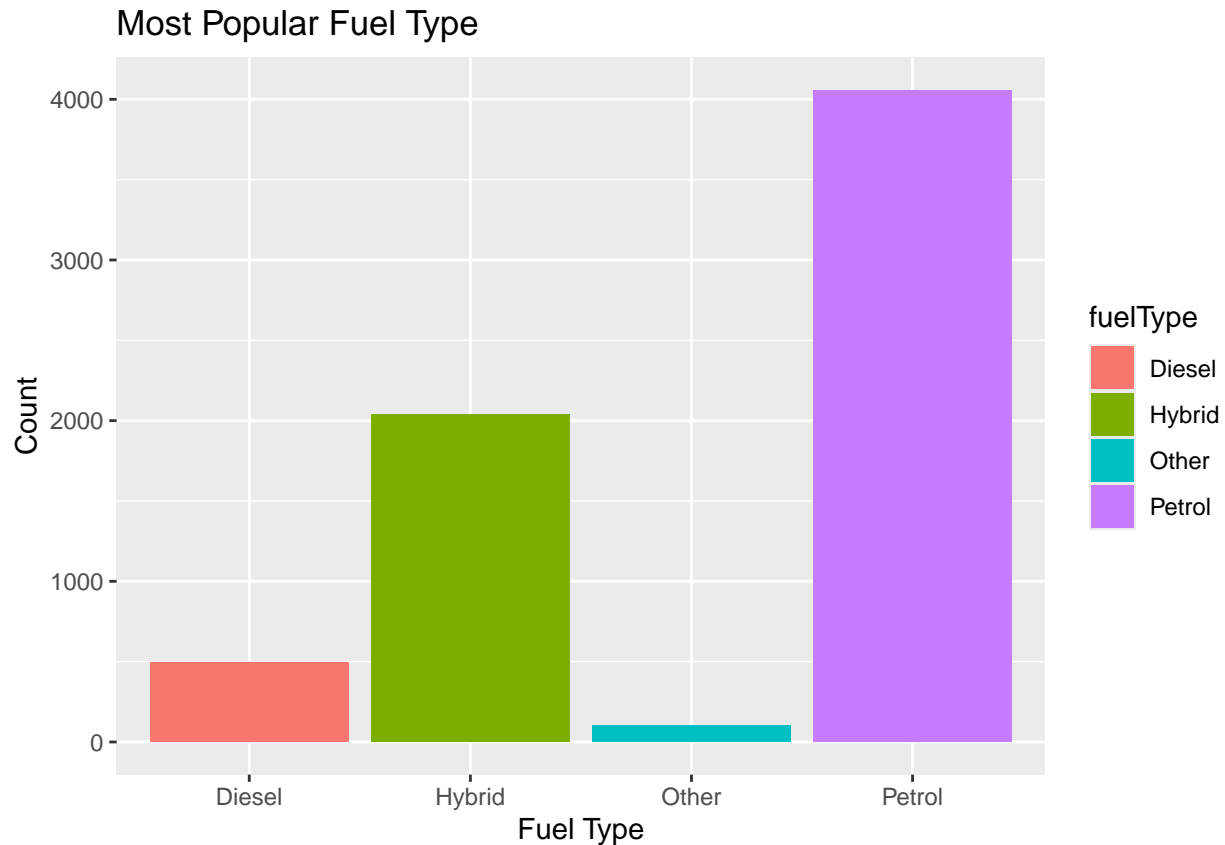
```
ggplot(df, aes(x = mileage, y = price)) +  
  geom_point(alpha = 0.5) +  
  geom_smooth(method = "lm", color = "red") +  
  labs(title = "Price vs Mileage", x = "Mileage", y = "Price")
```

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



### 3.4 Most Popular Fuel Type

```
ggplot(df, aes(x = fuelType, fill = fuelType)) +  
  geom_bar() +  
  labs(title = "Most Popular Fuel Type", x = "Fuel Type", y = "Count")
```



## 4. Insights & Recommendations

- **Price Trends:** Most used Toyota cars are priced between \$5,000 and \$15,000.
- **Fuel Preference:** Petrol cars dominate, but hybrid models are gaining popularity.
- **Mileage Effect:** Higher mileage leads to lower car prices, indicating depreciation trends.
- **Business Recommendation:** Dealerships should stock fuel-efficient hybrid models and price high-mileage cars competitively.

## 5. Conclusion

This analysis helps car dealerships and buyers make data-driven decisions. Data-driven pricing helps dealerships maximize profit. Petrol cars dominate, but hybrids are on the rise. Higher mileage reduces price significantly