Used Car Sales Analysis

Taaran Reddy

2023-10-28

Introduction

The used car market offers a platform where buyers and sellers of pre-owned vehicles can transact. The value of a used car depends on a myriad of factors, and understanding these can be crucial for both buyers and sellers. This analysis seeks to understand and predict the selling price of used cars based on various features.

Dataset Description

The dataset provides details about various used cars listed for sale. Below are the features available:

- **Car_Name**: Name of the car.
- **Year**: The manufacturing year of the car.
- **Selling_Price**: Offered price of the car (in lakhs).
- **Present_Price**: Current market price of the car (in lakhs).
- **Kms_Driven**: Total kilometers the car has been driven.
- **Fuel_Type**: The fuel type of the car (e.g., Petrol, Diesel).
- **Seller_Type**: The category of the seller (e.g., Dealer, Individual).
- **Transmission**: Car's transmission type (e.g., Manual, Automatic).
- **Owner**: Number of previous owners the car has had.

Data Loading and Exploration

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.3     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.4     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
s
```

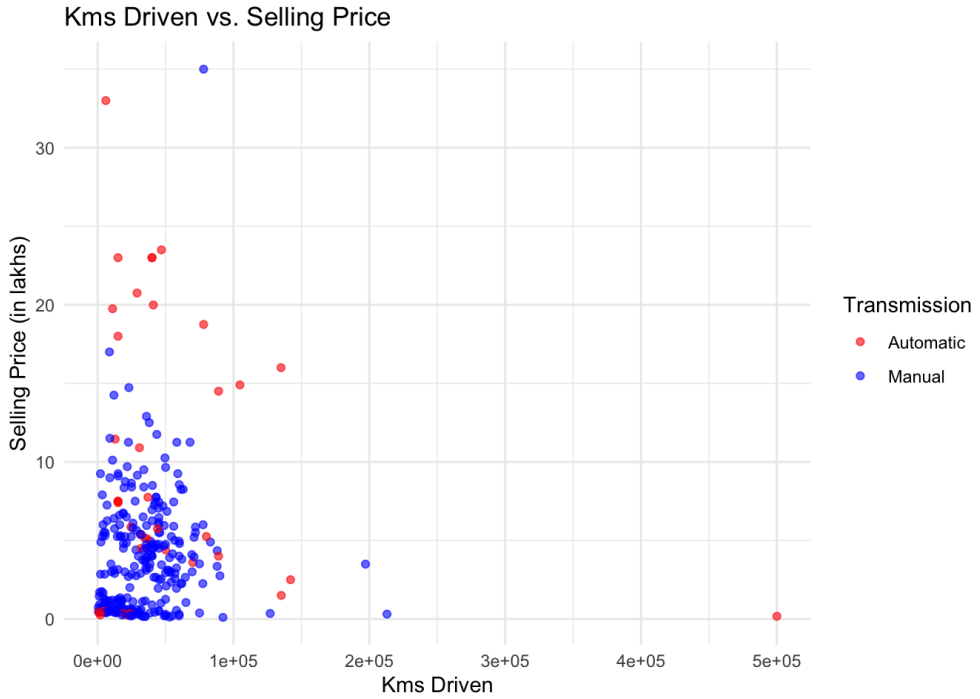
```
dataset <- read.csv("car_sales.csv")

head(dataset)
```

```
##        Car_Name Year Selling_Price Present_Price Kms_Driven Fuel_Type
## 1          ritz 2014          3.35          5.59      27000    Petrol
## 2           sx4 2013          4.75          9.54      43000    Diesel
## 3         ciaz 2017          7.25          9.85       6900    Petrol
## 4      wagon r 2011          2.85          4.15       5200    Petrol
## 5        swift 2014          4.60          6.87      42450    Diesel
## 6 vitara brezza 2018          9.25          9.83       2071    Diesel
##   Seller_Type Transmission Owner
## 1      Dealer       Manual     0
## 2      Dealer       Manual     0
## 3      Dealer       Manual     0
## 4      Dealer       Manual     0
## 5      Dealer       Manual     0
## 6      Dealer       Manual     0
```

```
ggplot(dataset, aes(x = Selling_Price)) +
  geom_histogram(fill = "blue", color = "white", binwidth = 0.5, alpha = 0.7) +
  labs(title = "Distribution of Selling Prices", x = "Selling Price (in lakhs)", y = "Number of Cars")
+
  theme_minimal()
```


```
ggplot(dataset, aes(x = Kms_Driven, y = Selling_Price)) +
  geom_point(aes(color = Transmission), alpha = 0.6) +
  scale_color_manual(values = c("Manual" = "blue", "Automatic" = "red")) +
  labs(title = "Kms Driven vs. Selling Price", x = "Kms Driven", y = "Selling Price (in lakhs)") +
  theme_minimal()
```

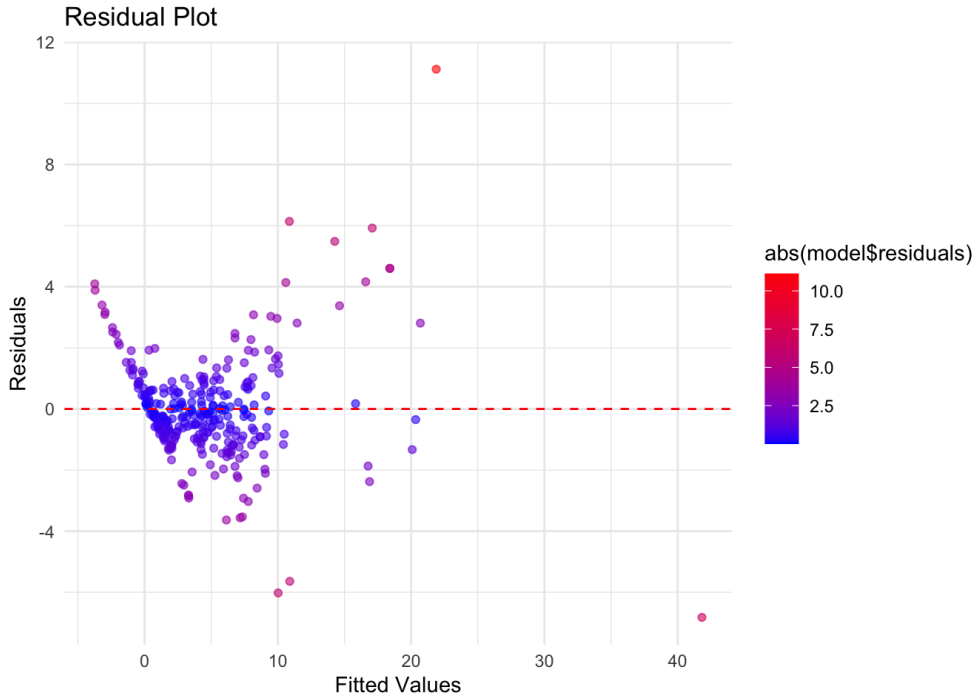



```
ggplot(dataset, aes(x = Transmission, y = Selling_Price, fill = Transmission)) +
  geom_boxplot(alpha = 0.7) +
  labs(title = "Impact of Transmission Type on Selling Price", x = "Transmission Type", y = "Selling Pr
ice (in lakhs)") +
  theme_minimal()
```


```
model <- lm(Selling_Price ~ . - Car_Name, data = dataset)
summary(model)
```

```
##
## Call:
## lm(formula = Selling_Price ~ . - Car_Name, data = dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8213 -0.8870 -0.2023  0.6766 11.1175
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -7.892e+02  8.652e+01  -9.122  < 2e-16 ***
## Year                   3.931e-01  4.292e-02   9.159  < 2e-16 ***
## Present_Price          4.372e-01  1.598e-02  27.355  < 2e-16 ***
## Kms_Driven            -7.013e-06  3.229e-06  -2.172   0.0307 *
## Fuel_TypeDiesel        2.470e+00  1.278e+00   1.932   0.0543 .
## Fuel_TypePetrol        6.077e-01  1.259e+00   0.483   0.6296
## Seller_TypeIndividual -1.121e+00  2.565e-01  -4.371 1.72e-05 ***
## TransmissionManual    -1.448e+00  3.279e-01  -4.417 1.41e-05 ***
## Owner                 -6.742e-01  4.226e-01  -1.595   0.1118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.765 on 292 degrees of freedom
## Multiple R-squared:  0.8826, Adjusted R-squared:  0.8794
## F-statistic: 274.3 on 8 and 292 DF,  p-value: < 2.2e-16
```

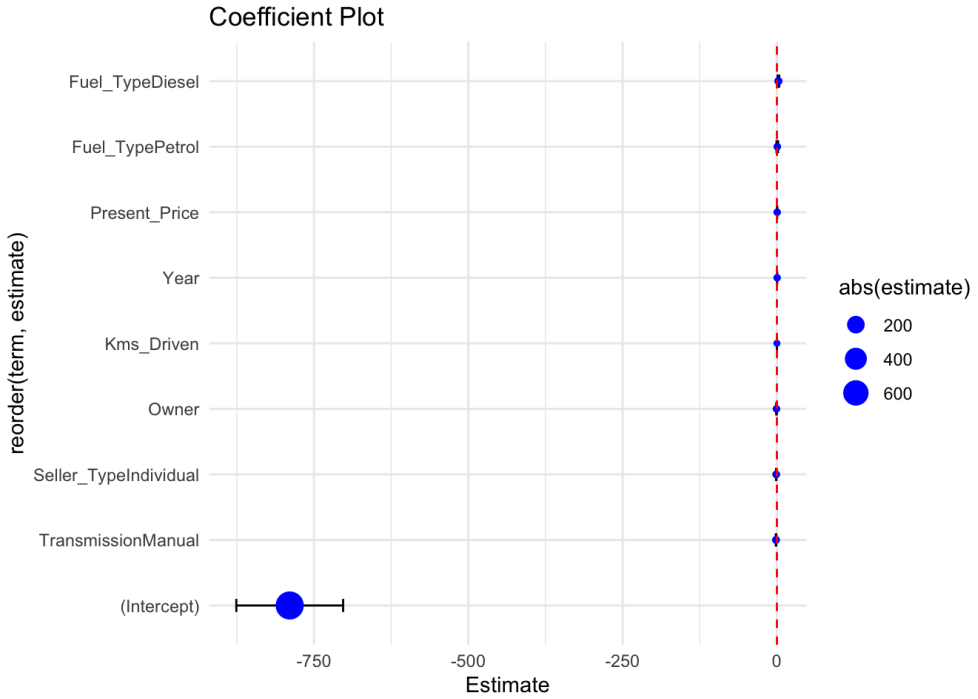
```
ggplot(data = NULL, aes(x = model$fitted.values, y = model$residuals)) +
  geom_point(aes(color = abs(model$residuals)), alpha = 0.6) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  scale_color_gradient(low = "blue", high = "red") +
  labs(title = "Residual Plot", x = "Fitted Values", y = "Residuals") +
  theme_minimal()
```




```
coef_data <- tidy(model)

ggplot(coef_data, aes(x = reorder(term, estimate), y = estimate)) +
  geom_errorbar(aes(ymin = estimate - std.error, ymax = estimate + std.error), width = 0.2) +
  geom_point(aes(size = abs(estimate)), color = "blue") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "red") +
  coord_flip() +
  labs(title = "Coefficient Plot", y = "Estimate") +
  theme_minimal()
```

- Introduction
- Dataset Description
- Data Loading and Exploration




```
dataset$predicted <- predict(model, dataset)

ggplot(dataset, aes(x = Selling_Price, y = predicted)) +
  geom_point(aes(color = abs(predicted - Selling_Price)), alpha = 0.6) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "red") +
  labs(title = "Actual vs Predicted Selling Price", x = "Actual Selling Price", y = "Predicted Selling
Price") +
  theme_minimal()
```

- Introduction
- Dataset Description
- Data Loading and Exploration