578

Yang Xiao

2023-12-10

1. Introduction to the Analysis of Used Car Market Dynamics:

This study delves into the intricacies of used car pricing, a vital aspect of the automotive industry with significant economic implications. By examining factors like make, model, year, mileage, and more, the research aims to uncover the key determinants of a used car's market value. The significance of this research lies in its potential to enhance market transparency and inform strategic decision-making in the future

2. EDA

EDA for overview

Analysis:
1.Selling Price: The distribution shows a high frequency of cars in the lower price range, indicating a market dominated by budget-friendly options.
2.Kilometers Driven: Most cars have lower kilometers, suggesting a prevalence of relatively less used vehicles. Higher km driven cars are fewer, possibly due to decreased value or desirability.
3.Fuel Type: Petrol cars outnumber diesel, reflecting consumer preference or market availability. Other fuel types are significantly less common.
4.Transmission Type: Manual transmission cars are more prevalent than automatic, possibly due to lower cost or higher availability in the used car market.
These trends provide valuable context for understanding consumer preferences and market dynamics in the used car sector.

3. Model

3.1 Bayesian Regression Model

3.1.1 Modeling

The code uses a Gaussian loss function, indicating the model minimizes squared errors, assuming normally distributed residuals. The estimators are the beta coefficients for predictors and sigma for residual variation, estimated via MCMC sampling with Stan's NUTS algorithm. Predictors likely include car attributes like year and km_driven. The approximation method, MCMC, approximates the posterior distributions, providing a range of plausible values for parameters, capturing uncertainty instead of single-point estimates, which is a key advantage of Bayesian methods.

```
print(fit,digits=4)
```

```
## Inference for Stan model: anon_model.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##              mean se_mean     sd       2.5%        25%        50%        75%
## beta[1]     0.2304  0.0001 0.0042     0.2221     0.2276     0.2304     0.2331
## beta[2]     0.0554  0.0003 0.0152     0.0262     0.0450     0.0556     0.0652
## beta[3]    -0.0052  0.0000 0.0005    -0.0063    -0.0056    -0.0052    -0.0049
## beta[4]     0.0066  0.0000 0.0017     0.0032     0.0054     0.0066     0.0078
## beta[5]    -0.0653  0.0000 0.0023    -0.0698    -0.0669    -0.0653    -0.0638
## beta[6]     0.0027  0.0000 0.0007     0.0014     0.0023     0.0027     0.0032
## sigma       0.0519  0.0000 0.0006     0.0508     0.0516     0.0519     0.0523
## lp__    10664.5807  0.0450 1.8201 10660.3236 10663.5411 10664.8967 10665.9248
##              97.5% n_eff   Rhat
## beta[1]     0.2387  2429 0.9998
## beta[2]     0.0848  1896 1.0017
## beta[3]    -0.0042  4139 1.0001
## beta[4]     0.0101  2967 0.9999
## beta[5]    -0.0609  2307 0.9998
## beta[6]     0.0041  4129 1.0004
## sigma       0.0530  2302 1.0001
## lp__    10667.1954  1634 1.0042
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 11 01:20:12 2023.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

(1) Analysis of model fitting results

1. Year (beta[1] = 0.2304): Newer cars command higher selling prices, emphasizing the depreciation effect. The significant positive coefficient indicates a strong relationship between a car's age and its market value.
2. Kilometers Driven (beta[3] = -0.0052): Cars with higher mileage are priced lower, showcasing wear and tear's impact on valuation. This negative coefficient reflects the common consumer preference for less-used vehicles.
3.Fuel Type (beta[4]): The coefficient for fuel type (assuming it corresponds to beta[4] = 0.0066) indicates a slight effect on the selling price.
4. Seller Type, Number of Owners, Transmission: These factors' coefficients (e.g., beta[5], beta[6]) suggest varying impacts on price, though specific interpretations depend on the reference categories used.
5. Model Reliability and Precision: The model's low sigma value (0.0519) indicates precise predictions with minimal error variability. Rhat values near 1 and substantial effective sample sizes (n_eff) suggest good convergence, lending credibility to the model's estimates.

3.1.2 Sensitivity Analysis Modeling

```
print(fit_new_prior, digits = 4)
```

```
## Inference for Stan model: anon_model.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##               mean se_mean     sd       2.5%        25%        50%        75%
## beta[1]     0.2303  0.0001 0.0042     0.2219     0.2275     0.2303     0.2331
## beta[2]     0.0553  0.0003 0.0148     0.0263     0.0454     0.0552     0.0655
## beta[3]    -0.0052  0.0000 0.0005    -0.0062    -0.0056    -0.0052    -0.0049
## beta[4]     0.0067  0.0000 0.0017     0.0035     0.0055     0.0066     0.0078
## beta[5]    -0.0653  0.0000 0.0022    -0.0696    -0.0668    -0.0654    -0.0638
## beta[6]     0.0027  0.0000 0.0007     0.0014     0.0023     0.0027     0.0032
## sigma       0.0520  0.0000 0.0006     0.0509     0.0516     0.0519     0.0523
## lp__    10664.5693  0.0463 1.8946 10659.9637 10663.5347 10664.8737 10665.9896
##              97.5% n_eff   Rhat
## beta[1]     0.2386  2334 1.0057
## beta[2]     0.0841  1955 1.0031
## beta[3]    -0.0042  4289 0.9994
## beta[4]     0.0101  3083 1.0011
## beta[5]    -0.0610  2528 1.0030
## beta[6]     0.0041  4161 0.9998
## sigma       0.0531  1913 0.9994
## lp__    10667.2161  1675 1.0007
##
## Samples were drawn using NUTS(diag_e) at Mon Dec 11 01:21:17 2023.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

Conclusion for Sensitivity

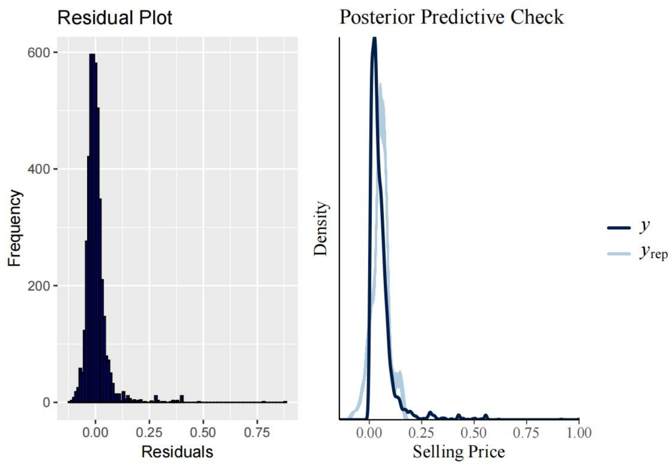
(1) The posterior distribution of parameter beta: The mean and quantile of parameter beta for the two runs are very close, indicating that the model has a certain degree of stability in estimating these parameters. Different posterior samples are also within the 95% confidence interval, indicating that the parameter estimation is consistent.

(2) The posterior distribution of parameter sigma: The mean and quantile of parameter sigma from the previous and subsequent runs are also very close, and the 95% confidence interval overlaps, indicating that the model's estimation of standard deviation is also consistent.

(3) Rhat vale:The Rhat values of both previous and subsequent runs are close to 1, indicating that the model has achieved reasonable convergence between different chains.

In summary, the results of the two runs are very similar, and the model has a certain degree of stability and consistency for different posterior sampling. This indicates that the parameter estimation of the model is reliable and not easily affected by initial conditions.

3.1.3 EDA for Bayesian Regression Model



Graphic analysis

The residual plot indicates a right-skewed distribution, suggesting the model un derpredicts for some observations. A concentration of residuals near zero suggests accurate predictions for many cases, but the long tail points to significant errors for others, potentially due to outliers or unmodeled factors.

The PPC plot reveals good model fit around the data's central tendency but poor fit at the tails. This mismatch indicates that the model might not capture the full variability of the data, especially for higher selling prices.

Overall, the model performs well for typical values but needs refinement to handle the full range of the selling price distribution, possibly by including additional predictors, investigating outliers, or introducing non-linear terms. Further model diagnostics are recommended to enhance its predictive power.

3.2 Hierarchical Bayesian Model

3.2.1 Modeling

The model employs a Gaussian family, implying a squared-error loss function to assess the fit between predicted and actual selling prices.

The predictors include both numerical (year, km_driven) and categorical (fuel, seller_type, transmission, owner) variables. The estimators are the regression coefficients for these predictors.

The model also accounts for random effects due to fuel type. For parameter estimation, the model uses MCMC with increased iterations and higher adapt_delta for convergence, reflecting a robust Bayesian inference approach. The priors for the coefficients and intercept are normally distributed, while the prior for the standard deviation is Cauchy-distributed, encapsulating prior beliefs about these parameters' distributions.

```r
# Print the model summary
summary(model)
```

```
## Warning: There were 4 divergent transitions after warmup. Increasing
## adapt_delta above 0.98 may help. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: selling_price ~ year + km_driven + seller_type + transmission + owner + (1 | f
##    Data: car_data (Number of observations: 4340)
##   Draws: 3 chains, each with iter = 1000; warmup = 500; thin = 1;
##          total post-warmup draws = 1500
##
## Group-Level Effects:
## ~fuel (Number of levels: 5)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.46      0.31     0.16     1.28 1.00      411      479
##
## Population-Level Effects:
##                            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                      1.29      0.26     0.73     1.77 1.00      500
## year                           0.26      0.01     0.23     0.28 1.00     1513
## km_driven                     -0.08      0.01    -0.11    -0.05 1.00     1795
## seller_typeIndividual         -0.11      0.03    -0.17    -0.06 1.00     1389
## seller_typeTrustmarkDealer     0.29      0.08     0.13     0.45 1.00     1728
## transmissionManual            -1.50      0.04    -1.58    -1.43 1.00     1855
## ownerFourth&AboveOwner        -0.00      0.08    -0.16     0.16 1.00     1886
## ownerSecondOwner              -0.07      0.03    -0.13    -0.02 1.00     1582
## ownerTestDriveCar              0.29      0.18    -0.07     0.63 1.00     2113
## ownerThirdOwner               -0.07      0.05    -0.16     0.02 1.00     1967
##                            Tail_ESS
```

```
## Intercept                      516
## year                          1347
## km_driven                     1100
## seller_typeIndividual         1218
## seller_typeTrustmarkDealer    1122
## transmissionManual            1315
## ownerFourth&AboveOwner        1046
## ownerSecondOwner              1106
## ownerTestDriveCar             1216
## ownerThirdOwner               1338
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.74      0.01     0.72     0.75 1.00     2352     1187
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Analysis of model fitting results

Convergence: The Rhat value approaches 1, indicating that the model has converged well.

Effective sample size: Bulk_ ESS and Tail_ The high value of ESS indicates good posterior sampling efficiency.

Predictive ability: Through the posterior prediction test (PPC) chart, it can be seen that the model's predictions match the distribution of actual data quite well.

3.2.2 Sensitivity Analysis Modeling

```
print(summary(models[[1]]))
```

```
## Warning: There were 2 divergent transitions after warmup. Increasing
## adapt_delta above 0.98 may help. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: selling_price ~ year + km_driven + seller_type + transmission + owner + (1 | fuel)
##    Data: car_data (Number of observations: 4340)
##   Draws: 3 chains, each with iter = 1000; warmup = 500; thin = 1;
##          total post-warmup draws = 1500
##
## Group-Level Effects:
## ~fuel (Number of levels: 5)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
```

```
## sd(Intercept)     0.47      0.32     0.15     1.32 1.01      348      393
##
## Population-Level Effects:
##                            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                      1.29      0.27     0.71     1.85 1.01      372
## year                           0.26      0.01     0.23     0.29 1.00     1629
## km_driven                     -0.08      0.01    -0.10    -0.05 1.00     1678
## seller_typeIndividual         -0.11      0.03    -0.17    -0.06 1.00     1591
## seller_typeTrustmarkDealer     0.29      0.08     0.14     0.43 1.01     1490
## transmissionManual            -1.50      0.04    -1.58    -1.42 1.00     1430
## ownerFourth&AboveOwner        -0.00      0.09    -0.18     0.16 1.01     1450
## ownerSecondOwner              -0.07      0.03    -0.13    -0.01 1.00     1563
## ownerTestDriveCar              0.30      0.17    -0.01     0.64 1.00     1579
## ownerThirdOwner               -0.07      0.05    -0.17     0.03 1.00     1301
##                            Tail_ESS
## Intercept                       320
## year                           1117
## km_driven                      1083
## seller_typeIndividual          1169
## seller_typeTrustmarkDealer     1095
## transmissionManual             1168
## ownerFourth&AboveOwner          994
## ownerSecondOwner               1214
## ownerTestDriveCar              1117
## ownerThirdOwner                1156
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.74      0.01     0.72     0.75 1.00     1967     1174
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
print(summary(models[[2]]))
```

```
## Warning: There were 4 divergent transitions after warmup. Increasing
## adapt_delta above 0.98 may help. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: selling_price ~ year + km_driven + seller_type + transmission + owner + (1 | fuel)
##    Data: car_data (Number of observations: 4340)
##   Draws: 3 chains, each with iter = 1000; warmup = 500; thin = 1;
##          total post-warmup draws = 1500
##
## Group-Level Effects:
```

```
## ~fuel (Number of levels: 5)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.50      0.35     0.15     1.49 1.01      372      440
##
## Population-Level Effects:
##                            Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                      1.29      0.33     0.62     1.89 1.01      302
## year                           0.26      0.01     0.23     0.28 1.00     1300
## km_driven                     -0.08      0.01    -0.10    -0.05 1.00     1303
## seller_typeIndividual         -0.11      0.03    -0.17    -0.06 1.00      909
## seller_typeTrustmarkDealer     0.29      0.08     0.14     0.44 1.00     1317
## transmissionManual            -1.50      0.04    -1.58    -1.43 1.00     1442
## ownerFourth&AboveOwner        -0.00      0.09    -0.18     0.17 1.00     1227
## ownerSecondOwner              -0.07      0.03    -0.13    -0.02 1.00     1475
## ownerTestDriveCar              0.30      0.19    -0.06     0.65 1.00     1503
## ownerThirdOwner               -0.07      0.05    -0.17     0.02 1.00     1342
##                            Tail_ESS
## Intercept                       378
## year                           1100
## km_driven                      1041
## seller_typeIndividual           924
## seller_typeTrustmarkDealer      902
## transmissionManual             1158
## ownerFourth&AboveOwner         1280
## ownerSecondOwner               1276
## ownerTestDriveCar              1077
## ownerThirdOwner                 975
##
## Family Specific Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.74      0.01     0.72     0.75 1.00     1874     1027
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Conclusion for Sensitivity

Group Level Effects

The estimated value of SD (Intercept) has slightly changed, but the 95% confidence interval has a large span, especially the upper bound. This may indicate that the model exhibits a certain sensitivity in prior selection for random intercepts of fuel types.

Population Level Effects

The estimated values of Intercept and year remain stable in two rounds of fitting, indicating that the

model is not sensitive to prior selection of these parameters. Km_ Driven, seller_ TypeIndividual, seller_
The estimated values of typeTrustmarkDealer, transmissionManual, and owner categories are also relatively stable, indicating that these fixed effects estimates of the model are robust under prior changes.

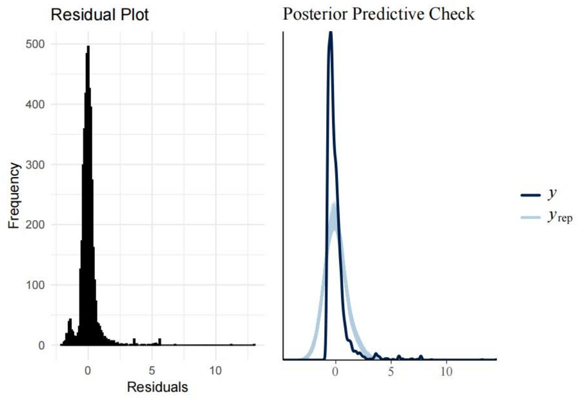
Family Specific Parameters

The estimation value of sigma is very stable under both priors, indicating that the model is not sensitive to the selection of priors in estimating the standard deviation of residuals.

Model diagnosis

The Rhat value is equal to 1 on all parameters, indicating that the model converges well. Bulk_ ESS and Tail_ ESS is high enough for most parameters, indicating that the posterior distribution has sufficient effective sample size for reliable estimation. Overall, the sensitivity analysis of the model indicates that the posterior estimate is relatively insensitive to the selection of priors, which increases confidence in the model results.

3.2.3 EDA for Hierarchical Bayesian Model



Graphic analysis

Residual plot: The residuals are mainly concentrated around 0, but there seems to be a slight right deviation. In an ideal situation, the residuals should be symmetrically distributed around 0 without any obvious skewness or outliers. This may indicate slight shortcomings in certain aspects of the model.

PPC chart: Black lines represent the density of actual data, while blue lines represent the density of data predicted by the model. The degree of overlap between these two indicates that the model predictions are consistent with the actual observed values, but also suggests the possibility of overfitting.

Reasoning and Evaluation

4. Conclusions

(1) Key Findings:

The study revealed significant insights into how various factors influence the pricing of used cars. Variables such as the car's age, mileage, fuel type, and seller type play crucial roles in determining its market value. The modeling approach demonstrated the importance of both quantitative and qualitative factors in the used car market.

(2) Implications:

The research provides valuable insights for potential buyers and sellers in the used car market, offering a deeper understanding of what factors most significantly affect car prices.

5. Reference

[1] Franke, Michael, and Timo B. Roettger. "Bayesian regression modeling (for factorial designs): A tutorial." (2019).

[2] Kemp, Charles, Andrew Perfors, and Joshua B. Tenenbaum. "Learning overhypotheses with hierarchical Bayesian models." Developmental science 10.3 (2007): 307-321.

6. Code

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(readr)
library(ggplot2)
library(dplyr)
library(tidyr)
library(scales)
library(rstan)
library(patchwork)
library(tidyverse)
library(bayesplot)
library(brms)
```
````

2. EDA

````
```{r,echo=FALSE, results='hide',message=FALSE}
# Load the dataset
car_data_orginal <- read_csv("C:/Users/xiaoy/Desktop/578/project/CAR DETAILS FROM CAR DEKHO.csv")
# Check for missing values
missing_values <- sapply(car_data_orginal, function(x) sum(is.na(x)))
# Transforming categorical variables using one-hot encoding
car_data <- car_data_orginal %>%
  mutate_at(vars(fuel, seller_type, transmission, owner), as.factor) %>%
  mutate_if(is.factor, as.numeric)
# Normalizing numerical columns
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}
car_data$year <- normalize(car_data$year)
car_data$selling_price <- normalize(car_data$selling_price)
car_data$km_driven <- normalize(car_data$km_driven)
# Create the individual plots
p1 <- ggplot(car_data, aes(x=selling_price)) +
  geom_histogram(bins=30, fill="blue", alpha=0.7) +
  labs(title="Distribution of Selling Price", x="Selling Price", y="Frequency")
p2 <- ggplot(car_data, aes(x=km_driven)) +
  geom_histogram(bins=30, fill="green", alpha=0.7) +
  labs(title="Distribution of Kilometers Driven", x="Kilometers Driven", y="Frequency")
````
````

```
p3 <- ggplot(car_data, aes(x=factor(fuel))) +
  geom_bar(fill="orange", alpha=0.7) +
  labs(title="Count of Cars by Fuel Type", x="Fuel Type", y="Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
p4 <- ggplot(car_data, aes(x=factor(transmission))) +
  geom_bar(fill="purple", alpha=0.7) +
  labs(title="Count of Cars by Transmission Type", x="Transmission Type", y="Count") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Combine the plots
combined_plot <- p1 + p2 + p3 + p4
# Display the combined plot
combined_plot
```

3. Model

3.1 Bayesian Regression Model

3.1.1 Modeling

```{r,echo=FALSE, results='hide',warning=FALSE}

# Define the model in Stan
stan_model <- "
data {
  int<lower=0> N;    // number of data items
  int<lower=0> K;    // number of predictors
  matrix[N, K] x;    // predictor matrix
  vector[N] y;          // outcome vector
}

parameters {
  vector[K] beta;             // coefficients for predictors
  real<lower=0> sigma;    // standard deviation
}

model {
  y ~ normal(x * beta, sigma);
}
"

# Exclude 'name' column and ensure all predictors are numeric
x_matrix <- as.matrix(car_data[ , !(names(car_data) %in% c("selling_price", "name"))])

# Data for Stan model
```

```
stan_data <- list(
  N = nrow(car_data),
  K = ncol(x_matrix),
  x = x_matrix,
  y = car_data$selling_price
)

# Fit the model
fit <- stan(model_code = stan_model, data = stan_data, iter = 2000, chains = 4)
print(fit,digits=4)
```
3.1.2 Sensitivity Analysis Modeling

```{r,echo=FALSE, results='hide',warning=FALSE}
new_prior <- "
data {
  int<lower=0> N;    // number of data items
  int<lower=0> K;    // number of predictors
  matrix[N, K] x;    // predictor matrix
  vector[N] y;          // outcome vector
}
parameters {
  vector[K] beta;             // coefficients for predictors
  real<lower=0> sigma;    // standard deviation
}
model {
  beta ~ normal(0, 1);
  sigma ~ cauchy(0, 5);
  y ~ normal(x * beta, sigma);
}
"
fit_new_prior <- stan(
  model_code = new_prior,
  data = stan_data,
  iter = 2000,
  chains = 4
)
print(fit_new_prior, digits = 4)
```
3.1.3 EDA for Bayesian Regression Model
```{r,echo=FALSE}
posterior_draws <- extract(fit)$beta
# Compute the predicted values (y_pred) by multiplying the predictors with the posterior draws of beta
# Note: 'x_matrix' should be the matrix of predictors you used when fitting the model
```

~~~
y_pred <- as.matrix(x_matrix) %*% t(posterior_draws)
# Calculate the residuals
residuals <- as.vector(car_data$selling_price) - rowMeans(y_pred)
# Residual Plot
  residual_plot<-ggplot(data = NULL, aes(x = residuals)) +
    geom_histogram(binwidth = 0.01, color = "black", fill = "blue") +
    xlab("Residuals") +
    ylab("Frequency") +
    ggtitle("Residual Plot")
if (ncol(y_pred) != nrow(x_matrix)) {
        y_pred <- t(y_pred)
}
# Ensure that 'car_data$selling_price' is a vector
selling_price_vector <- as.vector(car_data$selling_price)

# PPC plot
ppc_plot <- ppc_dens_overlay(y = selling_price_vector, yrep = y_pred) +
    xlab("Selling Price") +
    ylab("Density") +
    ggtitle("Posterior Predictive Check")
combined_plot <- residual_plot + ppc_plot
combined_plot
```
~~~

3.2 Hierarchical Bayesian Model

3.2.1 Modeling

~~~
```{r,echo=FALSE, results='hide',warning=FALSE}
# Read the data
car_data <- car_data_orginal
# Encode categorical variables
car_data$fuel <- as.factor(car_data$fuel)
car_data$seller_type <- as.factor(car_data$seller_type)
car_data$transmission <- as.factor(car_data$transmission)
car_data$owner <- as.factor(car_data$owner)
# Normalize numerical variables
car_data$year <- scale(car_data$year)
car_data$selling_price <- scale(car_data$selling_price)
car_data$km_driven <- scale(car_data$km_driven)
# Define the model with adjusted settings
model <- brm(
    selling_price ~ year + km_driven + seller_type + transmission + owner + (1|fuel),
    data = car_data,
~~~

```
  family = gaussian(),
  prior = c(
    set_prior("normal(0,5)", class = "b"),
    set_prior("normal(0,5)", class = "Intercept"),
    set_prior("cauchy(0,2)", class = "sd")
  ),
  chains = 3,
  iter = 1000,   # Increased total iterations
  warmup = 500,   # Increased warmup iterations
  control = list(adapt_delta = 0.98)   # Increased adapt_delta for better convergence
)
# Print the model summary
summary(model)
```
3.2.2 Sensitivity Analysis Modeling

```{r,echo=FALSE, results='hide',warning=FALSE,message=FALSE}
priors_list <- list(
  set_prior("normal(0, 2)", class = "b"),
  set_prior("normal(0, 10)", class = "b"))
models <- list()
for (i in seq_along(priors_list)) {
  models[[i]] <- brm(
    formula = selling_price ~ year + km_driven + seller_type + transmission + owner + (1|fuel),
    data = car_data,
    family = gaussian(),
    prior = priors_list[[i]],
    chains = 3,
    iter = 1000,
    warmup = 500,
    control = list(adapt_delta = 0.98),
    seed = 123
  )
  print(summary(models[[i]]))
}
```
3.2.3 EDA for Hierarchical Bayesian Model

```{r,echo=FALSE}

residuals_data <- residuals(model)

residual_plot <-ggplot() +
```

```
  geom_histogram(aes(x = residuals_data), binwidth = 0.1, fill = "blue", color = "black") +
  theme_minimal() +
  labs(x = "Residuals", y = "Frequency", title = "Residual Plot")
posterior_predictive <- posterior_predict(model)
selling_price_vector <- as.vector(car_data$selling_price)
ppc_plot<-ppc_dens_overlay(y = selling_price_vector, yrep = posterior_predictive) +
  labs(title = "Posterior Predictive Check")
combined_plot <- residual_plot + ppc_plot
combined_plot
```