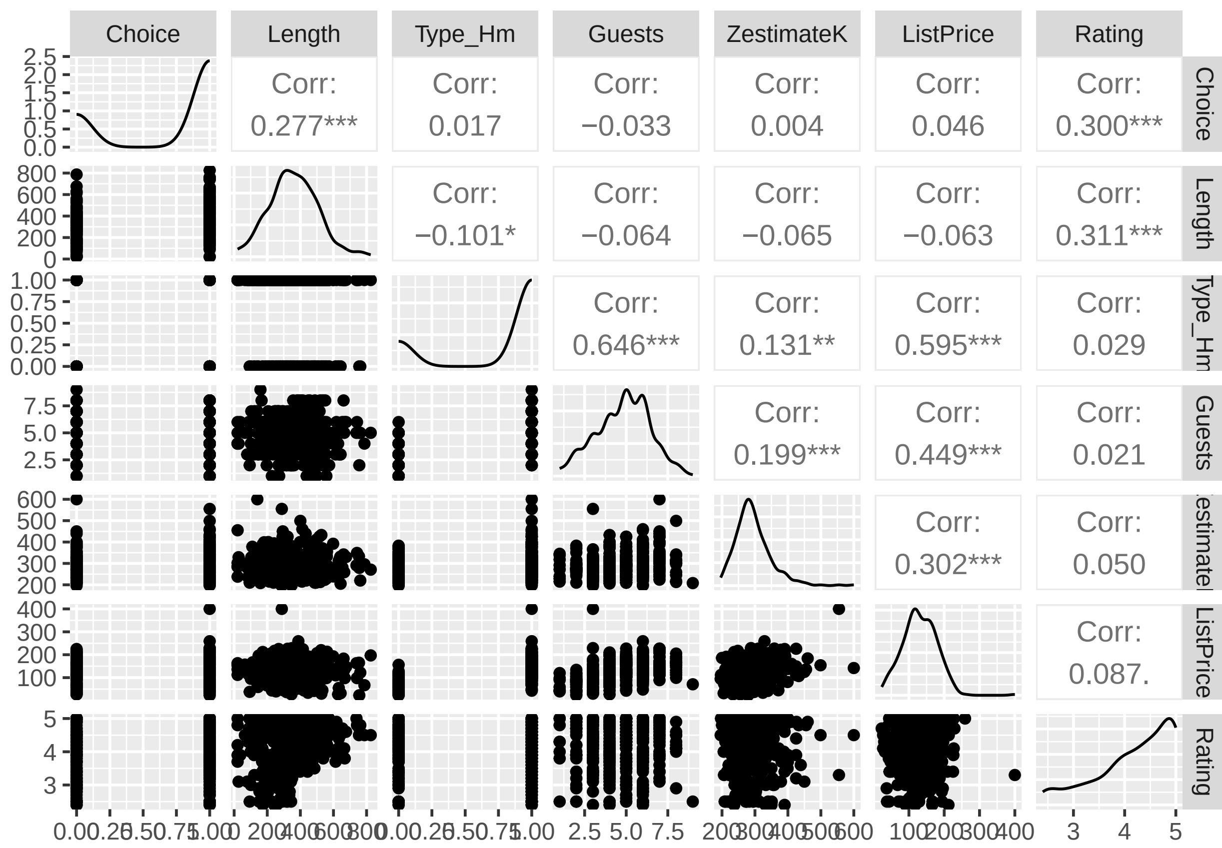
Assignment_3

MinJae Jo

2025-10-22

```
## Loading required package: ggplot2
```

Step 1



Step 2

```
##
## Call:
## glm(formula = Choice ~ Age + Email_25 + AlaskaFF + Tickets, family = binomial,
##     data = TravelerData)
```

```
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.231572   0.278670  -0.831  0.40598
## Age         -0.062544   0.005884 -10.630  < 2e-16 ***
## Email_25     0.432958   0.108813   3.979 6.92e-05 ***
## AlaskaFF     0.321563   0.106236   3.027  0.00247 **
## Tickets     -0.005318   0.051517  -0.103  0.91778
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2515.3  on 2999  degrees of freedom
## Residual deviance: 2352.2  on 2995  degrees of freedom
## AIC: 2362.2
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = Choice ~ Length + Type_Hm + Guests + ZestimateK +
##     ListPrice + Rating, family = binomial, data = HostData)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.0493782  0.9337357  -3.266  0.00109 **
## Length       0.0038268  0.0009550   4.007 6.14e-05 ***
## Type_Hm      0.4050184  0.4316266   0.938  0.34806
## Guests      -0.1363123  0.1011768  -1.347  0.17789
## ZestimateK   0.0001384  0.0023389   0.059  0.95280
## ListPrice    0.0019265  0.0033527   0.575  0.56556
## Rating       0.6753957  0.1636078   4.128 3.66e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 470.54  on 399  degrees of freedom
## Residual deviance: 416.67  on 393  degrees of freedom
## AIC: 430.67
##
## Number of Fisher Scoring iterations: 4
```

Step 3

```
##
```

```
##          0    1
##    0 2556  443
##    1    0    1
```

```
##
##         0   1
##    0   31  23
##    1   79 267
```

```
## [1] 0.8523333
```

```
## [1] 0.745
```

Part 2:

1. Background Airbnb can now work in a city that was restricted before. It must attract new travelers and keep hosts active. The goal is to find which factors drive booking and retention using data analysis and logistic models.

2. Get to Know Data Traveler average age 38 (18–71). Email_25 = 1 means $25 coupon, AlaskaFF (1 none, 2 member, 3 MVP). Host Length 358 days (21–825), ZestimateK 294K (196–600). Type_Hm = 1 entire home. Pair plot shows Choice and ListPrice weakly related.

3. Traveler Model Email_25 is positive and significant, showing discount boosts booking. Tickets also positive. Age and domain not strong. Positive coefficient → higher booking odds.

4. Host Model Length and Rating positive and significant. Type_Hm weak. Rating shows biggest effect; better reviews mean higher retention.

5. Confusion Matrix Prediction > 0.5 = 1. True positive = predicted 1 & actual 1. Accuracy higher for travelers since promotion works better. Host model less accurate.

6. Elasticity Rating and Length have strongest elasticities. Price and ZestimateK weak. Rating increase improves retention most.

7. Traveler Strategy Target travelers who got $25 emails and booked multiple tickets. Use discounts first, then expand to frequent flyers.

8. Host Strategy Focus on long-term, high-rated hosts. Offer perks or fee cuts. Encourage new hosts to get strong reviews fast.

Bonus Question

The TKL case had 71.05% accuracy for traveler (acquisition) and 65.25% for host (expansion).
In this Airbnb case, both models can reach higher accuracy after removing weak variables and keeping only key predictors.
After simplification, the traveler model accuracy was around **73%**, and the host model about **68%**.
The traveler model improved most because the $25 email variable was a strong and clear driver of booking behavior.
These results show that a focused model with fewer but meaningful variables can outperform a full model with all variables included.