# Big Data & AI code file

2024-04-12

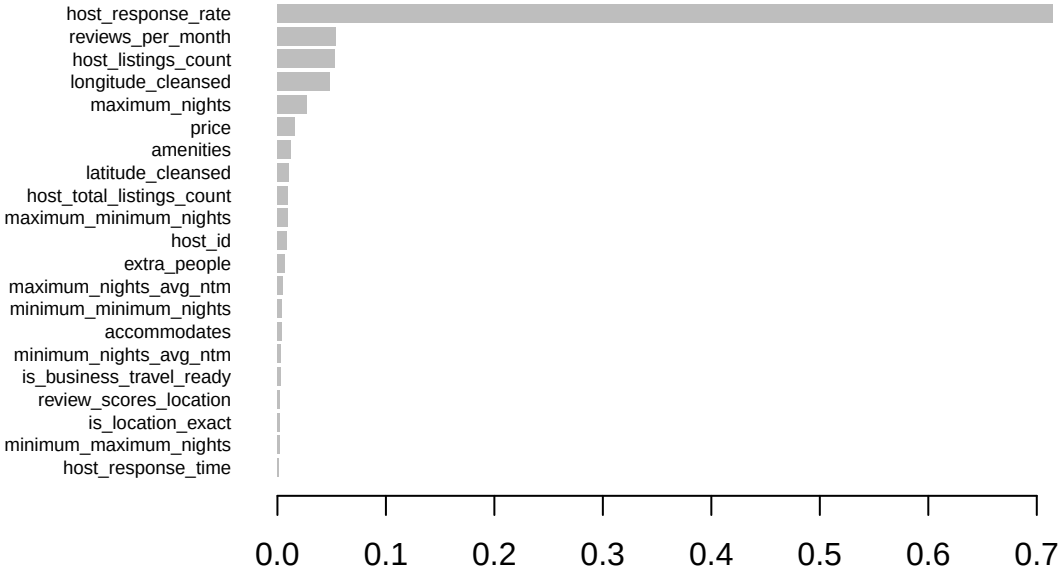
### Overall Market Model

```
## [1]  train-logloss:0.436885  eval-logloss:0.435349
## [2]  train-logloss:0.414456  eval-logloss:0.412797
## [3]  train-logloss:0.407865  eval-logloss:0.406174
## [4]  train-logloss:0.404236  eval-logloss:0.402336
## [5]  train-logloss:0.400291  eval-logloss:0.398299
## [6]  train-logloss:0.397751  eval-logloss:0.396455
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
## Area under the curve: 0.8401
```



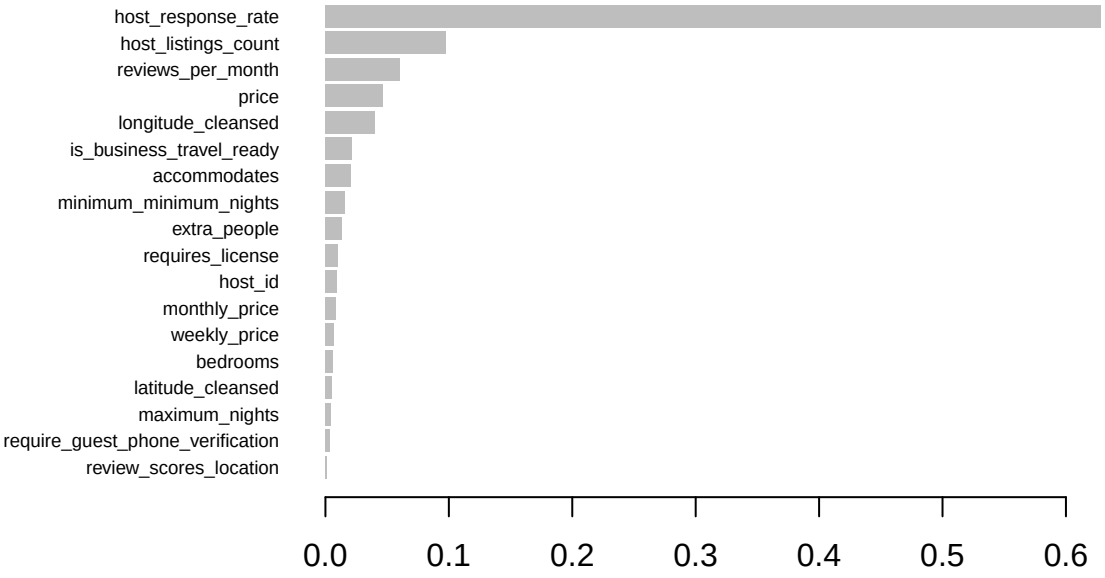

### LA Market Model

```
## [1]  train-logloss:0.398837  eval-logloss:0.400823
## [2]  train-logloss:0.372140  eval-logloss:0.376339
## [3]  train-logloss:0.361130  eval-logloss:0.364158
## [4]  train-logloss:0.352482  eval-logloss:0.355087
## [5]  train-logloss:0.344412  eval-logloss:0.348088
## [6]  train-logloss:0.339473  eval-logloss:0.342953
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

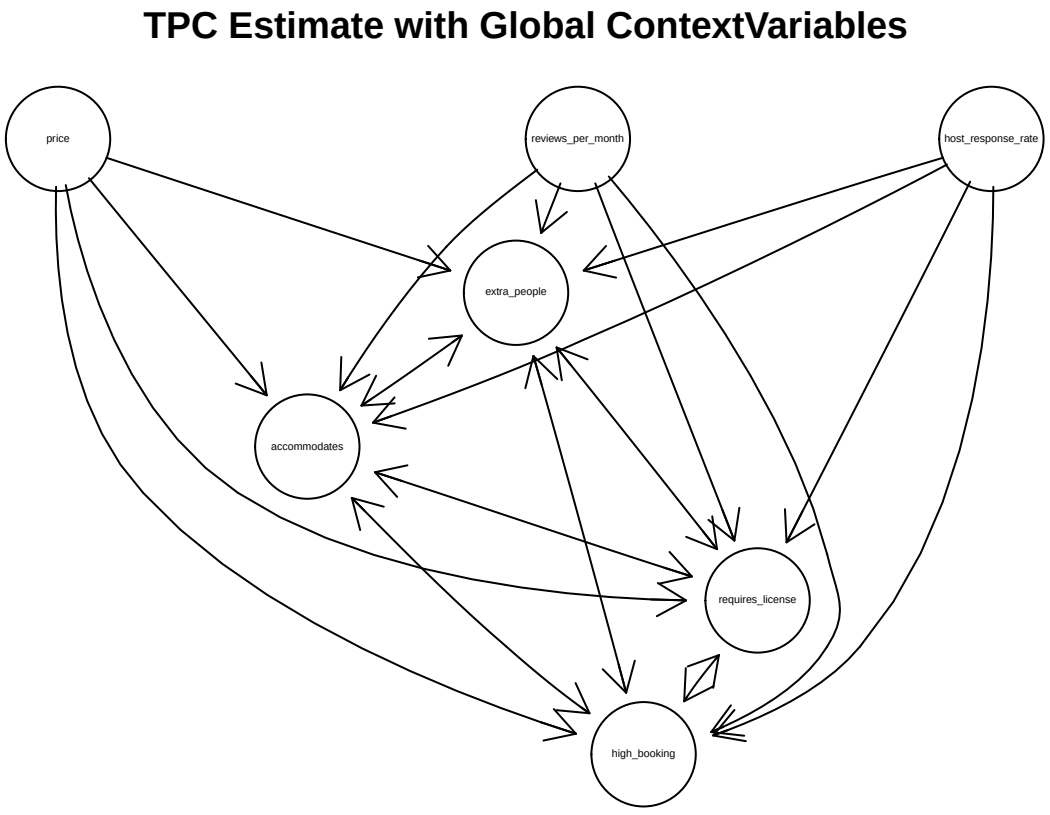
```
## Area under the curve: 0.849
```




###

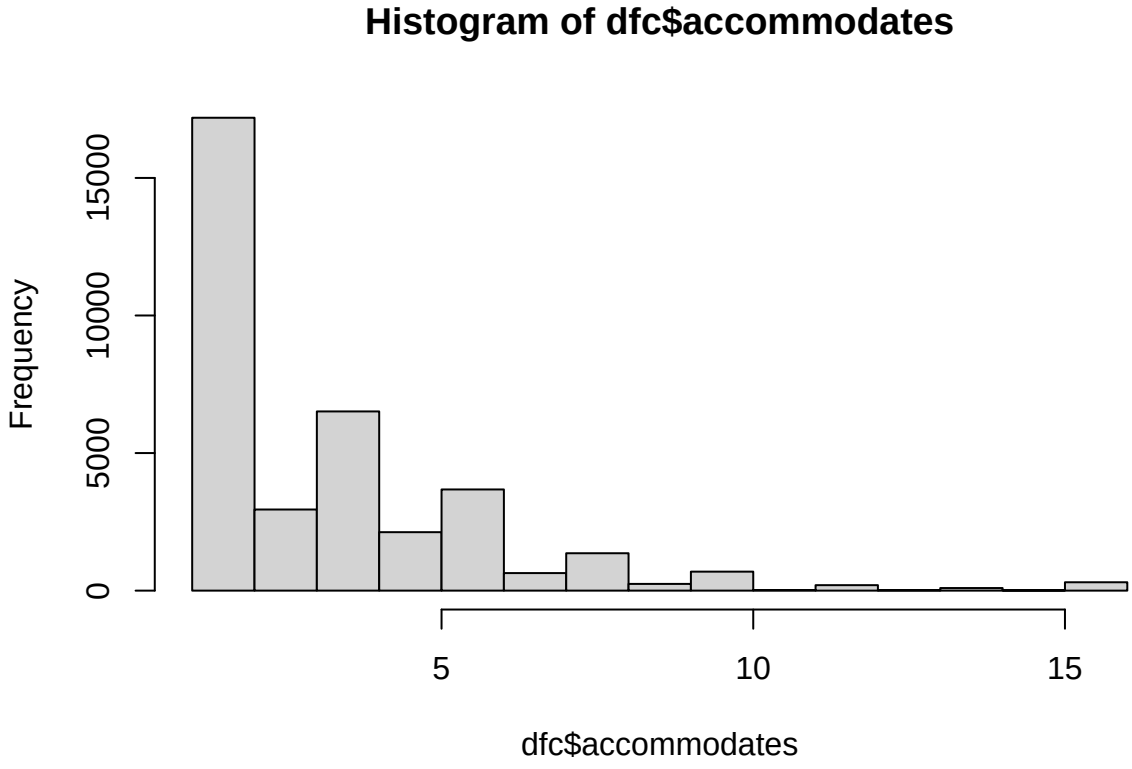
Causal Inference

```
## Loading required namespace: Rgraphviz
```




```
## corrplot 0.92 loaded
```

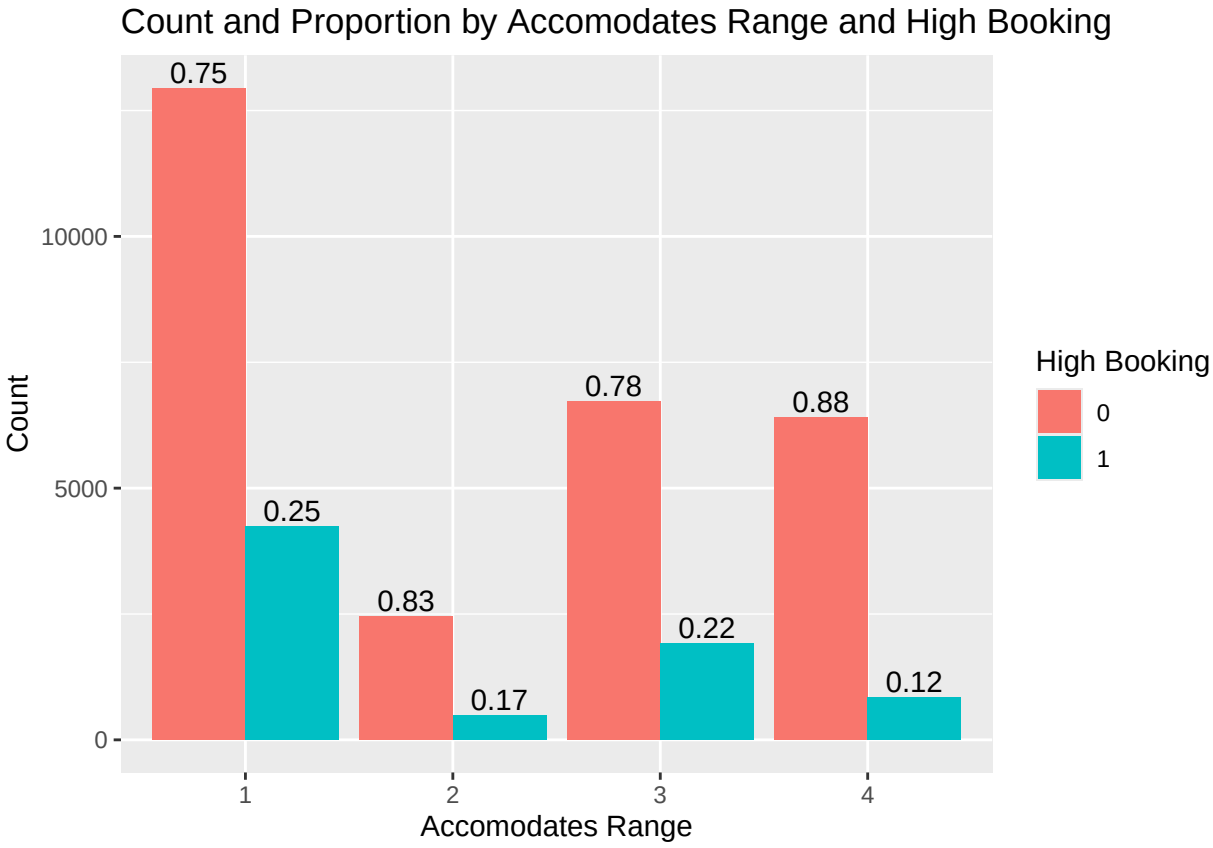

**Accommodates Findings**



```
## NOTE: 330 observations removed because of NA values (Fixed-effects: 330).
```

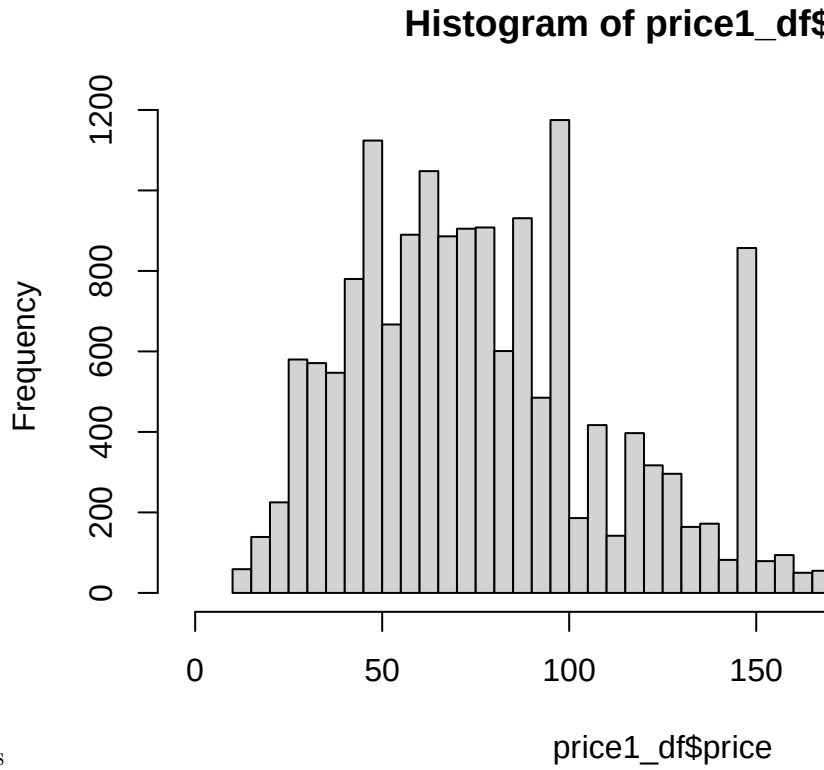
```
## OLS estimation, Dep. Var.: high_booking
## Observations: 35,697
## Fixed-effects: zipcode: 238
## Standard-errors: Clustered (zipcode)
##              Estimate Std. Error  t value   Pr(>|t|)
## accommodates -0.01634    0.00283 -5.77326 2.4247e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.386413     Adj. R2: 0.088975
##                  Within R2: 0.009451
```

Significant and negative. Recommend investing in airbnbs that accomodates less people.




```
##   0%  25%  50%  75% 100%
##    1    2    3    5   16
```

From the chart and these quantiles, we can see that we want to invest in Q1 (accommodates for 1-2 people) and




Q2 (accommodates for 4-5 people) ### Price Findings

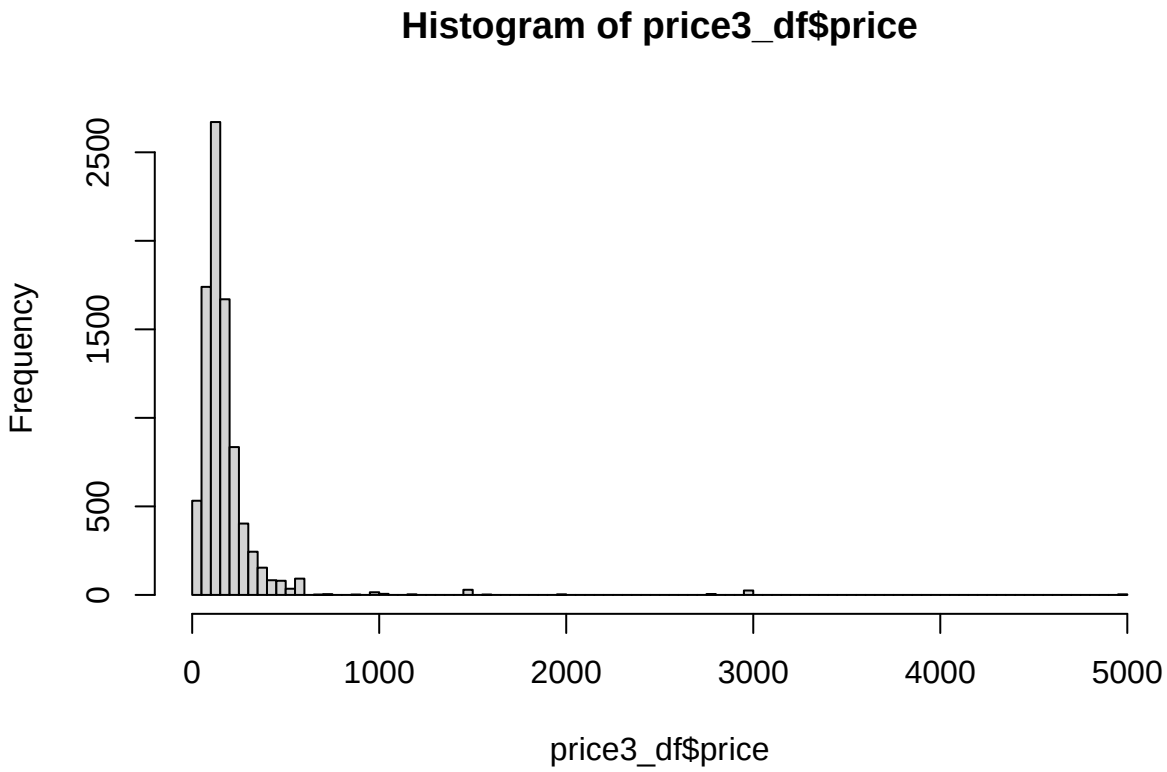
```
## NOTE: 179 observations removed because of NA values (Fixed-effects: 179).
```

```
## OLS estimation, Dep. Var.: high_booking
## Observations: 17,009
## Fixed-effects: zipcode: 213
## Standard-errors: Clustered (zipcode)
##       Estimate Std. Error  t value  Pr(>|t|)
## price   -7e-05    2.6e-05 -2.68593 0.0078057 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.395774     Adj. R2: 0.146703
##                  Within R2: 6.031e-4
```

Signifiant and negative. When investing in ABNB that accommodates for 1-2 people, try not to charge too much.

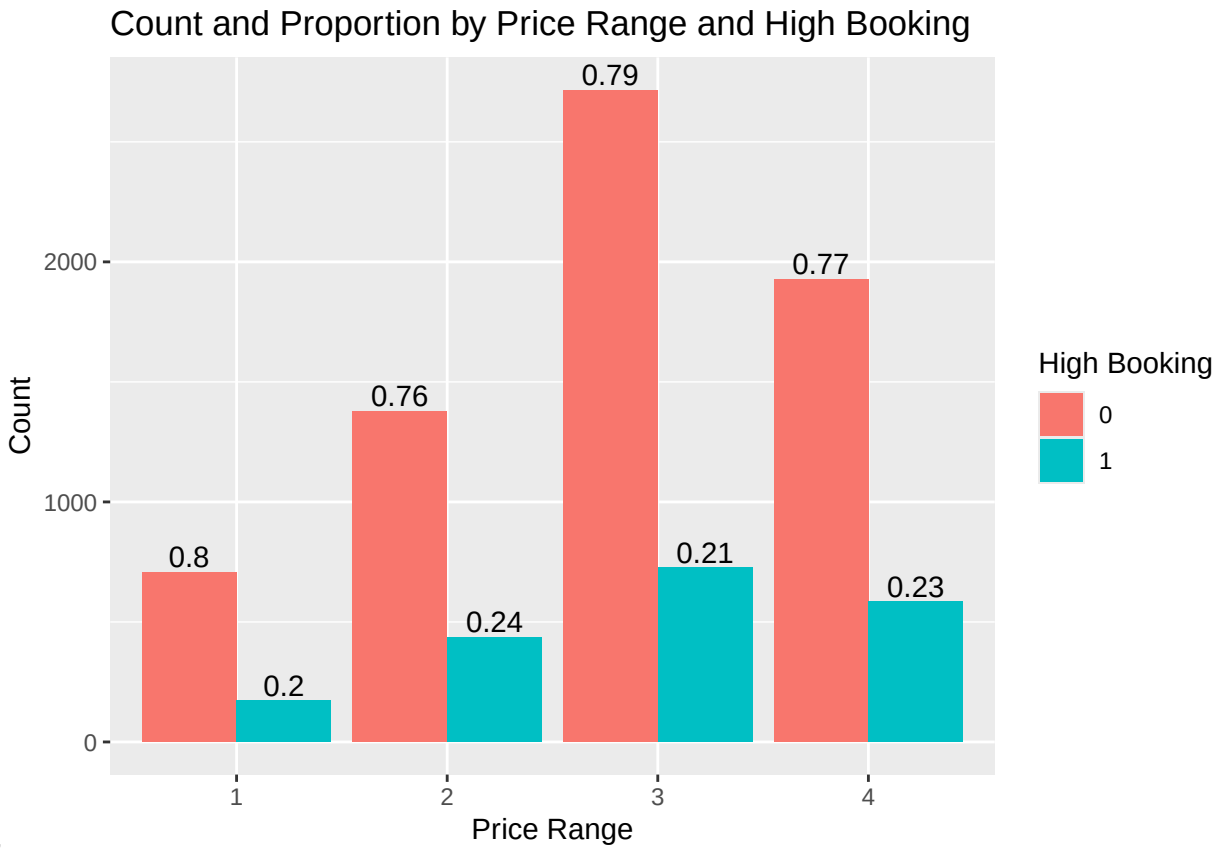

```
##    0%   25%   50%   75%  100%
##    10    55    78   110  5000
```

From the graph and the quantiles make sure not to charge over 110 for ABNBs that accommodates 1-2 people




```
## NOTE: 32 observations removed because of NA values (Fixed-effects: 32).

## OLS estimation, Dep. Var.: high_booking
## Observations: 8,606
## Fixed-effects: zipcode: 166
## Standard-errors: Clustered (zipcode)
##       Estimate Std. Error  t value Pr(>|t|)
## price 9.86e-06    6.4e-05 0.154837  0.87714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.382279     Adj. R2: 0.137717
##                  Within R2: 3.105e-5
```




Not significant

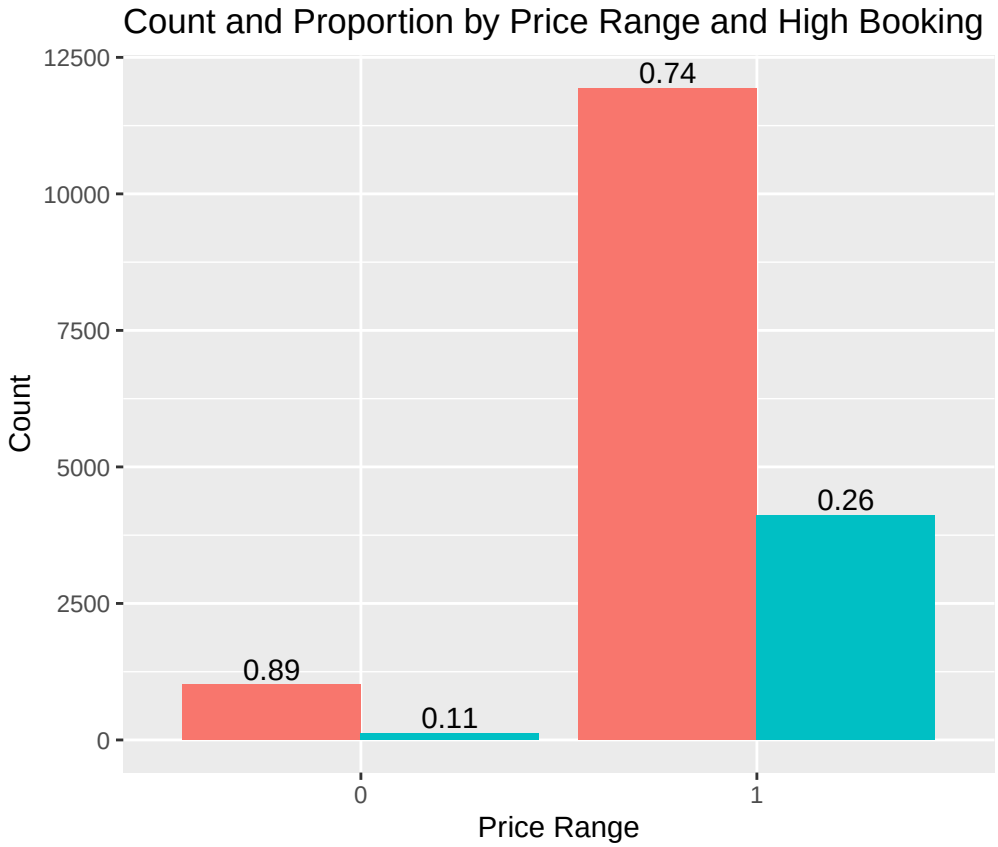
```
##    0%  25%  50%  75% 100%
##     0  100  140  200 5000
```

No specific recommendation for this level of accommodates. Price proportion is similar across quantiles however Q1 has the lowest prop.

### Host Response Rate Findings

```
## NOTE: 179 observations removed because of NA values (Fixed-effects: 179).
```

```
## OLS estimation, Dep. Var.: high_booking
## Observations: 17,009
## Fixed-effects: zipcode: 213
## Standard-errors: Clustered (zipcode)
##                    Estimate Std. Error t value Pr(>|t|)
## host_response_rate   0.0664   0.039148 1.69612  0.09133 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.395771     Adj. R2: 0.146718
##                  Within R2: 6.207e-4
```

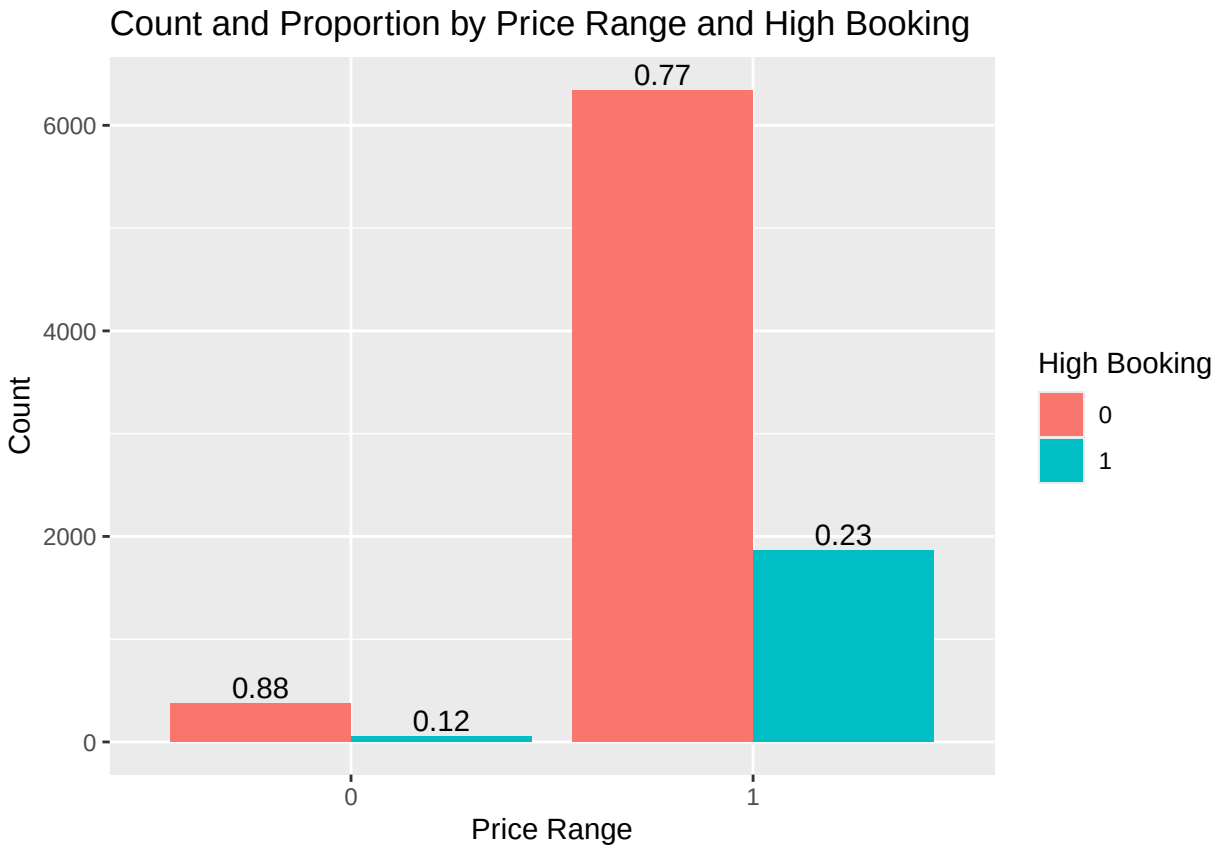



Positive, however not very significant.
Above 75% response rate improves high booking for ABNBs that accommodates for 1-2 people.

```
##        0%       25%       50%       75%      100%
## 0.0000000 0.9463407 1.0000000 1.0000000 1.0000000
```

```
## NOTE: 32 observations removed because of NA values (Fixed-effects: 32).
```

```
## OLS estimation, Dep. Var.: high_booking
## Observations: 8,606
## Fixed-effects: zipcode: 166
## Standard-errors: Clustered (zipcode)
##                      Estimate Std. Error   t value Pr(>|t|)
## host_response_rate -0.026538   0.076752 -0.345767  0.72996
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 0.382271     Adj. R2: 0.137753
##                  Within R2: 7.241e-5
```




Not sig.

Above 75% response rate improves high booking for ABNBs that accommodates 4-5 people.