Information and Airbnb, An Analysis on the Impact of Reviews and Ratings on Airbnb Prices in Canada*

Alaina Hu and Ahad Qureshi

February 14, 2024

Table of contents

1 Introduction

In the changing landscape of modern travel, Airbnb has emerged as a transformative force, reshaping traditional notions of accommodation and hospitality. No longer are travelers confined to hotels; people have the choice to choose between various forms of accommodation and find the option most suitable for their needs. With Airbnb's vacation rentals, travelers have the option to gain access to more space, kitchens, home amenities, and lower cost (Guttentag, 2016). Central to Airbnb's allure are the wealth of user-generated reviews and ratings, which serve as vital sources of information for prospective guests navigating a vast array of listings. These reviews not only offer insights into the quality and character of accommodations but also play a pivotal role in shaping consumer decisions. However, in this growing world of shared experiences, a fundamental question persists: What effect do these reviews and ratings have on Airbnb pricing strategies? This question lies at the heart of our inquiry, as we delve into the relationship between information, consumer behavior, and pricing dynamics within the Airbnb marketplace in Canada. Through statistical analysis and data visualization, this

*Code and data are available at: https://github.com/alainahu/airbnb_analysis paper endeavors to continue to explore the relationship between reviews, ratings, and Airbnb prices.

In an econometric analysis on Airbnb reviews and price in Boston, Lawani et al. find that reviews serve as a good proxy for rooms' quality and reviews affect both the host price and neighboring host price. Peng et al. (2020) build a machine learning model and use natural language processing to predict Airbnb prices in nine cities across the United States, Australia, and Europe. Their research concludes that customer reviews, house features, and geographical data are all predictive factors for Airbnb rental prices.

2 Data

```
Rows: 87059 Columns: 9
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): city
dbl (8): review, guest_satisfaction_overall, accuracy_rating, cleanliness_ra...

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.
```

```
Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.
```

```
Warning: Removed 102 rows containing non-finite values (`stat_bin()`).
```

```
Warning: Removed 2 rows containing missing values (`geom_bar()`).
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  26.00   68.67   93.83  107.66  126.33  634.33 
```

```
summary_df <- airbnb |>
  summarise(
    Mean_New_Price = mean(new_price, na.rm = TRUE),
    Median_New_Price = median(new_price, na.rm = TRUE),
    SD_New_Price = sd(new_price, na.rm = TRUE))
```

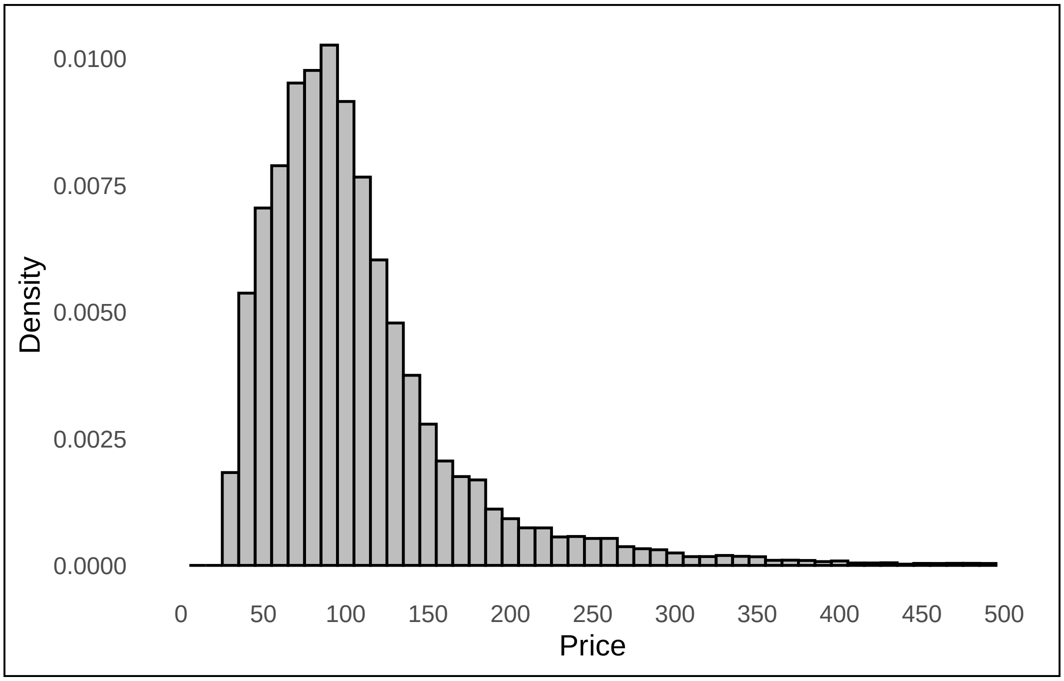


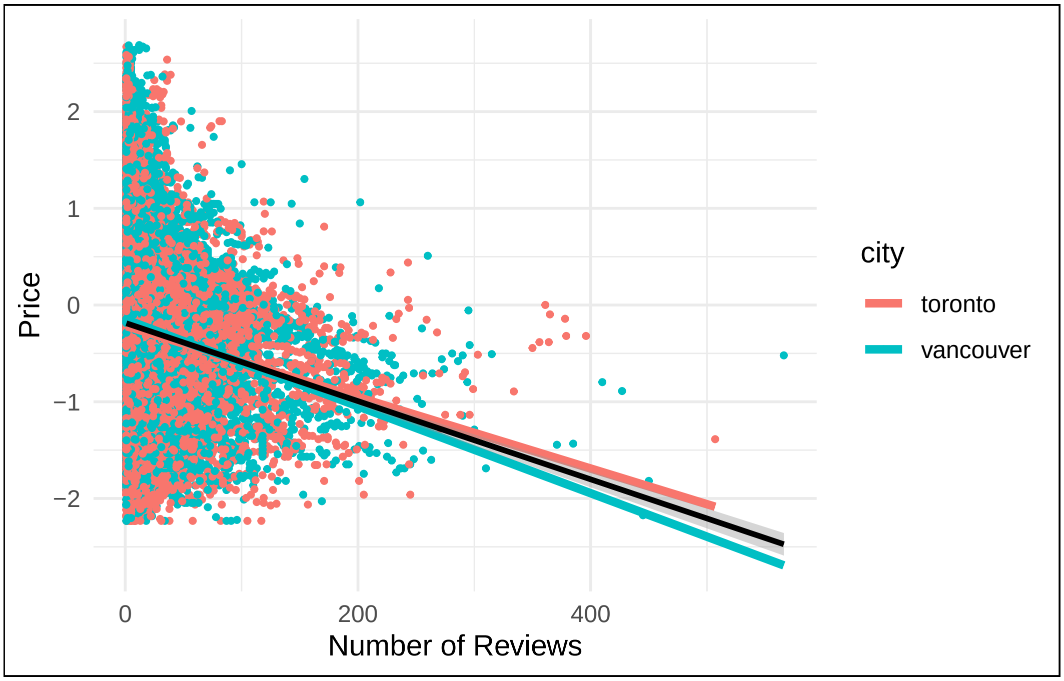
Figure 1: Distribution of Airbnb prices per night (including cleaning fees)

Table 1: Summary Statistics for the Reviews and Ratings

Variable	Mean	Median	Standard Deviation	Min	Max
Accuracy Rating	9.47	10	0.98	0	10
Cleanliness Rating	9.28	10	1.14	0	10
Guest Overall Satisfaction	93.47	95	7.78	20	100
Location Rating	9.48	10	0.95	0	10
Number of Reviews	14.90	7	23.86	1	566
Value Rating	9.29	9	1.00	0	10

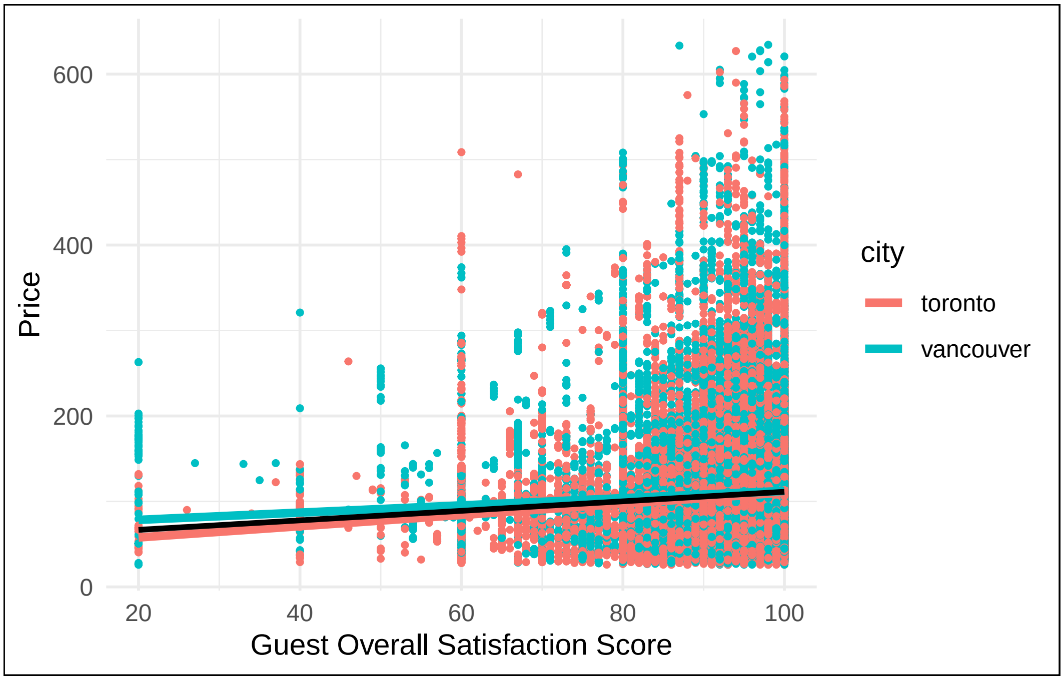
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

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




```
<ggproto object: Class ScaleDiscrete, Scale, gg>
    aesthetics: colour
    axis_order: function
    break_info: function
    break_positions: function
    breaks: waiver
    call: call
    clone: function
    dimension: function
    drop: TRUE
    expand: waiver
    get_breaks: function
    get_breaks_minor: function
    get_labels: function
    get_limits: function
    guide: legend
    is_discrete: function
    is_empty: function
    labels: waiver
    limits: NULL
    make_sec_title: function
    make_title: function
    map: function
```

```
    map_df: function
    n.breaks.cache: NULL
    na.translate: TRUE
    na.value: NA
    name: waiver
    palette: function
    palette.cache: NULL
    position: left
    range: environment
    rescale: function
    reset: function
    scale_name: brewer
    train: function
    train_df: function
    transform: function
    transform_df: function
    super:  <ggproto object: Class ScaleDiscrete, Scale, gg>

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

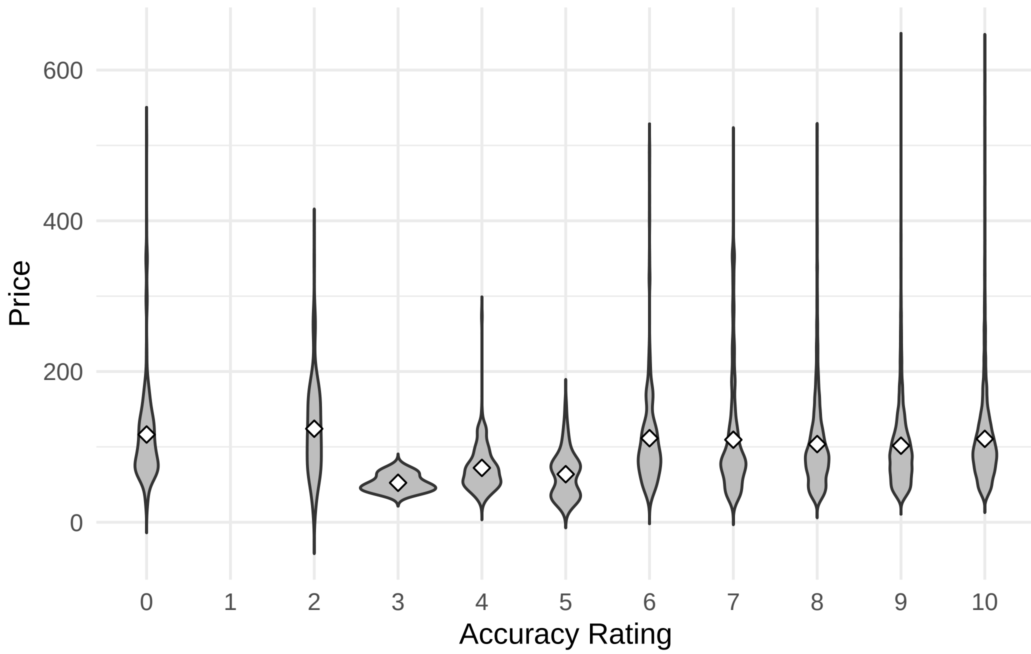


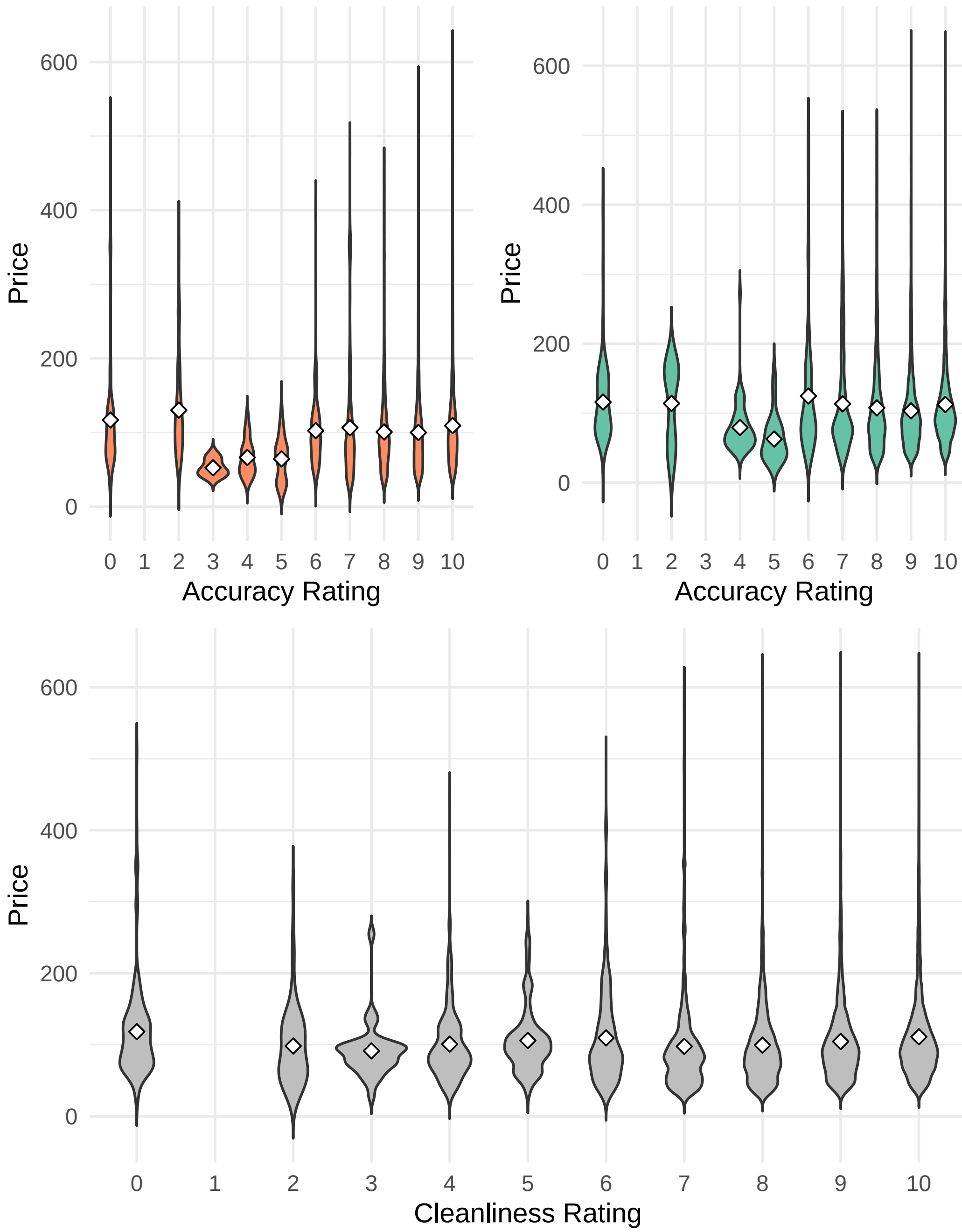

```
<ggproto object: Class ScaleDiscrete, Scale, gg>
```

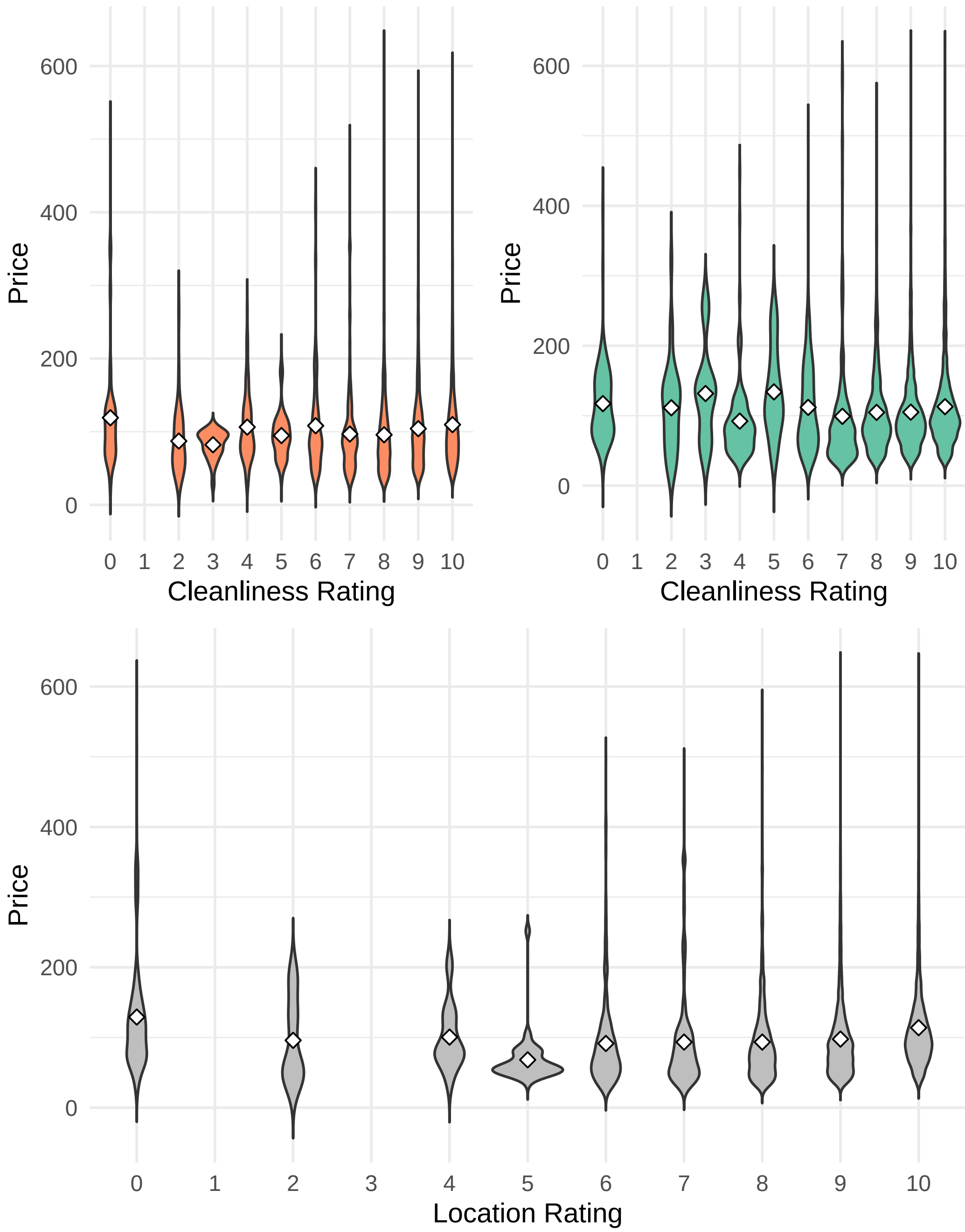
```
aesthetics: colour
axis_order: function
break_info: function
break_positions: function
breaks: waiver
call: call
clone: function
dimension: function
drop: TRUE
expand: waiver
get_breaks: function
get_breaks_minor: function
get_labels: function
get_limits: function
guide: legend
is_discrete: function
is_empty: function
labels: waiver
limits: function
make_sec_title: function
make_title: function
map: function
map_df: function
n.breaks.cache: NULL
na.translate: TRUE
na.value: grey50
name: waiver
palette: function
palette.cache: NULL
position: left
range: environment
rescale: function
reset: function
scale_name: manual
train: function
train_df: function
transform: function
transform_df: function
super:  <ggproto object: Class ScaleDiscrete, Scale, gg>
```

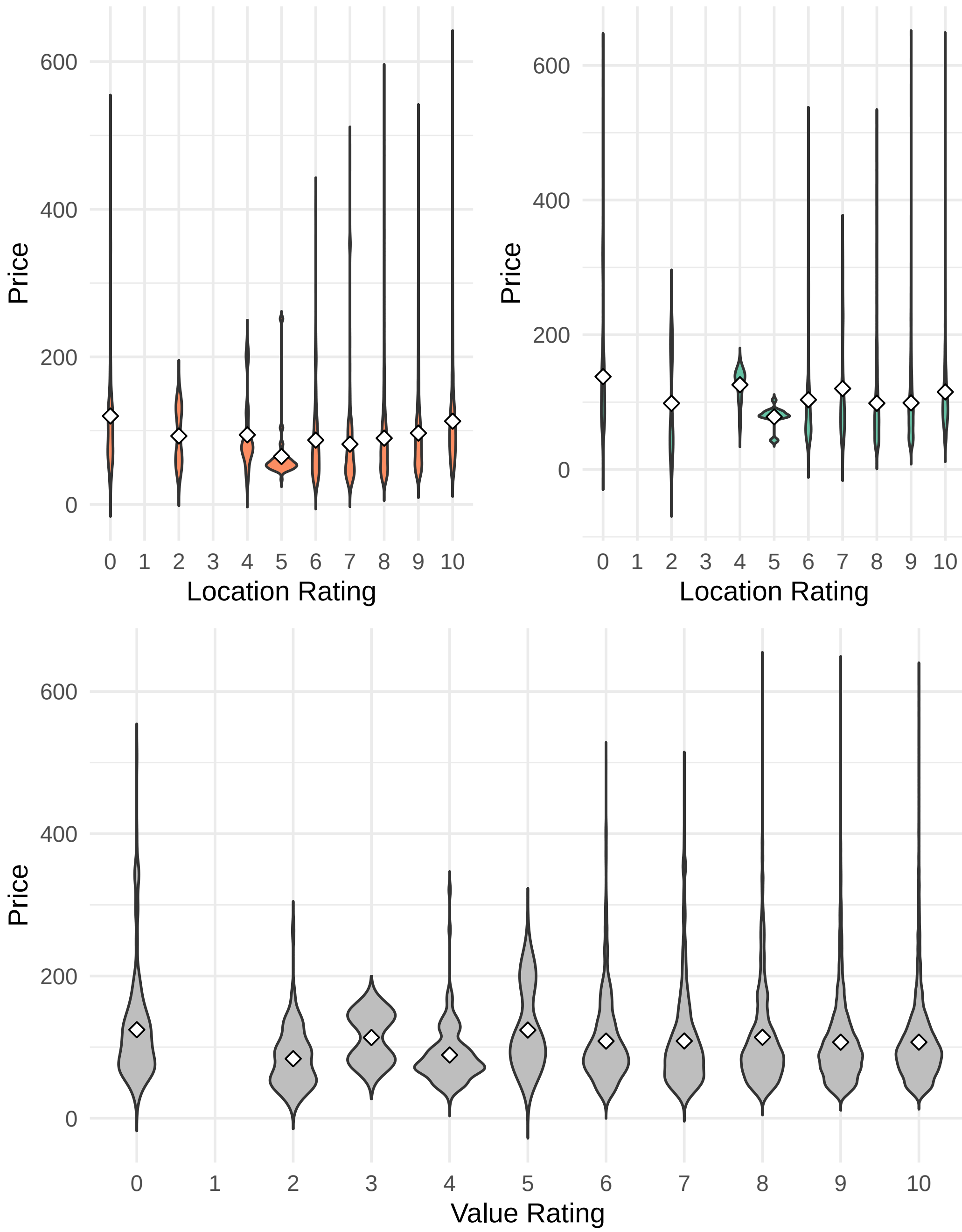
```r
unique_values <- unique(airbnb$accuracy_rating)

# Print the unique values
print(unique_values)
```

```
[1] 10  9  8  7  0  6  2  5  4  3
```

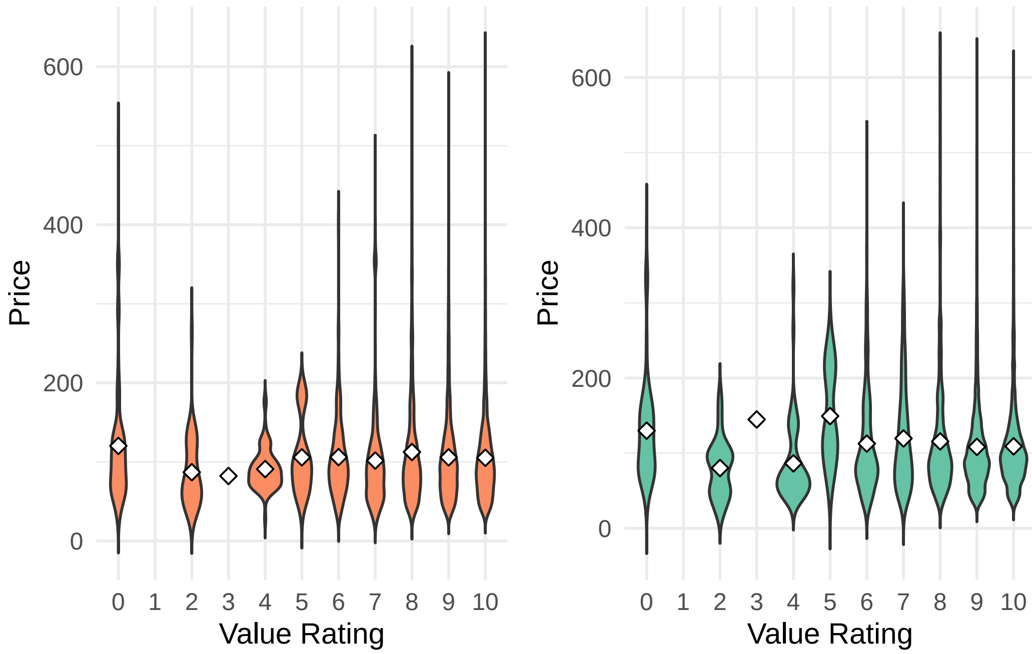



Price
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10
Accuracy Rating
Price
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10
Accuracy Rating
Price
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10
Cleanliness Rating


Price
Cleanliness Rating
Price
Cleanliness Rating
Price
Location Rating




```
Warning: Groups with fewer than two data points have been dropped.
Groups with fewer than two data points have been dropped.
```


Price
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10
Value Rating
Price
600
400
200
0
0 1 2 3 4 5 6 7 8 9 10
Value Rating