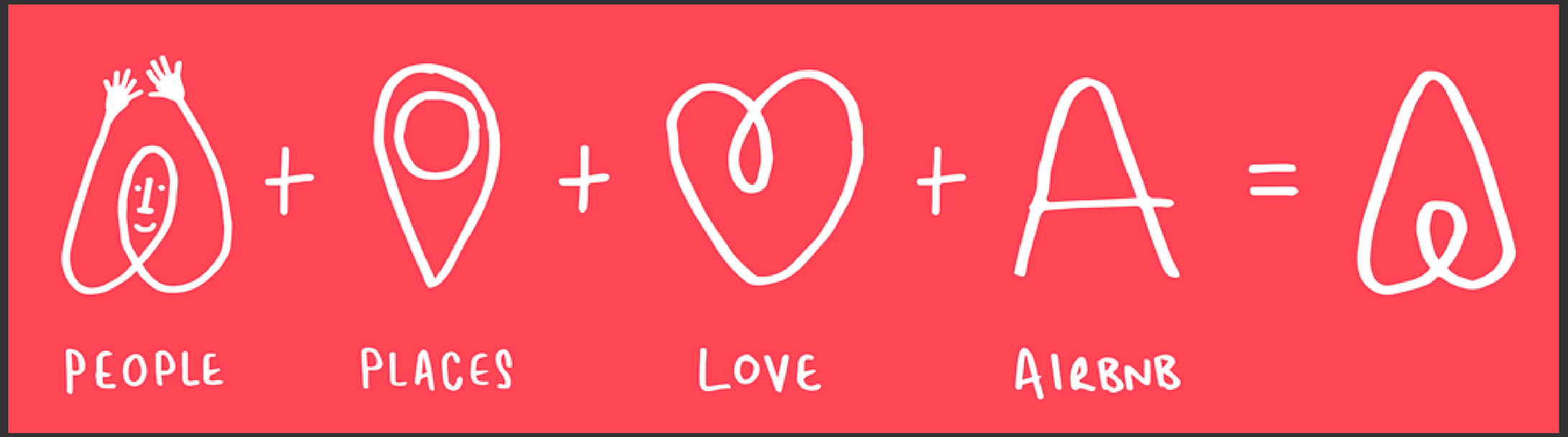
# Airbnb Pricing and Property Analysis




Aditi Poddar, Kristina Cinova, Sylvia Lam

# Project Questions

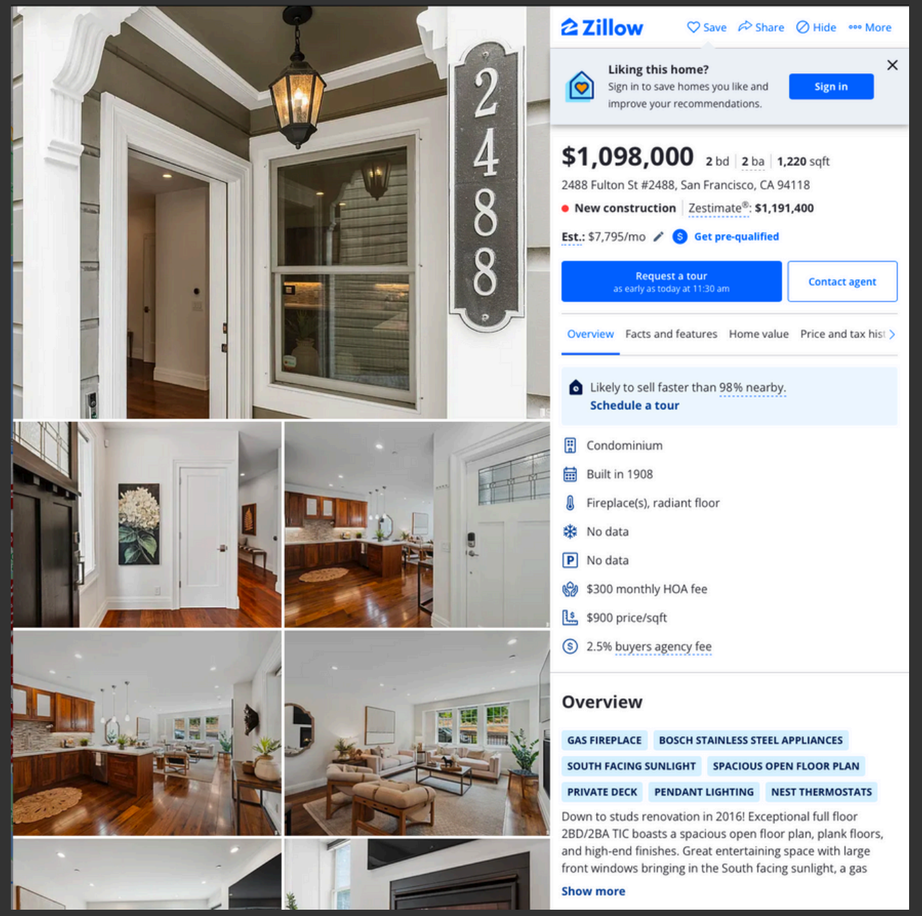
1. Imagine I purchased the property shown in the screenshots below to rent via Airbnb. How much should I charge per night to rent this entire property? Is there anything I should try to highlight in the listing

2. What percent of properties listed on Airbnb in San Francisco would you classify as “full-time rental properties”? In other words, properties in which the owners never reside.

3. Where are the cheapest and most expensive neighbourhoods to rent an Airbnb in the US and Canada?

4. Does crime rate of the area affect price of the Airbnb listings?

5. Should we list this property on Airbnb or Zillow?



# Q1: How much should I charge per night to rent this entire property?

## Hypothesis

Model 1: Price ~ bedrooms + bathrooms + location

Model 2: Price ~ bedrooms + bathrooms + location + amenities

Null Hypothesis:

There won’t be any difference between Model 1 & Model 2.

Alternate Hypothesis:

There will be a difference between Model 1 & Model 2 as amenities play a role in determining the price.

# Q1: How much should I charge per night to rent this entire property?

## Method I

1.Identify what variables determine the price of the property

*#bedrooms, bathrooms, location*

2. Established regression equation

*# price ~ bedrooms + bathrooms + location*

3. Read the data in R

4. Identified bathrooms are given as ‘2 baths’ i.e. numeric and text format

*# Used as.numeric function and extract to pull out the number*

*5.* Identified the location of the property i.e. Inner Sunset / Outer Sunset

*# Converted the location to a dummy variable using ifelse function keeping Inner Sunset = 1 and then Outer sunset = 1*

*Note: Build two different regression models with both areas*

6. Changed price from character to number

*# Used stringr and as.numeric*

*7.* Built a regression model

8. Predicted prices for Inner Sunset and Outer Sunset, concluded the average as the suggested price

Final Price ~ $ 424.8708 per night

# Q1: How much should I charge per night to rent this entire property?

## Method II

1.Identify what variables determine the price of the property

*#bedrooms, bathrooms, location, amenities*

2. Established regression equation

*# price ~ bedrooms + bathrooms + location + amenities*

3. Read the data in R

4. Identified bathrooms are given as ‘2 baths’ i.e. numeric and text format

*# Used as.numeric function and extract to pull out the number*

*5.* Identified the location of the property i.e. Inner Sunset / Outer Sunset

*# Converted the location to a dummy variable using ifelse function keeping Inner Sunset = 1 and then Outer sunset = 1*

*Note: Build two different regression models with both areas*

*6.* Extract all amenities and change it to dummy variable

*#Used strsplit, unlist and loop*

*7.* Changed price from character to number

*# Used stringr and as.numeric*

# Q1: How much should I charge per night to rent this entire property?

## Method II

8. Built a regression model
9. Predicted prices for Inner Sunset and Outer Sunset, concluded the average as the suggested price

Final Price ~ $ 372.36 per night

# Q2. Percentage of Full-Time Rental Properties on Airbnb in San Francisco

Method I

Filter with our criteria (must fufill all 3):

```
# must be listed as an entire home
# must pass a high availability threshold (more than 300 days)
# host must have multiple listings (more than 1)
```

Then...

```
percentage_full_time_rentals <- (full_time_rental_properties_count / total_listings) * 100
```

Results:

```
> cat(percentage_full_time_rentals, "%")
9.423025 %
```

# Q2. Percentage of Full-Time Rental Properties on Airbnb in San Francisco

Method II

Filter with our criteria (must fufill both):

```
# must have 5 or more reviews in the last 12 months
# must be a minimum stay of 30 days or more
```

Then...

```
percentage_full_time_rentals <- (full_time_rental_properties_count / total_listings) * 100
```

# Q2. Percentage of Full-Time Rental Properties on Airbnb in San Francisco

Results:

```
> cat(percentage_full_time_rentals, "%")
3.181451 %
```

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

***Goal: List out the cheapest and most expensive neighborhoods***

***Approach:***

1. Filter listings in the US and Canada
2. Group by Country, City, Neighborhood
3. Find the mean for price
4. Find the top 10 cheapest neighborhood
5. Find the top 10 most expensive neighborhood

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

***Code:***

```
airbnb_data <-read_csv('~/Desktop/Textbook/AppliedStat/DATA/Airbnb/Airbnb_Listings.csv')

filtered_data <- airbnb_data %>%
  filter(country %in% c("United States of America", "Canada"))%>%
  group_by(country, city, neighborhood) %>%
  summarize(mean.price = mean(price, na.rm = TRUE))

filtered_data %>%
  arrange(mean.price) %>%
  head(10)

#Replace head(10) with tail(10)
```

Alternative method for finding the most expensive neighborhoods :

```
filtered_data %>%
arrange(desc(mean.price)) %>%
 head(10)
```

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

***Results:***

Filtered_data (part of) for mean price

| | country | city | neighborhood | mean.price |
|---|---|---|---|---|
| 1 | Canada | Montreal | Ahuntsic-Cartierville | 101.98031 |
| 2 | Canada | Montreal | Anjou | 87.41509 |
| 3 | Canada | Montreal | Baie-d'Urfé | 253.16667 |
| 4 | Canada | Montreal | Beaconsfield | 118.47619 |
| 5 | Canada | Montreal | Côte-Saint-Luc | 151.03922 |
| 6 | Canada | Montreal | Côte-des-Neiges-Notre-Dame-de-Grâce | 113.87314 |
| 7 | Canada | Montreal | Dollard-des-Ormeaux | 132.34375 |
| 8 | Canada | Montreal | Dorval | 150.75862 |
| 9 | Canada | Montreal | Hampstead | 142.00000 |
| 10 | Canada | Montreal | Kirkland | 205.12500 |
| 11 | Canada | Montreal | L'Île-Bizard-Sainte-Geneviève | 200.02174 |
| 12 | Canada | Montreal | LaSalle | 115.56061 |
| 13 | Canada | Montreal | Lachine | 125.49020 |
| 14 | Canada | Montreal | Le Plateau-Mont-Royal | 146.58098 |
| 15 | Canada | Montreal | Le Sud-Ouest | 155.52098 |
| 16 | Canada | Montreal | Mercier-Hochelaga-Maisonneuve | 107.48408 |
| 17 | Canada | Montreal | Mont-Royal | 125.17391 |
| 18 | Canada | Montreal | Montréal-Est | 73.00000 |
| 19 | Canada | Montreal | Montréal-Nord | 110.49091 |
| 20 | Canada | Montreal | Montréal-Ouest | 87.10000 |
| 21 | Canada | Montreal | Outremont | 140.88372 |
| 22 | Canada | Montreal | Pierrefonds-Roxboro | 153.26087 |
| 23 | Canada | Montreal | Pointe-Claire | 129.07500 |
| 24 | Canada | Montreal | Rivière-des-Prairies-Pointe-aux-Trembles | 129.54098 |
| 25 | Canada | Montreal | Rosemont-La Petite-Patrie | 116.65699 |

Top 10 Cheapest Neighborhood

| | country | city | neighborhood | mean.price |
|---|---|---|---|---|
| | <chr> | <chr> | <chr> | <dbl> |
| 1 | Canada | Portland | Pleasant Valley/Powellhurst-Gilbert | 40 |
| 2 | United States of America | NYC | New Dorp | 40 |
| 3 | United States of America | LosAngeles | Desert View Highlands | 40.3 |
| 4 | United States of America | LosAngeles | Watts | 44.1 |
| 5 | Canada | NewBrunswick | Blissville | 45 |
| 6 | United States of America | SanMateo | Colma | 49.2 |
| 7 | United States of America | Oakland | Hegenberger | 51 |
| 8 | United States of America | LosAngeles | Historic South-Central | 51.9 |
| 9 | United States of America | Chicago | West Englewood | 53.5 |
| 10 | Canada | Portland | Centennial/Pleasant Valley | 54 |

Top 10 Most Expensive Neighborhood

| | country | city | neighborhood | mean.price |
|---|---|---|---|---|
| | <chr> | <chr> | <chr> | <dbl> |
| 1 | United States of America | SanDiego | Eastlake Woods | 661 |
| 2 | United States of America | NYC | Fort Wadsworth | 650 |
| 3 | United States of America | RhodeIsland | New Shoreham | 529. |
| 4 | United States of America | Oakland | Hiller Highlands | 524 |
| 5 | United States of America | LosAngeles | Aliso and Wood Regional Park | 513 |
| 6 | United States of America | LosAngeles | Hidden Hills | 506. |
| 7 | United States of America | Austin | 78712 | 500 |
| 8 | Canada | Portland | Healy Heights/Southwest Hills | 499 |
| 9 | United States of America | NYC | Hollis Hills | 497 |
| 10 | United States of America | SanDiego | Horton Plaza | 485 |

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

## Goal: Is there a relationship between neighborhood cost and ratings?

Approach:

1. Filter listings in the US and Canada
2. Group by Country, City, Neighborhood
3. Find the mean for price
4. Find the mean for ratings
5. Plot the relationship between the price & ratings
6. Run regression to test the relationship

Null Hypothesis: There is no relationship between neighborhood cost and ratings

Alternate Hypothesis: There is a relationship between neighborhood cost and ratings

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

***Code:***

```
airbnb_data <-read_csv('~/Desktop/Textbook/AppliedStat/DATA/Airbnb/Airbnb_Listings.csv')

filtered_data <- airbnb_data %>%
  filter(country %in% c("United States of America", "Canada"))%>%
  group_by(country, city, neighborhood) %>%
  summarize(mean.price = mean(price, na.rm = TRUE),
mean.ratings = mean(review_scores_rating, na.rm = TRUE))

ggplot(data=filtered_data,
       aes(x=mean.ratings, y=mean.price))+
  geom_smooth(method='lm')+
  geom_point()+
  geom_point(color = "black", size = 0.2,shape = 1)+
  theme_minimal()

summary(lm(data=filtered_data, mean.price~mean.ratings))
```

| | country | city | neighborhood | mean.price | mean.ratings |
|---|---|---|---|---|---|
| 1 | Canada | Montreal | Ahuntsic-Cartierville | 101.98031 | 4.667696 |
| 2 | Canada | Montreal | Anjou | 87.41509 | 4.475714 |
| 3 | Canada | Montreal | Baie-d'Urfé | 253.16667 | 4.752500 |
| 4 | Canada | Montreal | Beaconsfield | 118.47619 | 4.635882 |
| 5 | Canada | Montreal | Côte-Saint-Luc | 151.03922 | 4.208235 |
| 6 | Canada | Montreal | Côte-des-Neiges-Notre-Dame-de-Grâce | 113.87314 | 4.593798 |
| 7 | Canada | Montreal | Dollard-des-Ormeaux | 132.34375 | 4.325217 |
| 8 | Canada | Montreal | Dorval | 150.75862 | 4.481304 |
| 9 | Canada | Montreal | Hampstead | 142.00000 | 4.536667 |
| 10 | Canada | Montreal | Kirkland | 205.12500 | 4.646000 |
| 11 | Canada | Montreal | L'Île-Bizard-Sainte-Geneviève | 200.02174 | 4.696364 |
| 12 | Canada | Montreal | LaSalle | 115.56061 | 4.661667 |
| 13 | Canada | Montreal | Lachine | 125.49020 | 4.632235 |
| 14 | Canada | Montreal | Le Plateau-Mont-Royal | 146.58098 | 4.580131 |
| 15 | Canada | Montreal | Le Sud-Ouest | 155.52098 | 4.602987 |
| 16 | Canada | Montreal | Mercier-Hochelaga-Maisonneuve | 107.48408 | 4.560299 |
| 17 | Canada | Montreal | Mont-Royal | 125.17391 | 4.304103 |
| 18 | Canada | Montreal | Montréal-Est | 73.00000 | 4.260000 |
| 19 | Canada | Montreal | Montréal-Nord | 110.49091 | 4.561463 |
| 20 | Canada | Montreal | Montréal-Ouest | 87.10000 | 4.785000 |
| 21 | Canada | Montreal | Outremont | 140.88372 | 4.555660 |
| 22 | Canada | Montreal | Pierrefonds-Roxboro | 153.26087 | 4.779815 |
| 23 | Canada | Montreal | Pointe-Claire | 129.07500 | 4.773235 |
| 24 | Canada | Montreal | Rivière-des-Prairies-Pointe-aux-Trembles | 129.54098 | 4.515833 |
| 25 | Canada | Montreal | Rosemont-La Petite-Patrie | 116.65699 | 4.626067 |
| 26 | Canada | Montreal | Saint-Laurent | 124.28804 | 4.600763 |

# Q3. Where are the cheapest and most expensive neighborhoods to rent an Airbnb in the US and Canada?

***Results:***

- When one unit increases in Ratings, Price will be increased by 37.9
- P-value: super small, null rejected
- Significant (variable)

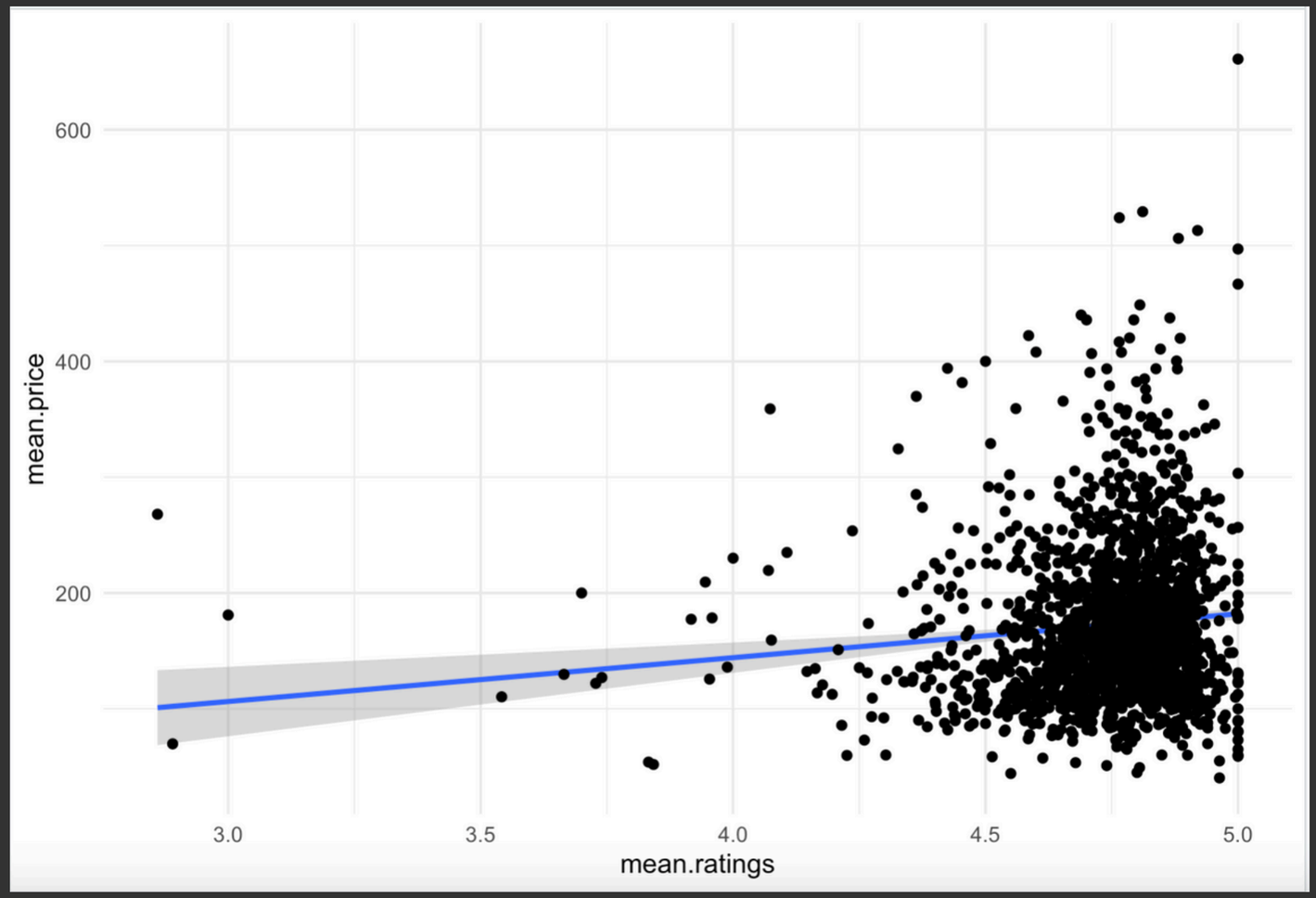
```
Call:
lm(formula = mean.price ~ mean.ratings, data = filtered_data)

Residuals:
    Min      1Q  Median      3Q     Max
-140.31  -43.79  -12.10   27.69  478.97

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    -7.285     41.334  -0.176     0.86
mean.ratings   37.862      8.701   4.351 1.42e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 66.54 on 1914 degrees of freedom
  (10 observations deleted due to missingness)
Multiple R-squared:  0.009795,  Adjusted R-squared:  0.009278
F-statistic: 18.93 on 1 and 1914 DF,  p-value: 1.425e-05
```

Relationship between mean Price & mean Ratings



# ----THE END ---