

# AIR BNB PRICE OPTIMIZATION

# Overview

1. Business problem
2. Dataset
3. Data Wrangling (Data Preprocessing)
4. Exploratory Data Analysis
5. Cluster Analysis
6. Sentiment Analysis
7. Machine Learning – K-best, Random Forest
8. Linear Regression
9. Logistic Regression
10. Final Output



## Business Problem

How can data-driven insights and analytics be used to assist current and future Airbnb listing owners in their pricing strategies?

# Dataset

## SOURCES: Inside Airbnb



- Listings.csv
- Calendar.csv
- Neighborhoods.csv
- Data sets from other cities: Boston, Seattle & Denver
  - Boston: 3862 rows, 75 columns
  - Seattle: 6376 rows, 75 columns
  - Denver: 5362 rows, 75 columns

# Data Wrangling/ Data Preprocessing

- Check Dataset
  - Data Profiling
  - `describe()`, `info()`
- Combine Datasets
- Process Duplicates, Missing and Null values
  - With specific value & Deleted
- Drop irrelevant columns
- Convert Data type
  - Encode and Dummy categorical data

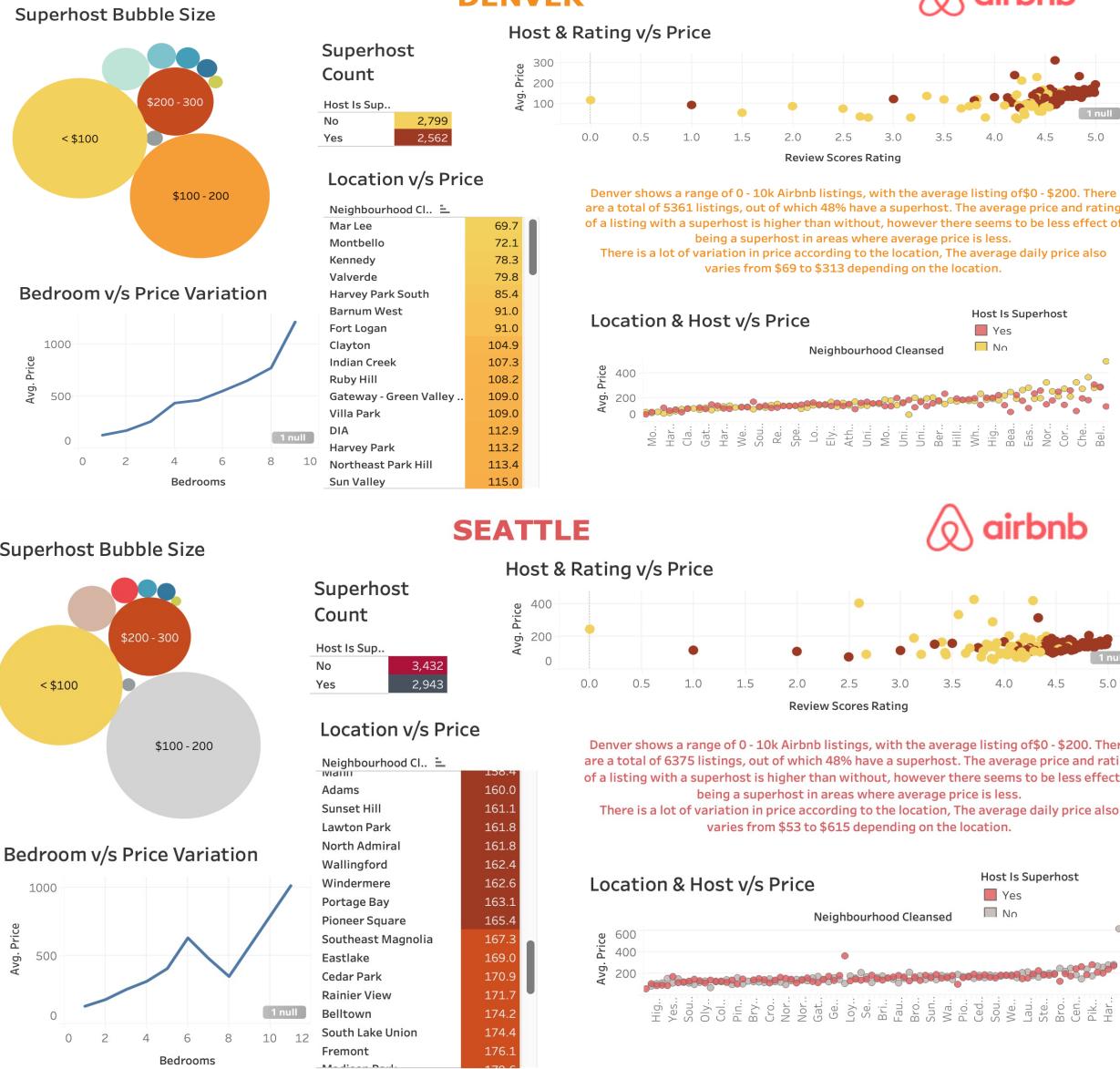
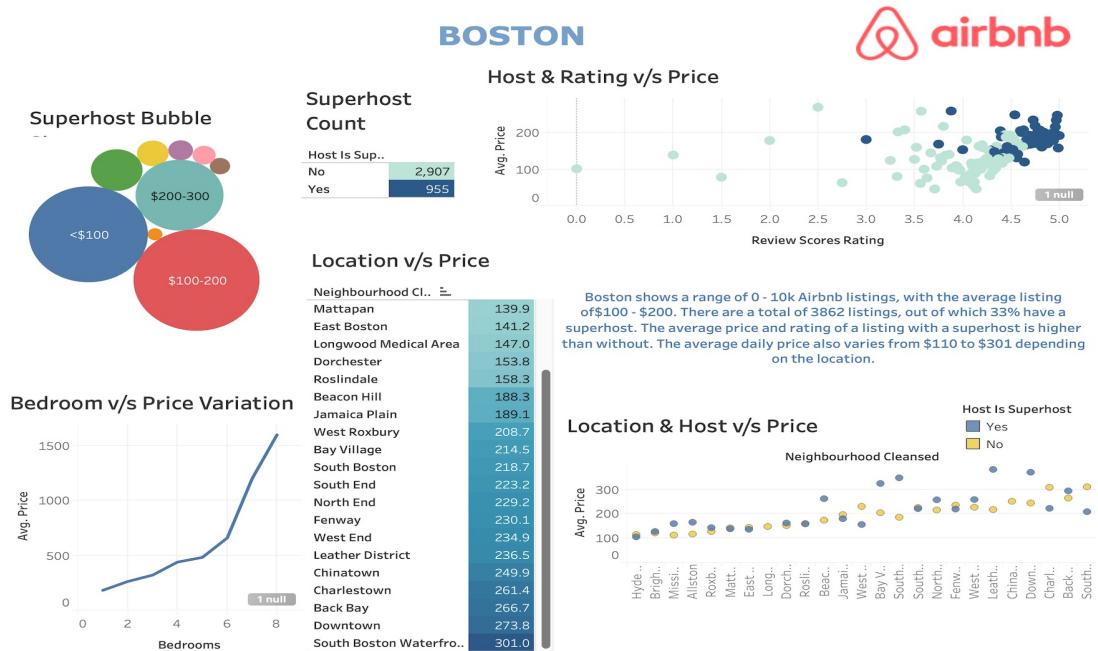


# Exploratory Data Analysis



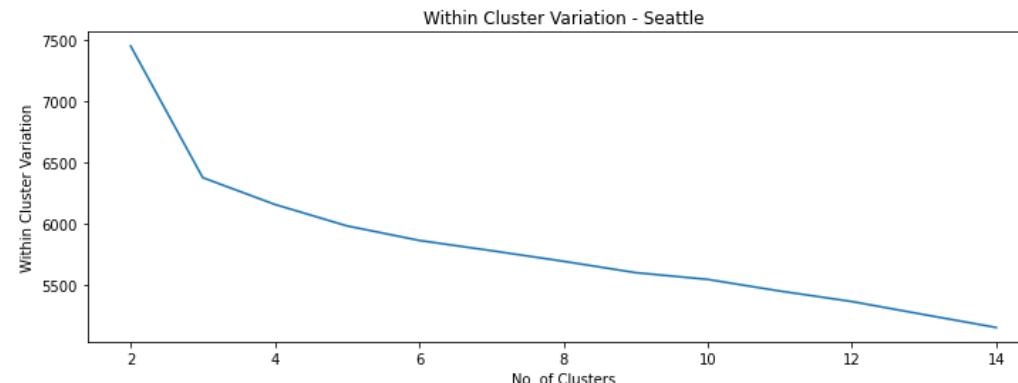
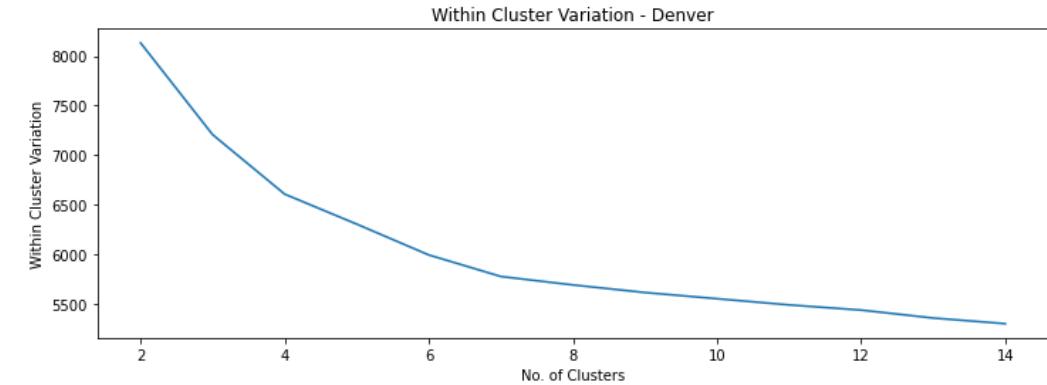
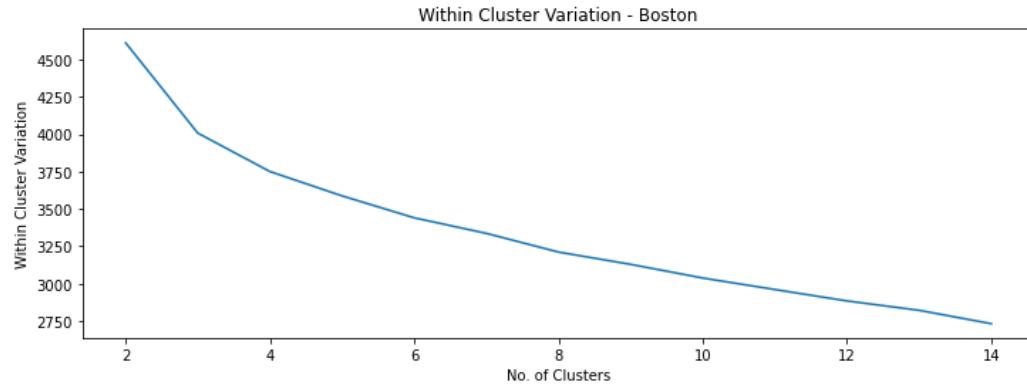
## Inferences:

- Superhosts are profitable with higher prices and better ratings, but the effect decreases with equal percentages.
- Prices vary with the number of bedrooms in different cities.
- Neighbourhood greatly impacts price, with some charging 2x more.
- The \$100 - \$200 range is the biggest market in all three cities.



# Cluster Analysis

Determining the ideal number of clusters



# Cluster Analysis

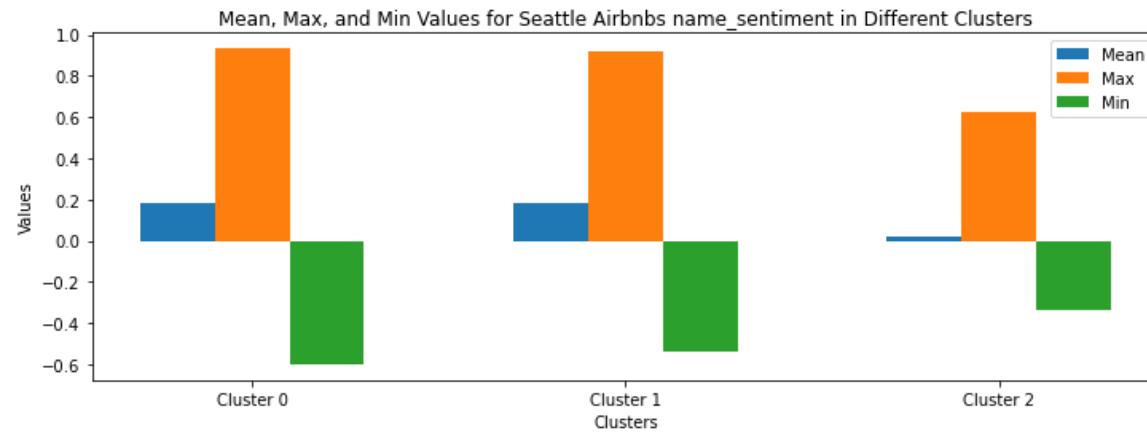
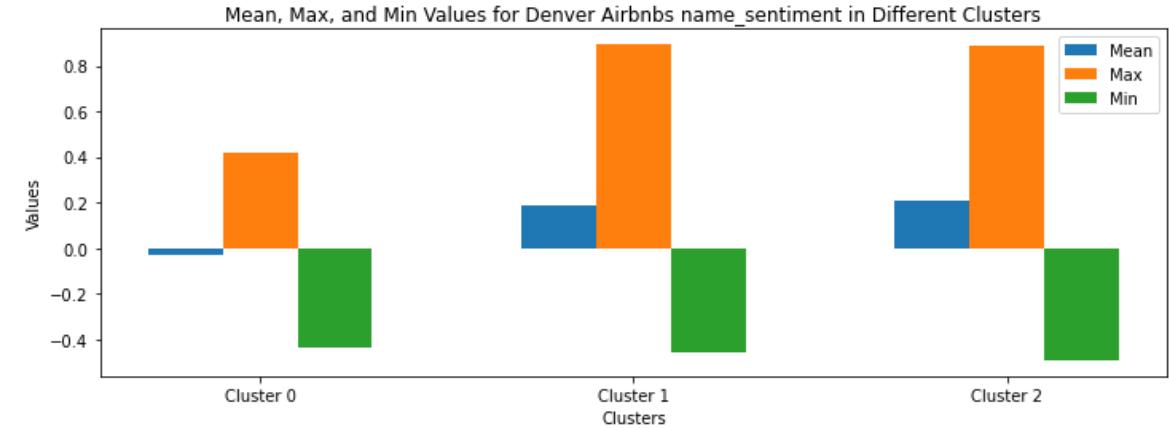
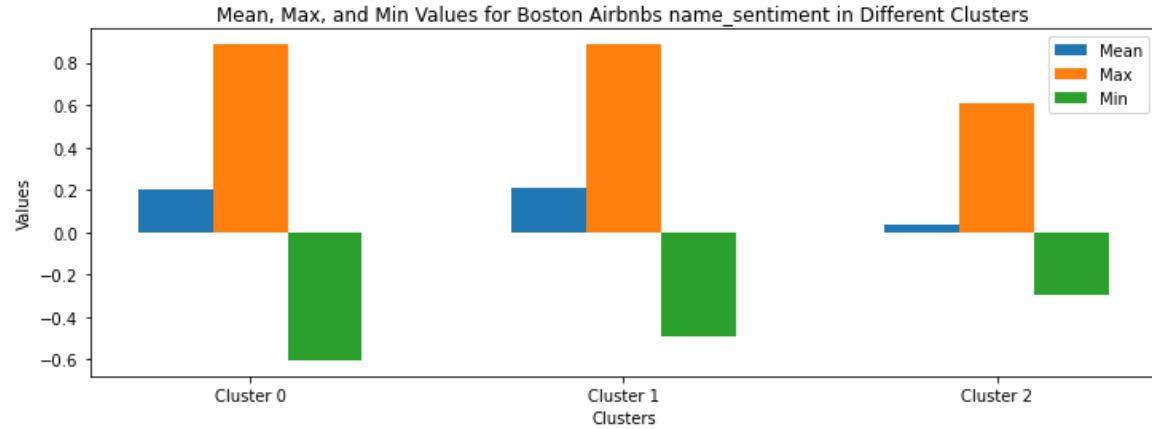
- We conducted Cluster Analysis to better assist future listers in understanding the competitive environment of current listings
- Sentiment Analysis results were incorporated in the dataframe to assist in cluster formation and analysis.
- For each city, using the metric of within cluster variation, ideal number of clusters chosen was 3.



# Sentiment Analysis

- Natural Language Tool Kit (NLTK) library in Python utilized
- Sentiment Analysis Performed on Columns:
  - 'name',
  - 'description',
  - 'neighborhood\_overview',
  - 'host\_about'
- Scores for each sentiment analysis used and stored in a new column

# Results of Sentiment Analysis



# Select Key Features on Price

- Correlation

- Higher absolute values indicate a stronger relationship

Correlation  
Coefficients

Feature	Correlation coefficients with price
price	1
accommodates	0.405735957
beds	0.378846149
2 baths	0.279289952
Entire home/apt	0.267555011
3+ baths	0.246283988
Private	
Bathroom	0.237793389
pets_allowed	0.14723263
TV	0.13090944
Downtown_Area	0.125160719

- Two Machine Learning Models

- Random Forest
  - Extracting feature importance
  - Accuracy: 0.13
- Select-K-Best
  - Using the chi-squared test
  - Accuracy: 0.08

Random Forest performs  
better

Feature	Importance
availability_365	0.06528
availability_90	0.054324
number_of_reviews	0.053303
availability_60	0.047694
availability_30	0.038547
number_of_reviews_ltm	0.03667
minimum_nights_avg_ntm	0.03474
accommodates	0.03014
host_total_listings_count	0.027294
maximum_minimum_nights	0.022508

# Linear Regression



- We built a Linear Regression model for Boston, Seattle and Denver:
  - Y : Price
  - X : Bathrooms, Beds, Super-host, Neighborhoods, Amenities.
- We incorporated distinctive characteristics of each city in the regression.

Boston
<ul style="list-style-type: none"><li>• Educational Hub: "dedicated_workspace", "long_stay"</li><li>• Waterfront: "beach_access", "waterfront_view"</li><li>• City view: "skyline_view"</li></ul>

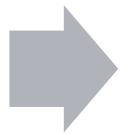
Seattle
<ul style="list-style-type: none"><li>• Tech Industry: "dedicated_workspace", "private_entrance"</li><li>• Waterfront: "beach_access", "waterfront_view"</li><li>• Outdoor activities: "patio", "bbq_grill", "park_view"</li></ul>

Denver
<ul style="list-style-type: none"><li>• Family trips: "parking", "cribs", "fireplace"</li><li>• Outdoor activities: "patio", "bbq_grill", "outdoor_furniture"</li><li>• Sunny weather: "pool", "ac_unit"</li></ul>

# Linear Regression - Neighborhoods

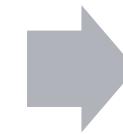


**Boston**



## Positive

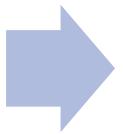
- Back Bay: 82.332
- Beacon Hill: 30.454
- Downtown: 64.017



## Negative

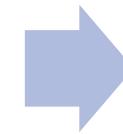
- East Boston: -49.858
- Mission Hill: -36.123
- Roxbury: -37.175

**Seattle**



## Positive

- Alki: 18.464
- Broadview: 66.694
- Central Business District: 56.038



## Negative

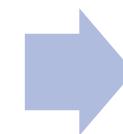
- Atlantic: -23.119
- Roxhill: -10.803
- South Beacon Hill: -65.734

**Denver**



## Positive

- Belcaro: 22.712
- Jefferson Park: 79.698
- North Park Hill: 8.817



## Negative

- Congress Park: -60.120
- Fort Logan: -112.413
- Kennedy: -104.772

# Linear Regression - Main Characteristics



Boston



Positive

- Yes\_Instant\_Bookable: 14.809
- Entire home/apt: 58.260
- Private room: 36.092
- 2 baths: 102.919
- beds: 29.380



Negative

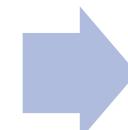
- maximum\_nights: -0.014

Seattle



Positive

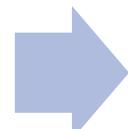
- Entire home/apt: 35.841
- Private room: 14.663
- 2 baths: 34.130
- beds: 4.695



Negative

- Yes\_Host\_ProfilePic: -12.059
- maximum\_nights: -0.008

Denver



Positive

- Entire home/apt: 181.506
- 2 baths: 20.109
- 3+ baths: 190.811
- beds: 27.486



Negative

- Yes\_Host\_ProfilePic: -30.551

\*Hosts might charge a premium to offset these risks. Risks: last-minute cancellations or potentially problematic guests.

\*Boston is a city with a high cost of living and expensive real estate market.

\*Owners might offer discounts for longer stays

\*Cost of living and real estate market might not be as expensive as Boston's.

\*Denver is known for its outdoor recreational opportunities and may have a large market for entire homes or apartments catering to families or groups looking for accommodations together.

# Linear Regression - Amenities



Boston

Positive

- skyline\_view: 52.249
- pets\_allowed: 19.203
- pool: 12.397
- elevator: 24.019

Negative

- dedicated\_workspace: -15.250

\*City views are desirable in Boston

\*Listings with workspace might be in University neighborhoods.

Seattle

Positive

- skyline\_view: 3.563
- beach\_acces: 2.366
- harbor\_view: 20.864
- dedicated\_workspace: 8.424

Negative

- bathtub: -8.427

\*Seattle is known for the tech-hub so listings might charge more for workspace.

Denver

Positive

- skyline\_view: 5.791
- fireplace: 44.761
- bbq: 21.286
- outdoor\_furniture: 8.405

Negative

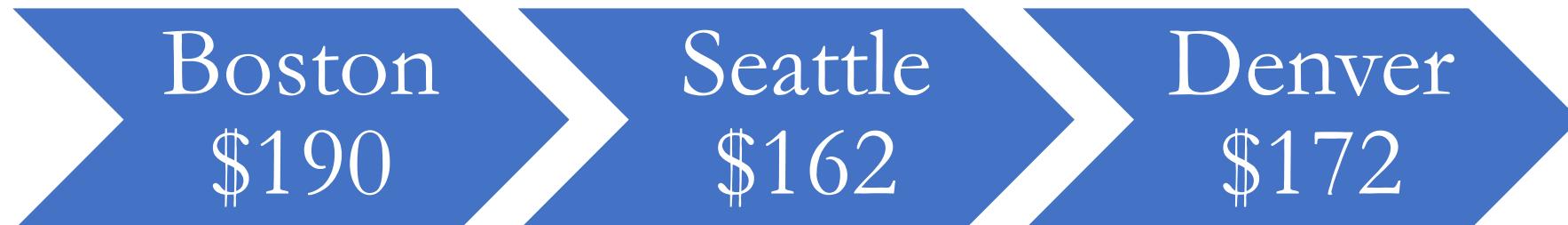
- ac: -9.009

\*Denver's skyline, while not as famous as some other cities, is still attractive and can add value to a property

\*Listings with AC unit window may be older places.

# Logistic Regression

- We built a Logistic Regression model for Boston, Seattle and Denver.
- Average Price for each city, and turned it into a dummy variable:
  - 1: Listing has a price higher than average
  - 0: Listing has a price below average



# Logistic Regression

Boston



R2: 0.76



Positive  
• 2 baths



Negative  
• Dorchester  
• East Boston  
• Mission Hill

Seattle



R2: 0.70



Positive  
• 3+ baths  
• Central Business District



Negative  
• Rainier Beach

Denver



R2: 0.67



Positive  
• 3+ baths



Negative  
• 2+ shared baths

# Results : Case-1, Boston



## (J8) Sophisticated studio in Heart of Back Bay!

★ 4.69 · 123 reviews · Superhost  
Boston, Massachusetts, United States

Lower price. Your dates are \$418 less than the avg. nightly rate of the last 60 days.

Entire rental unit hosted by Stone

2 guests · Studio · 1 bed · 1 bath

\$343 \$259 night  
Jun 28 – Jul 3

Reserve

### Original Listing

```
# specify the characteristics of the listing
listing = {
    'accommodates': 2,
    'availability_30': 5,
    'Is_Superhost': 1,
    'Yes_Host_IDVerified': 0,
    'Yes_Instant_Bookable': 1,
    'Entire home/apt': 1,
    'Back_Bay': 1,
    'wifi': 1,
    'skyline_view': 0,
    'pets_allowed': 1,
    'pool': 0,
    'patio': 1,
    '1_bath': 1,
    'beds': 1
}

# calculate the price by multiplying the characteristics by the coefficients and summing the products
price = model.intercept_
for feature, value in listing.items():
    if feature in coef_df['feature'].values:
        price += value * coef_df.loc[coef_df['feature'] == feature, 'coefficient'].values[0]

print(f"Estimated price of the listing: ${price:.2f}")

Estimated price of the listing: $288.34
```

### Optimized Price

Output from Regression Model for optimized price : \$ 288.34

### Features of Listing :

- Backbay
- Superhost
- Instantly Bookable
- Entire Home / Apt
- 1 bed

Original Listing  
Price : \$343

### Suggestions :

The host should re-adjust the price of the listing and decrease to average \$280- 290 (excluding holidays), for optimized pricing

# Results : Case-2, Denver



## Specious Denver Home w/ Large Backyard!

★ New · Superhost  
Denver, Colorado, United States

Entire home hosted by  
**Oshi**

8 guests · 4 bedrooms · 5 beds  
3.5 baths

 Dedicated workspace  
A common area with wifi that's well-suited for working.

\$315 night  
Jul 23 – 28

Reserve

### Features of Listing :

- Belcaro
- Superhost
- Entire Home / Apt
- Accommodates 8
- 4 beds , patio, fireplace

```
# specify the characteristics of the listing
listing = {
    'accommodates': 8,
    'yes_Host_ProfilePic': 1,
    'Entire home/apt': 1,
    'Belcaro': 1,
    '2 baths':1,
    'beds':4,
    'skyline_view': 1,
    'dedicated_workspace': 1,
    'pets_allowed': 1,
    'bbq': 1,
    'patio': 1,
    'fireplace':1
}

# calculate the price by multiplying the characteristics by the coefficients and summing the products
price = model.intercept_
for feature, value in listing.items():
    if feature in coef_df['feature'].values:
        price += value * coef_df.loc[coef_df['feature'] == feature, 'coefficient'].values[0]

print(f"Estimated price of the listing: ${price:.2f}")

Estimated price of the listing: $355.74
```

### Optimized Price

Output from Regression  
Model for optimized  
price : \$ 355.74

### Suggestions :

The price of the listing can be increased to \$355 (excluding holidays), for optimized price after it has a few reviews as it is a newly listed property.



# Results : Case-3, Seattle

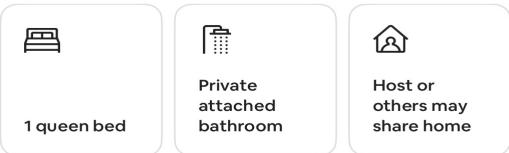


Original Listing

## Minerva Inn - Saffron Suite - Capitol Hill

★ 4.96 · 335 reviews · Superhost  
Seattle, Washington, United States

### Room in a guest suite hosted by Bill And Michael



Room in a guest suite  
Your own room in a home, plus access to shared spaces.

\$110 night  
Jul 19 – 21

Reserve

### Features of Listing :

- Capitol Hill/ South Lake Union
- Instantly Bookable
- Private Room
- Workspace

Original Listing  
Price : \$110

```
# specify the characteristics of the listing
listing = {
    'accommodates': 1,
    'Is_Superhost': 1,
    'Yes_Host_IDVerified': 0,
    'Yes_Instant_Bookable': 1,
    'Yes_Host_ProfilePic': 1,
    'Private_room': 1,
    'South_Lake_Union': 1,
    '1_baths': 1,
    'beds': 1,
    'wifi': 1,
    'dedicated_workspace': 1,
    'pets_allowed': 1,
    'harbor_view': 1,
    'patio': 1
}

# calculate the price by multiplying the characteristics by the coefficients and summing the products
price = model.intercept_
for feature, value in listing.items():
    if feature in coef_df['feature'].values:
        price += value * coef_df.loc[coef_df['feature'] == feature, 'coefficient'].values[0]

print(f"Estimated price of the listing: ${price:.2f}")

Estimated price of the listing: $112.92
```

Optimized Price



Output from Regression  
Model for optimized  
price : \$ 112.92

### Suggestions :

The original price is very close to the optimized price. The host does not need to make any changes.



# THANK YOU FOR LISTENING,

Contact for your Airbnb listings.