

Airbnb Analysis

Survey of Machine Learning and Artificial Intelligence

Kenneth Alexis, Meme Habel, Sam Moushey



Agenda



Variable Relationships



Analysis



Linear Regression



Why is this important?

Variable Exploration

Categorical Variables:

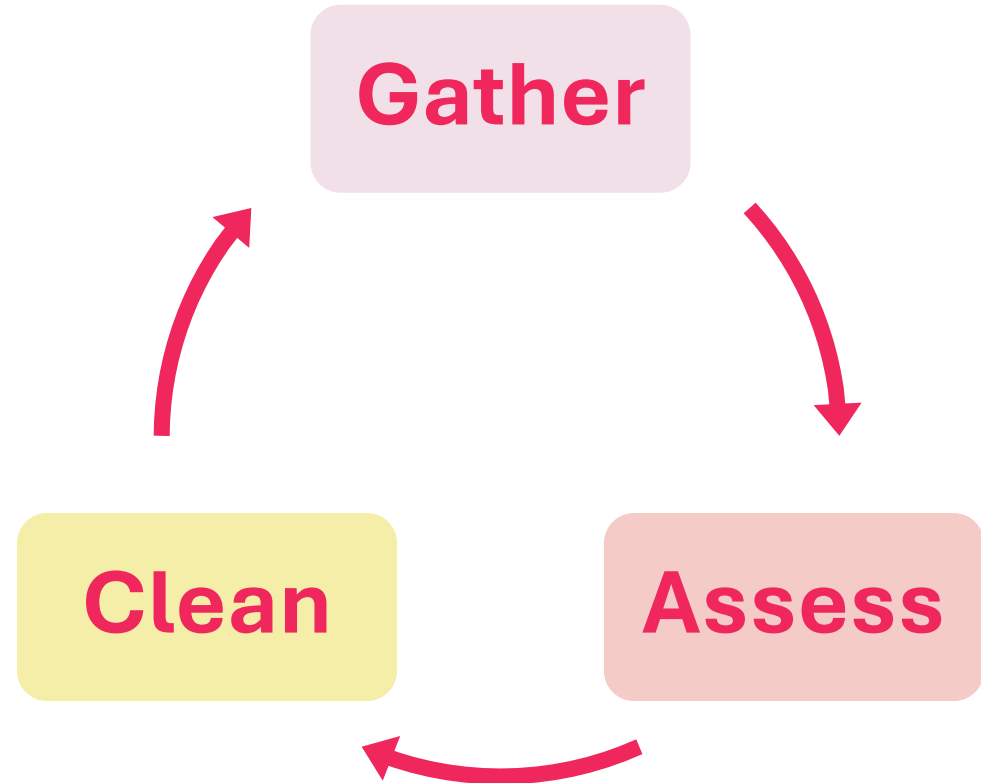
- host_is_superhost (t, f)
- host_has_profile_pic (t, f)
- host_identity_verified (t, f)
- is_location_exact (t, f)
- property_type
- neighbourhood
- room_type
- city
- bed_type
- has_availability (t) – ALL true
- requires_license (f) – ALL false
- instant_bookable (t, f)
- is_business_travel_ready (f) – ALL false
- require_guest_profile_picture (t, f)
- require_guest_phone_verification (t, f)

Empty Variables (contain no data):

- neighbourhood_group_cleansed
- license
- xl_picture_url
- experiences_offered

Data Cleaning + Wrangling

- Drop empty variables
- Remove trailing spaces from string columns
- Apply consistent capitalization to strings in 'city' column
- Convert 'price' column to numeric
- Dummy encode categorical variables



Analysis: Airbnb Locations

- The city with the most Airbnb listings is San Jose with a total of 3032 throughout its 35 zip codes.
- Danville and Watsonville have the lowest number of listings with 1 each
- The zip code containing the most listings is 95112 with 363 listings
- South Bay, California is the biggest market for Airbnb listings with 7,104 listings

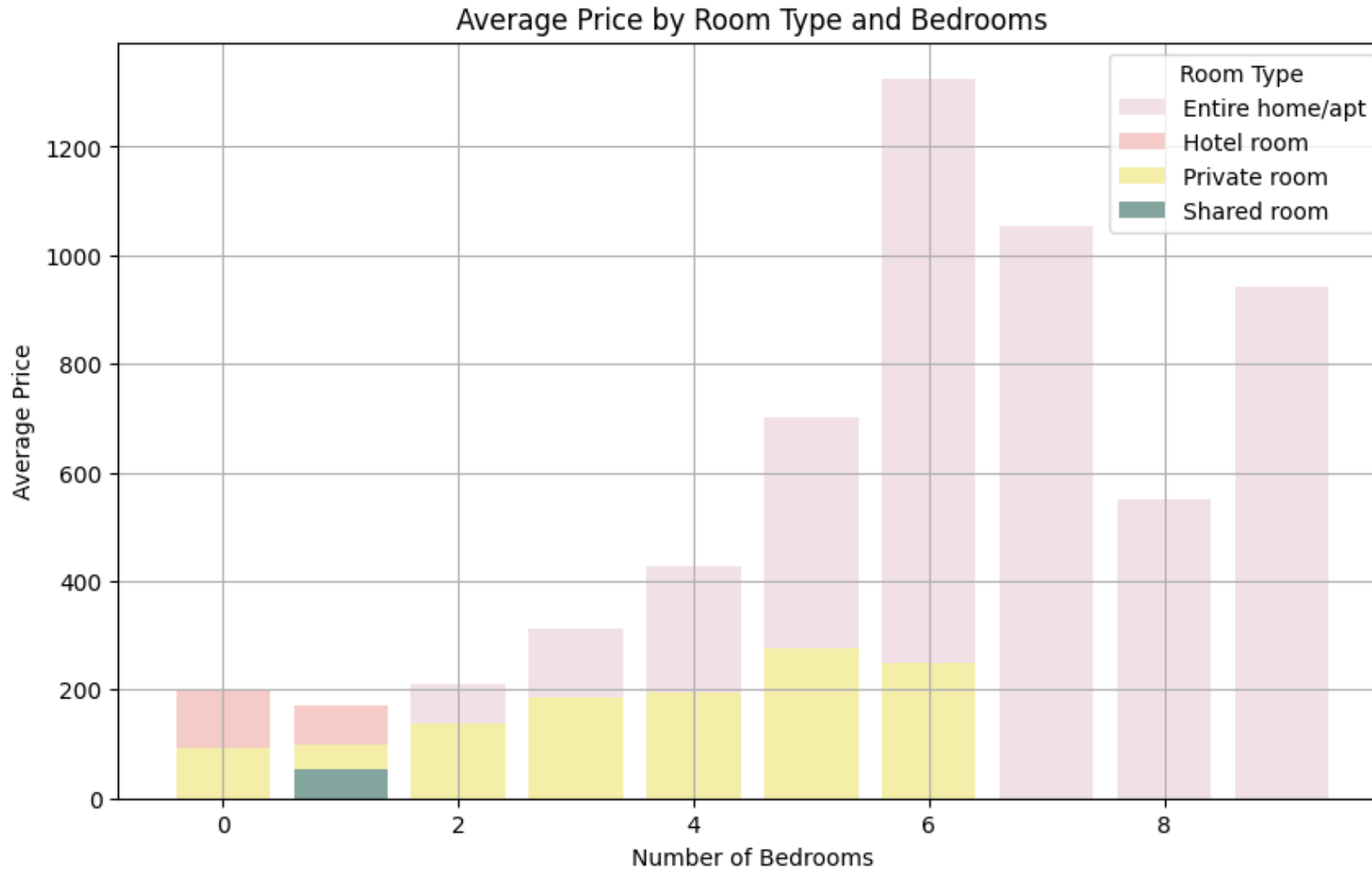


Analysis: Average Nightly Price by City

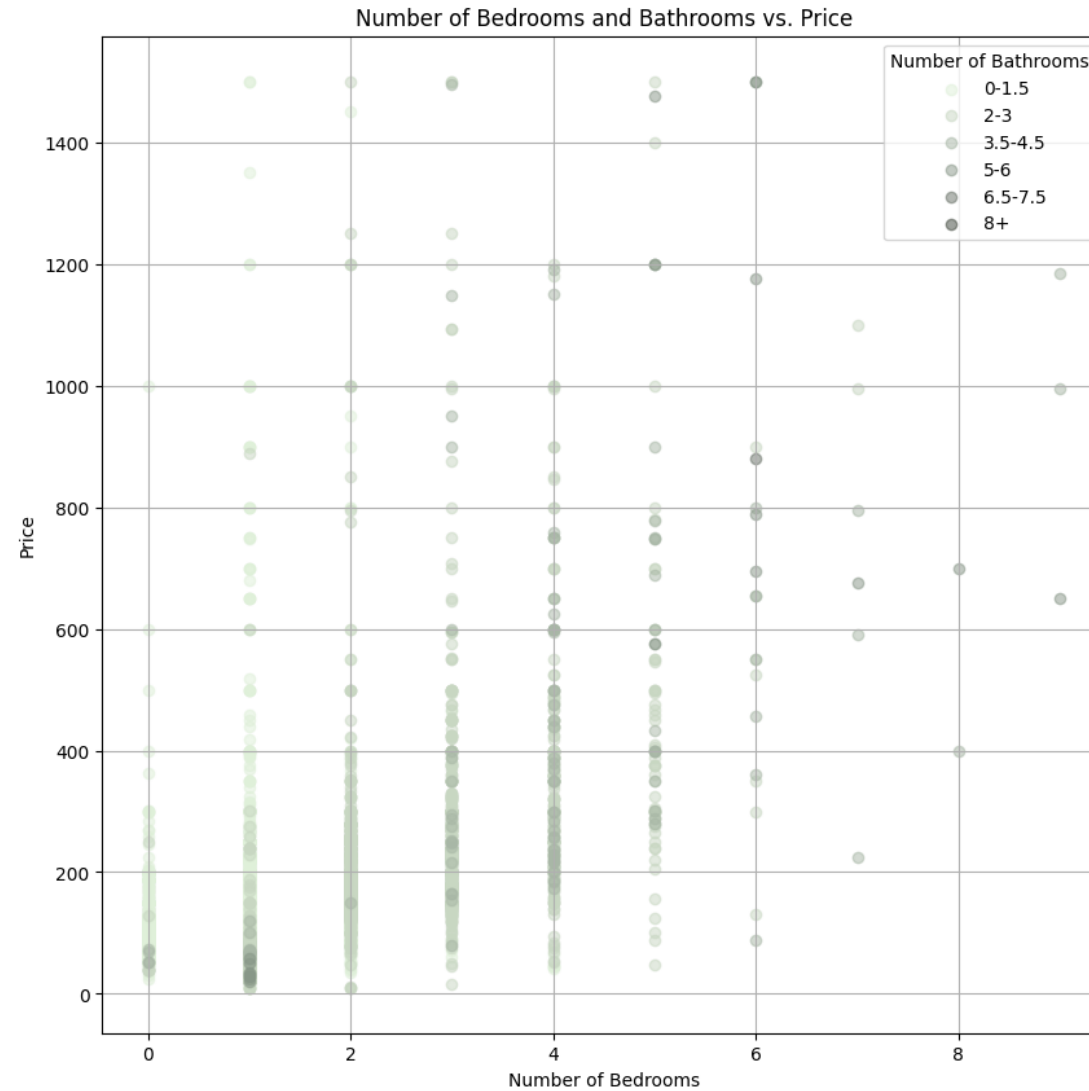
City	Average Price	Number of Listings
Saratoga	\$300.97	70
Danville	\$280.00	1
Los Gatos	\$261.45	134
Menlo Park	\$253.50	4
Los Altos Hills	\$253.15	53
Hollister	\$248.00	2
Los Altos	\$234.32	90
Stanford	\$228.67	3
San Martin	\$225.93	15
Mountain View	\$224.04	666
Palo Alto	\$221.73	792

City	Average Price	Number of Listings
Morgan Hill	\$162.94	63
Santa Clara	\$155.42	714
Campbell	\$153.04	130
Gilroy	\$145.19	32
San Jose	\$138.55	3034
Monte Sereno	\$130.56	16
Cupertino	\$129.87	335
Sunnyvale	\$124.90	765
Milpitas	\$120.28	297
Watsonville	\$76.00	1
Fremont	\$62.33	3

Analysis: Average Price by Room Type and Bedrooms



Analysis: Bedrooms vs. Price



Linear Regression: Variable Selection

- **Feature Variables:**

- 'city' (in the form of dummy variables)
- 'accommodates'
- 'bathrooms'
- 'bedrooms'
- 'beds'
- 'room_type' (in the form of dummy variables)

- **Target Variable:**

- 'price'

Linear Regression: Results ('city')

City	Regression Coefficient
Los Altos Hills	154.299849
Saratoga	149.372804
Los Gatos	138.367497
Stanford	101.603828
Mountain View	100.544595
Menlo Park	99.4108957
Palo Alto	87.0981975
Hollister	71.1223283
San Martin	68.5274211
Los Altos	66.9077577
Monte Sereno	42.7294916
Morgan Hill	35.899972
Campbell	23.9738214
Santa Clara	18.3312524
Watsonville	17.4976198
Sunnyvale	13.061682
San Jose	8.23326629
Cupertino	2.25268049
Danville	9.9476E-14
Milpitas	-5.6503648
Fremont	-9.5581079
Gilroy	-13.633891

Linear Regression: Results (remaining predictors)

City	Regression Coefficient
bathrooms	62.2780355
room_type_Hotel room	58.7343003
bedrooms	41.6822833
room_type_Entire home/apt	21.4772106
accommodates	15.1846635
room_type_Private room	-1.9646815
room_type_Shared room	-78.246829

Why is this important?

- Through data wrangling and analysis, we are able to come to conclusions about different aspects of Airbnb listings in California.
- Users can more easily digest the data and make important decisions on where to stay.
- Reduces time and makes the decision process more efficient to filter users' needs for an Airbnb.