Airbnb Global Price Analysis

December 5, 2023

Group 53

- **Richard Wang 509402**
- **Jimeng Yin 516867**
- **Bosan Hsu 519095**

Part 1: Introduction

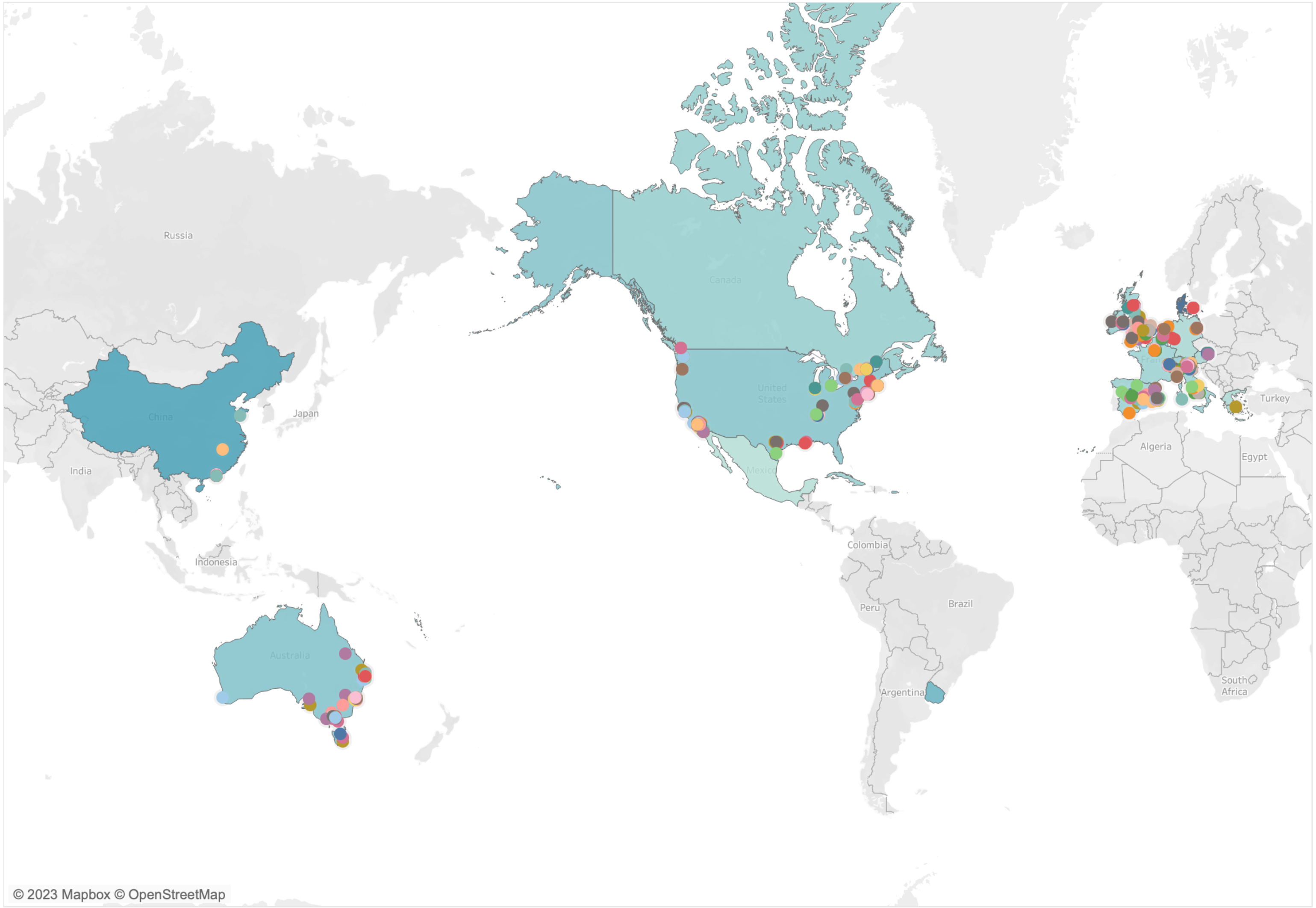
WashU Olin
Business School

Our Data

- **Data:** Airbnb's global price
- **Time Collected:** Last updated four months ago.
- **Data Size:** 1.94 GB
- **Data Source:** Kaggle by Joakim Arvidsson
- **Source Link:** https://www.kaggle.com/datasets/joebeachcapital/airbnb
- **Number of Columns:** 89
- **Number of Records:** 36,248,263
- **Data Type:** Structured

WashU Olin Business School

Our Data



WashU Olin Business School

Why This Data

- **Why is it big data?**
  - Large amount of data
  - Complex structure
  - Cannot be captured, managed, and processed with a single database tool
- **Why choose this data set?**
  - To understand market trends
  - To reveal patterns and correlations by price data
  - Inform decision-making and market strategies
  - this data can be applied to other domains

Problem Statement

- **Goal:** Explore whether
  - 1. the rating scores of Airbnb listings affect the prices of their corresponding listings
  - 2. the features affect the prices of Airbnb
- **What we are doing**: Predicting future Airbnb prices using machine learning methods and statistical modeling

Part 2: Analysis

WashU Olin
Business School

Our Method for Data Analysis

WORDCOUNT

- Transit
- Amenities

SQL

- Response Time
- Property Type

Our Method for Machine Learning

Linear Regression

- Model 1: Price & Number of Different Rooms
- Model 2: Price & Ratings on the website

Part 2-1: Analysis
Data Analysis

WashU Olin
Business School

Data Analysis -- Transit

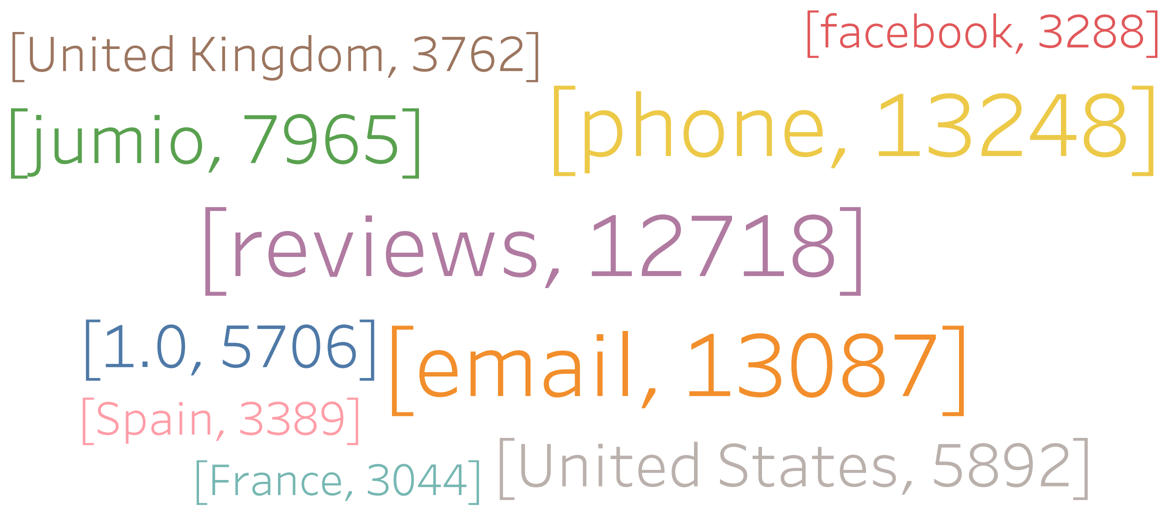
Ten most frequent words with their average price

Transit	count	AveragePrice
phone	13248	9.730272202364587
email	13087	9.733401240855635
reviews	12718	9.731131643948403
jumio	7965	9.738031119090365
United States	5892	9.904449741756059
1.0	5706	9.570974576271187
United Kingdom	3762	9.66472602739726
Spain	3389	9.617452440033086
facebook	3288	9.789454545454545
France	3044	10.288540534253647

WashU Olin Business School

Data Analysis -- Transit

- **Significantly frequent words: "phone", "reviews, and "email"**
  - customers take those three factors seriously
  - e.g. previous reviews, contact availability



Data Analysis -- Amenities

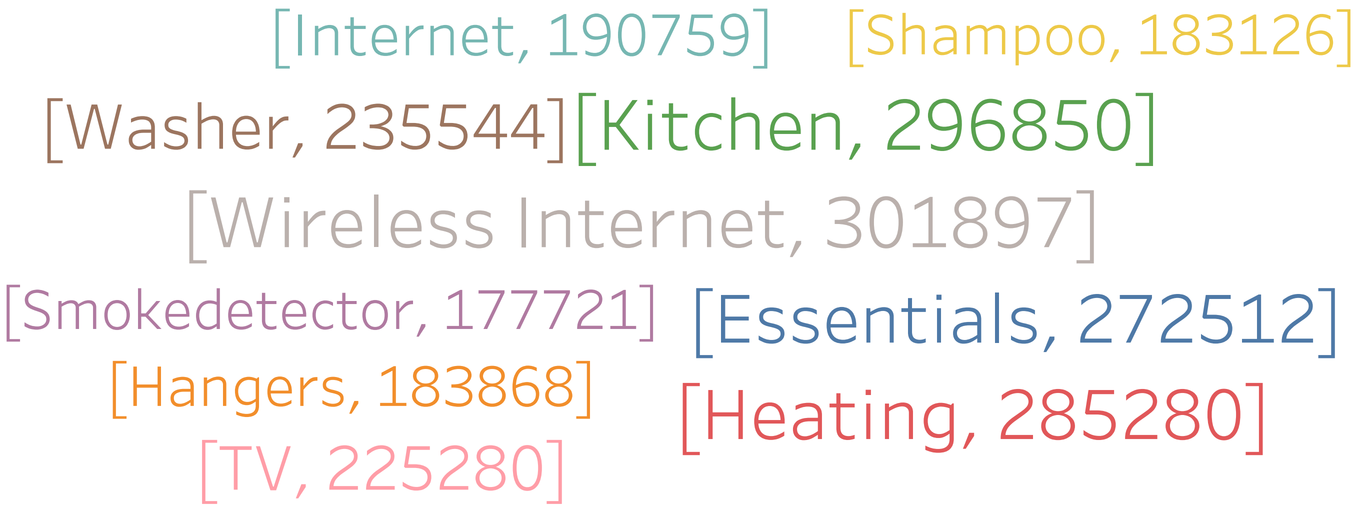
Ten most frequent words with their average price

Amenity	count	AveragePrice
Wireless Internet	301897	139.894383855269
Kitchen	296850	141.42693088455155
Heating	285280	138.53689722531155
Essentials	272512	140.51520093750932
Washer	235544	142.11130688508692
TV	225280	154.405368044634
Internet	190759	144.2520076196168
Hangers	183868	138.560177658159
Shampoo	183126	141.3421657196129
Smoke detector	177721	147.89176864200925

WashU Olin Business School

Data Analysis -- Amenities

- **TV leads to a higher price**
  - Hosts should provide TV to attract customers
  - Low-budget customers should seek rooms without TVs



Data Analysis- Response Time

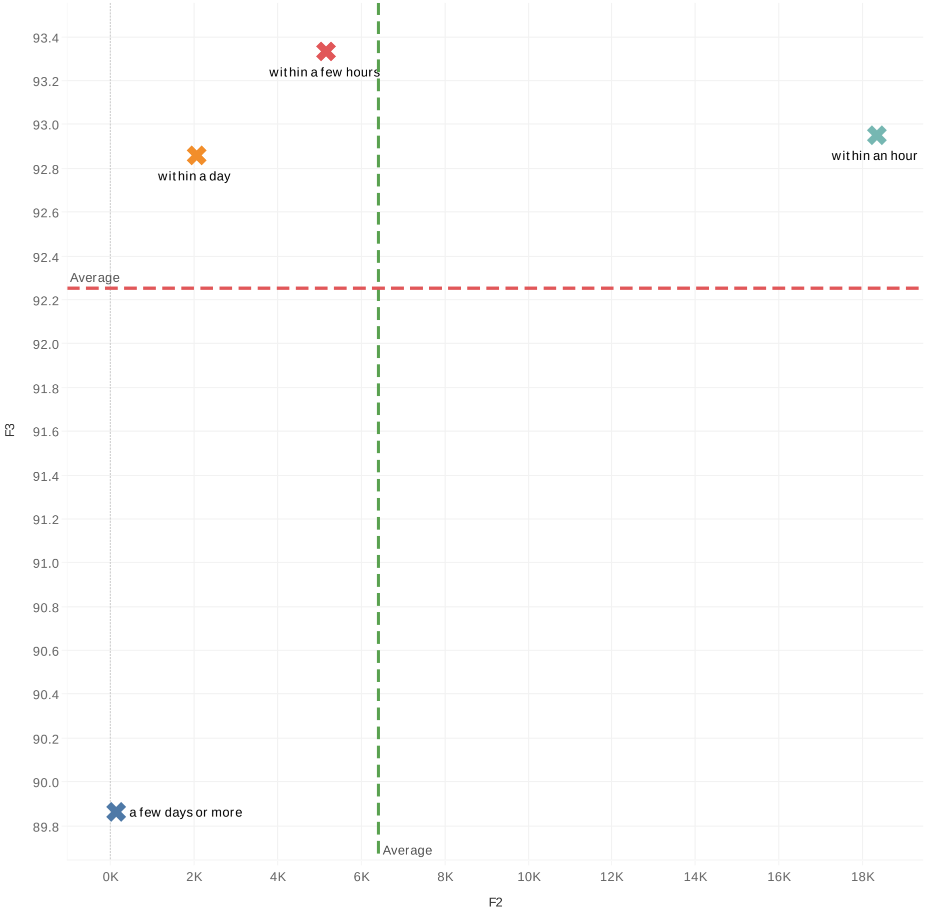
The relationship between the response time and the ratings

Host Response Time	Count_Number_of_Reviews	Avg_Review_Scores_Rating
within a few hours	5144	93.33547257876313
within an hour	18333	92.9530175706463
within a day	2047	92.8611246943765З
a few days or more	119	89.8655462184874

WashU Olin Business School

Data Analysis- Response Time

- **Fast, but not too fast.**
  - Hosts should respond customers within a few hours.



WashU Olin Business School

Data Analysis- Property Types

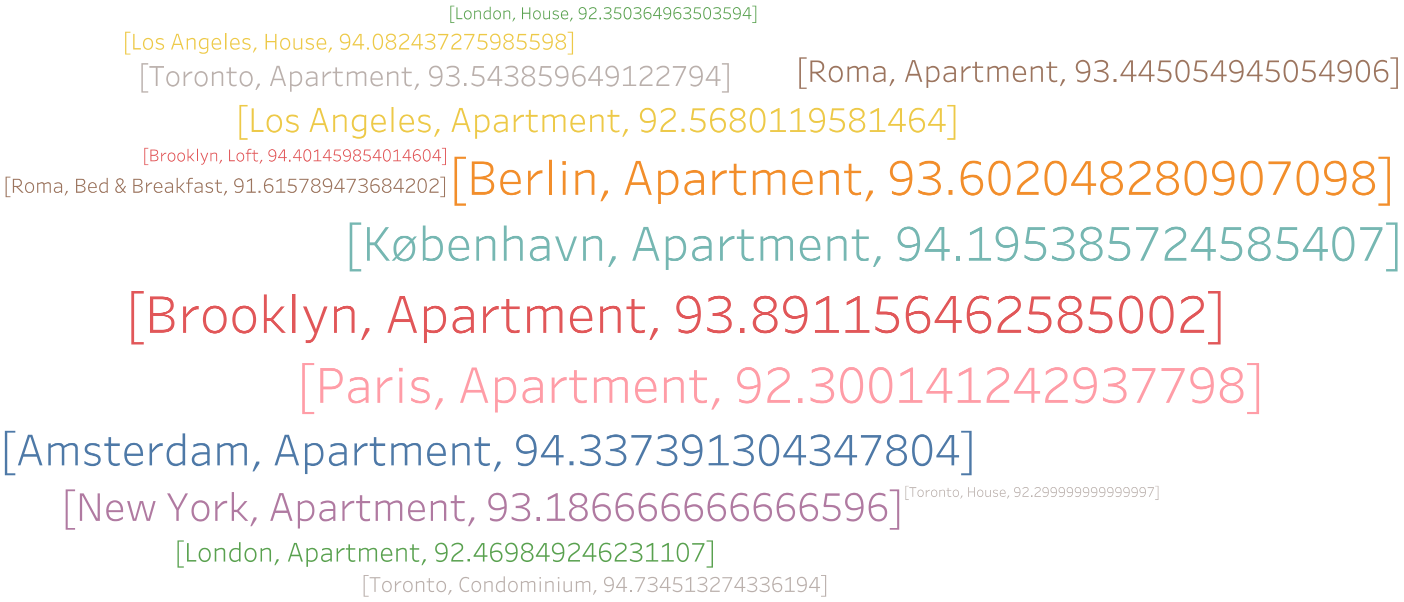
The most common property types in the top 10 city's top neighborhood.

City	Neighbourhood	Property Type	AvgRating	COUNTS
Amsterdam	Oud-West	Apartment	94.33739130434783	1150
Berlin	Neukölln	Apartment	93.6020482809071	1367
Brooklyn	Williamsburg	Apartment	93.89115646258503	1617
Brooklyn	Williamsburg	Loft	94.4014598540146	137
København	Nørrebro	Apartment	94.19538572458544	1387
London	LB of Islington	House	92.35036496350365	137
London	LB of Islington	Apartment	92.46984924623115	398
Los Angeles	Mid-Wilshire	Apartment	92.56801195814649	669
Los Angeles	Mid-Wilshire	House	94.08243727598567	279
New York	Upper West Side	Apartment	93.18666666666667	900
Paris	Montmartre	Apartment	92.30014124293785	1416
Roma	Prati	Bed & Breakfast	91.61578947368422	190
Roma	Prati	Apartment	93.44505494505495	546
Toronto	Downtown Toronto	House	92.3	100
Toronto	Downtown Toronto	Apartment	93.54385964912281	456
Toronto	Downtown Toronto	Condominium	94.73451327433628	226

WashU Olin Business School

Data Analysis- Property Types

- **Apartment is the majority type**
  - Apartment owners in big cities can consider entering the Airbnb market



WashU Olin Business School

Part 2-2: Analysis Machine Learning

WashU Olin Business School

Machine Learning -- Price & Rooms 1

- Independent Variables
  - Accommodates Acc
  - Bathrooms Bat
  - Bedrooms Ber
  - Beds Bed
- Dependent Variable
  - Price Pri
- Pri = 28.48 + 21.77 Acc + 13.33 Bat + 35.40 Ber - 13.65 Bed

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Price   R-squared:                       0.164
Model:                            OLS   Adj. R-squared:                  0.164
Method:                 Least Squares   F-statistic:                 2.374e+04
Date:                Sun, 03 Dec 2023   Prob (F-statistic):               0.00
Time:                        15:57:08   Log-Likelihood:            -3.0714e+06
No. Observations:              484544   AIC:                         6.143e+06
Df Residuals:                  484539   BIC:                         6.143e+06
Df Model:                           4
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const           28.4832      0.484     58.846      0.000      27.535      29.432
Accommodates    21.7748      0.184    118.073      0.000      21.413      22.136
Bathrooms       13.3307      0.429     31.072      0.000      12.490      14.172
Bedrooms        35.4008      0.360     98.417      0.000      34.696      36.106
Beds           -13.6465      0.252    -54.056      0.000     -14.141     -13.152
==============================================================================
Omnibus:                   311633.729   Durbin-Watson:                   0.850
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          3347489.466
Skew:                           3.051   Prob(JB):                         0.00
Kurtosis:                      14.339   Cond. No.                         14.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

WashU Olin Business School

Machine Learning -- Price & Rooms 2

- **Split the data:**
  - 70% training data
  - 30% testing data
- **Mean Squared Error:** 19624.42
- **Rsquared Error:** 16.51%
- **Reasons:**
  - Insufficient feature relevance
  - Underfitting
  - Lack of data

WashU Olin Business School

Machine Learning -- Price & Ratings 1

- Independent Variables
  - Review Scores Rating RSR
  - Review Scores Cleanliness RSC
  - Review Scores Location RSL
- Dependent Variable
  - Price Pri
- Pri = -34.57 + 0.58 RSR + 0.01 RSC + 12.09 RSL

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Price   R-squared:                       0.008
Model:                            OLS   Adj. R-squared:                  0.008
Method:                 Least Squares   F-statistic:                     936.1
Date:                Sun, 03 Dec 2023   Prob (F-statistic):               0.00
Time:                        16:40:00   Log-Likelihood:            -2.3070e+06
No. Observations:              361000   AIC:                         4.614e+06
Df Residuals:                  360996   BIC:                         4.614e+06
Df Model:                           3
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
const                       -34.5742      3.251    -10.634      0.000     -40.947     -28.202
Review Scores Rating          0.5834      0.042     13.982      0.000       0.502       0.665
Review Scores Cleanliness     0.0133      0.324      0.041      0.967      -0.621       0.648
Review Scores Location       12.0860      0.335     36.128      0.000      11.430      12.742
==============================================================================
Omnibus:                   227680.671   Durbin-Watson:                   0.948
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          2193573.632
Skew:                           3.020   Prob(JB):                         0.00
Kurtosis:                      13.458   Cond. No.                     1.28e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.28e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

WashU Olin Business School

Machine Learning -- Price & Ratings 2

- **Split the data:**
  - 70% training data
  - 30% testing data
- **Mean Squared Error:** 75000.15
- **Rsquared Error:** 34.73%
- **Reasons:**
  - Insufficient or Irrelevant Features
  - Non-linear Relationships
  - High Variance in the Target Variable

WashU Olin
Business School

Part 3: Conclusion

WashU Olin Business School

Conclusion

- Attach importance to phone communication, email communication, and reviews
- Providing TVs for customers to watch.
- Hosts should respond to customers within a few hours
- Apartment owners in big cities can consider entering the Airbnb market
- Accommodates, Bathrooms, Bedrooms, Beds, Review Scores Rating, and Review Scores Location are all statistically significant predictors of Price
- The linear regression models are not good machine learning method

WashU Olin
Business School