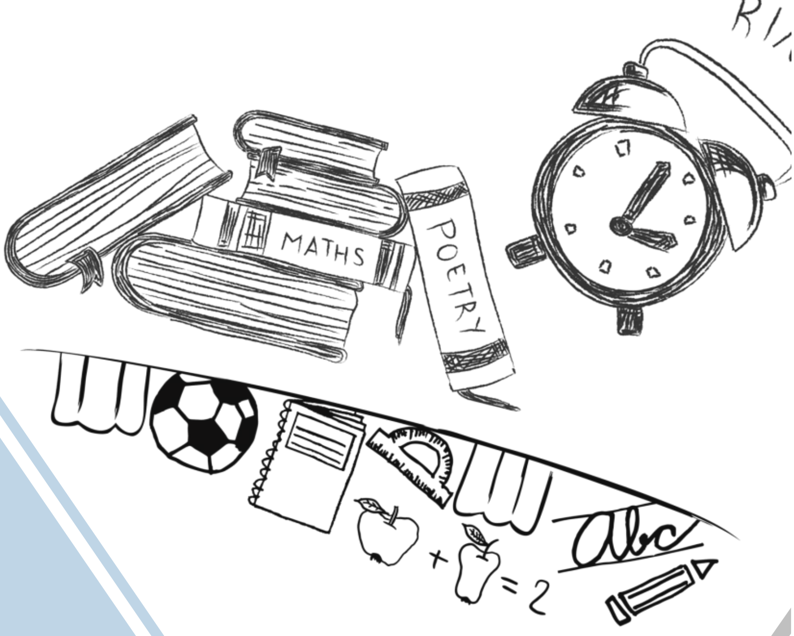
Airbnb's Demand in NYC:
An Econometric Analysis on an Airbnb's popularity



Presented by Qianshu Ni, Liangjiayi Wang, Haoyue Tan

Introduction

BIG IDEA: How can we boost the Airbnb popularity in NYC and optimize the potential tourist?

Motivation:

- Boost the Airbnb reviews helps to attract more potential consumer
- Provide pre-insights for travelers
- Measure the customers' and providers' behavior and performance on the platform

Possible considerations:

- The popularity of Airbnb in the NYC is not consistent with other cities
- We cannot predict the case during the COVID-19 time

Research Question

This research aims to find the factors that will influence the prosperity of house sharing economics in NYC, especially for the price of Airbnbs in the year of 2019.

Why asking:

- Provide information for Airbnb hosts to maximize their surplus and help the government to better track and manage tourism resources.

Why Airbnbs? Why the year of 2019?

- The Rise of "Sharing Economics"
- Emit the impact of the pandemic.

Outline

1. Literature Review: Literature Review: Inspiration from Sung and Lee’s work in 2018, they introduced a new economic model called “sharing economy” and aims to encourage people to make extra profit by sharing their unused properties with another person.

2. Data Resources：

- New York City Airbnb Open Data in Kaggle for 2019 and each Airbnb's specific demographics like room type, reviews per month, price (in USD), borough group, etc.
- RentHop New York Two Bedroom Median Rent Affordability from the NY Curbed website

3. Descriptive statistics: Basic summary for dependent variable and covariates.

4. Methodology: Use two stage least square method with instrumental variable ‘income per rent’

Literature Review

Interested in Sharing economics (Airbnb market)

- Sung, E., Kim, H., & Lee, D. (2018). Why Do People Consume and Provide Sharing Economy Accommodation? —A Sustainability Perspective

- The word “sharing” indicates the activities that people maximize the utilization of unused by sharing the good or services with others.

Existent Research

- Voltes-Dorta, A., & Sánchez-Medina, A. (2020). Drivers of Airbnb prices according to property/room type, season and location: A regression approach. Journal of Hospitality and Tourism Management
- Dudás, G., Vida, G., Kovalcsik, T., & Boros, L. (2017). A socio-economic analysis of Airbnb in New York City

- Aim to build the demand model that can be used to estimate a reasonable price based on the quantity demand of Airbnb in New York City. Researchers indicates that price of Airbnb is related with its location, neighborhood group and room type.

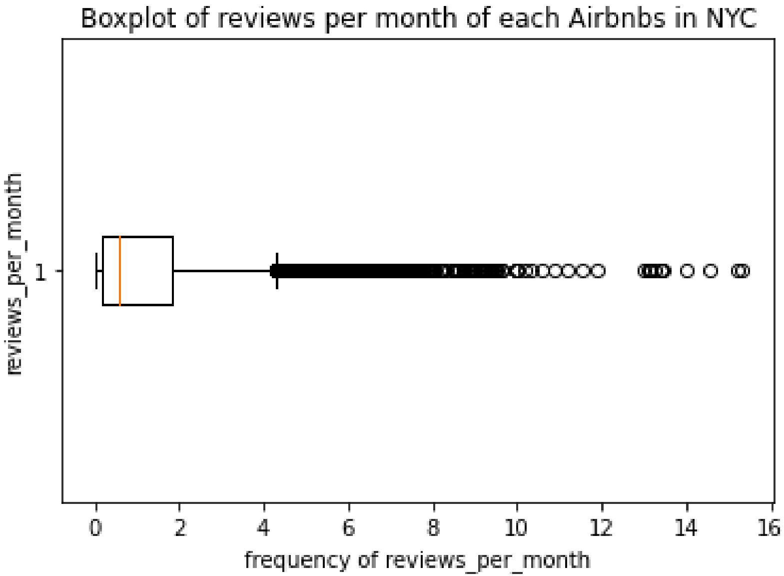
Preliminary analysis-Data cleaning

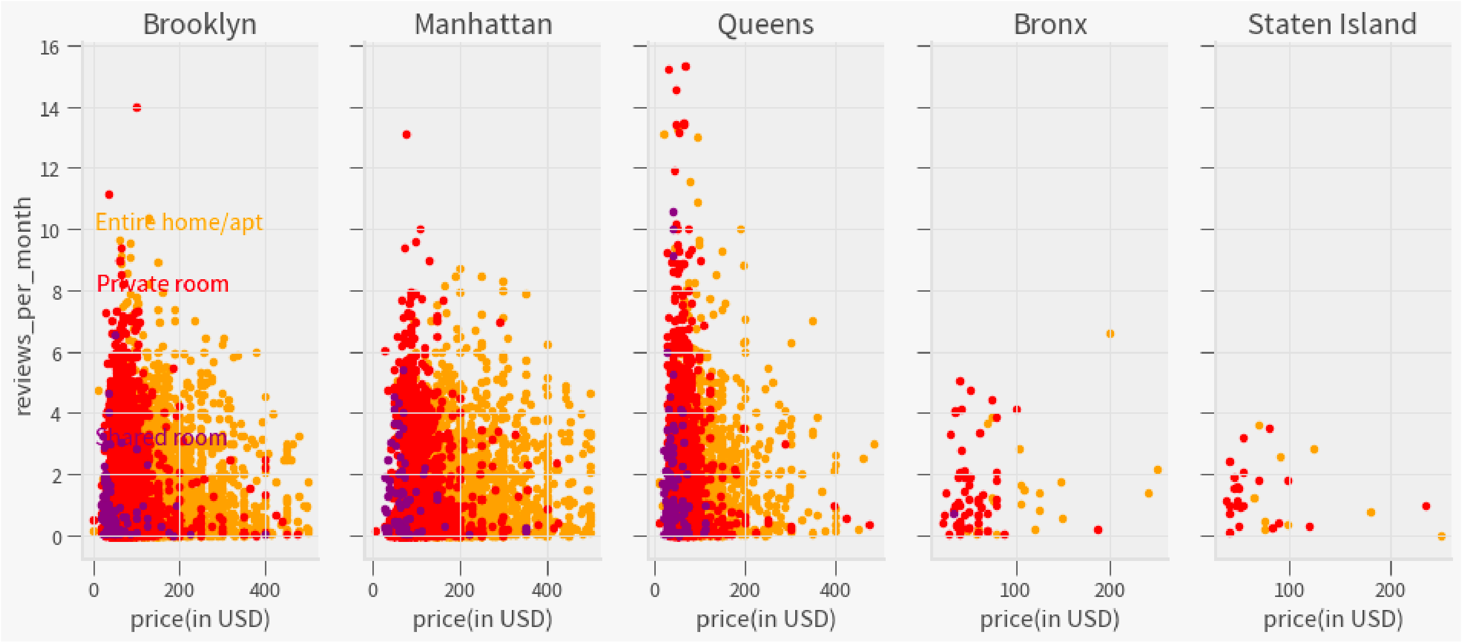
- price <= 500
- Merge with the table *“RentHop New York Two Bedroom Median Rent Affordability”*

```
RangeIndex: 14811 entries, 0 to 14810
Data columns (total 19 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              14811 non-null  int64
 1   name                            14810 non-null  object
 2   host_id                         14811 non-null  int64
 3   host_name                       14807 non-null  object
 4   neighbourhood_group             14811 non-null  object
 5   neighbourhood                   14811 non-null  object
 6   latitude                        14811 non-null  float64
 7   longitude                       14811 non-null  float64
 8   room_type                       14811 non-null  object
 9   price(in USD)                   14811 non-null  int64
 10  minimum_nights                  14811 non-null  int64
 11  advertisement                   14811 non-null  int64
 12  last_review                     14811 non-null  object
 13  reviews_per_month               14811 non-null  float64
 14  calculated_host_listings_count  14811 non-null  int64
 15  availability_365                14811 non-null  int64
 16  Coordinates                     14811 non-null  object
 17  income_per_rent                 14811 non-null  float64
 18  per_rent_borough                14811 non-null  float64
dtypes: float64(5), int64(7), object(7)
```

Descriptive Statistics - Dependent Variable

count	14811
mean	1.25
std	1.56
min	0.01
25%	0.16
50%	0.59
75%	1.82
max	15.32





Brooklyn
Manhattan
Queens
Bronx
Staten Island
reviews_per_month
price(in USD)
Entire home/apt
Private room
Shared room
0
2
4
6
8
10
12
14
16
0
100
200
400

Methodology - OLS

Why use OLS

- Building the OLS model with Airbnb's monthly review frequency and price and then gradually add other factors to view the changing coefficient significance.
- Whether all variables are significant and whether there are some potential problems such as endogenous
- We selected the model with the highest R-square

Controls include:

review_per_month: Average daily review in each month in 2019
price: Price in US dollarneighourhood_group: location in NYC
room_type: listing space type
advertisement: number of reviews cause by web-recommendation
availability_365: number of days when listing is available for booking
minimum_nights: amount of nights minimum for booking

OLS Regression Model

$$\text{Model1: } \widehat{\text{reviews\_per\_month}}_i = \beta_0 + \beta_1 \times \text{price(in USD)}_i$$

$$\text{Model2: } \widehat{\text{reviews\_per\_month}}_i = \beta_0 + \beta_1 \times \text{price(in USD)}_i + \beta_{2j} \times \text{neighbourhood\_group}_{ij}$$

$$\text{Model3: } \widehat{\text{reviews\_per\_month}}_i = \beta_0 + \beta_1 \times \text{price(in USD)}_i + \beta_{2j} \times \text{neighbourhood\_group}_{ij} + \beta_{3j} \times \text{room\_type}_{ij}$$

$$\begin{aligned}\text{Model4: } \widehat{\text{reviews\_per\_month}}_i &= \beta_0 + \beta_1 \times \text{price(in USD)}_i + \beta_{2j} \times \text{neighbourhood\_group}_{ij} + \beta_{3j} \times \text{room\_type}_{ij} \\ &+ \beta_4 \times \text{advertisement}_i\end{aligned}$$

$$\begin{aligned}\text{Model5: } \widehat{\text{reviews\_per\_month}}_i &= \beta_0 + \beta_1 \times \text{price(in USD)}_i + \beta_{2j} \times \text{neighbourhood\_group}_{ij} + \beta_{3j} \times \text{room\_type}_{ij} + \\ &+ \beta_4 \times \text{advertisement}_i + \beta_5 \times \text{availability\_365}_i + \beta_6 \times \text{minimum\_nights}_i\end{aligned}$$

OLS Regression Model Comparison

OLS Regressions Summaries for Model 1-5

	Model1	Model2	Model3	Model4	Model5
R-squared	0.003	0.031	0.032	0.339	0.350
	0.003	0.031	0.032	0.339	0.350
Intercept	1.375***	1.467***	1.355***	0.823***	0.809***
	(0.024)	(0.172)	(0.174)	(0.144)	(0.144)
price	-0.001***	-0.000	0.000*	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
C(neighbourhood_group)[T.Brooklyn]		-0.333*	-0.324*	-0.281**	-0.213
		(0.173)	(0.172)	(0.143)	(0.142)
C(neighbourhood_group)[T.Manhattan]		-0.386**	-0.379**	-0.335**	-0.240*
		(0.173)	(0.173)	(0.143)	(0.142)
C(neighbourhood_group)[T.Queens]		0.345**	0.353**	0.308**	0.314**
		(0.174)	(0.174)	(0.144)	(0.142)
C(neighbourhood_group)[T.Staten Island]		-0.145	-0.145	-0.187	-0.229
		(0.318)	(0.318)	(0.263)	(0.261)
C(room_type)[T.Private room]			0.116***	0.089***	0.037
			(0.032)	(0.026)	(0.026)
C(room_type)[T.Shared room]			0.001	0.114	-0.001
			(0.096)	(0.079)	(0.079)
advertisement				0.018***	0.017***
				(0.000)	(0.000)
availability_365					0.001***
					(0.000)
minimum_nights					-0.009***
					(0.001)
R-squared	0.00	0.03	0.03	0.34	0.35
No. observations	14811	14811	14811	14811	14811

Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

Methodology-2SLS

Why choose 2SLS:

- Endogenous between independent variable and dependent variable

Controls include:

- Income_per_rent: the instrument variable represent the percent of income spend and income required in a neighborhood
- Almost the same with OLS, since these variables are informative for the estimation
- Except for Location which cause is collinear with the instrument variable

Endogenity of price ~ reviews_per_month

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                  0.003
Method:                 Least Squares   F-statistic:                     38.66
Date:                Sun, 04 Apr 2021   Prob (F-statistic):           5.19e-10
Time:                        23:20:25   Log-Likelihood:                -85964.
No. Observations:               14811   AIC:                         1.719e+05
Df Residuals:                   14809   BIC:                         1.719e+05
Df Model:                           1
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept           133.6183      0.843    158.562      0.000     131.966     135.270
reviews_per_month    -2.6190      0.421     -6.218      0.000      -3.445      -1.793
==============================================================================
Omnibus:                     3735.111   Durbin-Watson:                   1.943
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8616.569
Skew:                           1.427   Prob(JB):                         0.00
Kurtosis:                       5.411   Cond. No.                         2.84
==============================================================================
```

Instrument Variable

- It shows the percent of income spent and income required in neighborhoods with sufficient data

RentHop New York Two Bedroom Median Rent Affordability

Search:

Neighborhood	Borough	Two-Bedroom Median Rent	Median Household Income	Income % for Median Rent	Income Required to Lease (40X Rule)
Queensbridge-Ravenswood-Long Island City	Queens	$3,300.00	$28,378	139.54%	$132,000
Williamsburg	Brooklyn	$2,499.00	$21,502	139.47%	$99,960
Lower East Side	Manhattan	$3,495.00	$31,273	134.11%	$139,800
Mott Haven-Port Morris	Bronx	$2,200.00	$20,334	129.83%	$88,000
East Harlem North	Manhattan	$2,495.00	$26,099	114.72%	$99,800

Instrument Variable- Validation

- Z has casual impact on X

- Z independent with Y

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  price   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     10.98
Date:                Sun, 04 Apr 2021   Prob (F-statistic):           0.000923
Time:                        23:18:57   Log-Likelihood:                -85978.
No. Observations:               14811   AIC:                         1.720e+05
Df Residuals:                   14809   BIC:                         1.720e+05
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
Intercept         125.7441      1.540     81.629      0.000     122.725     128.764
income_per_rent     0.0584      0.018      3.314      0.001       0.024       0.093
===================================================================================
Omnibus:                     3708.375   Durbin-Watson:                   1.940
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             8476.456
Skew:                           1.422   Prob(JB):                         0.00
Kurtosis:                       5.376   Cond. No.                         204.
==============================================================================
```

IV limitation - Collinearity with location factor

```
                          IV-2SLS Estimation Summary
==============================================================================
Dep. Variable:      reviews_per_month   R-squared:                      0.3413
Estimator:                    IV-2SLS   Adj. R-squared:                 0.3409
No. Observations:               14811   F-statistic:                 1.487e+04
Date:                Thu, Apr 08 2021   P-value (F-stat)                0.0000
Time:                        19:40:07   Distribution:                 chi2(11)
Cov. Estimator:                robust

                                   Parameter Estimates
=========================================================================================================
                                       Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
---------------------------------------------------------------------------------------------------------
C(neighbourhood_group)[Bronx]             0.5117     0.2840     1.8017     0.0716     -0.0449      1.0683
C(neighbourhood_group)[Brooklyn]          0.2060     0.3140     0.6560     0.5118     -0.4095      0.8215
C(neighbourhood_group)[Manhattan]         0.1088     0.3692     0.2948     0.7682     -0.6148      0.8325
C(neighbourhood_group)[Queens]            0.7981     0.2625     3.0400     0.0024      0.2835      1.3126
C(neighbourhood_group)[Staten Island]     0.2766     0.2701     1.0240     0.3058     -0.2528      0.8060
C(room_type)[T.Private room]              0.2476     0.1700     1.4564     0.1453     -0.0856      0.5807
C(room_type)[T.Shared room]               0.2765     0.2377     1.1631     0.2448     -0.1894      0.7424
advertisement                             0.0175     0.0004     46.367     0.0000      0.0168      0.0183
availability_365                          0.0007     0.0002     4.1899     0.0000      0.0004      0.0011
minimum_nights                           -0.0085     0.0027    -3.0918     0.0020     -0.0139     -0.0031
price                                     0.0025     0.0019     1.2831     0.1995     -0.0013      0.0062
=========================================================================================================

Endogenous: price
Instruments: income_per_rent
Robust Covariance (Heteroskedastic)
Debiased: False
```

Finalize regression

```
                         IV-2SLS Estimation Summary
==============================================================================
Dep. Variable:      reviews_per_month   R-squared:                      0.2233
Estimator:                    IV-2SLS   Adj. R-squared:                 0.2230
No. Observations:               14811   F-statistic:                 1.299e+04
Date:                Thu, Apr 08 2021   P-value (F-stat)                0.0000
Time:                        19:40:15   Distribution:                  chi2(7)
Cov. Estimator:                robust

                                Parameter Estimates
===============================================================================================
                              Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
-----------------------------------------------------------------------------------------------
C(room_type)[Entire home/apt]    2.1653     0.3895     5.5597     0.0000      1.4019      2.9286
C(room_type)[Private room]       1.4223     0.1769     8.0384     0.0000      1.0755      1.7690
C(room_type)[Shared room]        1.1696     0.1367     8.5578     0.0000      0.9017      1.4375
advertisement                    0.0169     0.0004     43.571     0.0000      0.0162      0.0177
availability_365                 0.0017     0.0002     10.704     0.0000      0.0014      0.0020
minimum_nights                  -0.0119     0.0037    -3.2000     0.0014     -0.0193     -0.0046
price                           -0.0088     0.0022    -3.9595     0.0001     -0.0132     -0.0045
===============================================================================================

Endogenous: price
Instruments: income_per_rent
Robust Covariance (Heteroskedastic)
Debiased: False
```

Considerations in future

- Due to the limited information provided by dataset, we only measure 5 factors that will influence the dependent variable which makes the R^2 small

Possible solution:
1. Find more information from online resources like: ‘whether the Airbnb is close to a tourist site’ and merge them into our dataset; then refit the model
2. We also have column ‘name’ , if possible, we can do the text analysis; find out top used listings' name words and see how entitle the Airbnb influence the reviews.

- Our analysis is limited in NYC, which cannot represent the low-tier cities’ Airbnb situation

- Due to the local epidemic prevention policy, we cannot get the sample data for Airbnb 2020 which we fail to use the hedonic regression to measure the effect of time period.

Thanks for Listening