Analysis of socio-economic variables and their impact on the Airbnb listings.

DS4A / Colombia

Team 7

Charles David Chavez

Diego Niño

Federico García Marín

Leonardo Vélez

Santiago Castellanos

Jorge Hernán Cárdenas

December 2020

Contents

Abstract: The sharing economy has produced revolutionary changes in the home rental market with the introduction of accommodation services like Airbnb. This phenomenon is attracting tourists who contribute to the sustainability of local commerce and the economic development of the city. Under this scenario, this research seeks to enrich the current debate on the variety of factors that influence the behavior dynamics of available Airbnb properties. To do this, an analysis was carried out to understand the relationship between the economic situation of the regions where Airbnbs are located and the number of properties available in a given period; considering variables such as GDP, personal income and the unemployment rate of 12 cities / states of the United States. A multivariate analysis technique is used to estimate a property availability model that adopts ordinary least squares for its adjustment. The findings on the impact of the regional economy on the number of properties listed are conclusive. The number of accommodations increases 0.55% when personal income increases 1% per quarter; However, for every 1% that GDP increases per quarter, the number of accommodations decreases 0.43% in the same period of time.

Keywords: Airbnb; GDP; personal income; unemployment rate; linear regression; sharing economy; COVID-19; listings available, adjust p-value; correlation matrix.

Introduction

Today's social and economic issues demand us to be able to understand how particular variables affect some business models, such as Airbnb and its occupancy level.

It is not just a matter of reacting to economic variables but using them to improve the performance of businesses. Such information could support a decision to deploy resources properly in locations that have a higher opportunity of growth, as could be the case of Airbnb and other companies in the same industry of shared economy.

The hypothesis that are subject of this research are focused on testing how an economic variable as Personal Income or Unemployment Rate have some significant effect on the number of listed properties in Airbnb. The added effect of those variables is studied by using exhaustive Exploratory Data Analysis and a regressive model to test the mentioned hypothesis.

There are benefits of directing this type or researches, one of them is giving Airbnb or companies with similar business models, an understanding on how economic variables can impact the number of listed properties. This finally translates into a framework they could use to prioritize investments and allocate resources in states or regions where economic variables favor the increase of listings.

This sort of framework can be initially obtained from this work but further improved by developing a number of new studies that could go deeper into the design of economic profiles across multiple regions or countries. Such regional economic profiles could be designed in a fashion that they are useful enough not just for Airbnb but for any other type of business or industry around the globe.

1. Hypothesis

The research intends to find if there is any impact or meaningful relation between unemployment levels, personal income levels and the listing of properties in Airbnb.

Ho: unemployment rate levels has no effect on the number of listed properties in Airbnb.

Ho: personal income level has no effect on the number of listed properties in Airbnb.

2. Literature Review

Airbnb is a community marketplace used to publish, advertise, and book accommodation in more than 190 countries via the internet or from your smartphone. It is based on the "Bed and Breakfast" model (where the "bnb" comes from). It is one of the most successful systems of the collaborative economy, allowing the user to find accommodation, with the difference that it will not be in a hotel but in the house of a person who may even be living in it. The interesting thing is that you can rent from common apartments to tree houses, igloos, geo domes, mills, etc.

Several investigations have been carried out on the variables that determine Airbnb's listing price, such as Chen and Xie (2017) and Gibbs et al. (2017), which concludes that the variables controlled by the host are price drivers. There is also the research by Brandon (2020), whose results showed that variables within the control of the host seem to have a greater impact on price than variables outside the control of the host, but that when these are combined, they are more accurate in predicting the price of a listing.

Other studies have attempted to identify key factors that affect commercial prices for variables not controlled by the host, such as proximity to some locations. Zhang et al. (2017) use general linear models (GLM) and geographically weighted regression (GWR) to analyze more than 794 price listings in Metro Nashville, Tennessee, noting that Airbnb listing prices are more sensitive to distance from the convention center in the central area than in other areas.

Perez et al. (2018) using a multivariate analysis technique to estimate a hedonic price model, conclude that accommodation prices progressively increase by 1.3% per kilometer from the tourist area, which in the four cases is located in the historic center of the city. However, at the same time, accommodation prices gradually decrease as distance from the coast increases.

Investigations like those of Dudás et al. (2017), analyzes the socioeconomic factors that influence the spatiality of Airbnb in New York, performing a statistical analysis (correlation, regression analysis) to determine the socioeconomic conditions of the areas and reveal those factors that may affect the spatial distribution of Airbnb listings in the city. The results highlight that (1) Airbnb's offering is concentrated in those parts of New York City with a young population, (2) there are a significant number of housing units, and (3) the number of points of interest is high.

This study proposes a new research framework, by analyzing the relationship between the variables that cannot be controlled by the host and that are associated at the macro level with the economic dynamics of the region, and the number of properties that are available per month in some states of the United States. The results of this study can provide implications for stakeholders, such as Airbnb hosts, to better understand the market situation and formulate an appropriate pricing strategy according to the economic situation of the state, in an unnormal situation such as that experienced with COVID-19, which seriously affected the world economy.

3. Data & Data Wrangling

3.1. Methodology

The goal is to assess how strong is the relationship - if there is any- between the personal income, economics of the states where the Airbnb properties are located, and the number of the listings, as well as prices and other significant parameters in the same area.

Methodology is focused on gathering and preparing the data in the first part of this research and, in a second stage, using a hypothesis test method that could provide information to prove or disprove the stated hypothesis.

3.2. Preliminary Exploration

The most meaningful parameters that could be extracted from the exploration of the data are explained in this section, using as the starting point the data provided by Correlation One as part of the "Lodging and rental industry" case.

3.2.1. Listings Dataset:

There is information regarding scores, amenities and characteristics of each Airbnb listing which could be relevant for the case and hypothesis testing. Also, there is data for the states and localities where Airbnbs are places as well as prices tracked through the period between 2016 and 2018.

Now, for the sake of testing the hypothesis, it was a must for the time frame for listings and economics variables to be the same. Unfortunately, it was found that the temporality of listings and economic data did not match, hence the decision of gathering data from Airbnb site which provided a comprehensive list of properties, including data through second quarter of 2020.

After careful analysis of this particular dataset, some columns can be removed keeping the features in the Table 1.

	id	quarter	property_type	state	price	accommodates	bathrooms	bedrooms	beds	availability_30	availability_60	availability_90	availability_365
0	100002	2Q-2018	Apartment	NY	$67.00	2.0	1.0	1.0	1.0	11	17	47	296
1	100002	2Q-2018	Apartment	NY	$67.00	2.0	1.0	1.0	1.0	3	25	43	304
2	100002	2Q-2018	Apartment	NY	$67.00	2.0	1.0	1.0	1.0	6	15	21	279
3	100002	3Q-2018	Apartment	NY	$67.00	2.0	1.0	1.0	1.0	3	10	24	299

Table 1 Listing table

3.2.2. Econ_state Dataset:

GDP, Personal Income, unemployment rate are variables provided within the dataset. These variables are of high importance for the research as they are the ones to be tested against the number of listings, to bring light about how those economic variables could induce a change in the number of Airbnbs.

After careful analysis and, in order to test the hypothesis, it is required to find complementary data from external sources that could have a temporal match with the listings time frame established from 2018 through 2020. Although the provided data was extensive in number of periods for each variable GDP, Unemployment and Personal Income, they did not temporally match between them.

The complementary economic data as the unemployment rate was obtained from *U.S. Bureau of Labor Statistics*[1]*;* GDP and Personal Income was obtained from the *U.S. Bureau of Economic Analysis (BEA)*.[2]

[1] https://beta.bls.gov/dataViewer/view/a053dd0d4632464bbdb51b05dc295ec2
https://beta.bls.gov/dataViewer/view/eda813873cac4da4be800035a2b9e355
https://beta.bls.gov/dataQuery/find?q=unemployment+rate+

[2] https://apps.bea.gov/regional/downloadzip.cfm

3.3. Data Preparation

Official data did not necessarily fit the requirements in terms of columns disposition or the time scale, which in most cases was provided quarterly or yearly, with the exception of unemployment rate, which was presented on a monthly basis.

One of the first tasks to perform was setting economic data frames in a proper format that could be easily merged with the listing's information. The keys that could be used to join all tables are '*state*' and '*date*' due to the commonality of those columns between all dataframes. Even so, it was required a preprocessing on the '*date*' as GDP and Personal Income dataframes had a quarterly frequency while unemployment rate had monthly frequency.

4. Exploratory Data Analysis

As part of the plan to work the hypothesis out, there must be an exploratory analysis with the expected outcome of getting insights about how the variables affect each other with the data at hand.

Initially, data frames are analyzed independently and then merged together to extract information from the interactions between the Airbnb variables- as prices and numbers of listings- and the economic ones.

4.1. Economic data Exploration

4.1.1. Unemployment Rate

In general, this data frame needed to be analyzed to ensure that the data for every month and every state was complete and that there were no missing values.

In the Table 2, can be seen the format for the final dataframe that is used to explore how unemployment rate behaves by its own and, later, with Airbnb variables.

	unemployment_rate	state	date
232	5.7	CA	2016-01-01
233	5.6	CA	2016-02-01
234	5.6	CA	2016-03-01
235	5.5	CA	2016-04-01
236	5.5	CA	2016-05-01
...	...	...	...
2547	8.4	TX	2020-06-01
2548	8.0	TX	2020-07-01
2549	6.8	TX	2020-08-01
2550	8.3	TX	2020-09-01
2551	6.9	TX	2020-10-01

Table 2 Unemployment rate header.

Some of the most relevant findings regarding unemployment rate, can be deducted from the graphics below.

It has been found that there are no considerable changes in unemployment rate between most of the states being part of the study, but there are particular cases as DC (District of Columbia), CO (Colorado) and VT (Vermont), which have the most extreme values of unemployment rate.

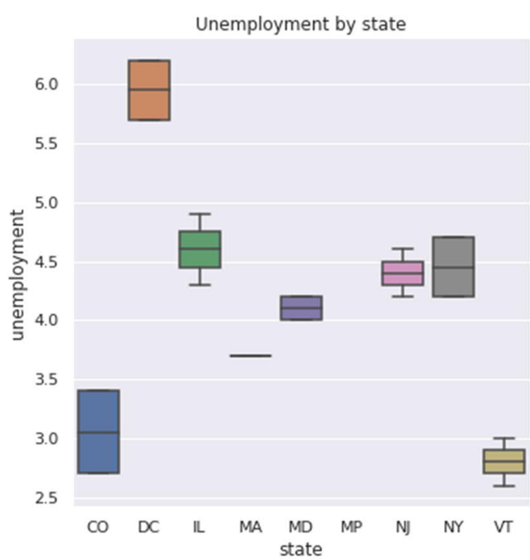


Figure 1 Unemployment rate by states box plot.

Also, after plotting the seasonal behavior for each state's unemployment rate, could be found some commonalities between states, like a smooth behavior through time and even some trends of lower rates unemployment through time. Interestingly enough is the case for 2020 data, were we find a shoot in the graph due to the **COVID-19** event.

There is also a clear stational behavior that could be of use for the sake of proving or disproving the hypothesis, which lead to consider that the data should be treated in a **time series manner**.

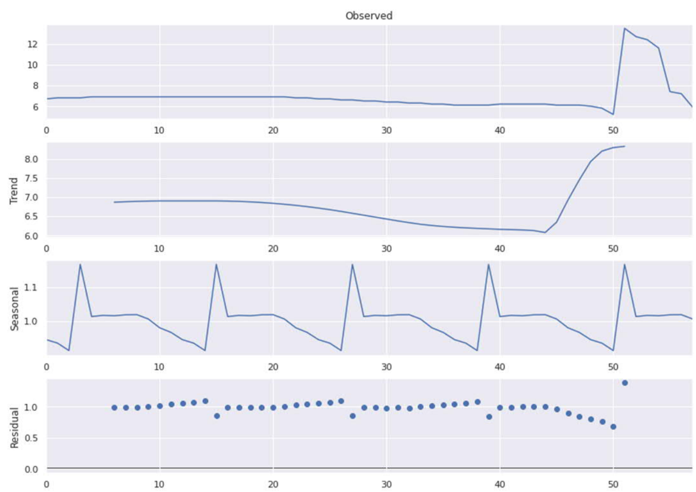


Figure 1 Seasonal Unemployment behavior for state of Alaska

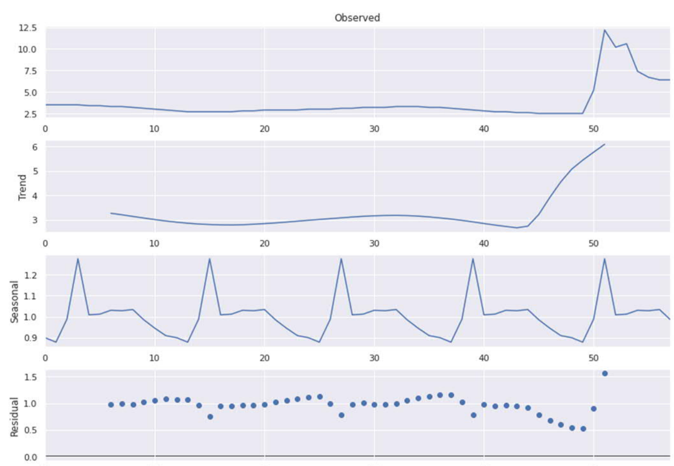


Figure 2 Seasonal Unemployment behavior for state of Colorado

According to the information derived from these graphs, the information for the unemployment rate for the last two quarters (2020), should be dropped from the dataframe as they could affect the analysis and the results of the hypothesis testing for being atypical data in the Covid 19 scenario.

4.1.2. Gross Domestic Product

In Table 3 is presented the GDP dataframe which was prepared in the same way as the unemployment rate so they could be easily merged in later steps for the data processing.

	gdp	state	date
736	1710394.1	CA	2005-01-01
737	1710394.1	CA	2005-02-01
738	1710394.1	CA	2005-03-01
739	1733114.0	CA	2005-04-01
740	1733114.0	CA	2005-05-01
...	...	...	...
8091	1861581.9	TX	2019-12-01
8092	1818394.5	TX	2020-01-01
8093	1818394.5	TX	2020-02-01
8094	1818394.5	TX	2020-03-01
8095	1628185.0	TX	2020-04-01

2208 rows × 3 columns

Table 3 GDP Header.

Regarding GDP and its behavior, it can be seen a growing trend through time for almost every state, although there are meaningful differences in the GDP value between states. Something to notice is the effect of COVID-19 pandemic on the last two quarters registered on the data.

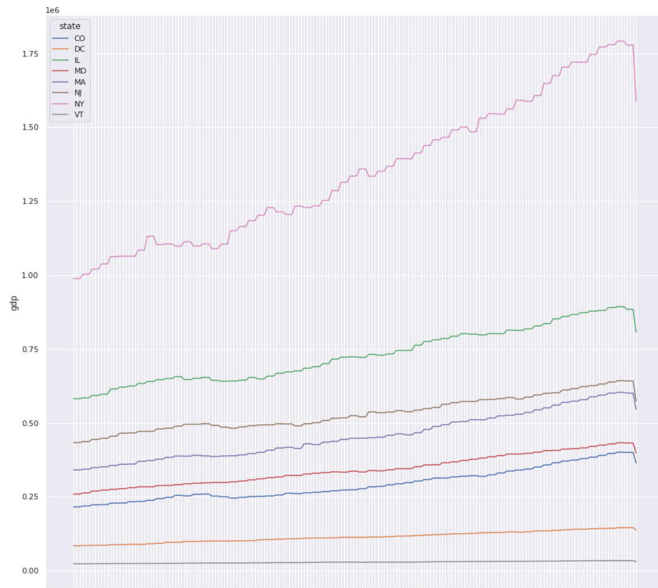


Figure 2 GDP for each state across time (2015-2020)

As can be seen in the Figure 2, the GDP behavior due to COVID-19 is odd and this could significantly affect the results, which means that those two last quarters should be removed for the final analysis.

All the prior information could be of use to relate the number of and the GDP value listings in a particular state, in case the research is performed by state.

Also, it is to be considered that the most significant feature, more than the GDP values, are the **variations or percentual change from one period to the next,** not just for the domestic product but also for the other economic variables.

One particular case worth studying is NY (New York) and CA (California), which have the highest GDP value between all states and the differences are high enough to be considered. Though it may be tempting to judge that specific GDP value as an outlier, it is not, as the research wants to consider the added effect of all selected states.

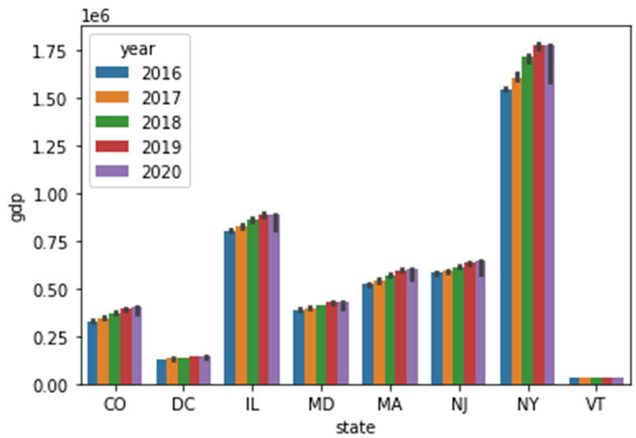


Figure 3 Yearly GDP for each state

4.1.3. Personal Income

Personal Income is defined as the after-tax income of one person and non-profit corporations, which includes payments to individuals (wages, salaries, and other income).

This specific variable should be of interest in the regard of how the number of listings is connected to the changes on the population and individual's income. Similarly, as previous dataframes, this one was prepared in a way that could be merged with the others by using *'date'* and *'state'* as keys.

	PI	state	date
256	2126849.0	CA	2015-01-01
257	2126849.0	CA	2015-02-01
258	2126849.0	CA	2015-03-01
259	2163796.5	CA	2015-04-01
260	2163796.5	CA	2015-05-01
...	...	...	...
2811	1550113.6	TX	2019-12-01
2812	1559485.2	TX	2020-01-01
2813	1559485.2	TX	2020-02-01
2814	1559485.2	TX	2020-03-01
2815	1678239.6	TX	2020-04-01

Table 4 Personal Income Header

The Personal Income, as is presented in the Figure 4, do not show a particular or meaningful change or odd behavior between quarters when the information for all states is aggregated.

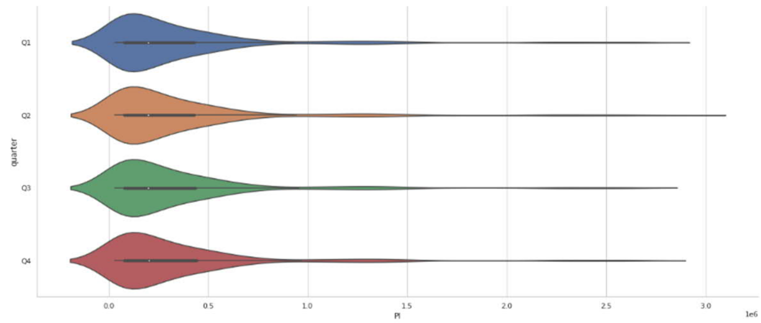


Figure 4 Personal Income by quarter across all states.

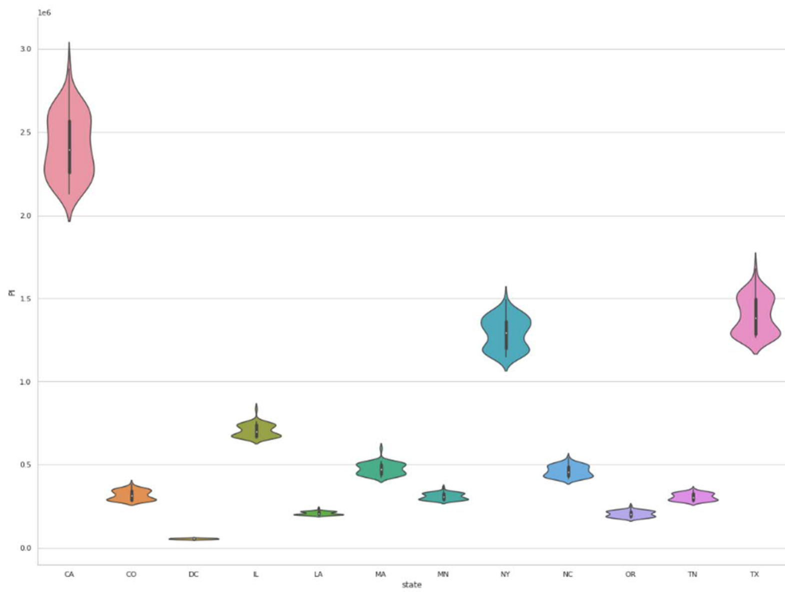


Figure 5 Personal Income by states.

Finally, states like California, New York and Texas, have a significant level of Personal Income which is relatively high when compared to the other states being studied in this case, as shown in the Figure 5. Although the previously mentioned states have a higher value, they could not be considered as outliers as the purpose of this study is to analyze the hypothesis in a global manner.

4.2. Listings Exploration

This data set is considered our main data source since it includes lot of information about the reviews, amenities, and what is called the *host-controlled variables* such as number of beds and numbers of guests.

The goal is to create a master table that stems from the joining of listings and other data sets such as GDP and Personal Income, that could provide the complementary data required to allow deeper insight.

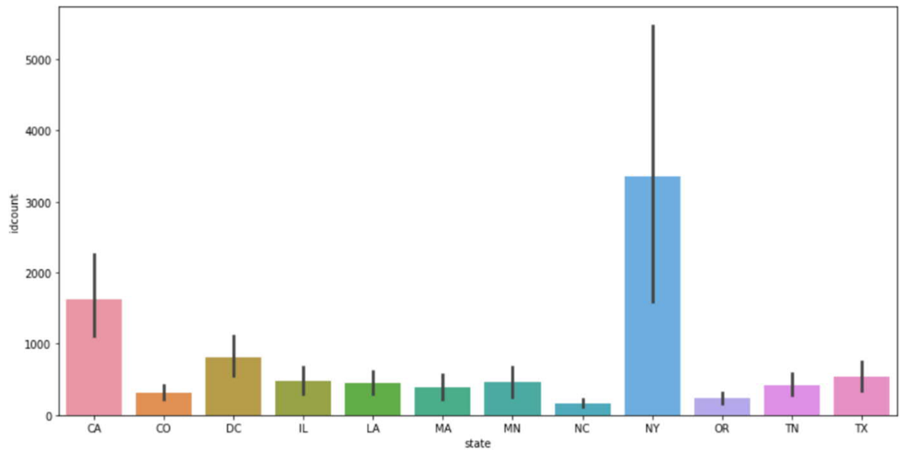


Figure 6 Number of listings by states through the entire time frame.

As can be seen in the Figure 6, the number of listed properties by state is not uniform, hence it is necessary to **weight all variables (price and availability) by the numbers of listings by state**.

Furthermore, when comparing the average number of listings per month against their prices and available days, a direct relationship between the price and the number of days available is clear in the Figure 7. That is, the more the price increases, the lower the number of days per month that the properties will be rented.

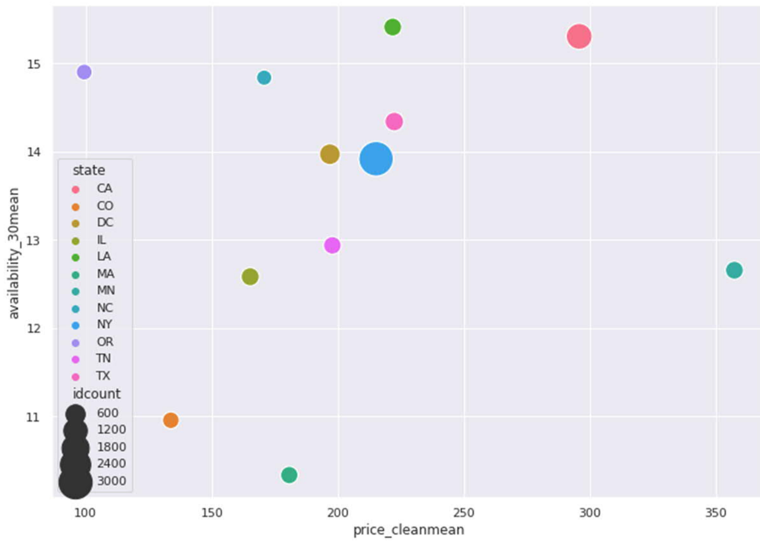


Figure 7 Number of listings by states through the entire time frame.

However, states such as Oregon and Minnesota seem to have different behaviors with respect to the other states. While in Oregon, a lower price causes less occupancy of days per month, in Minnesota the effect of a higher price causes a higher occupancy and therefore, less availability of days per month.

Comparing this information to the Personal Income charts, it is noticeable that the **two states with higher Personal Income are the ones with higher numbers of listings**.

Finally, the main transformation applied on listings data was grouping by state, quarter and property type with the goal of shaping the data frame to get information about property prices, time series and the number of listings by state. The other variables got aggregated in different manner each, for instance price was aggregated by mean and median, availabilities aggregated by mean and ‘id’ by ‘count’.

	state	quarter	property_type	price_cleanmedian	price_cleanmean	availability_30mean	availability_60mean	availability_90mean	availability_365mean	idcount
0	CA	1Q-2019	Aparthotel	140.0	147.069892	15.763441	39.876344	63.419355	194.849462	186
1	CA	1Q-2019	Apartment	103.0	136.355256	9.212825	22.588235	37.826760	142.725603	20740
2	CA	1Q-2019	Barn	173.0	173.000000	12.000000	29.666667	56.666667	327.666667	3
3	CA	1Q-2019	Bed and breakfast	100.0	111.071895	23.150327	49.254902	76.150327	306.568627	153
4	CA	1Q-2019	Boat	250.0	562.333333	16.703704	35.888889	55.037037	225.777778	27
...	...	...	...	...	...	...	...	...	...	...
›18	TX	4Q-2019	Tipi	19.0	62.666667	22.555556	50.222222	79.888889	319.555556	9

Table 5 Listing cleaned data frame.

For the next step, column types were properly ordered to make proper groupings and analysis by date and/or categories with the result of state and property type as categorical variables and quarter as time-period type.

Regarding the states that were used for the assessment, they were chosen by the criteria of being the ones available in Airbnb data and by having it complete for a period from 2018 through 2020, as well as their match with economic time frame.

LA: Lousiana
MN: Minnessota
CO: Colorado
TN: Tennessee
TX: Texas
NY: New York
OR: Oregon
IL: Illinois
DC: District of Columbia
CA: California
MA: Massachussets
NC: North Carolina

Table 6 List of states used for the research.

4.3. Merging of listings and economic data:

The data merging between ‘listings’, ‘GDP, ‘personal income and’, ‘unemployment’ was made by joining the dataframes by *‘quarter’* and *‘state’*, which yielded a new dataframe with all listings and their corresponding economic indexes for a given quarter. The previous steps of synchronizing the time periods between all data frames were of help to perform this joining.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2623 entries, 0 to 2622
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   state                 2623 non-null   object
 1   idcount               2623 non-null   int64
 2   month                 2623 non-null   int64
 3   year                  2623 non-null   int64
 4   quarter               2623 non-null   period[Q-DEC]
 5   property_type         2623 non-null   category
 6   price_cleanmedian     2623 non-null   float64
 7   price_cleanmean       2623 non-null   float64
 8   availability_30mean   2623 non-null   float64
 9   availability_60mean   2623 non-null   float64
 10  availability_90mean   2623 non-null   float64
 11  availability_365mean  2623 non-null   float64
 12  mean_ur               2623 non-null   float64
 13  std_ur                2623 non-null   float64
 14  mean_pi               2623 non-null   float64
 15  std_pi                2334 non-null   float64
 16  mean_gdp              2623 non-null   float64
 17  std_gdp               2334 non-null   float64
dtypes: category(1), float64(12), int64(3), object(1), period[Q-DEC](1)
memory usage: 373.0+ KB
```

Figure 8 Listings and economic data merged.

4.4. Feature Engineering:

4.4.1. Grouping and weighted values:

The goal of this section during the preparation of the data was to create a new data frame that could provide information about the number of listings by state and by quarter.

When doing the grouping by ‘*state’* and ‘*quarter’*, it was required to perform the calculation of the other values as price and availability in a way that would consider the number of listings for that specific state (weighting), that means calculating the weighted prices and availabilities, with the final result presented on the Table 7.

quarter	state	price_cleanmedian_W	availability_30mean_W	availability_60mean_W	availability_90mean_W	unemployment	income	gdp	count
2018Q2	CA	109.850066	10.613081	24.715921	24.715921	4.266667	2500871.9	2966250.1	51556
	CO	107.583909	10.002514	22.318668	22.318668	3.133333	332573.6	371053.9	1591
	DC	121.541098	6.825788	16.426102	16.426102	5.733333	56669.1	138232.6	10183
	IL	102.483753	10.618766	24.460817	24.460817	4.300000	724439.5	860142.0	5755
	LA	141.288080	12.162814	30.879488	30.879488	5.000000	214988.6	253237.1	6099
...	...	...	...	...	...	...	...	...	...
2020Q2	NC	103.533024	10.210506	25.307455	25.307455	11.066667	540565.9	546775.7	2589
	NY	104.245303	9.125906	19.019378	19.019378	15.133333	1493369.3	1587879.2	79575
	OR	90.125612	14.787751	32.213140	32.213140	13.600000	252884.6	233799.2	4490
	TN	169.389035	16.422667	33.664523	33.664523	12.033333	352529.0	333194.3	9485
	TX	164.832350	8.892711	19.705165	19.705165	11.633333	1678239.6	1628185.0	17793

108 rows × 8 columns

Table 7 Weighted prices and availabilities grouped by state and quarter.

4.4.2. Percentage change between time periods:

A percentage change transformation was applied to the numerical values of the data, due to the need of dealing with temporal trends in the data, and in consequence obtaining a more accurate linear regression.

quarter	state	price_cleanmedian_W	availability_30mean_W	availability_60mean_W	availability_90mean_W	unemployment	income	gdp	count
2018Q4	TN	0.011148	0.328618	0.344060	0.344060	-0.028571	0.008140	0.007199	0.062055
	TX	-0.018085	0.052018	0.086392	0.086392	-0.008929	0.011292	0.008479	0.064454
2019Q1	CA	0.000842	-0.064373	-0.056366	-0.056366	0.007874	0.014828	0.013120	0.002829
	CO	-0.006855	-0.106607	-0.108154	-0.108154	-0.051020	0.016586	0.012875	0.000163
	DC	-0.001190	-0.055315	-0.121172	-0.121172	0.017647	0.012678	0.011015	0.002483
	IL	-0.020066	0.068571	-0.014517	-0.014517	0.000000	0.006587	0.006149	0.008260
	LA	0.015717	-0.127263	-0.247925	-0.247925	-0.034965	0.001419	-0.002449	-0.001254
	MA	0.001686	-0.010585	-0.079401	-0.079401	-0.031579	0.018601	0.016913	0.008190
	MN	0.021967	0.106199	0.049680	0.049680	0.066667	0.004959	0.003468	-0.001071
	NC	0.007764	0.152526	0.037779	0.037779	0.033898	0.013006	0.008670	0.002005
	NY	-0.011995	0.500465	0.188721	0.188721	0.008403	0.018068	0.015070	0.004526
	OR	-0.013387	0.030167	-0.007951	-0.007951	0.000000	0.012561	0.013071	0.003382
	TN	-0.025166	-0.082191	-0.163311	-0.163311	0.009804	0.011957	0.007510	-0.010106
	TX	-0.001231	-0.119208	-0.144873	-0.144873	-0.027027	0.014841	0.001421	0.001450
2019Q2	CA	0.020470	-0.109428	-0.128668	-0.128668	-0.039063	0.010360	0.018276	0.000741

Table 8 Percentage change between quarters.

5. Analysis (Hypothesis Testing)

In order to analyse and test the hypothesis proposed at the beginning of this work, the selected method was the linear regression accompanied by the analysis of the correlation between the different variables, as can be seen in the Figure 9.

From the previously mentioned graph, can be inferred that the correlation between 'count' and 'income' is positive, which means that if income is increasing, the quantity of listings in that state tends to increase too. This is helpful in finding meaningful relationship for the hypothesis that states that *personal income level has no effect on the number of listed properties in Airbnb*.

For the second hypothesis about *unemployment rate levels has no effect on the number of listed properties in Airbnb*, seems to be true in the correlation matrix, as it shows a correlation of 0.065 with the quantity of listed properties (count column).

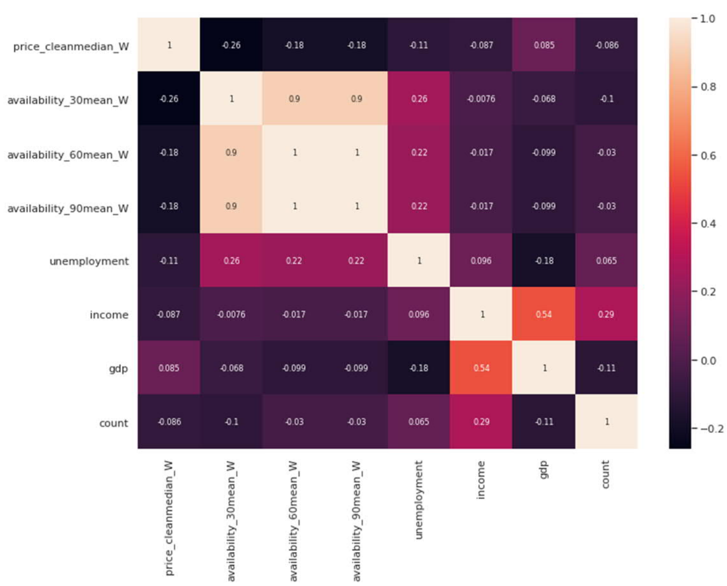


Figure 9 Correlation Matrix.

There is also a high correlation between GDP and Personal Income, due to the fact that, even though those variables are independent, it has been proven that they behave in a similar manner during periods of time.

The selected methodology used for testing the hypothesis was linear regression, in which the count of listings was compared with GDP, Unemployment Rate and Personal Income. Also, the analysis includes four performed tests for each of the Macro Economic variables.

The five tests performed are summarized in Table 9.

Count ~ GDP
Count ~ Unemployment
Count ~ Income
Count ~ GDP + Unemployment + Income

Table 9 List performed test for the regression model.

In order to properly evaluate the different results and their statistical significance, the *p-value* has to account for the number of tests performed by dividing the base value by the number of tests, which in this case yields an **adjusted p-value of 0.0125**.

A second consideration is the fact that the number of observations is 72, which was obtained after increasing the number of states being included in the data, originally being five and ended being 12 states.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  count   R-squared:                       0.184
Model:                            OLS   Adj. R-squared:                  0.148
Method:                 Least Squares   F-statistic:                     5.101
Date:                Mon, 14 Dec 2020   Prob (F-statistic):            0.00304
Time:                        02:18:11   Log-Likelihood:                -43.263
No. Observations:                  72   AIC:                             94.53
Df Residuals:                      68   BIC:                             103.6
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
                  coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       0.0551      0.134      0.411      0.682      -0.212       0.323
gdp           -43.3776     14.955     -2.901      0.005     -73.219     -13.536
unemployment   -0.7716      1.615     -0.478      0.634      -3.993       2.450
income         54.4809     14.510      3.755      0.000      25.526      83.435
==============================================================================
Omnibus:                       76.918   Durbin-Watson:                   1.632
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              674.083
Skew:                           3.212   Prob(JB):                    4.21e-147
Kurtosis:                      16.544   Cond. No.                         345.
==============================================================================
```

Figure 10 Linear Regression model.

Ho	unemployment rate level has no effect on the number of listed properties in Airbnb.
H1	Unemployment rate has effect on the number of listings in Airbnb.

After performing the linear regression, has been found the following evidence to keep the hypothesis:

- Though the coefficient related to unemployment shows an inverse relation to number of listings, the p-value (0.634) is higher than the expected p-value of reference what finally **supports keeping Ho**. There is no statistical significance to support the relation between unemployment and listings in Airbnb.

Ho	personal income level has no effect on the number of listed properties in Airbnb
H1	Personal Income level has effect on the number of listed properties in Airbnb.

As for the second hypothesis, the evidence found indicates that there is statistical significance for the coefficient relating Personal Income and the number of listings.

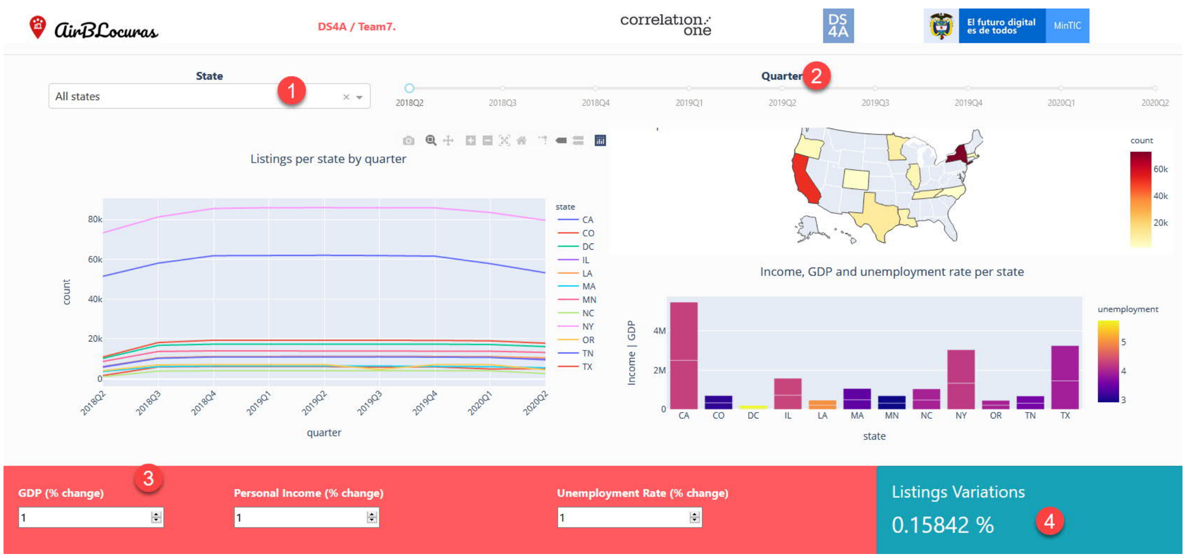
- There is positive relation between Personal income and the number of listings which supported by a p-value in the order of 10^{-4}, which is lower than the previously established 0.0125 value.
- This evidence supports the rejection of the **Ho** and provides enough ground to estate that there is a significant relation between Personal Income and the number of listed properties on Airbnb.

As a third finding worth to be mentioned is the statistically significant relation between GDP and the number of listing in Airbnb, but with a negative correlation, inversely as the positive reported by the Personal Income.

This a matter that requires further analysis as is well known that there is a relation between PI and GDP, they are both independent for the case of study and they behave differently with respect of the number of listings.

6. Interactive Tool: AirBLocuras

url: http://www.airblocuras.online/



In AirBLocuras you can analyze the Airbnb data and the Macro economic variables of the selected states (1), you can choose to see the information about a specific state or the entire state set, and a given Quarter of the year (2). At the bottom section, you can do a simulation of the listing variations (4) within the specified macroeconomic delta change in percentage (3).

7. Conclusions

At the beginning of this research, was considered that there were previous studies reporting significant evidence of relationship of prices and concentration of listed properties in Airbnb to a young population and the socio-economic factor of each locality. Most of them focused on prices related to the so called “host-controlled variables” such as number of beds, bathrooms and so forth.

This study puts on the table new information about how macro-economic (non-host-controlled) variables such as GDP, Personal Income and Unemployment Rate have impact on the rise or fall in the number of listed properties in Airbnb.

One of the most meaningful findings is that rise in personal income increments the properties listed in Airbnb in successive periods.

There could be explanations to that effect, for instance, as the people in a particular state has more available money (income), they could be interested on investing in new properties or remodeling the current ones with the goal of listing them in Airbnb.

A second finding is the correlation between the count of listings and GDP being negative, wich means that, in the states with a larger GDP the quantity of the properties listed in Airbnb decreases. This could be unexpected but important enough to be considered.

GDP is a number that includes not just income but also the company's incomes and government expending, which means that it informs about economic activity and growth of a state or nation.

This could be interpreted as a snapshot of the economic performance of a region and how it improves through time; the number of listings can be lower due to the people being less interested about getting into Airbnb business and being more prone to put their money into other types of business.

In particular, the issue of the relation between GDP and listings should be part of a future research that considers the components of GDP and which one of them are the actual factors having impact on listings as GDP stems from multiple variables.

Finally, regarding unemployment rate, this research has concluded that, based on the data for the time between 2018-Q2 and 2019-Q4, there is no significant relation between that variable and the number of listed properties.

One could expect to find a relation as the intuition suggests that people use Airbnb as a source of income when they are unemployed, but evidence suggests that this is not necessarily the case, furthermore, seems like people are being more interested about the business and investment side of listing a property in Airbnb.

Also, something to be considered in future studies is the assessment on how unemployment rate is measured; let us think about the case that an unemployed person, who is willing to list a property, never reported his or her employment status hence being not counted and part of the unemployment measurement.

As for the business side of this research, intending to **provide a framework for businesses like Airbnb to assess where to invest and where to place operations** that favor the growth in listings, the result indicates that they should put their effort to profiling regions or countries by their Personal Income and GDP, not just for the current value but for the variations through time. If they can track regions with successive periods of growth in personal income, they also could have a framework to prioritize where to invest in marketing strategies or establishing offices.

Glossary

GDP. Gross Domestic Product. Is the final value of the goods and services produced within the geographic boundaries of a country during a specified period of time, normally a year.

Personal Income. Is the amount of money collectively received by the inhabitants of a country. Sources of personal income include money earned from employment, dividends and distributions paid by investments, rents derived from property ownership, and profit sharing from businesses.

Unemployment Rate. Percentage of unemployed individuals in an economy among individuals currently in the labor force. It is calculated as Unemployed Individuals / Total Labor Force × 100 where unemployed individuals are those who are currently not working but are actively seeking work.

References

Brandon, Mc. (2020), "Price Prediction in the Sharing Economy: A Case Study with Airbnb data", *Honors Theses and Capstones. 504*, https://scholars.unh.edu/honors/504.

Chen, Y. and Xie, K. (2017), “Consumer valuation of Airbnb listings: a hedonic pricing approach”, International Journal of Contemporary Hospitality Management, Vol. 29 No. 9, pp. 2405-2424.

Dudás, G., Vida, G., Kovalcsik, T., & Boros, L., A socio-economic analysis of Airbnb in New York City. Regional Statistics, Vol. 7. No.1. 2017: 135–151; DOI: 10.15196/RS07108

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. and Goodwill, A. (2018), “Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings”, Journal of Travel & Tourism Marketing, pp. 1-11.

Zhang, Z., Chen, R., Han, L., & Yang, L. (2017). Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability*, *9*(9), 1635. https://doi.org/10.3390/su9091635

Perez, R., Serrano, L., Marti, P., & Mora, R. (2018). The What, Where, and Why of Airbnb Price Determinants. *Sustainability*, *10*(10), 4596; doi:10.3390/su10124596