

PROBLEM STATEMENT

Team 13

AirBnB is born in 2008 when it was founded by Brian Chesky, Nathan Blecharczyk and Joe Gebbia as a means to rent free space at home and generate additional income to meet month's endⁱ. Today it has grown to a company that will be going public in December 2020 with an IPO, for which some analysts forecast that the valuation of the company will go to \$30 billion. ⁱⁱ

Even though it has been successful, and it still is: a report mentioned in a recent article published in the New York Times in September, signals that home rentals have outperformed hotels in 27 global markets since the onset of Covid-19ⁱⁱⁱ it is not without complications: a murder in California on a rental property, charges of racism, its use for scams through fraudulent offers ^{iv} are some of their problems they have faced. Other issues are of concern too, accusations of triggering excess tourism in certain countries and transforming residential areas into tourist areas, has resulted in harsher regulation and in cities such as Paris, Barcelona, and Santa Monica, California short term rental has been restricted or even declared illegal^v. However, other cities such as Amsterdam, Berlin, London, San Francisco and New York have more permissive legislation and renting properties in them is an attractive business.

The AirBnB phenomenon has generated a new type of investment that consists of the purchase of properties to register on these platforms and take advantage of the profitability of short-term leases. In the research of Cocola-Gant and Gago ^{vi} in the Alfama neighborhood in the city of Lisbon they found that the short term rental instead of a sharing economy process, what was taking place was a "*buy-to-let investment in which*

different players make profits from rents". They also point : "Airbnb provides a set of advantages and benefits that makes the STR market increasingly attractive for both local and distant investors".

Why is that so?, there is a lot of money on this type of business. An article that appeared last January in Millionacres^{vii} states : *"Airbnb hosts earn over \$20,000 per year on average -- and that's just on two-bedroom properties. In cities like San Diego and Miami, you can even make upwards of \$28,000"*^{viii}

In the above-mentioned publication, the author states: *"Process-wise, buying property for Airbnb is pretty simple. You find a home, you make an offer, and you close on the property. The hard part? That's in finding the right property to begin with".*

We are going to work to solve the question raised on the previous paragraph. That is, the problem we are going to work will be:

<i>What are the best recommendations for an investor who has a budget and wants to buy a property to put it to rent in AirBnB?.</i>
--

The final product scenario is an investor who has a certain amount of dollars to invest and wants to know where he can put his money, buying a property to rent in AirBnB. We foresee that we will have to specify to city where the property will be located, the type of property to work with (condo, apartment, villa) and the details of the property (area, number of bedrooms, bathrooms, amenities...). On the other side we will have to forecast the income he will be getting by forecasting the occupation and the charges he will be perceiving when the property is occupied. We will not go into the detail of the costs of the maintenance, taxes etc. but rather on forecasting the property's gross income. In order to have all the

information the investor needs we will also have to establish the real estate prices on specific areas.

PRELIMINARY DATA ANALYSIS

1.- The **venues** data base has 267.958 rows and 7 variables, this base has some information about some points of interest in five different cities of USA (NY, Chicago, Denver, Washington and Boston). Each record in the table shows information about a specific place, contains the name of the establishment, the type of services that someone could find there, such as if it is a bus station, or a restaurant, or a clothing store, a grocery store... it has the latitude and longitude from the place, the city where the establishment is located and a rating that is a numeric qualification between 1 and 5 of the establishment. Some remarks about the table are that almost a half of the data has some value in the rating column but the rest have no information, and it also has about 12% duplicate rows. This data doesn't have a zipcode or some kind of date either, because of that it is difficult to know how recent the data is.

2.- The **Real Estate** data base has 29111 rows and 262 attributes.

The list of available features is as follows:

Type (ZHAVI – ZRI): Rent Indexes (Zillow Home Value Index and Zillow Rental Index) represents Zillow's median estimate ("Zestimate") valuation or a house or apartment of a certain zip or neighborhood. Dtype: Object

Zipcode: postal code used to designate a specific location of a Real estate properties: Dtype: Int64

City: Settlement to define the location of the Real Estate properties. Dtype: Object

State: Territory organization where the Real Estates properties are located. Dtype: Object

Metro: Metropolitan Area Definitions designed to identify the location of Real Estate Properties. Dtype: Object

County: Territorial division of the countries to identify where is located the Real Estate property. Dtype: Object

Size_rank: Ranking established for the Real Estate to bring a identification. Dtype: Int64

Real Estate cost range from 1996 to 2017 by Months: Real estate valuations that estimates the price a buyer should pay for a piece of property, these columns are organized by years and also by months. Dtype: Int64 and Float64. Dtype: Object

Real Este Info:

RangeIndex: 29111 entries, 0 to 29110

Columns: 262 entries, type to 2017-06

dtypes: float64(216), int64(41), object(5)

The following table, we can define it as an informative table that offers information of the prices through time of the real estate established in the rent of Airbnb, these are organized by Indexes of rent and sale like by the corresponding zip code, city, county, metropolitan area and ranking of size, being a suitable table to be able to infer the reach of the defined project, to be able to give to an investor the necessary information to make purchase of real estate for investment and to generate utilities from the application Airbnb.

3.- The **demographics** table has the following variables:

Mean_household_income	Integer
-----------------------	---------

Variable	Type
zipcode	Integer
age (13 ranges)	Integer
households	Integer
Income brackets	Float
Median_household_income	Integer
Mean_household_income	Integer

In the review of the demographic table, we consider that it will be useful to generate a description by region through the population and the level income. In this way and having presented the investigation focus, the investor would take decisions with more tools according to their preferences. The key field to the table will be the zip code. We are considering to build new variables that will generate ranges greater than those presented for age and income, in this way we will approach another dataset that provides principal information of the investigation

4.- The **Calendar** Data Base:

This dataset includes 5 variables as follows:

Variable Name	Data Type
Listing_id	String
Date	Datetime
Available	Boolean
Price	Float
Metro Area	Categorical

It has 21 million of rows roughly. A brief description of each variable can be found below:

Listing_ID: It is an unique ID by estate.

Date: It is the sample date.

Available: This field indicates if the estate is available or not.

Price: It is the estate Price per night.

Metro_area: City

First approach remarks:

- Only the rows with availability True has a Price data.
- 13M of rows are not available in the sample date.
- The data was got between May 2016 and May 2018.
- The Price distribution looks like heavy tailed distribution.
- The most expensive estates are located in NY.
- There are 5 cities (classes) in the metro area column.

5.- The **listings** database

This data source contains information about 58.824 Airbnb locations along the U.S. West coast. Each row gives relevant information about the property, number of bedrooms, number of bathrooms, type of property and other amenities that the property might have. With this data it is possible to get important insights out for possible investors. As we know the price of the property and its conditions it is possible to think on possible correlations between those and see if a certain property yield more revenue in comparison with the needed invest to make it ready to rent.

Other information as zipcode will let us connect this database with the other ones used in this case. Information related with latitude, longitude, metropolitan and state let us plot information in graphics and see where property yield more revenue in comparison with other cities or counties in the US.

ⁱ <https://www.nytimes.com/2020/08/19/technology/airbnb-ipo.html>

ⁱⁱ <https://www.forbes.com/sites/petercohan/2020/11/19/at-30-billion-valuation-airbnb-stock-will-pop-on-its-ipo/?sh=3a9370c21c80>

ⁱⁱⁱ <https://www.nytimes.com/2020/09/24/travel/airbnb-pandemic.html>

^{iv} AMES TEMPERTON, "*I stumbled across a huge Airbnb scam that's taking over London*", Wired, published 11 february 2020, accessed 27 November 2020

^v <https://www.investopedia.com/articles/investing/083115/top-cities-where-airbnb-legal-or-illegal.asp>

^{vi} Cocola-Gant, Agustín & Gago, Ana. (2019). *Airbnb, buy-to-let investment and tourism-driven displacement. A case study in Lisbon*. Environment and Planning A. 10.1177/0308518X19869012.

<https://www.researchgate.net/publication/334731376_Airbnb_buy-to-let_investment_and_tourism-driven_displacement_A_case_study_in_Lisbon>

^{vii} Millionacres is a service of the well-known stock adviser Motley Fool

^{viii} Aly J. Yale, "*Buying an Investment Property for Airbnb*", Millioacres, published January 22, 2020. Visited 27 november 2020

<<https://www.fool.com/millionacres/real-estate-investing/rental-properties/buying-property-airbnb>>