

# Project Name: “AIRBNB Insights Report”

Tableau server link- [https://tableau.admin.uillinois.edu/views/Project-AirbnbInsights/Story1?iframeSizedToWindow=true&:embed=y&:showAppBanner=false&:display\\_count=no&:showVizHome=no](https://tableau.admin.uillinois.edu/views/Project-AirbnbInsights/Story1?iframeSizedToWindow=true&:embed=y&:showAppBanner=false&:display_count=no&:showVizHome=no)

Contribution : Individual

## Why Airbnb?

Airbnb is a leading website where people can book beds, rooms, apartments, homes etc. all around the world. It provides a platform for the people to rent out their places at their convenience without involving complex channels or doing major investments. Users can also find a place to stay at much competitive prices as compared to hotels. Through Airbnb, people can find place to stay even in the areas where the likelihood of having hotels is very less. Many a times, even people prefer to stay in local settings, with local people.



Fig. Average rates for an entire home on Airbnb is the same or lower for the cost of a single hotel room.

## Airbnb Statistics

• Over 4 million listings worldwide	• 150 million users
• In over 191 countries	• Worldwide value is \$32 billion
• Global growth rate since 2009 - 153%	

These statistics clearly show that Airbnb is a leader in the business and drew our attention to study in depth and visualize its various parameters and functioning.

## Project proposal

Initially we wanted to study about Airbnb working in Chicago area. To make the study more interesting and insightful, we later decided to include one more city (New Orleans), to draw similarities, differences and special patterns these both cities possess.

## Dataset Selection

The datasets of both cities- **Chicago** and **New Orleans** have been taken from 'Inside Airbnb' which is an independent, non-commercial set of tools and data that allows to explore how Airbnb is being used around the world.

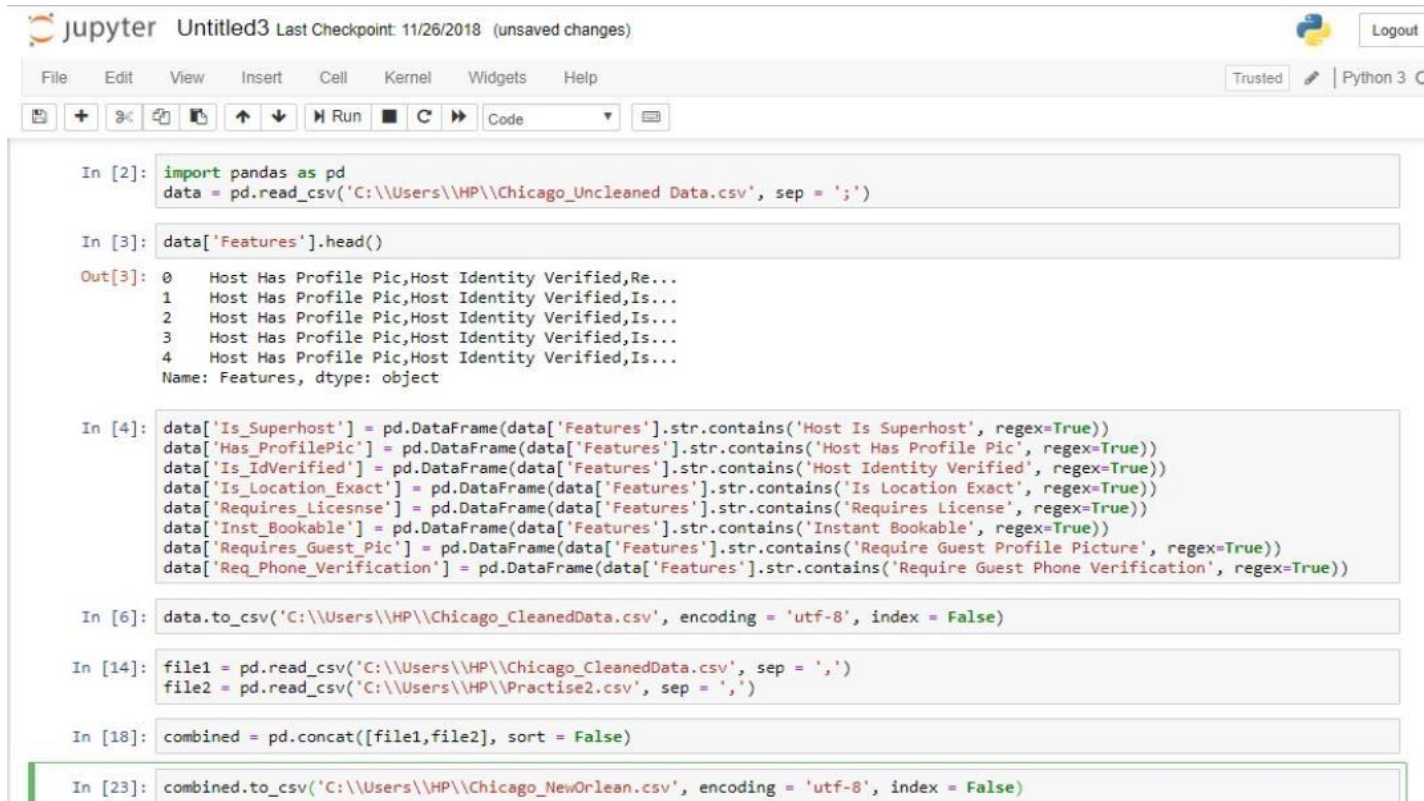
## Data Cleaning

Initially, Data was very messy and uncleaned, so we used external tools to make it ready for the Visualization. Data cleaning has been done using **Jupyter Notebook** and **Tableau Prep**.

### Jupyter Notebook

First Tool used was Python Language using Jupyter Notebook Environment, where we generated few extra columns with the help of already existing categorical column which was in actual long descriptive text and was hard to decipher and cannot be used in Visualization.

The extra columns that we created gave a much clear sense of how to approach and make an effective Visualization



The screenshot shows a Jupyter Notebook titled 'Untitled3' with a last checkpoint of '11/26/2018 (unsaved changes)'. The interface includes a top menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. Below the menu is a toolbar with icons for file operations, a 'Run' button, and a 'Code' dropdown. The notebook contains several code cells:

```
In [2]: import pandas as pd
data = pd.read_csv('C:\\Users\\HP\\Chicago_Uncleaned Data.csv', sep = ';')

In [3]: data['Features'].head()

Out[3]: 0    Host Has Profile Pic,Host Identity Verified,Re...
1    Host Has Profile Pic,Host Identity Verified,Is...
2    Host Has Profile Pic,Host Identity Verified,Is...
3    Host Has Profile Pic,Host Identity Verified,Is...
4    Host Has Profile Pic,Host Identity Verified,Is...
Name: Features, dtype: object

In [4]: data['Is_Superhost'] = pd.DataFrame(data['Features'].str.contains('Host Is Superhost', regex=True))
data['Has_ProfilePic'] = pd.DataFrame(data['Features'].str.contains('Host Has Profile Pic', regex=True))
data['Is_IdVerified'] = pd.DataFrame(data['Features'].str.contains('Host Identity Verified', regex=True))
data['Is_Location_Exact'] = pd.DataFrame(data['Features'].str.contains('Is Location Exact', regex=True))
data['Requires_Licesnse'] = pd.DataFrame(data['Features'].str.contains('Requires License', regex=True))
data['Inst_Bookable'] = pd.DataFrame(data['Features'].str.contains('Instant Bookable', regex=True))
data['Requires_Guest_Pic'] = pd.DataFrame(data['Features'].str.contains('Require Guest Profile Picture', regex=True))
data['Req_Phone_Verification'] = pd.DataFrame(data['Features'].str.contains('Require Guest Phone Verification', regex=True))

In [6]: data.to_csv('C:\\Users\\HP\\Chicago_CleanedData.csv', encoding = 'utf-8', index = False)

In [14]: file1 = pd.read_csv('C:\\Users\\HP\\Chicago_CleanedData.csv', sep = ',')
file2 = pd.read_csv('C:\\Users\\HP\\Practise2.csv', sep = ',')

In [18]: combined = pd.concat([file1,file2], sort = False)

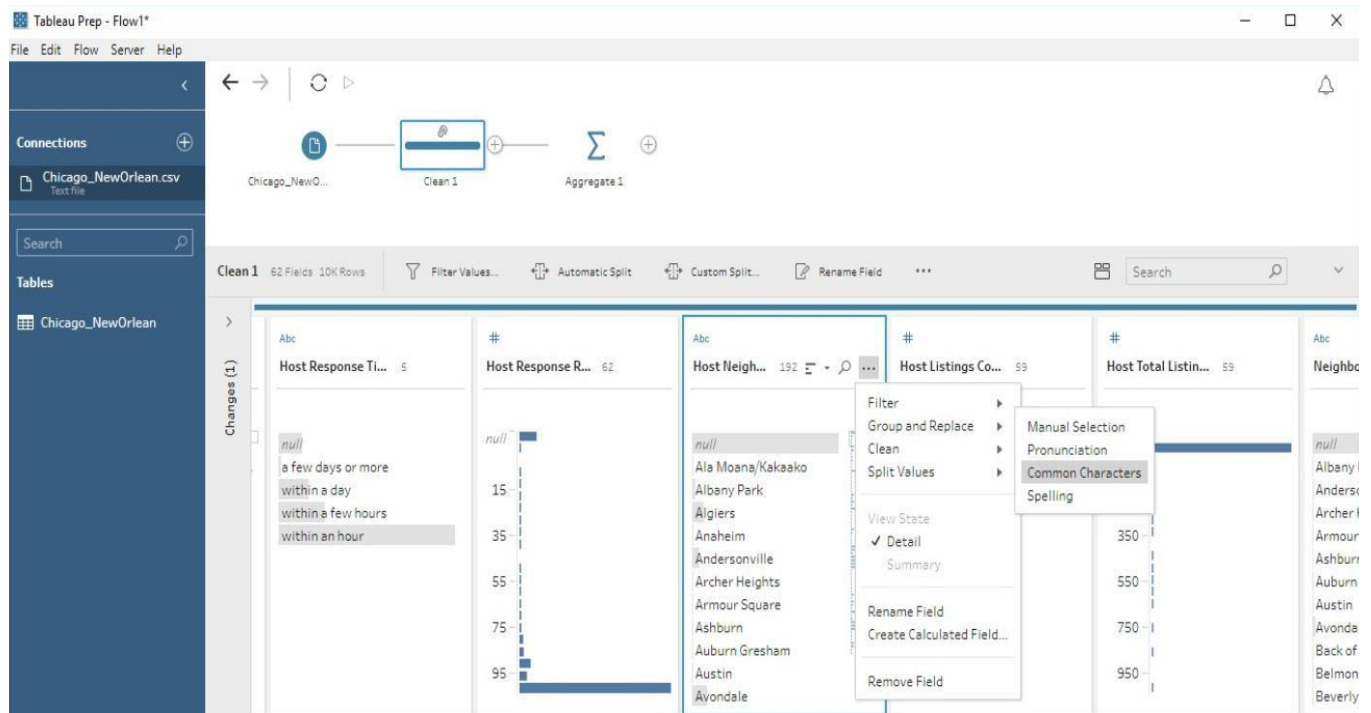
In [23]: combined.to_csv('C:\\Users\\HP\\Chicago_NewOrlean.csv', encoding = 'utf-8', index = False)
```

## Tableau Prep

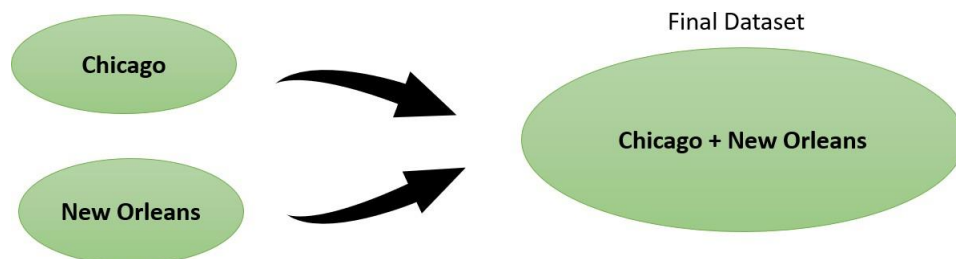
We also used Tableau Prep, to use the functionality of "Group and Replace" which it provided.

In our dataset under the column name say Neighbourhood, there were some names which were same but were written differently such as some of them were either in lower cases or upper cases or some were having different spellings but pronunciation was same.

So, Tableau Prep applied an algorithm to find all those similar words and replace it with one specific word which has actual Spelling and Pronunciation.



Finally, the cleaned datasets of both the cities are merged in Jupyter Notebook to get the 'Final Dataset' which has been used for doing the visualizations.



## Visualization Tool

**Tableau** has been used as the Data Visualization tool as it is a very powerful, secure and flexible end to end analytics platform where we can visualize our data quickly and easily by creating interactive dashboards and convey an overall story using storyboard. Tableau is been used predominantly in the industry for creating powerful visualizations.

## Target Audience

The insights drawn could be helpful for both Hosts and Users.

Hosts- People putting up their place for rental purpose.

Users- People searching for accommodation.

## Questions Addressed

1. Which are the popular **neighborhoods**, their average prices and no. of listings.
2. What are percent share of different **property type** and **room type**.
3. How the **pricing** is varying with location, property type, reviews.
4. What are the different correlations between type of **hosts** and factors like- reviews & price.

## Data Visualization

Our visualization findings has been categorized into four categories-

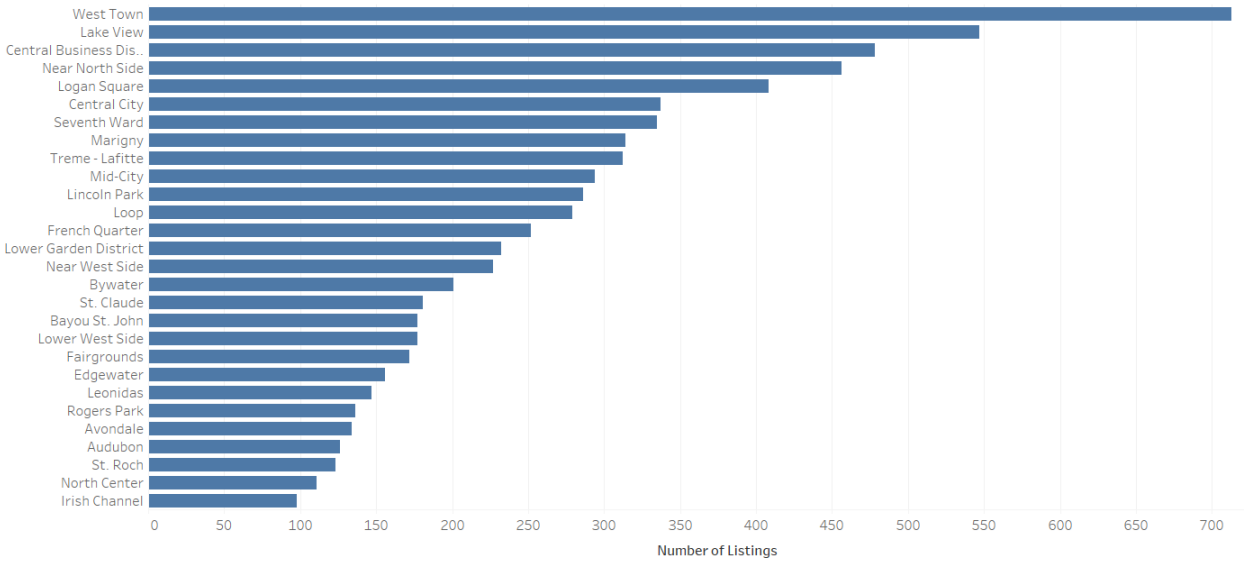
- a) Overview of Airbnb
- b) Property analysis
- c) Pricing analysis
- d) Host analysis

### a) Overview of Airbnb

1. Popular neighborhood areas in the ranking from highest number of listings to the least.

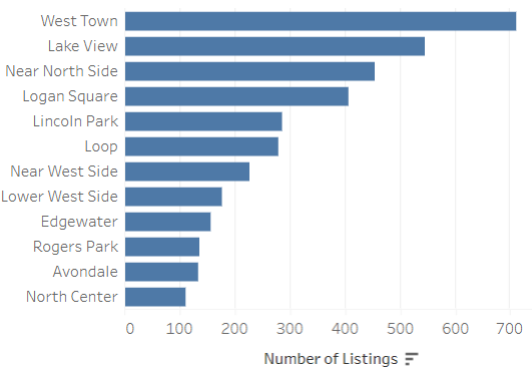
# Overall

## Popular Neighbourhood



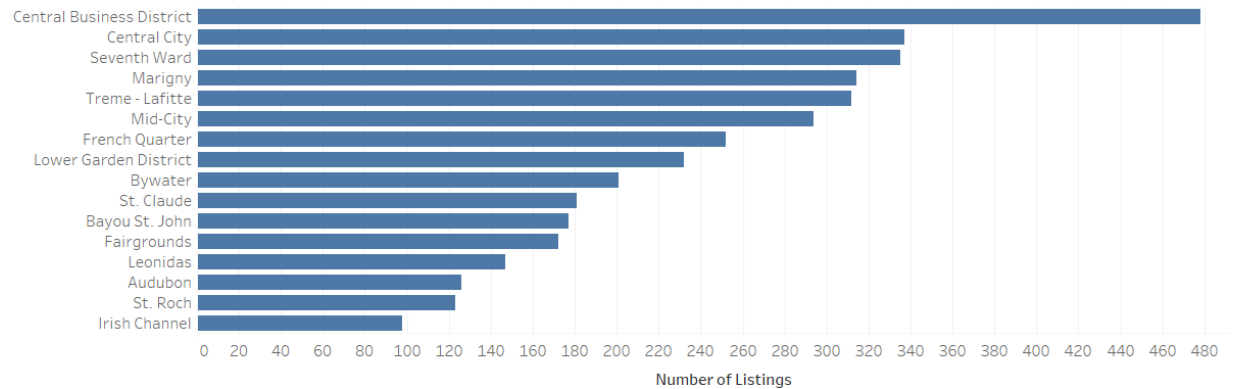
# Chicago

## Popular Neighbourhood



# New Orleans

## Popular Neighbourhood

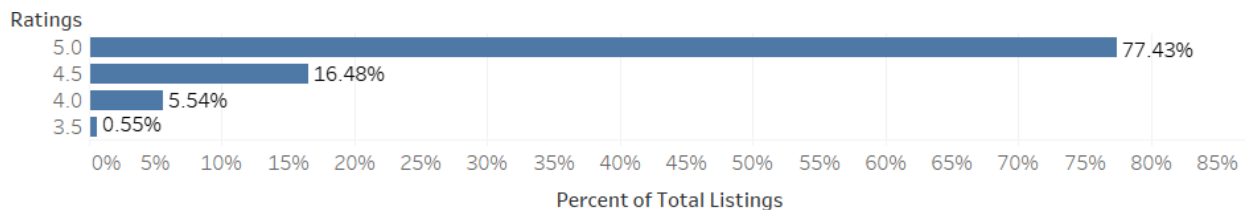


Overall, West Town area, in Chicago, has 713 number of listings (highest). West town is one of the best places to live in Illinois as it is very posh, located in west of downtown Chicago and is a center of attraction for people due to its liveliness.

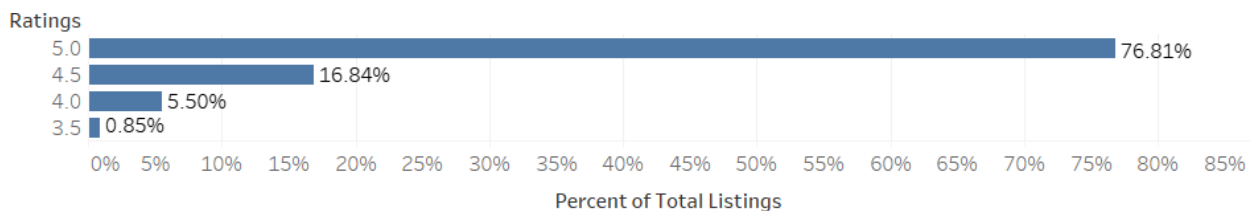
On the other hand, Irish Channel (in New Orleans) has least number of listings i.e. 98. One of the possible reasons could be that Irish Channel is predominantly known for its crime activities. It is not considered safe to live in the area, especially during night time. So, there are high chances that people don't prefer to take accommodation in this area which has affected the number of listings as well.

## 2. Ratings

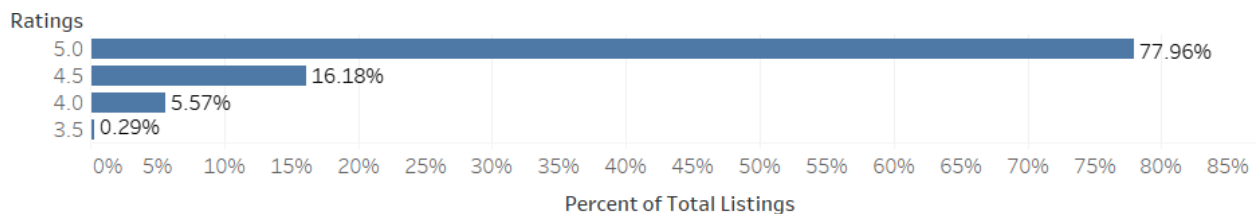
### Overall



### Chicago



### New Orleans



The rating pattern is quite similar in both Chicago and New Orleans. We can see that more than 75% of the listings have attained a rating of 5.0; and less than 1% have a rating of 3.5 or lower, which suggests that people are highly satisfied with the kind of accommodation, price and service they are getting through Airbnb.

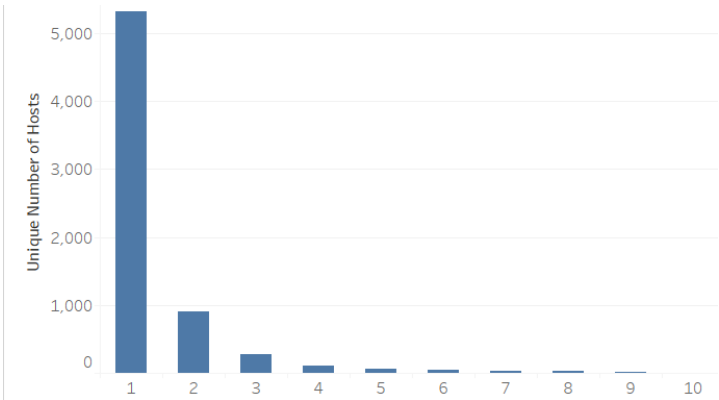
**Dashboard**



**b) Property Analysis**

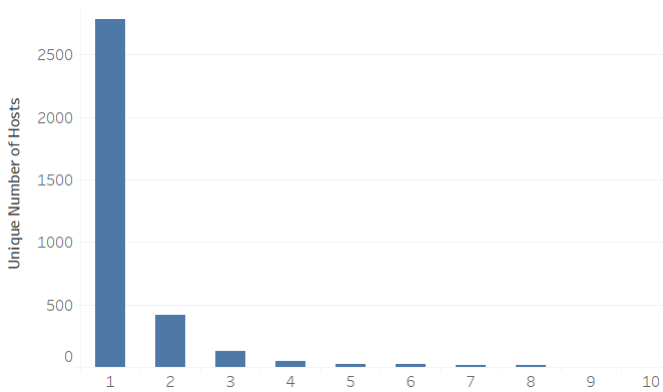
**1. Number of listings per host**

**Overall**

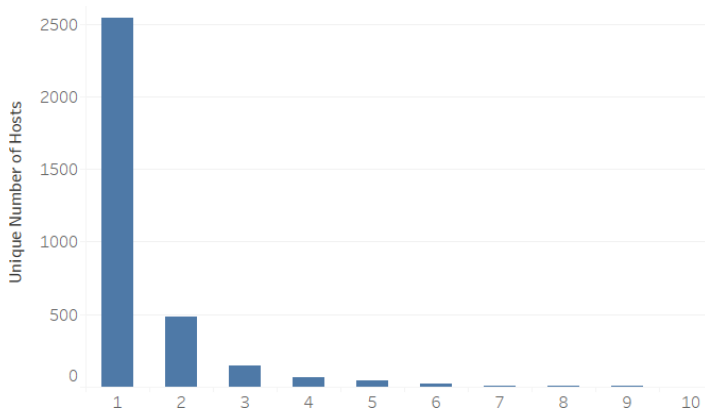




## Chicago



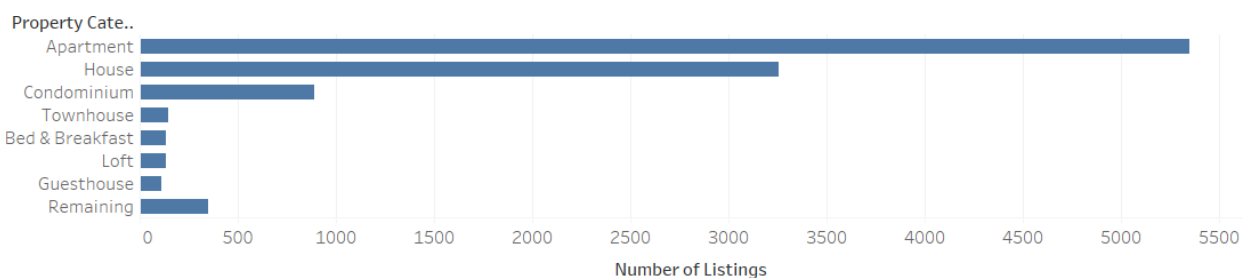
## New Orleans



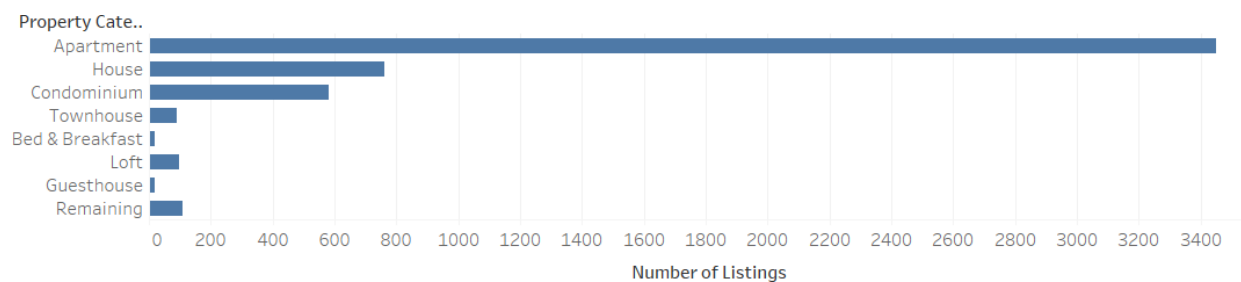
Both cities have the same pattern. Maximum number of hosts are having only one property. Also, there are only 4 hosts having 10 listings. Overall, more than 78% hosts have only 1 listing. This can be attributed by the fact that most of the hosts lend their place as a side business without getting into the main stream business, by lending their place or a room of their apartment at their convenience.

## 2. Property types

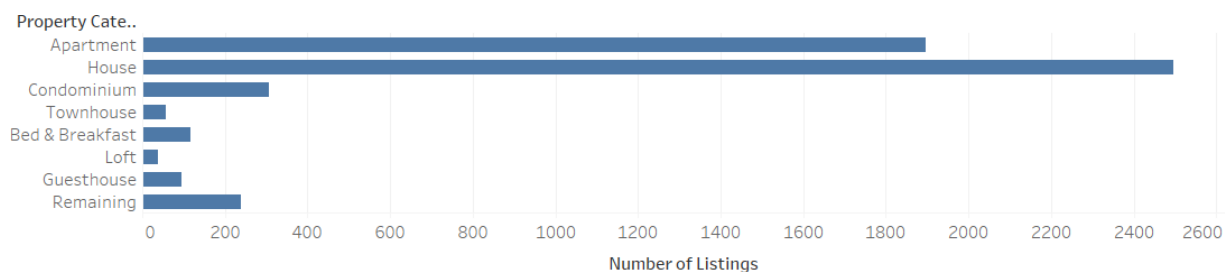
### Overall



## Chicago



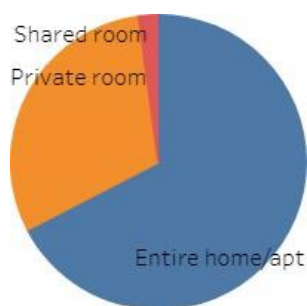
## New Orleans



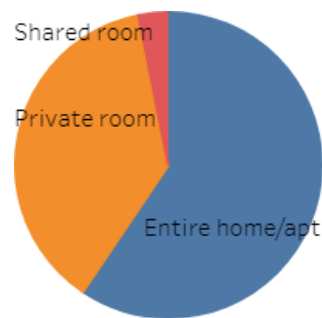
Predominantly, 'Apartments' and 'houses' have the highest number of listings among all the property types. Overall, there are around 50% apartments and 30% Houses. In Chicago, 'Apartments' outnumbers all the other property types whereas in New Orleans, majority of the property types are 'House'.

## 3. Room Type

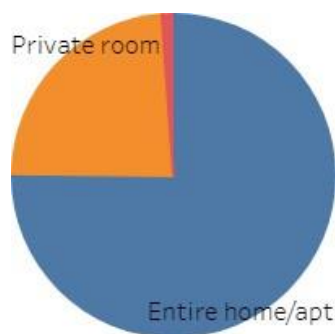
### Overall



Chicago



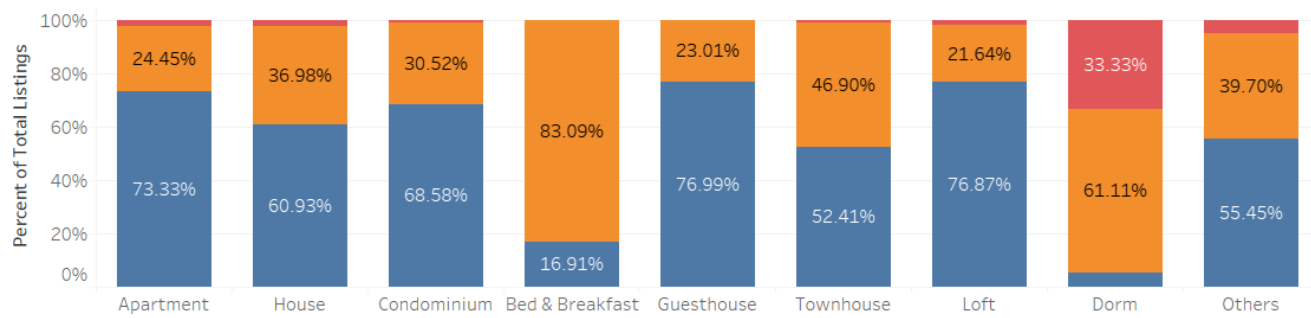
New Orleans



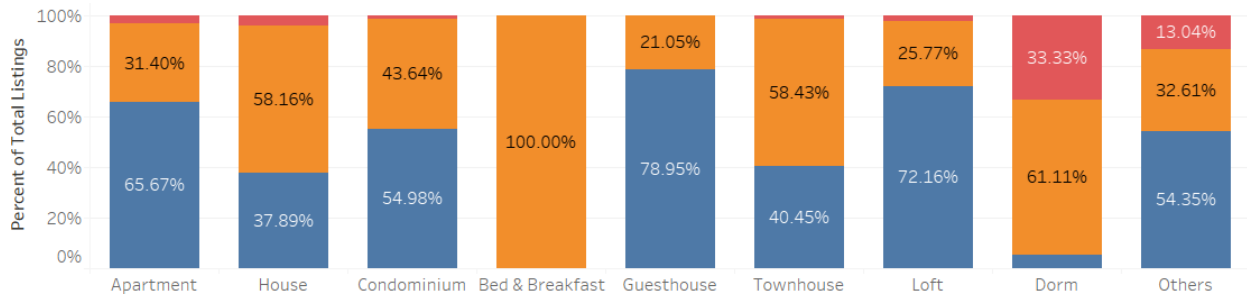
Entire home/apt have the highest share among ‘Room Types’ followed by Private Room and Shared Room. Both the cities depict same pattern.

4. Room Type by Property Type

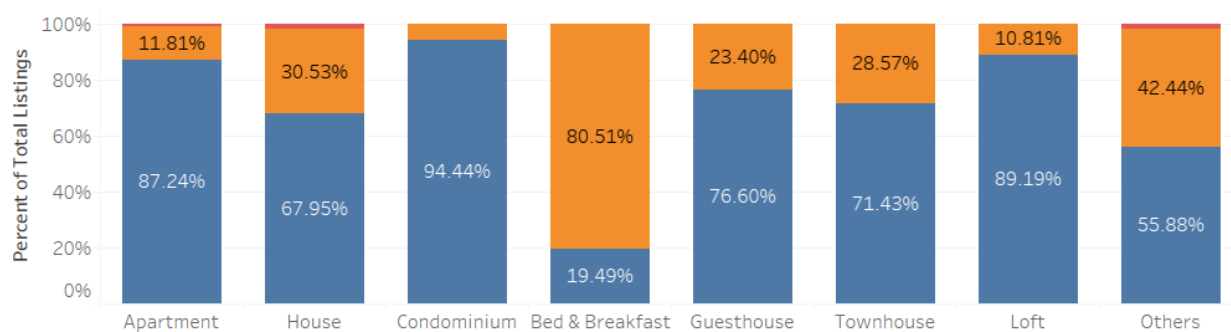
Overall



## Chicago

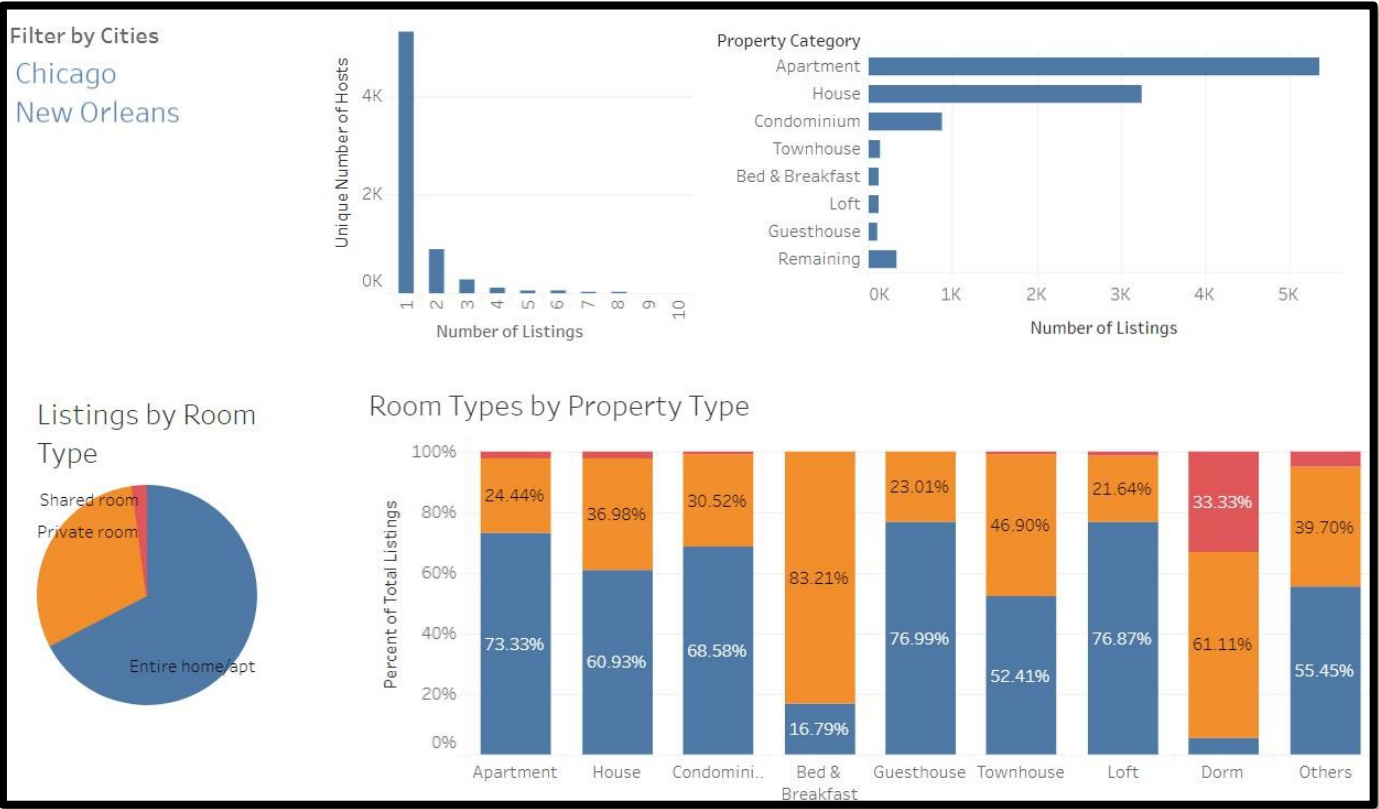


## New Orleans



All the property type, either in Chicago or in New Orleans, have the similar pattern that they all have more than 50% of the Room type as Entire Home/Apartment except Bed & Breakfast and Dorm type. These exceptional property type have more than 50% room type as Private Room, and this is very intuitive as people who are or may be solo travelers, would not like to book an entire home instead they want to live in private room plus it would be great for them if they are been provided by Breakfast, the next day.

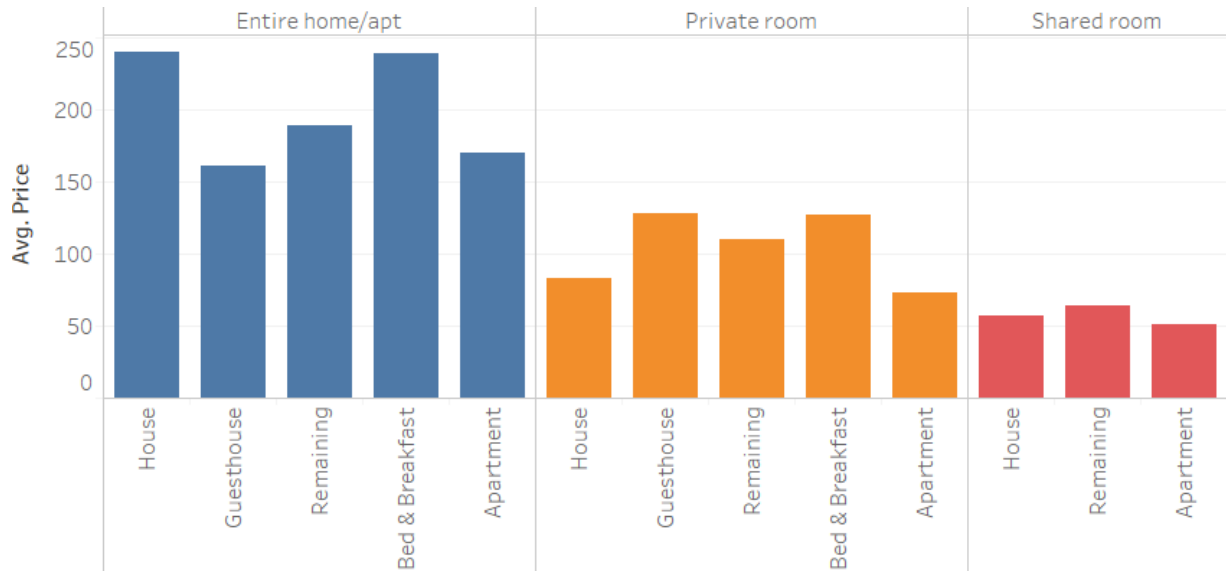
**Dashboard**



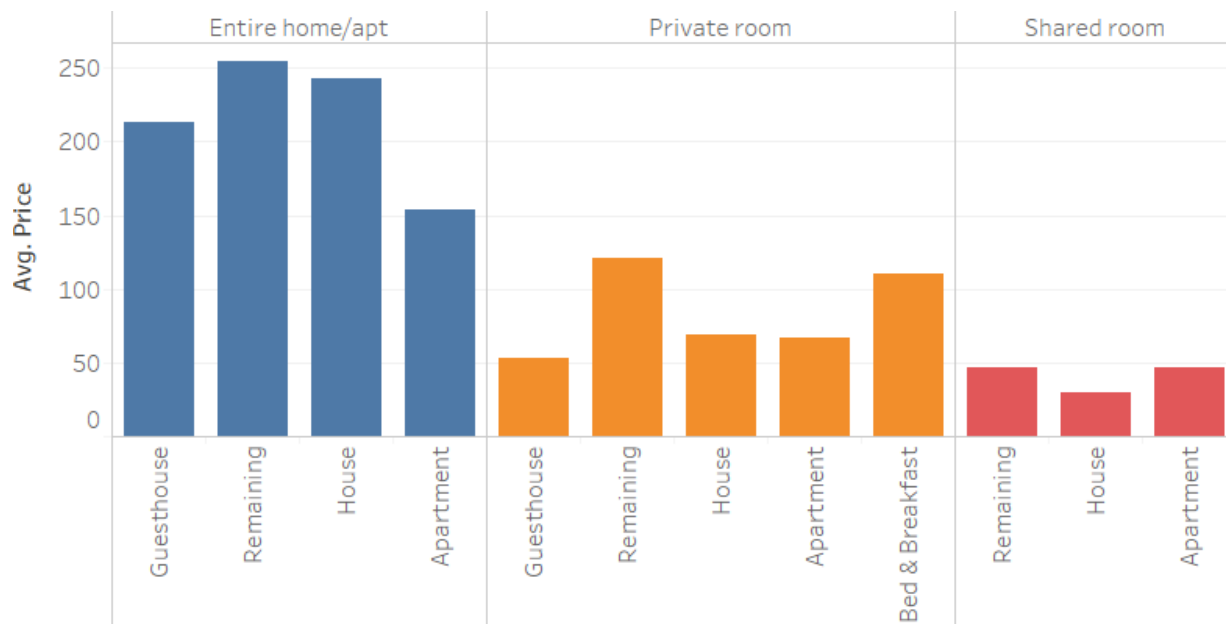
**c) Pricing analysis**

## 1) Room Type/ Property Type

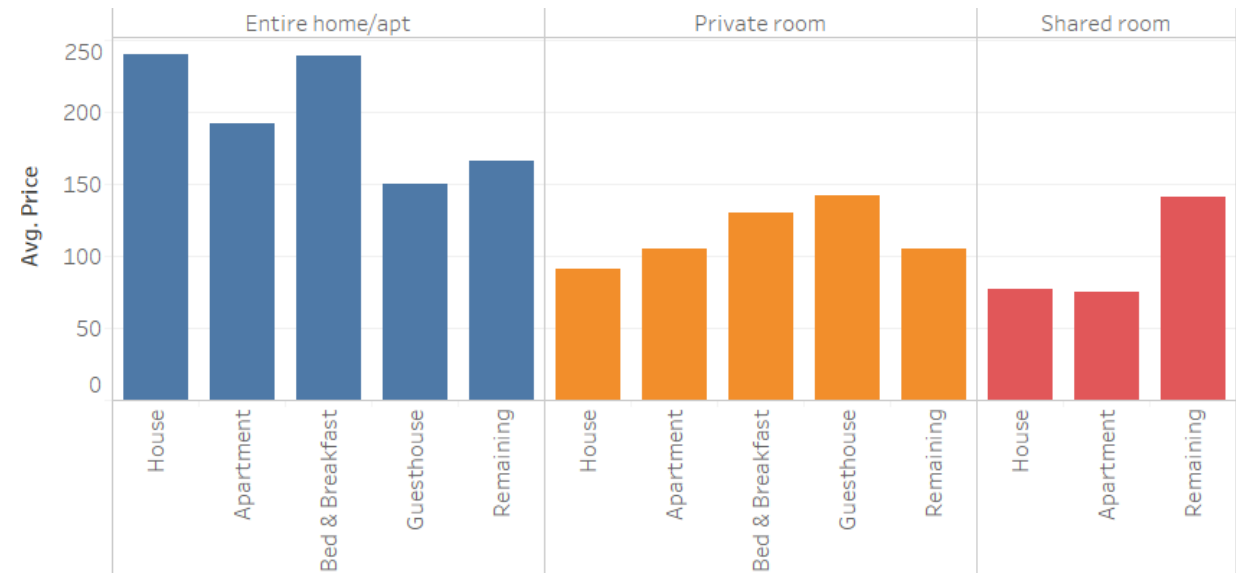
### Overall



### Chicago



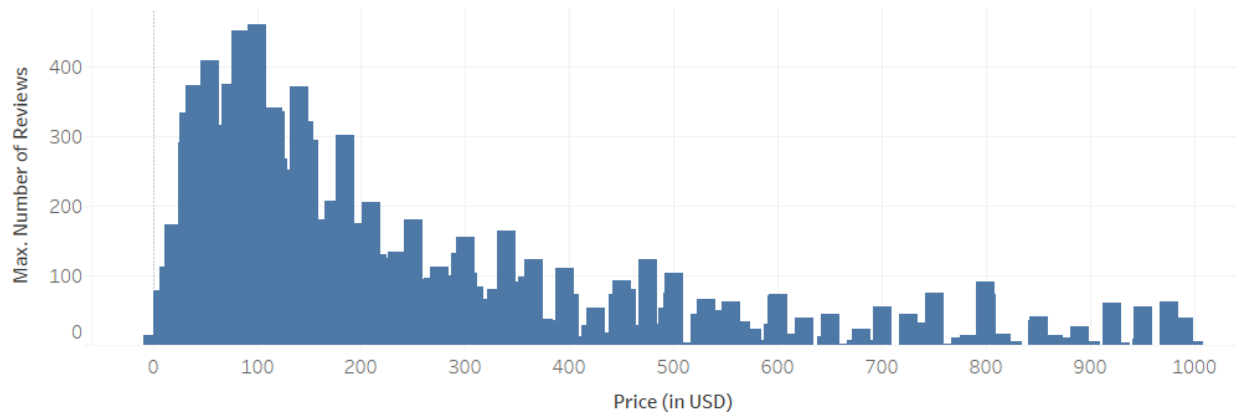
## New Orleans



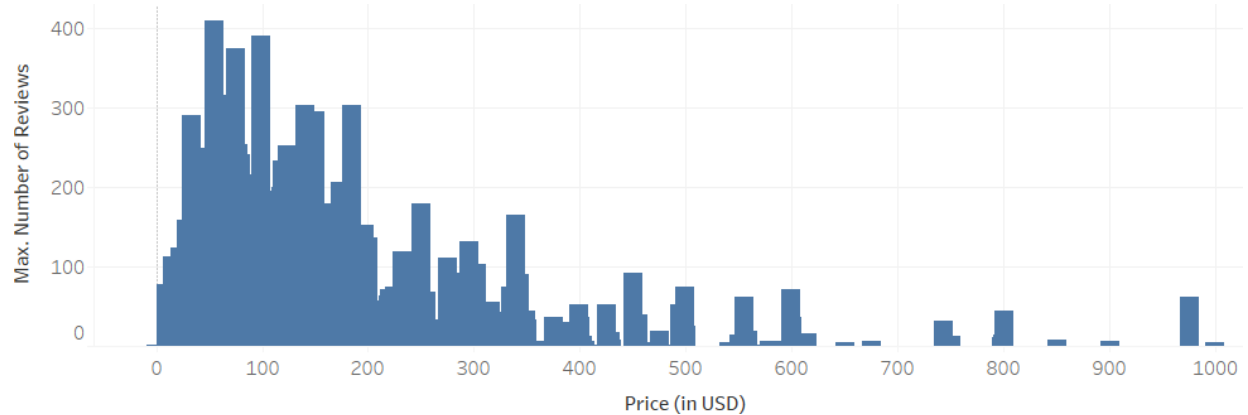
Average Price of Entire home/apt is the highest for both cities, followed by Private Room and Shared Room. An interesting finding to note is that though the number of listings for 'Apartment' type is the highest (~ 50%) as discussed in *Property Analysis- Dashboard 2*, Average price of 'Apartments' is never maximum in any of the 'Room Type'. This can be attributed by the fact that there are enough 'Apartments' available and hence there is no tight market for the 'Apartments' leading to no severe price competition.

## 2) Maximum number of Reviews

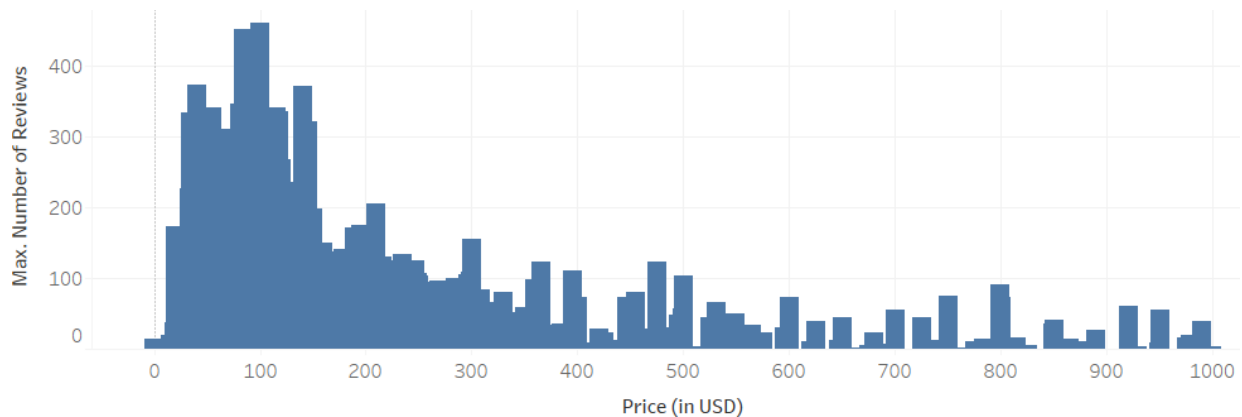
### Overall



## Chicago



## New Orleans



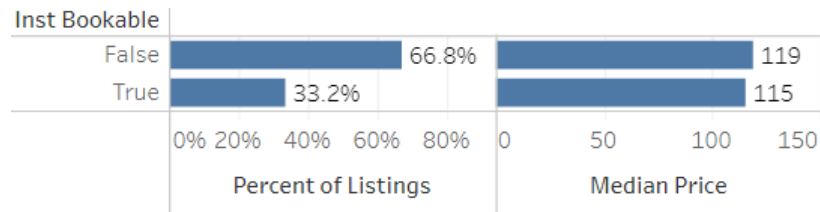
Both the cities show similar pattern where maximum number of reviews are for the listings which lie in the price range of \$25-\$200. There are two possible reasons for it.

- i) Most of the people prefer affordable and cheaper options to stay. Hence, there is high probability that very less number of people take up high priced accommodations and consequently have less number of reviews. On the other hand, major portion of people interesting in affordable accommodations will correspond to more number of reviews. We do not have any data regarding the number of times a listing has been booked. So, we cannot validate the fact.
- ii) Mostly, the elite/rich people opt for pricey options to stay and generally they are least bothered about posting their reviews to public forums. That could be a major reason for less reviews for high price accommodations. On the other hand, middle class people, opting for affordable options, care much more about writing reviews about the service they get.

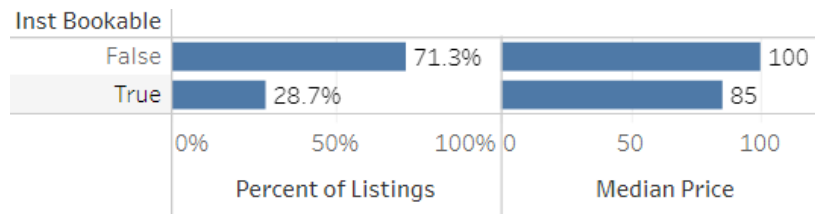


### 3) Instant Bookable v/s Non-Instant Bookable listings

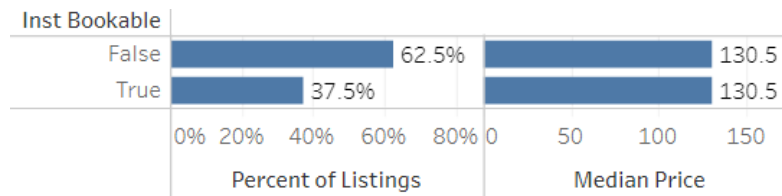
#### Overall



#### Chicago



#### New Orleans

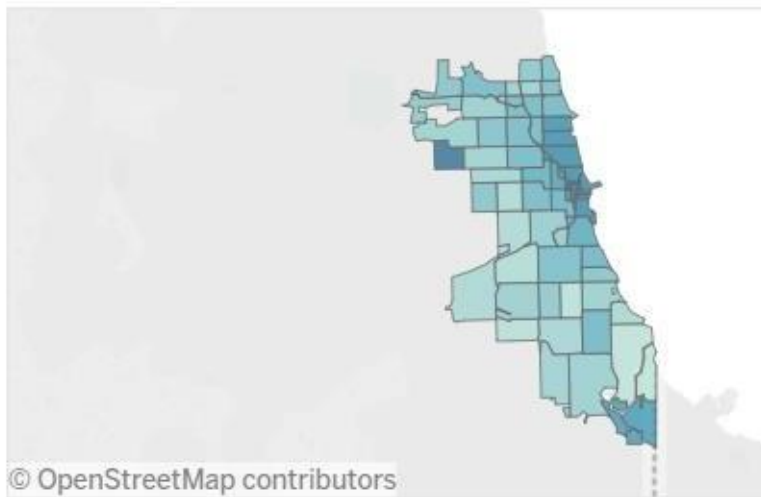


Accommodations available for 'Instant Booking' are usually considered to be lesser and the statistics also verifies the fact that overall there are only 33% listings which are available for instant booking. It is also considered a common notion that the instant bookable accommodations are usually pricey as compared to non-instant bookable ones. But, in contrast to our assumptions, we see that the prices of both instant bookable and non-instant bookable listings are comparable.

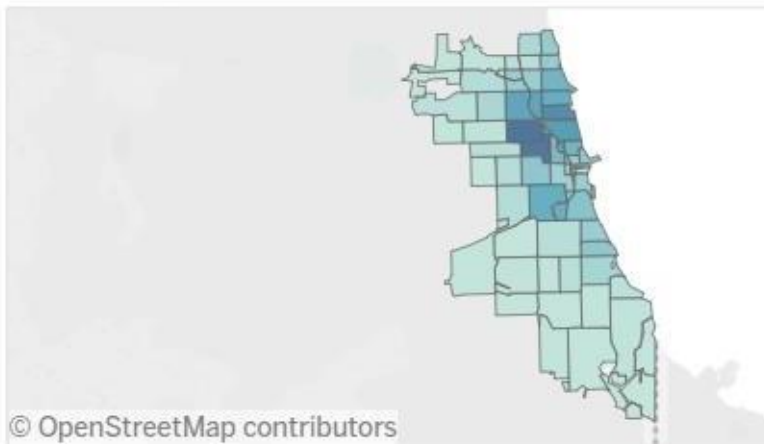
### 4) Average Price by Number of Listings

## Chicago

### Average Price By Area



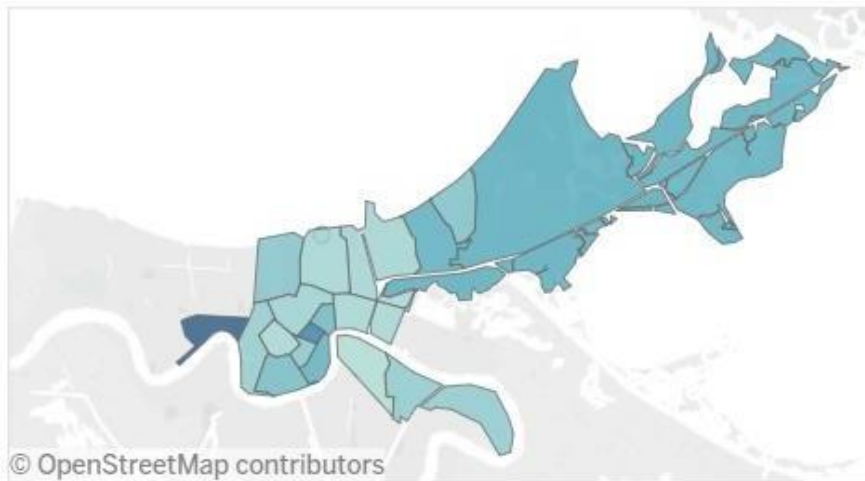
### Number of Listings By Area



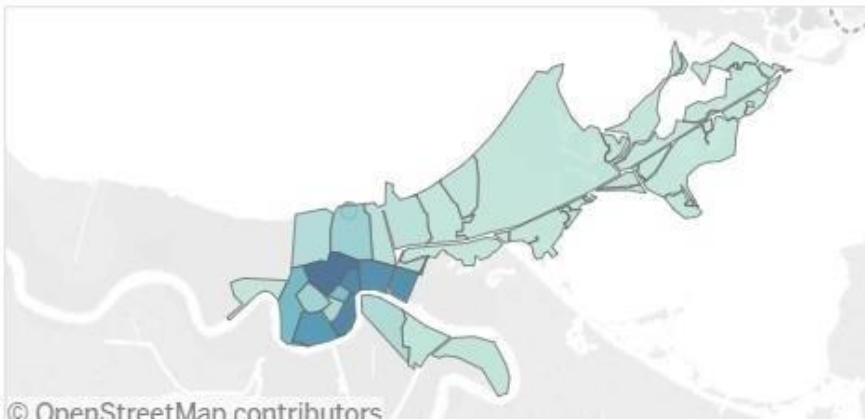
Here, we can see that the 'number of listings' are maximum along the coast line and gradually decreases as we move out of the coastal area. The average price is uniformly distributed around the city with slightly higher range around the coast line.

## New Orleans

### Average Price By Area

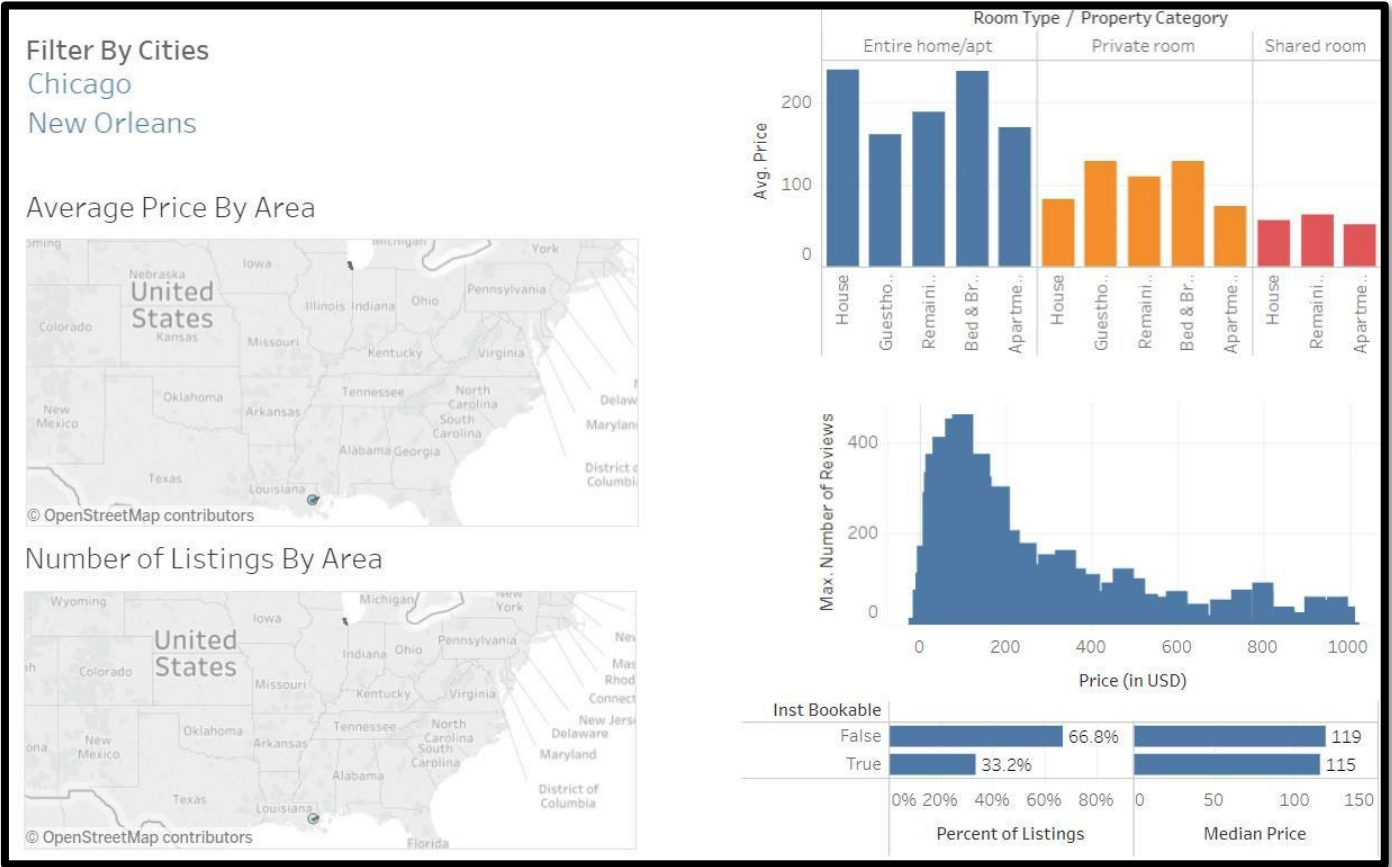


### Number of Listings By Area



Here, we can see that 'average price by area' and 'number of listing' are contradicting each other. The areas where the number of listings is comparatively more, the average price is less. This can be attributed by the fact that the competition is high and therefore the prices kept by the hosts are reasonable. In the areas where the listings are very less, prices are high. The hosts are probably getting benefitted by high demand due to less staying options in those areas.

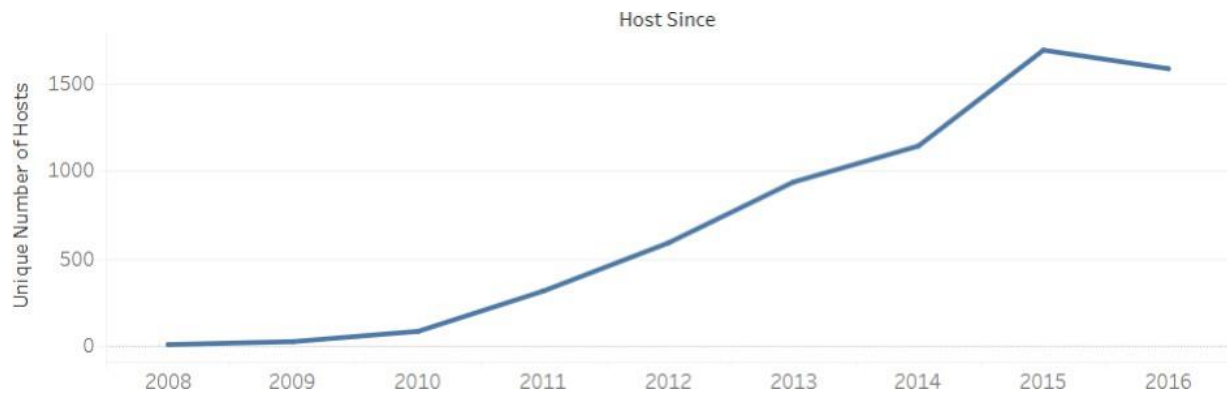
**Dashboard**



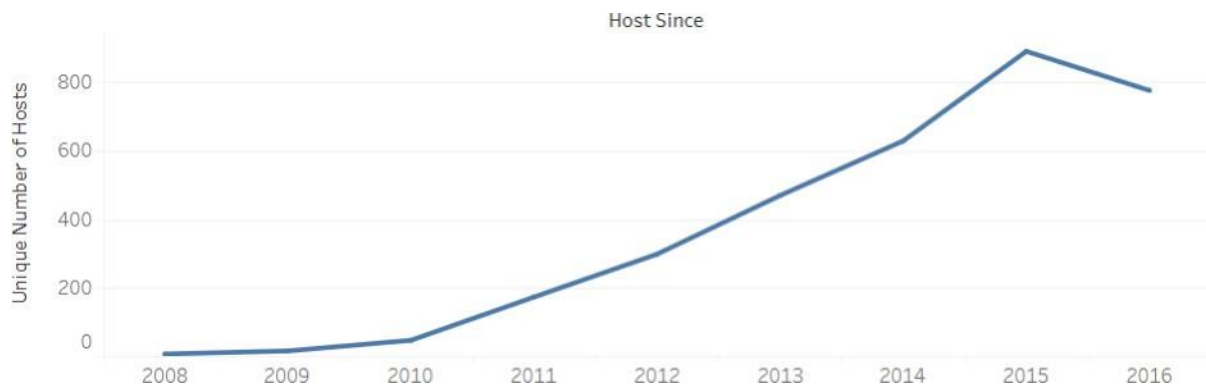
**d) Host analysis**

**1) Number of hosts added each year**

## Overall



## Chicago



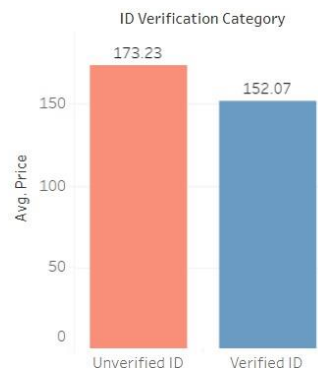
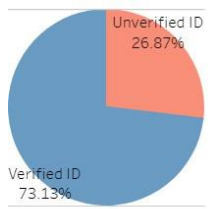
## New Orleans



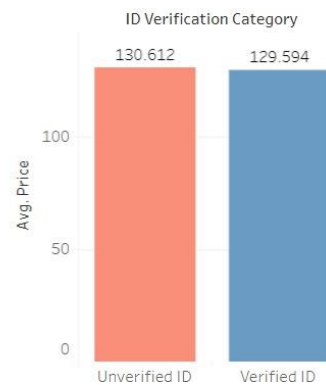
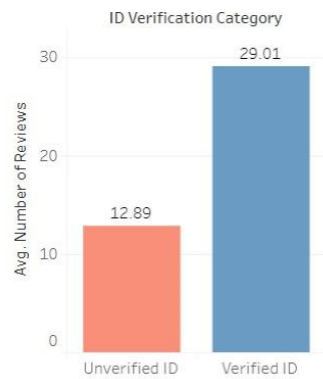
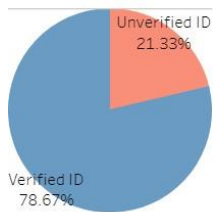
Here, we can see that the number of hosts joining Airbnb every year is continuously increasing at a higher rate, being only 8 hosts joining in 2008 to around 1500 hosts joining in 2016.

## 2) Verified/Unverified Hosts

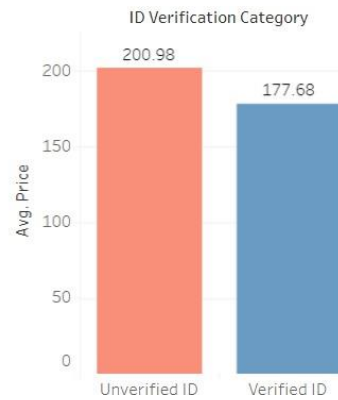
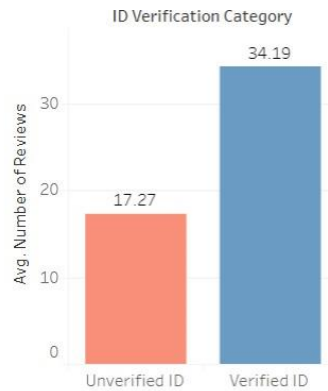
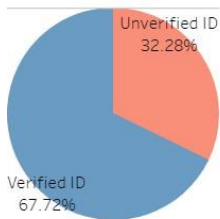
Overall



Chicago



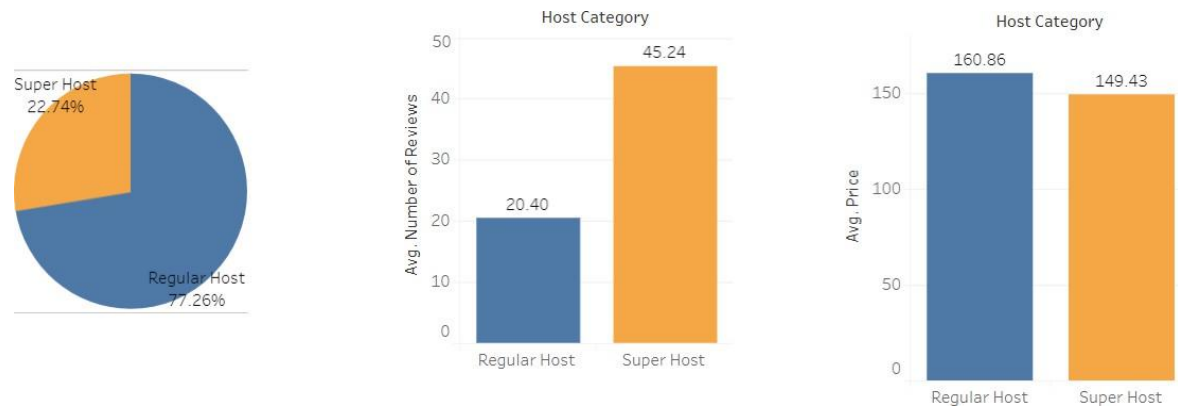
New Orleans



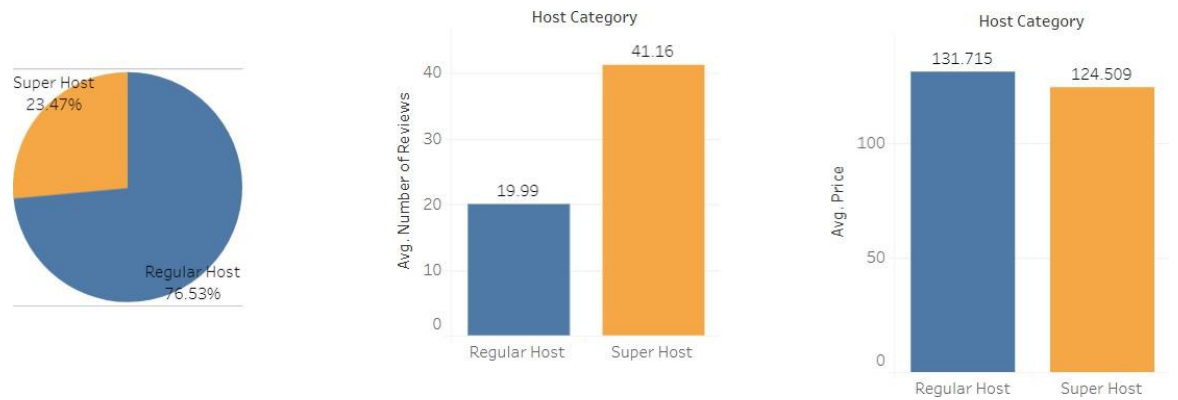
One of the findings which is very unexpected is that around one-third of the hosts are unverified on Airbnb. Although, unverified hosts possess higher prices as compared to the verified hosts, they get lesser reviews. There can be two possible reasons for that. First, maybe people do not prefer taking up places by the unverified hosts and hence lesser number of people booked the listings by unverified hosts which leads to comparatively lesser reviews. Since, we do not have any data for the number of times a listing has been booked, we cannot validate this fact. Second, even if people take up the places, they might be reluctant to write reviews for unverified hosts on public forums.

### 3) Super Hosts/Regular Hosts

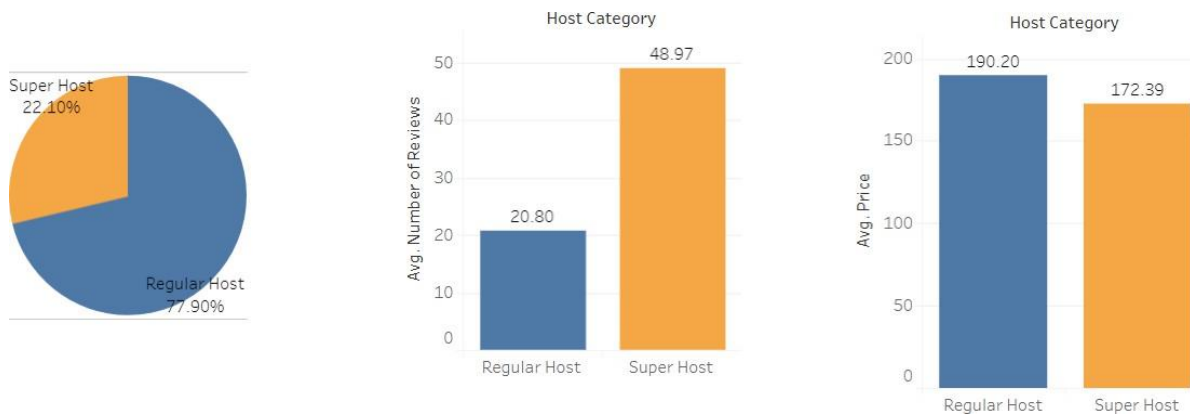
#### Overall



#### Chicago



## New Orleans



Airbnb lists its hosts into two categories- Super Hosts and Regular Hosts.

To become a Super Hosts, four criteria should be met by the hosts-

- Provide minimum 10 stays in a year.
- Maintain 90% response rate
- More than 80% 5-star reviews.
- Zero cancellations (with exceptions)

From the statistics, we can see that around one-third hosts have been designated as Super Hosts which suggests that hosts are doing pretty well in these areas. Also, Super hosts possess comparatively lesser prices and much higher reviews (around twice) as compared to regular hosts.



**Dashboard**

Filter By Cities

Chicago

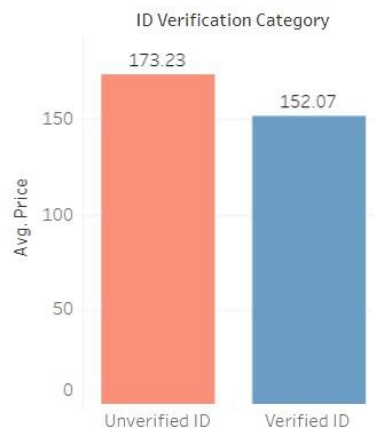
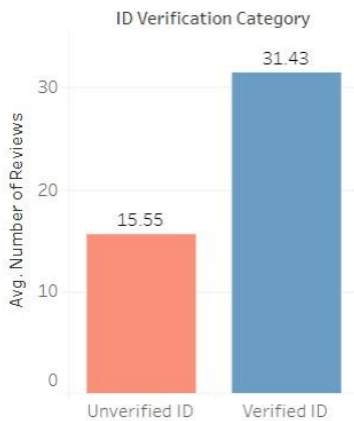
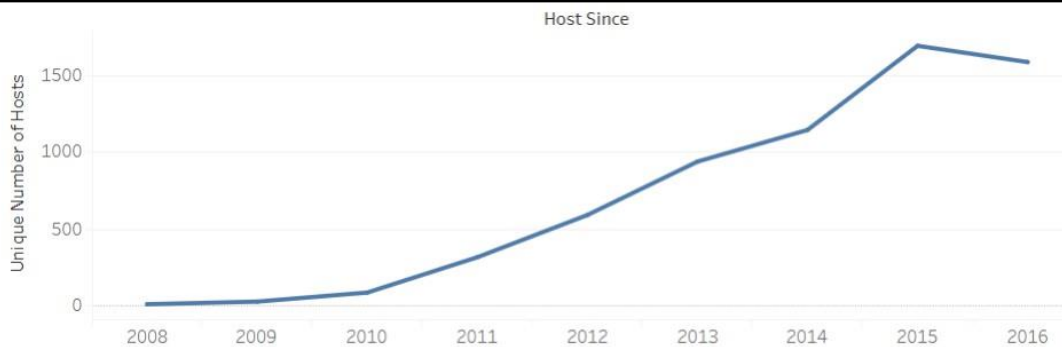
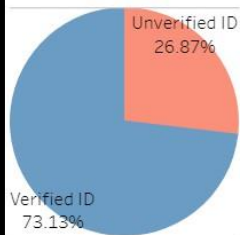
New Orleans

Features

ID

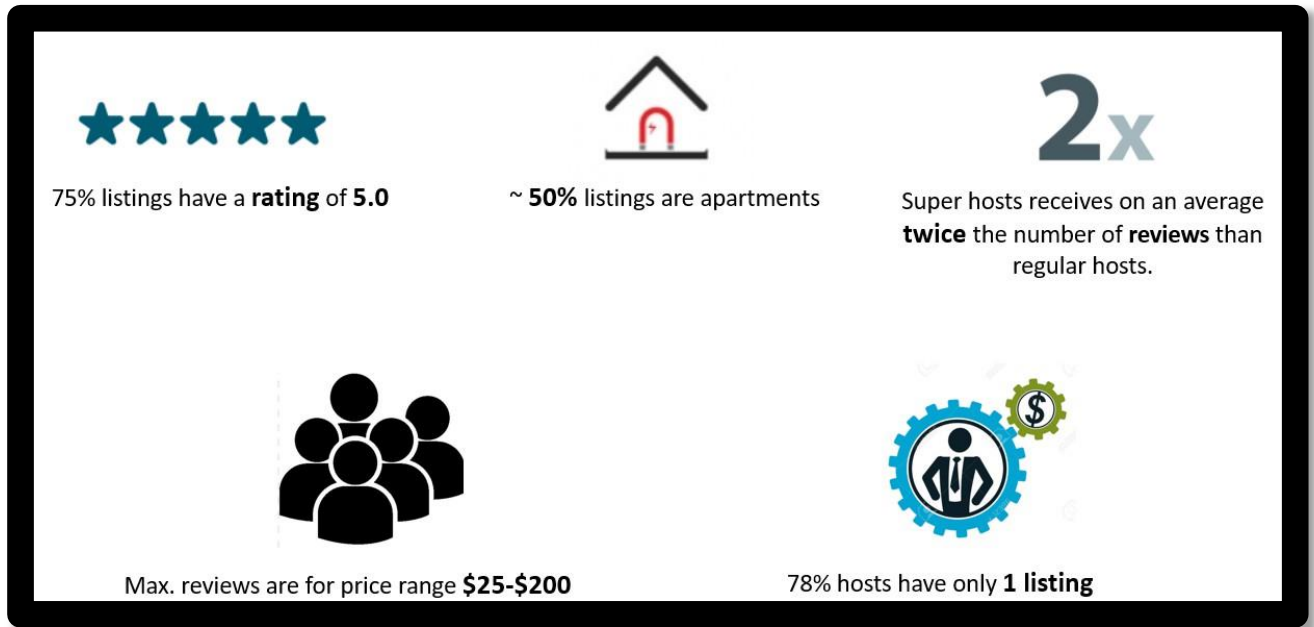
ID

Host



# Major Findings

Some of the major findings through the analysis have been listed below-



## REFERENCES

<https://ipropertymanagement.com/airbnb-statistics/>

<https://www.niche.com/places-to-live/n/west-town-chicago-il/>

<http://www.city-data.com/forum/new-orleans/1330835-irish-channel-area.html>

<https://www.airbnb.com/help/article/829/how-do-i-become-a-superhost>