

How did Air BNB fare across a number of major cities over time?

Abstract

Airbnb, which launched in 2009, has grown rapidly by generating value through the sharing economy business model. Property owners can profit from underutilized assets thanks to the Airbnb business model. However, it is time to examine how some of the world's most vibrant cities, such as London, Paris, Istanbul, New York, and Dublin, have grown in terms of revenue and accommodation capacity over time. We will also attempt to determine which factors are related to the amount of revenue generated by Air BNB. We will also look for patterns in how the business was affected by the Covid-19 pandemic. WAS it all the cities, or was there one that thrived despite the odds?

Data Collection

Data has been collected from: <http://insideairbnb.com/get-the-data>.

In my opinion, we have two 'V' of the Big Data in our dataset.

First is volume, the merged and cleaned data which we are dealing with is approximately 1 GB in size which contains 4,51,285 rows of data, thus we can categorize them into big data by the size of it.

Secondly, our data is diverse i.e., Variety. We have gathered a total of 16 tables, each with roughly 76 columns, for a variety of Air BNB-related data, such as latitude, host location, hotel amenities, check-in date, pricing, etc.

Thus, by the feature of 2 'V's in our dataset we can conclude that this project has used big data for the analysis.

Data Exploration, Processing, Cleaning and/or Integration

We have used Python for cleaning the datasets.

Since there were 16 separates .csv files for the respective cities, we have used Pandas to store the data in dataframes and merge them into a single dataframe. From that resultant dataset which had 76 columns, we trimmed down the unnecessary ones and took only the ones those are important to us. We have changed the formatting of the currency from comma separated at decimals to normal integer format.

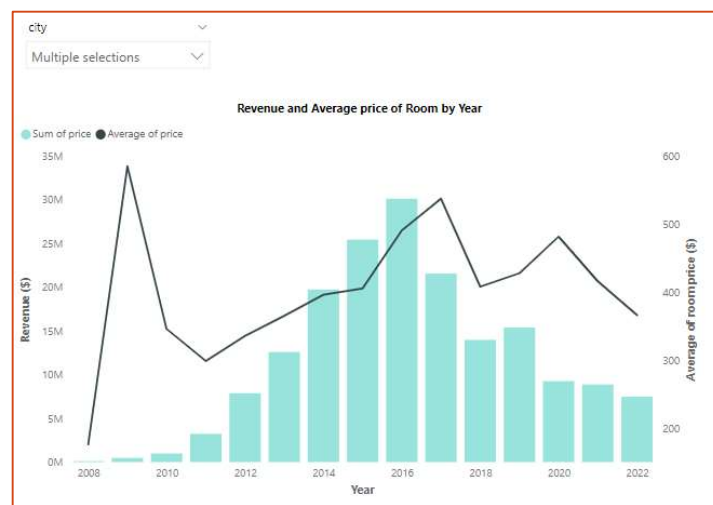
Data Visualization

We have selected Microsoft Power BI desktop as our visualization tool and have loaded our clean data to perform further analysis

We have plotted four graphs to stitch a story about Air BNB's performance at different happening cities of the world and what are the key factors affecting them.

1. Does the average room price per year correlate with the revenue generated by Air BNB per year?

- Average room price is calculated by applying the 'average' function on the room price attributes of all the unique hotels in Air BNB across major cities.
- Revenue, which means the total sales made, was calculated by applying the 'sum' function on the room prices of the Air BNB hotels booked.



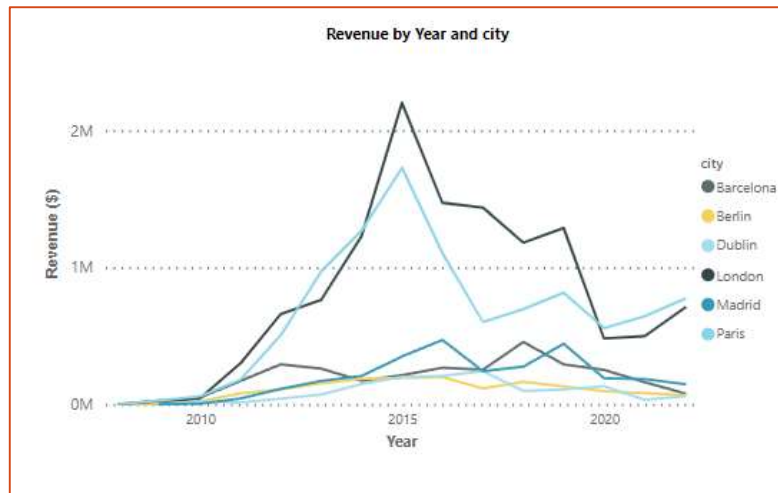
- Choice and design of the chart:
 - We have used: Line and Clustered Column Chart for our visualization. This is great choice when we are visualizing two types of data over same time-frame (X-axis).
 - In this chart we have leveraged the advantage of using 2 Y-axis to plot 2 different measures, the revenue and the average price of rooms.
 - We have chosen the revenue made to be represented in Vertical Bar charts because the values are clustered over the categories of different year.
 - Whereas, the line chart represented a continuous variation of the average price of the room over time.
 - Thus, the combination of the bar and line chart helps us to visualize the relation between the two attributes over same time frame.
 - Both the Y-axes start from the value zero since we were focusing to represent the magnitude of the change. Whereas, the X-axis start from year 2008 to 2022, i.e., the window from when Air BNB started operating to the current date.
 - We chose a subtle coloring for our graph because the bar graph takes up the majority of the area and the background of the graph, so we used a soft cyan color that soothes the viewer's eye. Whereas the line graph takes up very little

surface area and is in the foreground, we have colored it black to contrast with the background and draw the viewer's attention.

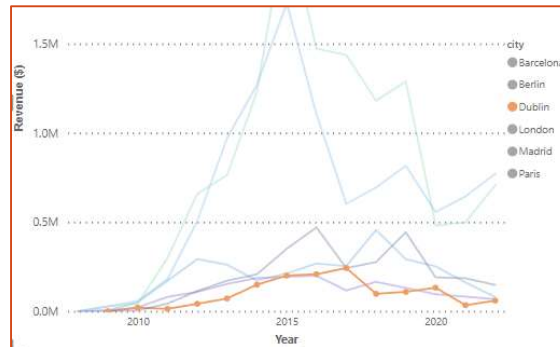
- We have removed the grid lines such that the graph remains minimal in design yet the trend of correlation will also be understood. Since Power BI graphs are responsive, we can hover over the graph to get the exact values as well.
 - We have labelled the axes with relevant names and prominent fonts such that it's easily readable by the users.
 - We have introduced a legend which shows the color used in the graph vs the attribute it represents.
- Findings from the chart:
 - We can see that the average price of the room, co-relates with the total revenue generated in certain patterns. We have anomaly from 2008-2010 maybe because the start-up was just founded and it took them some time to come-up with stable business plan. After that from 2011-2016, Air BNB has shown incredible growth in terms of revenue and adjacently we notice the increase in room prices as well.
 - From 2016-2018 we see a drastic fall of the business and also the average room price dropped from \$537.5 to \$ 408.01. Even after that we see a correlated trend till the year 2022.
 - Key Narrative:
 - Sum of price dropped from 13974896 to 7467848 during its steepest decline between 2018 and 2022.
 - Sum of price (63,251.27% increase) and Average of price (108.07% increase) both trended up between 2008 and 2022.

2. Covid-19 had jeopardized the hotel booking business, was Air BNB affected as well?

- Revenue, which means the total sales made, was calculated by applying the 'sum' function on the room prices of the Air BNB hotels booked.



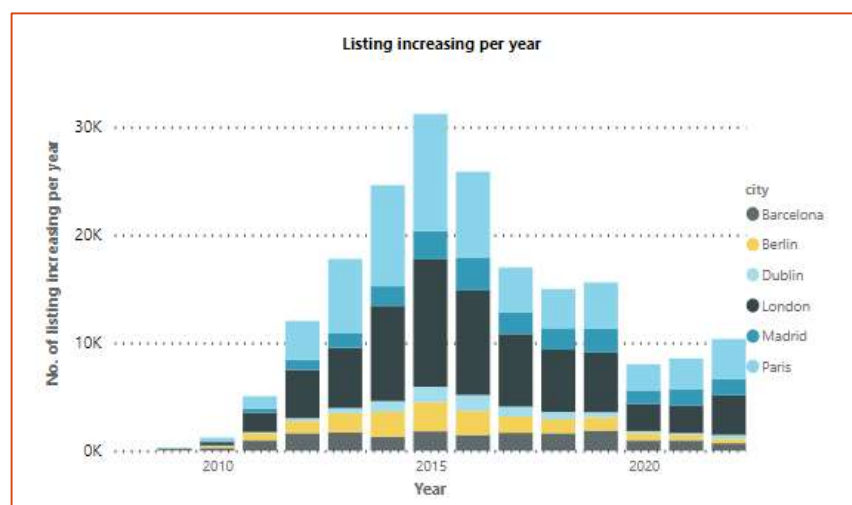
- Choice and design of the chart:
 - We have used Line chart for our graphical representation. Since we are trying to represent the revenue earned, which is a continuous variable, thus, the line chart would have been a suitable option.
 - In X-axis we have plotted the 'time' and in the Y-axis we have plotted the 'revenue' which is total sales made.
 - We have selected 6 European cities, namely Barcelona, Berlin, Dublin, London, Madrid and Paris to observe the pattern. We did not select a lot of cities because here we are trying to find a specific pattern, including a lot of lines in a chart will increase unnecessary noise.
 - We have used 6 different colors to represent the individual cities. And we have introduced legend for the ease of recognition.
- Findings from the chart:
 - We can see that 2015 was the golden year for Air BNB in terms of the total revenue generated. However, after that we see a slide in trend till 2017 and again started recovering till 2019, which was when covid-19 came into the picture.
 - Following 2019 we can see quite clearly that most of the cities have suffered massive loss in terms of revenue. Year 2020 was the year of trench and most of the cities have faced the steep fall in revenue.
 - Did all the cities face the loss in the year 2020? No, Dublin seems to be an exception. We have leveraged the responsive graphical analysis property of Power BI to see the performance trend of this city more clearly.



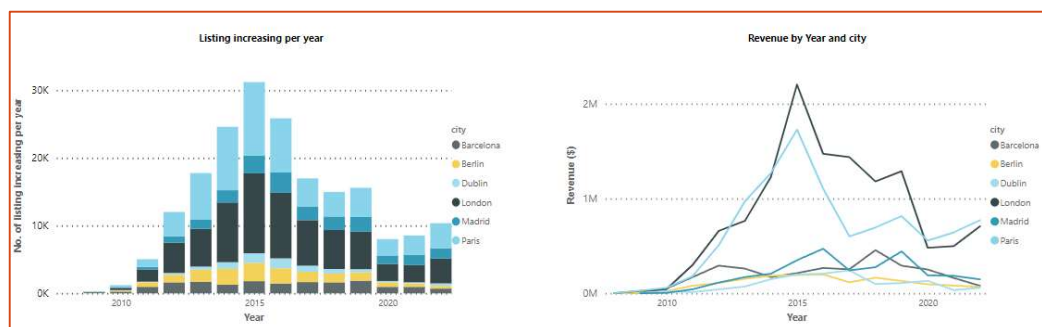
- Interesting Narrative:
 - Madrid experienced the longest period of growth in Revenue (+34549) between 2012 and 2022.
 - Count of id for city London was trending up between 2008 and 2017 with a rise of 6,707 but had a significant change in trend and dropped by 2,094 starting 2018.

3. London and Paris contribute to the main listings of Air BNB, but do they contribute to the revenue generated as well?

- Listings are calculated by the count of the unique hotel ids across the Air BNBS of different locations.



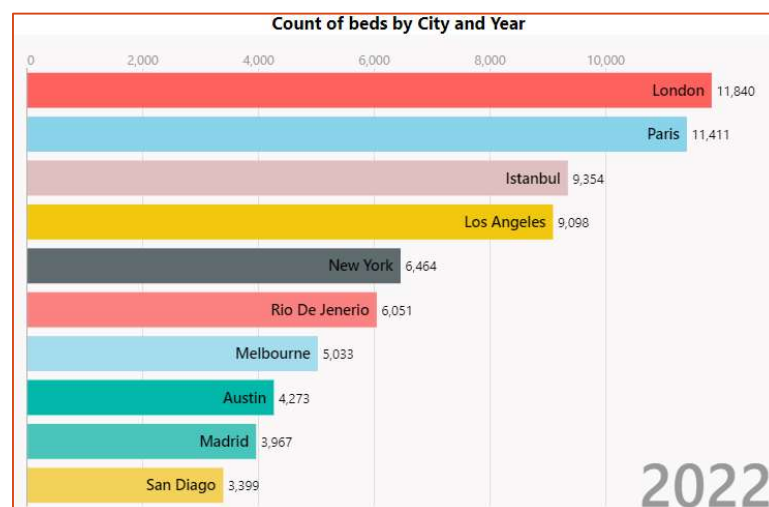
- Choice and design of the chart:
 - We have decided to make use of the Stacked Column Chart for our third graphical plot. The reason behind our choice is, it is a fantastic representation when we compare total values across different categories, here it's cities and the year.
 - In X-axis we have represented the Years and, in the Y-axis, we have plotted the listings made. And the bar charts are stacked by the magnitude of contribution made by different cities.
 - We have selected 6 European cities, namely Barcelona, Berlin, Dublin, London, Madrid and Paris to observe the pattern and also such that we can draw a relation with the previous graph made (Revenue by year and city).
 - We have used contrasting color theme, to make the key players of the graph (London and Paris) distinct.
- Findings from the chart:
 - London and Paris have contributed the most in terms of listings of new Air BNBS.
 - The number of listings in Air BNB across the selected cities have steadily increased from 2008 to 2015, however, it took a dip afterwards till 2018 and improved a bit in 2019. After that we see a dramatic fall of the listings in 2020 maybe due to the turmoil and stringent restrictions of Covid-19. Following years see a gradual increase trend in the listings.
 - Does the number of listings correlate with the revenue generated? (2nd graph of this project)?
 - According to our study, yes. We see a very interesting crest and trough pattern in our listing graph which matches to great extent with our line graph chart that represented the revenue generated.
 - Another interesting find is that, from the stacked graph we see that London and Paris have contributed the most in the number of listings, followed by Madrid, even we in the revenue contribution chart we have observed that these 3 cities have played a significant role in deciding the magnitude of the revenue generated.



- Interesting Narrative:
 - Listings for London started trending down on 2018, falling by 36.37% (2,094) in 4 years.
 - Listings for Paris started trending down on 2017, falling by 10.65% (441) in 5 years.

4. Istanbul's heroic run in the race for accommodations played by different cities around the world!

- The most granular way to find the magnitude of accommodations, is neither number of hotels nor the number of rooms. The count of beds is the building block of how many people can Air BNB accommodate in a city.
- Thus, to we have used the count function on the 'beds' to find the total count of beds in a particular city.



- Choice and design of the chart:
 - We have introduced Bar Chart Race to see how different cities have increased their accommodation power from 2008 to 2022. This is also a very interesting way to represent data where we can see how different category performed in the leader board specially when the competition is intense.
 - Since we have long labels (city names), to avoid collisions, we have chosen horizontal bar chart instead of vertical. Also, it's easier to spot the changes of the magnitude when a bar grows in horizontal way.

- Since it's an animated chart, which will be later explained in our presentation. We could not add all the time frames in our figure, so we have displayed the final standings only.
- Findings from the chart:
 - New York led the race from 2008 to 2012, after that they were overtaken by incredible growth of London's accommodations.
 - After 2012 to 2022, the race was mostly dominated by London and Paris. Paris provided the most accommodations from 2013-2016 and after that London took the lead till 2022 except the year 2021 when Los Angeles was at top.
 - Istanbul has had a phenomenal run. They were not on the leadership team until 2020. After that, they entered the scene and quickly rose to third place by the end of 2022, with a total of 9354 beds.
 - Interesting Narrative
 - In this graph we have seen how London and Paris have improved the number of their accommodations over time, maybe due to their vast capacity they have been able to be the major two European cities in terms of revenue generation (which can be observed from our second graph).
 - Because the most listings were added in London and Paris over time, as shown in our third stacked bar chart, the increase in the number of beds catches up for the two cities.

Conclusion

The process of plotting one graph after another and revealing intriguing patterns on how one feature affects another made this data visualization project quite engaging. A difficult aspect of our project was the bar chart race, which we completed by installing the Microsoft "Animated Bar Chart Race" plug-in.

We created all of our visualizations using Power BI, a potent tool for producing dynamic graphical charts. By filtering in and out of various cities, the Power BI idea of "slicers" enables us to quickly examine the patterns of the graphs.

Although the journey has been very intriguing; however, we have felt that if there had been some extra features, we could have performed even better visualization.

- I intended to add vivid color to the top 3 donors in the bar chart animation, which are London, Paris, and Istanbul. The remaining cities should have been colored in grey or gentle pastel tones. In this instance, too much color confuses the viewer's eye and might occasionally highlight unfavorable details.
- We had no control over the legend name in our first line and clustered column chart. Because we performed a sum operation on the 'price' column, revenue is represented as 'Sum of Price' in our graph. It would

have been more appropriate if we could have renamed it 'Revenue generated.'

- In our second stacked column chart, we could not control the individual color of the cities. Thus, due to auto color correction, the palate of Dublin and Paris remains similar. We wanted to add a distinctly different color like soft orange or dusty pink to make a more prominent distinction.

Aside from that, we attempted to weave a story using a very minimal graph design about how Air BNB grew in several major cities over time. And what were the key factors that determined their revenue growth? We've also seen the impact of covid on the organization. As we see an increasing trend in both revenue and listings for Air BNB in different cities, I assume Air BNB will gradually return to its peak and we will see more interesting data patterns in the coming years.

References:

- Data is gathered from: <http://insideairbnb.com/get-the-data>.
- We have learned the basics of Microsoft Power BI from: <https://app.datacamp.com/learn/courses/introduction-to-power-bi>
- From this video we understood how to install the plug-in for Bar Chart Animation in Power BI : <https://youtu.be/XiYliDFuPRQ>