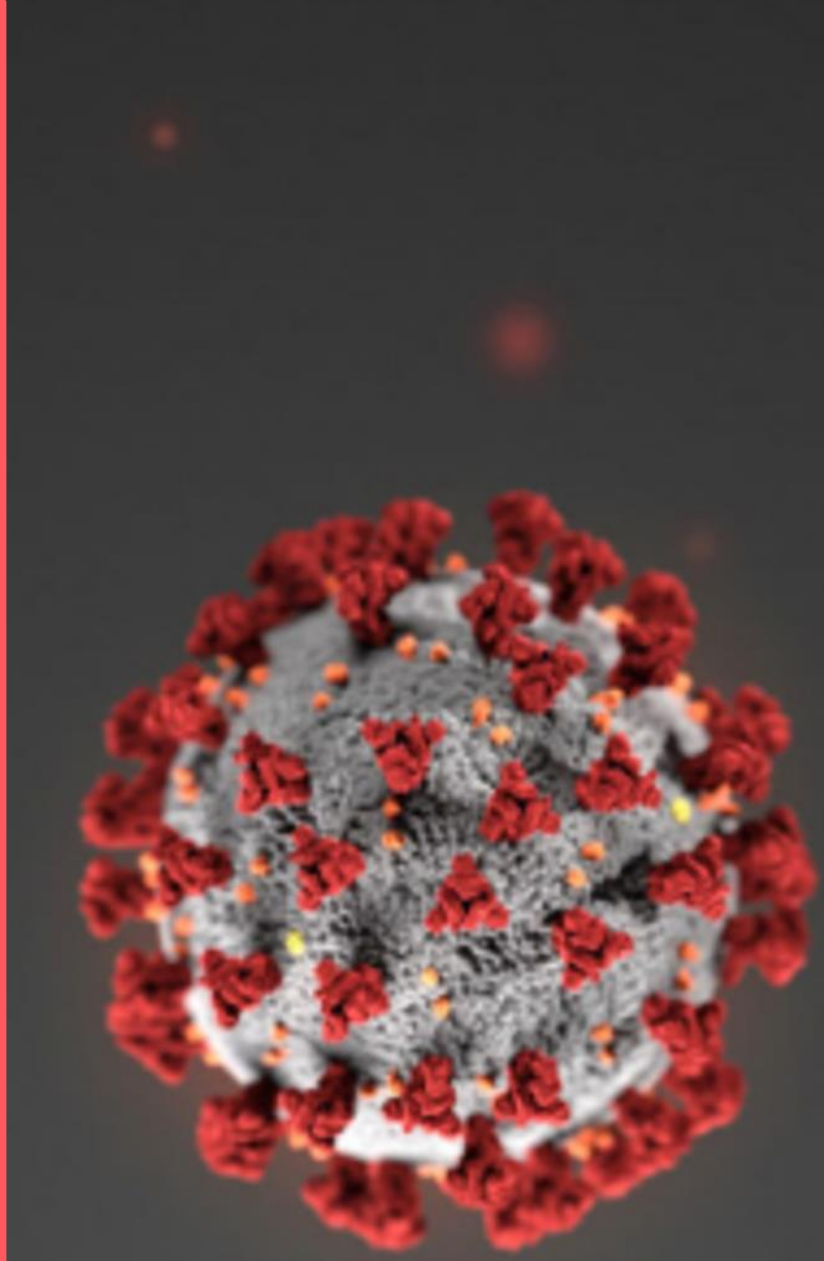
Airbnb Listings in Toronto: The Impact of COVID-19

Group Assignment: Big Data Technology

Group 2:
Aravind Kakarala
Daniel Adam Cebula
Dong Hyun Ko
Cynthia Fung
Juraj Polakovic
Su Wang



April 2020

Contents

Members 3
GROUP 2 3
Introduction 3
BACKGROUND 3
OBJECTIVE 3
Data Preparation 4
DATA SOURCE 4
DATA PROCUREMENT 4
Analysis 5
TOOLS 5
LISTINGS 5
LOCATION 6
BOOKINGS 6
LISTING PRICES 8
HOST ACCOUNT CREATION DATE AND PRICE 9
REGRESSION ANALYSIS-PREDICTING PRICES 9
Conclusion 10
Appendix 11
References 12

Members

GROUP 2

Aravind Kakarala - akakaral@uwaterloo.ca
Daniel Adam Cebula - dacebula@uwaterloo.ca
Dong Hyun Ko - dhko@uwaterloo.ca
Cynthia Fung - c27fung@uwaterloo.ca
Juraj Polakovic - jpolakov@uwaterloo.ca
Su Wang - s399wang@uwaterloo.ca

Introduction

BACKGROUND

Airbnb revolutionized the tourism and hospitality industry by taking a unique approach towards lodging. Founded in 2008, the company grew quickly with over half a billion guests that have stayed in an Airbnb home and being active in over 200 countries and 81,000 cities **(Curry)**. Unfortunately, Airbnb was no exception to unforgiving wreckage caused by the onset of a global pandemic from the relentless coronavirus. With the onslaught of travel plan cancellations, the unpredictable travel restrictions, and the evolving variants of the virus, COVID-19 has proven to be Airbnb's biggest challenge yet. In the early stages of the pandemic, Chesky, one of the three founders, wrote in a message to his staff "Travel in the new world will look different, and we need to evolve Airbnb accordingly…" **(Halpern)**.

OBJECTIVE

Brian Chesky, the CEO of Airbnb stated in an article published in Jan'2021 that Airbnb would need to evolve and adjust to a post pandemic world where travel would no longer be the same **(Halpern)**. It has been discussed that Airbnb has been made a model for resilience that has survived the related economic shock of the pandemic. Was Chesky and his fellow founders able to evolve Airbnb successfully to survive through a global pandemic? Group 2 would like to perform preliminary pandemic analysis by comparing Airbnb listings in Toronto, pre and post COVID-19 pandemic. The comparative analysis in which we would like to present our findings would entail a point in time comparison of months January and February in the year 2020 and 2021. To perform our analysis, we will be leveraging several toolsets that we have not used before.

Data Preparation

DATA SOURCE

Inside Airbnb - *link*

- Inside Airbnb provides an independent, non-commercial set of tools and data allowing a user to explore their platforms usage in cities around the world.
- By analyzing publicly available information about a city's Airbnb listings, this tool provides filters and key metrics to inform users how their usage competes with the residential housing market.
- Data used by Inside Airbnb is source from publicly available information from the Airbnb website.

DATA PROCUREMENT

Tools

For the procurement process, as previously mentioned, toolsets that are new to Group 2 were used. Koalas, a Pandas API for PySpark for DataFrame manipulations was leveraged instead of PySpark. SQL Hive in Azure Databricks for ELT was also used accompanied by connections to Microsoft Power BI.

For further information on the tools used for this analysis, please reference the appendix.

Extracting and Loading Data

After data was extracted from Inside Airbnb in .csv format, Databricks Community Edition was utilized to load the data files.

For more details on this step, please refer to the following published notebook: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/4473477877994595/4138564470759536/latest.html

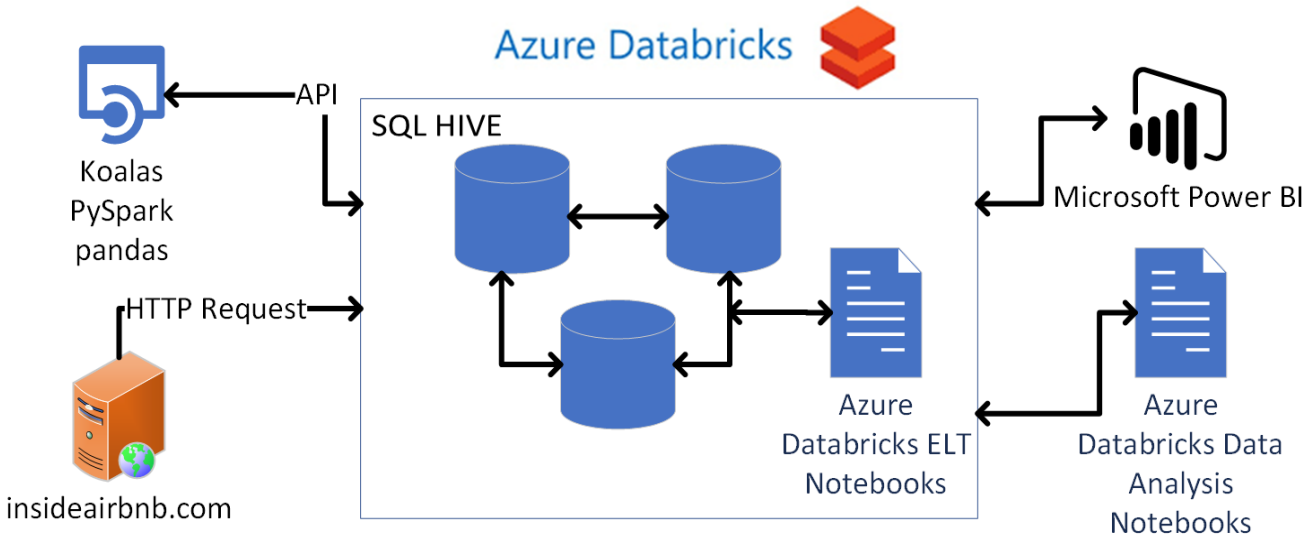
Data Cleansing

The initial state of the data required cleansing before it could be used for analysis. Throughout the data, many of the variables were categorical which needed to be converted to a numerical value to be fed into an analytical model. To ensure proper formatting was applied, column checks were also made along the way which facilitated data merges. For efficiency, any data tables not relevant or required for the analysis were dropped.

For more details on this step, please refer to the following published notebook: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871553/4138564470759536/latest.html

Preliminary Analysis
Prior to our analysis, a preliminary overview of the data was performed. This allowed a better understanding of the data to perform more intricate analyses and what can be explored.
For more details on this step, please refer to the following published notebook: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871594/4138564470759536/latest.html

For better understanding please find below a flow diagram of the data procurement process.



Analysis

TOOLS

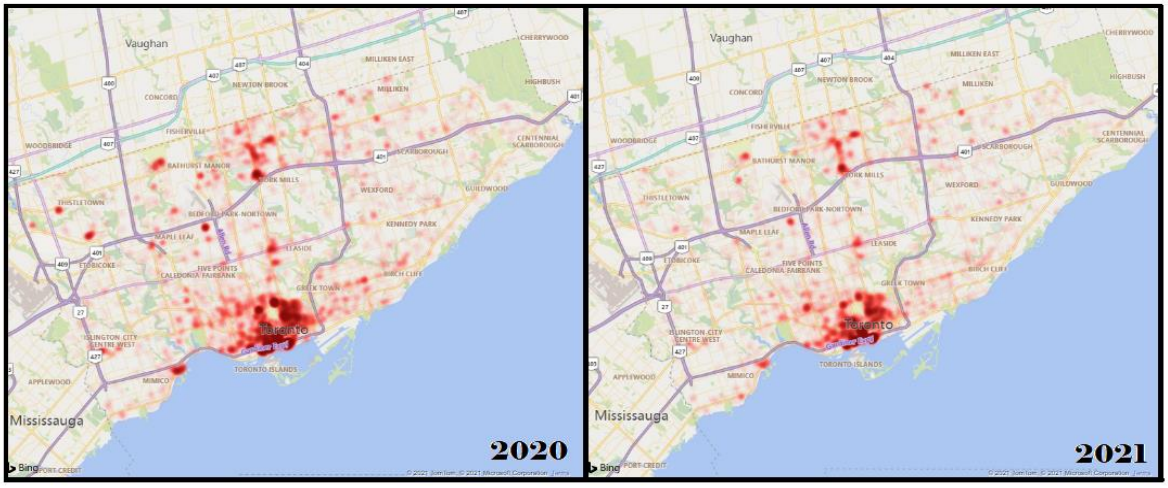
For the analytics stage, machine learning was applied using MLlib. For further information on this tool please reference the following link: https://spark.apache.org/docs/latest/api/python/reference/pyspark.ml.html
A connection was also established to Azure Databricks instance to Power BI Desktop to generate visuals to better interpret our findings.

For more details on how the analysis was performed, please find all notebooks in the appendix.

LISTINGS

Airbnb listings from across Toronto were studied. To have a pre pandemic picture of Airbnb activity, the months of January and February 2020 were selected for analysis. Post pandemic activity included listing from January and February 2021. Data was then visualized using Power BI creating a heat map.

Airbnb Listings Heatmap
2020 vs 2021

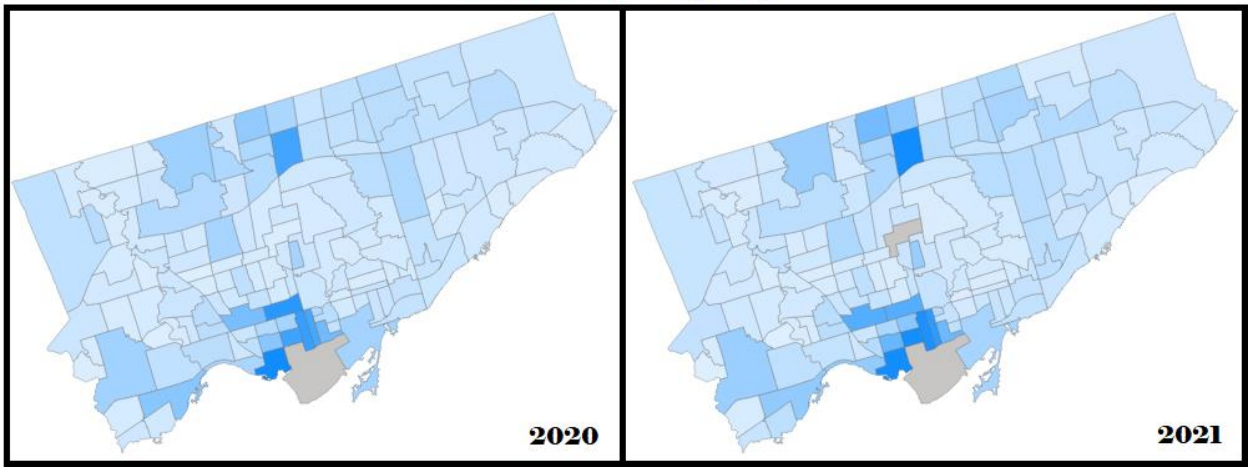


As we can see in the heatmap above, there were more Airbnb listings in 2020 than 2021, as expected. Surprisingly, even with travel bans, cancelled events, and many other restrictions affecting the tourism and hospitality industry originating from COVID-19, there is still a decent amount of Airbnb listings throughout Toronto.

LOCATION

To further the analysis, listings were also plotted on a base map. Higher concentrations of Airbnb listings were located in the Financial District, Fashion District, Distillery district, Cabbage town, and seemed to stay consistent before and after the onset of the pandemic.

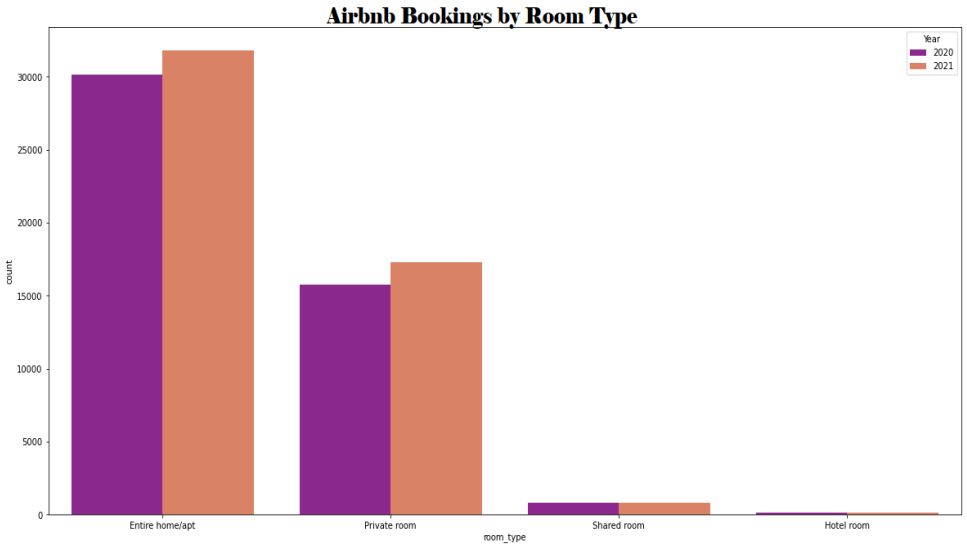
Airbnb Listings Shape Map



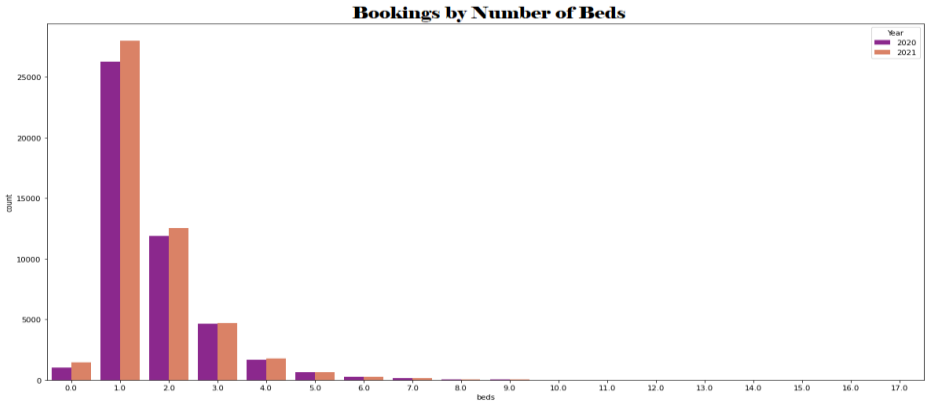
BOOKINGS

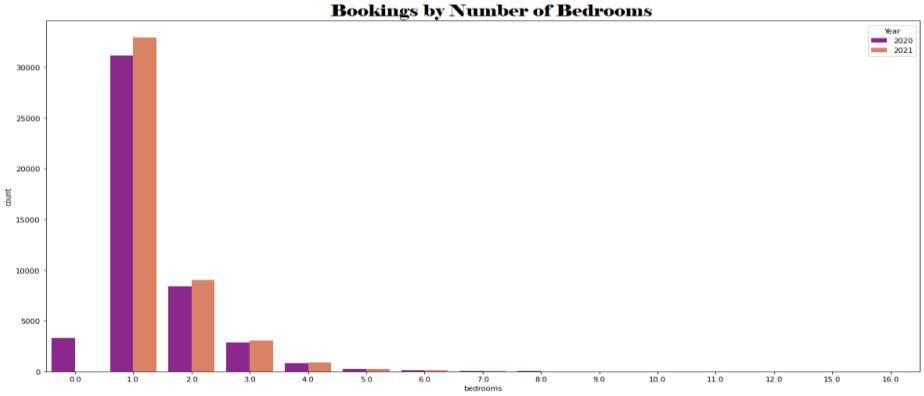
To further the analysis, the investigation of bookings was pursued. It was interesting that the data revealed, although 2021 might have had fewer listings, there were more bookings in 2021 than in 2020. In 2021 there were 50,103 bookings, which is a 6.8% increase compared to the 46,922 bookings made in 2020. The data also revealed that 14,437 of the bookings had booked a listing in both years.

Bookings were categorized by room type, number of beds, and number of bedrooms. As we can see to the right, entire homes and private rooms were preferred. Given the nature of the pandemic, the prominent preference to more isolated accommodations are understandable. The data demonstrates that accommodations where an individuals or group can isolate themselves was much more preferred. This would allow travelers to self-isolate for 14 days given the government's travel restrictions put in place to prevent the spread of COVID-19.



Furthermore, the same effects of the pandemic can be shown when the data is categorized by number of beds and bedrooms. Significantly more bookings are made for one bedroom and/or one bed accommodations opposed to any listing that includes multiple beds or multiple bedrooms.





LISTING PRICES

Based on the previous findings, the supply of Airbnb listings in Toronto had decreased amid the pandemic but the demand was still there. A deeper investigation was made on how this affected price listings.

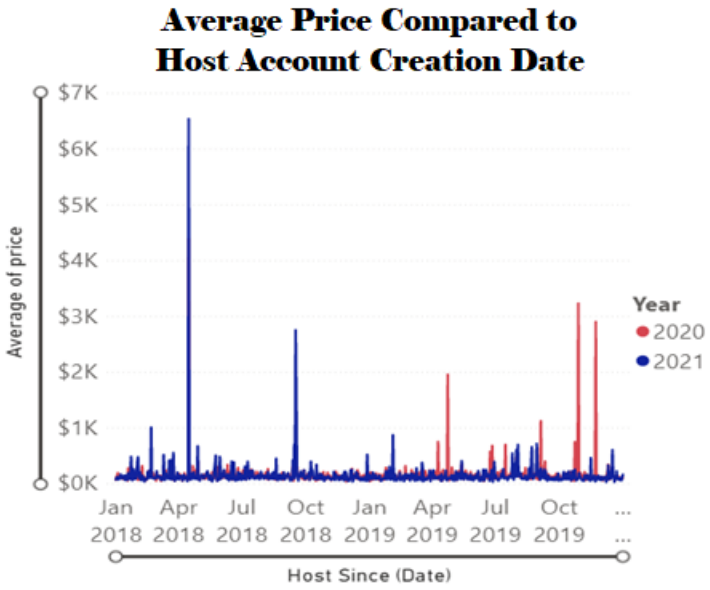
To start, listings price change was examined by performing a comparison between listing prices before and after the COVID-19 pandemic. In 2020, the average price of listings was $147.51 CAD per night and in 2021 the average price of listings was $136.52 CAD per night. To further analyze the decrease in average listing price, prices of the same listings were compared between 2020 and 2021. In total, 32,342 listings had lower prices in 2021 compared to their 2020 prices. However, there were 25,141 listings had higher prices in 2021.

To further investigate the decreasing prices, the distribution of price change was plotted into a bar chart, as seen above to the right. As the chart reveals, more than 10,000 listings have a small price change in 2021. A small price change was defined as a change of less than $10. Subsequently, most price decreases varied between the ranges of $10 to $20 with approximately 6,000 listings, then followed by changes of $20 to $30 and $30 to $40, with approximately 4,000 and 2,000 listings, respectively. It was noteworthy, that although the number of listings seemed to decrease as price change range increased, there were around 3,000 listings that had over $100 decrease, most likely as a result from the COVID-19 pandemic.

In addition, increasing price changes was also plotted and analyzed as seen to the right. There were over 12,000 listings that increase by at least $10 in 2021. This was followed by approximately 4,000 listings with increases between $10 and $20. In relation to decreasing price changes, increase in Toronto Airbnb price listings was not a common observation.

HOST ACCOUNT CREATION DATE AND PRICE

While looking at the average listing price, host account creation date was incorporated into the analysis. Adding this element would reveal whether more experienced host, an older host account, had any correlation with the average listing price in 2020 compared to 2021. Once the data was plotted into a line chart, as seen below, it revealed that there were no real correlations between host account creation date and average listing price.



REGRESSION ANALYSIS-PREDICTING PRICES

As mention above, machine learning was applied using MLlib. Data was used in a regression analysis to predict rental prices.

For more details on this step, please refer to the following published notebook: https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2492692934235656/3846416337159354/1930296091766500/latest.html

It was noted earlier on that the average listing price in 2020 was $147.51 and had decrease post pandemic to $136.52 in 2021. Both years' worth of data combined, the average listing price came to $141.40. Using data from both years, it was put into a vector type for Spark MLLib and Linear Regression estimator was used to build the first model.

After using average price as the prediction column, the RMSE was 347.44. Based on this, improvements were made by adding additional features to the model using One-Hot Encoding for categorical variables was used which brought RMSE to 331.31 and an R^2 of 0.09.

Once again, improvements were made by using a logarithmic scale. This modification rendered an RMSE of 336.37 and an R^2 of 0.06.

Conclusions

We find that Airbnb adapted well to the changing circumstances in travel due to the Global Pandemic. Key findings and conclusions from the comparative analysis of Airbnb Toronto data for the periods Jan-Feb'2020 and Jan-Feb'2021 are as follows:

1) There were more Airbnb listings for Toronto during Jan-Feb 2020 compared to Jan-Feb 2020.
2) There was no major change in the number of listings by neighborhood. The popular neighborhoods by concentration of Airbnb listings were the same between the two years (Financial District, Fashion District, Distillery district, Cabbage town etc.)
3) It was interesting that the data revealed, although 2021 might have had fewer listings, there were more bookings in 2021 than in 2020. In 2021 there were 50,103 bookings, which is a 6.8% increase compared to the 46,922 bookings made in 2020.
4) In 2021, Bookings were more popular with accommodations that constituted entire homes or listed private rooms. Given the nature of the pandemic, the preference to more isolated accommodations is understandable.
5) Significantly more bookings are made for one bedroom and/or one bed accommodations opposed to any listing that includes multiple beds or multiple bedrooms.
6) There was a small price reduction in the listings during the Jan-Feb 2021 period compared to Jan-Feb 2020. During the first two months of 2020, the average price of listings was $147.51 CAD per night and in 2021 the average price of listings was $136.52 CAD per night.
7) Majority of the decrease in the price for listings was in the $10 to $20 range.

Airbnb was able to take advantage of the changing travel preferences of the consumer that included a shift away from large hotels with their numerous shared spaces and elevators and to more private accommodations. As the consumer preferences for travel accommodations continue to evolve, we find that Airbnb able to respond in a nimble manner to cater to the changing needs and live up to its status as a "Disruptor" in the Travel and Accommodation industry.

Appendix

GitHub Repository
https://github.com/cebulada/Data-Science-Group-2

Reference Material
https://community.cloud.databricks.com/
https://koalas.readthedocs.io/en/latest/index.html
https://spark.apache.org/docs/latest/api/python/reference/pyspark.ml.html
https://docs.microsoft.com/en-us/azure/databricks/integrations/bi/power-bi
https://docs.databricks.com/integrations/bi/power-bi.html

Published Databricks Notebooks
Initial Data Loading
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/4473477877994595/4138564470759536/latest.html

Clean the Data
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871553/4138564470759536/latest.html

Preliminary Data Analysis
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871594/4138564470759536/latest.html

Price Change Analysis
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/1044481630572696/3266341553662997/2373503380995683/latest.htm

Data Analysis P1 notebook - data checks, basic exploratory data analysis, plotting
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2492692934235656/576067449046085/1930296091766500/latest.html

Data Analysis P2 notebook - Data Analysis - Lin Regression/ML/Log, Price prediction
https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2492692934235656/3846416337159354/1930296091766500/latest.html

References

Curry, David. *Airbnb Revenue and Usage Statistics (2021)*. 28 February 2021. 09 April 2021.

Halpern, Ashlea. *How Airbnb Could Change After the Pandemic–For the Better*. 20 May 2020. 10 April 2021.