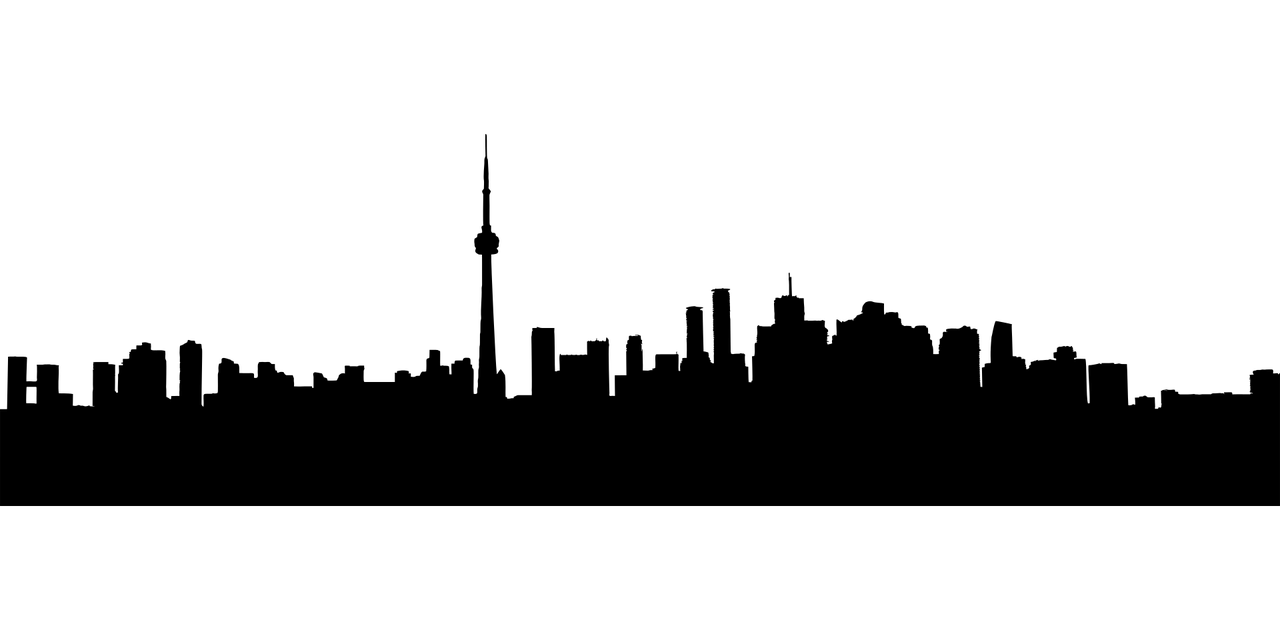


**Airbnb Listings in Toronto:**

**The Impact of COVID-19**



April 2020

**Group 2:**

Aravind Kakarala

Daniel Adam Cebula

Dong Hyun Ko

Cynthia Fung

Juraj Polakovic

Su Wang

Group Assignment: Big Data Technology

Contents

[Members 3](#_Toc68910447)

[GROUP 2 3](#_Toc68910448)

[Introduction 3](#_Toc68910455)

[BACKGROUND 3](#_Toc68910456)

[OBJECTIVE 3](#_Toc68910458)

[Data Preparation 4](#_Toc68910460)

[DATA SOURCE 4](#_Toc68910461)

[DATA PROCUREMENT 4](#_Toc68910462)

[Data Cleansing 4](#_Toc68910469)

[Analysis 5](#_Toc68910474)

[BACKGROUND 5](#_Toc68910475)

[Conclusion 5](#_Toc68910478)

[References 6](#_Toc68910479)

|  |
| --- |
| MembersGROUP 2Aravind Kakarala - akakaral@uwaterloo.caDaniel Adam Cebula - dacebula@uwaterloo.caDong Hyun Ko - dhko@uwaterloo.caCynthia Fung - c27fung@uwaterloo.caJuraj Polakovic - jpolakov@uwaterloo.caSu Wang - s399wang@uwaterloo.ca |
| IntroductionBACKGROUNDAirbnb 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 having 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)OBJECTIVEAs Chesky stated, Airbnb would need to evolve accordingly to adjust to post pandemic world where travel would no longer be the same. 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 PreparationDATA SOURCE **Inside Airbnb -** [*link*](http://insideairbnb.com/get-the-data.html)   * 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 PROCUREMENTToolsFor the procurement process, as previously mentioned, toolsets that are new to us were used. Koalas, a Pandas API for PySpark for DataFrame manipulations was leveraged instead of PySpark.For further information on Koalas, please reference the following link:<https://koalas.readthedocs.io/en/latest/index.html>Extracting and Loading DataAfter 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 CleansingThe 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>Prior to our analysis, a preliminary analysis was performed.<https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871594/4138564470759536/latest.html> |
| AnalysisToolsFor 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. |
| Conclusion |
| 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. |