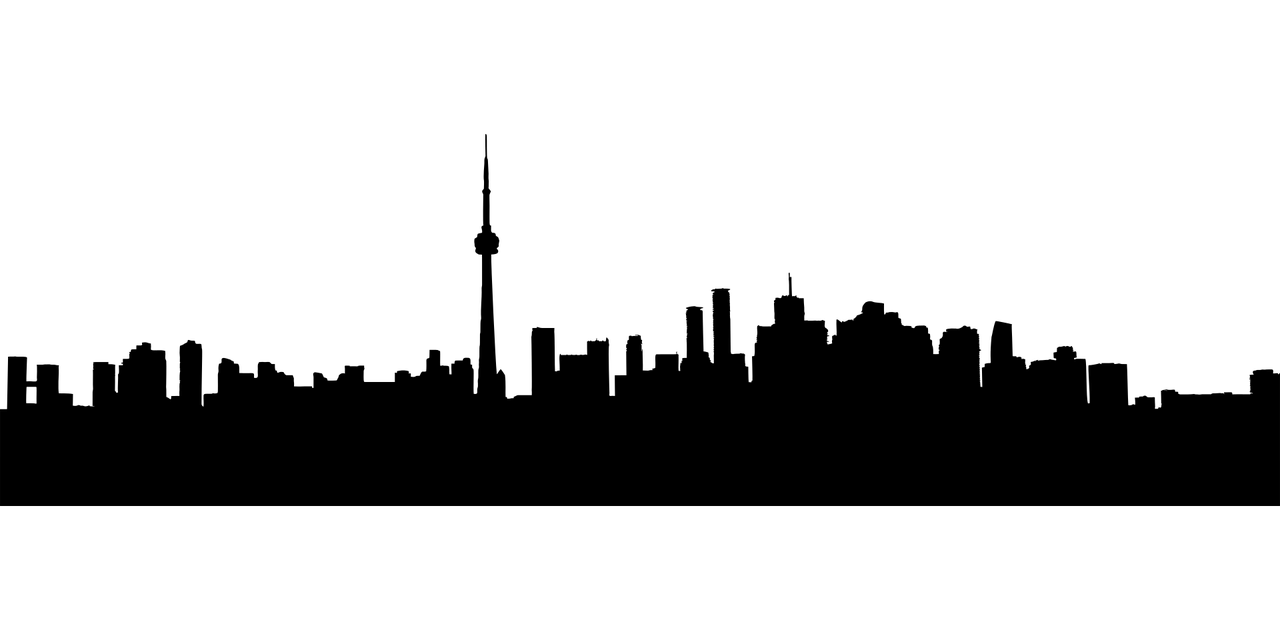


**Airbnb Listings in Toronto:**

**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

Table of Contents

**Type chapter title (level 1)1**

Type chapter title (level 2)2

Type chapter title (level 3)3

**Type chapter title (level 1)4**

Type chapter title (level 2)5

Type chapter title (level 3)6

# Objectives

Overview of Topic

AirBnB was being discussed as a model for resilience that survived the Pandemic related economic shock. We propose to analyze and compare AirBnB Listings in Toronto pre and post COVID (Jan and Feb 2020 to Jan and Feb 2021) and present our findings.

Description of Data Source

Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world.

By analyzing publicly available information about a city's Airbnb's listings, Inside Airbnb provides filters and key metrics so you can see how Airbnb is being used to compete with the residential housing market.

The data behind the Inside Airbnb site is sourced from publicly available information from the Airbnb site.

**Data Se**t: <http://insideairbnb.com/get-the-data.html>

# 

# Data Preparation

Data Load / Cleaning

The nature of this dataset required us to do some cleaning. There are a lot of categorical variables as listed above, which we converted to numerical values to feed the analytical model.

1. Initial Data Load: <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/4473477877994595/4138564470759536/latest.html>
2. Clean the Data: <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871553/4138564470759536/latest.html>
3. Preliminary Data Analysis: <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/2514828024743454/1148597687871594/4138564470759536/latest.html>

Data Analysis

# 

1. We will leverage Databricks Community Edition and load in the data as .csv files

<https://community.cloud.databricks.com>

1. We will leverage Koalas, a Pandas API for PySpark for DataFrame manipulations instead of PySpark

<https://koalas.readthedocs.io/en/latest/index.html>

1. We will leverage MLlib for machine learning

<https://spark.apache.org/docs/latest/api/python/reference/pyspark.ml.html>

1. We will connect Azure Databricks instance to Power BI Desktop to generate visuals for data analysis

https://docs.microsoft.com/en-us/azure/databricks/integrations/bi/power-bi) and [link](<https://docs.databricks.com/integrations/bi/power-bi.html>

Results

# Conclusions

**References**

1. <https://www.forbes.com/sites/deniselyohn/2020/11/10/how-airbnb-survived-the-pandemic--and-how-you-can-too/?sh=87544a293845>
2. <http://insideairbnb.com/get-the-data.html>