

# CMPT 732 Project Proposal

Team (LZY): Brian Zhou (bzz1), Hongying Yue (hya134), Yiran Lu (yla826)

## Subject

Airbnb Big Data Mining

## Summary

Our project aims to investigate and analyze Airbnb's listing data and the broader housing market within the Greater Vancouver Area, and explore the relationships between Airbnb listings' pricing, location, durations of stay, and its profitability in different neighborhoods. Additionally, we will leverage Vancouver's rental data statistics to assess the impact of Airbnb on the affordability of housing in various neighborhoods.

## Objectives

- Explore the relationship between different elements of an Airbnb listing and its profitability
- Assess the impact of Airbnb on the rental prices of different neighborhoods
- Visualize our findings in a way that's interactive and easy to understand

## Dataset

- [Inside Airbnb](http://insideairbnb.com/get-the-data/) (<http://insideairbnb.com/get-the-data/>)
- [CMHC Housing Market Information](https://www03.cmhc-schl.gc.ca/hmip-pimh/) (<https://www03.cmhc-schl.gc.ca/hmip-pimh/>)
- [REBGV Market Watch](https://www.rebgv.org/market-watch/) (<https://www.rebgv.org/market-watch/>)
- ... and more

## Technology

- We plan to utilize PySpark as our main data analysis tool.
- We'll likely utilize cloud computing services like Amazon S3 and EC2/EMR to help process our data.
- D3.js is chosen as our data visualization framework.
- More technologies will be involved if necessary as we take on our project.

## Interesting questions as a glimpse

- If Airbnb seeks for market expansion in the Greater Vancouver Area, which area will be the most attractive and profitable?
- In order to help the house owners improve the attractiveness of their house, what suggestions could Airbnb give?
- How significant are the seasonal impacts on the renting market?
- ... and more