Alek Carlson & Adam Tassabehji

Professor Milan Miric

DSO 574 - Using Big Data: Challenges and Opportunities

17 February 2021

# *Airbnb Project Part I – Methodology Summary*

## Overview:

The purpose of our analysis was to explore the impact of COVID-19 on the Airbnb market and the wider home-sharing economy in Los Angeles. Our approach was to examine the volume of listings, average minimum number of nights required, and average listing price for both the entirety of the Los Angeles listings and those listings belonging to the 5 most popular neighborhoods (popular referring to the number of listings in each neighborhood).

## Data:

The *Combined Listing Data [Summary].csv* file was the base dataset used for our analysis.

## Definitions and Variables:

March 2020 was defined as the time point in which Covid-19 emerged in Los Angeles for our analysis. To ensure our data is recent, relevant, and accurate; we filtered to only retain data from March 2019 and later. Pre covid months were defined as prior to March 2020, and post covid months were considered to be March 2020 and beyond.

The primary variables we examined were:

**Volume of Listings:** Defined as the count of id records

**Mean # of Minimum Nights:** Defined as the average value of minimum nights required

**Mean Price:** Defined as the average of listing price

From a business perspective, these variables are highly indicative and insightful of customer behavior on the platform. From a technical standpoint, the fields also had a very limited number of missing values and a healthy degree of variance for more concrete analysis.

## Slide One: Macro-level Analysis

In order to get an accurate snapshot of the pandemic's impact, our approach was to compare the months July to September of 2019 and July to September 2020. The reasoning for selecting these months are that these are summer months and should be some of the most popular months for Airbnb stays in Los Angeles, exploratory data analysis confirmed this.

This was done to address certain factors that might've negatively impacted the accuracy of our analysis such as:

- Initial Uncertainty: The first few months of the pandemic were completely unchartered territory for all Airbnb stakeholders. Little was known about the disease at the time, and several travel restrictions as well as lockdowns were introduced. Data from these months is unique and not likely to be truly reflective of the Airbnb market during Covid.

- Seasonality : As a popular tourist destination, certain listing variables such as price are greatly influenced by the month or season in which the listing appears on the platform. For instance, it would not be reasonable to compare the winter months to summer months.

- Relevance: The data provided extended five years prior to Covid. When examining various trends over time, we saw that many variables such as listing price, volume, and minimum nights were increasing. This could just be due to the popularity of Airbnb rising. To have more relevant/accurate data, we filtered for just the year prior to covid.

## Slide Two: Top Five Neighborhoods

To examine the data through a more micro level lens, analysis was conducted for the five neighborhoods with the greatest number of Airbnb listings.

These turned out to be Venice, Hollywood, Long Beach, Downtown, and Santa Monica.

The same three variables were analyzed over a time frame from March 2019 onwards. Using data from further beyond this date would increase the risk of unwanted factors such as inflation or regulation affecting our analysis.

## Slide Three: Linear Regressions

To further assess how Covid has changed the Airbnb market, simple linear regression models were built.

It was apparent that the volume of listings decreased dramatically, so instead we decided to focus on analyzing Covid's impact on the average number of minimum nights and mean price.

A binary variable was created to indicate the presence of Covid. All data records that came from a scrape file before March 2020 were assigned a 0, and those beyond were assigned a 1.

The Covid variable was used in a simple linear model to predict minimum nights and mean price. This was done at a macro level for all Los Angeles properties, as well as for the top five neighborhoods. In total 12 models were built.