Alek Carlson & Adam Tassabehji

Professor Milan Miric

DSO 574 - Using Big Data: Challenges and Opportunities

16 March 2021

Airbnb Project Part 2 – Methodology Summary

Overview:

The purpose of our analysis for part two was to discover trends that may provide insight into what the future of home sharing will be in Los Angeles. Our approach was to examine the trends regarding the number of hosts and host types, minimum nights required per stay, monthly reviews, and number of listings.

Data:

The *Combined Listing Data [Detailed].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:

Host Type: Created using “host_is_superhost” as a binary indicator.

Number of New Hosts: Created by aggregating “host_since” variable.

Number of Unique Hosts: Defined as the count of unique “host_id” values.

Mean Number of Minimum Nights: Created from the average value for “minimum_nights”.

Mean Number of Monthly Reviews: Created from the average value for “monthly_reviews”.

Cumulative Reviews Per Month: Aggregation of “monthly_reviews” for top five neighborhoods.

Volume of Host Listings Count: Aggregation of “host_listings_count” variable by neighborhood.

From a business perspective, these variables are highly indicative and insightful of host and user 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: Host Analysis

In order to see how host behavior has changed we wanted to take a look at the proportion of hosts that were categorized as superhosts, the number of new hosts joining the platform, and the number of unique hosts at a given time. We discovered that the proportion of hosts categorized as superhosts was increasing over time, while the number of new hosts joining the platform and the number of unique hosts on the platform were both decreasing. We found that these trends were in place before the pandemic, however, they seemed to intensify following the onset of Covid.

Slide Two: Number of Minimum Nights and Monthly Reviews

To dive more into the listings themselves, we examined how the mean number of minimum nights and the mean number of monthly reviews have changed from pre to post Covid. As expected we saw an increase in the mean number of minimum nights required. We anticipated this because we felt that hosts would try to avoid frequent high cost cleaning fees and that there was a possibility hosts could be trying to capitalize on the remote work trend. We also saw a decrease in the mean number of monthly reviews, which was an indicator that there have been fewer bookings post Covid.

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: Volume of Bookings and Host Listings Count

To further assess how Covid has changed the Airbnb market, we examined the trend of booking volume and total listings volume.

To analyze booking volume, we used the sum of reviews per month as a proxy for booking volume across the top five neighborhoods. When analyzing booking volume, we saw a slight down trend pre Covid that intensified once the pandemic occurred.

When analyzing the total listings volume, we took the total number of host listings across the top five neighborhoods. In this analysis, we saw a steady uptick in listings pre Covid, then a sharp drop immediately following the pandemic, and then high volatility in the months since March 2020.