

Impact of Covid-19 on Tourism in Clark county

Dhruv Relwani
Data 512 UW MSDS
11/10/2021

Motivation/problem statement

As a software engineer, my work was disrupted to a lesser extent during the pandemic and hence I was privileged to continue my livelihood/ income because of the nature of my job; however, I am aware that a majority of the world did not have this privilege and hence the pandemic had a direct impact on the livelihoods of a majority of the people. In this project, I am planning to explore the effect Covid-19 pandemic had on the tourism industry in Clark County, Nevada. I got interested in exploring this question since Clark county (specifically Las Vegas) has one of the biggest tourism based service sectors in the world and this industry is one of the primary employers in the region. This research domain is human-centered since it will help us uncover the details of human-interactions during a world-wide crisis and help us understand how human lives were practically impacted by the trends we see in our Covid-19 data. We will look at this pandemic's impact on the tourism industry through the lens of Airbnb data in Clark county.

Research questions and/or hypotheses

- How were the descriptions of Airbnb listings in Clark county affected by the pandemic?
- How was the income of Airbnb hosts in Clark county impacted by the pandemic?
- Hypothesis: For every 10% increase in the confirmed covid cases in Clark county, we notice a 2% decrease in Airbnb listings in that region.
- Hypothesis: The Airbnb listings in Las Vegas were quicker to bounce back after vaccinations were rolled out, than the listings in other parts of Clark county.

Data used

Inside Airbnb data: The data is a collection of Airbnb data that is publicly available on its website, across multiple cities. The data is aimed at providing a 360-degree insight into Airbnb's presence in a city. Dataset URL: [Get the Data - Inside Airbnb. Adding data to the debate.](#)

This dataset is composed of host listing details, aggregated by city and broken down by individual listings. Users are allowed to have multiple listings, so uniqueness is determined by a numerical identifier with a matching link URL. The dataset is compiled of entries spanning individual cities from Airbnb. The Clark county data has 96 columns and well over 10,270 rows. There is a broad mix of numerical, boolean, character values, web links and percentages across the data. There are some fields which explicitly have zero values and others which appear to use empty values to indicate 'N/A'. Analysis of certain fields will likely need data extrapolation for

aggregation. The location data for the listings is a bit redundant, but it may be due to the data being split from a central source prior to archiving. Other data available includes the listing amenities, various pricing stages, average host response times, listing types, review ratings, availability and geospatial coordinates.

Ethical concern: One ethical concern about accessing housing and income related data about Airbnb hosts and users, could be a breach of their privacy. Hence, sensitive data has been scrubbed to protect the privacy of hosts and Airbnb users in this dataset.

Terms of Use:

- Not allowed to republish the data.
- Attributions in research should include a detailed citation.
- Required to be mindful of calls to the database (scraping not allowed).
- Additional details summarized here: [Data Policies - Inside Airbnb. Adding data to the debate.](#)

Unknowns and dependencies

- Dataset is available for October 2020 to September 2021. If I need access to data at the early stages of the pandemic in early 2020, I will need to request data from the archives by emailing the team responsible for maintaining it. This dependency and the unknown response time from the team could potentially limit my research to the time frame available in the public dataset.
- Completeness of the available data is unknown at this stage, I will learn about it in detail when I start performing my hypothesis tests.

Methodology

1. I will gather my data by downloading the required CSVs from the URL I pasted above and then work in Tableau Prep Builder to combine and clean this data. This would be a quick way for me to verify the data cleaning process while I am completing each step and also for me to have a version history of every step. One other way to clean the data could be to load everything in Pandas dataframes in a notebook and then scrub the tables to prepare them for analysis.
2. I will use t-tests and [ANOVA](#) techniques for my hypothesis tests and to answer my research questions. These techniques are appropriate in this project because they will help me in proving/ disproving some of my research questions sequentially (leading up to a final list of conclusions about the pandemic's impact on Clark county's tourism industry as reflected by this dataset). These techniques will also help me quantify the impact of the pandemic (for example: how much of the variance in Airbnb listings can be explained by the variance in confirmed Covid cases in the region).
3. I will utilize matplotlib/ Tableau visualizations for representing insights that are tied to Geospatial data within Clark county. I will also make use of bar graphs and timeseries where necessary. This is because it would help anyone reading my report intuitively understand my conclusions.

Timeline to completion

1. **Data gathering and Cleaning:** This is one of the largest steps and I aim to complete the milestone of cleaning and combining my Airbnb data by **11/14**. Then by **11/17** I will combine this dataset with the confirmed covid cases dataset to prepare my analysis.
2. **Hypothesis Testing and ANOVA:** I will allocate the next 5 days (by **11/23**) to thoroughly answer my research questions and perform the hypothesis tests.
3. **Visualizations and Presentations:** I will allocate the next 2 days (**11/25**) to complete the graphs required in my presentation and to ensure that they are intuitive.
4. **Writing the final report:** I will draft my final report from **11/25 to 11/30** and leave some time to verify my analysis with the course staff. This will allow me to pivot or make edits if required.