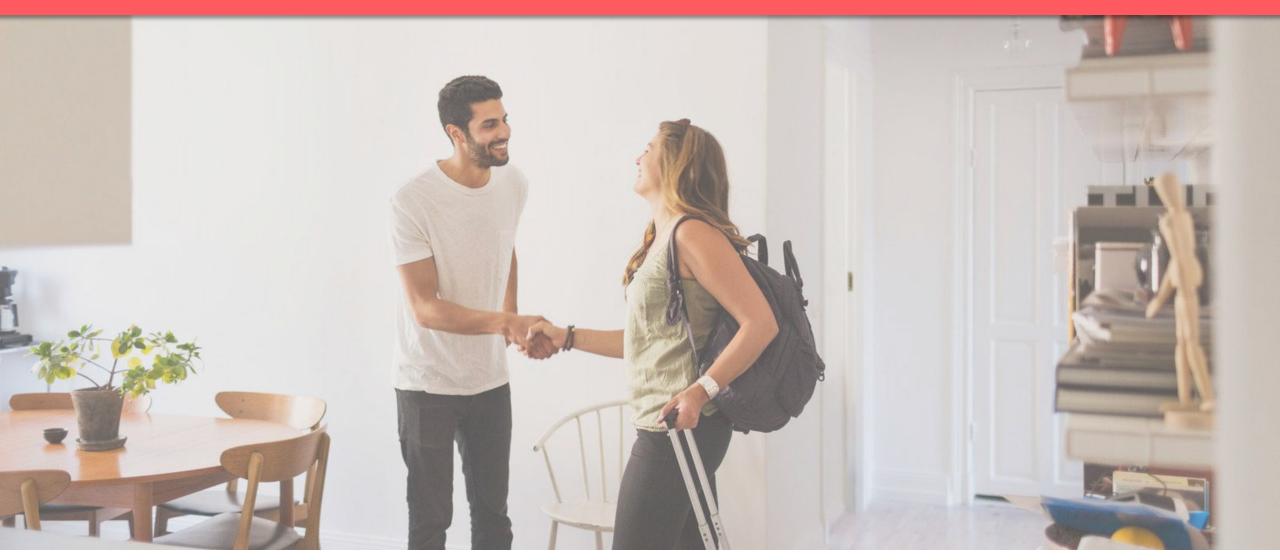


Where should you book your next Airbnb?



Group 5 High Five A++

Abhishek Saha Sameen Haroon Sharon

Sharon Xu Zeyu Zhu



Given the competitive and data-rich nature of the travel industry, our project focuses on Airbnb



- There is availability of data at scale in the online travel and hospitality space
- Data is being passed digitally to consumers in real-time along with an immediate feedback loop
- Increased competition provides consumers with more choices but also creates the need for informed decision-making
- The industry is highly relevant to everyday consumers, including us



- Airbnb is a disruptive and growing force in the online hospitality space
- The number of Airbnb users in the United States is 38 million right now and is forecasted to reach 43.3 million by 2020
- However, there is an opportunity to better enable Airbnb customers to make decisions about where they stay
- There is rich data on Airbnb which is available for public use



More specifically, we chose to leverage city-level Airbnb listing data for Los Angeles

Primary data from Inside Airbnb, http://insideairbnb.com/index.html

Focused on a single city, Los Angeles, with 43,000+ listings and 100+ attributes available

Distilled data to key relevant features such as



Location

- lat/long
- neighborhood



Customer feedback

- ratings
- reviews



Rental details

- room type
- capacity



Amenities

- wireless internet
- air conditioning



Service details

- Host flexibility
- Host experience

Los Angeles was deemed a good choice given the high level of tourism it receives and its urban sprawl. It also has a large number of Airbnb listings but is not as widely studied as other cities such as New York and San Francisco



Our analysis focuses on granular location data due to the critical role of location in driving travel rentals

Location matters, but it can be hard to tell what good is

- As travelers, we know that location is a major factor in finding a place to stay when visiting a new city
- However, it can be challenging to tell apart better vs. worse neighborhoods, and desirable locations can span across predefined boundaries
- Airbnb's interface provides a map view, but doesn't help a user identify which areas of a city travelers desire

Opportunity to use rich Airbnb and geographic data to help users differentiate between locations



Made in Python using WordCloud and matplotlib



We've sought to combine relevant data and data vis. tools to help users find the right area and the right listing

Right Area

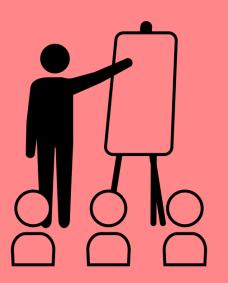
- Looked at ways to leverage available Airbnb data such as location ratings to create geographic clusters
- Paired these clusters with custom map layers in Tableau to ensure users could see these clusters in context
- Calculated and categorized area proximity to major tourist attractions
- Added in insights on how these areas stack on price and overall ratings

Right Listing

- Used online research and friendly conversations to gauge what customers prioritize in listing search
- Used natural language processing of reviews to explore additional traits that customers value
- Sought to overcome some limitations in current Airbnb filtering capabilities to better improve listing results
- Introduced tabular comparison of listings across key measures and characteristics



Putting it all together...

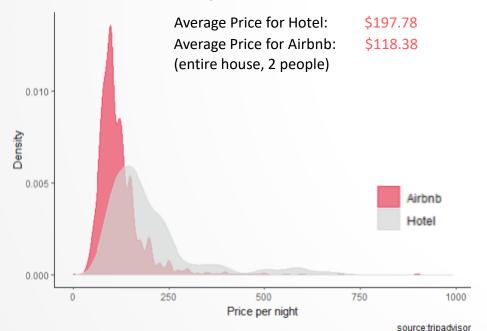




As with most projects, we learned some valuable insights along the way

- 1 Acquiring all the data you want can be challenging, possibly unfruitful, but always interesting
 - Three packages, dozens of attempts and numerous how-to-articles later....we could get hotel prices, just not latitudes and longitudes

Airbnb Listing Prices vs. Hotel Prices in LA



- 2 At the same time, its important to think through the right use-cases and audiences using the data available
 - While we thought through the perspectives of Airbnb hosts, business managers and customers, we realized our data was richest for the latter
- There are also meaningful tradeoffs between technology and functionality that are worth keeping in mind
 - Each additional filter or data layer we added steadily increased our load times (we had 3 cities to start, DC and Chicago got bumped)

Computing filters for 'OneSheet_Attempt' within 'Dashboard'	
Computing filters for 'OneSheet_Attempt' within 'Dashboard'	
Elapsed time 00:04	Cancel

Don't ever use Comic Sans MS. Except in pink. At the end of a semester in a class presentation on data visualization.

Thank you!

Questions?

