airbnb vs Hotels

Suprabhat Gurrala, Marc Marone, Aniruddha Nadkarni, Farhan Tejani, and Kexin Zhang
CSE 6242/CX 4242, Georgia Institute of Technology

Motivation

What is the problem?

The sharing economy has taken over a number of consumer industries, providing customers with many alternatives to traditional housing or transportation. Airbnb has quickly become a contender in the short term housing market. However, this has only increased the number of options facing consumers.

What options do short-term stay customers have?

Options like Airbnb provide a wide variety of choices for consumers, but they also vary dramatically in the amenities provided and level of comfort, often unbeknownst to the customer. Hotels have a far more rigid pricing model, but the level of care is easily anticipated.

How do they decide?

By investigating Airbnb and Hotel data in New York City, we are able to identify similar bookings across the two platforms. Our project provides comprehensive analysis and insightful visualizations to help consumers make an informed decision in a short amount of time.

Data Gathering

Airbnb

Pricing Over Time

Airbnb pricing data was gathered by scraping the site using Selenium, a web scraping tool for Python.

Reviews

We used Inside Airbnb, a data collection platform that gathered the listing reviews from Airbnb itself and compiled them into a dataset (44,318 Listings).

Location

The Inside Airbnb dataset also included columns for location, in the form of latitude and longitude coordinates.

Amenities

Amenities for Airbnbs include kitchens, televisions, air conditioning, and others, and were included in the Inside Airbnb dataset.

Hotels

Pricing Over Time

Amadeus, a travel technology company, provided us with pricing data over time for the hotels through their API.

Reviews

We scraped hotel reviews and amenity information from TripAdvisor (265 Hotels).

Location

We utilized the OpenStreetMap API to convert hotel addresses to latitude and longitude coordinates.

Amenities

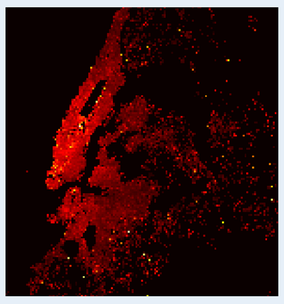
The hotel amenities data was provided by Amadeus and by TripAdvisor, and included data such as gyms, refrigerators, kid-friendliness, and others.

Approaches and Experiments

Cluster Analysis

Attempted to cluster by:

- **Latitude/longitude grid:** Airbnb listings cover New York but hotels are sparse.



- **Voronoi – Nearest subway:** Some cells were too large and regions with many hotels were broken up into several smaller cells

Solution:

- **Voronoi – K Means Centroid:** Clustered hotel locations with k = 130 and assigned each Airbnb listing to a cluster if it was within 1 kilometer of the centroid. 36% of Airbnb listings were disregarded for being too far away from a cluster.

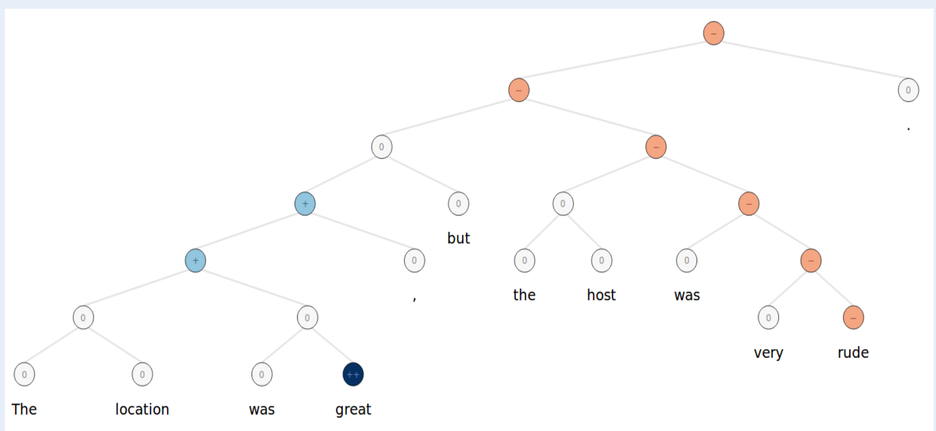
Directed Review Sentiment

- Our main visualization shows pricing data, but we also want to analyze subjective features.
- Reviews are useful, but don't have a standard format.
- Our method allows us to obtain sentiment scores representing specific aspects of each listing:
  1. Location
  2. Hospitality
  3. Room Quality

Method:

- Used the Stanford CoreNLP Library to obtain "sentiment trees."
- Broke the trees into subtrees (phrases).
- Classified each subtree to determine the aspect.

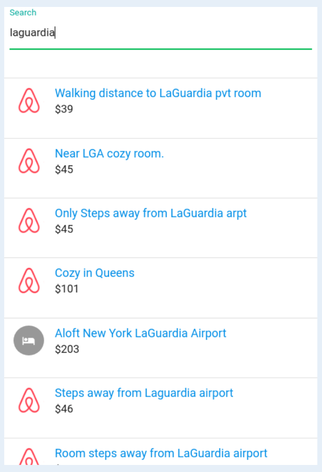
** We can handle a single sentence with multiple sentiments towards multiple aspects! (see figure below)*



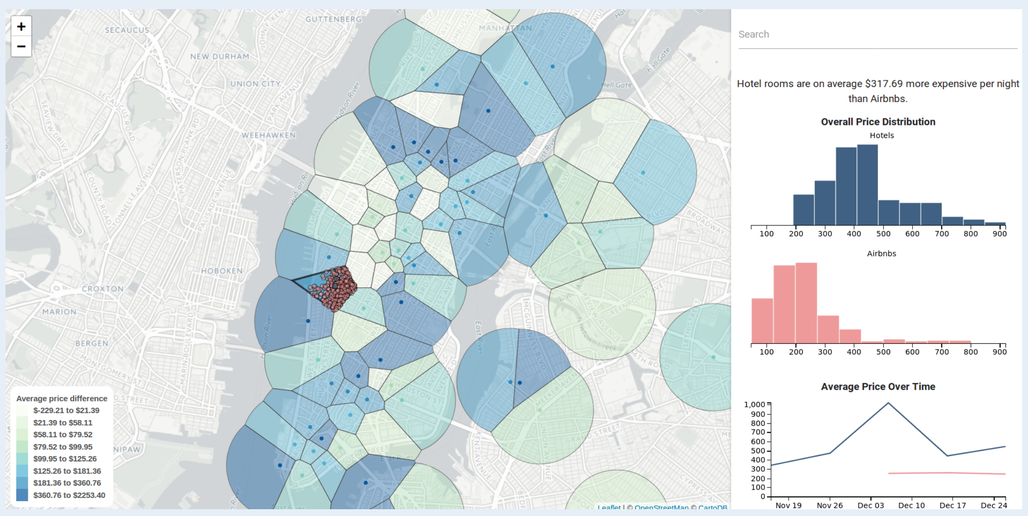
Unified Search

Method:

- Used ElasticSearch as a database and search backend.
- Converted search strings into latitude/longitude pairs using the OpenStreetMap API.
- Fed pairs into ElasticSearch location queries that return unified lists of nearest Airbnbs and hotels.



Results



Our interface provides an easy way to compare Airbnb listings to hotels:

- The choropleth map shows voronoi cells limited in radius and shaded by the average price difference.
- When a cell is clicked it shows the average price difference, a histogram for all listings in a group, and variations over time.
- Sentiment scores allow the user to easily compare subjective qualities relative to the average housing option.
- Links to the relevant listings are provided to aid the consumer with in-depth research and to allow for easy booking.

The combination of these features and intuitive visualizations help the end user quickly make a well-informed decision.