

Lodging and Rental Industry Problem Statement

Background

Human hospitality has been a cultural staple for thousands of years, with [civilizations ranging from Ancient Rome to early industrial England regularly offering basic accommodations to travelers](#). However, higher-end accommodations did not come about until 1768, when the Royal Clarence Hotel, the world's first "hotel" in the modern sense, was opened in the cathedral city of Exeter. Even then, the "hotel" concept only really caught on half a century later, when [Claridge's, the first 5-star hotel, opened in 1812](#). Throughout the 19th century, numerous hotel chains sprung up in Western Europe and North America, including the now famous Ritz luxury hotel chain (currently known as the Ritz-Carlton chain). The hoteling industry enjoyed a period of great expansion in the 20th century, as it spread its reach both economically and geographically.

But in 2007, recent college graduates Brian Chesky and Joe Gebbia embarked on a mission to solve a distinctly [Millennial](#) problem: affording rent in an expensive city. A little known company called AirBed & Breakfast. Although initially just an air mattress in their living room with breakfast service, it soon evolved into thousands of listings and tens of thousands of users. Soon enough, AirBed & Breakfast extended its reach across the globe, ballooning to a [valuation of over \\$30 billion as of March 2017](#). Airbnb, as it came to be known since March 2009, is now posing a serious challenge to the traditional hoteling industry by leveraging the distinctly 21st century "[sharing economy](#)" concept.

Airbnb has been a great boon to renters on their platform, solving for them the same problem that its founders had a decade ago when they graduated from college: affording their apartment rent via subletting it out to others when they weren't using it. However, it has forced hoteling industry incumbents to find new ways of differentiating their services, as the increased competition has driven down prices and put a sizable dent in the low- to middle-end business. With [the war between hotels and Airbnb reaching a boiling point in NYC](#), it seems like we are on the cusp of a watershed event that could revolutionize the rental market.

Your Task

Your goal is to analyze the U.S. Airbnb rentals data (described below), potentially in combination with supplementary datasets, in order to increase understanding of patterns in the lodging and rentals industry and how they relate to demographic, economic, and geographic trends at large.

We have partially pre-cleaned several supplementary datasets for your use. In addition to the Airbnb listings and calendar data, we have collected information on hundreds of thousands of metropolitan area venues. We also provide information on age and income brackets, as well as economic and housing-related data at the state and zipcode level.

You are asked to pose your own question and answer it using the available datasets in the available time. What is important is the insightfulness and depth of your conclusions and analysis. **You need not be comprehensive; quality data analysis is more important over breadth of the question posed.**

Submissions may be predictive, using machine learning and/or time series analysis to predict or model rental trends. Submissions may also be illuminating, through the use of thoughtfully chosen data visualizations or sound statistical tests.

Consider exploring one of the sample questions below, or creating your own variation. Creativity in formulating your own question is encouraged; **however, it should not be at the expense of analytical depth, precision, and rigor, which are far more important.**

Sample Question 1: What trends can you find in the Airbnb rental calendar over time, and how might these be explained by listing-specific and/or neighborhood-level factors?

Sample Question 2: Investigate the neighborhoods that have higher or lower Airbnb rental activity. Are there any notable demographic patterns that emerge between such neighborhoods?

Sample Question 3: Are there any interesting differences in Airbnb rental activity between areas with different real estate values and economic characteristics?

Sample Question 4: Analyze the locational patterns of Airbnb rentals. What relationships can you find between these and the locations of various types of metropolitan area venues?

Datasets

The provided datasets are spread across six tables. Your team should only use the tables that are relevant to your chosen question/topic. The raw data sources are noted; however, we encourage you to use our tables since they have been organized and cleaned to “play nice” with each other.

listings

Descriptive information on tens of thousands of Airbnb listings along the U.S. East Coast. 59,824 rows & 29 columns. Size: ~30MB. Source: [Airbnb](#).

calendar

Basic information on the Airbnb listing calendar. ~20 million rows & 5 columns. Size: ~55MB zipped, ~675MB unzipped. Source: [Airbnb](#).

demographics

Demographic data (population, age, income level, etc.) by zip code, taken from 2011 – 2015. 33,120 rows & 26 columns. Size: ~0.5MB. Source: [U.S. Census](#).

econ_state

Economic data (GDP, personal income, unemployment rate, etc.) for all 50 states and the District of Columbia, taken from 1980 – 2016. 51 rows & 519 columns. Size: ~0.2MB. Source: [U.S. Bureau of Economic Analysis](#).

real_estate

Monthly data that represent real estate prices, organized alphabetically by city and state. Specifically, monthly values from the Zillow Home Value Index (ZHVI) and the Zillow Rent Index (ZRI) for all homes (single-family residence & condo/co-op). 29,111 rows & 261 columns. Size: ~10MB zipped, ~40MB unzipped. Source: [Zillow](#).

venues

Important details about hundreds of thousands of venues in the metropolitan areas included in *listings*.

267,958 rows & 7 columns. Size: ~8MB zipped, ~30MB unzipped. Source: [Google Places](#).

Other Materials

We also provide you the schema for each of the data tables. You'll find them in the same project folder.