DARDEN
BUSINESS PUBLISHING
UNIVERSITY of VIRGINIA

UV7220
Rev. Dec. 6, 2016

Have Text, Will Travel: Can Airbnb Use Review Text Data to Optimize Profits?

Hundreds of thousands of would-be hoteliers have been popping up all around you. One might have been your unassuming 60-year-old neighbor, looking to pad her pension—others might include the young couple down the hall from your Chicago high-rise apartment.

The one thing all these individuals had in common was that they wanted to list their properties so they could invite complete strangers into their homes and charge them a premium to treat the space as their own. More often than not, these entrepreneurs had zero experience in the hospitality industry and no idea how to run a successful guesthouse. Welcome to the Airbnb era.

The goal of Airbnb's aspiring hosts was to use the Airbnb website (www.airbnb.com) to attract guests who were willing to pay the highest rates to stay in their homes for a short time. For Airbnb, the goal was to improve customer review performance so it could, in turn, increase profits. How could the company achieve its goal? Enter text mining, a technique that allowed businesses to scour Internet pages, decipher the meaning of groups of words, and assign the words a sentiment proxy through the use of a software package.

In order for text mining to be useful for Airbnb, its marketing professionals first had to gain access to customer review data on the vacation rental firm's own website. The team then had to analyze the data to find ways to improve property performance. Was the text data in the reviews adequate to the team's purposes? Was the team going to be able to leverage this large amount of data to determine a strategy (e.g., a location-specific approach) going forward?

If the marketing team was successful, hosts would be more likely to continue to list their properties on the Airbnb site, rather than being attracted away by the growing number of competitors in the market.

Taking Flight: The Rise of Airbnb

Founded in 2008 and based in San Francisco, California, Airbnb was an online platform that connected owners of homes, condos, apartments, villas, and even castles to prospective renters. This platform overturned the hospitality industry in the half decade after its founding and continued to gain traction in the market. Indeed, Airbnb's rise came at a time when shared economies—wherein the creation, production, distribution, trade, and consumption of goods and services were performed by a disparate group of individuals—were increasingly popular. The strategy had become entrenched in retail settings (e.g., eBay and Craigslist) and was gaining a foothold in other areas such as transportation (e.g., Zipcar and Uber).

This case was prepared by Shea Gibbs, Research Assistant; and Rajkumar Venkatesan, Bank of America Research Professor of Business Administration. It was written as a basis for class discussion rather than to illustrate effective or ineffective handling of an administrative situation.

Airbnb's specific model involved charging both its hosts and guests a fee for using its online connection service. Anywhere from 6% to 12% of the reservation subtotal went to Airbnb, and this reservation fee decreased as the price of the accommodation increased. Hosts paid a 3% service fee. To grow revenues, the company had, since its inception, looked for ways to help its hosts market their homes. For example, the company encouraged homeowners to use professional photography services,[1] and in winter 2014, it launched *Pineapple*, a magazine in which users in several key markets told their personal stories.[2]

It was a strategy that made Airbnb the world's leading peer-driven home rental website, and as of 2015, the company offered properties in more than 34,000 cities in 190 countries. Hosts who rented their homes through Airbnb—more than 800,000 people worldwide—had housed more than 20 million guests. A 2012 study by real estate consulting firm HR&A Advisors indicated that Airbnb provided a significant economic boost to both its users and the locations they visited.[3] In an examination of the city of San Francisco, the firm found that people who rented their homes on Airbnb used the income they earned to stay afloat in difficult economic times, and that travelers who used Airbnb spent more money in the cities they visited and brought income to less-visited neighborhoods than travelers who used traditional accommodations.[4]

"Airbnb represents a new form of travel," said Brian Chesky, Airbnb's CEO and cofounder, at the time of the study. "This study shows that Airbnb is having a huge positive impact—not just on the lives of our guests and hosts, but also on the local neighborhoods they visit and live in."[5]

The HR&A study showed that, from April 2011 to May 2012, guests and hosts utilizing Airbnb contributed $56 million in total spending to San Francisco's economy. Further examinations of Airbnb's economic impact have shown that the service generated $61 million in one year in Portland, Oregon;[6] $175 million in Barcelona, Spain;[7] $240 million in Paris, France;[8] and $632 million in New York.[9]

Airbnb was not alone in its dominion over the user-generated vacation rentals market. HomeAway, which operated a number of rental platforms such as VRBO and VacationRentals.com, was a large and growing competitor that had shifted its business model to more closely match that of Airbnb. HomeAway had previously charged its hosts on a per-listing basis, but it changed a portion of its services to charge users only when they successfully rented their properties.[10]

Airbnb was unique, however, in that all its listings were located in one place: its website. This made it an ideal candidate for collecting large amounts of text data using a web-scraping tool and represented a way for the company's marketing team to gain an advantage on its fast-growing competition.

[1] "Get Access to a Professional Photographer—It's Free," Airbnb website, https://www.airbnb.com/info/photography (accessed Mar. 24, 2015).
[2] "*Pineapple*—A Magazine from Airbnb," Airbnb website, https://www.airbnb.com/pineapple (accessed Mar. 24, 2015).
[3] "Study Finds that Airbnb Hosts and Guests Have Major Positive Effect on City Economies," Airbnb website, November 9, 2012, https://www.airbnb.com/press/news/study-finds-that-airbnb-hosts-and-guests-have-major-positive-effect-on-city-economies (accessed Nov. 7, 2014).
[4] https://www.airbnb.com/press/news/study-finds-that-airbnb-hosts-and-guests-have-major-positive-effect-on-city-economies.
[5] https://www.airbnb.com/press/news/study-finds-that-airbnb-hosts-and-guests-have-major-positive-effect-on-city-economies.
[6] "New Study: Airbnb Community Generates $61 Million in Economic Activity in Portland," Airbnb website, https://www.airbnb.com/press/news/new-study-airbnb-community-generates-61-million-in-economic-activity-in-portland (accessed Nov. 7, 2014).
[7] "New Study: Airbnb Community Contributes $175 Million to Barcelona's Economy," https://www.airbnb.com/press/news/new-study-airbnb-community-contributes-175-million-to-barcelona-s-economy (accessed Nov. 7, 2014).
[8] "New Study: Airbnb Community Contributes €185 Million to Parisian Economy," https://www.airbnb.com/press/news/new-study-airbnb-community-contributes-185-million-to-parisian-economy (accessed Nov. 7, 2014).
[9] "New Study: Airbnb Generated $632 Million in Economic Activity in New York," https://www.airbnb.com/press/news/new-study-airbnb-generated-632-million-in-economic-activity-in-new-york (accessed Nov. 7, 2014).
[10] Miguel Helft, "Growing Quietly in Airbnb's Shadow," *Fortune*, March 12, 2014, http://fortune.com/2014/03/12/growing-quietly-in-airbnbs-shadow/ (accessed Nov. 7, 2014).

Gathering Text Data

The idea of gathering and using text data from Internet-based sources was not a new one. It had been used with some success to examine other large web marketplaces such as Yelp, and it proved useful in automotive market segmentation.[11]

Recently, several commercially available tools made text mining far more practical for businesses such as Airbnb, as well as smaller businesses the world over that were looking to turn their web text into actionable information. Import.io was a fully web-based tool that was simple to use and could be customized to individual data-mining projects.[12] By deploying the tool on a set number of sample pages, the software understood the type of information available and could use that model to extract data from a large number of pages in a short amount of time. The data output was in a simple CSV export that could be used in the subsequent analysis.

Although text mining tools still had their limitations as of 2015, they were becoming smarter day by day, and the latest packages were capable of sophisticated sentiment analysis that could turn written words into quantifiable consumer preferences.

Analyzing Sentiment and Developing a Revenue Model

Imagine that Airbnb marketing professionals have received text data from the company's IT department and asked that the data be cleaned (i.e., have all text changed to lowercase, and have punctuation, numbers, and special characters removed) and whittled down to only two sample cities: Paris, France, and Miami, Florida. (A sample of the raw data is shown in **Exhibit 1**.)

Text itself was not usable in a regression model, so in order to use consumer reviews in a model designed to optimize property performance, Airbnb's marketing team needed to apply a numerical value to the sentiment implied in a group of related words. Several forms of sentiment-analysis tools were available. The software chosen for this particular analysis was called qdap in R. Once the raw data was imported into this sentiment-analysis tool, all the data could be mapped to the variables shown in **Exhibit 2**.

Using qdap in R, the Airbnb marketing team had two options when determining how to represent sentiment in its model: granular (e.g., at the level of each review) and high level (e.g., using multiple reviews). The team in this case selected the high-level analysis and elected to create a polarity metric to represent sentiment.

The qdap in R polarity algorithm used a prespecified dictionary of positive and negative words, as well as context shifters—words around the positive and negative words. The context shifters could be neutral, amplifiers, or deamplifiers. Neutral words did not add to the polarity score, and the weight of the positive and negative words was the net sum of the number of amplifiers and deamplifiers.

The polarity score computed the weighted average of positive and negative words in a sentence, and the weights were dependent on the combination of the words and the context shifters. For example, if a reviewer stated, "Nicolas is a great host everything was perfect and the flat is amazing and the location is great in a quiet area close to the subway I would definitely come back here," the polarity algorithm, after preprocessing, recognized "great," "perfect," "amazing," "great" again, and "quiet" as positive words. It identified no negative words and produced a polarity score of 1.22.

[11] Oded Netzer, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko, "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science* 31, no. 3 (2012).

[12] Import.io was available at https://www.import.io/ (accessed Oct. 23, 2015).

Alternatively, consider the following sentence: "Nicolas is a bad host everything was horrible and the flat is dirty and the location is great in a quiet area close to the subway I will not come back here." Balancing the two predefined positive words—"great" and "quiet"—with the three negative ones—"bad," horrible," and "dirty"—qdap in R returned a polarity score of −0.036.

Once a numerical value was assigned to the sentiment present in each review, Airbnb's marketing managers could use it as a variable in a regression analysis designed to optimize revenues just as it would any other variable. The Airbnb marketing team might consider *price*, *reviews*, *saved_wish*, and *min_stay*, for example. The team might then consider drivers of these metrics such as price itself; whether people rent out entire homes, private rooms, or shared rooms; and number of bedrooms.

Optimizing Price and Beyond

The goal of Airbnb's marketing team in this exercise was to improve its users' performance so it could reap the benefits of ongoing host and renter fees. If the company's hosts were not happy, they were not likely to continue listing their properties through Airbnb, and in a competitive and burgeoning marketplace, such attrition could be devastating.

What could the Airbnb marketing team offer to improve its users' experience? Should it rank properties it suggested to users based on some metric such as review sentiment? How would review sentiment compare to summary-rating value in terms of its ability to predict revenues? Given what we know about the performance of properties in Miami and Paris, did Airbnb need a region-specific strategy? Could the company suggest optimal pricing for hosts, or suggest other ways hosts could improve overall earnings?

Exhibit 1

Have Text, Will Travel:
Can Airbnb Use Review Text Data to Optimize Profits?

Sample Data

	price	reviews	rating	accommodates	extpeop	savwish	min_stay	sentiment	secdep	cleanfee	weekfee	monthfee	bedroom	bathroom	beds
1	$70	45	4.5	2	0	934	3	3.704471	1	0	0	0	1	1	1
2	$100	13	5	3	0	171	5	3.355278	1	0	1	1	1	1	1
3	$90	0	NA	4	1	0	1	2.962161	0	0	0	0	0	1	1
4	$125	20	4	4	0	460	7	2.139501	0	1	0	0	1	1	2
5	$99	10	5	6	0	589	1	3.628548	1	1	1	1	2	2	3
6	$129	22	5	4	1	560	7	3.334242	1	1	1	1	1	1	1
7	$99	72	4.5	4	0	1464	1	0.772571	1	1	0	0	1	1	1
8	$1,300	0	NA	4	0	0	5	NA	1	1	0	0	2	2	2

Data source: Airbnb public website, accessed April 2014.

Exhibit 2

Have Text, Will Travel:
Can Airbnb Use Review Text Data to Optimize Profits?

Data Dictionary

Variable	Description
rating	Average user rating on a scale of 1 to 5, with 5 being the top rating
reviews	Number of user reviews
price	Daily rental price of the property
saved_wish	Number of times property was saved to the wish list
review_text	Raw dump of all the user reviews
sentiment	Numeric value assigned to review text describing its positivity or negativity
accommodates	Number of people the property can accommodate
bedroom	Number of bedrooms
bathroom	Number of bathrooms
beds	Number of beds
min_stay	Minimum number of days required for rental
secdep	Security deposit necessary?
cleanfee	Fees for cleaning services
weekfee	Discount for weekly stay
monthfee	Discount for monthly stay
extpeop	Fees for extra people

Source: Created by case writer; variables are from Airbnb public website, accessed April 2014.