The life and death of the *bar de barrio*?

Francisco Villamil*

Social Data Science course (ETH Zurich, February 2019)
March 11, 2019

1 Introduction

The promise of the so-called sharing economy was to facilitate peer-to-peer transactions among people, increasing the efficiency of the market by giving new uses to unexploited goods and services. Both sides of each transaction would benefit from this arrangement, earning some side money and getting access to more accessible services. In the case of AirBnB, the stated outcome was to make it easier and cheaper to get accommodation when travelling and make it possible for guests to experience local life through hosts. Beyond this, local businesses could benefit from an influx of new visitors, who would now go to areas or neighborhoods that were only previously visited by locals. AirBnB promised to democratize travel, and to tear down the walls separating local inhabitants from temporary travellers.

Yet, this new model is not free from conflict. In a similar vein to the conflicts brought about by Uber and other tech companies, AirBnB – as the biggest online marketplace for hospitality services– has been blamed for the worst excesses of international tourism. AirBnB is probably the latest chapter in a story that started with the liberalization of short-haul air travel and that brought about a dramatic increase in tourism and the weekend-long "city breaks," at least within Europe. This process has put a lot of pressure on a few popular cities, which now cope with an unbearable number of visitors.[1] AirBnb allows landlords to earn way more money from their properties than normal long-term renting, but, as critics claim, at the cost of breaking apart the social fabric of neighborhoods and replacing loyal customers of local bars for one-time buyers from abroad. The traditional *bar de barrio* has then to close down, and its clientele of local drinkers is now being replaced by exchange students drinking kale smoothies in a new Brooklyn-inspired café.

But is this true? Is there a link between peer-to-peer online platforms and a local increase of international tourists? In particular, does an increase of AirBnB listings change the clientele of local businesses? This is the main question in this report. Focusing on the city of Madrid and using disaggregated data from AirBnB, I explore whether the number and increase of AirBnB listings in each neighbourhood is related to the number of national vs international visitors local bars and restaurants receive. Following the argument above, the expectation would be that an increase in AirBnB listings is negatively correlated with the share of reviews in Spanish in each *barrio*.

2 Methods[2]

To answer the question above, I build a database for the city of Madrid at the level of neighbourhoods (*barrios*), covering all but one of the 131 official neighbourhoods[3] and including data on

*francisco.villamil@icr.gess.ethz.ch

[1] Elle Hunt, 'Tourism kills neighbourhoods': how do we save cities from the city break?, *The Guardian*, August 4, 2017. https://bit.ly/2h2tKPs (accessed 11/03/2019).

[2] The R code and further information on the TripAdvisor data can be found at https://github.com/franvillamil/tripadvisor_reviews

[3] In order to aggregate the different variables spatially, I rely on the official *Distritos y barrios de Madrid* shapefile offered by the *Instituto de Estadística* of Madrid, available at https://bit.ly/2Cf0a0u (accessed 11/03/2019).

AirBnB listings, the share of Spanish reviews in the online travel website TripAdvisor, and a number of control variables.

Data on AirBnB listings comes from the Inside Airbnb project[4], which periodically scraps AirBnB website and makes the data public. I downloaded the data for the city of Madrid at three different points in time (November 7, 2018, April 8, 2017, and October 2, 2015) and, using the coordinates of each listing, aggregate the number of listings for each neighbourhood.

To measure the number of local and international visitors to bars and restaurants, I rely on the online reviews posted at TripAdvisor for all 10,000+ bars and restaurants listed for Madrid. I scrap all reviews in Spanish or English[5] for every bar and restaurant posted before January 1, 2019, removing the automatic translations, and registering the posting date. In total, this results in 854,237 reviews for 9,518 businesses. I measure the share of reviews in Spanish respective to the number of reviews in Spanish and English roughly two months after each of the three points in time for which the AirBnB data was retrieved, i.e., on December 31, 2018, June 30, 2017, and November 30, 2015.

This data structure was designed to run two types of analyses. First, I use the static picture in each of three points in time and run a regression of the (log) number of AirBnB listings on the share of TripAdvisor reviews in Spanish. Second, I calculate the change in both AirBnB listings and reviews in Spanish between 2015–2017, 2017–2018, and 2015–2018 and run dynamic analyses using these variables.

In addition, I include a number of control variables. First, using the TripAdvisor data, I control for the number of restaurants and the total number of reviews in each barrio. In the dynamic analyses, I include the change version of these variables. Second, using again data from the *Instituto de Estadística de Madrid*, I include the total number of (a) 4- and 5-star hotels and (b) 1-, 2-, and 3-hotels and all hostels in each neighbourhood. Finally, I also control for the (log) distance in meters from each neighborhood centroid to Madrid downtown (Puerta del Sol) and include district fixed effects. In the static, three-period analysis I also include year fixed effects.

3 Results

Table 1 shows the results of the static analysis, i.e., including a single observation for each *barrio* at each of the three different points in time. Contrary to the expectation spelled out at the beginning of this report, the number of AirBnB listings in each barrio is, if any, positively correlated with the share of TripAdvisor reviews in Spanish.

Table 1: TripAdvisor reviews in Spanish (static)

(Intercept)	0.409^{***}	0.424^{***}
	(0.051)	(0.051)
ln AirBnB listings	0.006^{+}	0.002
	(0.003)	(0.004)
ln Hotels 4/5*	-0.028^{***}	-0.028^{***}
	(0.004)	(0.004)
ln Hostels/Hotels 1/2/3*	-0.003	-0.003
	(0.004)	(0.004)
ln Distance to Sol (m)	0.057^{***}	0.057^{***}
	(0.006)	(0.006)
ln total no. of reviews	-0.008^{*}	-0.008^{*}
	(0.003)	(0.003)
ln no of bars/rests	0.012^{*}	0.011^{+}
	(0.006)	(0.006)
2017	0.016^{***}	-0.002
	(0.005)	(0.012)
2018	0.017^{**}	-0.010
	(0.005)	(0.012)
AirBnB x 2017		0.006^{+}
		(0.003)
AirBnB x 2018		0.008^{*}
		(0.003)
Observations	378	378
R^2	0.801	0.805
Adjusted R^2	0.785	0.788

Note: $+p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001$. District FE not shown.

Although the relationship in the global analysis (first column) is fairly weak (the effect is small and only significant at the 90% level), including

[4] Inside AirBnB, a data project by Murray Cox. Available at http://insideairbnb.com (accessed 11/03/2019).

[5] Although some reviews are written in other languages, the great majority of them are either in Spanish or English.

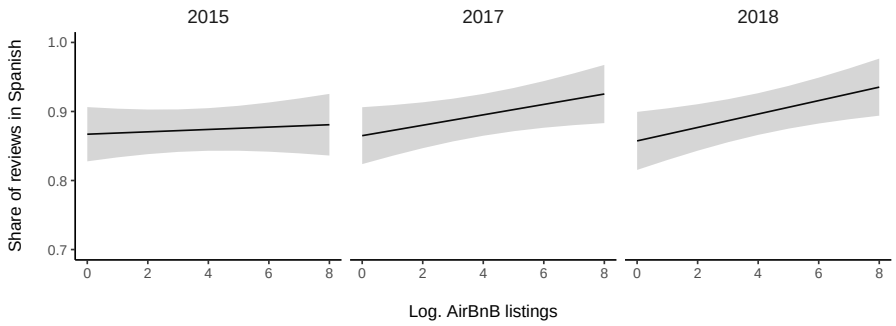

Figure 1: Number of AirBnB listings and share of TripAdvisor reviews in Spanish at three points in time

an interaction term with the year (second column) shows that the effect increases and is at its highest in the latest period (late 2018). Figure 1 shows this result graphically.

Table 2: TripAdvisor reviews in Spanish (change)

	2015-17	*2017-18*	*2015-18*
(Intercept)	1.216^{***}	0.996^{***}	1.287^{***}
	(0.066)	(0.021)	(0.070)
Δ AirBnB	−0.003	0.005^{***}	−0.002
	(0.003)	(0.001)	(0.002)
Δ all reviews	0.005	0.035^{**}	−0.002
	(0.004)	(0.010)	(0.001)
Δ bars/rests	−0.003	−0.005	−0.004
	(0.006)	(0.004)	(0.005)
ln Hotels 4/5*	0.008	0.005^{**}	0.010
	(0.006)	(0.002)	(0.006)
ln Host/Hotels	−0.002	−0.001	−0.004
	(0.006)	(0.002)	(0.006)
ln Dist Sol (m)	-0.027^{**}	−0.002	-0.032^{***}
	(0.008)	(0.003)	(0.009)
Observations	123	125	124
R^2	0.318	0.424	0.417
Adjusted R^2	0.133	0.271	0.261

Note: $+p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001$. District FE not shown.

Table 2 shows the results for the dynamic analyses, using both the dependent variable (share of reviews in Spanish) and the main covariates as changes between two points in time. Again, results are weak and, if anything, go against the argument explained in the introduction. In particular, between mid-2017 and late 2018, an increase in AirBnB listings in Madrid neighbourhoods was related to an increase in Spanish reviews relative to the ones in English. This result is in line with the findings from the static analyses, which show that the correlation between AirBnB listings and reviews in Spanish was increasing over time.

4 Discussion and critique

Results from both the static and dynamic analyses show the the number of AirBnB listings in each *barrio* of Madrid is not related to an increase in foreign visitors to local bars and restaurants, but rather the opposite. Although the correlation is relatively weak, results show that there is an increasingly strong relationship between AirBnB listings and reviews in Spanish, which probably started to be present after 2017.

The naive conclusion is that an increase in AirBnB hosts is bringing Spanish tourists who then visit and review local bars and restaurants. However, this might be an spurious correlation. The general share of reviews in Spanish has been increasing over the years, which could explain these findings if results are being driven by neighbourhoods that both attract local visitors and make landlords more likely to post listings online. Without a clear link between AirBnB guests and TripAdvisor reviewers, it is hard to know the specific causal chain without falling into an ecological fallacy.

Moreover, not every bar and restaurant has a site on TripAdvisor, and the share of reviews in Spanish vs English is highly influenced by the popularity of a business. Thus, analyzing the data at a lower level of aggregation or even at the level of individual bars/restaurants would help to know better the impact of AirBnB on local clienteles.