

Airbnb - New York's Analysis

Jackson A. Prado Lima

Department of Computer Science, Federal University of Paraná

CP:19081, CEP: 81531-980, Curitiba, Paraná, Brazil

japlima@inf.ufpr.br

ABSTRACT

Airbnb is an outstanding hosting service used worldwide and provides information to its users in a simple and efficient manner, although this information is limited only to listings of its products. In this work, an analysis is made taking into account if incidents and cultural organizations influence the prices practiced in Airbnb. Besides, the prediction of new prices is carried out that the user can come to inform taking into account information of the current listings. The price forecast has some advantages when advising the user: on the price practiced (is too large or small), the mode of advertisement, or even on the investment decisions of the market analysis. To help in this task, the present work performs an analysis using graphs, maps and different types of machine learning.

1 INTRODUCTION

Airbnb was founded in August of 2008 and based in San Francisco, California. Airbnb is an online marketplace and hospitality service, enabling people to lease or rent short-term lodging including vacation rentals, apartment rentals, homestays, hostel beds, or hotel rooms.

The company does not own any lodging; it is merely a broker and receives percentage service fees (commissions) from both guests and hosts in conjunction with every booking. It has over 3,000,000 lodging listings in 65,000 cities and 191 countries, and the host sets the cost of accommodation. Like all hospitality services, Airbnb is a form of collaborative consumption and sharing. Airbnb has two million rooms available - nearly three times more than that of the largest hotel chain in the world.

How can we predict if a guest will continue booking rooms through Airbnb, and how can we predict if a host will continue listing out their property on Airbnb? Unlike traditional hotels, Airbnb relies on one group of its customers (hosts) to generate the supply of rooms for its guests. Thus, predicting customer loyalty, both on the host side and guest side, is arguably more important to Airbnb than to traditional hotels. In this paper, we try and capture this loyalty to Airbnb.

The data reveals that guests who stay with highly rated hosts are far more likely to return to Airbnb. For hosts, the best predictors of whether or not they will continue to list through Airbnb is the recency and frequency of their listing.

The paper is organized as follows. Section 2 describes how the experimental evaluation was conducted. Section 3 presents and analyses the results obtained. Finally, Section 4 concludes the paper and discusses the future work.

2 EXPERIMENT DESCRIPTION

The hypothesis of this work is that our approach is capable to recommend a local (district/borough) or a price to a user. To evaluate this, we consider uses maps, charts and machine learning.

According to our goals the experiment was guided by the following research question: **RQ:** "How the incidents and cultural organizations impact on listing price?".

To answer the RQ we conducted the evaluation in three phases: (i) Charts; (ii) Maps; and (iii) Machine Learning. In the first phase, we analyses the data distribution. In the second phase, we understand the impact of different factors using maps. In the last phase, we applies machine learning to predict some data.

From the main question we divided in other questions to understand different points:

- **RQ1:** How the the cultural organizations impact on listing price?
- **RQ2:** How the the incidents impact on listing price?
- **RQ3:** How can we use average value rating, average location rating, etc. in the data?

2.1 Target Data

The investigation focused on four dataset collected in open data sites. These dataset are as follows: 1) Listings, a dataset that is a Airbnb snapshot of 02 April, 2017; 2) Incidents, contains all incidents from New York for the year 2016; 3) Cultural Organizations, contains all cultural organizations from New York in 2017; and 4) NYC Borough, contains informations about boundaries of boroughs (districts). Further information about these datasets are available in the Table 1.

2.2 Settings

In this step was used three different approaches to read the data (csv files): (i) SqlContext with databricks; (ii) parquet; and (iii) pandas. After, it was create a unique dataframe with the necessary informations. To help in the analysys, the Cultural Organizations and Incidents dataset received a score a they are like a column unique in the new dataframe. How we have 5 districts in New York, and to be easy, the Cultural Organizations and Incidents dataset were grouped and ranked in relation to the order (1-5), where 1 is the work and 5 the best. In this case, the datasets were ordered by the number in each district, ascendant for cultural organizations and descendant for the incidents. In the end, we have a unique dataset with 37043 rows.

Table 1: Characteristics of the data sets used in the experiment.

#	Dataset	Description	Size	Download
1	Listings [1]	Detailed listings data, including various attributes (features) of each listing such as number of bedrooms, bathrooms, location, etc.	153,7MB	csv
2	Incidents [3]	All valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) for all complete quarters in 2016.	124,2MB	csv
3	Cultural Organizations [4]	Listing of all Cultural Organizations in the Department of Cultural Affairs directory.	333,kB	csv
4	NYC Borough [2]	Polygon boundaries of boroughs (water areas excluded).	400,kB	GeoJson

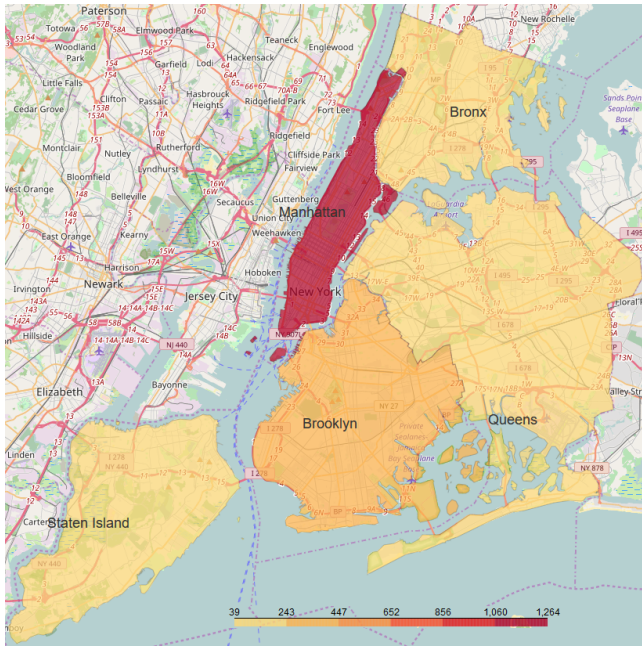


Figure 2: Number of Cultural Organizations per District

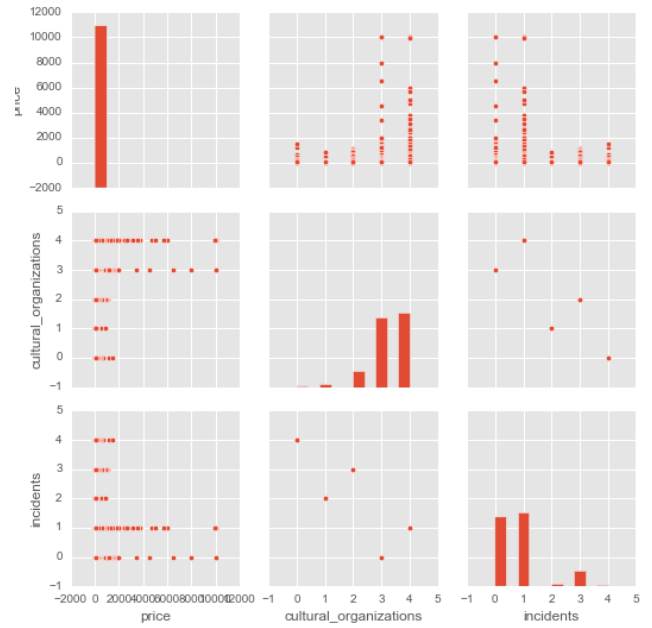


Figure 4

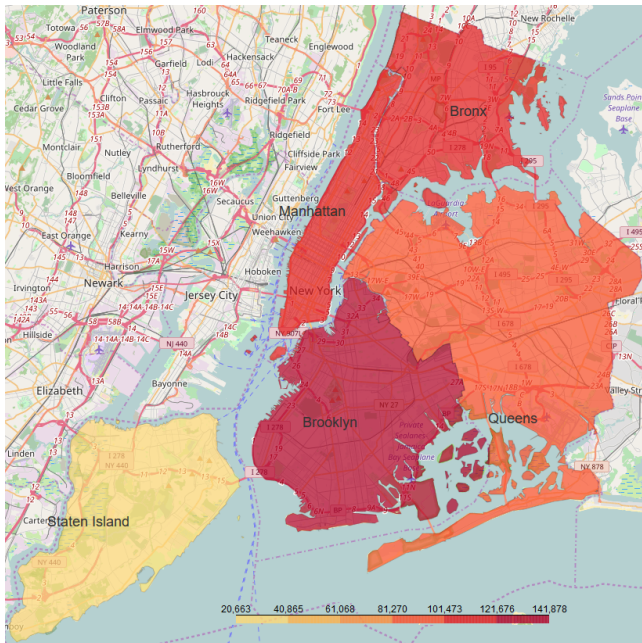


Figure 3: Number of Incidents per District

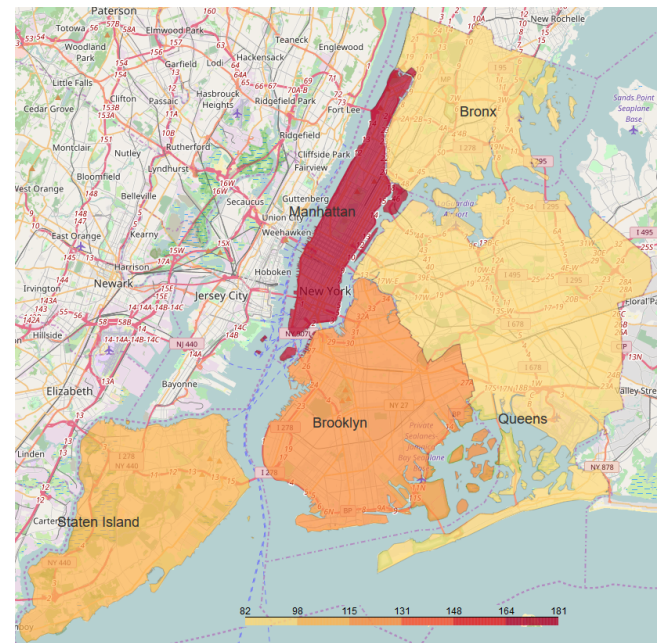


Figure 1: Mean Price per District

3 RESULTS AND ANALYSIS

3.1 RQ1 - Comparing

3.2 RQ2 - UCB

3.3 Threats To Validity

4 CONCLUDING REMARKS

REFERENCES

- [1] Inside Airbnb. 2017. Listings. <http://data.insideairbnb.com/united-states/ny/new-york-city/2017-04-02/data/listings.csv.gz>. (2017). Accessed Apr 2017. Date

- Compiled April 2, 2017.
- [2] BetaNYC. 2015. NYC Borough. http://data.beta.nyc/dataset/nyc-borough-boundaries/resource/7c164faa-4458-4ff2-9ef0-09db00b509ef?view_id=8c05eff7-54d1-4359-8183-0e278e91b01a. (2015). Accessed Apr 2017.
- [3] New York City Police Department. 2016. Incidents. <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-YTD/5uac-w243/data>. (2016). Accessed Apr 2017.
- [4] Department of Cultural Affairs. 2017. Cultural Organizations. <https://data.cityofnewyork.us/Recreation/DCLA-Cultural-Organizations/u35m-9t32>. (2017). Accessed Apr 2017. Last Update April 7, 2017.