

What makes an airbnb Superhost in NY



Introduction

Data mining on listings

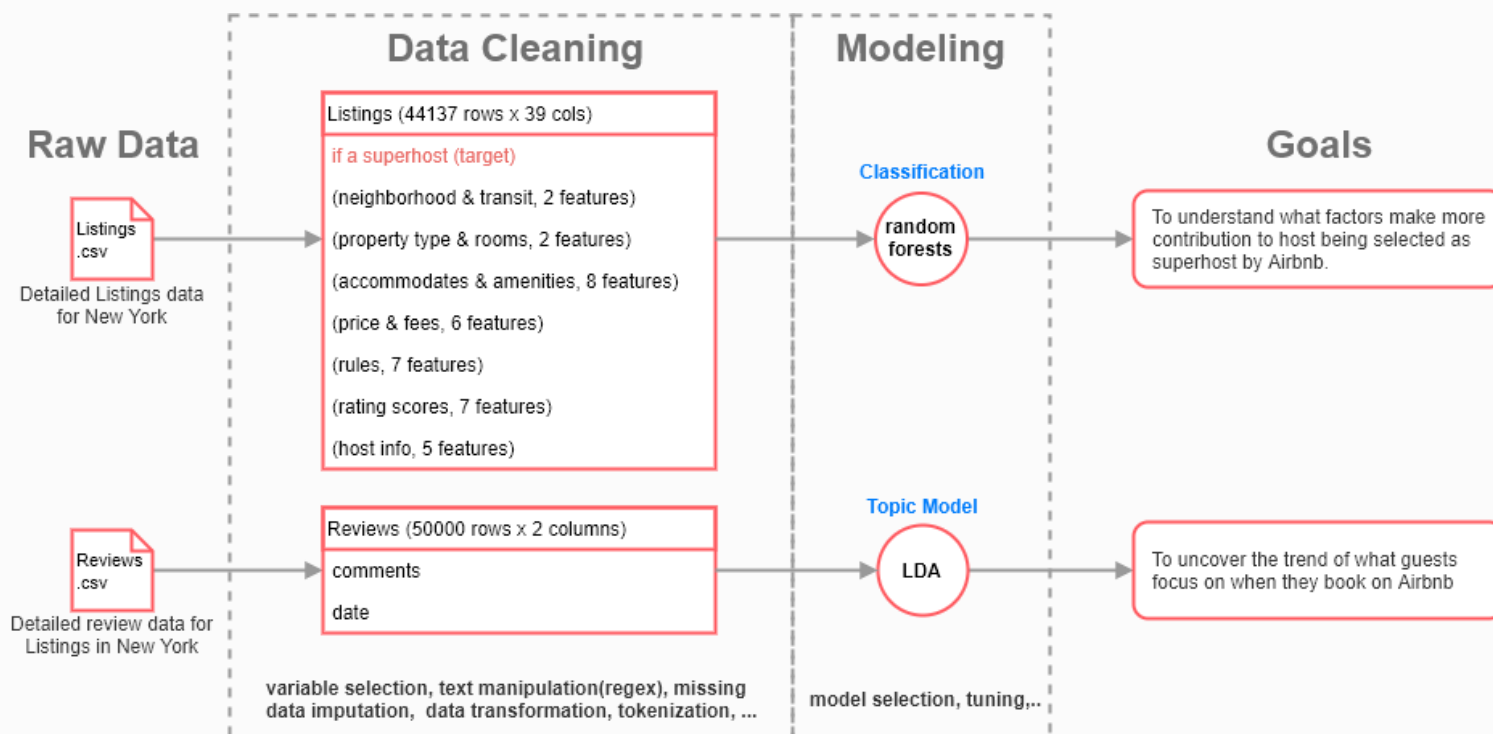
Text mining on reviews

Yiqiao Chen

INTRODUCTION

If you ever scrolled through Airbnb, you have surely seen some renters labeled as *Superhosts*. However, no exact rules specify how to get the badge. The project aims to understand what characteristics existing *Superhosts* have, and what key topics guests care about. Therefore, we would have a better understanding of its business model.

- A random forests model was applied to listings data to classify if a host can be selected as *Superhost*. The model with an feature importance could help achieve the first goal.
- The second goal was met by fitting topic model to extract trends of topics over time within guest reviews.



What makes an airbnb Superhost in NY



Introduction

Data mining on listings

Text mining on reviews

Yiqiao Chen

DATA MINING ON LISTINGS

- Applied Random Forests (500 trees, 'gini') to listings data and calculated the importance for each feature;
- Selected the top 15 important features and plotted the bar chart;
- Obtained the model with 7.78% misclassification rate, 3.4% Type I Error, and 6.2% Type II Error;
- According to the plot, a Superhost would fulfill following standards.

1) reviews

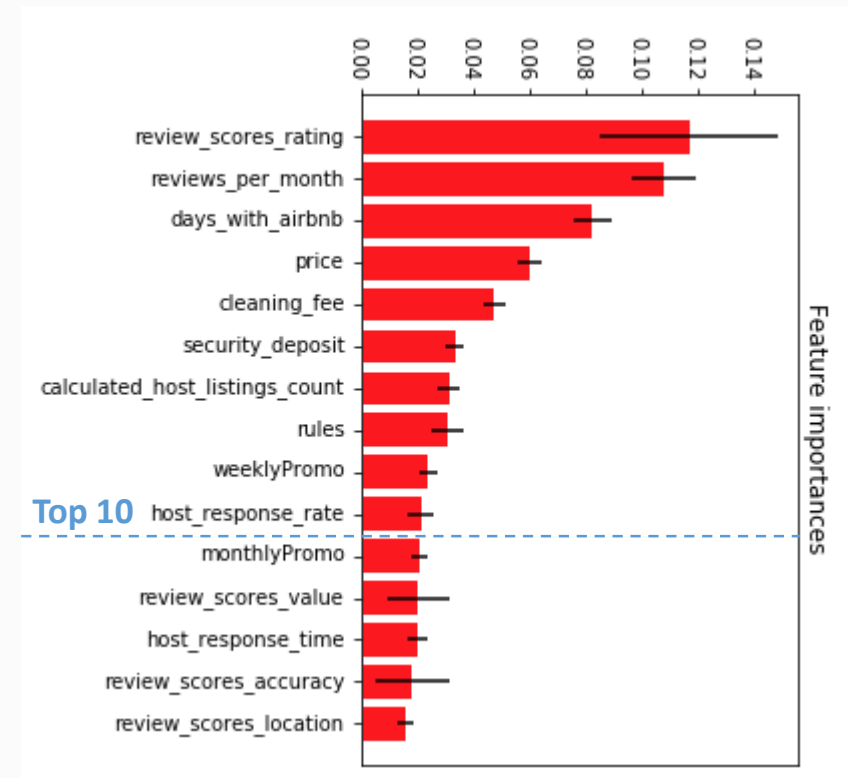
- Having sufficient reviews and a good overall rating;
- Sharing accurate information;
- Providing a valuable booking and well located place;
- Specifying rules clearly but not excessively;
- Being helpful and responsive.

2) price and promotions:

- Offering attractive daily price options;
- Charging reasonable cleaning fee and deposit;
- Giving discounts for staying longer (week/month).

3) activities:

- Hosting for a long time on Airbnb;
- Avoiding excessive listings under his/her name.



What makes an airbnb Superhost in NY



Introduction

Data mining on listings

Text mining on reviews

Yiqiao Chen

TEXT MINING ON REVIEWS

- 50k comments were randomly sampled from 840k (too big), 10k for each year (2013 – 2017, 5 years);
- Documents (comments) were tokenized, cleaned and converted to bag-of-words;
- Then LDA was applied to get topics (topics = 10, iterations = 400, passes = 20, chunk size = 2000);
- Most meaningful topics (5 of out 10) were selected and trends were plotted;
- Guests care more about:
 - 1) Convenient external environment;
 - 2) Comfortable interior design;
 - 3) Efficient host service.

Trends of Topics 2013-2017 in New York
(% of mentioning in sampled reviews)

