



Airbnb: Evaluating Factors of Rental Listing Quality

Kevin Liu, Rohit Jagga

Overview

Data

- Airbnb listings
- Crime reports (New York, San Francisco)
- Home value (Zillow)
- Hotel value (Trivago)

Techniques

- Sentiment analysis of comments, descriptions, etc.
- Multiple linear regression
- Target: User ratings (overall, location)

Models (Predictors)

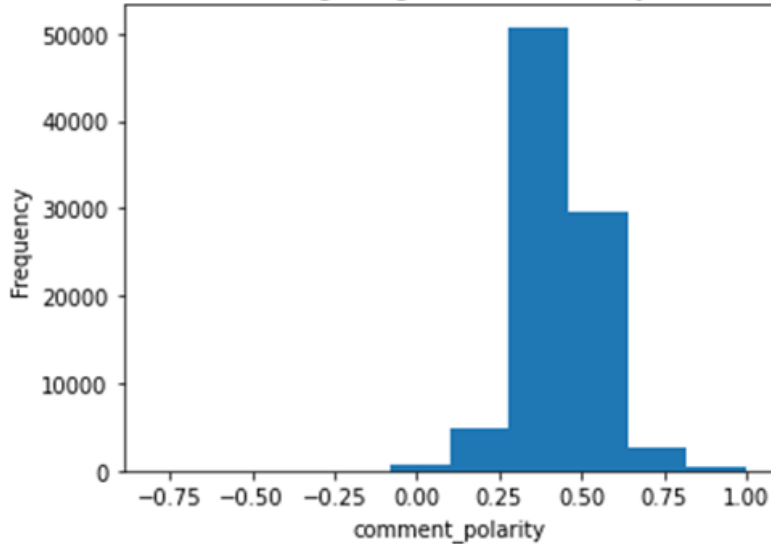
- Direct (listing-specific, sentiment)
- External (crime, home/hotel value)
- Combined (all)

Exploratory Data Analysis

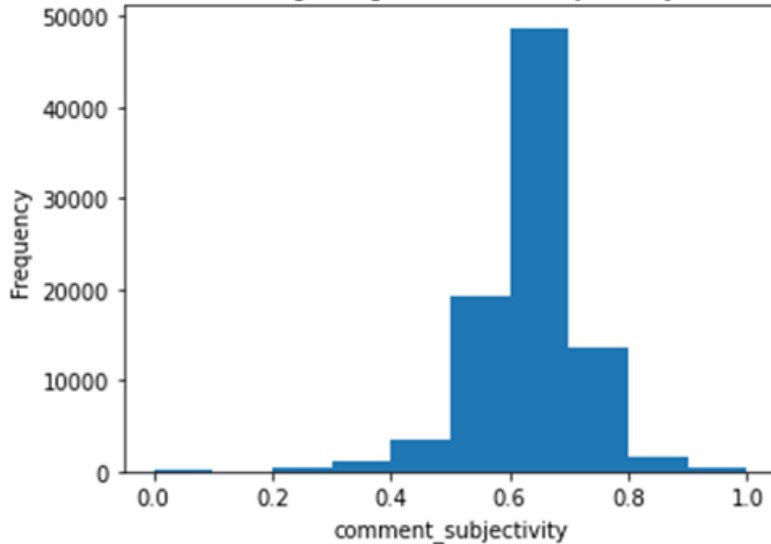
Sentiment analysis: Higher polarity/subjectivity → higher overall scores

Crime analysis: Higher crime rate → lower overall scores

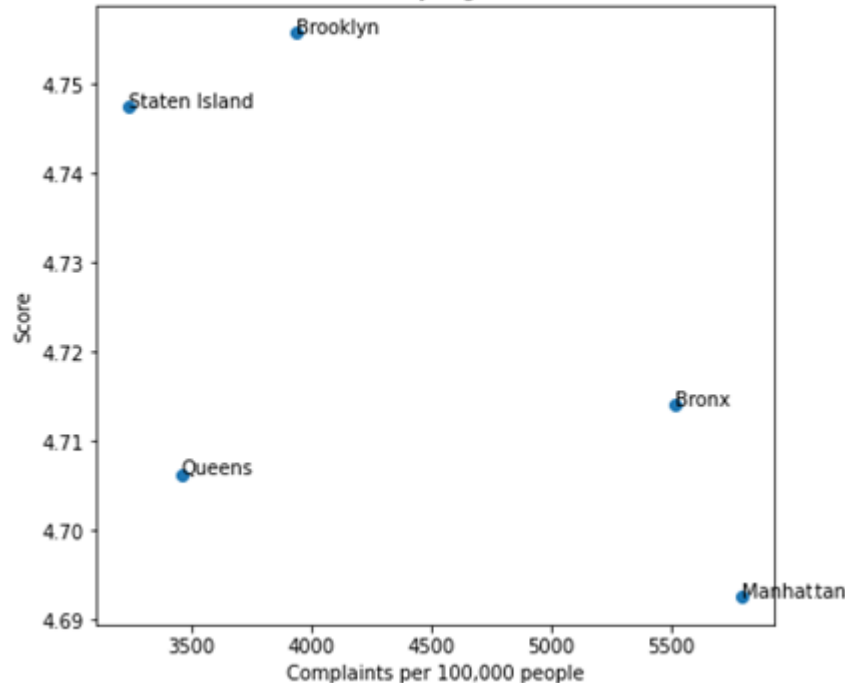
Listing's avg Comment Polarity



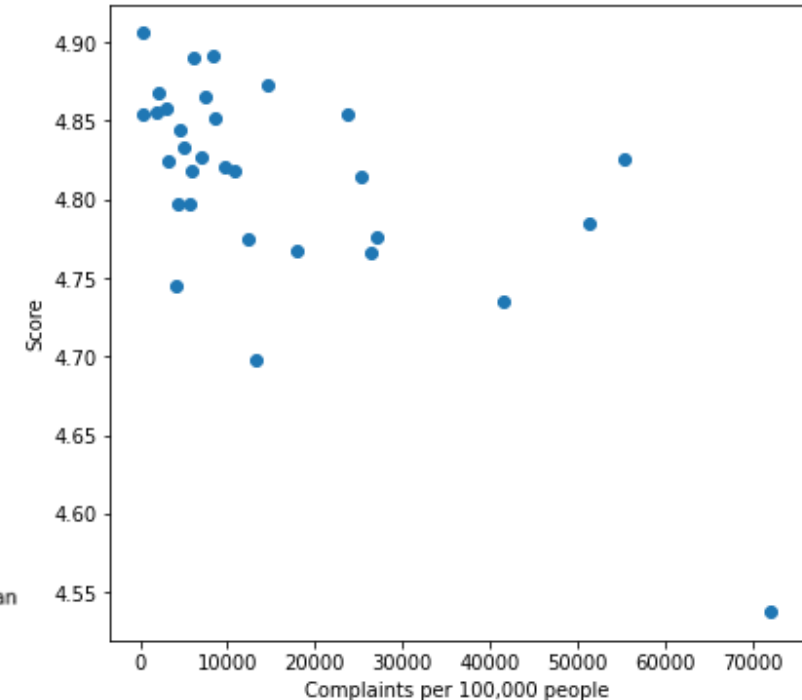
Listing's avg Comment Subjectivity



Score (overall) by region (New York)



Score (overall) by region (San Francisco)



Initial Models

Direct Model

- Some significant features: Price, amenities, comment polarity, comment subjectivity
- Comment polarity and subjectivity were positively correlated with overall score

External Model

- Two cities: New York, San Francisco
- Two targets: Overall rating, location rating
- Crime rate was negatively correlated with scores
- Home value (Zillow) was positively correlated with location score

Combined Model: Significant Predictors

Model type → City ↓	Overall score	Location score
New York	Amenities (+) Instant bookable (-) Comment polarity (+)	Home value (+) Number of people accommodated (-) Comment polarity (+)
San Francisco	Availability (-) Calculated host listings count (-) Price (+)	Home value (+) Calculated host listings count (-) Price (+)