



Price Prediction & Feature Determination

Keith Tan

Agenda



Problem Statement & Approach



Visualizations

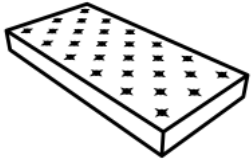


Results



Future Improvements

Problem Statement



Founded Aug 2008

Online marketplace for
accommodation and
experiences



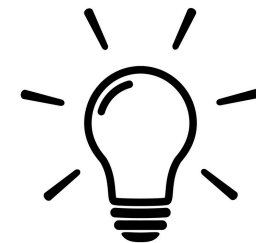
Income loss for
homeowner and Airbnb

Using New York as a case study, I would like to predict the daily price of listings, and in so doing, understand the factors that have significance influence on the price (e.g. location, quality of host)



Lack of checkbox, guidelines

Best guess on appropriate price



Smart Pricing feature

Not so smart



Problem Statement



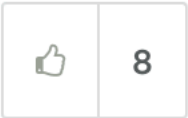
Christine615 in Kansas City, MO

Level 10

11-14-2017 11:17 AM

I'm with you. Airbnb suggested I lower my minimum price to \$68 a night. Our apartment is nine rooms (including three bedrooms). My friends have a carriage house for \$119 and are busy. They're space is about 25% of the size of mine. We're a block apart. I'm at \$125 and I'm not going lower. The algorithm should take into account size of space and not which newcomer is advertising a fake teaser rate. One Airbnb advertises \$68/night only for you to find its only available for a single night and the regular price is as high as \$210.

I decided what makes it worth my while to list my place and for the most part I ignore the suggested changes.



Inaccurate representation

Recommended to halve price



Blind mimic to similar listing

Reply

Approach



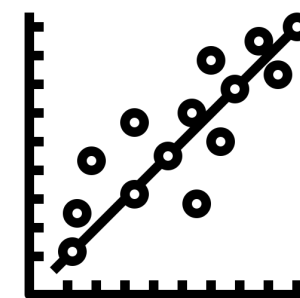
Data Scraping

1 year worth of data from
Inside Airbnb



Merge, clean, analyze

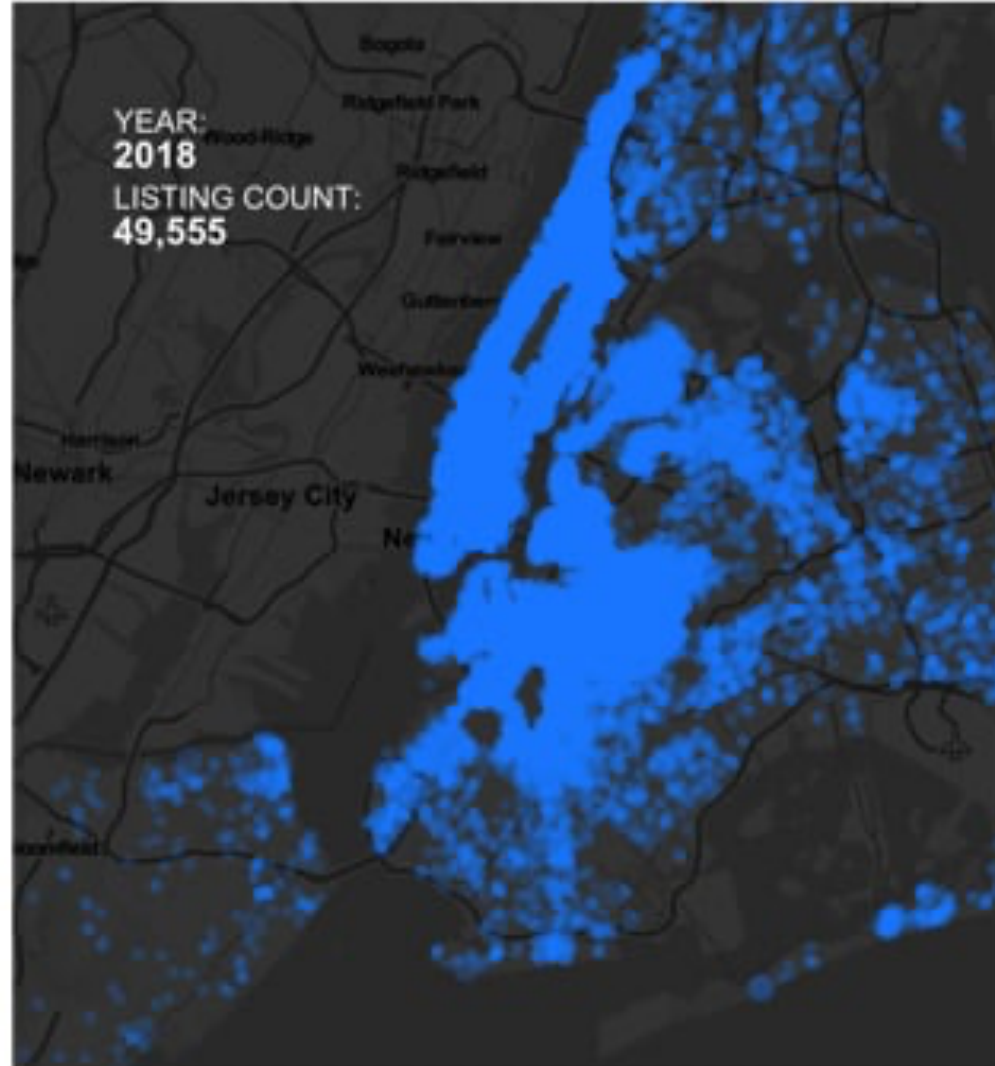
Reviews polarity analysis



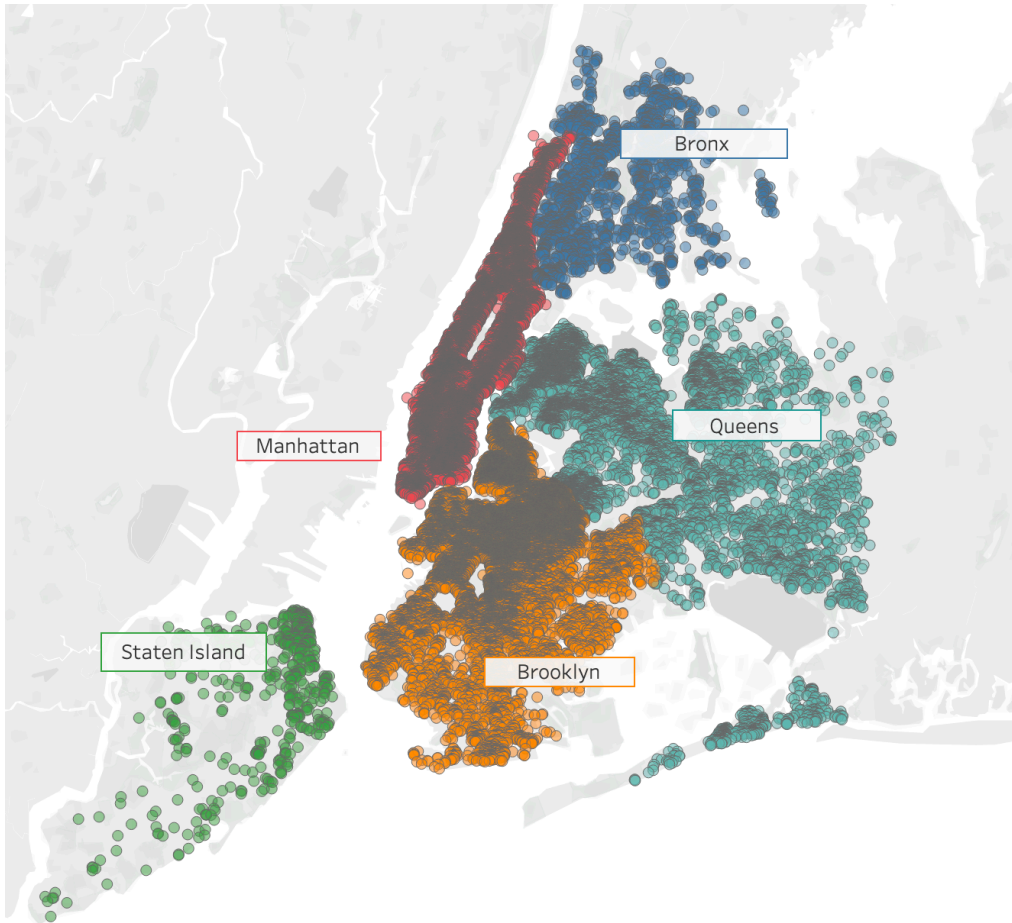
Model

GridSearch on multiple
regression models

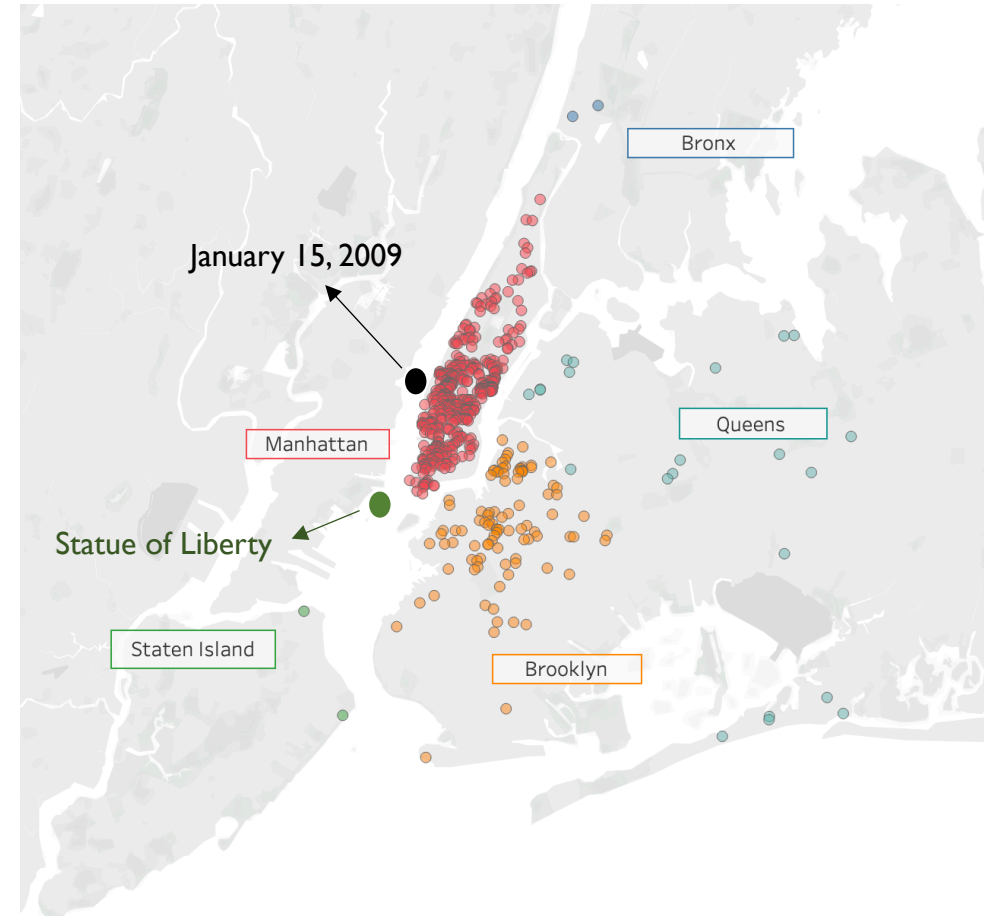
Visualizations



Visualizations



Listings with price < \$700



Listings with price > \$700

Visualizations



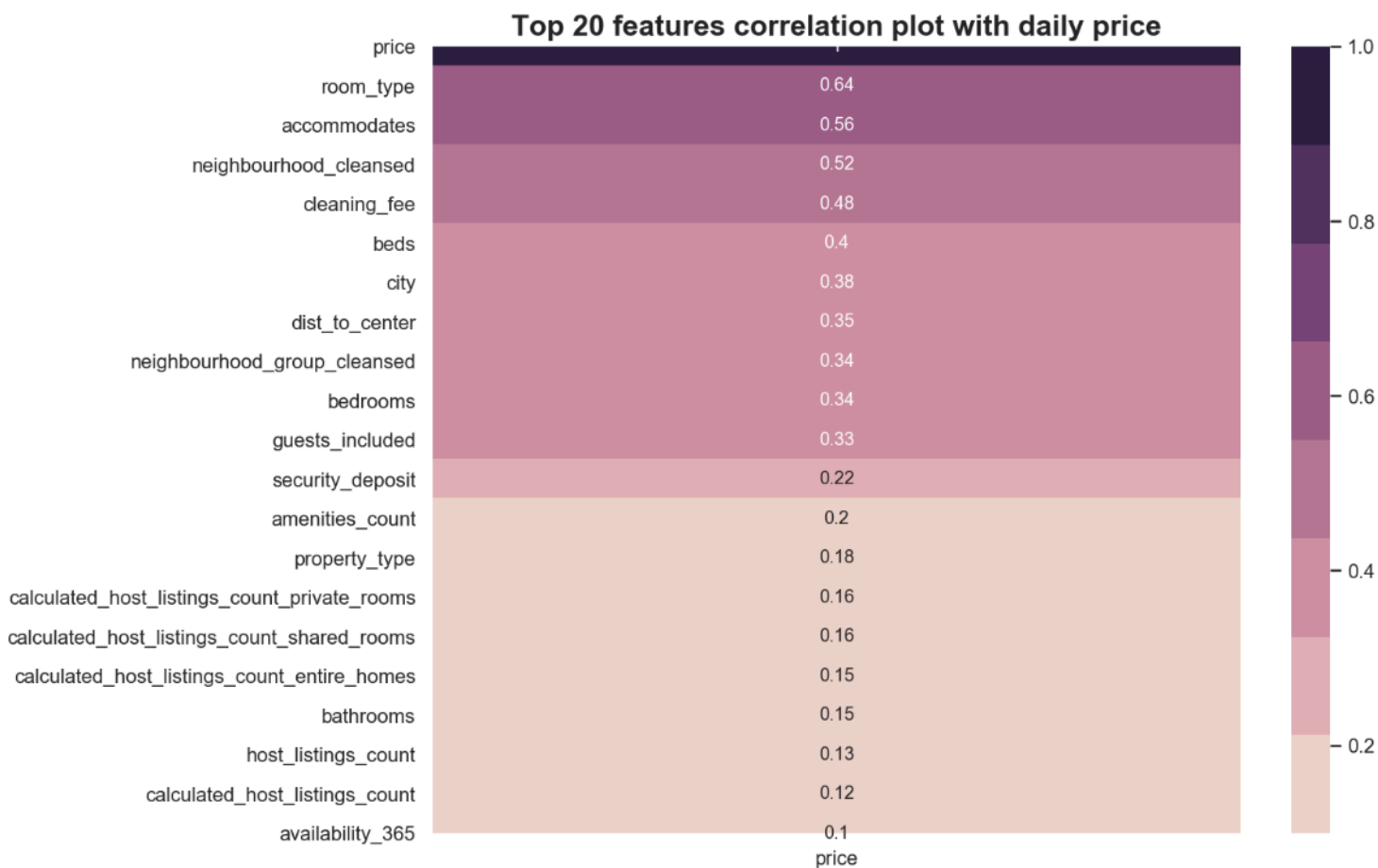
Word cloud for Positive Reviews



Important factors:

- 1) Locations, especially proximity to key facilities, popular spots
- 2) Quality of listing
- 3) Host-specific factors

Visualizations



Expected
Vs
Unexpected

Feature Grouping

Results



	Model	Train Score	Test Score	Best Parameters
8	XGBRegressor	0.7746	0.7859	{'xg__learning_rate': 0.1, 'xg__loss': 'lad', ...}
3	KNeighborsRegressor	0.6834	0.6988	{'knr__n_neighbors': 8}
6	RandomForestRegressor	0.6656	0.6669	{'rf__max_depth': 5, 'rf__min_samples_leaf': 3...
7	ExtraTreeRegressor	0.6651	0.6655	{'et__max_depth': 5, 'et__min_samples_leaf': 1...
0	LassoCV	0.6753	0.6652	{'la__n_alphas': 10}
5	DecisionTreeRegressor	0.6506	0.6552	{'dt__max_depth': 5, 'dt__min_samples_leaf': 2...
9	AdaBoostRegressor	0.6164	0.6176	{'abr__learning_rate': 0.1, 'abr__n_estimators...
2	ElasticNetCV	0.4933	0.4947	{'elas__l1_ratio': 0.1}
4	SupportVectoRegressor	0.671	0.2776	{'svr__C': 100.0, 'svr__gamma': 1e-05, 'svr__k...
1	RidgeCV	NaN	NaN	NaN

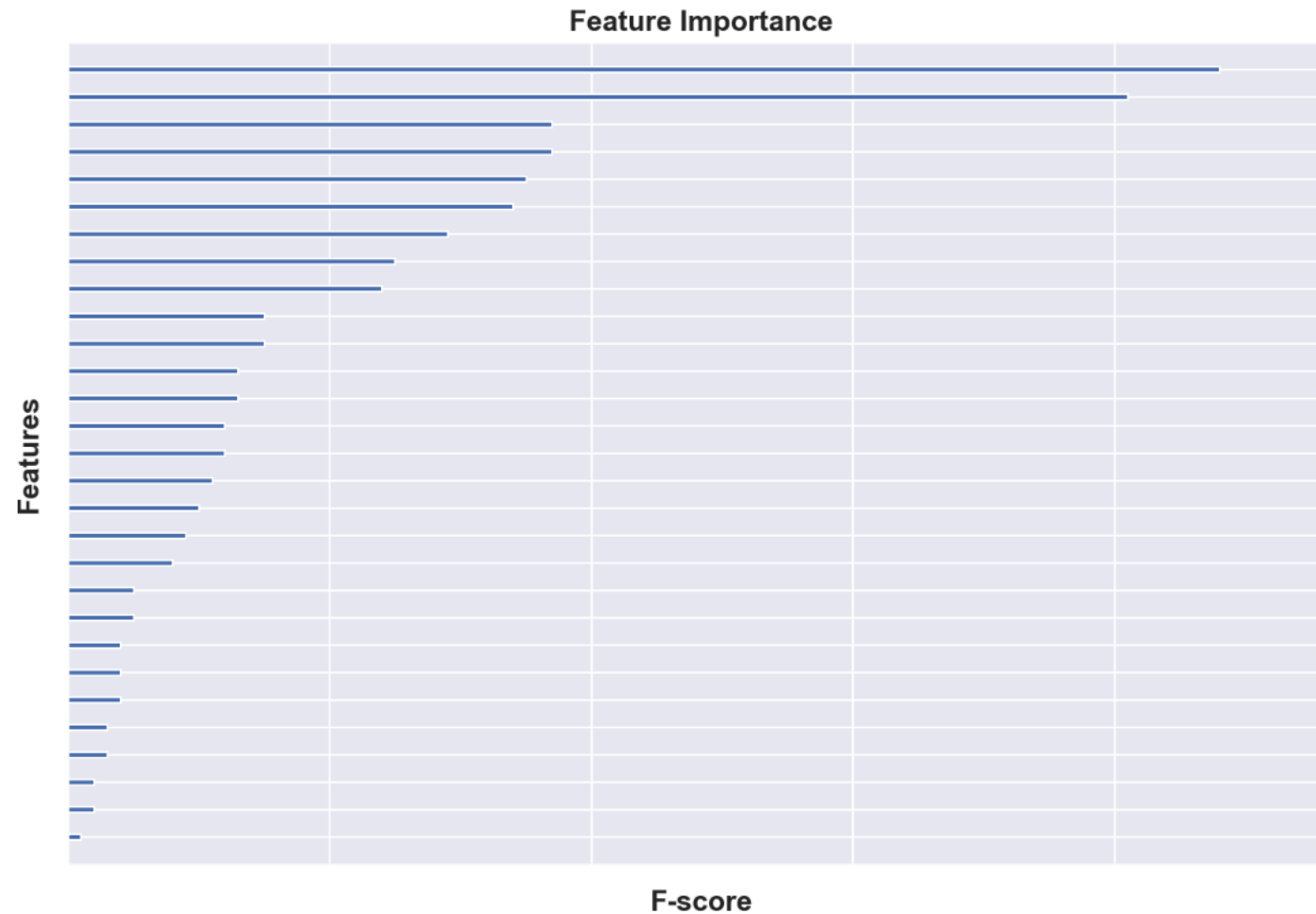
XGBoostRegressor
established the best results...
By a mile.

Model evaluation metrics
include:

R² score: **0.786**

RMSE: **\$ 2.08**

Results



Top 5 Features:

- 1) Occupancy size
- 2) Neighbourhood
- 3) Cleaning fee
- 4) Number of bedrooms
- 5) Minimum nights

Future Improvements



Account for price seasonality



Calling PolynomialFeatures to explore feature interactions



Text analysis for descriptive features



Geographic-based analysis



Thank You!