Popularity Rankings on TripAdvisor

How can hotels improve theirs?

Sort by: Traveler Ranked | Best Value + Top Rated | Lowest Price | Distance

Sponsored

Hotel Beacon

3,398 Reviews

#57 of 468 hotels in New York City

"My second home" 04/11/2017

"My go-to hotel in NYC" 04/07/2017

Mid-range | Upper West Side

Check In | Check Out

Show Prices

414 Hotel

1,294 Reviews

#1 of 468 hotels in New York City

"Home Away From Home" 04/15/2017

"Making the best of an odd situation" 04/12/2017

Mid-range | Midtown | Breakfast included

Check In | Check Out

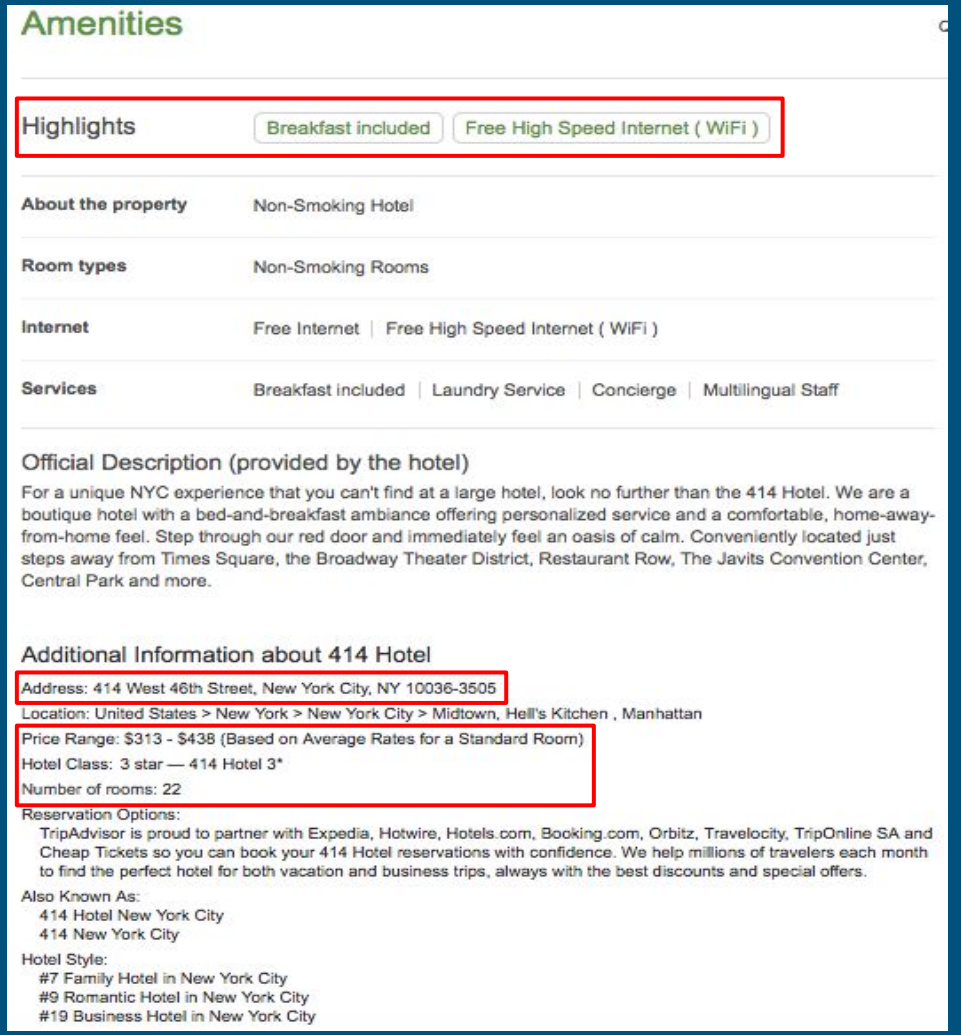
Show Prices

Problem

- Popularity is important because the closer you are to #1, the more likely it is that travelers will see your property when they search your area.
- The three factors used in TripAdvisor's proprietary ranking formula:
  - Quality of reviews
  - Quantity of reviews
  - Age of reviews
- But do hotels have much influence over those factors?
  - Ranking, by definition, is relative
  - Are there tangible and modifiable factors that distinguish high and low performers?
- Data Product: Model predicting rankings with insight into significant/important predictors of those rankings

Data Collection

- TripAdvisor.com
  - Hotel characteristics (e.g., name, location, amenities, price range)
  - Reviews (e.g., text, review date, month of stay, season of stay, type of travel, rating (1-5))
- 467 hotels identified under US > New York (NY) > New York City
- 100,000 reviews scraped (used a pipeline to save data in PostgreSQL)



Amenities

Highlights: Breakfast included | Free High Speed Internet (WiFi)

About the property	Non-Smoking Hotel
Room types	Non-Smoking Rooms
Internet	Free Internet | Free High Speed Internet (WiFi)
Services	Breakfast included | Laundry Service | Concierge | Multilingual Staff

Official Description (provided by the hotel)

For a unique NYC experience that you can't find at a large hotel, look no further than the 414 Hotel. We are a boutique hotel with a bed-and-breakfast ambiance offering personalized service and a comfortable, home-away-from-home feel. Step through our red door and immediately feel an oasis of calm. Conveniently located just steps away from Times Square, the Broadway Theater District, Restaurant Row, The Javits Convention Center, Central Park and more.

Additional Information about 414 Hotel

Address: 414 West 46th Street, New York City, NY 10036-3505

Location: United States > New York > New York City > Midtown, Hell's Kitchen , Manhattan

Price Range: $313 - $438 (Based on Average Rates for a Standard Room)

Hotel Class: 3 star — 414 Hotel 3*

Number of rooms: 22

Reservation Options:

TripAdvisor is proud to partner with Expedia, Hotwire, Hotels.com, Booking.com, Orbitz, Travelocity, TripOnline SA and Cheap Tickets so you can book your 414 Hotel reservations with confidence. We help millions of travelers each month to find the perfect hotel for both vacation and business trips, always with the best discounts and special offers.

Also Known As:

414 Hotel New York City

414 New York City

Hotel Style:

#7 Family Hotel in New York City

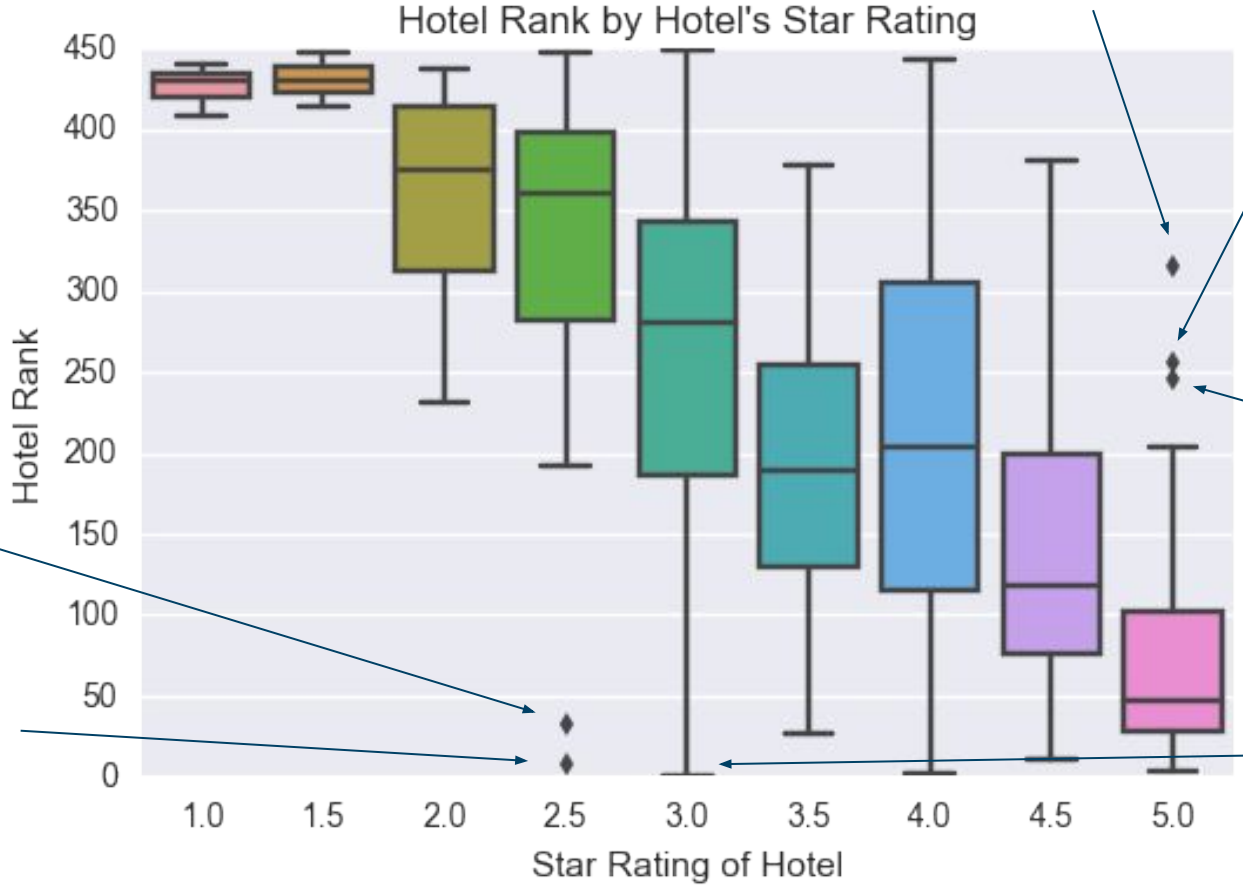
#9 Romantic Hotel in New York City

#19 Business Hotel in New York City

Pre-processing

- Removed 17 hotels without a rank and 4 hotels outside of Manhattan
- Cleaned hotel addresses and geocoded using Google Maps API
- Imputed missing values of number of rooms in hotel (6), hotel class (28), and price range (22) using k-Nearest Neighbors
  - Hotel class values were then rounded to nearest 0.5 to allow for categorical treatment
- Removed reviews without text content or without hotel stay summary (< 2%)
- Removed 6 hotels without reviews
- Final hotel count: 440
  - Shifted ranks to account for gaps created by removing hotels
- Cleaned review text & aggregated to hotel level

EDA



Four Seasons
Opened 1993
60 reviews (earliest Dec 2016)

The Lowell (1927)
270 reviews

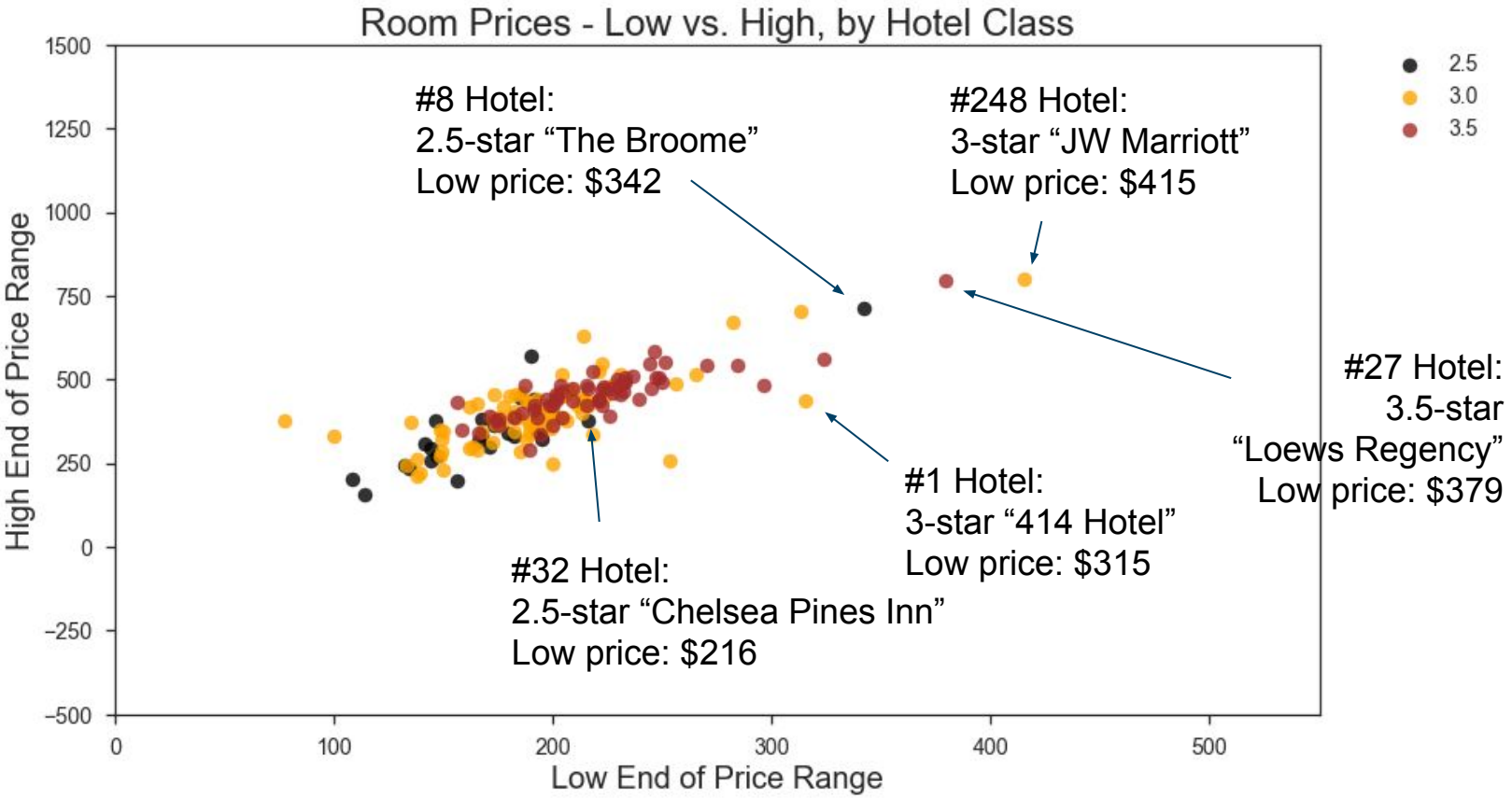
The Carlyle
Upper East Side
"Old School
Elegance"

#32 "Chelsea Pines Inn"
B&B, 26 rooms
"Friendly wonderful staff"
"Great location"
"Cute", "Unique"

#8 "The Broome":
14-room hotel
"small/boutique"
"European feel"
"Inner courtyard"
"Beautiful rooms"

#1 Hotel:
"414 Hotel"
22 rooms
"welcoming"

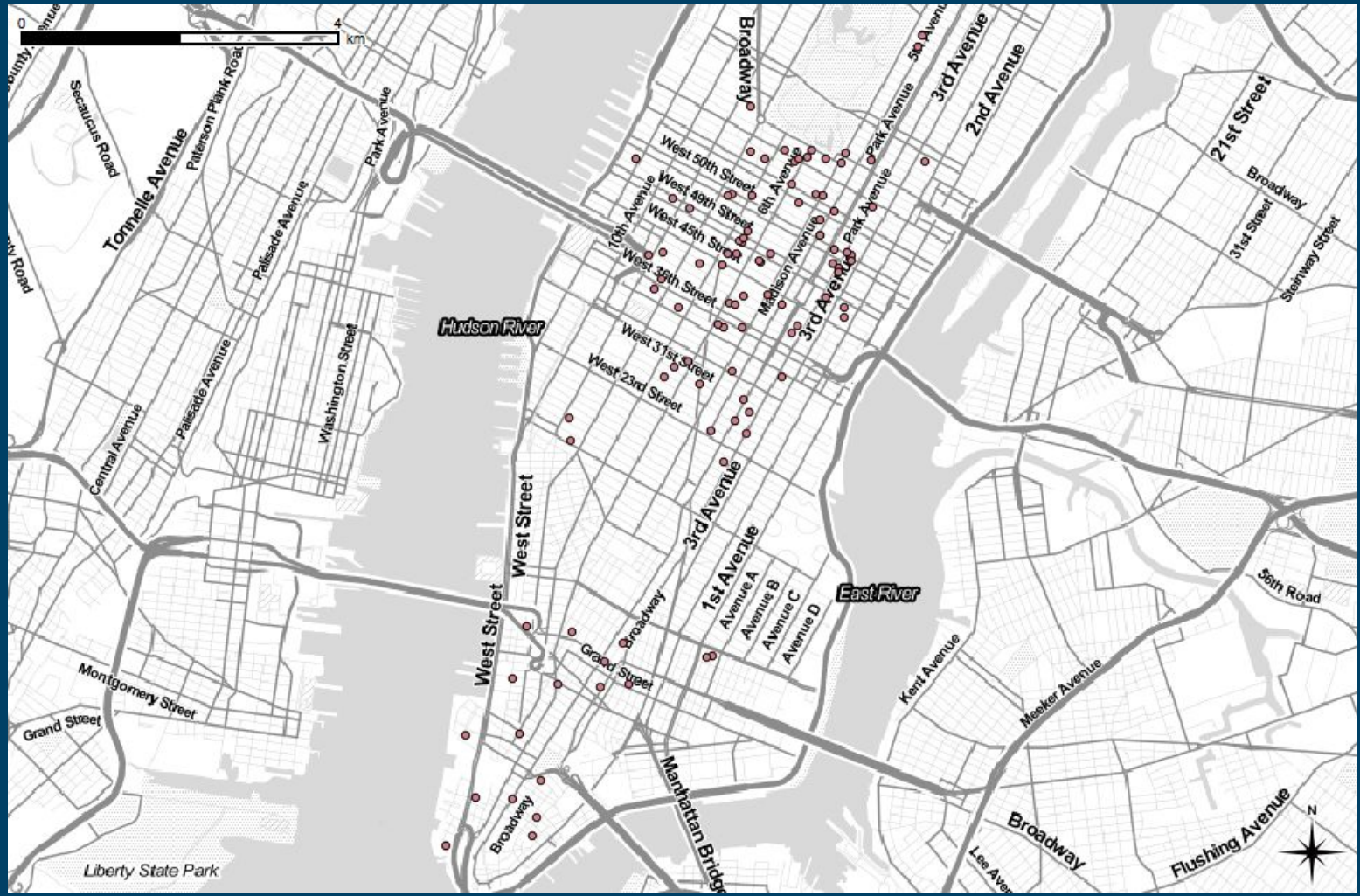
EDA


Room Prices - Low vs. High, by Hotel Class
High End of Price Range
Low End of Price Range
1500
1250
1000
750
500
250
0
−250
−500
0
100
200
300
400
500
2.5
3.0
3.5
#8 Hotel:
2.5-star “The Broome”
Low price: $342
#248 Hotel:
3-star “JW Marriott”
Low price: $415
#27 Hotel:
3.5-star
“Loews Regency”
Low price: $379
#1 Hotel:
3-star “414 Hotel”
Low price: $315
#32 Hotel:
2.5-star “Chelsea Pines Inn”
Low price: $216

Map of NYC

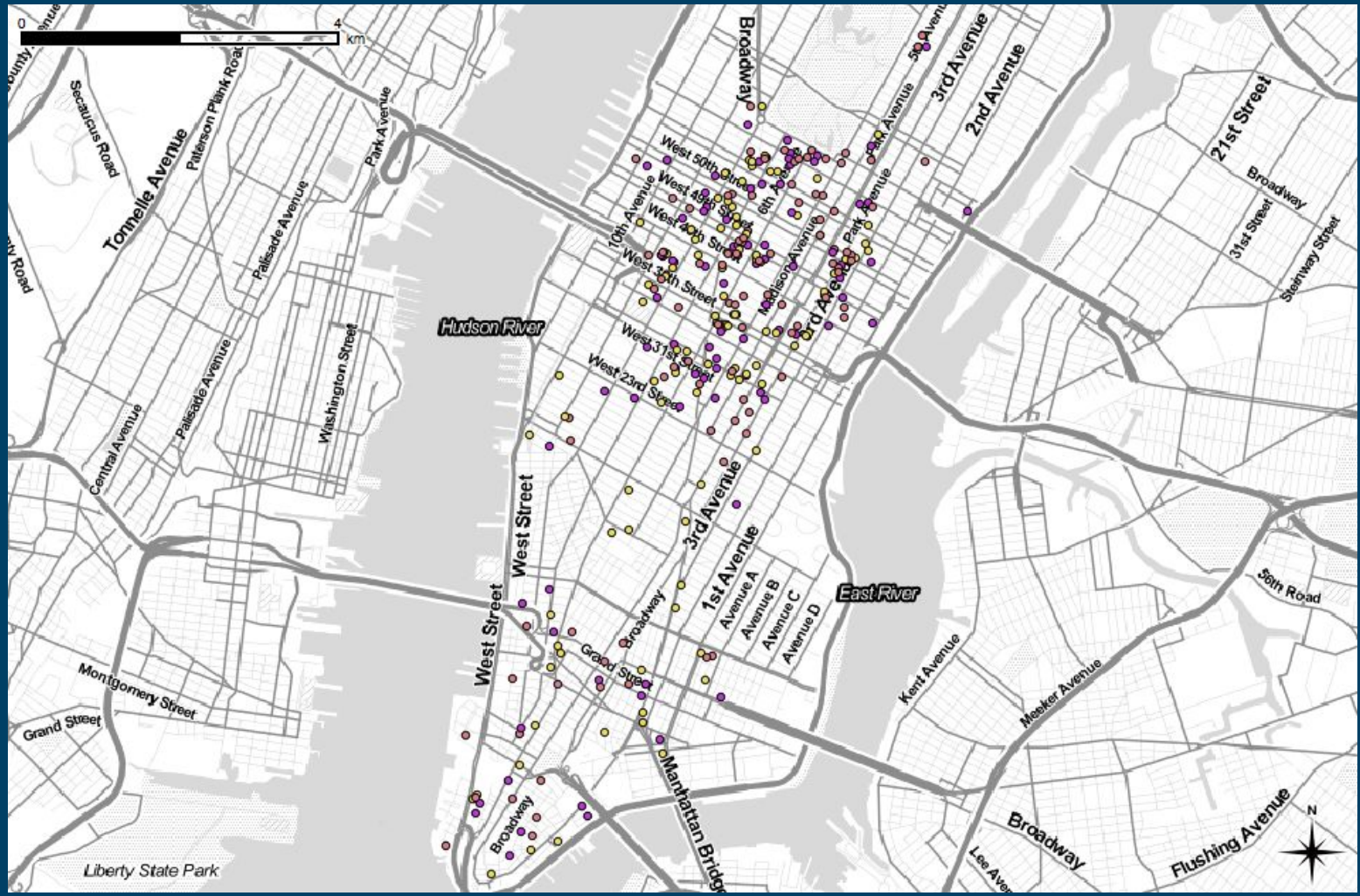

0
4
km
Secaucus Road
Tonnelle Avenue
Paterson Plank Road
Park Avenue
Palisade Avenue
Washington Street
Central Avenue
Hudson River
West 50th Street
West 49th Street
West 45th Street
West 36th Street
West 31st Street
West 23rd Street
10th Avenue
6th Avenue
Madison Avenue
Park Avenue
3rd Avenue
2nd Avenue
5th Avenue
Broadway
21st Street
31st Street
Steinway Street
West Street
1st Avenue
Avenue A
Avenue B
Avenue C
Avenue D
East River
Grand Street
Montgomery Street
Manhattan Bridge
Kent Avenue
Meeker Avenue
58th Road
Flushing Avenue
Liberty State Park
N

Map of NYC Hotels Ranked 1 - 100

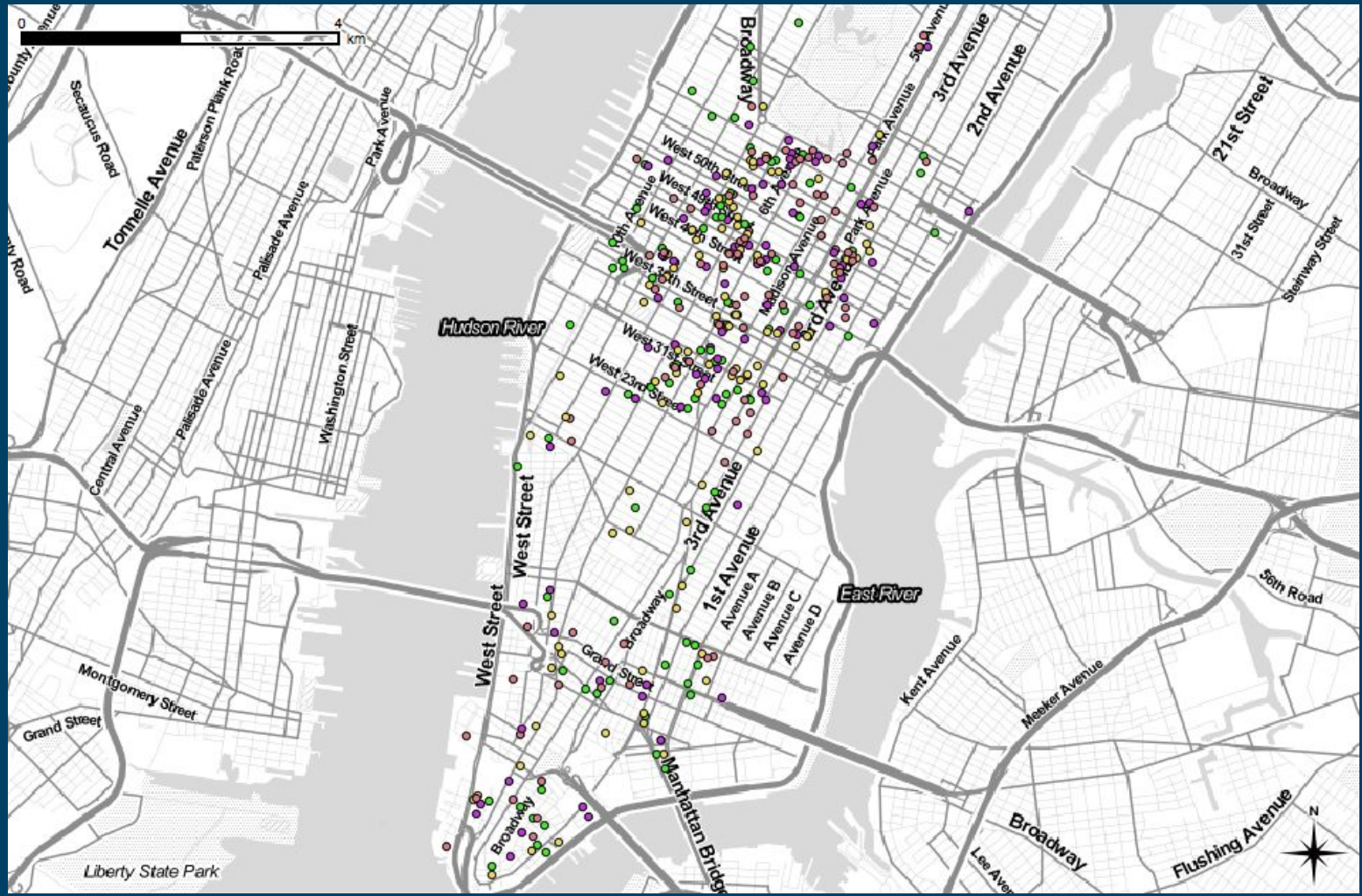


Map of NYC Hotels Ranked 1 - 200

Map of NYC Hotels Ranked 1 - 300



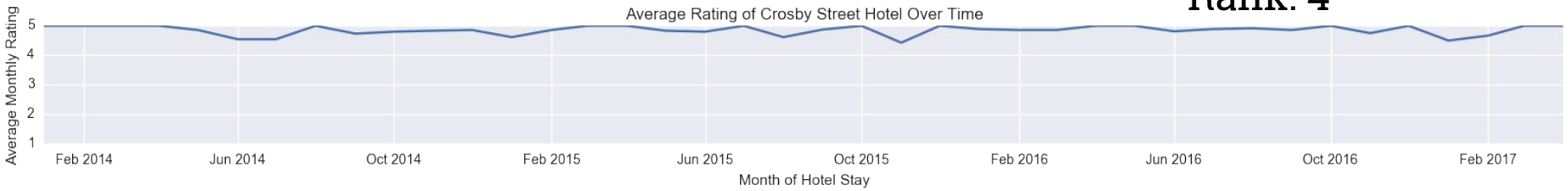
Map of NYC Hotels Ranked 1 - 400



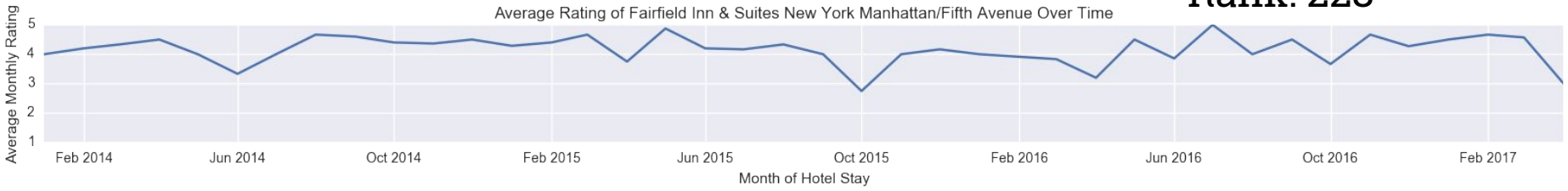
Map of NYC Hotels Ranked 1 - 446

Quality of Reviews Over Time

Rank: 4



Rank: 228



Rank: 417

Pairwise Comparison Model

- Set-up:
  - 440 hotels converted into (440 choose 2) combinations ⇨ 96,580 pairings
  - Label: 1/0 for whether Hotel 1 or Hotel 2 in each pair is ranked higher
  - Feature set: Hotel 1's and Hotel 2's characteristics
    - 53 features per hotel: 31 zip code dummies, 9 hotel class dummies, 10 amenities dummies, number of rooms, low and high ends of price range
  - ⅔ - ⅓ Train-Test split
- Deriving predicted ranks from pairwise classification:
  - For each hotel, count hotels that are ranked lower, then sort the hotels by that count

Pairwise Comparison Model

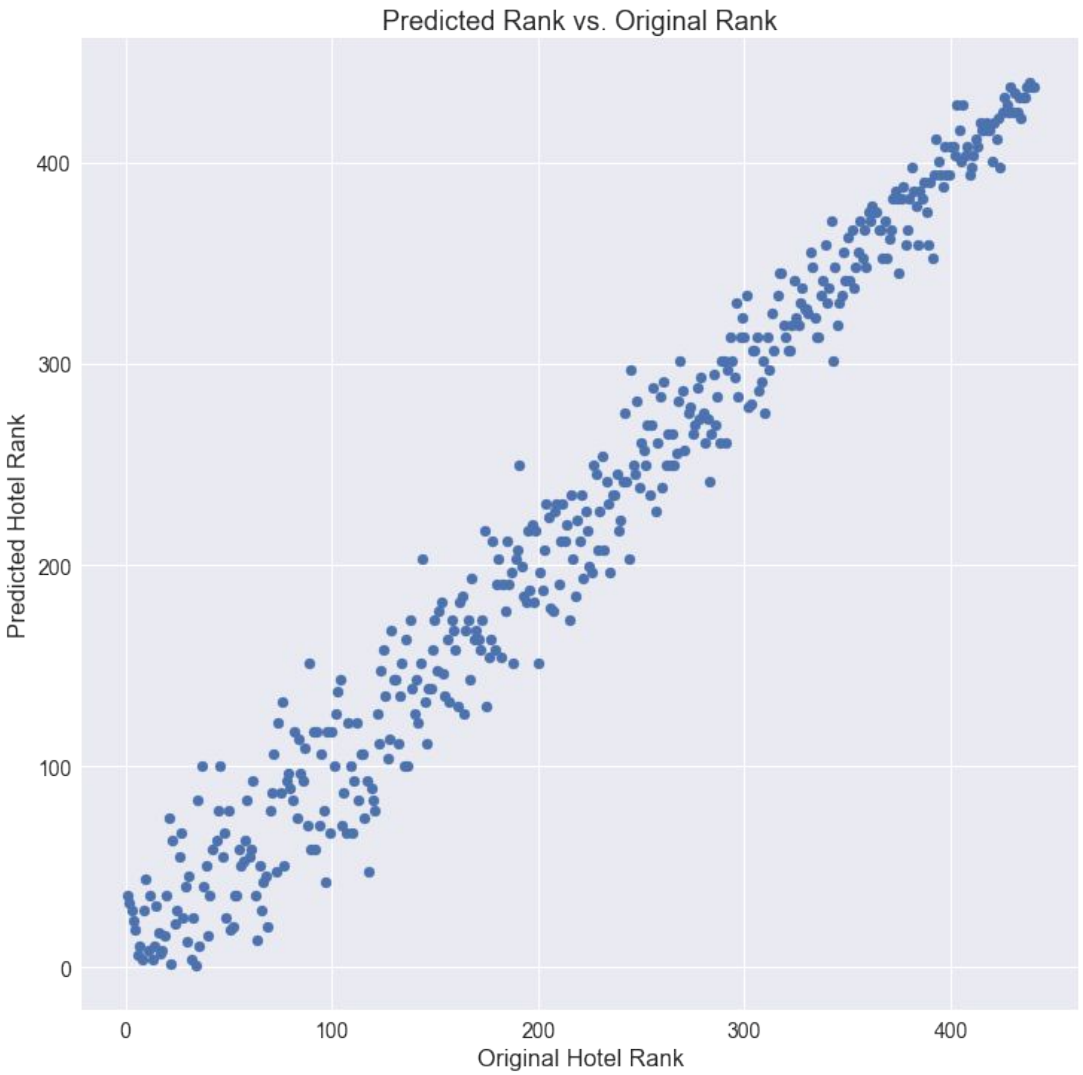
- Results (with hyperparameter search):
  - Baseline: 50.44% (Test data)
  - Logistic Regression:
    - Best model CV score: 76.02%
    - Full training data score: 76.19% Test score: 76.10%
  - Stochastic Gradient Descent:
    - Best model CV score: 75.96%
    - Full training data score: 76.08% Test score: 75.89%
  - Random Forest:
    - Best model CV score: 89.74%
    - Full training data score: 97.44% Test score: 90.10%
  - XGBoost:
    - Best model CV score: 96.24%
    - Full training data score: 99.04% Test score: 96.72%

Results

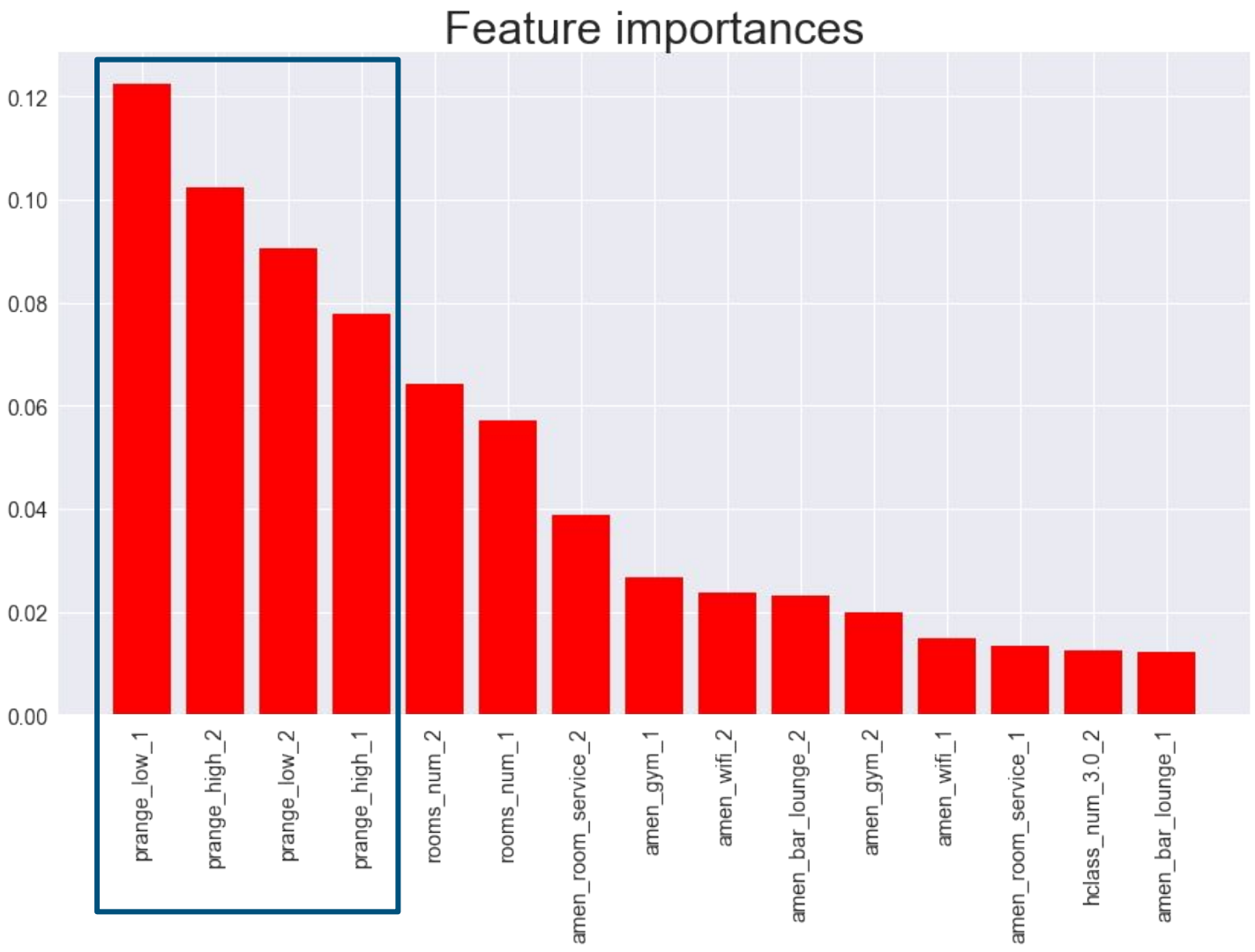
Spearman's Rank-Order Correlation (measures the strength and direction of the monotonic relationship between two variables):

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

= 0.9862 (p-value: 0.0) for Test data



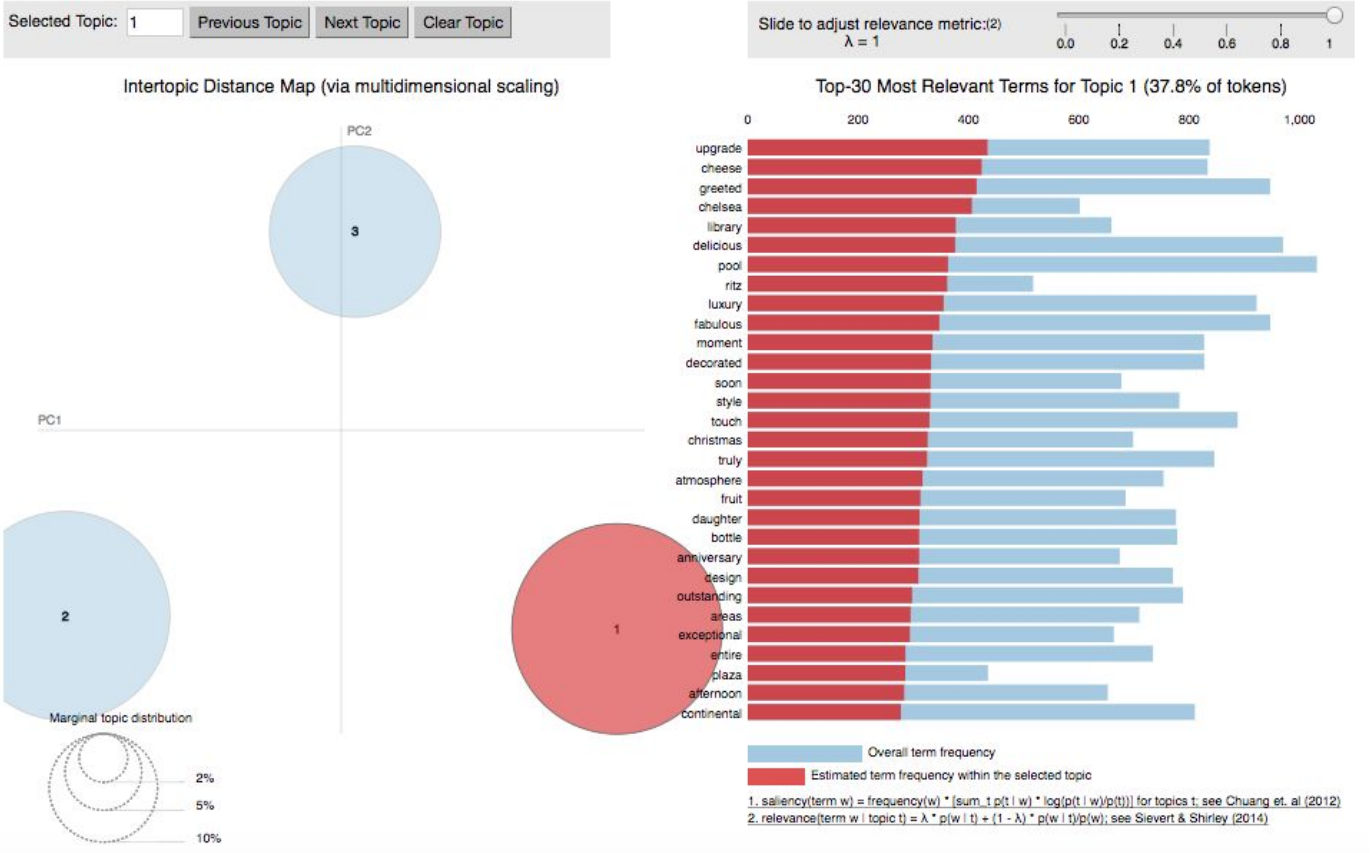
Prices Are Best Predictors



Topic Modeling

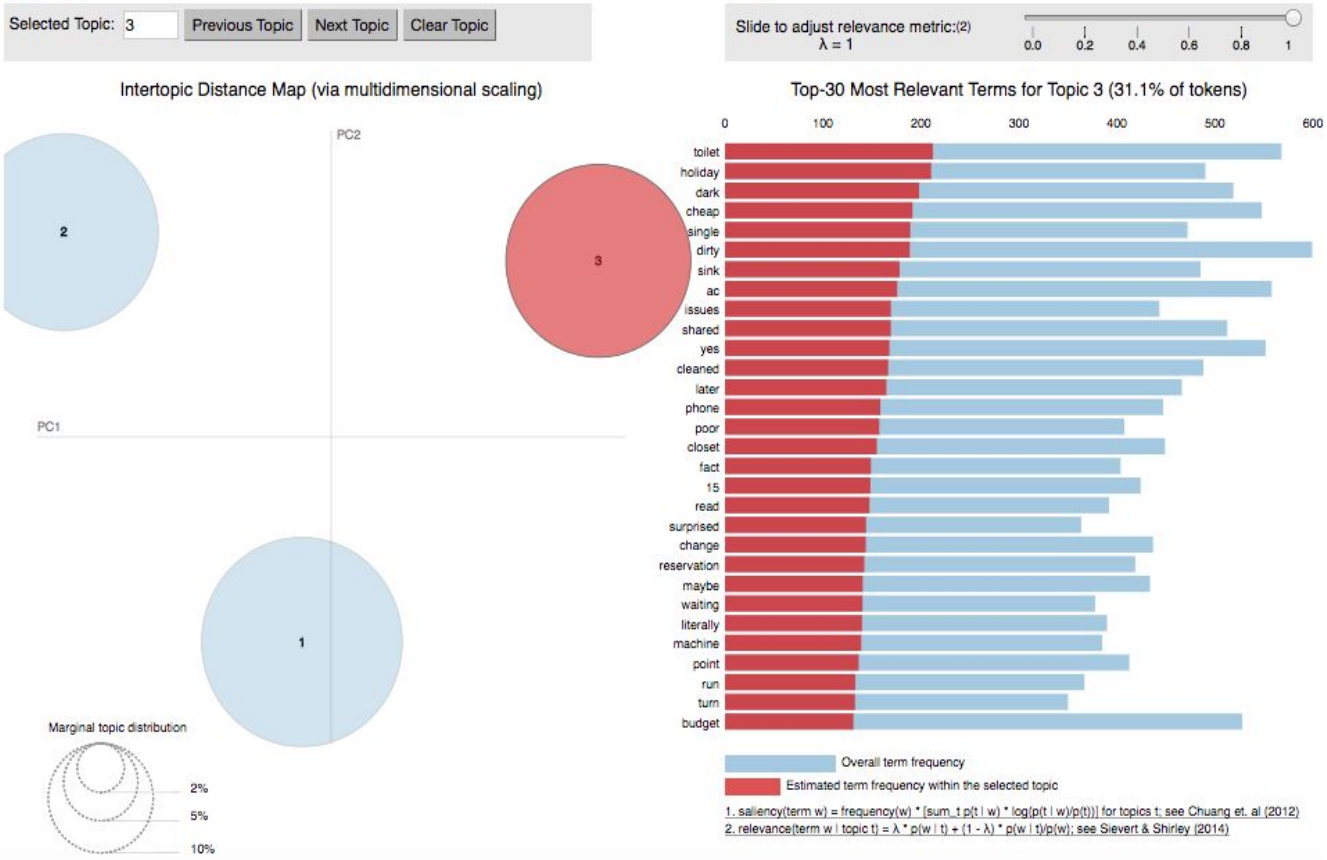
- Topic Models:
  - Text-mining tool for discovering abstract “topics” that occur in a collection of documents
  - Topics produced by topic modeling are clusters of similar words:
    - E.g., document about dogs may have a cluster of “dog” and “bone”
- Set-up/pre-processing:
  - Change text to lowercase, remove punctuation, remove stopwords
  - Split hotels into four tiers of ranks (1-100, 101-200, 201-300, 301-400)
  - Remove words appearing more than 2,500 times across all reviews in one tier
    - For reference, “hotel” appears 142K times and “comfort” appears 1,500 times in full review corpus

Top Tier Hotels: Review Topic #1



Upgrade
Cheese
Greeted
Chelsea
Library
Delicious
Pool
Ritz
Luxury
Fabulous
Moment
Decorated
Soon
Style
Touch
Christmas
Truly
Atmosphere
Fruit

Bottom Tier Hotels: Review Topic #3



Toilet
Holiday
Dark
Cheap
Single
Dirty
Sink
AC
Issues
Shared
Yes
Cleaned
Later
Phone
Poor
Closet
Fact

Conclusions and Next steps

- Price/luxury ⇨ Better ranking
- Hotels can use these results to help them decide:
  - which amenities/features to offer to guests,
  - whether to just buy sponsored listings (and to offer deals in support of those listings),
  - or, whether it even makes sense to place themselves on TripAdvisor

- Product improvements:
  - Incorporate time-dependent features (e.g., rankings over time, rating volatility)
  - Train/validate models with data from other cities