Check-in To Reality

Beyond the Star Rating

FACULTY OF DATA AND DECISION SCIENCES
MAKING SENSE OF IT

TECHNION Israel Institute of Technology

Amit Gertner 209399336 Dekel Elimelech 205852486 Shoval Zvieli 314947797

Why Check-in To Reality?

Ratings hide important details

Hotels with similar scores often deliver very different experiences

Overwhelming information

Amenities and marketing language create false expectations

Our mission

Reveal systematic **gaps between what hosts promise** and **what guests** actually experience

Data Sources

Booking.com listings & reviews

External Enrichment ~200k records

- **OpenStreetMap**

Transport, Leisure, Dining, Tourism and Culture

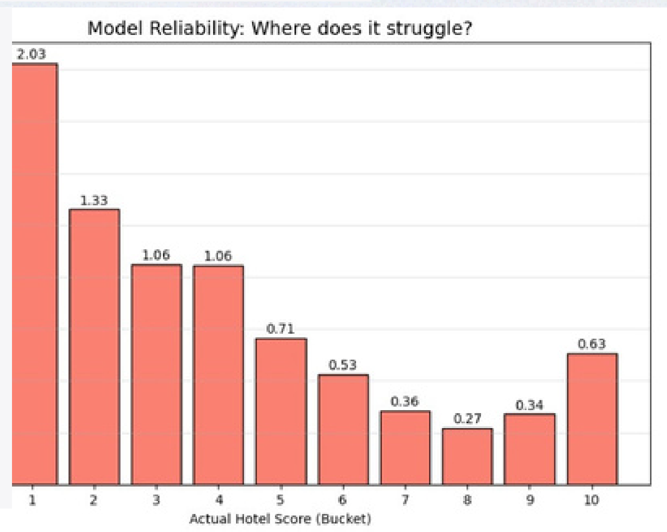
- **Google Maps**

Reviews and Ratings

- **Numbeo**

Crime and Safety

Prediction errors increase for extreme low ratings, highlighting reduced reliability for rare negative experiences.



Our Methodology

Quantifying the Expectation—Reality Gap

1 Expectation (Host Claims)

NLP applies to titles, descriptions and ammenties

KeyWord-based claim detection:

Location: central, accessible

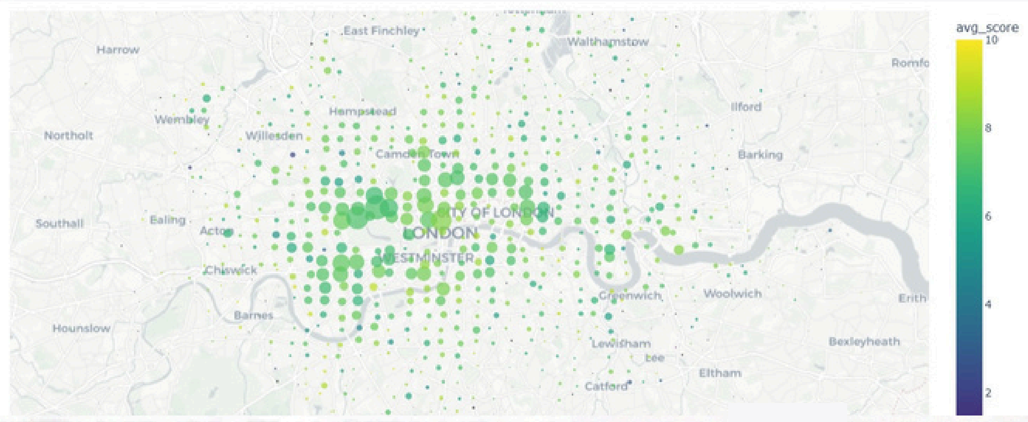
Atmosphere: quiet, peaceful

Quality: luxury, modern, spacious

2 Reality Modeling

Spatial Data Intergation between core dataset and External datasets

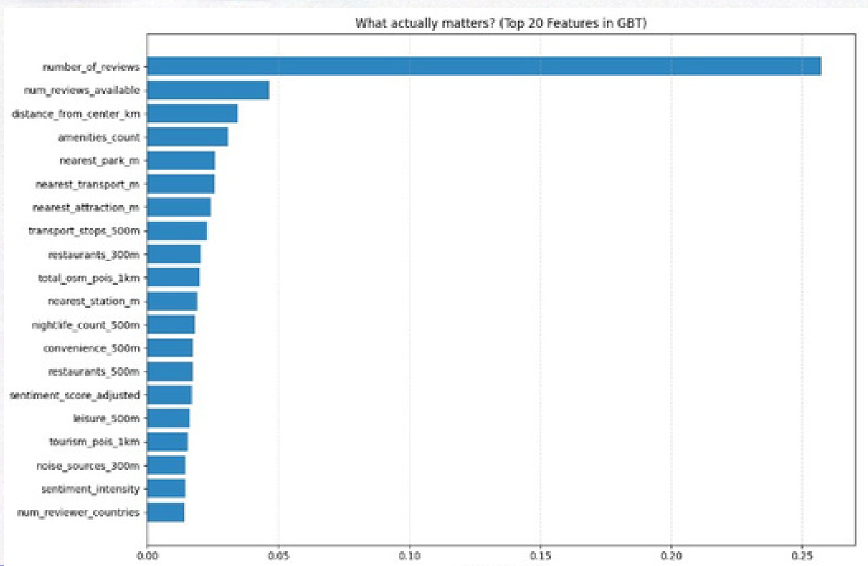
Process: Unfied spatial join using a Geohash-based grid strategy to merge hotel data with external POIs

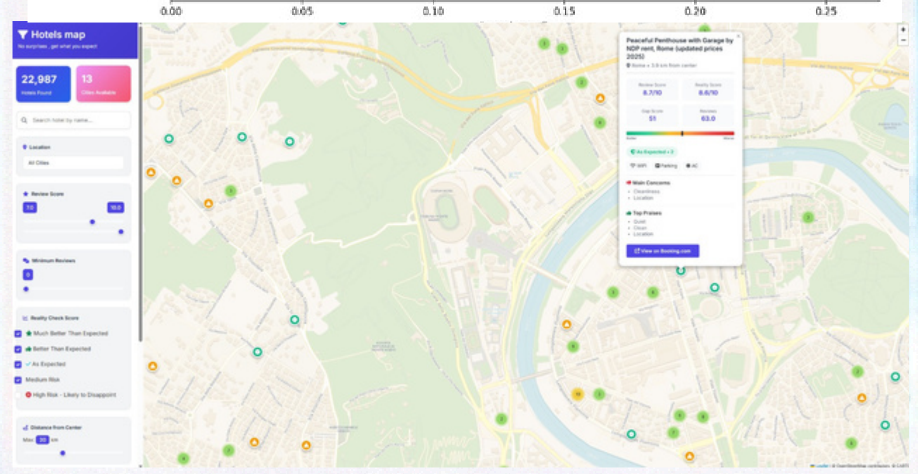


3 Spatial Feature Engineering

Reality & Experience Engineering: Calculating objective metrics such as nightlife density, transport proximity, and safety indices.

Guest Experience: Hybrid NLP on reviews to extract sentiment and specific complaint ratios.





4 Gap Signal Engineering

Explicitly encoding the discrepancies between Step 1 Expectations and Step 3 Real-world Features

Noise Gap claims quiet vs nightlife or noise complaints

Location Gap claims central vs actual transport proximity

Sentiment Gap claims numeric rating vs actual sentiment

These signals reveal hidden dissatisfaction, providing a data-driven layer of truth that goes beyond standard numerical ratings.

Our Process & Results

5 Predictive Modeling

Model: Gradient boosted Trees (GBT) Regressor

Architecture: Sequential boosting stages using Spark ML

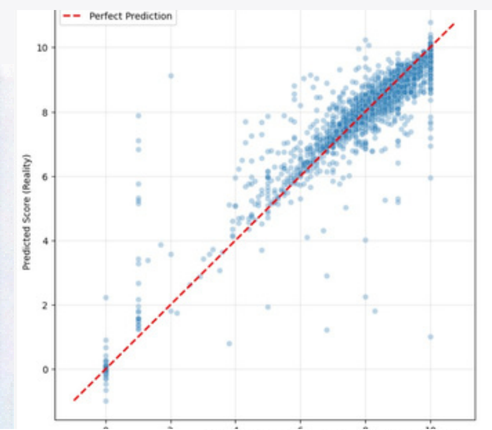
Input: approxamialy 150 features

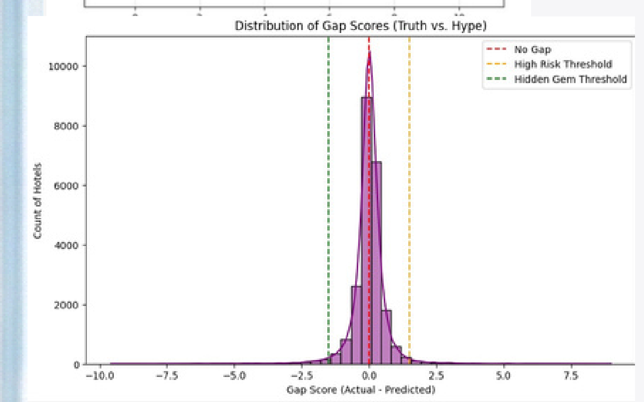
Training: 80/20 train-test split on 0-10 ratings

Model performance:

$R^2 = 0.87$. Explains rating variance

$RMSE = 0.84$ Prediction error < 0.9





Key Insights

Not all amenities matter: Only specific amenities and transport access consistently impact satisfaction

Location beats marketing: Objective neighborhood context explains guest experience better than listing claims

Ratings hide frustration: High scores can still mask negative review sentiment

Expectation gaps are common: "Quiet" and "central" claims are often contradicted by reality

Conclusion: Check-in To Reality turns hidden gaps between hotel promises and guest experiences into clear, actionable insights–directly at booking time.