Prycer

Helping first-time Airbnb hosts list right

Andrea Sy, Charlene Chen,
David He, Rory Liu

Prycer Objective

Pricing guidance

$10 difference in listing price sometimes lead to $200 difference in monthly revenue

Listing suggestion

Highlighting the right features that resonate with guests can significantly increase click-through and therefore occupancy

Prycer's mission is to be a **comprehensive resource** for **first time hosts**

The Prycer MVP

Minimum Viable Product

User needs

- Understanding of **competition**
- Suggestion on **base price** & corresponding **revenue**
- Suggestion on **amenities / features to highlight**
- Identification of **improvement potential**

Product flow

- Enter property features
- Neighbor dashboard
- Base price engine: Price-occupancy chart
- Post highlight suggestion
- Improvement assessment
- New price engine

Live Demo

We use structured and unstructured data to achieve our mission

Structured data to understand pricing

- **Listing data** that contains property features, price and review per month
- **Calendar data** for pricing over 12 months
- **160k+ listings** analyzed

Unstructured data to understand listing highlights

- **Review data** that contains all review verbiage for all listings
- **5 million reviews** analyzed for 11 cities

Our architecture is designed to provide actionable insights to hosts

Calendar

Listings

Data cleaning & feature engineering

Occupancy model: *how often would a property be booked given location, price and amenities?*

+

User Input

Location, property characteristics, amenities

Optimization curve
Using occupancy model to predict occupancy for the same listing at different prices to

Reviews

LDA* topic modeling by city

Manual topic processing by city

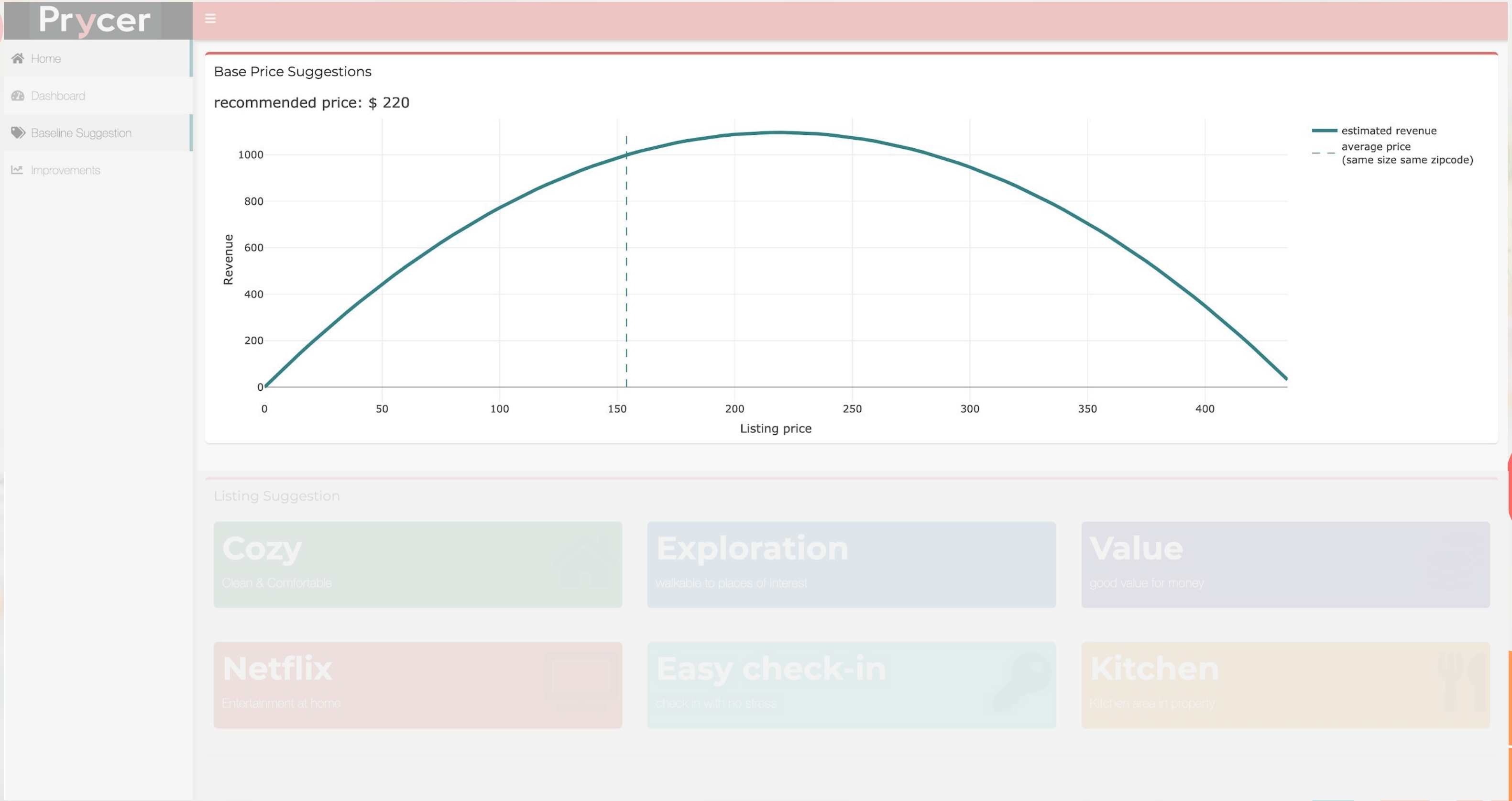
+

Listing specific highlights (e.g. amenities)

General highlights (applies to all listings in city)

*LDA: Latent Dirichlet Allocation

Let's first look at how we provided pricing guidance



First: why do we predict occupancy (and not price directly?)

Price prediction

- Use listing features to predict price
- Like predicting real estate price

No insights for revenue potential

Result is a single price point

Our approach

- Use price as an input to predict occupancy
- Feed model with same listing features but different prices to obtain estimated occupancy for each price
- Estimated revenue = occupancy * price

Monthly revenue estimation

Results of full price spectrum

We believe that there are 5 main drivers for occupancy

	Hypothesis	Feature engineering	
Price	Higher price -> lower occupancy	Market price	← Average price for same-sized listing in same zip code
		Listing premium	← Listing price - market price
Location	Better locations (i.e. touristy, convenient) -> higher occupancy	City	← Binary features for each city
		Neighborhood	← Clustered zip codes into “downtown”, “suburbs”, etc.
Type	Hotel / service apartment -> higher occupancy (vs. apartments)	Binary features for main property types	
Amenity	More amenities -> higher occupancy	Binary features for main amenity groups (e.g. kitchen appliances)	
Size	Property Size (all else equal) -> TBD occupancy	# of bedrooms and # of bathrooms	← Larger # lower occupancy
		max. guest accommodated	← Larger # higher occupancy

We chose linear regression with regularization as our final model because of its ability to interpolate

Baseline: occupancy ~ market price + listing price + bedrooms

Adjusted R2: 0.0057, RMSE: 0.229

LASSO regression

with 40 features across the 5 categories

Adjusted R2: 0.2853

RMSE: 0.1925

Random Forest

with 40 features

Adjusted R2: 0.7014

RMSE: 0.1791

Xgboost

with 40 features

Adjusted R2: 0.0485

RMSE: 0.2222

Stacked models

LASSO -> Xgboost

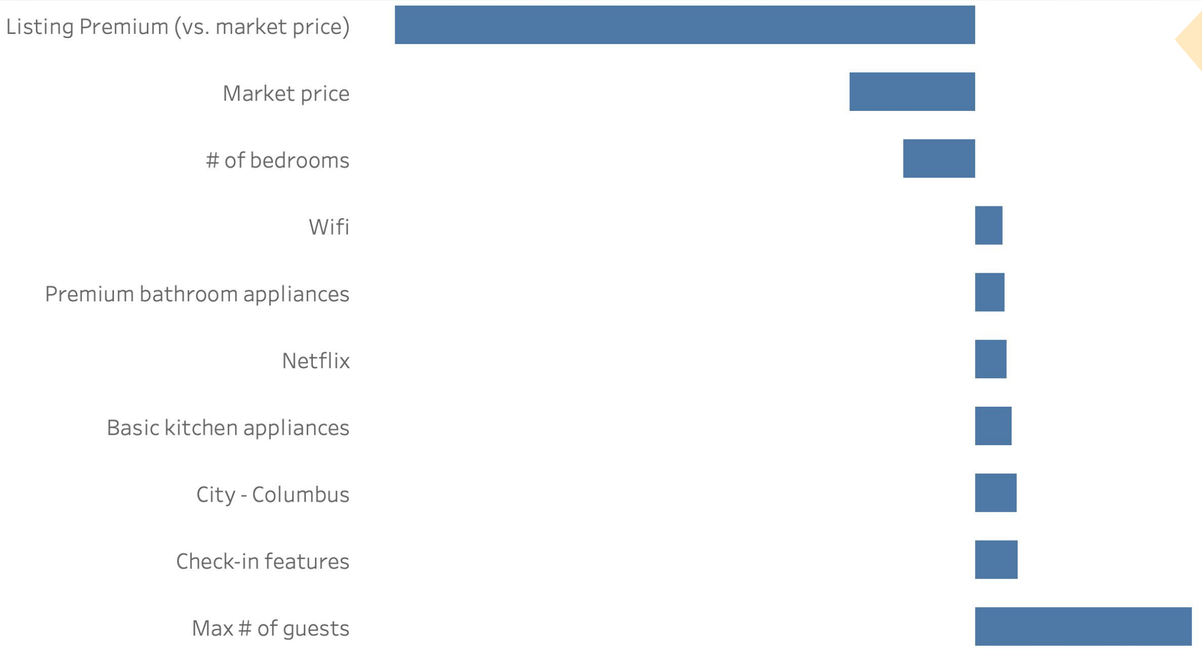
Adjusted R2: 0.0853

RMSE: 0.2199

We selected **LASSO regression** because it is:

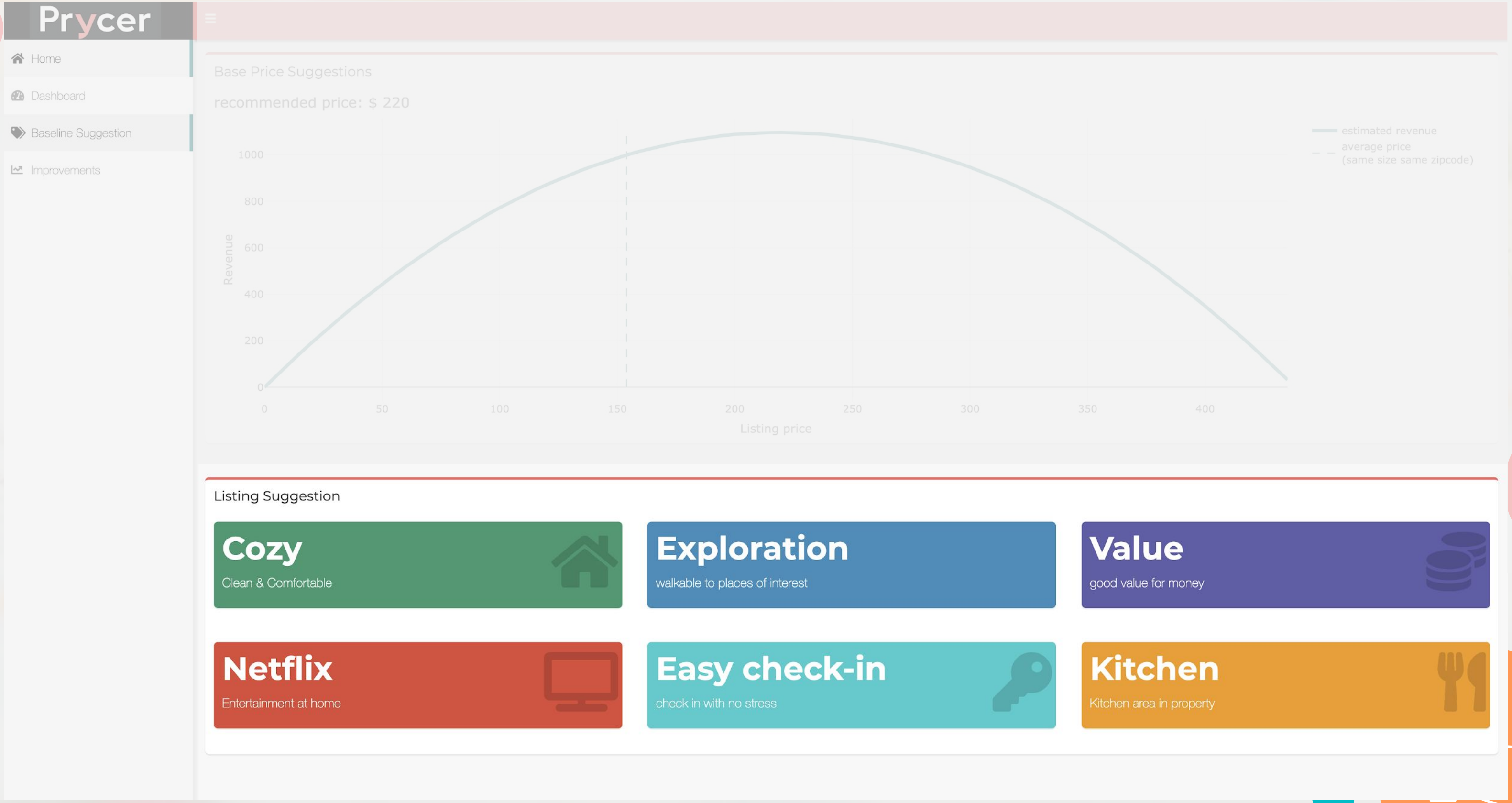
(1) interpretable
(2) regularization helps with generalization
(3) works well with optimization linear regression model.

Our hypotheses was validated in our regression model

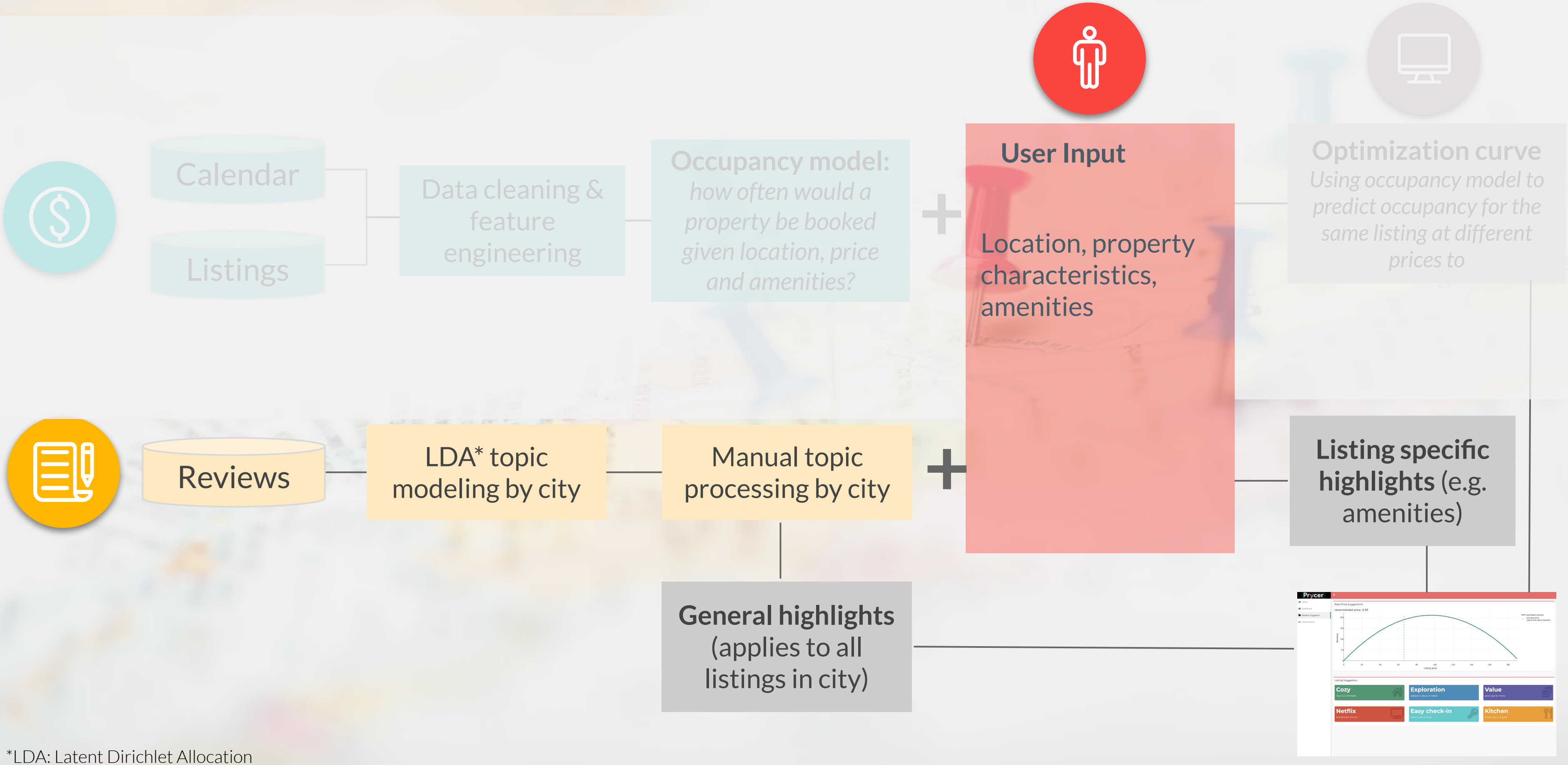


Listing premium is the most important predictor of occupancy

Let's switch gears to look at the listing suggestion section



Recap: we provide listing suggestions via topic processing of reviews



*LDA: Latent Dirichlet Allocation

We use topic modeling to understand the themes of the large volumes of review data

What is topic modeling

- Unsupervised Learning
- Group of words represents topic

Why topic modeling

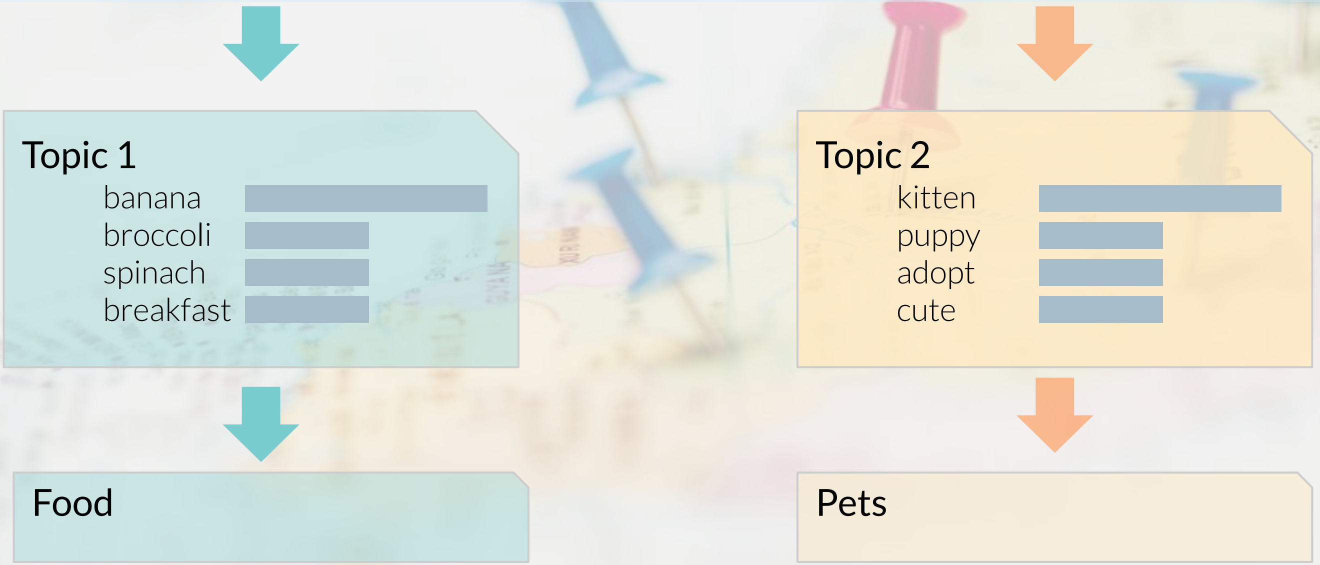
- Huge amount of text
- Distill important topics

How to prepare data

- Tokenize each sentence
- Remove punctuations, stop words, and lemmatize
- Create text corpus

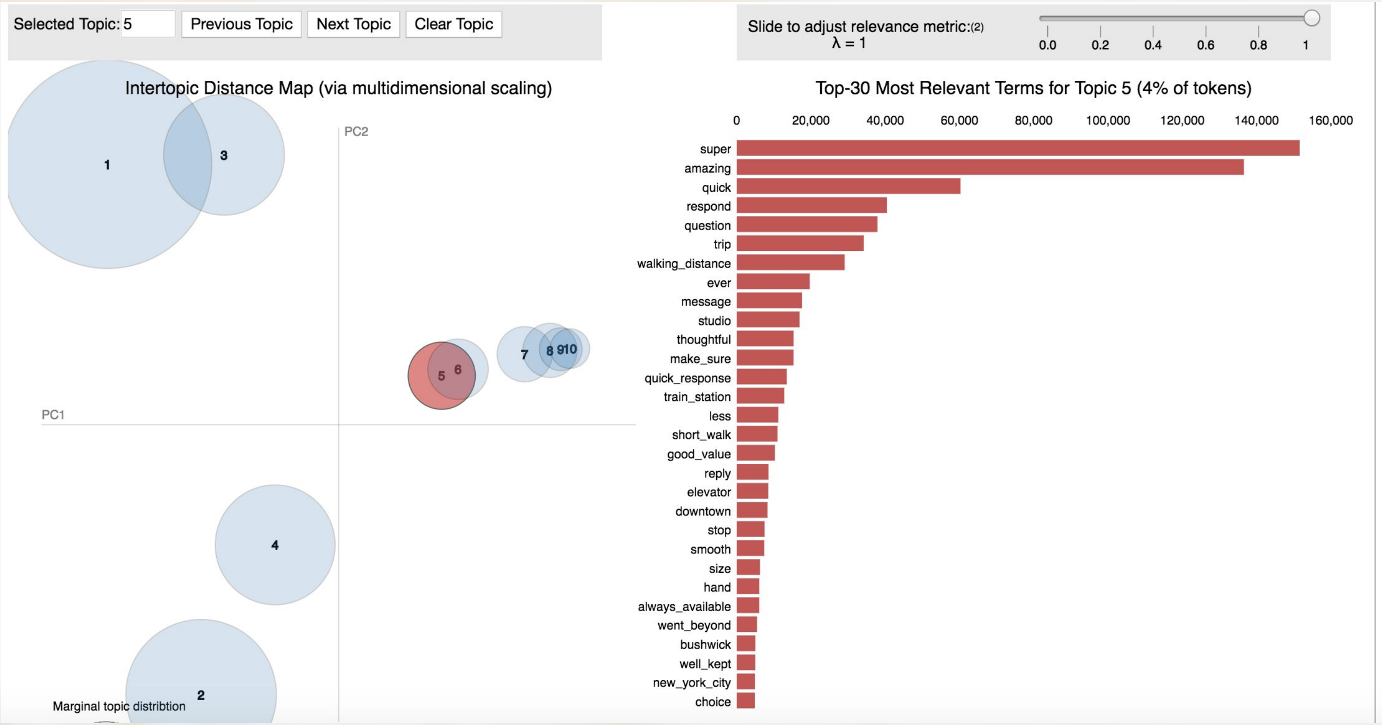
LDA* is a method that assigns words to a given number of topics

Document:
I like to eat broccoli and bananas. I ate a banana and spinach smoothie for breakfast. Puppies and kittens are cute. My sister adopted a kitten yesterday…..



*LDA: Latent Dirichlet Allocation

We reviewed the relevant words generated by LDA and manually synthesized topics



Topic: Accessible, convenient location

Topic Modeling Approach | Interpretation | Result Highlights

San Francisco

Listing Suggestion

Cozy
Clean & Comfortable

Exploration
walkable to places of interest

Value
good value for money

Easy check-in
check in with no stress

Kitchen
Kitchen area in property

Seattle

Listing Suggestion

Location
convenient and accessible

Amenities
full-featured property

View & vibe
great environment

Netflix
Entertainment at home

Easy check-in
check in with no stress

Next Steps

Further Pricing Model Analysis & Improvement
- Further optimization of the Base Price Suggestion
- Price suggestion adaptability to listing improvement suggestions

Listing Suggestion Improvement
- Listing Suggestion prioritization
- Cost assumption for each listing suggestion

Dashboard & Visualization
- Pull more up to date information into the Neighborhood Dashboard
  - Would need access to Airbnb data

Product Testing & Gathering Feedback
- Launch & have Airbnb hosts test it
- Gather user feedback via survey or focus group

Quick Recap

What is Prycer?

An Airbnb pricing & listing suggestion tool for first-time hosts

Why is it important?

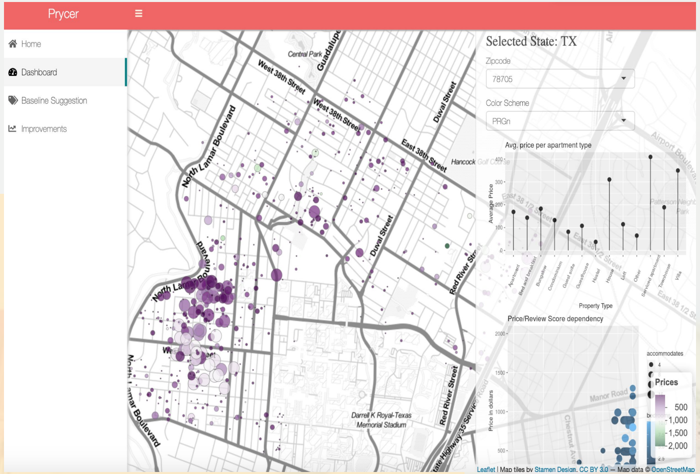
First-time hosts are not given the right tools to optimize revenue or their booking rates.

How we make an impact

We take away the hassle of listing your property by giving you an idea of how to price based on your neighborhood, suggest a base price & key words to make your place appealing!

Thank you

Descriptive analysis for neighborhood dashboard



Process:

- Preprocessed the dataset (utilizing Pandas for the sake of efficiency)
  - -deleted missing columns, values
  - -reformatted zipcodes & prices
  - -made cities uniform
- Loaded the dataset into R Shiny
  - Utilized leaflet package
  - Created functions to filter out necessary data
    - Filtration by state
    - Filtration by zipcode
    - Filtration of zipcodes by state
  - Styling
    - CSS file
  - Aggregation visualization for zipcodes, charts, reviews, prices utilizing ggplot
  - Property Display
    - Color palettes to circles for property
    - Sizing of circles to correlate with accomodation