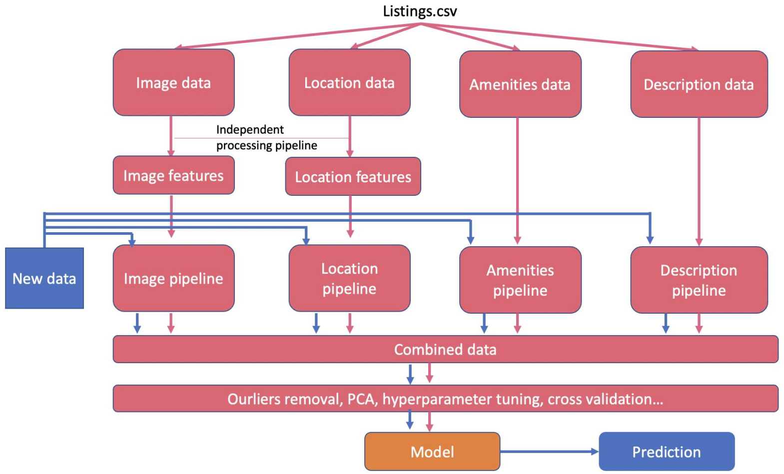
Price Recommender Web Application for New Properties on Airbnb

Team GMT8-9: Yangkang Chen, Wei Li Tan, Masato Ando, Dongyao Wang

Project Overview

How should I charge my Airbnb listing? Owners typically consider several factors including prices around the neighborhood, the configuration of the property, and the cost of owning and listing the unit. While some systems have provided solutions to a reasonable pricing recommendation, they are either less comprehensive in features input or did not mention the "competitiveness" at that price. This project aims to create a web application to allow users to figure out the optimal price point that is competitive with the market while not undervaluing the property for new property owners looking to list on Airbnb. Besides that, we aim to provide users with further helpful information about why they should charge at this range and what additional features they can include improving their pricing power.

Workflow



Data Sources

- Airbnb Listings data in Los Angeles (http://insideairbnb.com/get-the-data/)
- "Amenity Universe" dataset
- COCO128 dataset (https://github.com/ultralytics/yolov5)
- RealEstate dataset (https://www.redfin.com/news/data-center/)

Key methodology

Image pipeline

- CNN: extracting price related image features
- Aesthetic features: brightness, contrast, etc.
- Object detection: common daily object identified
- Rule of third: structural values in photography

Amenity pipeline

- Binary encoding of Airbnb standard amenities
- Bathrooms, bedrooms, property types & counts encoding.

Location pipeline

- Summary of neighborhood average prices with similar conditions
- Real estate pricing by zipcode
- Count & distance of facilities in the neighborhood
- Distance to the famous tourist attractions

Description pipeline

- Named entities recognition form host descriptions
- Whole Sentence embedding using transformers

Modeling

- Five quantile regression models with LightGBM
- Outlier removal and dimensionality reduction
- Fine-tuning using GridSearchCV

Interpretation

- Feature importance interpretation with SHAP package beewarm plot
- prediction result explanation using waterfall plot
- Learning curve, sensitivity analysis and failure analysis with matplotlib & searbon presentation

Application

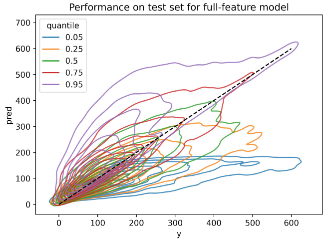
- Web application: Streamlit

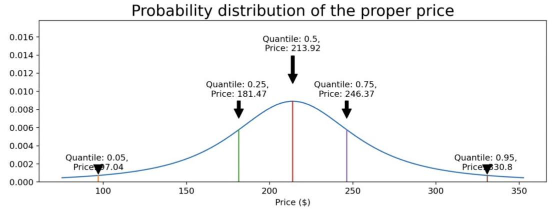
Results

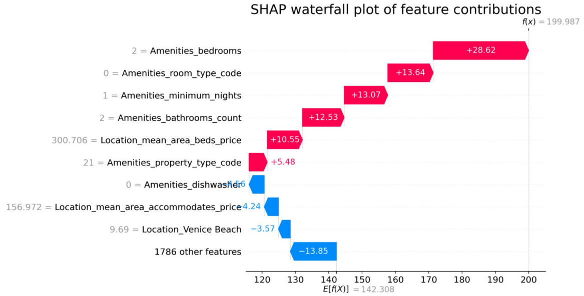
Overall good model fitting with **R2 around 0.64** for the 0.5 quantile model. Web application generate prediction report within 10 seconds.

Suggested price: $ 213

	Quantiles	Price ($)
0	0.05	97.0399
1	0.25	181.4735
2	0.5	213.9194
3	0.75	246.3652
4	0.95	330.7988

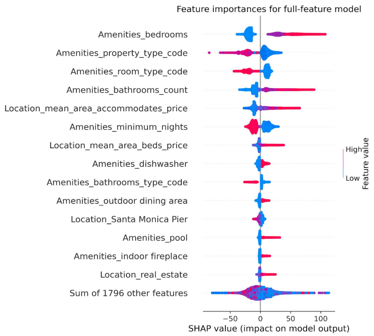






Feature importances

As "hardware", House property, amenities and location determined the most percentages of price variance, with description and image as decoration effects. Other unexplained price variance could be due to complexity, dynamism and subjectivity in Airbnb pricing.



On the web application side, our product successfully provide insightful information on instructing the hosts to price their listing properly, with further model interpretation to facilitate the understanding of pricing mechanism. This is how we "go beyond a single price".

Project Resources:
Web application: http://18.205.39.151:8502/my_app
GitHub: https://github.com/foye501/Capstone_GMT89
Report: https://docs.google.com/document/d/161fEv0t4Ops9SG5NPMAXZgnTigGPNvgPrR8gCyeM7x0/
Video explanation: https://www.youtube.com/playlist?list=PL-Ih8lEqwhvFGjcMphHh4x3e4OoW4smun
Medium blog: https://medium.com/@chenyk_80392/beyond-a-single-price-pricing-range-improvement-suggestions-for-new-airbnb-hosts-a33ffa718fbb

UNIVERSITY OF MICHIGAN