West Coast Market Insights

Los Angeles · San Diego · San Francisco · Seattle · Portland

How Pricing, Amenities & Property Types Drive Revenue and Improve Review Scores

The West Coast Laboratory

Market Context

SF & Seattle

High average daily rates, shorter stays, business travelers

LA & San Diego

Seasonality, luxury amenities, high competition

Portland

Consistent, lower entry barrier, unique "culture-driven" travel

* 55,000+ combined listings
* Extreme variation in density, regulation, and guest type

Key Question: For AirBnb

How can Airbnb use pricing, reviews, amenities, and property type data to increase revenue, improve guest satisfaction, and improve listing performance across five West Coast cities?

Key Question: For Hosts

What drives host revenue, guest satisfaction, and listing performance?

West Coast Markets: Los Angeles · San Diego · San Francisco · Seattle · Portland

Hypotheses

01

Revenue is influenced by pricing, responsiveness, & city-level factors.

02

Listings with richer amenity bundles achieve higher review scores.

03

Property type significantly influences annual revenue.

What To Expect

Actionable insights for hosts and strategic signals for Airbnb.

Data & Methods

Cleaning and Preparation

Data Overview

1. Coverage

55,000+ Airbnb listings (2025 snapshots) of 5 major West Coast markets: LA, San Diego, SF, Seattle, and Portland.

2. Source

Data Source: http://insideairbnb.com.

3. Key Variables

Price • Reviews • Ratings • Amenities • Property Type • Host Behavior • Occupancy • Estimated Revenue

4. Unit Analysis

Each row represents 1 listing.

Cleaning and Preparation

Import & Load

Brought in all libraries and Airbnb listings for 5 West Coast cities.

Exploratory Data Analysis

Used .info(), .describe(), and plots to inspect structure, distributions, and outliers.

Cleaning

Dropped irrelevant columns, fixed missing/inconsistent values, converted price strings, standardized binary fields, and filtered extreme outliers.

Feature Engineering

Encoded property types & amenities, grouped amenities into 6 categories, created *amenity_category_count* and *response_speed* buckets, consolidated rare types into "Other."

Combined Master Dataset

Merged all city-level datasets into a single master dataset with a city identifier for cross-city comparisons.

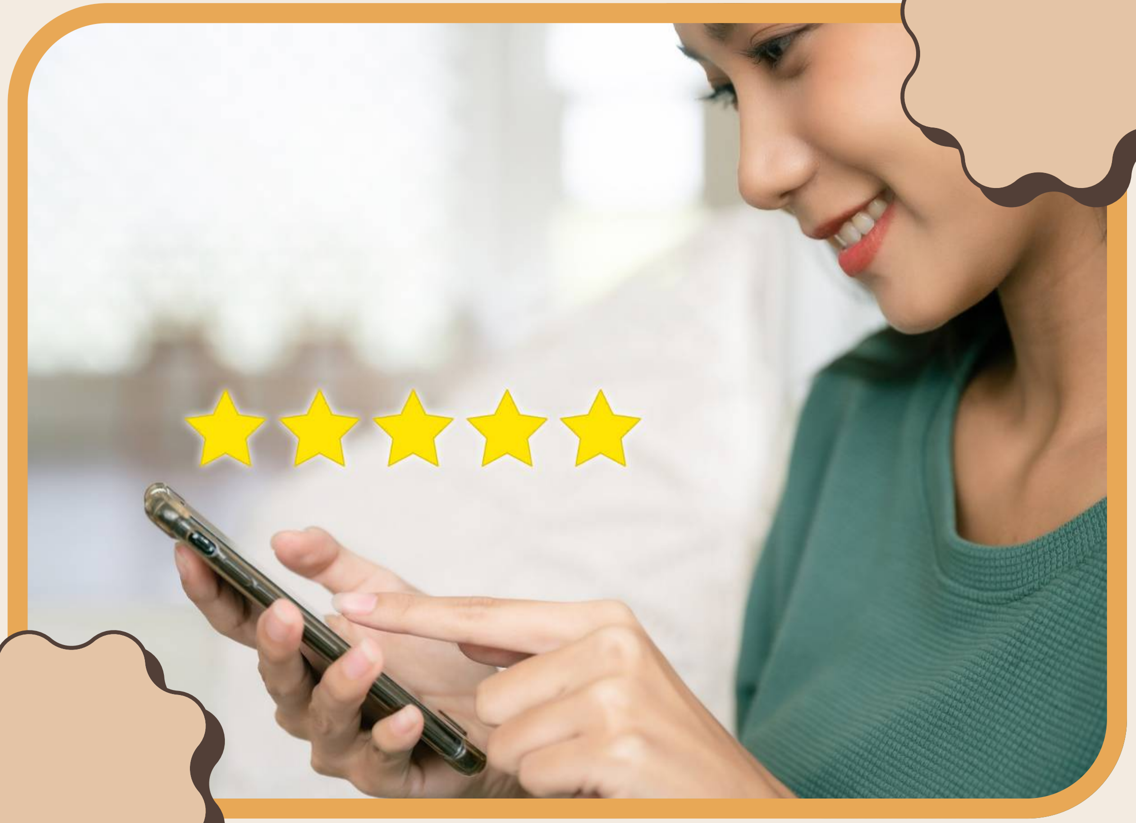
Methods

Descriptive Analysis & EDA
Distributions, city-level summary statistics, boxplots

Inferential Statistics
Correlations (Pearson), OLS regression, ANOVA

Machine Learning
Decision Tree to classify High Rated listings (rating ≥ threshold). Random Forest to predict *estimated_revenue_l365d* and compute feature importances

Hypothesis 1



Cities, Neighborhoods, Responsiveness, and Reviews

Hypothesis 1

How can hosts increase revenues?

Key Questions

- How do prices and revenue differ by city, neighborhood density and guest capacity?
- Do review scores or rating affect price?
- Do responsiveness and superhost status significantly impact revenue?

Hypothesis 1

Definition

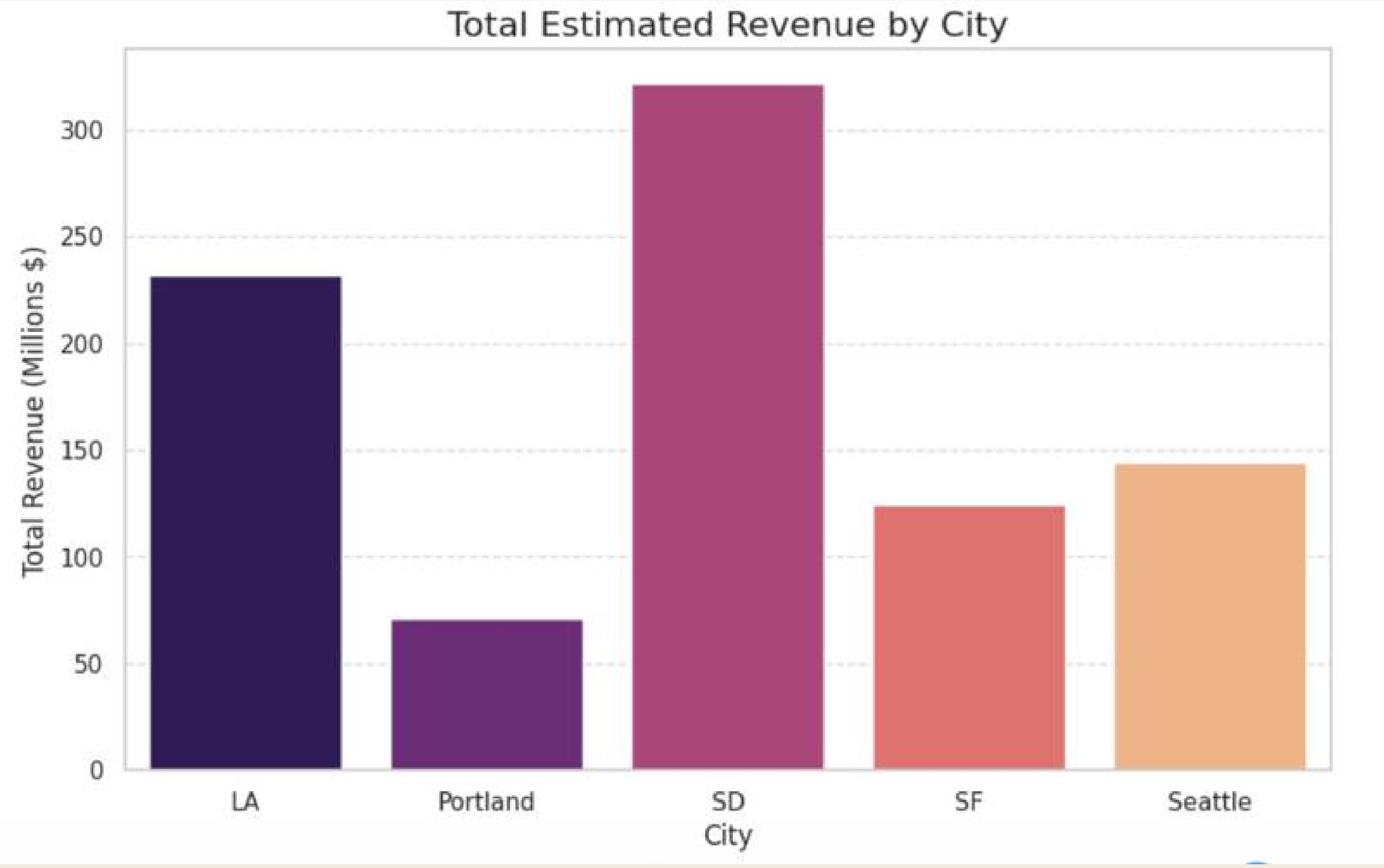
- **H_0:** No significant relationship between pricing, availability, reviews, responsiveness, and revenue.
- **H_1:** There is a significant relationship between these factors and revenue.

Sub-Components

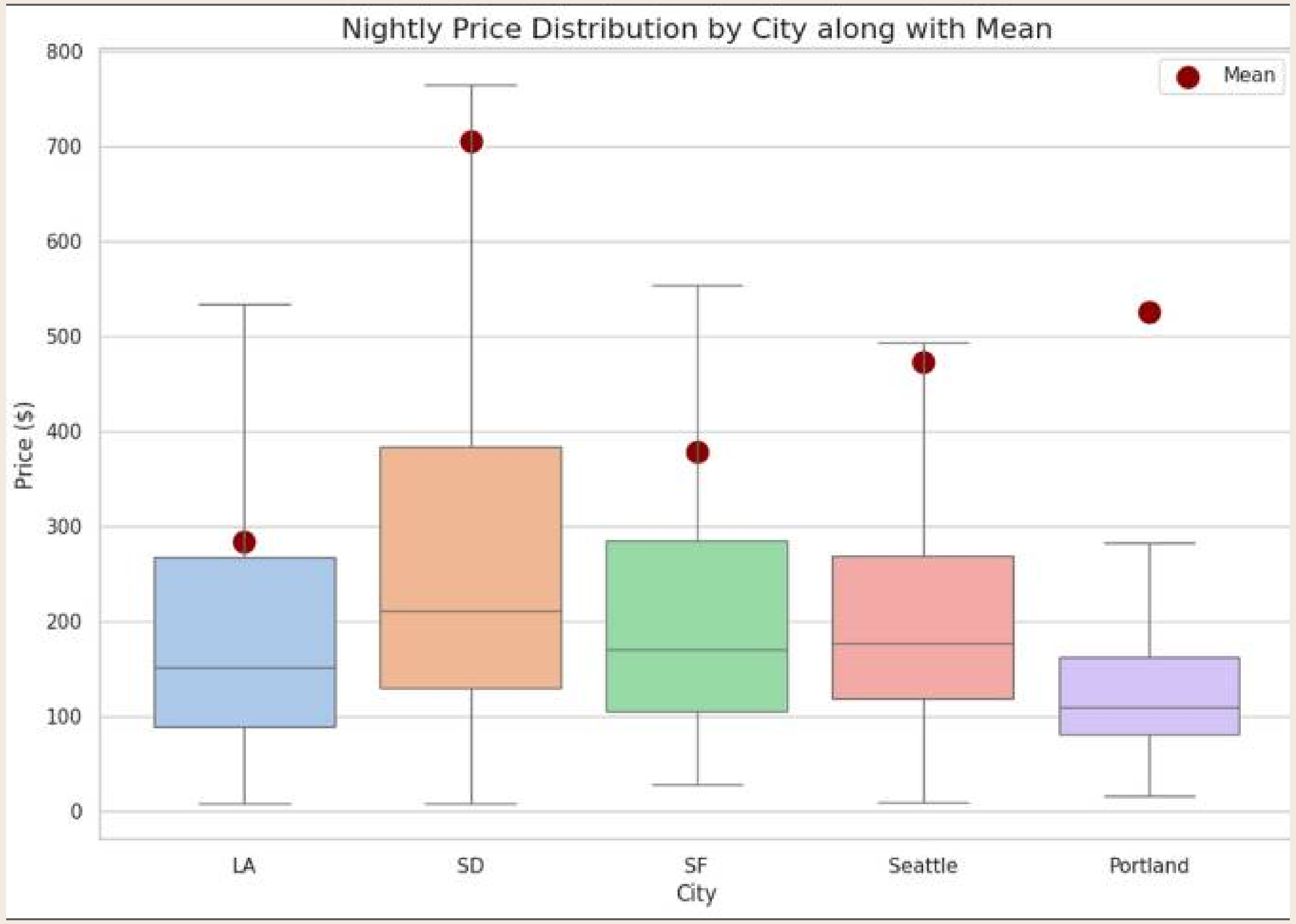
- **H_1a:** Price & revenue patterns by city and density.
- **H_1b:** Relationship between review scores/rating and price.
- **H_1c:** Impact of host responsiveness & Superhost status on revenue.

H1 Data Analysis: Revenue by cities

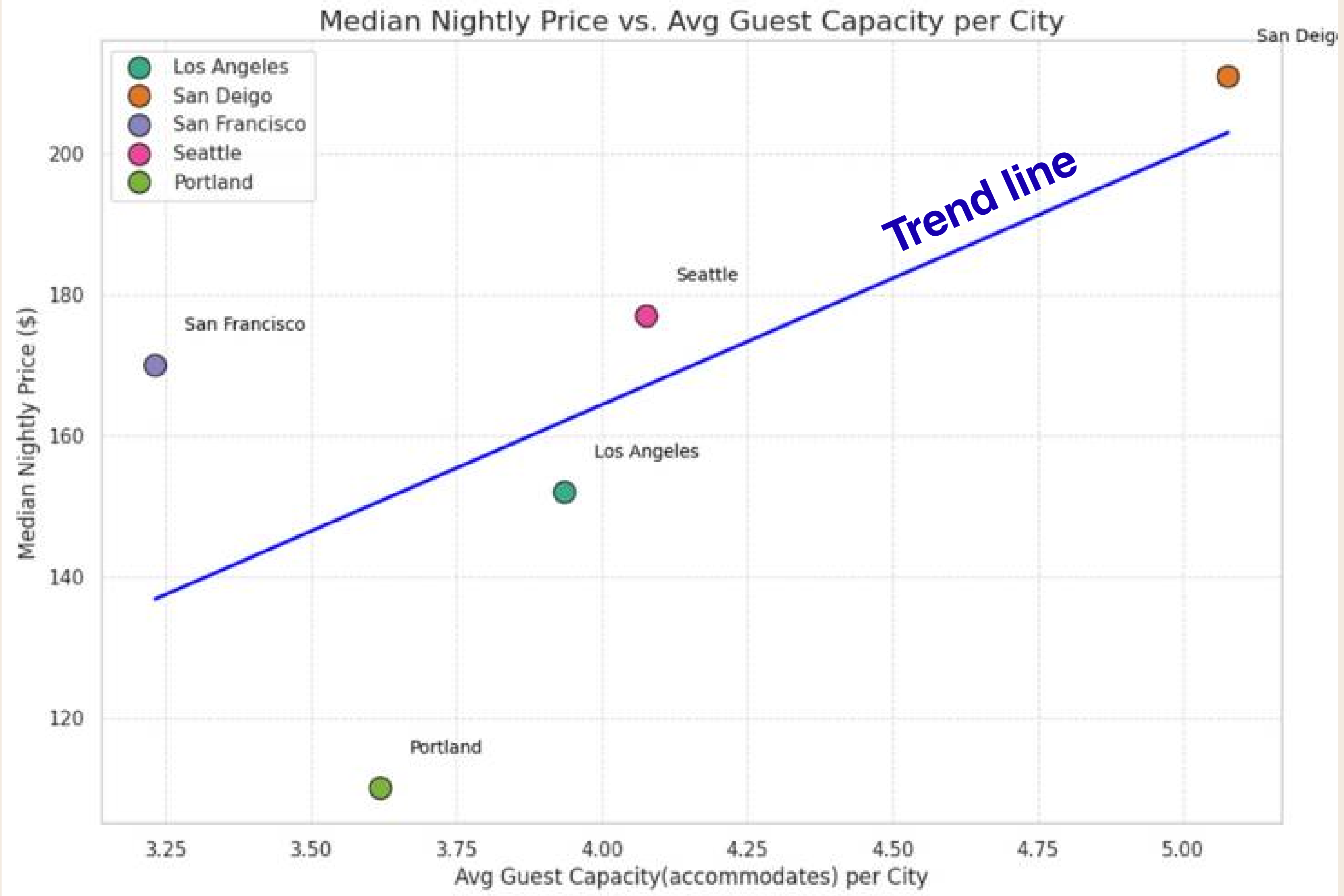
- Analyzed cities based on total estimated revenue and mean revenue per listing to **distinguish large markets from the most profitable ones**
- **LA & SD:** Highest total revenue (~$230M and $350M+), driven by large volumes of listings and popularity amongst tourists → **beneficial the company**, not necessarily individual hosts
- **SF, Seattle & SD:** Higher revenue per listing ($21,000+/yr) → **more lucrative opportunities for hosts**
- **Big markets ≠ most profitable**. Per-listing revenue helps hosts choose better locations



H1 Data Analysis: Nightly Prices by City



- Analyzed **price distributions** for different cities
- Average price sit far above medians indicating a **strong skew** in most cities due to luxury outliers
- **Long upper whiskers** indicate high-priced outliers pulling the mean upward
- **Median better reflects typical nightly prices**



- How Guest Capacity Influences Median Nightly Price
- Guest capacity shows a **moderate–strong positive correlation with price (r = 0.66)**
- However, capacity isn't the sole pricing driver
- Smaller listings can still perform well by offering strong amenities, leveraging seasonality, earning Superhost status, and adding value-enhancing features

H1 Data Analysis: Demand, Density and Price

Neighborhood density vs Median price

- Weak positive correlation between neighborhood density vs median nightly price : 0.27.

H1 Data Analysis: Reviews vs Price

Reviews & price

- **Correlation**:
  - Price is not strongly correlated to review scores: 0.07 (very weak correlation).
- **OLS regression:**
  - Model's $R^2 = 0.054$, so only 5% of the price variability is explained by reviews.
  - Regression analysis also suggests a 1-point increase in review score can demand a ~$54 increase in price.
  - Review scores have a small positive effect on price.
  - City (or location) has a bigger impact on price than reviews.
    - For example, a listing in San Diego tends to cost much more than a similar listing in Portland, even if their ratings or number of reviews are the same.
  - Further analysis of amenities and property characteristics will be explored in the next sections.

review_scores_rating

OLS Regression Results

Dep. Variable:	price	R-squared:	0.054
Model:	OLS	Adj. R-squared:	0.054
Method:	Least Squares	F-statistic:	332.5
Date:	Fri, 28 Nov 2025	Prob (F-statistic):	0.00
Time:	21:21:13	Log-Likelihood:	-2.3923e+05
No. Observations:	34835	AIC:	4.785e+05
Df Residuals:	34828	BIC:	4.785e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-14.3101	16.459	-0.869	0.385	-46.571	17.950
C(City)[T.Portland]	-80.7943	4.666	-17.315	0.000	-89.940	-71.648
C(City)[T.SD]	74.0730	3.199	23.155	0.000	67.803	80.343
C(City)[T.SF]	-7.8045	4.045	-1.929	0.054	-15.733	0.124
C(City)[T.Seattle]	-15.6567	3.848	-4.069	0.000	-23.199	-8.114
number_of_reviews	-0.1866	0.010	-18.571	0.000	-0.206	-0.167
review_scores_rating	53.6849	3.416	15.716	0.000	46.989	60.380

Are there city-specific patterns in host responsiveness and superhost status that impact revenue?

56.7%

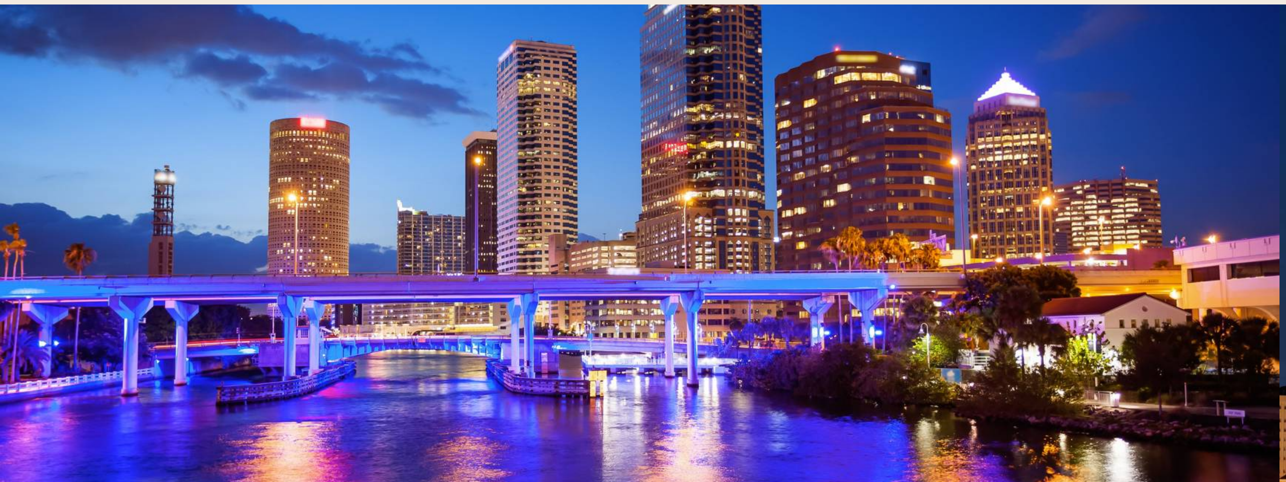
of hosts in Portland are Superhosts which is the most of the 5 cities!

$27,727

is San Diego's average estimated revenue which is the highest among the cities.

34,931

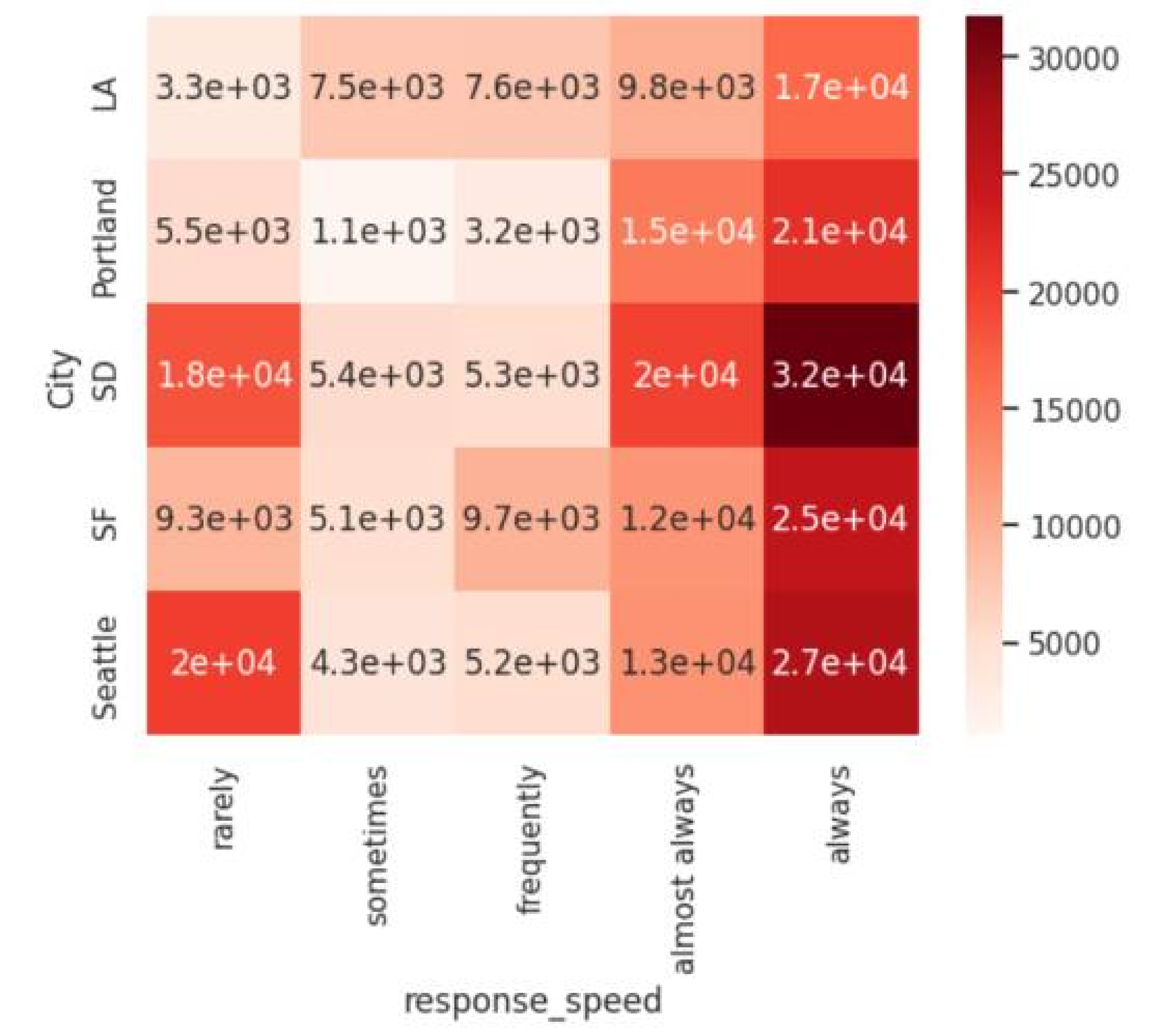
of hosts 'always' reply (99-100% response rate) across all cities.



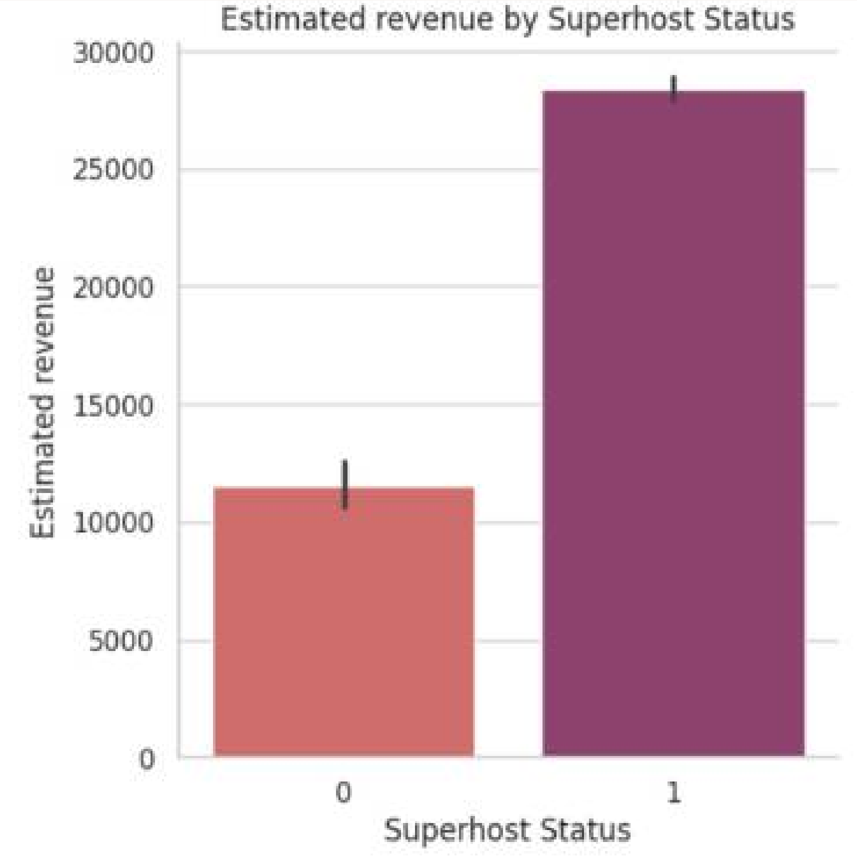
Heatmap of revenue broken down by response speed and city

Higher response speed is strongly associated with higher revenue across all cities.

- Those who always respond have higher values (darker shade of red) than those who rarely or sometimes respond.
- San Diego and Seattle have consistently high values across all responses, but those who always respond in this cities still have higher revenue.



Superhost Status affect on Revenue



Across all cities, we see superhosts have higher estimated revenue.

Broken down by city, we still see each individual's cities Superhosts have higher estimated revenue. San Diego and Seattle having the highest.

Hypothesis 1: Findings

Prices and Markets

- San Diego has the highest median nightly price; LA the lowest despite being the largest market.
- Portland & Seattle show higher average revenue per listing, hinting at better host ROI.

Reviews

- Reviews have a weak relationship with price (very weak R^2).
- Higher review scores enable a slight price premium, but city and property traits dominate.

Responsiveness and Superhosts

- Faster response-speed buckets consistently show higher revenues across cities.
- Superhosts earn more on average than non-Superhosts in every city.

Hypothesis Outcome

- Based on our analysis, we **rejected the null hypothesis**.

Managerial Recommendations

For Hosts

- Consider markets in Seattle and San Diego for stronger per-listing revenue potential.
- Maintain very high response rates (99–100%) to maximize revenue.
- Obtain Superhost status
- Focus on rating quality (especially cleanliness and communication) to justify small price premiums.

For Airbnb

- Reward fast response rates and Superhost status in search rankings.
- Provide city-specific pricing guidance to new hosts.

Hypothesis 2

Amenities

Hypothesis 2

Does the Number of Amenity Types Matter?

- Amenities are a major differentiator on the Airbnb platform.
- Hosts see long, unstructured amenity lists with little guidance on which combinations matter most.
- Research goal: Does offering more (and the right mix of) amenity types actually improve ratings?

Hypothesis 2

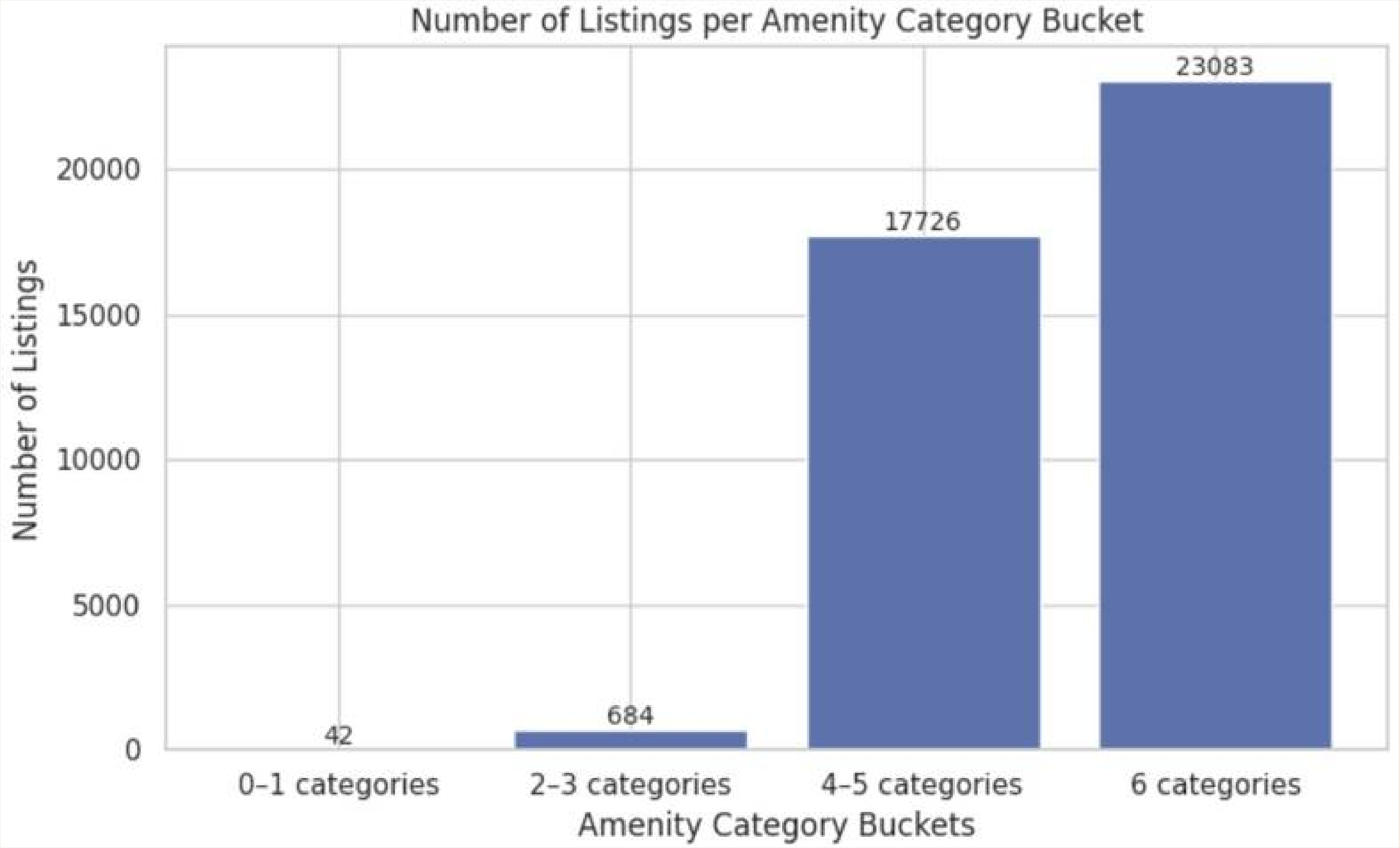
Definition

- **H_0:** No significant relationship between the number of amenity types and review scores.
- **H_1:** Listings with richer amenity bundles (more categories) have significantly higher review scores.

Key Questions

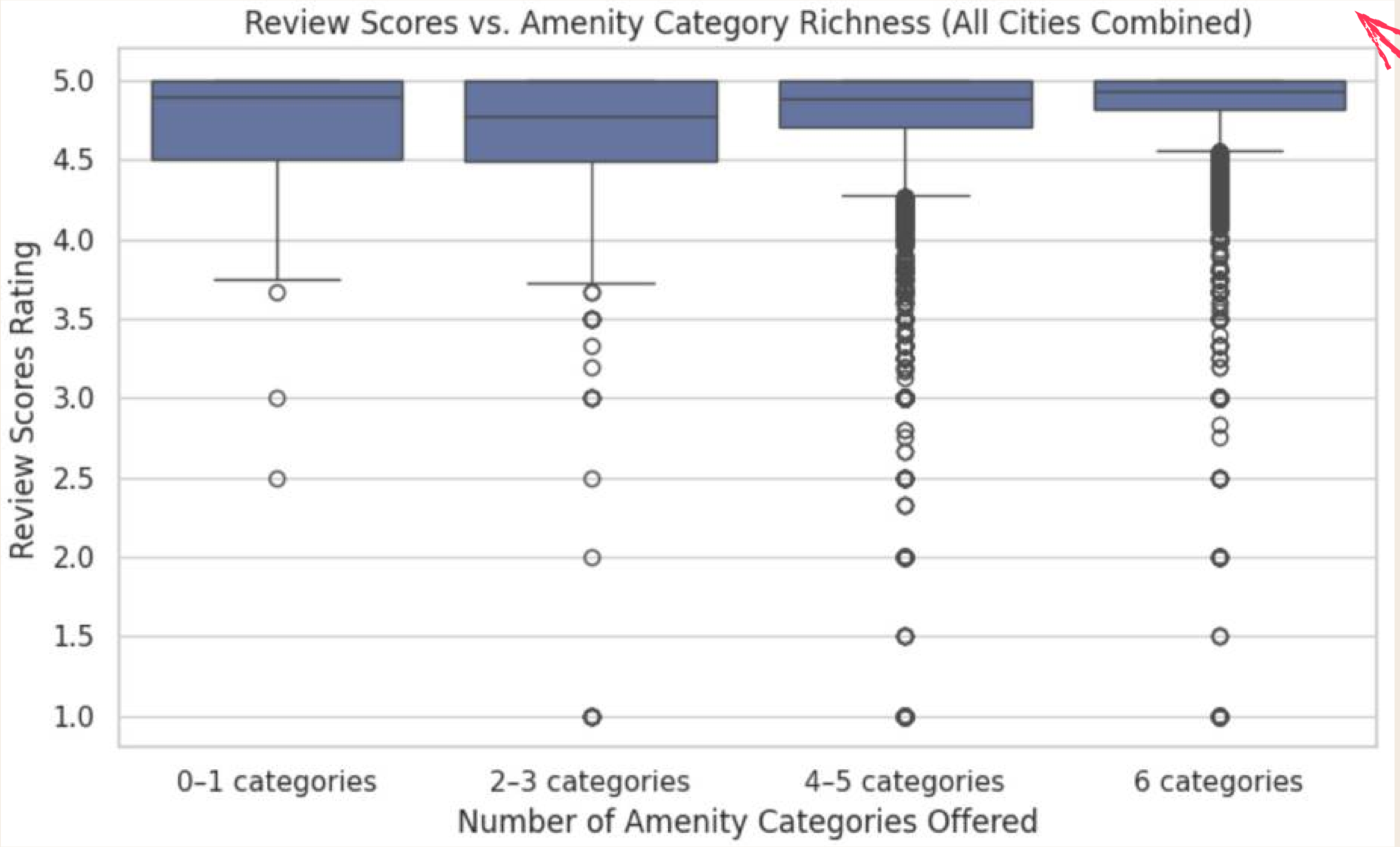
- Does amenity richness improve average ratings?
- Which amenity categories contribute most to higher scores?
- Can we predict "High Rated" listings using amenity features?

Data Preparation



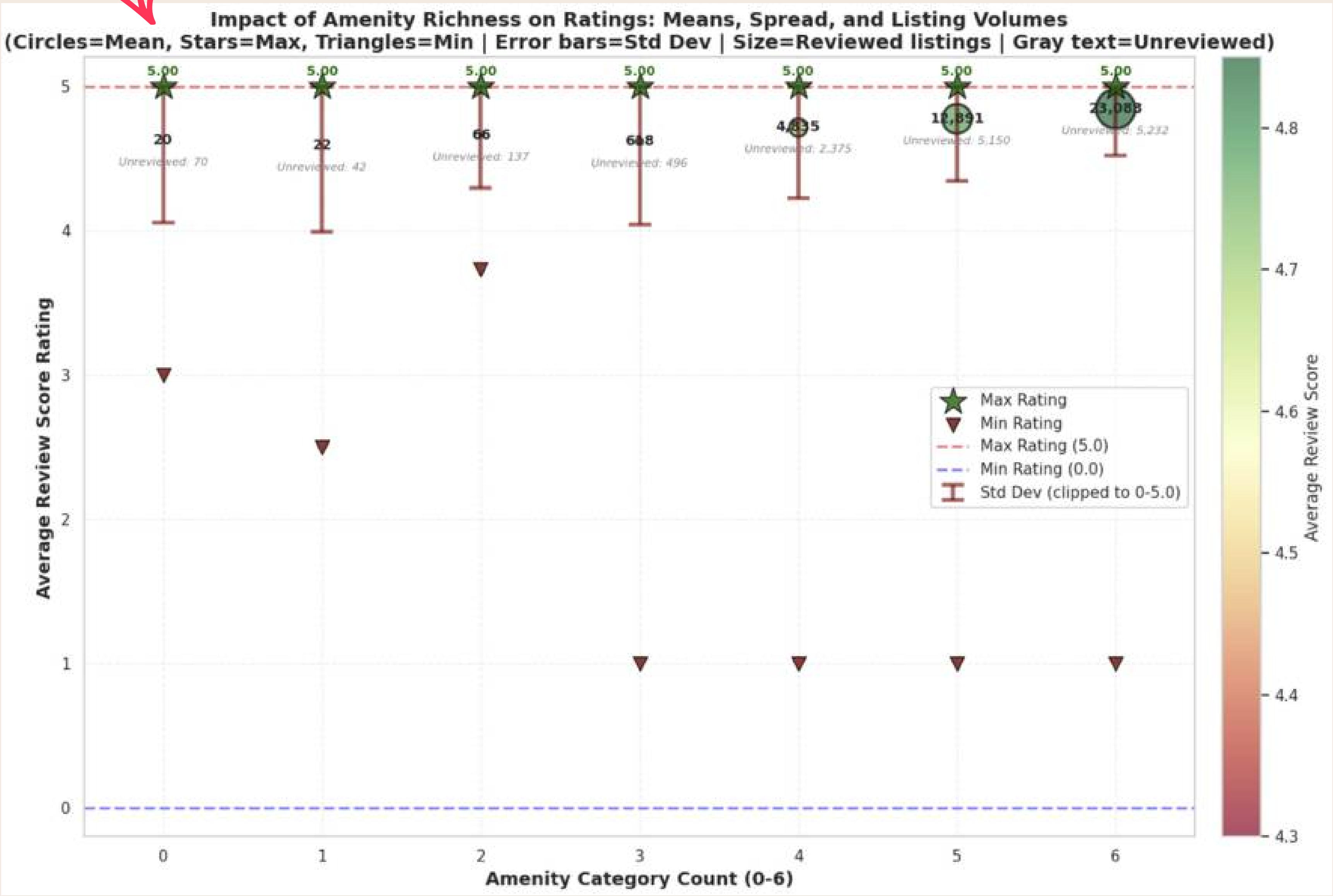
- Parsed amenities into 6 functional categories: Comfort, Luxury, Business, Safety, Kitchen, and Outdoor.
- Created *amenity_category_count* (0–6) per listing.
- Binned listings into buckets: 0–1, 2–3, 4–5, 6 categories.

Data Visualization Techniques: Boxplot & Bubble Plot



Boxplot for *review_scores_rating* vs *amenity_category_count*.

Bubble Plot for mean, min, max ratings, and rating variance vs amenity count.



Statistical Analysis Techniques: Correlation & T-Tests

	amenity_category	count_with	count_without	avg_rating_with	avg_rating_without	difference	p_value
0	amenity_comfort	41417	118	4.803025	4.564915	0.238110	1.869909e-04
1	amenity_luxury	31931	9604	4.811657	4.771399	0.040258	9.812899e-19
2	amenity_business	41310	225	4.803179	4.649867	0.153312	5.058394e-05
3	amenity_safety	40817	718	4.805529	4.621560	0.183969	1.200911e-13
4	amenity_kitchen	40629	906	4.805393	4.665817	0.139576	7.069618e-16
5	amenity_outdoor	28197	13338	4.839539	4.723726	0.115813	2.245999e-132

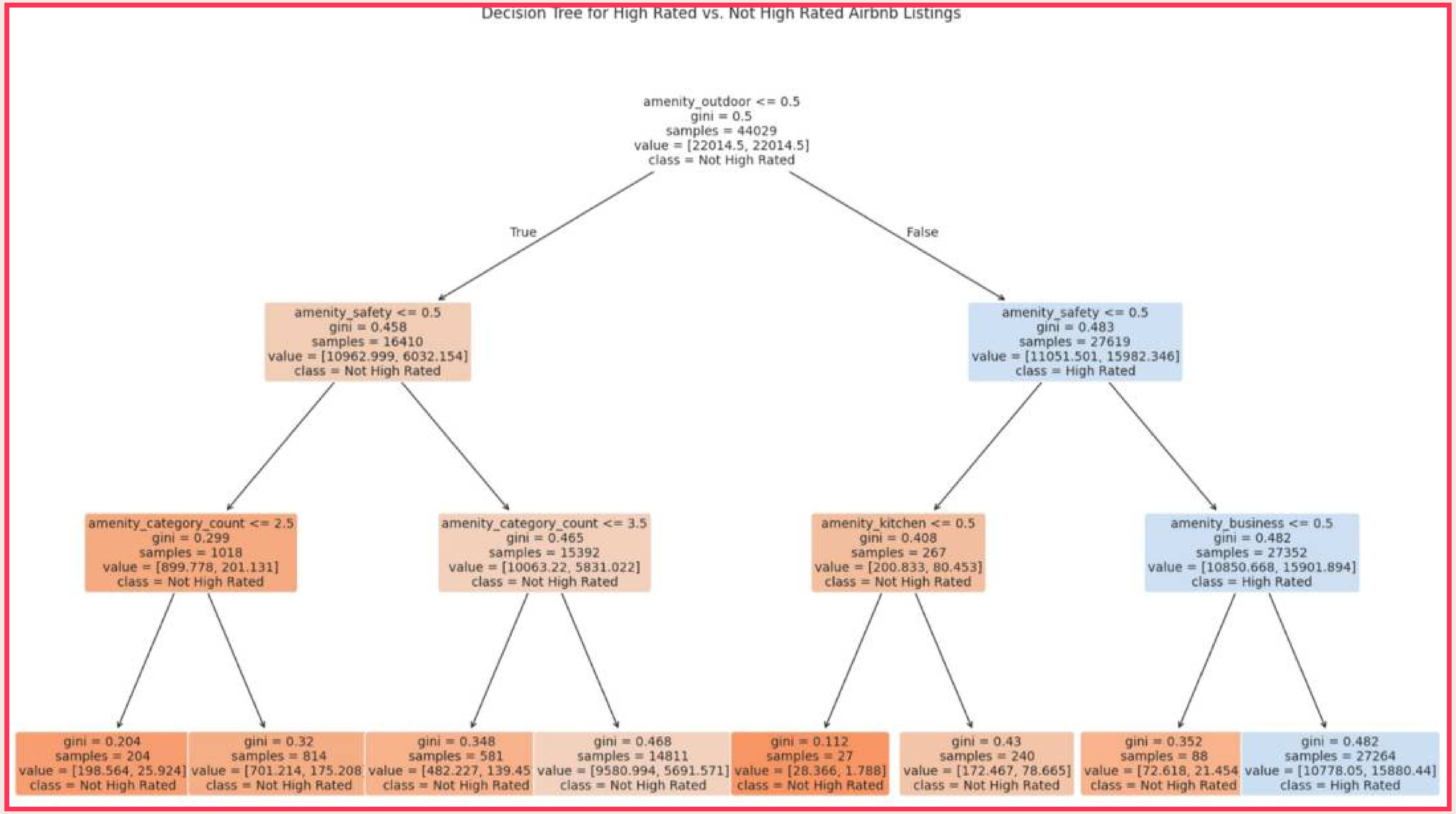
T-tests for each amenity category (with vs without) to compare average ratings.

- Computed summary stats:
  - Count of listings per bucket.
  - Mean and std of review_scores_rating per bucket.

0.132

Correlation between *amenity_category_count* and *review_scores_rating*.

Predictive Modeling Technique: Classification Tree



Process

- Defined target: High Rated (rating ≥ threshold of 4.79).
- Split data into train (80%)/test (20%); fitted a Decision Tree with constrained depth (3 levels) and class-weighting.
- Interpreted the tree to see which amenity features drive “High Rated” predictions.

Findings

- *amenity_outdoor* is the strongest predictor.
- For listings meeting this criterion, adding Safety and Business amenities increases the likelihood of being “High Rated” (~59% at the final node).

Hypothesis 2: Findings

More Amenity Categories → Higher Average Scores

- Listings with 6 amenity categories have the highest average review scores, while listings with 0–1 categories have the lowest.
- Diverse amenities are associated with better guest experiences.

Hypothesis Outcome

- Based on our analysis, we **rejected the null hypothesis**.

Managerial Recommendations

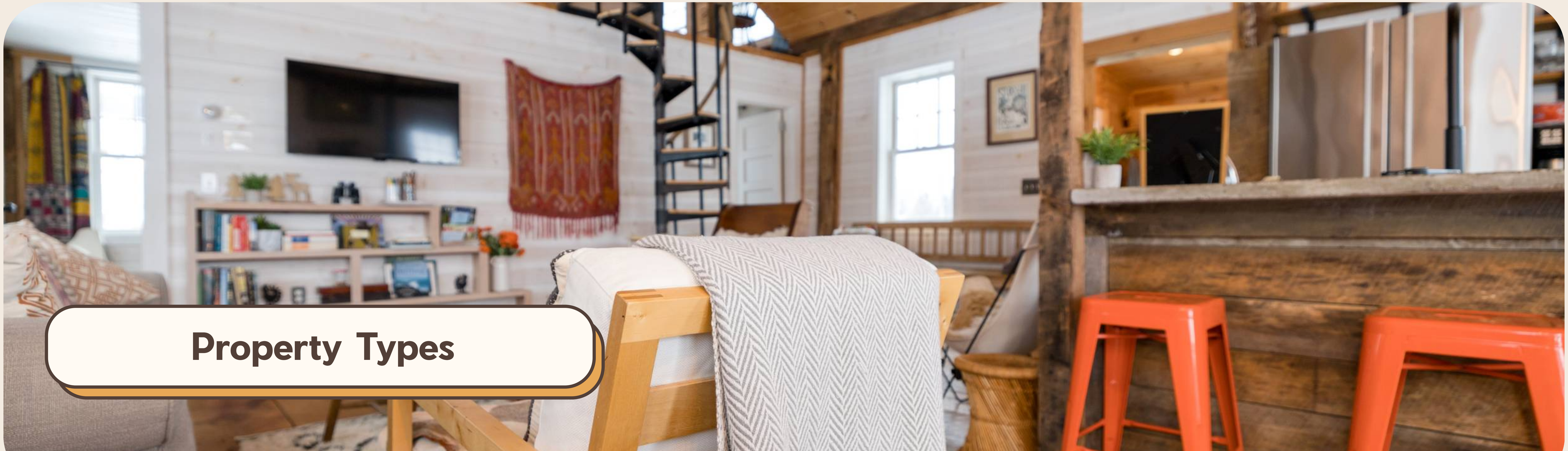
For Hosts

- Aim for at least 4 amenity categories to improve ratings and reduce risk of low scores.
- Prioritize Outdoor, Safety, and Business amenities (e.g., outdoor seating, secure locks, workspaces, strong Wi-Fi).

For Airbnb

- Suggest "amenity bundles" during listing creation and updates (e.g., Work-Friendly, Outdoor & Leisure).
- Highlight amenity-rich listings with filters, tags, or badges.

Hypothesis 3



Property Types

Hypothesis 3 - Background

Is there a relationship between Revenue and property type?

- Multiple property types exist across markets: entire homes, condos, townhouses, guest suites, private rooms, etc.
- Hosts lack clarity on which property types deliver the best revenue in each city.

Hypothesis 3

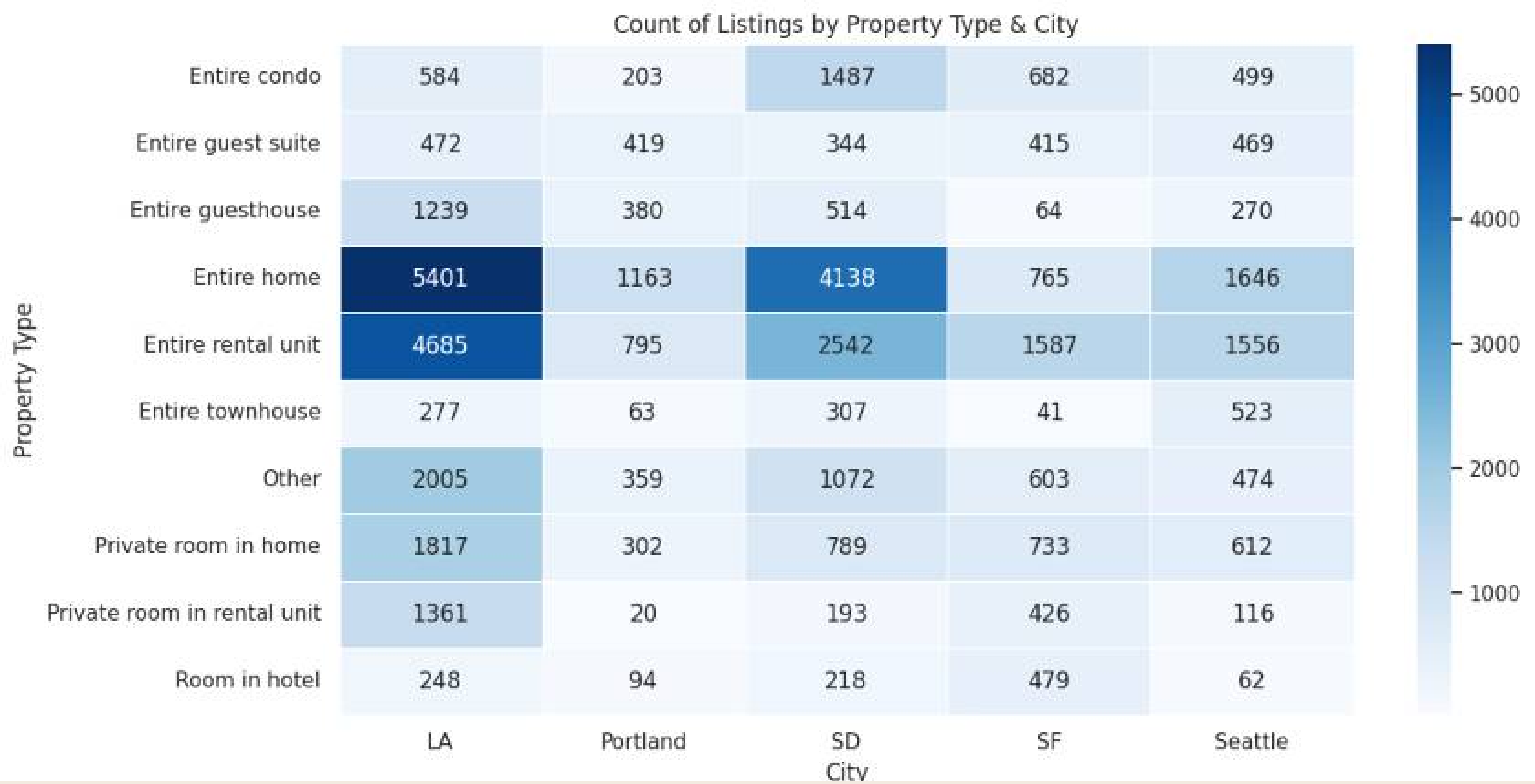
Definition

- **H_0:** No meaningful difference in revenue across property types and cities.
- **H_1:** Certain property types deliver significantly higher revenue, and this varies by city.

Key Questions

- Which property types dominate inventory in each city?
- Which property types earn the highest median revenue per listing?
- Are these differences statistically significant?

Data Preparation & Analysis - Visualizing Inventory Saturation Across Cities

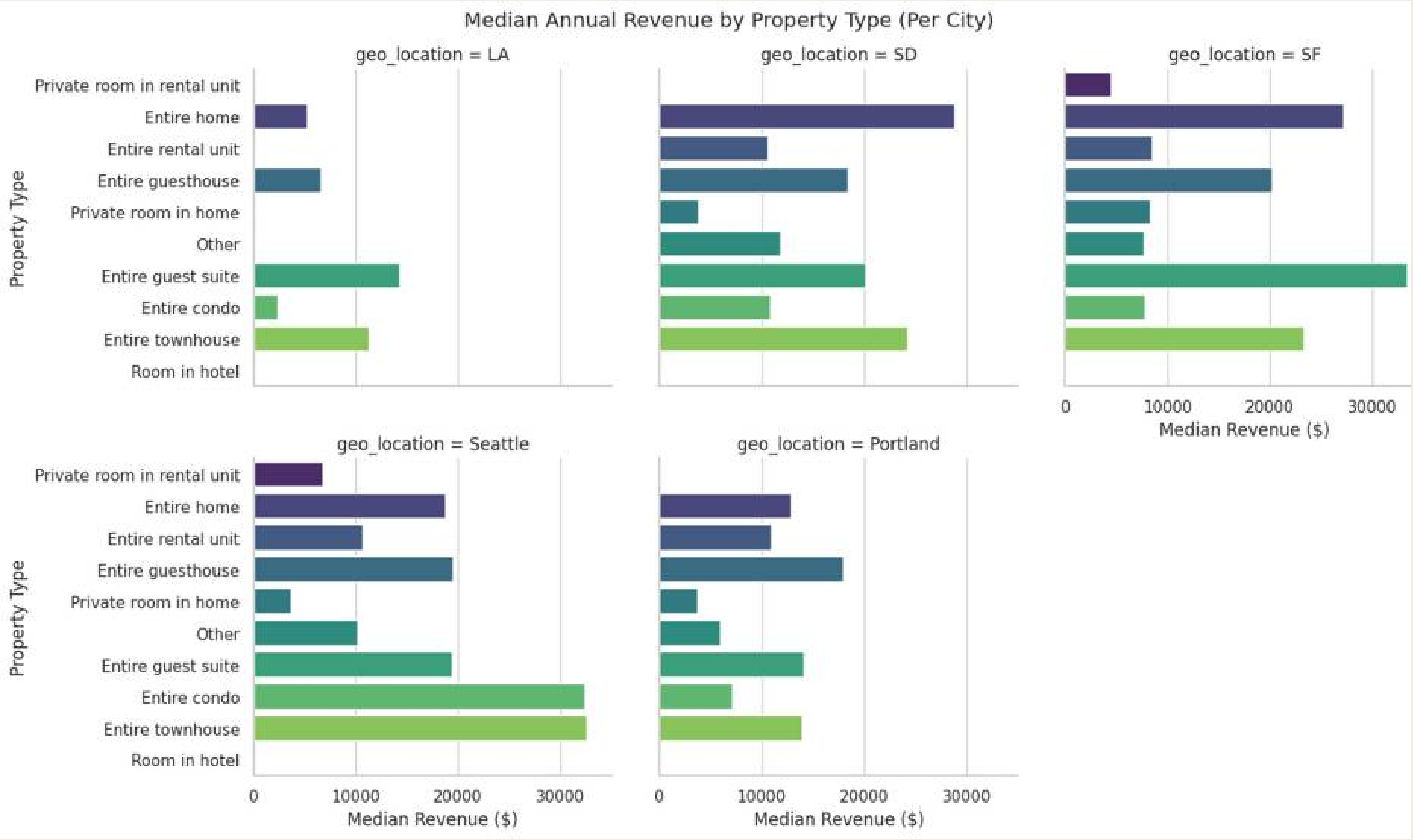


Los Angeles Dominance: Highest saturation implies intense host competition.

San Diego's Vacation Density: A strong second driven by robust vacation rental demand.

"Entire Home and Rental Unit" remains the dominant asset class across all five cities.

Data Preparation & Analysis - Comparative Volatility Analysis by City



Los Angeles: Saturated; harder to achieve high yields.

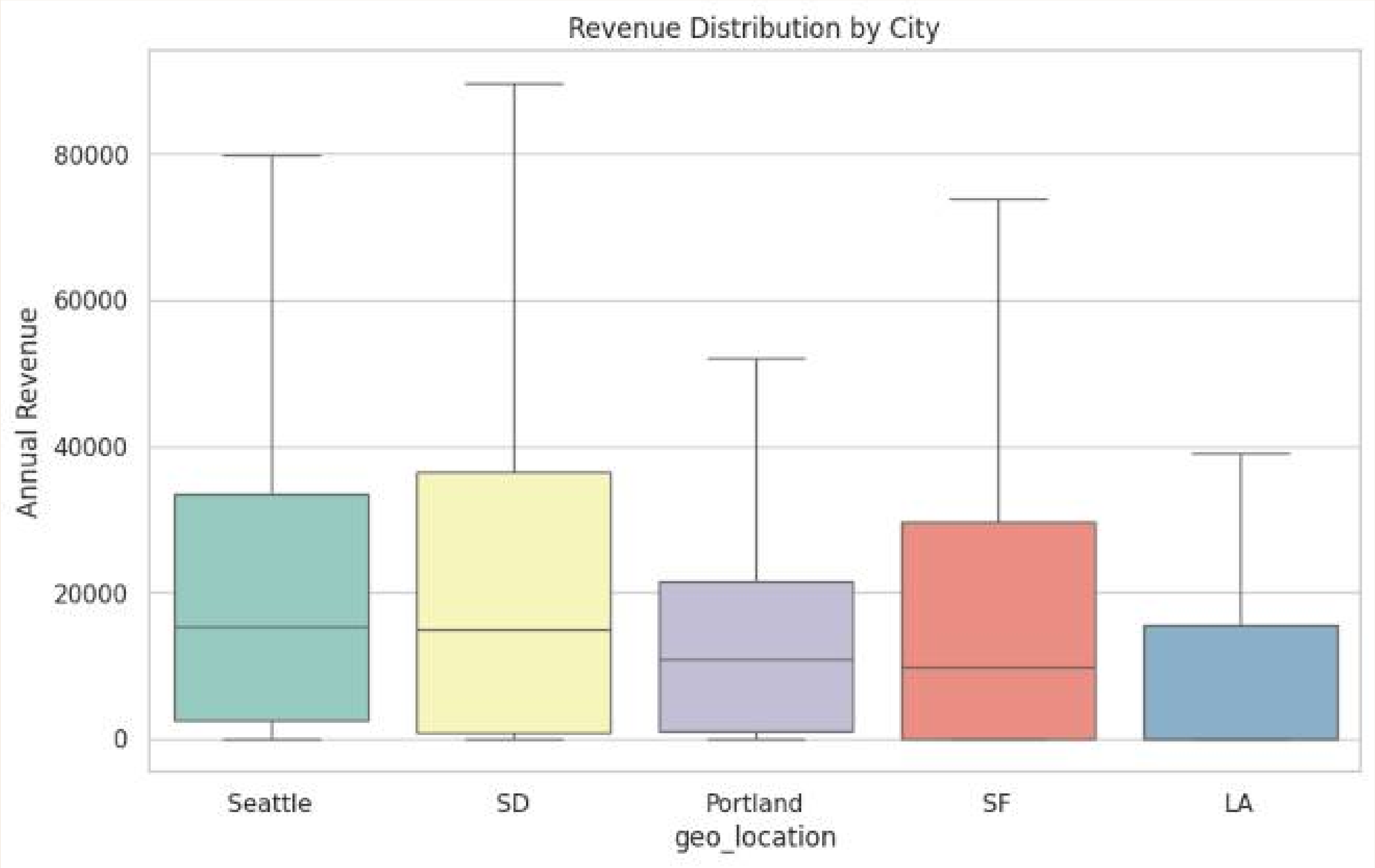
San Diego: Best for aggressive investors (High variance).

San Francisco: Best for selective strategies (Guest Suites & Homes deliver top-tier yields).

Seattle: Best for balanced returns (High median).

Portland: Best for safety (High stability).

Which City pays the best?



Seattle (Top Pick): Highest median income; reliable and consistent.

San Diego (High Risk/Reward): Highest upside, but volatile results.

Portland (The Safe Bet): Predictable, steady income with a lower ceiling.

San Francisco: Offers a high revenue , but lower median earnings (make returns more volatile).

Los Angeles (Saturated): High competition dilutes typical earnings.

Strategy: Target Seattle for reliable, highreturns.

Statistical Analysis Technique: ANOVA

City	P-Value	Significance
LA	0.0000	Significant
SD	0.0000	Significant
SF	0.0000	Significant
Seattle	0.0000	Significant
Portland	0.0001	Significant

Statistic	Value	Conclusion
F-Stat	115.30	
P-Value	5.1546×10^{-98}	P-Value $\ll 0.05$

Purpose

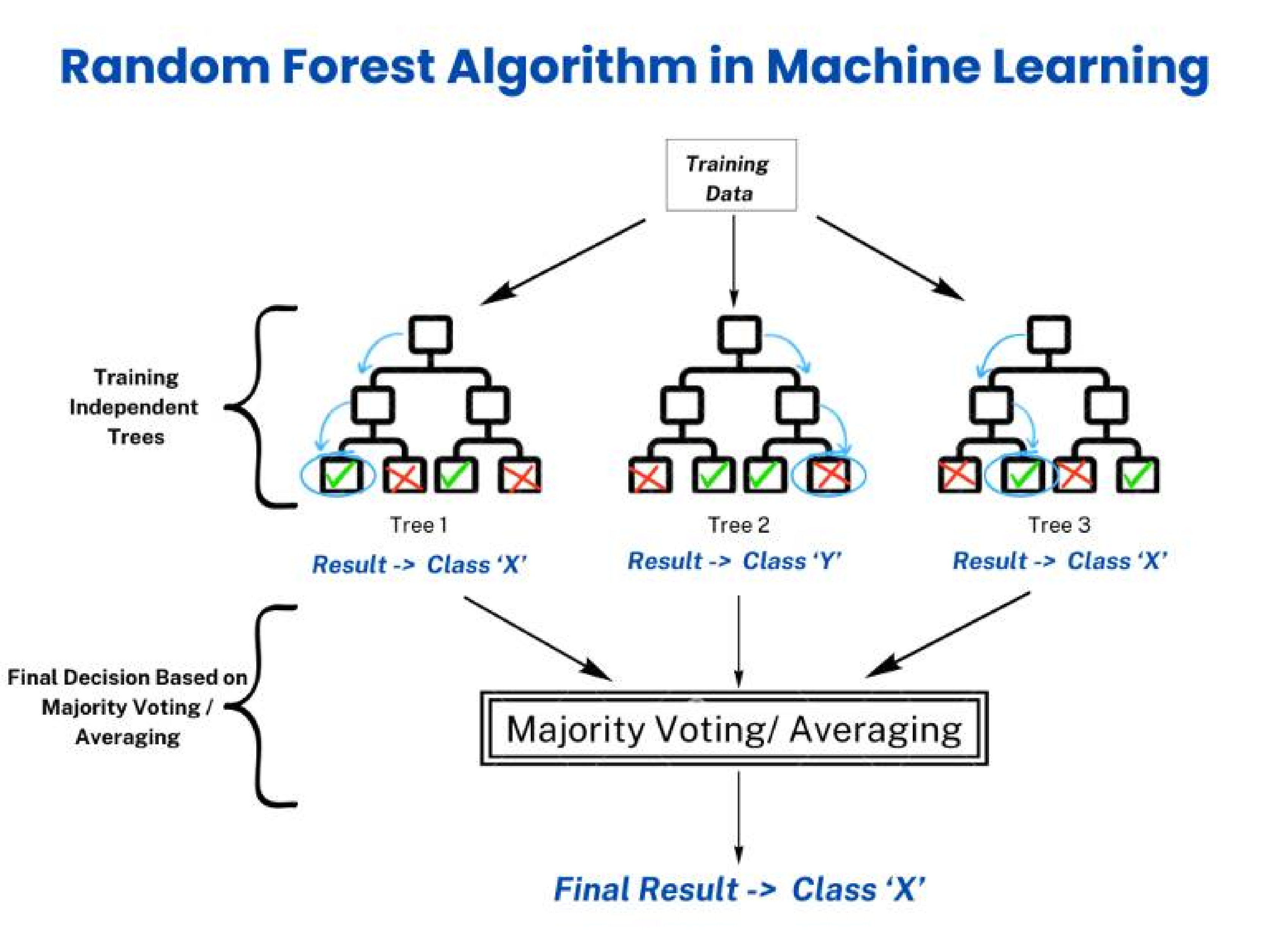
- Tested whether City significantly affects revenue.
- Tested whether Property Group significantly affects revenue within cities.

Findings

- ANOVA confirmed that Property Type is a statistically significant driver of revenue in all five cities (P < 0.0001). We reject the null hypothesis.

- Niche properties (Townhouses, Guest Suites, Condos) consistently outperform generic "Entire Home" listings.

Predictive Modeling Technique: Random Forest



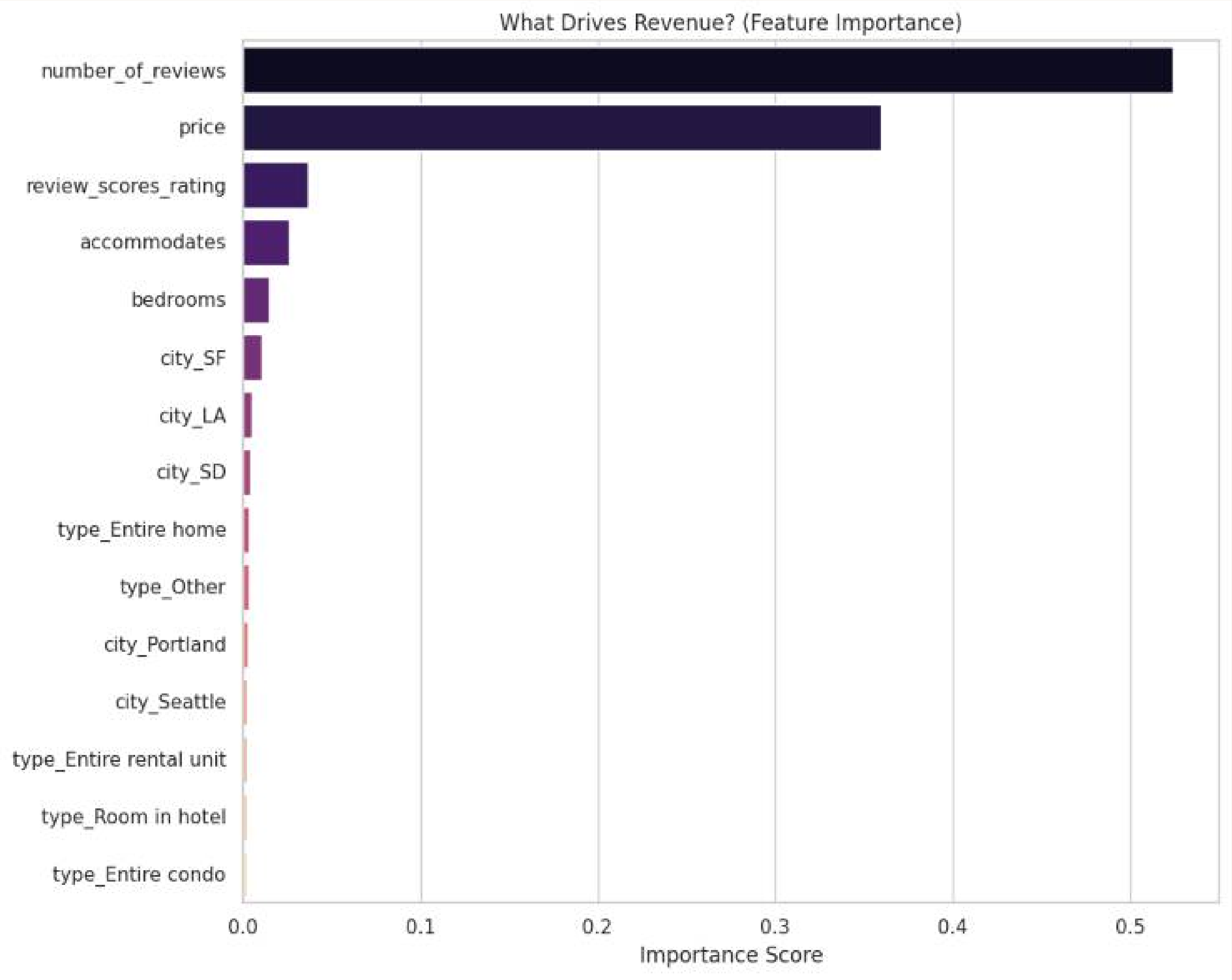
Features

- Features: *price*, *number_of_reviews*, *review_scores_rating*, *city*, *property_group*, etc.
- Target: *estimated_revenue_l365d*.
- Evaluated R^2 and feature importances to understand major drivers.

Findings

- The two most critical factors influencing annual revenue (explaining 87% of variance) are:
  - Number of Reviews (Feature Importance: 0.52)
  - Daily Price (Feature Importance: 0.35)

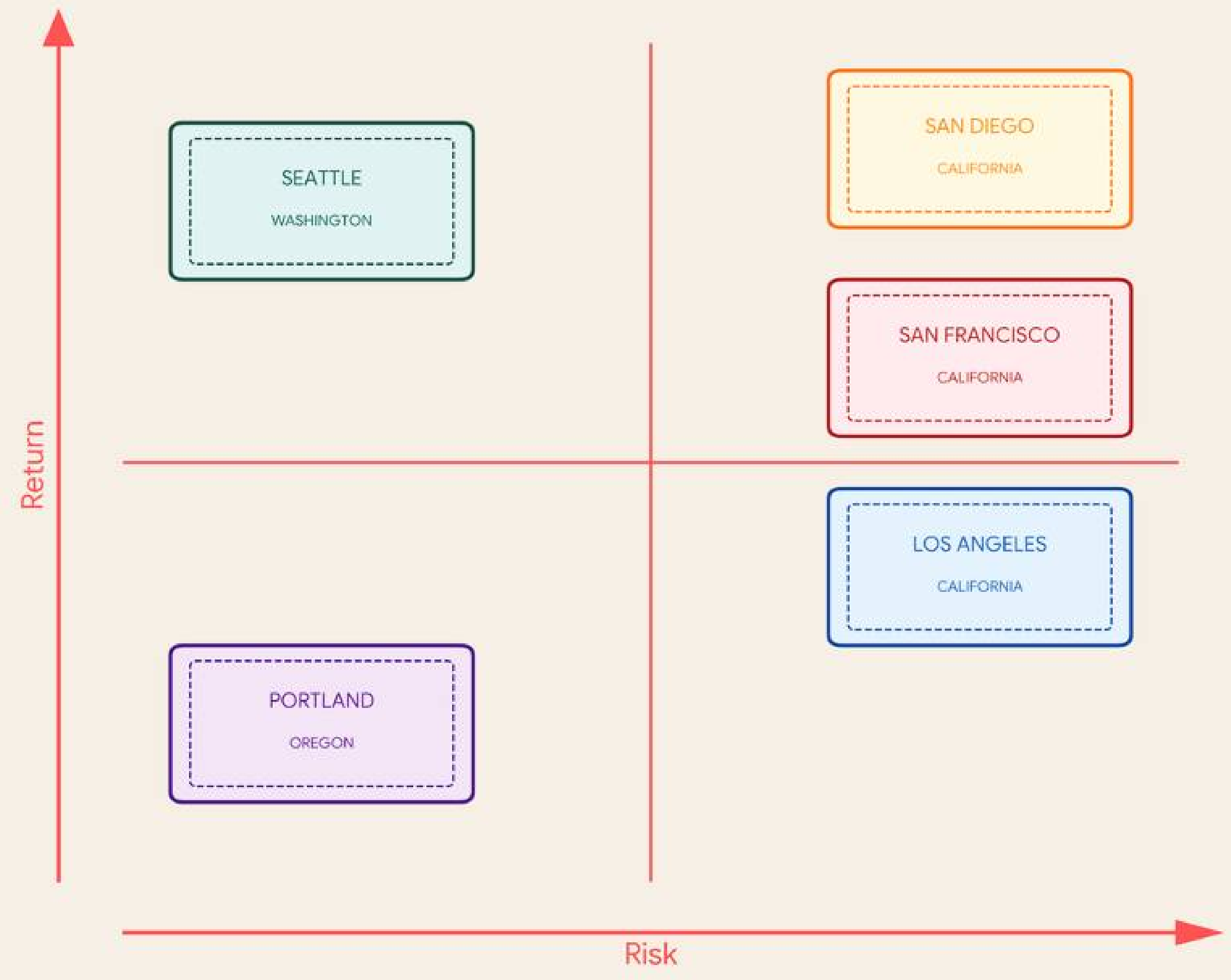
Model Performance & Drivers



Interpretation

The model explains **86%** of the variance in revenue, which is a strong fit. Interestingly, "City" has a low direct score because its value is largely captured by the Price and Review differences inherent to each location.

R-Squared (Accuracy)	0.8557
Mean Absolute Error	$10,728



Managerial Recommendations

For Hosts

- Where possible, invest in higher-performing property types (townhouses, condos, guest suites) rather than generic entire homes in saturated markets.
- Consider Seattle for safe, steady growth, while targeting San Diego and San Francisco for high-yield returns

For Airbnb

- Provide city-specific property-type guidance in onboarding.
- Highlight underserved, high-demand property types to encourage differentiated supply.

Learnings & Next Steps

H1, H2 & H3

Findings Summary

H1

Revenue varies sharply by city; Responsiveness and Superhost status consistently boost earnings.

H2

Richer amenity bundles, especially Outdoor/Safety/Business, improve ratings and reduce low-score risk.

H3

Property type is a significant revenue lever; niche home types outperform generic entire homes in several markets.

Strategic Roadmap

@Airbnb

Embed amenity bundle guidance, response-speed nudges, and city/property-type recommendations into host tools

@Airbnb

Adjust ranking algorithms to reward responsiveness, amenity richness, and historically high-performing property types

---

@Hosts

Optimize city + property-type choice where flexible

@Hosts

Maintain top-tier responsiveness and invest in 4+ amenity categories

@ Hosts

Calibrate data-driven pricing strategies to city demand/review strength

Limitations & Next Steps

One-year snapshot; does not fully capture seasonality.

Revenue estimates are model-based, not actual financial statements.

Future Improvements

- Time-series and seasonality analysis.
- Incorporate regulatory constraints & longer time horizons.
- Use richer demand/booking funnel data if available.

thank you!