Airbnb Pricing Optimization and Revenue Analytics – Group 6 Team Report

DNSC 4281: Pricing and Revenue Management Analytics

Jonathan Beck jonathan.beck@gwu.edu, Maya Serna mayalserna@gwu.edu,
Alp Sarioz alp.sairoz@gwu.edu, Martin Urruela martinurruela@gwu.edu

Professor Zhengling Qi

George Washington University

December 22, 2025

Introduction

The pricing challenge in DC for Airbnb is setting nightly rates, fees, and stay rules that maximize revenue and occupancy rate on the property rental market in a city with large, varied listings, fluctuating demand, and high seasonality. In order to avoid underpricing and missing opportunities, the platform and hosts both need to consider neighborhood variations, high-seasonality, listing features, and price elasticity.

Methodology & Data

Our analysis will optimize Airbnb's pricing strategies for Washington, DC using real-world data from Inside Airbnb, a publicly available repository that regularly scrapes, cleans, and publishes Airbnb listings. We used the March 13th 2025 snapshot to ensure the data reflects recent market conditions. The three datasets - Listings (62,257 properties and 79 features), Calendar (over 2 million entries of listing prices and availability), and Reviews (over 400,000 guest reviews) - will be wrangled to induce models and ultimately infer demand, occupancy, and pricing relationships.

Moreover, we will follow a standard analytical data mining approach: data cleaning and preprocessing → EDA → feature engineering → modeling and validation → interpretation of results. Throughout our process, we will be careful to immediately document any noteworthy insights we find.

To conduct our price optimization strategy, we will be using Python to clean the data and fit and evaluate several price-response models and assess its generalizability using RMSE, MAPE, and WMAP. Additionally, we are considering using R to support seasonal-trend decomposition. Lastly, to extend our analysis, we plan on applying LLM-based GenAI tools, such as ChatGPT and Gemini, to perform automated text analysis and classify the listings (e.g., budget vs luxury) to ultimately assess the relationship between guest satisfaction and occupancy and pricing outcomes.

Analysis and Results

The objective of this analytical stage was to quantify how prices, listing attributes, booking constraints, availability patterns, and reputational signal influence realized demand for Airbnb listings in Washington, DC. Because InsideAirbnb's public data repository includes substantial forward-looking availability that isn't indicative of demand, we implemented a careful time based-split to induce our training, validation, and holdout datasets: our training sample consisted of listings from March 13 2025 to May 1 2025; validation sample contained listings from May 2 2025 to July 1 2025; and a final test set was set aside, containing listings from July 2 2025 to July 22 2025 (the date that a booking was realized in the data). This was done to ensure that our models were induced using actual market outcomes and reflected true occupancy patterns rather than future projections and supply. Moreover, we leveraged Z-scores and boxplots to remove listings with outlier pricing and stay length characteristics. For example, because Stay Length values ranged from 1 to 270+ nights upon import, we decided to only keep bookings in which the guest(s) stay did not exceed 10 sequential nights, as including these extreme outliers would only distort the relationship between our demand indicators and stay duration. *Please see Appendix A.1 and A.2 for an example to illustrate how we applied these techniques in Python.*

Ultimately, our derived feature set incorporated various dimensions of demand determinants, including: *Pricing, capacity (beds, bedrooms, accommodations, room, and property types), policies (minimum/maximum night stays), availability pressures (the number of days available in the next 30, 90, and 365 days), reputation (cleanliness, location, and value scores, as well as total number of reviews), and temporal indicators.* This design ensured that pricing effects were evaluated in context of other relevant demand signals.

To examine how price and non-price attributes shape guest behavior, four demand specifications were estimated using Lasso-regularized regression: ***1. Linear Model****: Stay Length = f(Price + Features)* ***2. Exponential (log-linear)****: ln(Stay Length + 1) = f(Price + Features)* ***3. Constant Price Elasticity (log-log)****: ln(Stay Length + 1) = f(ln(Price + 1) + Features)* ***4. Logistic Model****: ln(Overall Booking Probability/(1 - Overall Booking Probability + Epsilon) = f(Price + Features).*

Model performance was evaluated using *RMSE, MAPE, and WMAPE* on our hold-out validation set. *Please see Appendix A.3 for table of performance metrics across all four models.* The *Logistic Model* achieved the lowest MAPE and WMAPE by a wide margin, reflecting strong discriminatory power (high AUC of 84.15%) and highly accurate booking-probability predictions. Although this model also returned the lowest RMSE, we did not use this metric to compare the Logistic Model with the other models, as they were used to predict *Stay Length* while the logistic model was, of course, used to derive a probability. Across the models predicting stay length, our *constant price elasticity* and *exponential models* performed best on percentage-based error metrics. However, because all three models predicting stay length returned RMSE values (~1.9 days) that were high relative to the (trimmed) standard deviation of stay length for listings in the validation set (which was ~2.95 days), we inferred that these models supplied limited ability in predicting stay length for guests. Therefore, because our stay-length demonstrated constrained explanatory power (*see Appendix A.4*) whilst our logistic model demonstrated much better predictive performance as demonstrated by the strong AUC (*see Appendix A.5*), predicting the likelihood of a booking using the Logistic Model became central to our objective of optimizing revenue across Washington DC Airbnb Listings.

Across our selected models, price and non-price attributes exerted meaningful (though sometimes modest) effects on realized demand:

Constant Price Elasticity Model *(see Appendix A.6 for visualization of influence of predictors)*

- Estimated an average price elasticity of -0.08, indicating weakly inelastic demand:
  - A 1% increase in price leads to only a 0.08% reduction in stay length.
    - This implies room for strategic upward price adjustments with limited demand erosion.
  - Availability features were among the most influential predictors:
    - *Availability_30* coefficient ~-0.104 indicates that each additional available day over the next month is associated with a ~10% decrease in stay length, holding all else constant. Interpreted economically, tighter short-term supply is naturally associated with longer stays.
  - Minimum-night requirements consistently increased expected stay lengths. This pattern suggests that policy rules structurally shape guest behavior and can be used to stabilize occupancy and reduce turnover.
  - Although mode in magnitude, review-based reputation effects were consistent:
    - Higher cleanliness, location, and value scores increased stay length and booking probability.
  - The binary indicator for private room returned meaningful shifts in stay length - a private-listing decreases stay length by ~2.5% relative to the baseline property type, holding all else constant.

Logistic Model *(see Appendix A.7 for visualization of influence of predictors)*

- The price coefficient in the logistic model returned a coefficient of ~-0.07, implying a one-unit increase in price decreases the likelihood of a booking by ~6.8%, on average and holding all else constant.

- Review scores for "Value" were strongly positive - a one-unit increased in this customer review metric increases the overall booking probability of a listing by ~21.8%, on average and holding all else constant.
- The binary indicator for an "Entire Serviced Apartment" exhibits ~20% lower likelihood of a booking relative to the baseline property type, on average and holding all other factors constant.

To quantify how price sensitivity varied across neighborhoods, we first derived micro-level elasticities per listing using our logistic model on our validation data, rather than our constant price elasticity model, due to our logistic model's more robust and stable performance metrics. Because the logistic model predicts the probability of a booking, its standardized price coefficient reflects how a one-unit change in the scaled price affects the log-odds of a booking occurring. Therefore, to convert this into an interpretable price elasticity for each listing, we applied a closed-form equation that adjusts for 1) the standard deviation used during feature scaling, 2) the listing's actual unscaled price, and 3) the fact that logistic response functions depend on predicted probabilities. *The implementation of this formula using Python can be seen in Appendix A.8, where scaled_price_coef is the standardized logistic price coefficient, price_std_dev is the price standard deviation, original_price_val is the original rate for the listing, logit_val_preds is the models predicted booking probability for that listing.* This formulation yields a true economic elasticity: the percent change in booking probability resulting from a 1% change in price, holding all other listing attributes constant.

After computing elasticities for all 570K+ listings in our validation set,we aggregated them to the neighborhood level by taking the average elasticity within each neighborhood cluster. *As the table at Appendix A.9 and accompanying KDE overlaid Histogram plot at Appendix A.10 illustrates, every neighborhood exhibited inelastic demand, with elasticities ranging from approximately -0.03 to -0.12.* This indicates that across Washington, DC, from May 2 2025 to July 1 2025, booking probability generally decreases only modestly in response to incremental price increases. The significance of these results is substantial: in an inelastic environment, moderate price increases should raise expected revenue because the percentage drop in booking probability is smaller than the percentage increase in price. Thus, these neighborhood-level elasticity profiles served as the foundation for our revenue optimization simulations, which explore how price adjustments translate into higher expected revenue for Airbnb.

LLM Application

Using our test dataset of listings (280K+ booking instances from July 2 2025 to July 22 2025), we worked closely with Gemini to derive and execute our revenue optimization strategy. We first estimated each listing's expected revenue under its current pricing strategy. Because our logistic model provides the predicted probability that a booking will occur at the observed price, expected revenue for each listed was calculated as:

*Expected Revenue per Listing = Price * Stay Length * Predicted Probability of Listing being Booked*

Summing across each listing grouped by neighborhood produced the baseline revenue level against which our optimization strategy was evaluated (*Appendix A.11 - Top 15 Neighborhoods Expected to generate the most revenue shown*).

The empirical finding that all neighborhoods in Washington, DC exhibit inelastic demand (*Appendix A.9 & A.10*), meaning that a percentage increase in price leads to a smaller percentage decrease in booking probability, was instrumental and guided our revenue optimization strategy. Specifically, we could reasonably expect that, given the inelastic environment, guests are relatively insensitive to incremental price increases. Thus, since the percentage decrease in booking probability (our proxy for demand) will be less than the percentage increase in price *across each neighborhood* we could expect that the total revenue for across each listing per neighborhood would increase at a faster rate than quantity demanded would decrease given an modest price increase.

As this section insinuates, Gemini was used to implement this logic by applying elasticity-adjusted booking probabilities under new (simulated by Gemini) price increases, allowing us to effectively simulate revenue outcomes without manually coding new prediction pipelines (*See Appendix A.12 for Subtask provided*). In turn, Gemini applied a 5% price increase to each listing rate, predicted new booking probabilities using our Logistic Model, calculated the new expected revenue, and then aggregated these metrics by neighborhood to illustrate the effects of our optimized revenue strategy per neighborhood. In applying the uniform 5% price increase to all listings in the test data, we can infer that this choice was guided by a couple considerations: 1) a 5% price increase is large enough to generate meaningful revenue differences yet small enough to remain competitive within Washington DC's B&B environment and 2) with elasticities between -0.03 and -0.12 across each neighborhood, the predicted reduction in booking probability from such an increase remains modest. The formulas applied by Gemini to calculate the updated expected revenue using the 5% price increase can be observed in the code snippet (*See Appendix A.13).*

Ultimately, the simulation demonstrated that applying a uniform 5% price increase resulted in:

- An overall increase of approximately $5.07 million in expected total revenue, representing a 4.62% gain relative to baseline levels.
- Positive revenue growth in every neighborhood, consistent with fully inelastic elasticities
- Neighborhoods such as Eastland Gardens/Kenilworth exhibiting the largest percentage gains in revenue (~4.87%), reflecting the neighborhoods extremely weak demand sensitivity.
- Neighborhoods such as Georgetown/Burleith-Hillandale showing slightly smaller but still positive gains (~4.45%), indicating relatively higher - but still inelastic - price sensitivity.

See Appendix A.14 for cumulative table illustrating these result and revenue increases across each neighborhood.
As a whole these results validate our core assumptions from our demand modeling endeavors: Airbnb demand in Washington, DC can tolerate modest price increases with limited detrimental booking loss effects, making upward price adjustments a viable component of revenue optimization strategies.

Recommendations and Business Impact

Our analysis demonstrates that both pricing and non-pricing attributes meaningfully shape booking outcomes Washington, DC Airbnb listings. Specifically, the Logistic Model's strong predictive performance paired with the consistently inelastic demand patterns we observed across all neighborhoods provide a clear foundation for actionable, data-driven recommendations for hosts and platform operators.

1. **Strategic Price Adjustments:** Because neighborhood-level elasticities were uniformly inelastic, in such an environment, moderate price increases lead to proportionally smaller reductions in booking probability, resulting in overall revenue growth. Moreover, the LLM-backed simulation confirms this relationship: a uniform 5% price increase yielded a 5.62% increase in total expected revenue across all listings, or roughly $5.07M, in the test dataset. Every neighborhood experienced positive revenue growth, with especially strong gains in the most inelastic areas such as Eastland Gardens/Kenilworth (~4.87%).

   Recommendation: Hosts should consider implementing controlled, moderate price increases, particularly during strong-demand periods or in neighborhoods with the lowest elasticities. Airbnb could develop automated "elasticity-aware pricing nudges" to help hosts make profitable adjustments without harming guest booking behaviors.

2. **Enhance Listing Quality and Reputational Signals:** The logistic models returned strong effects for non-price indicators:
   - Higher review score values, cleanliness, and location scores increase booking probability.
   - Number of reviews and superhost status also signal enhanced trust and credibility.
   - Property features such as (number of) accommodations/guests, bedrooms, beds, and room types influence demand.

   These factors meaningfully shift booking probabilities even when price is constant.

   Recommendation: Hosts should invest in improvements that directly enhance these properties.

3. **Leverage Availability and Forward-Looking Supply Signals to Guide Pricing:** Availability features, such as *availability_30, availability_90,* and *availability_365* were among the most influential predictors of both our booking probability and stay length models. Lower availability in the near future corresponded with longer stays and likely reflects stronger demand due to limited supply.

   Recommendation: Hosts should adjust prices based on near-term availability conditions:
   - Low availability → strategic price increases to capitalize on heightened customer demand.
   - High availability → more friendly and competitive pricing to capture missed booking opportunities.

   Airbnb might consider integrating these insights into a customer-facing booking assistant.

The combined modeling and LLM-based revenue optimization framework demonstrates that Airbnb's Washington DC market offers substantial opportunity for revenue enhancement by moderately increasing prices, improving non-pricing listing features, leveraging availability signals, and implementing predictive booking models to inform pricing strategies. These insights can drive the development of more advanced, elasticity-aware pricing assistants, and improve host recommendation systems and guidance to enhance the overall efficiency of the market. For individual hosts, adopting these recommendations may lead to hider revenue, steadier occupancy, and improved guest satisfaction, all which contribute to enhanced profitability margins.

Limitations and Future Work

This analysis uses Airbnb's public calendar and listing datasets to simulate realized demand and estimate the price–response relationship. While the workflow successfully cleans, structures, and enriches the data to infer booking behavior, several methodological and data-related limitations constrain the validity and interpretability of the results:

1. **Snapshot Bias**
   - The dataset represents a single Airbnb calendar scrape from March 13, 2025, rather than continuous booking data. As a result, bookings that began before this date are not visible, leading to undercounting of early demand and biasing occupancy and booking rate estimates.
2. **Availability Does Not Equal Booking**

- A central assumption in this analysis is that available=False indicates a booking. In reality, it can also mean the listing was blocked by the host, removed, or never made available. Similarly, available=True only shows that a listing could be booked. It may reflect future supply, not actual demand.
- This assumption likely overestimates demand when listings are blocked, and blurs the distinction between demand and supply when listings are newly posted or inactive.

3. **Future Projections vs. Realized Demand**
   Airbnb publishes listings up to a year in advance. Data after March 13, 2025, especially into 2026, represents future scheduled availability, not realized market behavior. Including these observations would distort demand analysis by:
   - Making it appear that occupancy is lower than reality,
   - Biasing elasticity estimates (demand appearing less responsive to price), and
   - Skewing revenue projections.
     - To reduce this bias, the analysis excludes all 2026 listings and retains only postings from early 2025 that likely reflect actual observed demand.
4. **Overcounting Continuous Bookings**
   - Each booking spans multiple daily rows in the dataset. Without careful transition logic, the same stay can be counted multiple times. The notebook mitigates this using new_booking and booking_id tracking, but irregular date gaps or missing records can still cause double-counting or missed bookings.
5. **Endogeneity of Price**
   - Prices on Airbnb are set by hosts based on expected demand and market trends. As a result, observed relationships between price and occupancy do not imply causation. Higher prices may reflect high-demand periods, not reduced demand sensitivity.
6. **Cross-Listing and Temporal Variation**
   - The analysis does not yet account for differences between listings (e.g., property type, location, amenities) or seasonal effects (e.g., weekends, holidays). These factors can confound the price–demand relationship and bias elasticity estimates.
7. **Granularity of Elasticity Estimations and Limited Segmentation**
   - Although our models revealed meaningful insights, we did not fully account for the heterogeneity arising from differences in property type, accommodations, beds, bedrooms, amenities characteristics, and other non-pricing variables. Grouping listings by these additional attributes could uncover more nuanced pricing sensitivities and customized pricing strategies. Currently, a one-size-fits all neighborhood elasticity may may substantial variation among luxury, budget, multi-bedroom, and one-room listings.

8. **Interpretation Challenges for Stay-Length Models**

   - All three stay-length models exhibited high RMSE relative to the standard deviation of stay length, despite trimming extreme outliers (any listing with a stay length greater than 10 days) from the dataset before modeling. As such, predicting stay duration remained difficult. It might be beneficial, going forward, if Inside Airbnb incorporated additional contextual variables such as *trip purpose proxies* to improve interpretability and prediction accuracy.

Appendix A- Primary Figures and Tables

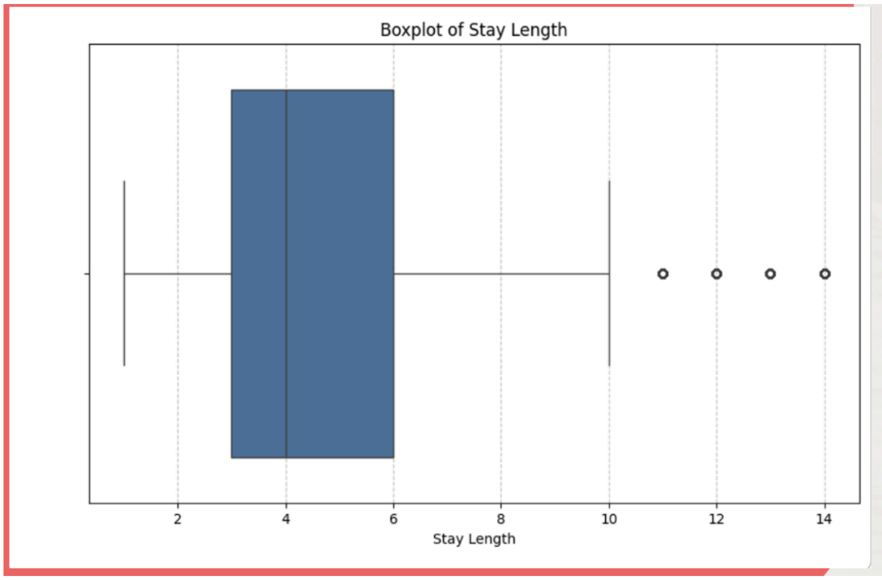
```
# Third pass of outlier removal
drop_rows_third_pass = booking.assign(
    stay_length_zscore = lambda x: np.abs((x.stay_length - x.stay_length.mean())/np.std(
        x.stay_length, ddof=1))
).loc[lambda x: np.abs(x.stay_length_zscore) >= 3].index

booking.drop(drop_rows_third_pass, axis=0, inplace=True)
```

```
booking.stay_length.describe()
```

	stay_length
count	1.982834e+06
mean	6.337001e+00
std	6.051852e+00
min	1.000000e+00
25%	3.000000e+00
50%	4.000000e+00
75%	7.000000e+00
max	3.800000e+01

dtype: float64

A.1



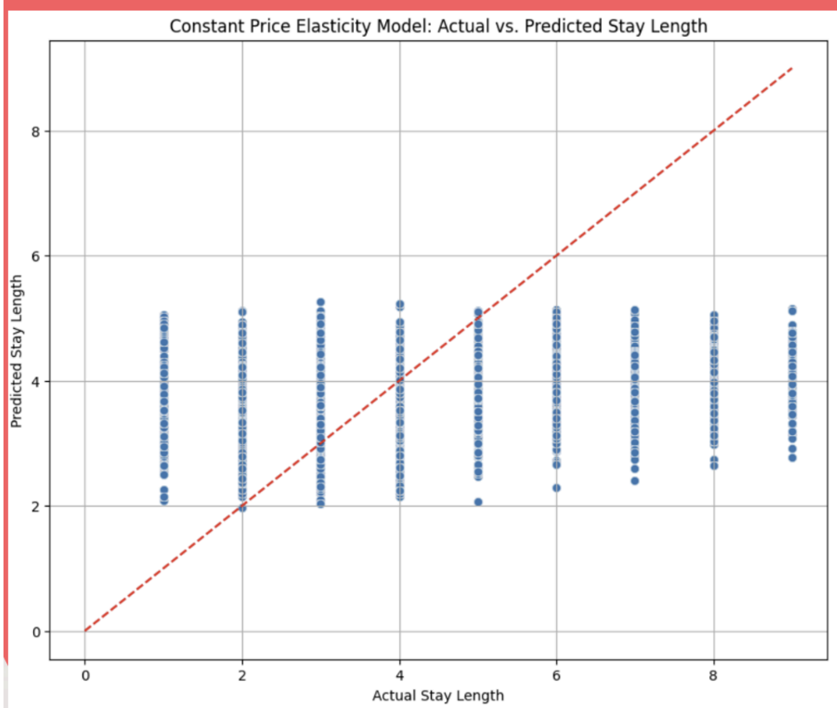
A.2

Model Performance Comparison:

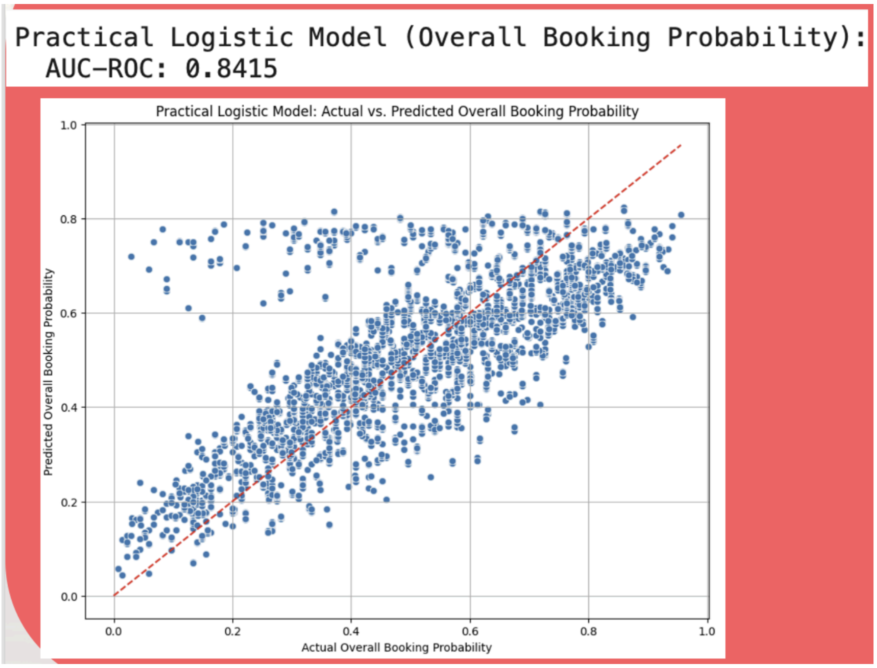
	RMSE	MAPE	WMAPE
Linear Model	1.889772	0.498545	0.360110
Exponential Model	1.915795	0.471634	0.359438
Constant Price Elasticity Model	1.915102	0.471546	0.359340
Practical Logistic Model	0.135964	0.279769	0.196993

A.3

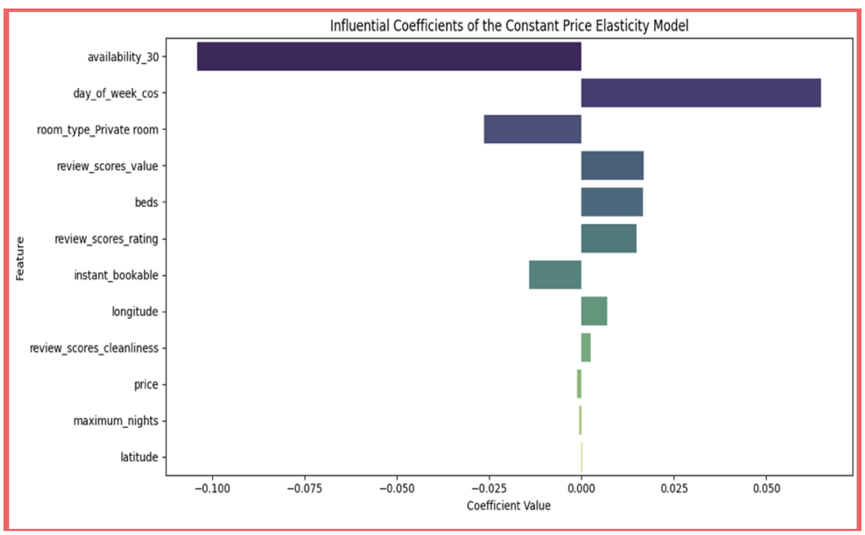
```
Constant Price Elasticity Model (Stay Length):
  R-squared: 0.0330
  Mean Squared Error (MSE): 3.6676
```



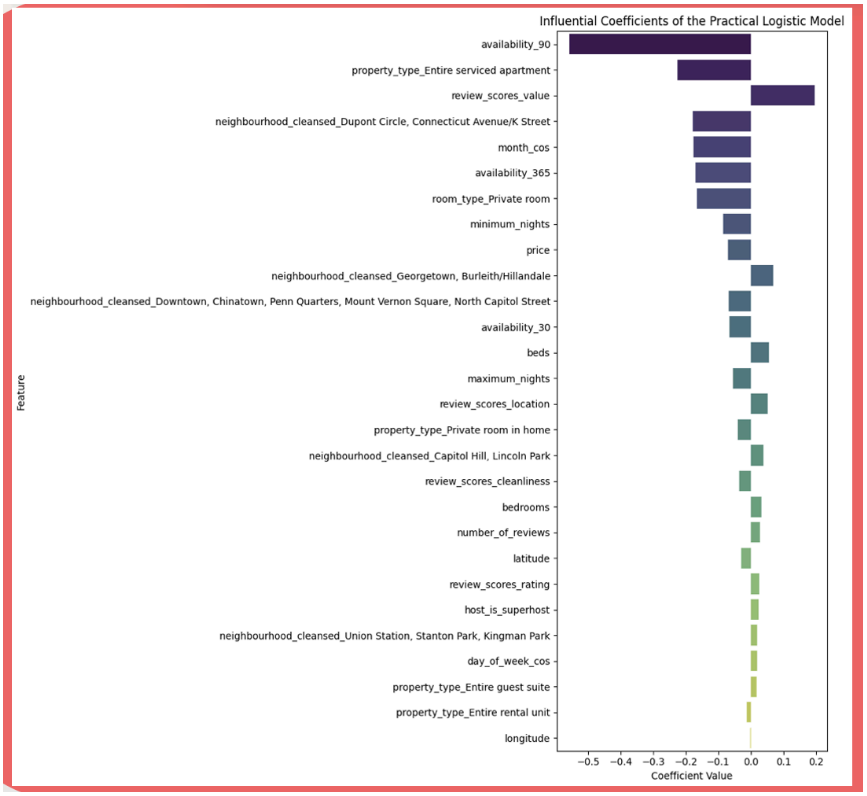
A.4


Practical Logistic Model (Overall Booking Probability):
AUC-ROC: 0.8415
Practical Logistic Model: Actual vs. Predicted Overall Booking Probability
Predicted Overall Booking Probability
Actual Overall Booking Probability
0.0
0.2
0.4
0.6
0.8
1.0

A.5


Influential Coefficients of the Constant Price Elasticity Model
Feature
availability_30
day_of_week_cos
room_type_Private room
review_scores_value
beds
review_scores_rating
instant_bookable
longitude
review_scores_cleanliness
price
maximum_nights
latitude
−0.100
−0.075
−0.050
−0.025
0.000
0.025
0.050
Coefficient Value

A.6



A.7

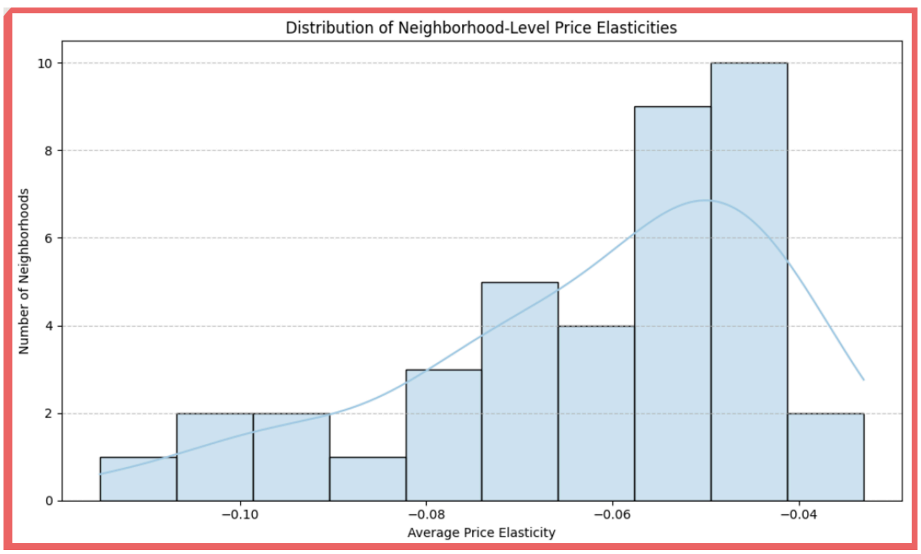
```
# Elasticity = (beta_scaled_price / price_std_dev) * original_price * (1 - predicted_probability)
elasticities = (scaled_price_coef / price_std_dev) * original_price_val * (1 - logit_val_preds)
```

A.8

Neighborhood-Level Price Elasticities (Average):

neighbourhood_cleansed	price_elasticity
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point	-0.115017
Cathedral Heights, McLean Gardens, Glover Park	-0.100308
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	-0.100138
Douglas, Shipley Terrace	-0.094022
Dupont Circle, Connecticut Avenue/K Street	-0.092085
Georgetown, Burleith/Hillandale	-0.084919
Near Southeast, Navy Yard	-0.078200
Ivy City, Arboretum, Trinidad, Carver Langston	-0.076289
West End, Foggy Bottom, GWU	-0.075300
Sheridan, Barry Farm, Buena Vista	-0.072850
Colonial Village, Shepherd Park, North Portal Estates	-0.072315
Kalorama Heights, Adams Morgan, Lanier Heights	-0.070860
Brookland, Brentwood, Langdon	-0.070574
Historic Anacostia	-0.067347
Mayfair, Hillbrook, Mahaning Heights	-0.065471
Woodridge, Fort Lincoln, Gateway	-0.060019
Woodland/Fort Stanton, Garfield Heights, Knox Hill	-0.059910
Congress Heights, Bellevue, Washington Highlands	-0.059080
Howard University, Le Droit Park, Cardozo/Shaw	-0.057151
Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir	-0.055736
Shaw, Logan Circle	-0.055568
Eastland Gardens, Kenilworth	-0.052172
Capitol Hill, Lincoln Park	-0.051921
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	-0.051595
Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont	-0.051482
Union Station, Stanton Park, Kingman Park	-0.051187
North Cleveland Park, Forest Hills, Van Ness	-0.050476
Cleveland Park, Woodley Park, Massachusetts Avenue Heights, Woodland-Normanstone Terrace	-0.047113
Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights	-0.045860
Hawthorne, Barnaby Woods, Chevy Chase	-0.044906
Takoma, Brightwood, Manor Park	-0.044667
Fairfax Village, Naylor Gardens, Hillcrest, Summit Park	-0.044511
North Michigan Park, Michigan Park, University Heights	-0.044133
Capitol View, Marshall Heights, Benning Heights	-0.043314
Friendship Heights, American University Park, Tenleytown	-0.043235
Edgewood, Bloomingdale, Truxton Circle, Eckington	-0.042415
Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill	-0.041245
Brightwood Park, Crestwood, Petworth	-0.038732
River Terrace, Benning, Greenway, Dupont Park	-0.033030

A.9



A.10

```
Initial Expected Revenue per Neighborhood:
```

neighbourhood_cleansed	expected_revenue
Capitol Hill, Lincoln Park	2.958056e+07
Union Station, Stanton Park, Kingman Park	2.128601e+07
Shaw, Logan Circle	8.507773e+06
Georgetown, Burleith/Hillandale	7.698340e+06
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	5.822381e+06
Howard University, Le Droit Park, Cardozo/Shaw	4.582482e+06
Edgewood, Bloomingdale, Truxton Circle, Eckington	3.900246e+06
Dupont Circle, Connecticut Avenue/K Street	3.509890e+06
Ivy City, Arboretum, Trinidad, Carver Langston	3.300779e+06
Kalorama Heights, Adams Morgan, Lanier Heights	2.956821e+06
Brightwood Park, Crestwood, Petworth	2.795950e+06
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point	2.737950e+06
Brookland, Brentwood, Langdon	2.393761e+06
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	2.144282e+06
Friendship Heights, American University Park, Tenleytown	9.805065e+05

A.11

Subtask:

Based on the previously calculated neighborhood-level price elasticities, simulate a simple revenue optimization strategy on the `X_test` data. This involves defining a percentage price adjustment (e.g., a 5% increase across all neighborhoods, given the observed inelasticity), calculating new prices, predicting new booking probabilities using the `best_logit_model` with these adjusted prices, and then computing the new expected revenue (`optimized_price * stay_length * new_predicted_probability`). This will be aggregated to show the total optimized expected revenue per neighborhood.

A.12

```
X_test_optimized_revenue = X_test_initial_revenue.copy()
X_test_optimized_revenue['price'] = X_test_optimized_revenue['price'] * 1.05

predicted_logit_values_optimized = best_logit_model.predict(X_test_optimized_revenue[model_features])
X_test_optimized_revenue['predicted_overall_booking_prob'] = 1 / (1 + np.exp(-predicted_logit_values_optimized))

X_test_optimized_revenue['optimized_expected_revenue'] = X_test_optimized_revenue['price'] * \
                                                         X_test_optimized_revenue['stay_length'] * \
                                                         X_test_optimized_revenue['predicted_overall_booking_prob']

optimized_revenue_per_neighborhood = X_test_optimized_revenue.groupby('neighbourhood_cleansed')['optimized_expected_revenue'].sum().sort_values(asc

print("Optimized Expected Revenue per Neighborhood (with 5% price increase):")
display(optimized_revenue_per_neighborhood)
```

A.13

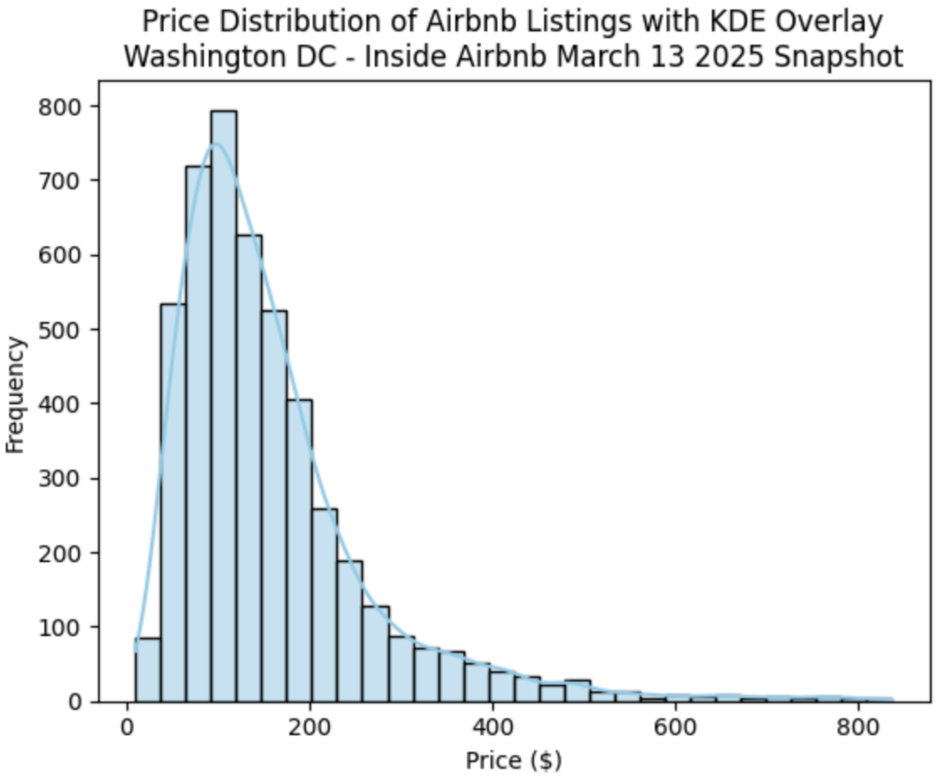
venue Comparison (Initial vs. Optimized with 5% Price Increase) per Neighborhood:

neighbourhood_cleansed	Initial Expected Revenue	Optimized Expected Revenue	Revenue Change	Percenta Chan
Eastland Gardens, Kenilworth	3.773176e+04	3.956753e+04	1.835771e+03	4.8653
Capitol View, Marshall Heights, Benning Heights	6.931428e+05	7.268136e+05	3.367089e+04	4.8577
Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir	1.987621e+05	2.083625e+05	9.600367e+03	4.8300
Cleveland Park, Woodley Park, Massachusetts Avenue Heights, Woodland-Normanstone Terrace	5.008264e+05	5.250054e+05	2.417901e+04	4.8278
Sheridan, Barry Farm, Buena Vista	3.573498e+04	3.745980e+04	1.724820e+03	4.8266
Fairfax Village, Naylor Gardens, Hillcrest, Summit Park	3.895014e+04	4.082999e+04	1.879851e+03	4.8263
River Terrace, Benning, Greenway, Dupont Park	7.329865e+04	7.682579e+04	3.527136e+03	4.8120
West End, Foggy Bottom, GWU	3.805688e+05	3.988417e+05	1.827289e+04	4.8014
Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont	4.803695e+05	5.033864e+05	2.301694e+04	4.7915
Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights	4.712561e+04	4.937547e+04	2.249857e+03	4.7741
Edgewood, Bloomingdale, Truxton Circle, Eckington	3.900246e+06	4.086442e+06	1.861967e+05	4.7739
Takoma, Brightwood, Manor Park	2.120041e+05	2.221242e+05	1.012012e+04	4.7735
Friendship Heights, American University Park, Tenleytown	9.805065e+05	1.027232e+06	4.672512e+04	4.7654
Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill	2.235016e+05	2.341178e+05	1.061612e+04	4.7499
Woodridge, Fort Lincoln, Gateway	5.496393e+05	5.756784e+05	2.603906e+04	4.7374
Union Station, Stanton Park, Kingman Park	2.128601e+07	2.229281e+07	1.006802e+06	4.7298
North Michigan Park, Michigan Park, University Heights	3.480662e+05	3.645062e+05	1.643992e+04	4.723217
Congress Heights, Bellevue, Washington Highlands	1.767337e+04	1.850607e+04	8.327035e+02	4.711628
Shaw, Logan Circle	8.507773e+06	8.907757e+06	3.999833e+05	4.701387
Brightwood Park, Crestwood, Petworth	2.795950e+06	2.927339e+06	1.313881e+05	4.699230
North Cleveland Park, Forest Hills, Van Ness	3.913650e+05	4.097477e+05	1.838261e+04	4.697050
Woodland/Fort Stanton, Garfield Heights, Knox Hill	2.798317e+05	2.929582e+05	1.312652e+04	4.690865
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View	5.822381e+06	6.095134e+06	2.727531e+05	4.684562
Mayfair, Hillbrook, Mahaning Heights	3.298043e+05	3.452162e+05	1.541183e+04	4.673023
Brookland, Brentwood, Langdon	2.393761e+06	2.505084e+06	1.113236e+05	4.650574
Cathedral Heights, McLean Gardens, Glover Park	1.050736e+05	1.099416e+05	4.867967e+03	4.632909
Kalorama Heights, Adams Morgan, Lanier Heights	2.956821e+06	3.093048e+06	1.362277e+05	4.607235
Near Southeast, Navy Yard	7.827469e+05	8.187442e+05	3.599729e+04	4.598842
Howard University, Le Droit Park, Cardozo/Shaw	4.582482e+06	4.792934e+06	2.104520e+05	4.592532
Hawthorne, Barnaby Woods, Chevy Chase	9.038214e+05	9.452308e+05	4.140940e+04	4.581591
Ivy City, Arboretum, Trinidad, Carver Langston	3.300779e+06	3.451724e+06	1.509452e+05	4.573018
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street	2.144282e+06	2.241857e+06	9.757531e+04	4.550489
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point	2.737950e+06	2.862503e+06	1.245528e+05	4.549126
Capitol Hill, Lincoln Park	2.958056e+07	3.092019e+07	1.339627e+06	4.528742
Dupont Circle, Connecticut Avenue/K Street	3.509890e+06	3.668368e+06	1.584785e+05	4.515199
Historic Anacostia	3.952356e+05	4.129749e+05	1.773925e+04	4.488272
Douglas, Shipley Terrace	6.072466e+05	6.342751e+05	2.702845e+04	4.450985
Georgetown, Burleith/Hillandale	7.698340e+06	8.040761e+06	3.424208e+05	4.447982

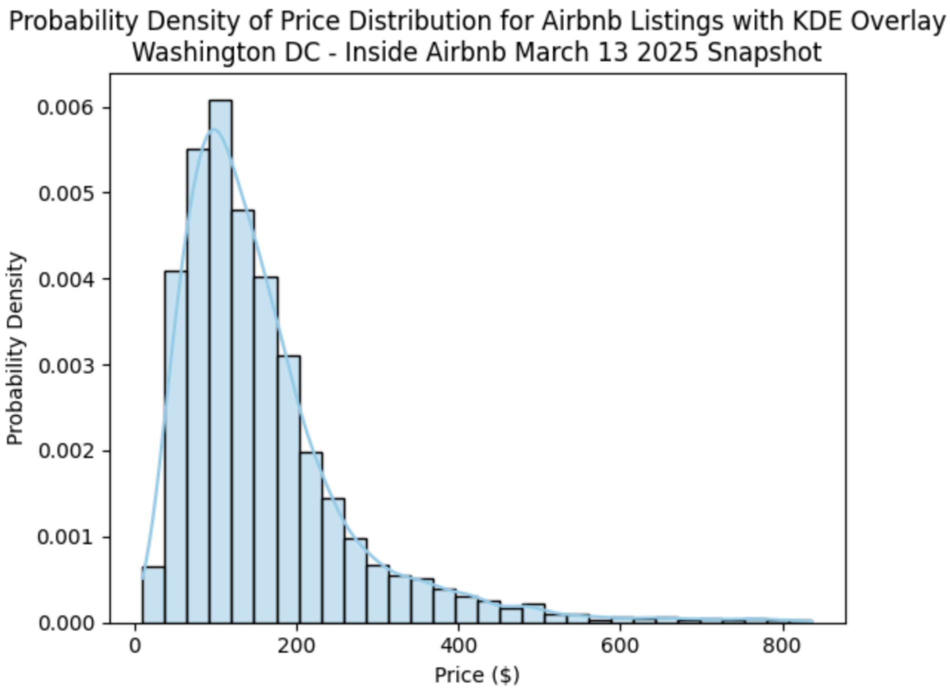
```
Total Initial Expected Revenue: 109830252.39679603
Total Optimized Expected Revenue: 114903672.64505917
Total Revenue Change: 5073420.248263136
Total Percentage Change: 4.619328588933605
```

A.14

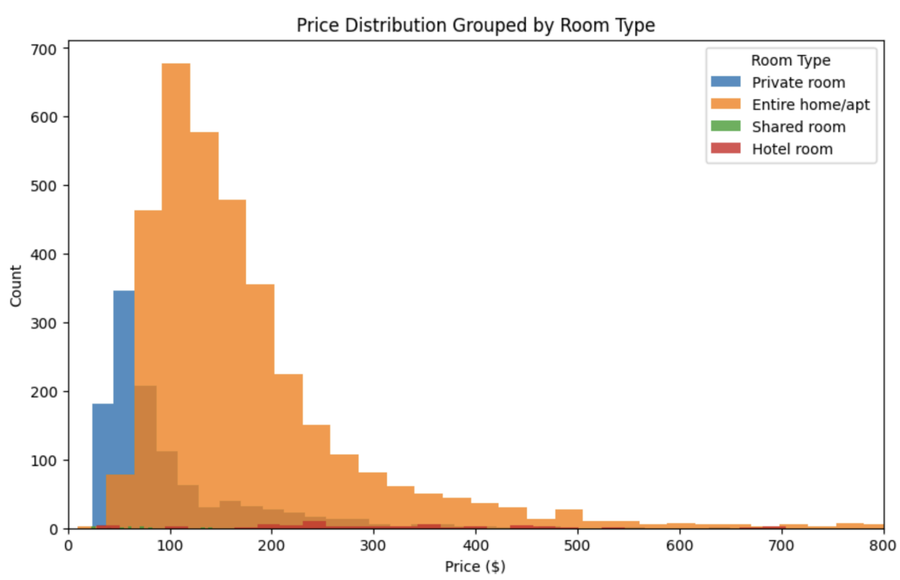
Appendix B - Additional Figures and Tables



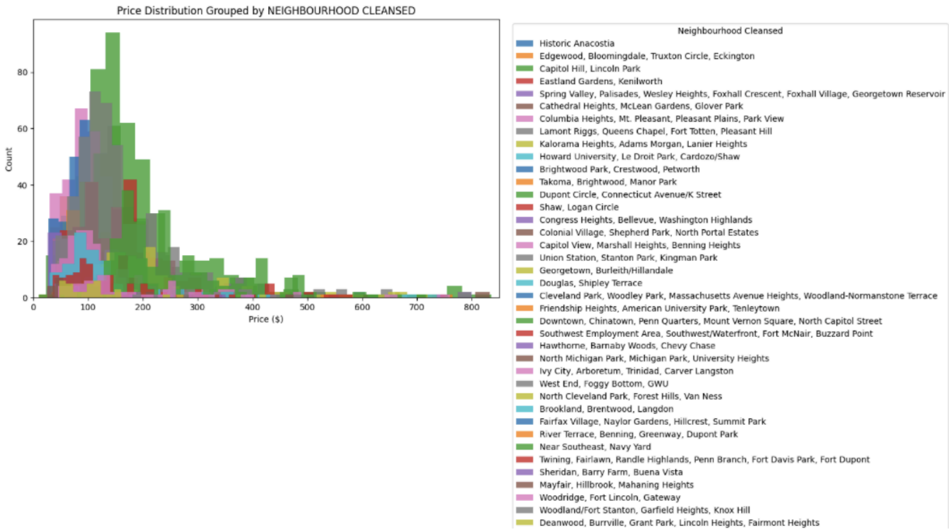
B.1


Probability Density of Price Distribution for Airbnb Listings with KDE Overlay
Washington DC - Inside Airbnb March 13 2025 Snapshot
Probability Density
Price ($)
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0
200
400
600
800

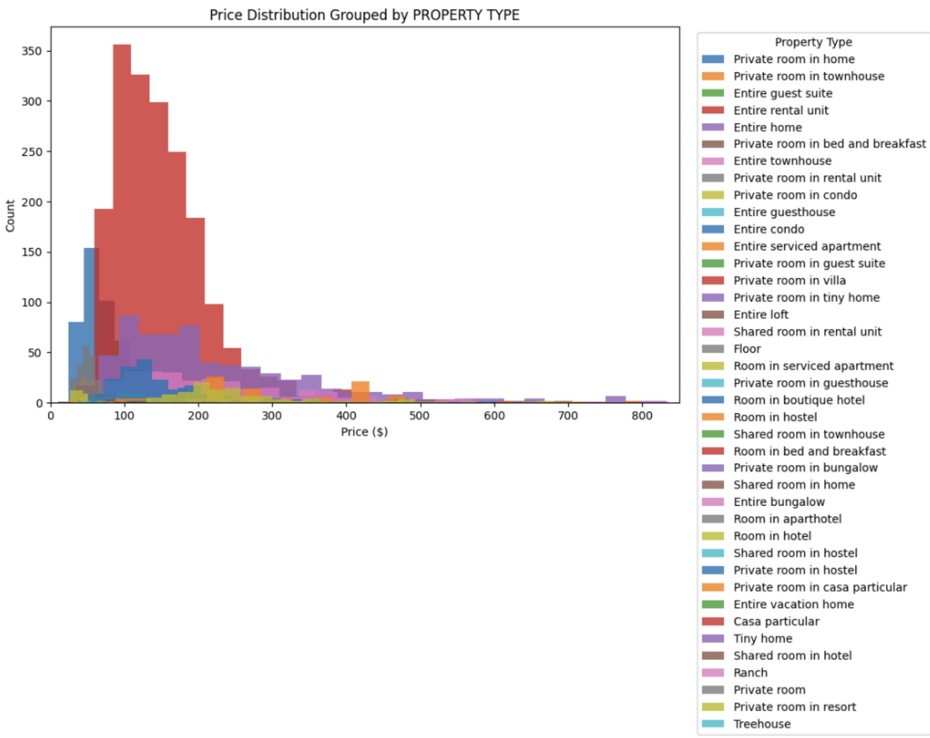
B.2


Price Distribution Grouped by Room Type
Room Type
Private room
Entire home/apt
Shared room
Hotel room
Count
Price ($)
0
100
200
300
400
500
600
700
800

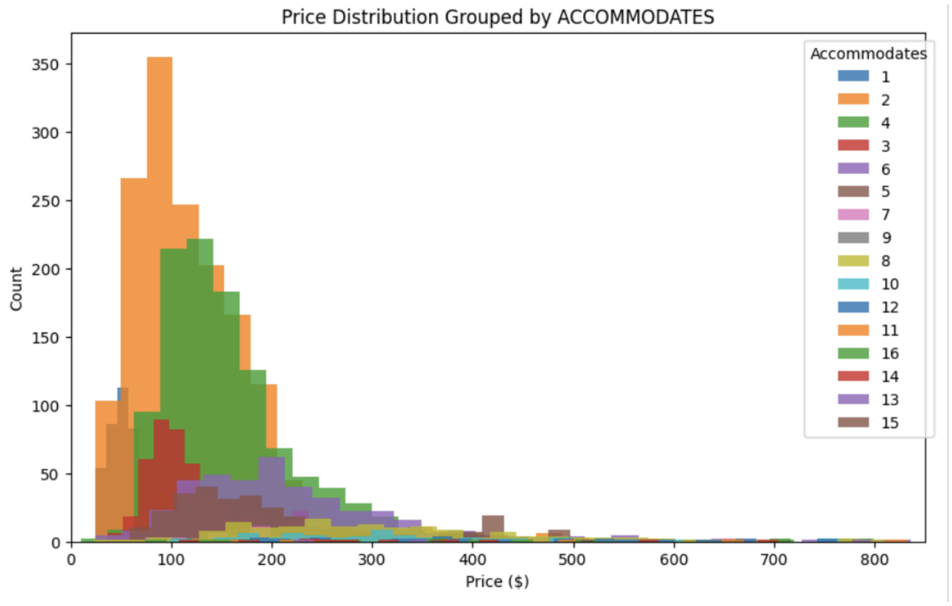
B.3


Price Distribution Grouped by NEIGHBOURHOOD CLEANSED
Count
Price ($)
Neighbourhood Cleansed
Historic Anacostia
Edgewood, Bloomingdale, Truxton Circle, Eckington
Capitol Hill, Lincoln Park
Eastland Gardens, Kenilworth
Spring Valley, Palisades, Wesley Heights, Foxhall Crescent, Foxhall Village, Georgetown Reservoir
Cathedral Heights, McLean Gardens, Glover Park
Columbia Heights, Mt. Pleasant, Pleasant Plains, Park View
Lamont Riggs, Queens Chapel, Fort Totten, Pleasant Hill
Kalorama Heights, Adams Morgan, Lanier Heights
Howard University, Le Droit Park, Cardozo/Shaw
Brightwood Park, Crestwood, Petworth
Takoma, Brightwood, Manor Park
Dupont Circle, Connecticut Avenue/K Street
Shaw, Logan Circle
Congress Heights, Bellevue, Washington Highlands
Colonial Village, Shepherd Park, North Portal Estates
Capitol View, Marshall Heights, Benning Heights
Union Station, Stanton Park, Kingman Park
Georgetown, Burleith/Hillandale
Douglas, Shipley Terrace
Cleveland Park, Woodley Park, Massachusetts Avenue Heights, Woodland-Normanstone Terrace
Friendship Heights, American University Park, Tenleytown
Downtown, Chinatown, Penn Quarters, Mount Vernon Square, North Capitol Street
Southwest Employment Area, Southwest/Waterfront, Fort McNair, Buzzard Point
Hawthorne, Barnaby Woods, Chevy Chase
North Michigan Park, Michigan Park, University Heights
Ivy City, Arboretum, Trinidad, Carver Langston
West End, Foggy Bottom, GWU
North Cleveland Park, Forest Hills, Van Ness
Brookland, Brentwood, Langdon
Fairfax Village, Naylor Gardens, Hillcrest, Summit Park
River Terrace, Benning, Greenway, Dupont Park
Near Southeast, Navy Yard
Twining, Fairlawn, Randle Highlands, Penn Branch, Fort Davis Park, Fort Dupont
Sheridan, Barry Farm, Buena Vista
Mayfair, Hillbrook, Mahaning Heights
Woodridge, Fort Lincoln, Gateway
Woodland/Fort Stanton, Garfield Heights, Knox Hill
Deanwood, Burrville, Grant Park, Lincoln Heights, Fairmont Heights

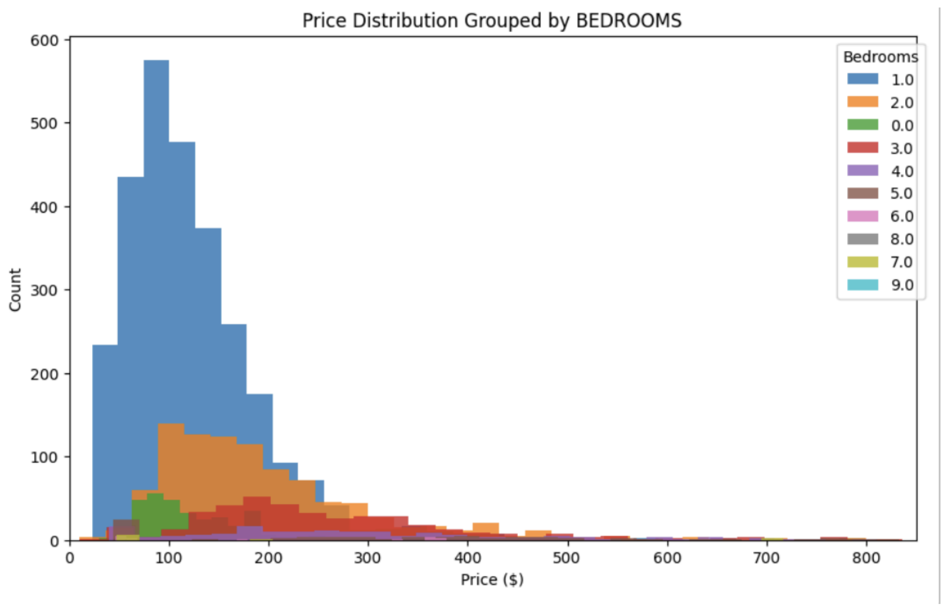
B.4


Price Distribution Grouped by PROPERTY TYPE
Count
Price ($)
Property Type
Private room in home
Private room in townhouse
Entire guest suite
Entire rental unit
Entire home
Private room in bed and breakfast
Entire townhouse
Private room in rental unit
Private room in condo
Entire guesthouse
Entire condo
Entire serviced apartment
Private room in guest suite
Private room in villa
Private room in tiny home
Entire loft
Shared room in rental unit
Floor
Room in serviced apartment
Private room in guesthouse
Room in boutique hotel
Room in hostel
Shared room in townhouse
Room in bed and breakfast
Private room in bungalow
Shared room in home
Entire bungalow
Room in aparthotel
Room in hotel
Shared room in hostel
Private room in hostel
Private room in casa particular
Entire vacation home
Casa particular
Tiny home
Shared room in hotel
Ranch
Private room
Private room in resort
Treehouse

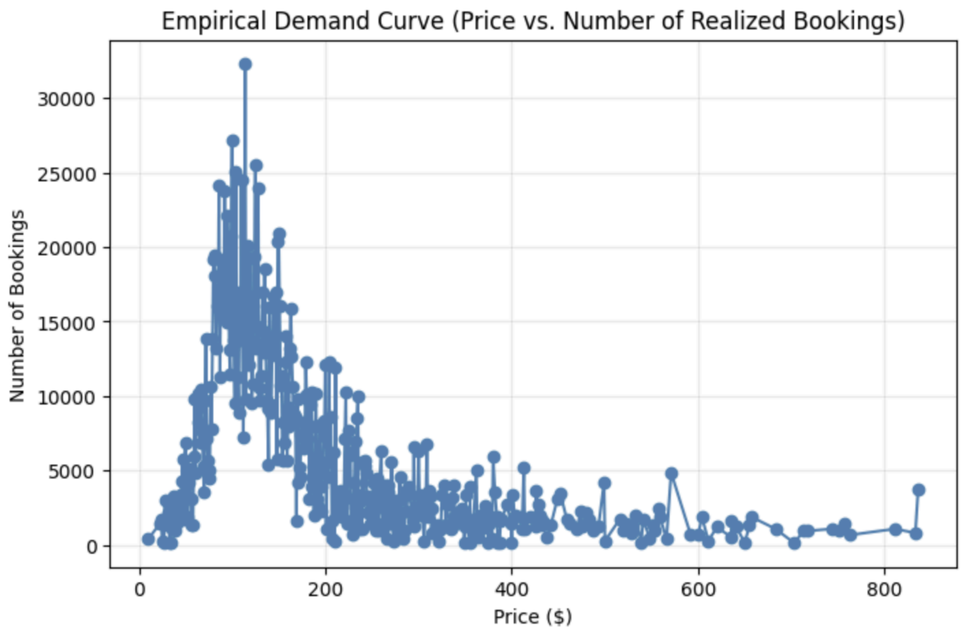
B.5


Price Distribution Grouped by ACCOMMODATES
Accommodates
1
2
4
3
6
5
7
9
8
10
12
11
16
14
13
15
Count
350
300
250
200
150
100
50
0
100
200
300
400
500
600
700
800
Price ($)

B.6


Price Distribution Grouped by BEDROOMS
Bedrooms
1.0
2.0
0.0
3.0
4.0
5.0
6.0
8.0
7.0
9.0
Count
600
500
400
300
200
100
0
100
200
300
400
500
600
700
800
Price ($)

B.7


Empirical Demand Curve (Price vs. Number of Realized Bookings)
Number of Bookings
30000
25000
20000
15000
10000
5000
0
0
200
400
600
800
Price ($)

B.8