MEMBERS & ROLES

Team Member	**Role**
Sai Akshith Bandari	Analytical Lead
Leslie Buckholtz	PM & Conceptual Design Lead
Cibi Siddarth	Technical Lead

1. *Data Description with Tables*

The dataset is a comprehensive collection of AirBnB property listings enriched with detailed metadata. It includes key identifiers such as listing and host IDs, which uniquely distinguish each property and its associated host. The dataset also captures various property features like capacity (accommodates, bedrooms, beds, bathrooms), descriptive textual data regarding the bathrooms, and categorical details such as property type and room type. Additionally, an extensive list of amenities offered per listing is provided, illuminating the various features that make each property unique. This detailed account ensures that every variable is well-defined, allowing for robust analysis of market trends, host performance, and guest preferences.

<table>
<tr><th>Tables</th><th>Parameters</th><th>Table Names</th><th>Description</th></tr>
<tr><td>Timeline</td><td>last_scraped
host_since
calendar_last_scraped
first_review
last_review
date_diff</td><td>tbl_Timeframe</td><td>
<table>
<tr><th>Description</th><th>Listing ID</th><th>last_scraped</th><th>host_since</th><th>calendar_last_scraped</th><th>first_review</th><th>last_review</th><th>Date Diff First and Last Review</th></tr>
<tr><td>Mean</td><td></td><td></td><td>5/30/2017</td><td></td><td>3/21/2021</td><td>4/28/2024</td><td>2.569025385</td></tr>
<tr><td>Standard Error</td><td></td><td></td><td></td><td></td><td>5.645197165</td><td>2.016556574</td><td>0.013718933</td></tr>
<tr><td>Median</td><td></td><td>9/21/2024</td><td>8/19/2016</td><td>9/21/2024</td><td>3/19/2022</td><td>8/7/2024</td><td>1.44109589</td></tr>
<tr><td>Mode</td><td></td><td>12/23/2024</td><td>7/22/2012</td><td>12/23/2024</td><td>5/27/2024</td><td>9/2/2024</td><td>0</td></tr>
<tr><td>Standard Deviation</td><td></td><td></td><td></td><td></td><td>1095.531927</td><td>391.3418865</td><td>2.927409447</td></tr>
<tr><td>Sample Variance</td><td></td><td></td><td></td><td></td><td>1200190.202</td><td>153148.4721</td><td>8.569726073</td></tr>
<tr><td>Kurtosis</td><td></td><td></td><td></td><td></td><td>-0.03293072</td><td>25.2453725</td><td>0.645814775</td></tr>
<tr><td>Skewness</td><td></td><td></td><td></td><td></td><td>-0.89602129</td><td>-4.39634381</td><td>1.204631147</td></tr>
<tr><td>Range</td><td></td><td></td><td></td><td></td><td>6096</td><td>5031</td><td>16.33972603</td></tr>
<tr><td>Minimum</td><td></td><td>6/24/2024</td><td>3/3/2008</td><td>6/24/2024</td><td>6/22/2008</td><td>5/23/2011</td><td>0</td></tr>
<tr><td>Maximum</td><td></td><td>3/2/2025</td><td>2/19/2025</td><td>3/2/2025</td><td>3/1/2025</td><td>3/1/2025</td><td>16.33972603</td></tr>
<tr><td>Count</td><td>21,208</td><td>45,534</td><td></td><td>45,534</td><td>37,661</td><td>37,661</td><td>45,533</td></tr>
<tr><td>Confidence Level(95.0%)</td><td></td><td></td><td></td><td></td><td>11.06473874</td><td>3.952505288</td><td>0.02688933</td></tr>
</table>
</td></tr>
</table>

Occupancy Rates	availability_30 availability_60 availability_90 availability_365 availability_eoy has_availability estimated_occupancy_l365d	tbl_Occupancy	(statistics table below)

	estimated_occupancy_l365d	availability_30	availability_60	availability_90	availability_365	availability_eoy
Mean	14.05681469	12.97373391	30.86219089	51.20599991	199.9913691	153.0114928
Standard Error	0.238974981	0.050324896	0.098697982	0.143800704	0.57004211	1.259332085
Median	0	12	32	57	212	159
Mode	0	0	0	0	0	0
Standard Deviation	50.99414763	10.7386772	21.06086354	30.6851969	121.6395601	111.4419956
Sample Variance	2600.403093	115.319188	443.5599729	941.5813085	14796.18258	12419.31839
Kurtosis	14.16365909	-1.37195014	-1.369655502	-1.169283847	-1.34216632	-1.461166015
Skewness	3.884479569	0.236568641	-0.127584086	-0.394206419	-0.18904029	-0.049129053
Range	255	30	60	90	365	306
Minimum	0	0	0	0	0	0
Maximum	255	30	60	90	365	306
Confidence Level(95.0%)	0.468394806	0.098637606	0.193449632	0.281851693	1.117291706	2.468627132

Pricing Strategy	price minimum_nights maximum_nights minimum_minimum_nights maximum_minimum_nights minimum_maximum_nights maximum_maximum_nights minimum_nights_avg_ntm maximum_nights_avg_ntm instant_bookable estimated_revenue_l365d	tbl_Pricing	(statistics table below)

	estimated_revenue_l365d	maximum_nights	price	minimum_nights	minimum_minimum_nights	maximum_minimum_nights	minimum_maximum_nights	maximum_maximum_nights	minimum_nights_avg_ntm	maximum_nights_avg_ntm
Mean	17862.12	449.9525849	306.4677553	12.91015505	12.32445079	15.70452548	378041.8177	991385.4211	13.21918717	913433.4781
Standard Error	331.7791263	22.04218818	3.120178075	0.133698626	0.127961335	0.178007287	133425.5767	216143.2586	0.132072099	205696.474
Median	8431.5	365	189	3	2	4	365	365	3	365
Mode	0	365	150	2	2	2	1125	1125	2	1125
Standard Deviation	25725.18728	4703.515802	631.0632472	28.52954497	27.30108381	37.97859628	28466902.58	46115064.6	28.17813253	43886199.6
Sample Variance	661785260.6	22123060.9	398240.822	813.9349361	745.349177	1442.373775	8.10365E+14	2.1266E+15	794.0071527	1.926E+15
Kurtosis	18.96642682	44794.83046	2123.432261	287.3229525	265.6243831	120.6986158	5685.624497	2162.857161	267.4460235	2373.803761
Skewness	3.179565576	210.7848489	33.63100194	12.70942644	12.15392298	9.050727554	75.4146838	46.52700409	12.33539757	48.68180425
Range	330225	999998	49991	1124	1124	1124	2147483646	2147483646	1124	2147483646
Minimum	0	1	9	1	1	1	1	1	1	1
Maximum	330225	999999	50000	1125	1125	1125	2147483647	2147483647	1125	2147483647
Sum	107387047	20488141	12536370	587851	561009	714870	17208463541	45127864368	601737.4	41579491923
Count	6012	45534	40906	45534	45520	45520	45520	45520	45520	45520
Confidence Level(95.0%)	650.4061026	43.2030434	6.115617611	0.262051457	0.250806276	0.34889715	261516.2787	423644.2672	0.258863441	403168.4012

Guest Reviews and Ratings	number_of_reviews number_of_reviews_ltm number_of_reviews_l30d number_of_reviews_ly reviews_per_month review_scores_rating review_scores_accuracy review_scores_cleanliness review_scores_checkin review_scores_communication review_scores_location review_scores_value	tbl_Reviews	(statistics table below)

	number_of_reviews	number_of_reviews_ltm	number_of_reviews_l30d	number_of_reviews_ly	review_scores_accuracy	review_scores_checkin	review_scores_cleanliness	review_scores_communication	review_scores_location	review_scores_rating
Mean	62.77596961	13.88808363	1.016690824	8.316306985	4.827540256	4.886424149	4.796473402	4.882340968	4.833751329	4.80857837
Standard Error	0.505254789	0.100001841	0.008568819	0.20959413	0.001649226	0.001346419	0.001695252	0.001431385	0.00143173	0.001669715
Median	18	4	0	1	4.91	4.96	4.89	4.96	4.92	4.89
Mode	0	0	0	0	5	5	5	5	5	5
Standard Deviation	107.8147898	21.33909012	1.828474326	18.54760024	0.319941427	0.261194951	0.328870238	0.277681321	0.277740876	0.324032204
Sample Variance	11624.02889	455.3567671	3.343318362	344.0134748	0.102362517	0.068222802	0.108155633	0.077106916	0.077139994	0.104996869
Kurtosis	15.05942695	151.0997408	100.6752244	49.33082441	56.11607872	85.3532596	43.66239689	77.04748155	48.13951011	49.76160958
Skewness	3.264472527	6.913027425	5.835963379	5.170023904	-6.20100162	-7.602410506	-5.243157035	-7.290638974	-5.319621404	-5.695310412
Range	1334	683	60	320	5	5	5	4	5	4
Minimum	0	0	0	0	0	0	0	1	0	1
Maximum	1334	683	60	320	5	5	5	5	5	5
Count	45534	45534	45534	7831	37634	37633	37634	37634	37632	37661
Confidence Level(95.0%)	0.990307514	0.196005217	0.016795023	0.410860457	0.003232528	0.002639018	0.003322741	0.002805553	0.00280623	0.003272686

<table>
<tr>
<td>Location</td>
<td>neighbourhood
neighbourhood_cleansed
neighbourhood_group_cleansed
host_location
host_neighbourhood
latitude
longitude</td>
<td>tbl_Location</td>
<td>
<table>
<tr><th>Most represented neighborhoods</th><th></th></tr>
<tr><td>Mission Beach</td><td>4192</td></tr>
<tr><td>Pacific Beach</td><td>3562</td></tr>
<tr><td>La Jolla</td><td>2181</td></tr>
<tr><td>North Park</td><td>1525</td></tr>
</table>
<table>
<tr><th>Least represented neighborhoods</th><th></th></tr>
<tr><td>San Mateo Park</td><td>1</td></tr>
<tr><td>Fraccionamiento Torres Del Lago</td><td>1</td></tr>
<tr><td>Forestland</td><td>1</td></tr>
<tr><td>Flamingos</td><td>1</td></tr>
<tr><td>Five Points</td><td>1</td></tr>
</table>
</td>
</tr>
<tr>
<td>Property Type and Amenities</td>
<td>property_type
room_type
accommodates
bathrooms
bathrooms_text
bedrooms
beds
amenities</td>
<td>tbl_PropertyType</td>
<td>
<table>
<tr><th></th><th>accommodates</th><th>bathrooms</th><th>bedrooms</th><th>beds</th></tr>
<tr><td>Mean</td><td>4.747265779</td><td>1.432731585</td><td>1.824614574</td><td>2.340361049</td></tr>
<tr><td>Standard Error</td><td>0.015071361</td><td>0.004965295</td><td>0.006639071</td><td>0.010758373</td></tr>
<tr><td>Median</td><td>4</td><td>1</td><td>1</td><td>2</td></tr>
<tr><td>Mode</td><td>2</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Standard Deviation</td><td>3.216032024</td><td>1.059529348</td><td>1.416691243</td><td>2.295696725</td></tr>
<tr><td>Sample Variance</td><td>10.34286198</td><td>1.122602439</td><td>2.007014077</td><td>5.270223453</td></tr>
<tr><td>Kurtosis</td><td>2.044409551</td><td>22.78499607</td><td>189.0568688</td><td>86.05815525</td></tr>
<tr><td>Skewness</td><td>1.415135755</td><td>2.688836282</td><td>5.455964523</td><td>4.672849646</td></tr>
<tr><td>Range</td><td>15</td><td>23</td><td>66</td><td>85</td></tr>
<tr><td>Minimum</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>Maximum</td><td>16</td><td>23</td><td>66</td><td>85</td></tr>
<tr><td>Confidence Level(95.0%)</td><td>0.029540109</td><td>0.009732059</td><td>0.013012686</td><td>0.021086585</td></tr>
</table>
<table>
<tr><th colspan="2">Frequency of each room type</th></tr>
<tr><td>room_type</td><td>frequency</td></tr>
<tr><td>Entire home/apt</td><td>36889</td></tr>
<tr><td>Hotel room</td><td>530</td></tr>
<tr><td>Private room</td><td>7908</td></tr>
<tr><td>Shared room</td><td>207</td></tr>
</table>
</td>
</tr>
<tr>
<td>Seasonality</td>
<td>calendar_updated
calendar_last_scraped
availability_30, availability_60,
availability_90, availability_365
first_review, last_review
minimum_nights_avg_ntm,
maximum_nights_avg_ntm</td>
<td>tbl_Seasonality</td>
<td>
<table>
<tr><th></th><th>minimum_nights avg_ntm</th><th>maximum_ nights_avg_ntm</th><th>availability_30</th><th>availability_60</th><th>availability_90</th><th>availability_365</th></tr>
<tr><td>Mean</td><td>13.21918717</td><td>913433.4781</td><td>12.97373391</td><td>30.86219089</td><td>51.20599991</td><td>199.9913691</td></tr>
<tr><td>Standard Error</td><td>0.132072099</td><td>205696.474</td><td>0.050324896</td><td>0.098697982</td><td>0.143800704</td><td>0.57004211</td></tr>
<tr><td>Median</td><td>3</td><td>365</td><td>12</td><td>32</td><td>57</td><td>212</td></tr>
<tr><td>Mode</td><td>2</td><td>1125</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>Standard Deviation</td><td>28.17813253</td><td>43886199.6</td><td>10.7386772</td><td>21.06086354</td><td>30.6851969</td><td>121.6395601</td></tr>
<tr><td>Sample Variance</td><td>794.0071527</td><td>1.926E+15</td><td>115.319188</td><td>443.5599729</td><td>941.5813085</td><td>14796.18258</td></tr>
<tr><td>Kurtosis</td><td>267.4460235</td><td>2373.803761</td><td>-1.371950137</td><td>-1.369655502</td><td>-1.169283847</td><td>-1.342166318</td></tr>
<tr><td>Skewness</td><td>12.33539757</td><td>48.68180425</td><td>0.236568641</td><td>-0.127584086</td><td>-0.394206419</td><td>-0.189040288</td></tr>
<tr><td>Range</td><td>1124</td><td>2147483646</td><td>30</td><td>60</td><td>90</td><td>365</td></tr>
<tr><td>Minimum</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>Maximum</td><td>1125</td><td>2147483647</td><td>30</td><td>60</td><td>90</td><td>365</td></tr>
<tr><td>Sum</td><td>601737.4</td><td>41579491923</td><td>590746</td><td>1405279</td><td>2331614</td><td>9106407</td></tr>
<tr><td>Count</td><td>45520</td><td>45520</td><td>45534</td><td>45534</td><td>45534</td><td>45534</td></tr>
<tr><td>Confidence Level(95.0%)</td><td>0.258863441</td><td>403168.4012</td><td>0.098637606</td><td>0.193449632</td><td>0.281851693</td><td>1.117291706</td></tr>
</table>
</td>
</tr>
</table>

<table>
<tr>
<td>Host Responsiveness</td>
<td>host_response_time
host_response_rate
host_acceptance_rate
host_is_superhost</td>
<td>tbl_Host</td>
<td>
<table>
<tr><th colspan="2">host_response_rate</th></tr>
<tr><td>Mean</td><td>0.971278882</td></tr>
<tr><td>Standard Error</td><td>0.000595694</td></tr>
<tr><td>Median</td><td>1</td></tr>
<tr><td>Mode</td><td>1</td></tr>
<tr><td>Standard Deviation</td><td>0.120406857</td></tr>
<tr><td>Sample Variance</td><td>0.014497811</td></tr>
<tr><td>Kurtosis</td><td>44.14388068</td></tr>
<tr><td>Skewness</td><td>-6.355740333</td></tr>
<tr><td>Range</td><td>1</td></tr>
<tr><td>Minimum</td><td>0</td></tr>
<tr><td>Maximum</td><td>1</td></tr>
<tr><td>Count</td><td>40856</td></tr>
<tr><td>Confidence Level(95.0%)</td><td>0.001167574</td></tr>
</table>
</td>
</tr>
<tr>
<td>Market Demand</td>
<td>host_listings_count
host_total_listings_count
calculated_host_listings_count
calculated_host_listings_count_entire_homes
calculated_host_listings_count_private_rooms
calculated_host_listings_count_shared_rooms
estimated_occupancy_l365d
estimated_revenue_l365d</td>
<td>tbl_MktDemand</td>
<td>
<table>
<tr><th></th><th>estimated_occupancy_l365d</th><th>estimated_revenue_l365d</th><th>host_listings_count</th><th>host_total_listings_count</th><th>calculated_host_listings_count</th><th>calculated_host_listings_count_entire_homes</th></tr>
<tr><td>Mean</td><td>14.05681469</td><td>17862.11693</td><td>112.5316389</td><td>165.501671</td><td>16.78793868</td><td>13.17870163</td></tr>
<tr><td>Standard Error</td><td>0.238974981</td><td>331.7791263</td><td>2.575797747</td><td>3.554649682</td><td>0.16080069</td><td>0.129186651</td></tr>
<tr><td>Median</td><td>0</td><td>8431.5</td><td>4</td><td>6</td><td>3</td><td>2</td></tr>
<tr><td>Mode</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Standard Deviation</td><td>50.99414763</td><td>25725.18728</td><td>549.3277468</td><td>758.082696</td><td>34.31277251</td><td>27.56674831</td></tr>
<tr><td>Sample Variance</td><td>2600.403093</td><td>661785260.6</td><td>301760.9734</td><td>574689.3739</td><td>1177.366357</td><td>759.9256123</td></tr>
<tr><td>Kurtosis</td><td>14.16365909</td><td>18.96642682</td><td>51.756702</td><td>42.57526367</td><td>19.44017357</td><td>13.97050391</td></tr>
<tr><td>Skewness</td><td>3.884479569</td><td>3.179565576</td><td>6.963972468</td><td>6.353016696</td><td>3.922410727</td><td>3.451462789</td></tr>
<tr><td>Range</td><td>255</td><td>330225</td><td>5254</td><td>9054</td><td>263</td><td>178</td></tr>
<tr><td>Minimum</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>Maximum</td><td>255</td><td>330225</td><td>5255</td><td>9055</td><td>264</td><td>178</td></tr>
<tr><td>Sum</td><td>640063</td><td>107387047</td><td>5118164</td><td>7527347</td><td>764422</td><td>600079</td></tr>
<tr><td>Count</td><td>45534</td><td>6012</td><td>45482</td><td>45482</td><td>45534</td><td>45534</td></tr>
<tr><td>Confidence Level(95.0%)</td><td>0.468394806</td><td>650.4061026</td><td>5.048605172</td><td>6.967170769</td><td>0.31517194</td><td>0.253207914</td></tr>
</table>
</td>
</tr>
</table>

2. *Key findings from your exploratory data analysis.*

A comprehensive exploratory data analysis (EDA) was conducted to uncover meaningful insights, trends, anomalies, and relationships within the Airbnb dataset. The goal of this analysis is to guide the development of predictive and clustering models while informing host strategy and business decisions. Key findings are summarized below:

2.1 Temporal and Group-Based Trends

- Seasonal Demand Patterns: Listings experience a clear seasonal trend, with review counts peaking between June and August, indicating increased guest activity during the summer. This suggests opportunities for dynamic pricing and targeted marketing during high-demand periods.
- Listing Growth Over Time: A year-over-year increase in total listings was observed, particularly post-2019. This growth aligns with Airbnb's rising popularity and may suggest an increasingly competitive market landscape in recent years.
- Superhost Advantage: Superhosts consistently outperform regular hosts across key metrics. They typically:
    - Charge 10–20% higher nightly rates
    - Receive more frequent and higher-rated reviews
    - Maintain lower cancellation rates and faster response times

These findings emphasize the value of achieving Super Host status to maximize occupancy and revenue.

2.2 Outliers and Anomalies

- Extreme Pricing: A small subset of listings priced over $1,000 per night significantly skew the overall price distribution. These listings likely represent luxury accommodations or potential data entry errors. Box plots and histograms helped identify and flag these outliers for further cleaning or segmentation.
- Zero-Review Anomalies: Numerous listings show full-year availability (365 days) but have zero guest reviews, which may indicate inactive or placeholder listings. These anomalies can distort average availability and occupancy calculations and should be filtered in modeling.
- Review Count Extremes: While the average listing has ~30–40 reviews, some listings exceed 500+ reviews, indicating highly established or corporate-managed properties. These high-volume listings require normalization when evaluating typical host performance.

2.3 Variable Relationships and Correlations

- Price vs. Review Count: Listings priced between $100–$200 generally receive the most reviews, suggesting this is the market's "sweet spot." Both ultra-low and high-priced listings tend to have lower engagement.
- Availability vs. Popularity: Listings with higher availability show a strong positive correlation with review count and guest feedback, implying that being available year-round increases occupancy potential.
- Review Score Breakdown:
- Cleanliness, Communication, and Location ratings have the highest correlation with overall score (Pearson's $r > 0.75$).
- These three factors should be considered critical KPIs (Key Performance Indicators) for hosts looking to improve listing quality and ratings.
- Room Type Impact: "Entire home/apartment" listings consistently outperform private or shared rooms in both price and rating, indicating a clear guest preference. These room types command higher average nightly prices and contribute to higher guest satisfaction.

2.4 Preliminary Sentiment Insights from Guest Reviews

- An initial sentiment analysis of guest reviews revealed overwhelmingly positive sentiment, with common keywords including:
- "Clean," "Convenient location," "Responsive host," "Comfortable stay," and "Easy check-in."
- These terms reflect guest priorities and should be highlighted in listings to attract more bookings.
- Full sentiment scoring (e.g., using TextBlob or VADER) and word cloud visualization will be included in later stages to identify themes that contribute most to 5-star reviews.

Conclusion

The EDA provided valuable insights into pricing dynamics, seasonal demand, host performance, and guest expectations. These findings lay the groundwork for selecting the most appropriate quantitative techniques in the next phase of the project, including:

- Regression models for price prediction
- Clustering algorithms for identifying high-performing neighborhoods
- Time series analysis for seasonality and occupancy forecasting
- Sentiment analysis to link guest feedback with listing performance

3. Propose and justify potential methods for your next analysis or modeling.

Based on the exploratory data analysis (EDA) and understanding of the Airbnb dataset, we propose a multi-faceted modeling approach tailored to address the core business objectives: accurate price prediction, neighborhood performance clustering, and actionable host insights through visualizations.

	Proposed Method	Justification
1. Accurate Predictions: Achieve high accuracy in predicting listing prices using machine learning models.	Multiple Linear Regression Random Forest Regressor XGBoost Regressor Artificial Neural Networks (ANN)	These models can effectively handle complex relationships in pricing data. Treebased models (Random Forest, XGBoost) are robust to outliers and non-linearities. ANN captures complex feature interactions.
2. Effective Clustering: Successfully identify high-performing neighborhoods through clustering analysis.	K-Means Clustering DBSCAN Hierarchical Clustering	Clustering helps group neighborhoods based on metrics like price, review scores, and availability. K-Means is effective for structured groups; DBSCAN detects noise/outliers; Hierarchical shows nested clusters.
3. Insightful Visualizations: Create clear and informative visualizations of geographic trends.	**POWER BI DASHBOARD** **Visualizations for General Information Purposes** **1. Price Distribution**: A histogram or box plot showing the distribution of listing prices. This can help users understand the range and average prices in different neighborhoods **2. Availability Heatmap**: A heatmap displaying the availability of listings throughout the year. This can help users identify peak and off-peak seasons **3. Geographic Trends**: A map visualization showing the geographic distribution of listings and their prices. This can help users visualize trends and identify high-demand areas **4. Review Sentiment Analysis**: A word cloud or sentiment score chart showing the most common words and overall sentiment in guest reviews. This can help users	**Visualizations to Help a User Become a Super Host** **1. Host Response Time**: A bar chart showing the average response time of hosts. Faster response times can lead to better guest experiences and higher ratings **2. Review Scores Breakdown**: A detailed breakdown of review scores for accuracy, cleanliness, check-in, communication, location, and value. This can help hosts identify specific areas to improve **3. Monthly Review Trends**: A line chart showing the number of reviews per month. This can help hosts track their performance over time and identify trends **4. Amenities Analysis**: A bar chart showing the frequency of different amenities offered by super hosts. This can help hosts understand which amenities are most valued by guests

	understand guest feedback and areas for improvement **5. Neighborhood Clusters**: A scatter plot or cluster map showing different neighborhoods based on rental performance and availability. This can help users identify high-performing neighborhoods	**5. Occupancy and Revenue Trends**: A line chart showing the estimated occupancy and revenue over the past year. This can help hosts understand their financial performance and identify opportunities for growth
4. Sentiment Analysis: Provide meaningful insights from guest reviews and their impact on listing success.	• TextBlob or VADER for sentiment scoring • TF-IDF or CountVectorizer for keyword extraction • WordCloud for visual sentiment representation	Using Natural Language Processing (NLP), we will extract sentiment scores from guest reviews and analyze their impact on: • Price variation • Overall Ratings • Booking frequency

Conclusion:

The proposed methods align with the key objectives of the project:

- Predictive models will optimize pricing strategy.
- Clustering will uncover high-performing neighborhoods.
- Power BI visualizations will provide actionable insights for hosts and business users.
- Sentiment analysis will connect guest feedback with listing performance and success metrics.