MULTI-CITY AIRBNB HOST STRATEGY & REVENUE OPTIMIZATION ANALYSIS

Travel & Hospitality Data Analytics Report

Submitted By:
Archit Gosain
Shivam Dixit
Yashi Agrawal
Price Singh
Kavya Mukhija
Pratyush Parida
Samay Samrat

Newton School of Technology
Data Visualization and Analysis

EXECUTIVE SUMMARY

Problem

The short-term rental market is highly competitive across major U.S. cities. Host performance varies significantly depending on pricing strategy, room type, occupancy management, and host professionalism. Understanding the drivers of revenue and occupancy is essential to optimize listing performance and improve overall platform profitability.

Approach

A dataset of 10,000 listings across multiple cities was analyzed using Google Sheets. Key performance indicators (KPIs) such as Average Price, Occupancy Rate, Estimated Annual Revenue, Revenue per Available Day, and Host Performance Score were developed. Exploratory Data Analysis (EDA) was conducted to evaluate pricing trends, host segmentation, room type performance, city-level revenue distribution, and demand patterns. An interactive dashboard was created to support strategic decision-making.

Key Insights

• Individual hosts show higher average occupancy as well as performance score and reviews as compared to professional hosts.
• Premium listings generate higher revenue but do not always achieve higher occupancy.
 • Budget and average price categories maintain more stable occupancy rates.
 • Revenue concentration is significantly higher in top metropolitan cities.
 • Minimum night requirements influence listing demand distribution.
• All the room types have occupancy rates similar to each other, but the revenue through Enitre Property listing is higher.

Key Recommendations

• Optimize pricing bands based on occupancy elasticity and have a wider category pricing.
 • Encourage professional host & commercial operators to improve pricing strategies and maintenance.
 • Focus acquisition efforts on high-revenue metropolitan markets.
 • Promote balanced minimum night policies to increase demand.
 • Use performance-based host segmentation for strategic planning.

SECTOR & BUSINESS CONTEXT

Travel & Short-Term Rental Industry Overview

The short-term rental industry has experienced significant growth over the past decade, driven by the expansion of digital platforms such as Airbnb and increased demand for flexible, experience-driven travel. Major metropolitan cities including New York, Los Angeles, San Francisco, and Chicago dominate revenue generation due to high tourism and business travel demand.

The market is characterized by intense competition among hosts, price variability, and demand fluctuations influenced by seasonality, location, and listing type. As supply increases across cities, host performance optimization becomes critical for sustaining revenue growth.

Business Challenges in the Industry

Despite strong market growth, several structural challenges exist:
• Revenue concentration in a small percentage of hosts
 • Pricing inefficiencies across different market segments
 • Variability in occupancy rates across host types
 • Imbalance between premium pricing and demand stability
 • Impact of minimum night requirements on booking behavior
Data-driven strategy is essential to identify performance drivers and optimize listing outcomes.

Why This Problem Was Chosen

The dataset of 10,000 listings across multiple U.S. cities reveals substantial differences in:
• Host type performance
 • Revenue distribution across cities
 • Price category impact on occupancy
 • Demand variation based on listing structure
Given the competitive nature of the travel and hospitality sector, understanding the relationship between price, occupancy rate, host segmentation, and estimated annual revenue is critical for strategic decision-making.
This analysis aims to bridge the gap between raw listing data and actionable business insights that can improve host performance and maximize revenue efficiency.

PROBLEM STATEMENT & OBJECTIVES

Problem Statement

The short-term rental market shows significant variation in revenue and occupancy across cities, pricing segments, and host types. Even within the same platform, listings generate different booking rates and annual revenues. The core problem addressed in this project is:
 How can revenue and occupancy be optimized by identifying high-performing hosts, effective pricing strategies, and demand patterns using data-driven analysis?

Project Scope

This analysis focuses on:
• 10,000 Airbnb listings across multiple U.S. cities
 • Host segmentation (Individual, Professional, Small and Commercial Operator)
 • Pricing categories (Budget, Average, Premium)
 • Occupancy rate and estimated annual revenue
 • Revenue distribution across cities
 • Minimum nights and demand behavior
External factors such as seasonality and competitor pricing are not included.

Project Objectives

• Compare host performance across segments
 • Analyze the relationship between price and occupancy
 • Identify high-revenue cities
 • Evaluate demand based on minimum night policies
 • Determine top revenue-generating hosts
 • Optimise pricing strategies and booking types
 • Generate actionable business insights

Success Criteria

The project will be successful if:
• Clear KPIs are defined
 • Revenue drivers are identified
 • Insights are supported by visual analysis
 • Strategic recommendations are data-backed
 • The dashboard enables performance monitoring

DATA DESCRIPTION

Dataset Overview

The dataset contains 10,000 Airbnb listings across multiple U.S. cities.
Each record represents an individual listing with details related to pricing,
location, host type, availability, and performance metrics.

Structure & Variables

The dataset includes a mix of numerical, categorical, and date variables.
Key variables include:
• Price – Nightly listing price
 • Calculated Host Listings - No. of listings of a articular host
• Number of reviews - No. of customer reviews received for the listing
• Availability 365 - No. of days the listing is available
 • Room Type – Entire Home/Apt, Private Room, Shared Room, Hotel Room
 • City – Listing location
 • Minimum Nights – Required booking duration

Derived Metrics

Additional performance indicators and metrics were calculated:
• Price Category– Budget, Premium or Average (Normalising wide range of values)
• Occupancy Rate – Rate of bookings
• Host type – Normalised category of hosts based on the number of listings they own
• Annual Revenue – Yearly revenue generated by listing
• Demand & Performance Score – Performance indicators based on price, review and occupancy

Limitations

• Revenue values are estimated
 • No seasonal breakdown available
 • No customer demographic data
 • No cancellation or competitor pricing data
 • No positive/negative review data for deeper analysis of property listing

DATA CLEANING & PREPARATION

Data Preparation

- Original dataset was received in ARFF format
- Converted to CSV format for analysis
- Randomly sampled and reduced from ~200,000 rows to 10,000 rows
- Dataset prepared in Google Sheets for EDA and dashboard development
- Preparation done using python → Google Colab Link

Cleaning Process

- Standardized categorical fields (Host Type, Room Type, City)
- Converted neighbourhood zip codes to text for consistency
- Removed duplicate and incomplete records
- Removed Outliers using IQR

Missing Values Handling

neighbourhood_group:
- 5000+ missing values detected
- No clear frequency pattern across neighbourhoods or cities
- Treated missing values as "Not Specified" category

neighbourhood:
- Mixed formats (text + zip codes)
- Converted all entries to text format
- Missing values in numerical columns replaced with zero

Minimum Nights Treatment

- Valid range defined: 1 – 90 nights
- Values ≤ 0 replaced with median
- Values > 90 replaced with median
- Ensured column is numeric and analysis-ready

Feature Engineering

- Calculated Occupancy Rate
- Estimated Annual Revenue
- Revenue per Available Day
- Price Category Segmentation
- Host Performance Score and Host Type Segmentation
- Price and Minimum Nights requirement Categorised

KPI & METRIC FRAMEWORK

Total Revenue

Definition

Total projected revenue generated by all listings over one year.

Formula

Total revenue = Sum of all estimated revenues.

Why it Matters

- Core financial health indicator
- Measures total economic output of the platform
- Reflects combined impact of price, occupancy, and supply

Objectives concerned

- Identify high-revenue cities
- Determine top revenue-generating hosts
- Generate actionable business insights

Average Occupancy Rate

Definition

Average percentage of available days booked across all listings.

Formula

Avg occupancy Rate = AVG(Occupancy Rate)

Why it Matters

- Indicates overall demand strength
- Reflects listing attractiveness
- Major driver of total revenue

Objectives concerned

- Analyze relationship between price and occupancy
- Evaluate demand based on minimum night policies
- Optimize pricing strategies

KPI & METRIC FRAMEWORK

Average Price

Definition
Average nightly price across all listings.

Formula

Avg Price = AVG(Price)

Why it Matters

- Reflects overall pricing positioning of the marketplace
- Directly impacts booking behavior
- Helps assess affordability vs revenue tradeoff

Objectives concerned

- Analyze relationship between price and occupancy
- Optimize pricing strategies

Revenue per Available Day (RevPAD)

Definition
Average revenue generated per available listing day across the platform.

Formula

RevPAD = Total Revenue ÷ Total Available Days

Why it Matters

- Measures revenue efficiency
- Standardizes performance beyond raw totals
- Highlights monetization effectiveness

Objectives concerned

- Compare host performance across segments
- Optimize pricing strategies
- Generate actionable business insights

KPI & METRIC FRAMEWORK

Average Minimum Nights

Definition
Average minimum stay requirement across listings.

Formula
avg minimum requirement = AVG(Minimum Nights)

Why it Matters
- Indicates booking flexibility
- Influences occupancy and demand
- Impacts short-term vs long-term rental strategy

Objectives concerned
- Evaluate demand based on minimum night policies
- Optimize booking types

Total Listings (Supply Volume)

Definition
Total number of active listings on the platform.

Formula
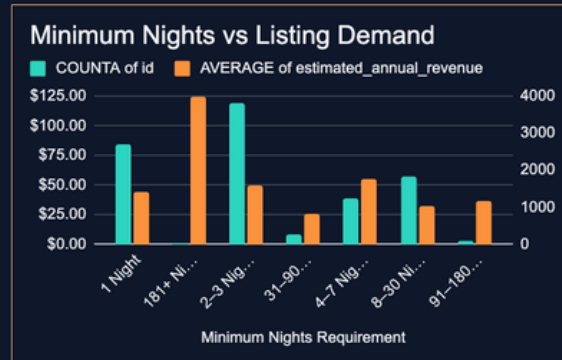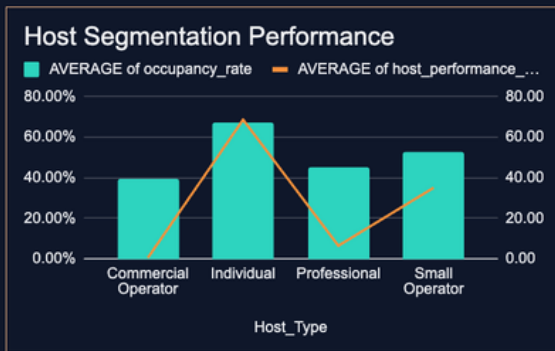Total listings = COUNT(Listing ID)

Why it Matters
- Measures platform scale
- Provides context for revenue performance
- Indicates supply-side growth

Objectives concerned
- Identify high-revenue cities
- Determine top revenue-generating hosts

EXPLORATORY DATA ANALYSIS

Trend Analysis
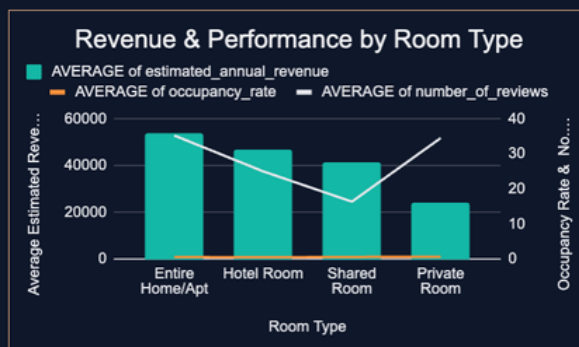
Analyses Conducted

• Revenue trends across cities
 • Occupancy variation across room types
 • Performance variance by host segment
 • Minimum night trends across segments

Key Insights

- Revenue is concentrated in a few high-performing cities, indicating market dominance in metropolitan cities rather than evenly distributed growth.
- Entire homes generate higher average revenue but may not always show proportionally higher occupancy.
- Private Rooms give out higher and stable occupancy rate depicting customer loyalty.
- Professional and commercial hosts contribute disproportionately to total revenue. This states either over pricing or questionable host behaviour.
- However, individual and small operator hosts tend to attract decent amount of customers and maintain balance between pricing and occupancy.
- Listings with very high minimum night requirements tend to show lower occupancy rates and thus lower estimated revenue

EXPLORATORY DATA ANALYSIS

Comparison Analysis
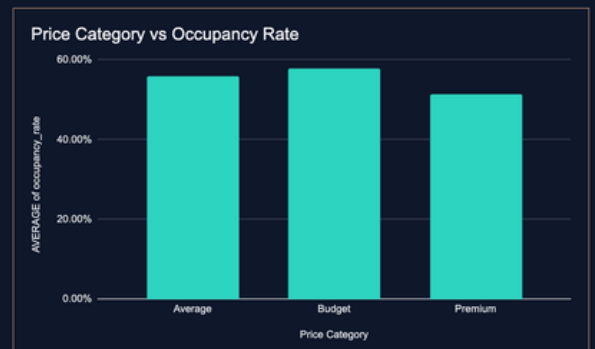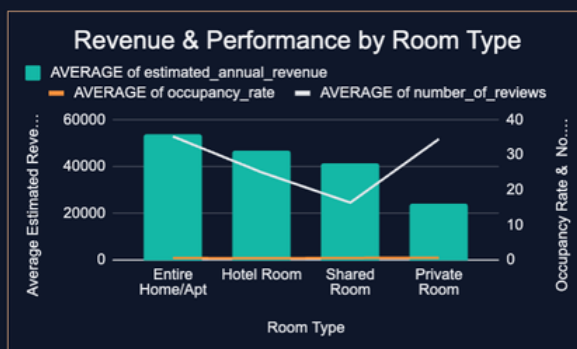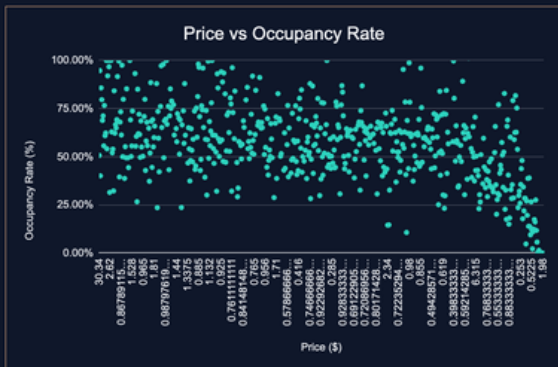
Analyses Conducted

- Host performance comparison
- Room type performance comparison
- City-wise revenue comparison
- Price vs occupancy comparison

Key Insights

- Commercial operators generate higher revenue per listing but may operate with lower occupancy efficiency.
- Private rooms tend to have higher occupancy rates but lower revenue per listing compared to entire homes.
- High-revenue cities do not always have the highest occupancy, suggesting pricing power.
- Higher pricing does not always correlate with higher revenue, indicating demand elasticity.

EXPLORATORY DATA ANALYSIS

Distribution Analysis
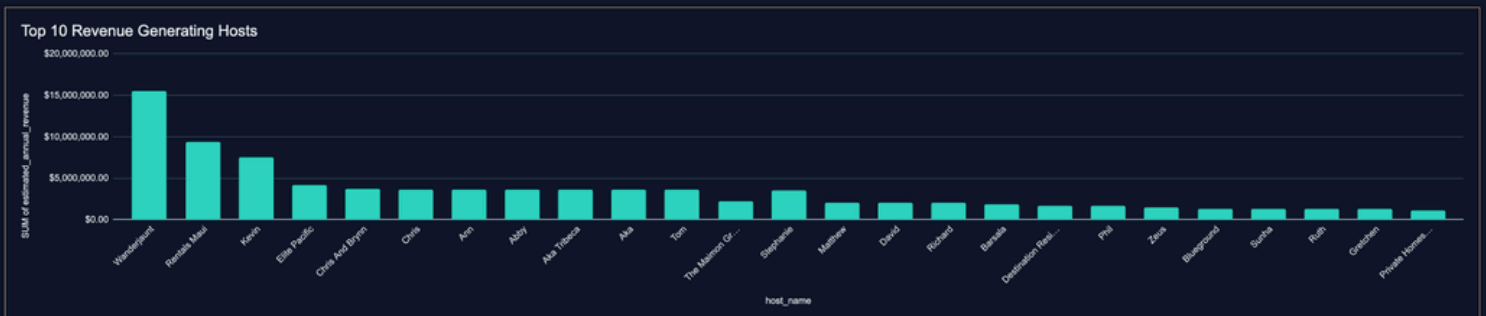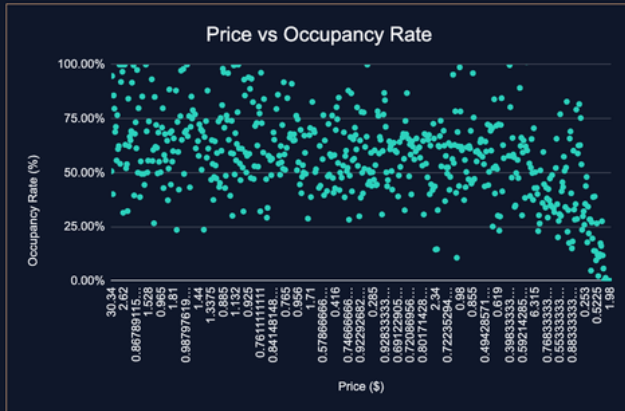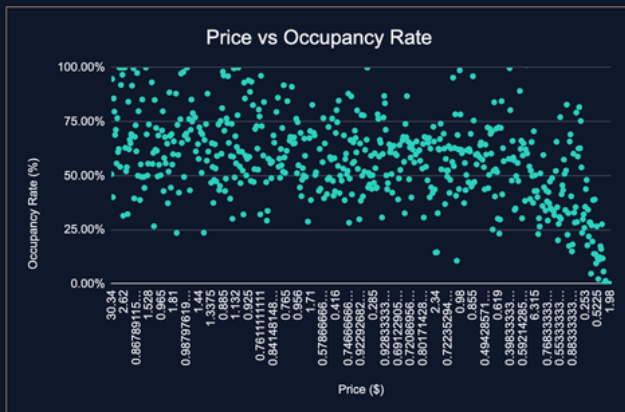

Analyses Conducted

• Revenue distribution
 • Price distribution
 • Occupancy distribution
 • Listing count distribution per host

Key Insights

- Revenue distribution is right-skewed, with a small percentage of listings contributing disproportionately to total revenue.
- Price distribution shows clustering within mid-range values, indicating competitive pricing behavior.
- Most hosts own only 1 listing, while a small number of commercial operators control a large share of inventory.
- Occupancy rates show moderate variance, suggesting demand stability across listings.
- Premium price category demonstrates higher revenue but low demand score.

EXPLORATORY DATA ANALYSIS

Correlation Analysis

Price vs Occupancy Rate

Listing Level Demand

Minimum Nights vs Listing Demand

Analyses Conducted

- Price vs Occupancy
- Minimum Nights vs Occupancy
- Reviews vs Revenue
- Listings per Host vs Revenue
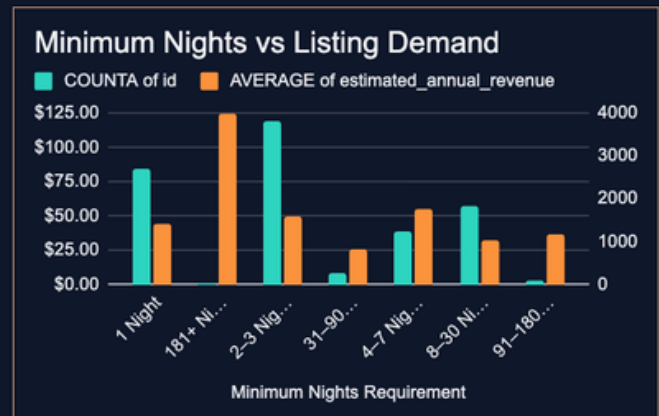
Key Insights

- Weak to moderate inverse relationship between price and occupancy, indicating price sensitivity.
- Higher minimum night requirements generally reduce occupancy.
- Positive relationship between number of reviews and revenue, suggesting social proof drives bookings.
- Hosts with multiple listings generate higher total revenue but not necessarily higher efficiency per listing.

ADVANCED ANALYSIS

Revenue Concentration

Revenue is concentrated in a few major metropolitan cities, with a small percentage of listings contributing disproportionately to total revenue. This indicates strong demand clustering and market dominance in top cities.

Host Performance Insights

Individual hosts show higher average occupancy compared to professional segments. This suggests operational flexibility and localized management may influence booking efficiency.

Pricing & Demand Dynamics

Higher pricing does not guarantee higher occupancy. Budget and mid-range segments maintain more stable demand patterns, indicating the importance of price-demand balance.

Key Revenue Drivers

Primary factors influencing performance:
• City-level demand
• Pricing category
• Host type
• Minimum night policy
• Listing availability

DASHBOARD DESIGN

The dashboard was developed in Google Sheets using pivot tables, formulas, and interactive filters. It integrates KPIs, segmentation insights, and comparative visualizations into a single analytical interface. The objective is to enable performance monitoring and support strategic decision-making.

Dashboard Structure

- KPI Summary Panel (Revenue, Listings, Occupancy, Pricing)
- Analytical Charts (City comparison, Pricing vs Occupancy, Host Segmentation)
- Demand & Policy Insights (Minimum Nights vs Listings)
- Top Revenue-Generating Hosts

Filters & Interactivity

Charts and Plots can be filtered by:-

- Neighbourhood
- Host Type
- Minimum Nights
- Price Categories
- Room Type

INSIGHTS SUMMARY

Key Findings

- Revenue is highly concentrated in a few major metropolitan cities. This demonstrates skewed distribution of industry.
- Individual hosts achieve higher average occupancy compared to professional   segments.
- Professional and Commercial operator hosts give out good annual revenue but a low performance score and customer retention.
- High Minimum Night policies lead to low occupancy and customer retention, leading to lower revenue.
- Budget and mid-range pricing categories maintain more stable demand levels and occupancy rate.
- Higher pricing does not guarantee higher occupancy.
- A small percentage of listings contribute disproportionately to total revenue.

STRATEGIC RECOMMENDATIONS

1. Optimize Pricing Strategy

Adopt dynamic pricing within budget and mid-range segments to maintain stable occupancy while maximizing revenue. Avoid excessive premium pricing without strong demand support.

2. Focus on High-Revenue Cities

Prioritize operational expansion and marketing efforts in top-performing metropolitan cities where demand and revenue concentration are highest. Expansion in low revenue generating areas is only advisable with strong existing revenue conversion and demand support.

3. Improve Host Performance

Provide performance optimization support for professional and commercially operating hosts, including pricing guidance and occupancy improvement strategies to enhance booking consistency. Conduct regular inspections for hosts based on their revenue and reviews.

4. Adjust Minimum Night Policies

Encourage flexible minimum night requirements to increase booking frequency and improve overall occupancy levels.

5. Monitor Performance via Dashboard

Continuously track KPIs using the interactive dashboard to identify demand shifts, revenue gaps, and underperforming segments in real time.

IMPACT ESTIMATION

Revenue Growth Potential

Optimizing pricing strategies and improving occupancy consistency across mid-range segments could increase overall revenue performance by leveraging stable demand patterns.
Focusing on high-revenue metropolitan markets may further enhance total revenue contribution.

Occupancy Improvement

Reducing restrictive minimum night requirements and improving professional host performance could increase booking frequency, leading to higher average occupancy rates across segments.

Operational Efficiency

Segment-based performance monitoring through the dashboard enables faster identification of underperforming listings and markets, reducing revenue leakage.

Risk Reduction

Data-driven pricing and segmentation strategies reduce the risk of overpricing, low occupancy, and inefficient market allocation.

LIMITATIONS

Data Constraints

The analysis is based on a sample of 10,000 listings and may not fully represent the entire short-term rental market.

Estimated Revenue Values

Revenue figures are calculated estimates based on occupancy and pricing assumptions, not actual realized transaction data.

Lack of Seasonal Insights

The dataset does not include monthly or seasonal breakdowns, limiting time-based performance analysis.

External Factors Not Included

Macroeconomic conditions, competitor pricing, local regulations, and tourism trends were not incorporated into the analysis.

Geographic Granularity

Neighborhood-level insights were limited due to missing or inconsistent location data.

FUTURE SCOPE

Seasonal & Time-Series Analysis

Incorporating monthly or seasonal booking data would enable forecasting of demand patterns and more accurate revenue prediction.

Dynamic Pricing Model

Developing a data-driven pricing optimization model could help adjust nightly prices based on demand trends, city performance, and occupancy behavior.

Customer & Review Analytics

Integrating review sentiment and customer ratings could provide deeper insights into factors influencing occupancy and host performance.

Neighborhood-Level Insights

Improved location data could enable granular neighborhood analysis for more precise market targeting.

Predictive Performance Modeling

Machine learning models could be developed to predict high-revenue listings and identify potential underperforming segments.

CONCLUSION

This project analysed 10,000 Airbnb listings across multiple U.S. cities to evaluate revenue performance, occupancy behaviour, pricing strategy, and host segmentation.

Through structured KPI development, exploratory analysis, and dashboard implementation, key revenue drivers were identified.

The findings highlight:-

- Significant revenue concentration in major metropolitan markets
- Stronger occupancy performance among individual hosts
- Importance of balanced pricing strategies.

Demand patterns are influenced by pricing categories, minimum night policies, and host management approach.

The interactive dashboard provides a centralized platform for monitoring performance, identifying revenue gaps, and supporting strategic decision-making.

Overall, this analysis demonstrates that revenue optimization requires a balanced approach across pricing, segmentation, geographic focus, and operational efficiency. Data-driven strategy remains critical for sustainable growth in the short-term rental market.

APPENDIX

Appendix A – Dataset Overview

- Original dataset size: ~226,000 records
- Sample used for analysis: 10,000 listings
- Total features: 17 variables
- Data types: Numeric and String

Appendix B – Key Feature Summary

Feature	Type	Distinct	Missing
id	Numeric	226,029	0
host_id	Numeric	130,425	0
neighbourhood_group	String	34	115,845
neighbourhood	String	1,450	0
room_type	String	4	0
price	Numeric	1,975	0
minimum_nights	Numeric	169	0
reviews_per_month	Numeric	1,242	48,602
availability_365	Numeric	366	0
city	String	28	0

Calculated Field Name	Description	Formula
occupancy_rate	For how much percentage of the year is the property booked?	(365-availability)/365
estimated_annual_revenue	How much is the annual revenue that the property earns?	price*(365 - availability_365)
revenue_per_available_day	How efficiently the listing converts its available days into revenue?	price*occupancy_rate
host_type	Segregated the host into 3 types based on number of listings he has	Range based segregation
host_performance_score	Performance score of the host based on the number of listings he has, reviews and occupancy rate	(no_of_reviews/host_listings)*occupancy_rate*reviews_per_month
neighbourhood_average_price	average price of properties in the same neighbourhood	average(price_in_same_neighbourhood)
demand_score	estimate of demand of the property based of reviews and price	z-score(number_of_reviews) + z-score(reviews_per_month) + z-score(price)
days_since_last_review	how many days since the last review has been given	today - last_review
city_average_price	average price of listing in the same city	average(price_in_same_city)
price_index	index of price between the city average price and the price of the listing	price/city_average_price
price_category	Segregated the price into three buckets	Range based Segregation
minimum_nights_bucket	Segregated the minimum nights into range values	Range based Segregation

Appendix C – Data Processing Notes

• Original dataset received in ARFF format
 • Converted to CSV for analysis
 • Randomly sampled from ~200,000 rows to 10,000 rows
 • Missing neighbourhood_group treated as "Not Specified"
 • minimum_nights restricted to valid range (1–90)
 • Dataset processed in Google Sheets
 • Final dataset used for analysis: 10,000 cleaned records

Contribution Matrix

Team Member	Dataset and Sourcing	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Yashi Agrawal	Converted arff to csv using python script	Curation of KPI list and pivot table list	Charts & pivot tables	Few changes with content		Creation and Analysis
Archit Gosain			Colour theme and design idea	Complete Design and creation		Dashboard design and report
Shivam Dixit	Randomisation and decrement of rows	Calculated fields & pivot tables	Charts & pivot tables			Data cleaning & Dashboard
Prince Singh	Randomisation and decrement of rows	Calculated fields & pivot tables	Charts & pivot tables			Data cleaning & Dashboard
Kavya Mukhija	Identification of outliers and missing values		Slicer connection		Complete Design and creation	Data cleaning & PPT
Samay Samrat	Data Cleaning		Charts & Pivot tables			Data cleaning & Dashboard
Pratyush Parida			Connection of pivot tables with slicers			Dashboard Interactivity

Contribution Matrix

Declaration

We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Signature Block

Name 1: Yashi Agrawal

Name 2: Archit Gosain

Name 3: Shivam Dixit

Name 4: Prince Singh

Name 5: Kavya Mukhija

Name 6: Samay Samrat

Name 7: Pratyush Parida