

Airbnb New York Market Analysis

Sector: *Short-Term Rental / Hospitality Analytics*

Team Members:

1. Soumya Tiwari (2401020111)
2. Avneet Singh (2401010117)
3. Kushagra Bhardwaj (2401010247)
4. Vaibhav Singh (2401020074)
5. Alisha Gupta (2401010057)
6. Dhruv Kumar (2401010152)

Executive Summary

Problem

Airbnb hosts struggle to determine which neighbourhoods, room types, and pricing strategies maximise revenue and occupancy while maintaining strong demand and ratings.

Approach

We cleaned and transformed a New York Airbnb dataset using Google Sheets, engineered key KPIs (Estimated Revenue, Occupancy Rate, Property Age), and built an interactive dashboard with pivots, slicers, and advanced visualisations.

Key Insights

- Brooklyn and Manhattan dominate revenue generation.
- Entire homes generate the highest revenue.
- Mid-range pricing maximises occupancy and demand.
- New properties do not always outperform older ones.
- Host verification has a limited impact on revenue.

Key Recommendations

- Focus investment in Brooklyn & Manhattan.
- Optimise pricing in the mid-range zone.
- Encourage flexible cancellation for demand boost.
- Use an occupancy-based pricing strategy.

Sector & Business Context

Sector Overview

The short-term rental industry has grown significantly, with platforms like Airbnb transforming urban hospitality markets.

Current Challenges

- Price optimization
- Demand volatility
- High competition
- Occupancy management
- Quality consistency

Why This Problem

Hosts lack data-driven guidance on where and how to invest.

Problem Statement & Objectives

Problem Statement

Which neighbourhoods, room types, and pricing strategies should Airbnb hosts focus on to maximise revenue and occupancy while maintaining strong demand and ratings?

Objectives

- Identify top-performing neighbourhoods
- Analyse pricing vs demand relationship
- Evaluate property lifecycle performance
- Measure occupancy efficiency

- Provide actionable investment insights

Data Description

Dataset

Airbnb New York Open Data (10,000 listings)

Source: Kaggle

Link: <https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata>

Original Dataset Columns

The original dataset contained the following raw columns:

- ID
- NAME
- HOST ID
- HOST_ID_VERIFICATION
- host name
- neighbourhood group
- Neighbourhood
- lat
- long
- Country
- country code
- instant_bookable
- cancellation policy
- room type
- Construction year
- Price
- service fee
- minimum nights
- number of reviews

- last_review
- reviews per month
- review rate number
- calculated host listings count
- availability 365
- house_rules

The dataset contained inconsistencies such as:

- Missing values
- Inconsistent text formatting
- Blank categorical fields
- Invalid availability values
- Missing ratings despite reviews

Cleaned Dataset Columns

After cleaning and transformation, the dataset includes:

Column Name	Description	Data Type
ID	Unique listing identifier	Numeric
Name	Property name	Text
Host Id	Unique host identifier	Numeric
Host Id Verified	Boolean (TRUE/FALSE)	Boolean

Host Name	Cleaned host name	Text
Neighbourhood Group	Borough (Manhattan, Brooklyn, etc.)	Categorical
Neighbourhood	Specific locality	Categorical
Latitude	Geographic coordinate	Numeric
Longitude	Geographic coordinate	Numeric
Country	Country name	Text
Instant Bookable	Whether booking is immediate	Boolean
Cancellation Policy	Flexible/Moderate/Strict	Categorical
Room Type	Entire home/Private room/Shared room	Categorical
Number of Reviews	Total lifetime reviews	Numeric

Reviews (per month)	Monthly review rate	Numeric
Average Ratings	Average customer rating (1–5)	Numeric
Price in Dollars	Listing price	Numeric
Service Fee in Dollars	Platform service fee	Numeric
Minimum Nights	Minimum booking duration	Numeric
Availability 365	Available days per year	Numeric
Construction Year	Year property built	Numeric /Text
Property Age	Derived lifecycle category	Categorical
Estimated Annual Revenue	Derived revenue proxy	Numeric
Occupancy Rate	Derived utilization metric	Percent age

Derived Variables (Feature Engineering)

Several important metrics were engineered:

1. **Estimated Annual Revenue:** Used as revenue proxy.

Formula:

- $\text{Price} \times \text{Reviews per Month} \times 12$

2. **Occupancy Rate:** Indicates booking efficiency.

Formula:

- $(\text{Availability 365}) / 365$

3. **Property Age Category**

Grouped based on construction year into:

- New Property
- Mid Age
- Old Property

Data Cleaning Summary

S	Cleaning Action	Column(s) Affected	Method Used	Business Justification
1	Filled missing host verification	Host Id Verified	Replaced NULL with FALSE	Absence of verification logically treated as unverified

2	Replaced missing price values	Price	Median Imputation	Median prevents distortion from extreme price outliers
3	Filled missing reviews per month	Reviews (per month)	Replaced NULL with 0	No reviews implies no booking activity
4	Corrected invalid availability values	Availability 365	Capped values between 0 and 365	Availability cannot be negative or exceed 365 days
5	Imputed missing ratings (when reviews existed)	Average Ratings	Median rating of same neighbourhood	Maintains local quality consistency and avoids bias
6	Standardized text formatting	Host Name, Cancellation Policy, Neighbourhood	Applied PROPER() and TRIM()	Ensures uniform formatting for pivot analysis
7	Removed irrelevant or inconsistent columns	Country Code, House Rules, Last Review, etc.	Dropped unused columns	Removed non-analytical fields to improve clarity and performance

KPI & Metric Framework

KPI Name	Formula Used	Type of Metric	Business Meaning	Why It Matters for the Problem Statement
Properties Listed	COUNT(ID)	Descriptive	Total number of active listings	Indicates market size and competition intensity
Average Price	AVERAGE(Price in Dollars)	Pricing Metric	Average listing price across NYC	Helps evaluate pricing strategy effectiveness
Average Occupancy	(Availability 365) / 365 (Average)	Utilization Metric	Percentage of days properties are booked	Measures booking efficiency and demand strength
Estimated Revenue	Price × Reviews per Month × 12 (SUM)	Revenue Metric	Annual revenue proxy	Indicates total earning potential of the market
Average Rating	AVERAGE(Average Ratings)	Quality Metric	Overall customer satisfaction score	Reflects service quality and brand positioning

Exploratory Data Analysis (EDA)

This section analyzes patterns, trends, distributions, and relationships observed in the dashboard.

Overall Market Overview

From the KPI cards:

- **Total Listings:** 9,999
- **Average Price:** \$624.75
- **Average Occupancy:** 50.30%
- **Total Estimated Revenue:** \$63.78M
- **Average Rating:** 3

Key Observations:

- The NYC Airbnb market is large and competitive.
- Average occupancy at ~50% indicates moderate utilization.
- Revenue generation is highly concentrated in certain boroughs.

Revenue by Neighbourhood Group

- **Brooklyn generates the highest revenue**, closely followed by Manhattan.
- Queens contributes moderate revenue.
- Bronx and Staten Island contribute minimal revenue.

Interpretation:

Revenue is geographically concentrated, suggesting:

- High tourist demand in Brooklyn and Manhattan.
- Lower demand or lower listing density in outer boroughs.

Revenue by Room Type

Observations:

- Entire home/apartment generates highest revenue.
- Private rooms generate slightly lower but still substantial revenue.
- Shared rooms contribute minimal revenue.

Interpretation:

Entire home listings are more profitable due to:

- Higher pricing power
- Longer stays
- Higher booking value

Shared rooms are low-margin and low-demand.

Price vs Demand Relationship

(From Occupancy scatter plot & demand distribution)

Observations:

- As price increases, occupancy tends to decrease.
- Demand clusters in mid-price range.
- Extremely high prices show lower booking frequency.

Interpretation:

There exists an optimal pricing zone where:

- Price is competitive
- Occupancy remains strong
- Revenue is maximized

Overpricing reduces booking efficiency.

Property Age & Revenue

(From Neighbourhood-wise Revenue by Age Category)

Observations:

- Mid-age properties contribute highest revenue.
- New properties do not consistently outperform older properties.
- Property age has moderate but not dominant impact.

Interpretation:

Revenue depends more on:

- Location
- Pricing
- Demand

rather than the construction year alone.

Host Verification

(From Host Verification donut chart)

Observations:

- Verified: ~49.4%
- Unverified: ~50.6%

Revenue distribution does not strongly depend on verification.

Verification status alone is not a strong revenue driver.

Year-wise Trend Analysis

(From Year Wise Trend chart)

Observations:

- Number of properties fluctuates across years.
- Average price remains relatively stable.
- No extreme price volatility observed.

Interpretation:

Market supply changes do not drastically impact average pricing.

Diagnostic Analysis (Why Did It Happen?)

Why Brooklyn & Manhattan Lead Revenue

Reasons:

- Higher tourist footfall
- Business hubs
- Strong neighborhood brand value

- Higher willingness to pay

High demand combined with pricing power drives revenue.

Why Entire Homes Perform Better

- Privacy preference
- Family/group travelers
- Higher nightly rates
- Longer booking duration

This creates higher total revenue per listing.

Why High Price Reduces Occupancy

As price increases:

- Price elasticity increases
- Customers shift to substitutes
- Competitive listings capture demand

Demand is price-sensitive beyond a threshold.

Why Ratings Have Moderate Impact

Average rating is around 3.

Moderate ratings indicate:

- Market is competitive
- Quality differentiation is limited

- Pricing influences booking more than ratings alone
-

Predictive Insights (What Will Likely Happen?)

If Price Increases Excessively

- Occupancy will decline
 - Revenue may stagnate
-

If Hosts Focus on Brooklyn & Manhattan

- Higher revenue probability
 - Stronger demand stability
-

If Mid-Range Pricing Strategy Is Adopted

- Higher occupancy
 - Balanced revenue
 - Better long-term sustainability
-

If Flexible Cancellation Is Encouraged

- Increased booking confidence
 - Higher demand
-

Prescriptive Recommendations

1. Focus on High-Revenue Boroughs

Hosts should prioritize listings in **Brooklyn and Manhattan**, as these boroughs generate the majority of total revenue. High tourist demand, strong brand value, and greater pricing power make these locations more profitable and stable for investment.

2. Promote Entire Home Listings

Entire home listings generate the strongest revenue compared to private or shared rooms. Hosts should consider converting private or shared accommodations into full-unit offerings where feasible to increase booking value and overall earnings potential.

3. Adopt a Mid-Range Pricing Strategy

Extreme high pricing reduces occupancy and demand. Hosts should align pricing competitively with the borough average and continuously monitor occupancy performance to maintain a balance between revenue and booking frequency.

4. Improve Occupancy Rather Than Just Increasing Price

Since occupancy strongly influences total revenue, hosts should focus on improving booking efficiency. This can be achieved by reducing minimum night requirements, offering flexible cancellation policies, and enabling instant booking to enhance customer confidence and demand.

