

New York Airbnb EDA Project - 15 Point Summary

1. Dataset Overview: The dataset includes Airbnb listings in NYC with details like price, location, room type, and availability.
2. Initial Inspection: Used head, info, shape, and describe to understand structure and summary statistics.
3. Data Cleaning: Removed missing values, 12 duplicate rows, and converted ID columns to object type.
4. Price Filtering: Listings with prices above \$1500 were excluded to remove outliers.
5. Price Distribution: Most listings priced under \$400; visualized with boxplot and histogram.
6. Availability Analysis: Identified patterns in availability - low (0-50), medium (100-150), high (300-350) days.
7. Neighborhood Group Analysis: Manhattan is the most expensive, followed by Brooklyn.
8. Feature Engineering: Created 'price_per_bed' to normalize price by bed count.
9. Price Per Bed by Neighborhood: Manhattan had the highest price per bed.
10. Room Type & Neighborhood Impact: Hotel rooms and entire homes in Manhattan are the priciest; shared rooms are cheapest.
11. Price vs. Number of Reviews: Most listings under \$400 and 500 reviews; no strong correlation found.
12. Market Insights: Brooklyn and Manhattan are the busiest markets; others are smaller and cheaper.
13. Visualization Tools Used: Used matplotlib and seaborn for various plots.
14. Pair Plot Analysis: Explored relationships between price, nights, availability, and reviews.
15. Overall Insights: Provided clarity on price trends, neighborhood differences, and market activity in NYC Airbnb rentals.