Airbnb Data Analysis — Findings & Recommendations

Executive Summary

- Objective: Analyze Airbnb listings to identify pricing patterns, location insights, room-type effects, and sentiment from reviews.
- Data: listings.csv (primary). occupancy estimated from availability_365 when available.
- Deliverables: Cleaned data, EDA visualizations, interactive Streamlit dashboard, business recommendations, slide deck.

Key Findings

- Top price drivers: Property type (Entire homes), neighbourhood, and number_of_reviews (proxy for popularity).
- Price distribution: Wide long-tail; median price significantly lower than the mean due to high-end outliers.
- Location insights: A small set of neighbourhoods account for a large share of listings and revenue.
- Room types: Entire homes command highest average price; private rooms are most numerous and budget-friendly.
- Availability: Higher-priced listings often show lower availability (booked more), where availability_365 exists.
- Sentiment: Reviews generally positive (replace with computed average). Negative sentiment keywords often relate to cleanliness and check-in experience.

Recommended Visualizations (to include)

- Price distribution histogram (log-scale recommended for long-tail prices).
- Boxplots: Price by neighborhood and price by room type.
- Scatter: Price vs. number_of_reviews (size by review_scores_rating).
- Map: Geo distribution of listings colored by price or occupancy proxy.
- Heatmap: Correlation matrix of numerical features.
- Sentiment charts: Review sentiment distribution and average sentiment by neighbourhood/room type.

Business Recommendations — Hosts (left) & Airbnb (right)

- Optimize listing photos: Listings with professional photos attract higher prices and more bookings.
- Dynamic pricing: Adjust prices seasonally and by occupancy trends; experiment with promotional rates in low season.
- Encourage reviews: Offer small gestures to guests to increase review count — higher review counts improve trust and bookings.
- Improve guest experience: Address common negative feedback (cleanliness, check-in) to boost sentiment and ratings.
- Listing enhancement: Add popular amenities (WiFi, essentials, self-check-in) to improve competitiveness.

- Enhance pricing algorithm: Use ML models including location, property features, reviews, seasonality; integrate host-level behavior.
- Market insights tools: Provide hosts with neighborhood demand heatmaps and suggested price ranges.
- Quality controls: Identify low-rated clusters and offer targeted host training or support.
- Fraud/Outlier detection: Flag extreme-price or low-quality listings to protect marketplace reputation.

Dashboard Features

- Interactive filters: neighborhood, room type, price range, min reviews.
- KPI cards: total listings, avg price, avg occupancy (proxy), avg rating.
- Map view with hover tooltips, histograms, scatter plots, and heatmaps.
- Export filtered data to CSV for further analysis.
- Optional ML: Price prediction and k-means clustering to group similar properties.

Next Steps (left) & Appendix (right)

- Replace placeholders with real computed values and charts from your dataset.
- Validate ML models using cross-validation and report RMSE / R^2.
- Add time-series analysis when calendar data is available to study seasonality.
- Deliver final notebook, Streamlit app, and a short 3–5 minute demo script for presentation.

- Data sources: Inside Airbnb (listings.csv, calendar.csv, reviews.csv).
- Assumptions: Occupancy estimated from availability_365 if calendar.csv missing.
- Contact: (add your name and email here).