



UrbanVista Analytics

From Raw Data to Real Insights: Storytelling Airbnb's Urban Footprints

Formative Assessment 1 (FA-1) |
Strategic Data Storyboard

Prepared for the Course Instructor & UrbanVista Stakeholders



Analyst Note:

This deck serves as the foundational design brief for the upcoming Tableau Dashboard (FA-2).

Operational Mission and Stakeholder Analysis



Project Purpose

The goal is to transform the raw 'AB_US_2023.csv' dataset into a coherent narrative. We are moving beyond simple data entry to clear, actionable storytelling. This process establishes the cleaning protocols and visual standards necessary for a robust interactive dashboard.



Who Needs This?

1. **The Assessor:** Evaluating technical rigour in data cleaning, EDA, and narrative flow.
2. **UrbanVista Stakeholders:** City Planners, Airbnb Representatives, and Tourism Researchers requiring accurate insights into urban accommodation density and pricing trends.

Inventory of Intelligence: The Dataset

Analysis built upon the 'AB_US_2023.csv' file, capturing the 2023 Airbnb landscape.



Economic
price
(Nightly cost)
availability_365
(Market capacity)



Product
room_type
(Privacy level)
minimum_nights
(Stay duration)



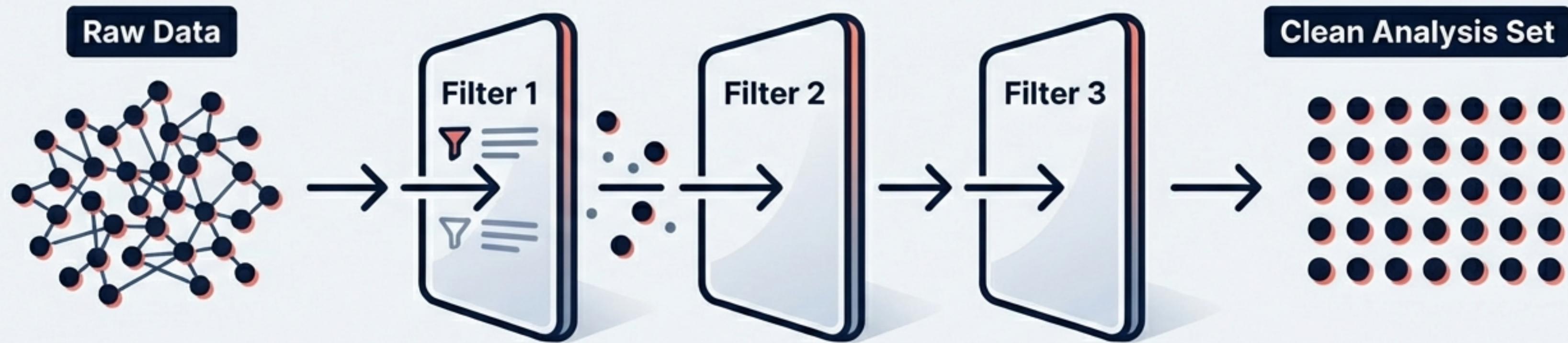
Geographic
neighbourhood
city
(Location clustering)



Engagement
number_of_reviews
reviews_per_month
last_review
(Guest activity)

Analyst Note: Understanding these variables is the prerequisite for isolating urban tourism hotspots.

Ensuring Integrity: Data Hygiene Strategy



Filter 1: Handling Nulls

Target: `reviews_per_month` & `last_review`.
Action: Imputation. Nulls imply zero activity. Filled with 0 or flagged as 'No Reviews' to preserve listing count.

Filter 2: De-Duplication

Target: Unique `id`.
Action: Identified and removed duplicate listing IDs to prevent double-counting inventory.

Filter 3: Formatting

Target: `price`.
Action: Converted from string/currency format to numeric integers for statistical calculation.

Quality Control and Anomaly Detection

Pricing Extremes



Observation: Extremely high prices detected.

Resolution: Listings flagged. While luxury assets exist, prices beyond statistical probability (e.g., >\$10,000/night for a shared room) are treated as input errors and excluded to normalise the mean.

Duration Constraints



Observation: High `minimum_nights` values.

Resolution: Listings requiring >365 days are re-categorised or excluded, as they represent long-term leases rather than the short-term tourism market UrbanVista analyses.

Analyst Note: Rigorous outlier management ensures our descriptive statistics reflect the 'true' market, not the errors.

The Market Pulse: Descriptive Statistics

Price (\$)	Availability (Days)	Review Activity
<p><u>Mean:</u> \$ [Calculated Mean]</p> <p><u>Median:</u> \$ [Calculated Median]</p> <p>→ Median is the preferred metric due to luxury skew.</p> <p>Range: Min \$10 - Max \$5,000</p>	<p>Avg. Availability: [Value] / 365</p> <p>Indicates market saturation and openness for new bookings.</p>	<p>Avg. Reviews per Listing: [Value]</p> <p>A proxy for listing popularity and guest turnover.</p>

Key Takeaway: These baselines provide the context required to interpret the distribution graphs that follow.

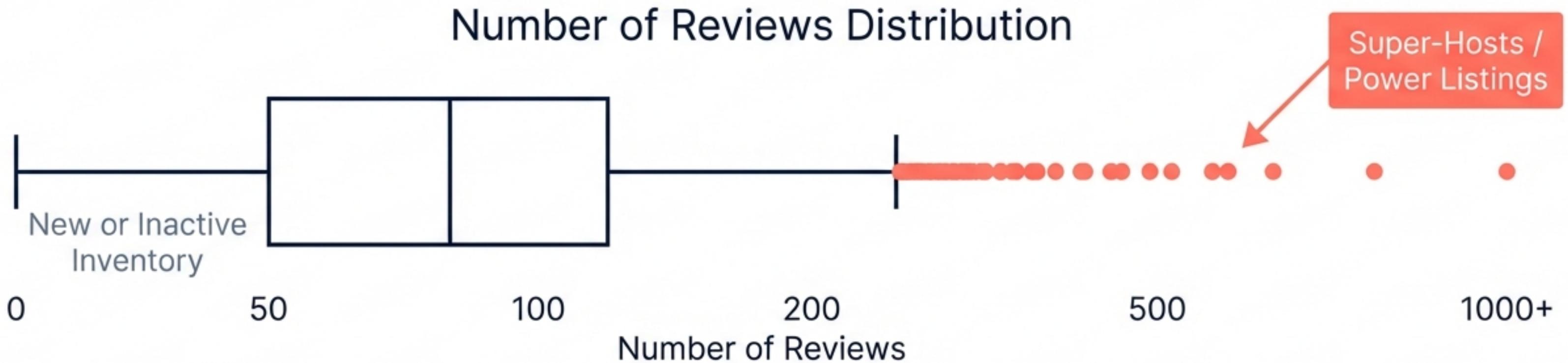
The Affordability Spectrum: Price Distribution



Analyst Note

The market is heavily weighted towards budget and mid-range accommodation. The long tail to the right represents luxury assets. This skew confirms that the median is a more reliable measure of "typical" cost than the mean.

Volume of Engagement: Identifying Market Activity



Interpretation: The boxplot reveals a significant disparity in engagement. While the median listing has moderate activity, there is a distinct tier of "Power Listings" with review counts far exceeding the norm.

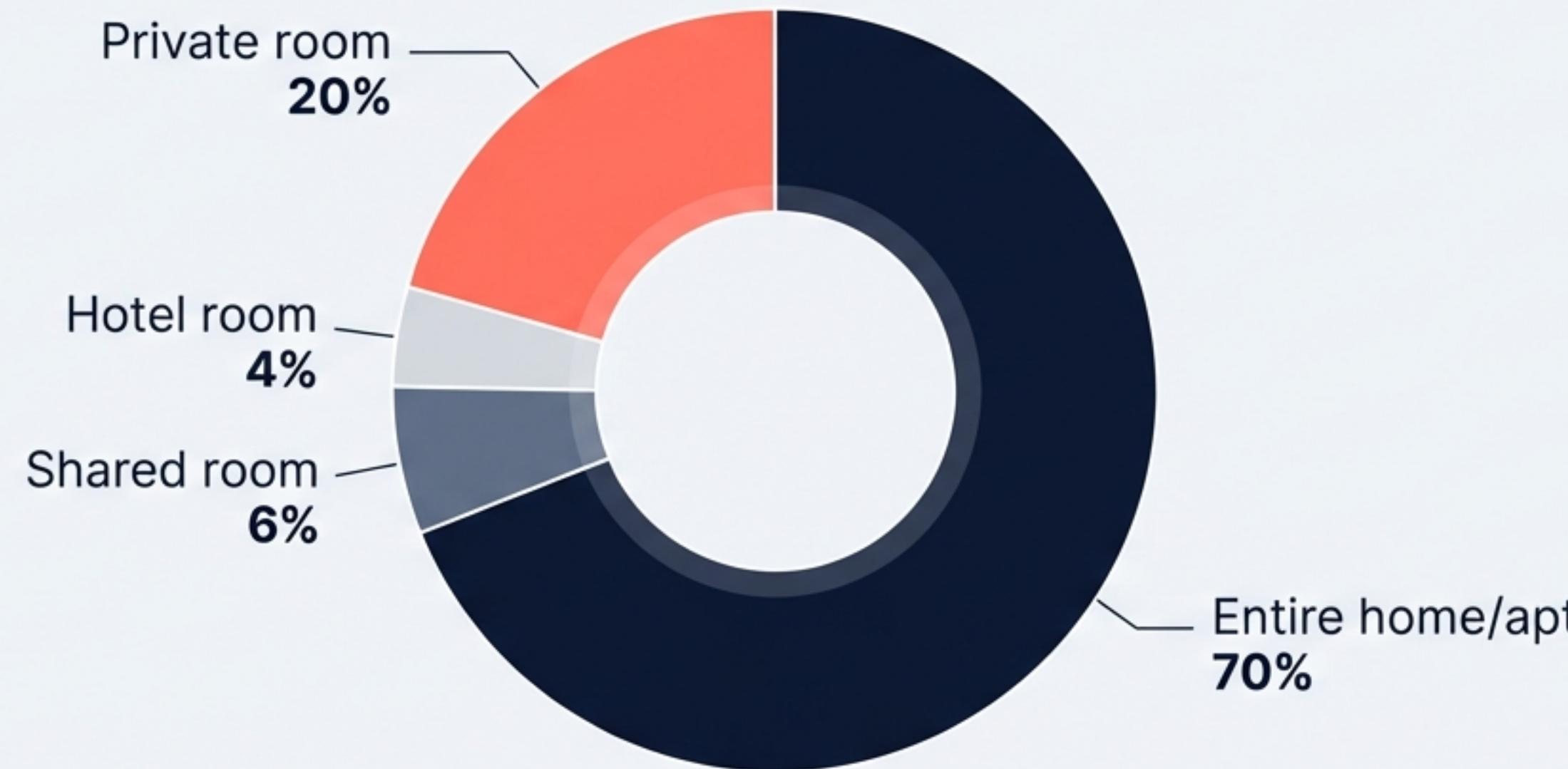
Urban Density: The Geographic Footprint



Key Observation

Listing volume is not evenly distributed. Specific urban centres dominate the dataset, representing the primary tourism “hotspots.” These high-volume cities require distinct sub-analysis compared to lower-volume regions.

The Accommodation Mix: Space and Privacy



Interpretation: The ratio of 'Entire homes' to 'Private rooms' is a critical metric for stakeholders. A dominance of Entire Homes suggests a **professionalised short-term rental market**, contrasting with the traditional 'homestay' economy.

Strategic Synthesis: Key Insights from 2023

Affordability



The market is accessible but skewed. Pricing strategies must account for the high density of budget options to remain competitive.

Engagement Gap



Success is unequal. A small percentage of listings capture a disproportionate amount of guest activity, suggesting a “winner-takes-all” dynamic.

Geographic Concentration



Tourism is hyper-localised. Resources and policy decisions should be focused on the top identified cities where volume is highest.

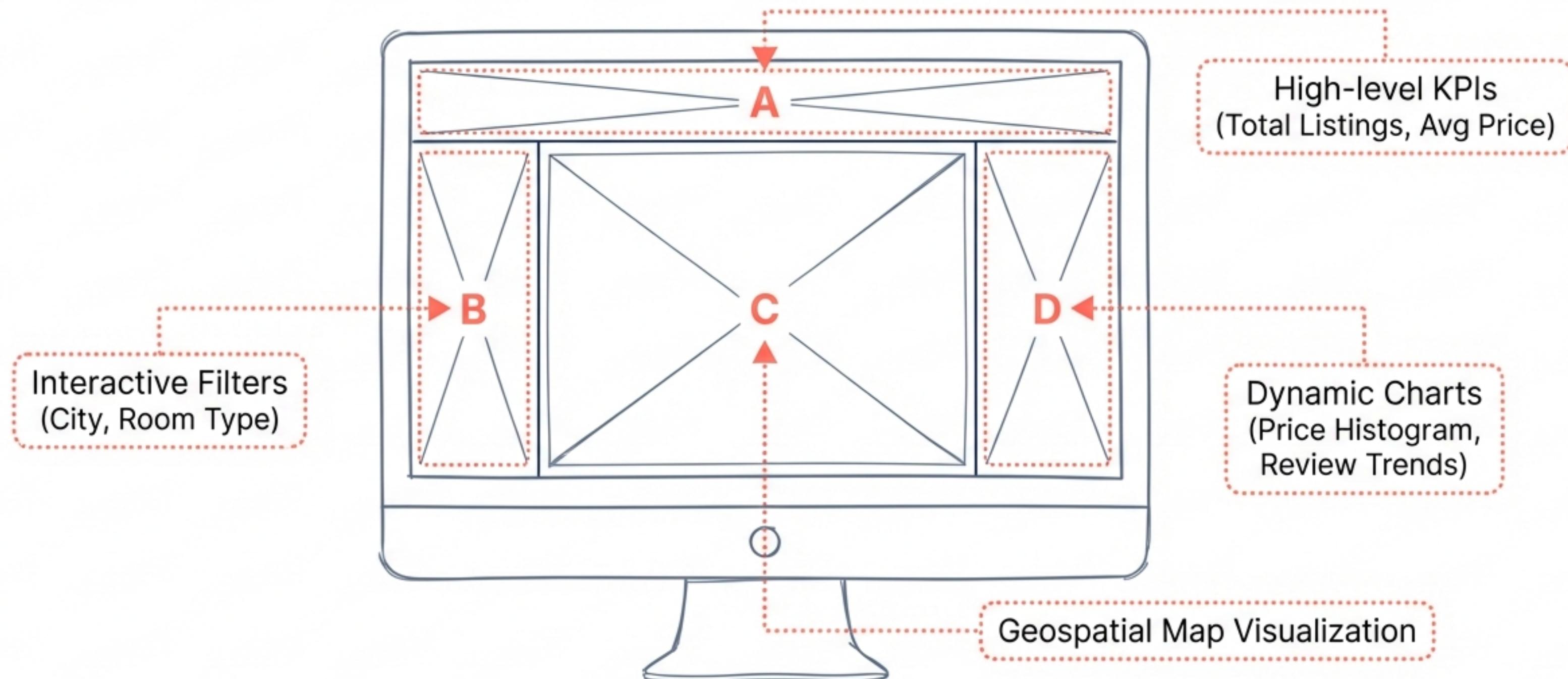
Inventory Impact



The distribution of Room Types indicates the primary use-case of the platform in 2023: dominated by professionalised entire-home rentals.

The Future Interface: Dashboard Prototype Architecture

Modern Consultancy Editorial



This layout prioritises the ‘Shneiderman Mantra’: Overview first (KPIs), zoom and filter (Selectors), then details-on-demand (Charts).

Reflection and Forward Momentum



The Transition: This analysis has successfully audited the data quality and established the core narrative arcs.

Next Steps (FA-2):

1. **From Static to Dynamic:** Moving from fixed charts to interactive filtering in Tableau.
2. **User Empowerment:** Allowing stakeholders to drill down into specific neighbourhoods or price brackets themselves.

Final Thought: Reliable data storytelling is not just about showing the numbers; it is about curating the path to discovery. The cleaning and EDA performed here ensure that the path is safe and accurate.

Appendix A: Glossary of Terms

EDA

: Exploratory Data Analysis – the process of investigating data to discover patterns.

Availability_365

: The number of days in the future that the listing is available for booking.

Room_Type

: The classification of the listing (Entire place, Private room, Shared room).

Ghost Listing

: A listing with zero reviews and low availability, potentially inactive.

Outlier

: A data point that differs significantly from other observations (e.g., \$10k price).

Appendix B: References and Acknowledgements

Data Source:

AB_US_2023.csv - Airbnb Listings Data (United States, 2023).

Tools Used:

Data cleaning and EDA performed using Python/Excel.

Visualisation conceptualisation for Tableau.

Declaration:

This presentation was created for the '**Art of Storytelling with Data**' Formative Assessment 1. All insights are derived directly from the provided source file.