

Milestone 3: Prototyping Design & Visual Encoding

Hüseyin Soykök - Group 24

Project: NYC Airbnb Market Analysis

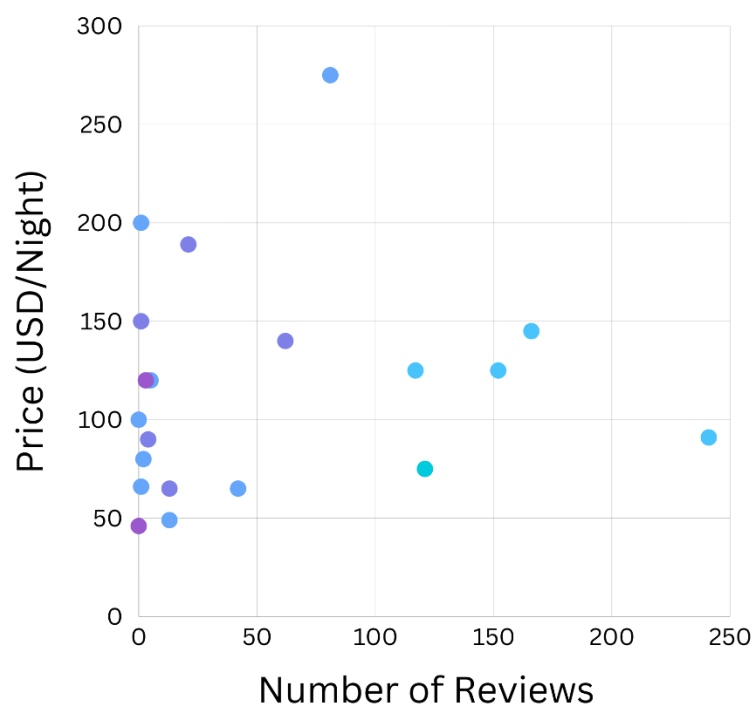
Dataset: New York City Airbnb Open Data (2019)

User Story 1: The Value-Seeking Traveler

This story focuses on Hüseyin, a budget-conscious traveller who wants to find the best "price-performance" accommodation rather than just the cheapest option. His goal is to filter for specific "value gems" listings that are affordable (\$100-\$150) yet highly vetted (>100 reviews). The design must enable the user to "Filter" the dataset based on these strict criteria and immediately "Locate" the geographic "Distribution" of the remaining targets on a map to identify where these value options are clustered.

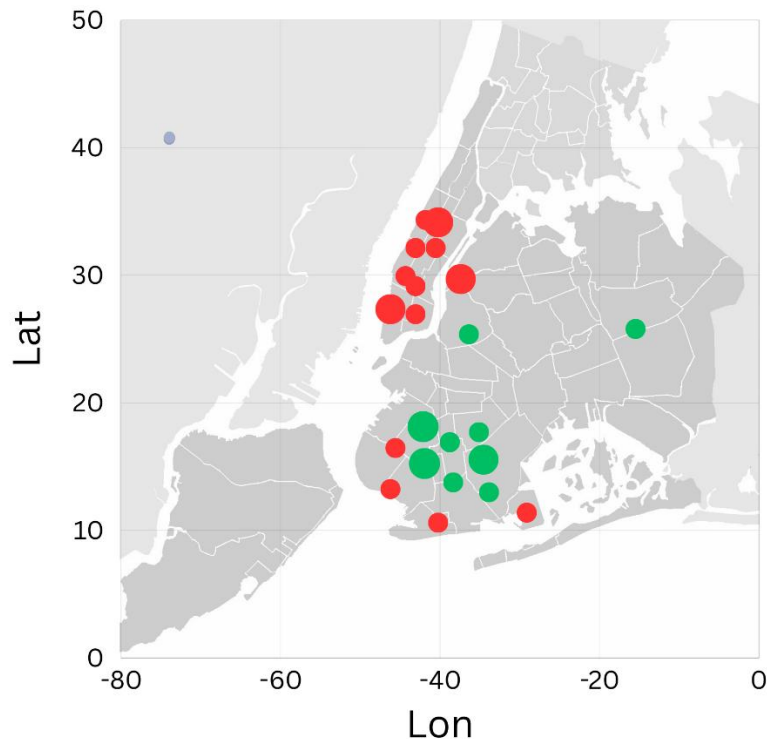
Alternative 1

A Scatterplot with "Price" on the Y-axis and "Number of Reviews" on the X-axis. Points are coloured by "Neighbourhood Group".



Alternative 2

A *Symbol Map (Geographical Dot Map)* where each listing is a dot. Colour represents "Price" (diverging colour scale) and size represents "Number of Reviews".



Baseline View: The baseline design focuses on the user's need to "Locate" listings geographically while "Filtering" for specific budget and popularity constraints. The primary view is a **Geographical Symbol Map** because the user explicitly asks, "Where are these... geographically clustered?" and is unfamiliar with NYC geography.

Interactions:

- **Dynamic Filtering:** Sliders for Price (\$100-\$150) and Number of Reviews (>100). Checkbox for Room Type ('Entire home/apt').
- **Geometric Zooming/Panning:** To inspect specific streets in Brooklyn or Queens.
- **Details-on-Demand:** Hovering over a dot reveals the specific Name, Price, and Review Score.

How multiple linked views can be incorporated into the design:

- **View 1 (Baseline): Geographical Symbol Map.** Encodes Latitude/Longitude (Position), Price (Colour Hue: Green for cheap, Red for expensive), and Review Count (Size).
- **View 2 (Baseline): Bar Chart** showing "Number of Listings per Neighbourhood".
- **Linking:** This is a "Filter + Correlate" task. When the user filters the map (View 1) by price and reviews, the Bar Chart (View 2) updates to show which neighbourhoods have the most *remaining* listings. Conversely, clicking a bar (e.g., "Williamsburg") filters the map to zoom into that area.

Storyboard:

- 1. Hüseyin opens the dashboard and sees all NYC listings.
- 2. He adjusts the "Price" slider to \$100-\$150 and "Reviews" slider to >100.
- 3. The Map (View 1) updates, showing clusters of available dots.
- 4. He notices a cluster in Queens and clicks the "Queens" bar in View 2.
- 5. The map zooms into Queens; he hovers over a large green dot to see the listing details.

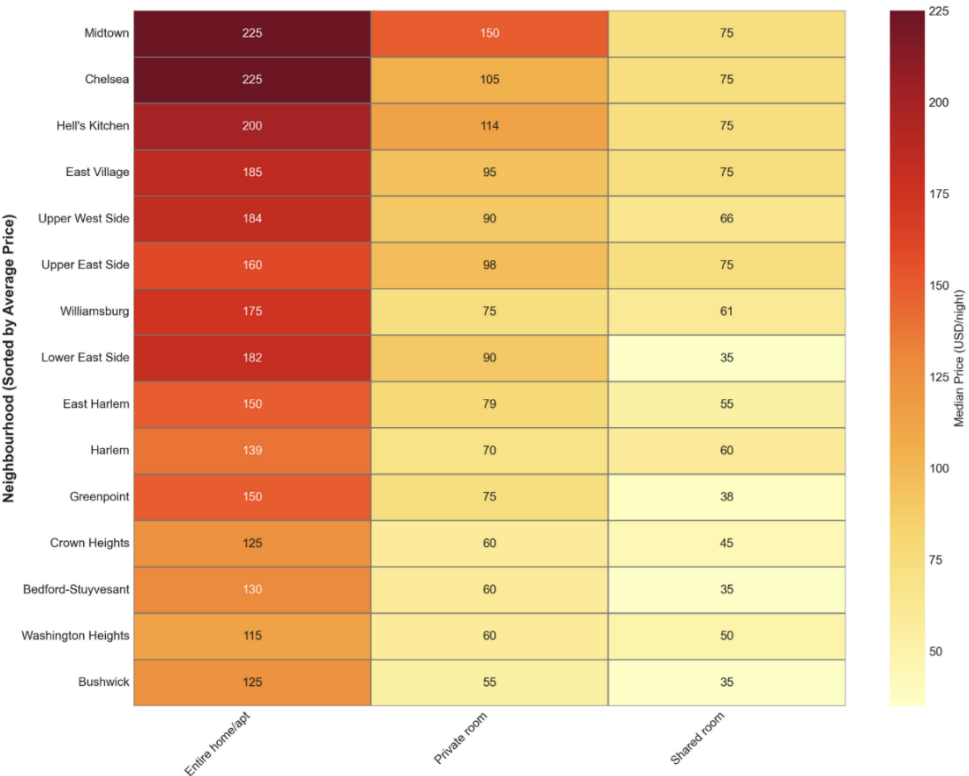
How the interface addresses the task: The interface directly addresses the "Discover value hotspots" task. The map encoding allows for immediate spatial recognition of clusters, while the linked filtering allows the user to remove noise (expensive or unpopular listings) to find the "value gems".

User Story 2: The New Host (Investor)

This story centers on David, a local resident and potential investor aiming to maximize his Return on Investment (ROI). Unlike a traveller, he is not interested in individual listing details but needs to "Summarize" the data to find typical (median) prices and availability trends. The design needs to facilitate a "Compare" task, allowing the user to evaluate different market segments (e.g., Manhattan vs. Brooklyn, or different room types) to "Discover" the most profitable area for investment.

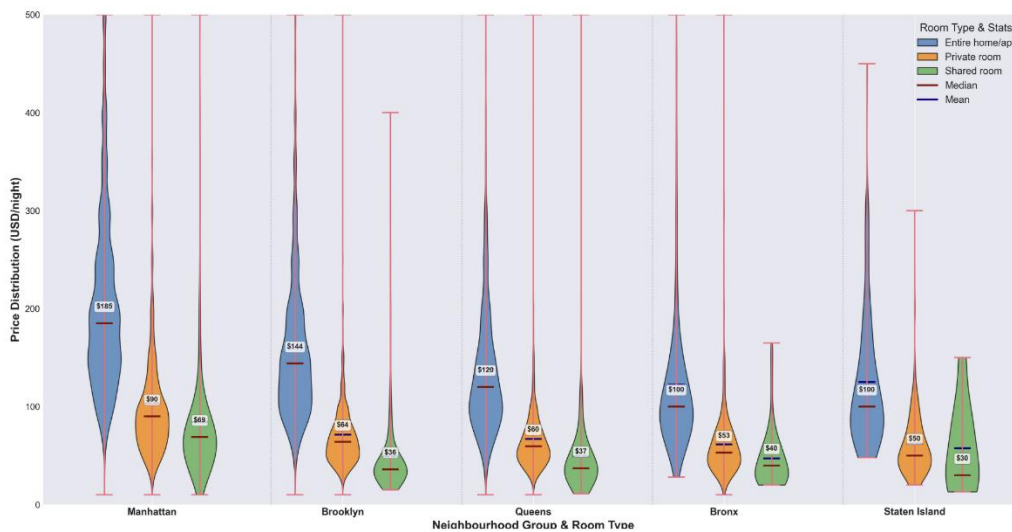
Alternative 1

A Heatmap Matrix. Rows are Neighbourhoods, Columns are Room Types. Cell colour intensity represents "Median Price".



Alternative 2

A Split Violin Plot. X-axis is Neighbourhood Group. Y-axis is Price. The 'violins' show the probability density of prices, split by Room Type (Left side: Entire home, Right side: Private room).



Baseline View: The baseline design is a **Violin Plot** because the user (David) needs to minimize investment risk. Unlike a simple Bar Chart that only shows the median (averages), a Violin Plot reveals the **shape of the distribution**. It allows David to see if prices in a neighbourhood are stable (short, fat shape) or highly volatile (long, thin shape), which is critical for making a safe investment decision.

Interactions:

- **Sorting:** Sort violins by "Median Price" or "IQR (Interquartile Range)" to find the most expensive or most stable markets.
- **Tooltip on Hover:** Since violin shapes can be abstract, hovering over a violin displays exact statistical summaries: Median, Q1, Q3, and Min/Max values.
- **Split/Group Toggle:** A control to switch between "Split Violin" (comparing two room types of side-by-side in one shape) or separate violins.

How multiple linked views can be incorporated into the design:

- **View 1 (Baseline): Violin Plot.** Shows Price Distribution density per Neighbourhood.
- **View 2 (Baseline): Choropleth Map (Regional Map).** Areas are coloured by "Investment Risk" (calculated from the standard deviation shown in the violin).
- **Linking:** Hovering over a specific "bulge" (price concentration) in the Violin Plot (View 1) highlights the corresponding neighbourhood on the Map (View 2). Clicking a violin filters the map to show only listings within that specific price range.

Storyboard:

1. David wants to invest safely. He looks at the Violin Plot (View 1).
2. He notices that Manhattan has a very long, thin violin shape, indicating extreme price volatility (high risk).
3. He looks at Queens and sees a "short and fat" violin shape around \$100, indicating a very stable and predictable market.
4. He hovers over the thickest part of the Queens violin to verify the exact price range (\$80-\$120).

5. He decides to invest in Queens for reliable returns and clicks the violin to see those specific properties on the map.

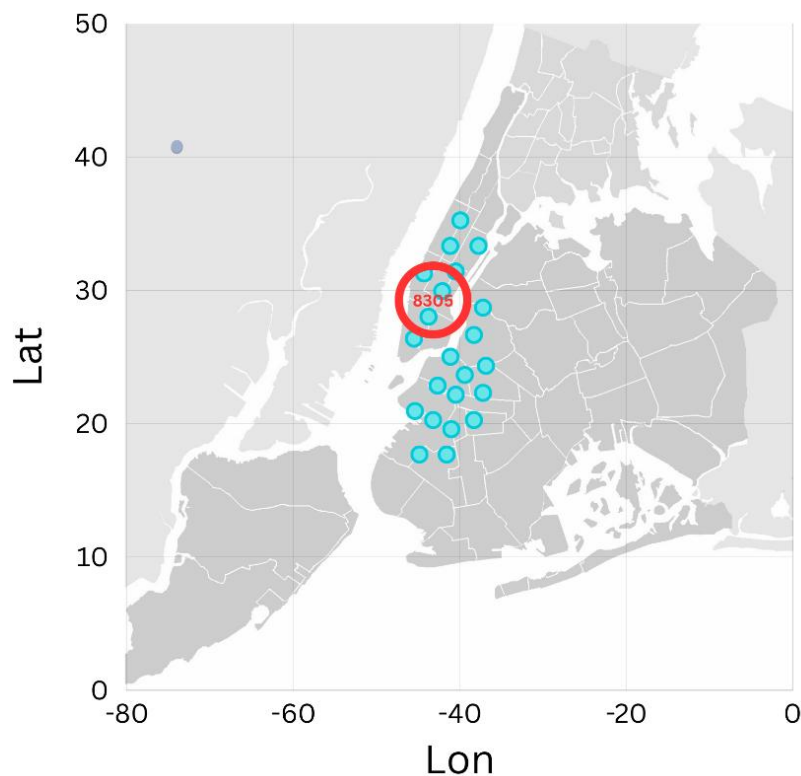
How the interface addresses the task: The interface directly addresses the "Compare" and "Discover" tasks for an investor. By using a Violin Plot instead of a bar chart, the design elevates the user's ability to analyse **market variance**. It prevents the user from being misled by a simple "average" price in unstable areas, directly supporting the goal of maximizing ROI while managing risk.

User Story 3: The City Regulator

Representing the perspective of Dr. Chen, an urban planner, this story aims to detect "commercial" operations or "ghost hotels" that affect the long-term housing market. The user needs to "Identify" listings that meet specific high-availability or multi-host criteria (availability > 300 days or host listings > 5). The visualization must focus on the spatial "Distribution" of these specific items to reveal "hotspots" where commercial activity is concentrated.

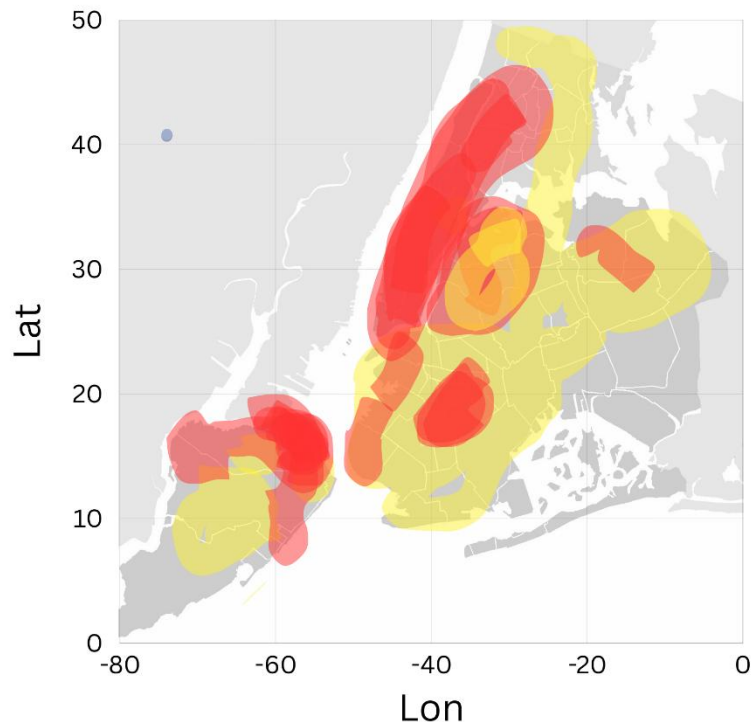
Alternative 1

A Dot Map using clustering. As you zoom out, individual points aggregate into numbered circles showing the count of listings in that area.



Alternative 2

A *Density Map (Heatmap)*. Smooth colour gradients (Yellow to Red) indicate the spatial concentration of listings matching the filter criteria.



Baseline View: Dr. Chen's task is to "Identify" and "Locate" concentrations of commercial activity. A **Density Map (Spatial Heatmap)** is the best baseline because she is looking for "hotspots" and general "zones" rather than specific addresses.

Interactions:

- **Pre-defined Filter Queries:** A "Commercial Filter" button that automatically applies the logic: $\text{availability_365} > 300$ OR $\text{calculated_host_listings_count} > 5$.
- **Opacity/Threshold Slider:** Adjusting the intensity of the heatmap to isolate only the most dense areas.

How multiple linked views can be incorporated into the design:

- **View 1 (Baseline): Density Map.** Visualizes the "Distribution" of filtered commercial listings.
- **View 2 (Baseline): Histogram of Host Listings Count.** Shows the distribution of how many listings hosts own.
- **Linking:** Brushing the "tail" of the histogram (hosts with >5 listings) updates the Density Map (View 1) to show *only* where those multi-listing hosts are operating.

Storyboard:

1. Dr. Chen wants to find "ghost hotels".
2. She clicks the "Commercial Listing Mode" button.
3. The Density Map (View 1) turns dark red in specific areas like Midtown and Williamsburg.

4. She brushes the Histogram (View 2) to select only hosts with 10+ listings.
5. The map updates to reveal a very tight cluster in a specific block, identifying a potential illegal hotel operation.

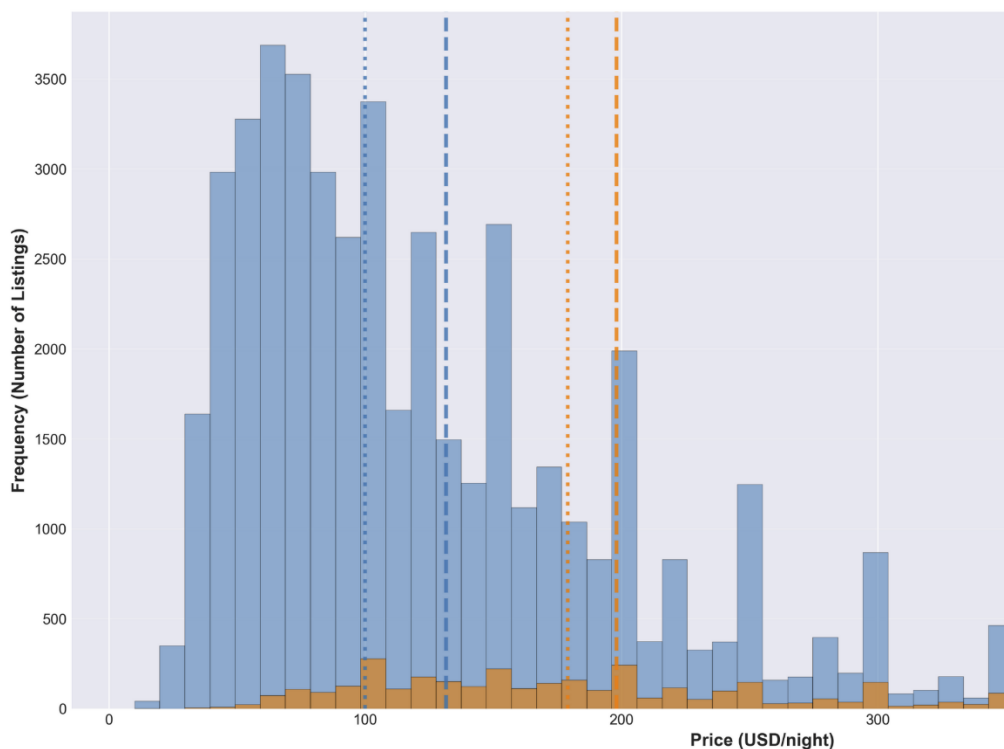
How the interface addresses the task: The interface facilitates the "Discover patterns" task. By using a density encoding, the user can instantly see "concentration" without being overwhelmed by 48,000 individual dots. The linked histogram allows her to investigate the *nature* of these commercial hosts (multi-owners vs. highly available units).

User Story 4: The Hotel Competitor

This story addresses the needs of Maria, a hotel manager in "Midtown," who is conducting a competitive analysis of the local Airbnb market. She is not interested in the whole city but specifically needs to "Explore" the competitive landscape within her own neighbourhood. The interface must allow her to "Filter" by location and "Summarize" the "Distribution" of prices and review counts to "Compare" her hotel's pricing strategy against direct local competitors.

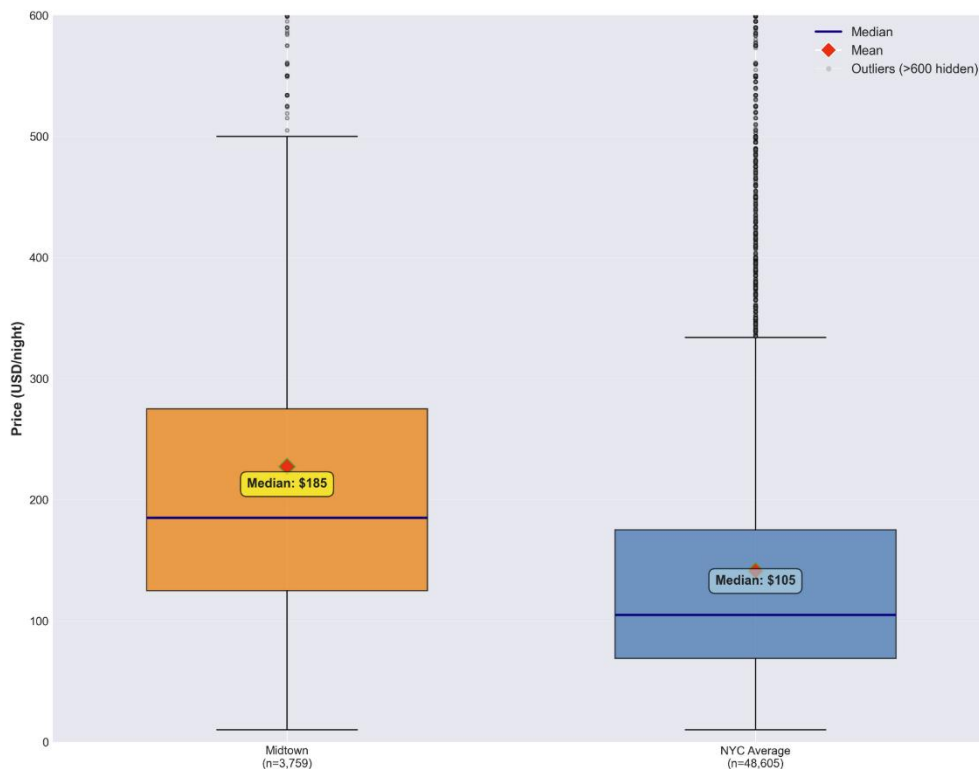
Alternative 1

A Histogram showing the frequency distribution of prices. Vertical lines indicate the mean and median.



Alternative 2

Box and Whisker Plots. One box for "Midtown" and one for "NYC Average". Y-axis represents Price.



Baseline View: Maria needs to understand the "range (distribution)" and "outliers" of prices in her specific neighbourhood compared to the city. A **Box Plot** is the most effective visual idiom here because it explicitly summarizes the min, max, median, and interquartile range (distribution spread).

Interactions:

- **Neighbourhood Dropdown Selection:** To filter the analysis specifically to 'Midtown' or other competitor areas.
- **Reference Line Input:** Maria can input her hotel's room price (e.g., \$250) to draw a horizontal line across the charts for direct comparison.

How multiple linked views can be incorporated into the design:

- **View 1 (Baseline): Side-by-Side Box Plots.** Comparison of Price Distribution: [Selected Neighbourhood] vs. [All NYC].
- **View 2 (Baseline): Scatterplot.** X-axis: Price, Y-axis: Number of Reviews.
- **Linking:** Filtering the Neighbourhood to "Midtown" in the dropdown updates both the Box Plot (View 1) and the Scatterplot (View 2). The Box Plot shows the price range, while the Scatterplot shows the "Popularity" (reviews) relative to that price.

Storyboard:

1. Maria selects "Midtown" from the dropdown menu.
2. The Box Plot (View 1) shows the median Airbnb price in Midtown is lower than her hotel, but the upper quartile ("Entire home") is similar.

3. She enters her hotel rate (\$250) to see a reference line.
4. She looks at the Scatterplot (View 2) to see if listings at \$250 have high review counts (high popularity).
5. She identifies that Airbnb's at her price point have fewer reviews, suggesting her hotel has a competitive advantage in service/trust.

How the interface addresses the task: The design solves the "Explore" and "Compare" tasks. The Box Plot statistically summarizes the "range" she requested, allowing her to position her pricing strategy against the quartiles of the Airbnb market. The linked scatterplot adds the necessary context of "popularity".