**Airbnb Open Data**

**DSCI 5360 Data Visualization for Analytics**

**Assignment 3**

**Name:** Indu Battula

**Student Id:** 11821110

**Reflection Report: Airbnb Open Data**

**Dataset Source:** <https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata>

**Tableau Public Link:** <https://public.tableau.com/app/profile/indu.battula/viz/Indu_BattulaAssignment_3/Dashboard1?publish=yes>

**Dataset Summary** For this project, I used the **Airbnb Open Data** dataset, which includes over 100,000 listings primarily located in **New York City**. It contains information on listing prices, availability, host details, reviews, location coordinates, and room types. This dataset was selected because of its real-world relevance and the variety of attributes available for spatial and analytical visualization.

**Purpose of Visualizations:** These visualizations explore key patterns in the Airbnb NYC dataset. A map highlights average prices across boroughs, while a boxplot compares price ranges by room type. The line chart shows the relationship between minimum stay and price. A bar chart identifies top neighbourhoods by listing count, and the treemap visualizes how listings are distributed across room types and boroughs. Together, they support spatial analysis, formatting clarity, and interactive dashboard design.

**Visualization Summary and Expanded Questions:**

**Visualization 1: Average Price by Neighbourhood Group (Map)**

* **Question Addressed:**  
  Which NYC boroughs have the highest average Airbnb prices?
* **Design:**
  + Filled map using neighbourhood group with colour encoding for AVG (price).
  + Geographic role assigned for spatial mapping.
* **Insight:**  
  Queens and Brooklyn have significantly higher average prices compared to outer boroughs like the Manhattan and Staten Island.

**Visualization 2: Price by Room Type (Box-and-Whisker Plot)**

* **Question Addressed:**  
  How does the price vary across different Airbnb room types?
* **Design:**  
  Box-and-whisker plot with room type on the x-axis and price on the y-axis, showing spread, median, and outliers.
* **Insight:**  
  Entire homes have the widest price range and highest outliers, while shared rooms are generally the most affordable with minimal price variation.

**Visualization 3: Minimum Nights vs Price (Line chart with Trend Line)**

* **Question Addressed:**  
  Is there a relationship between minimum stay requirements and price?
* **Design:**  
  Line chart with minimum nights vs price, filtered to remove outliers, and enhanced with a linear trend line.
* **Insight:**  
  Listings requiring longer minimum stays tend to have higher prices, especially for entire home rentals, indicating a targeted long-stay strategy.

**Visualization 4: Top Neighbourhoods by Number of Listings (Bar Chart)**

* **Question Addressed:**  
  Which NYC neighbourhoods have the highest concentration of Airbnb listings?
* **Design:**  
  Horizontal bar chart showing neighbourhood on y-axis and CountD(id) on x-axis, sorted in descending order.
* **Insight:**  
  Neighbourhoods like Bedford-Stuyvesant and Harlem have the most listings, revealing high Airbnb activity in these areas.

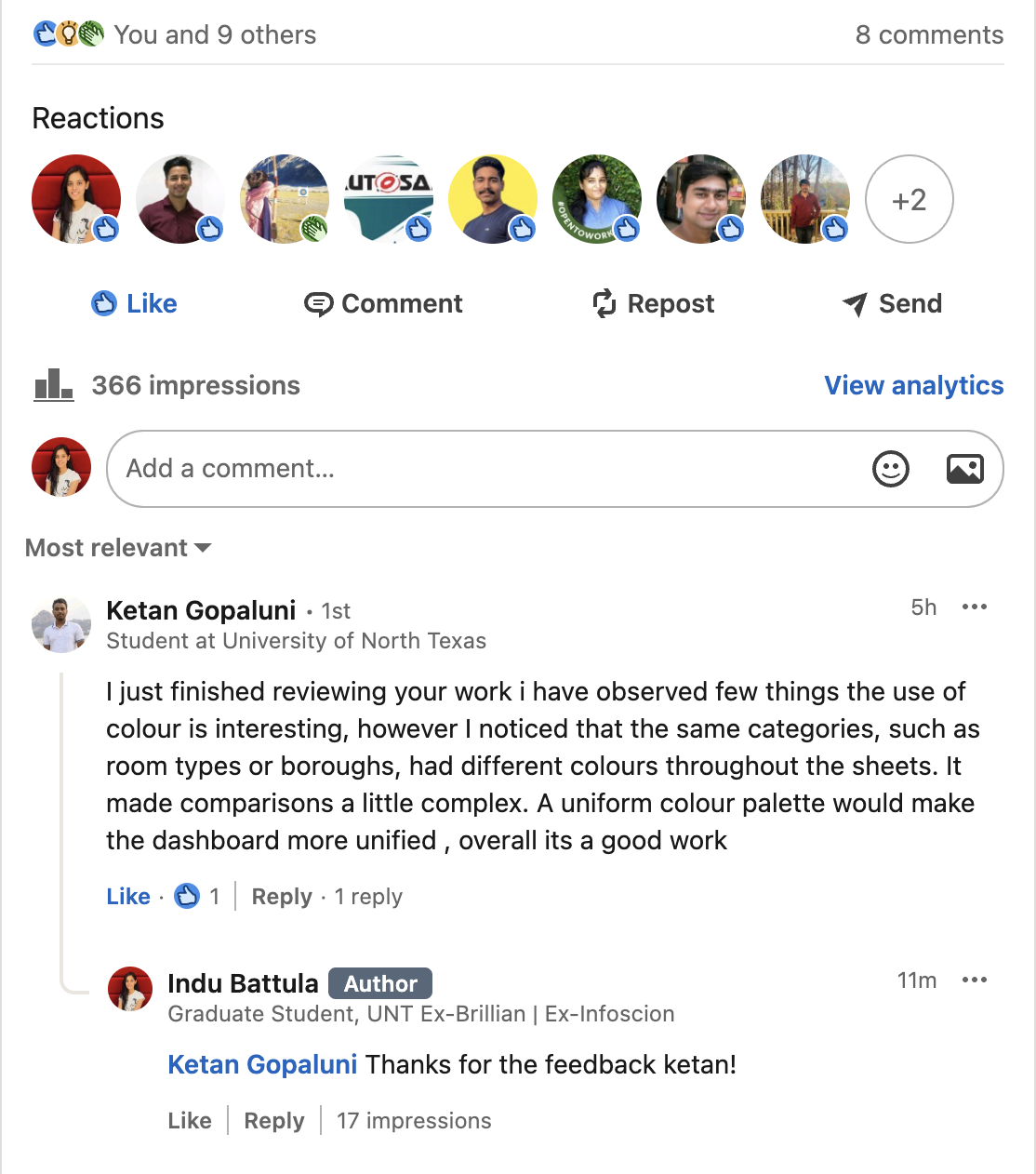
**Visualization 5: Listings by Room Type and Neighbourhood Group (Treemap)**

* **Question Addressed:**  
  How are Airbnb listings distributed by room type and borough?
* **Design:**  
  Treemap with room type as label, neighbourhood group as colour, and Count(id) as size.
* **Insight:**  
  Private rooms dominate in Brooklyn, while entire homes are more common in Manhattan—clearly showing market distribution patterns.

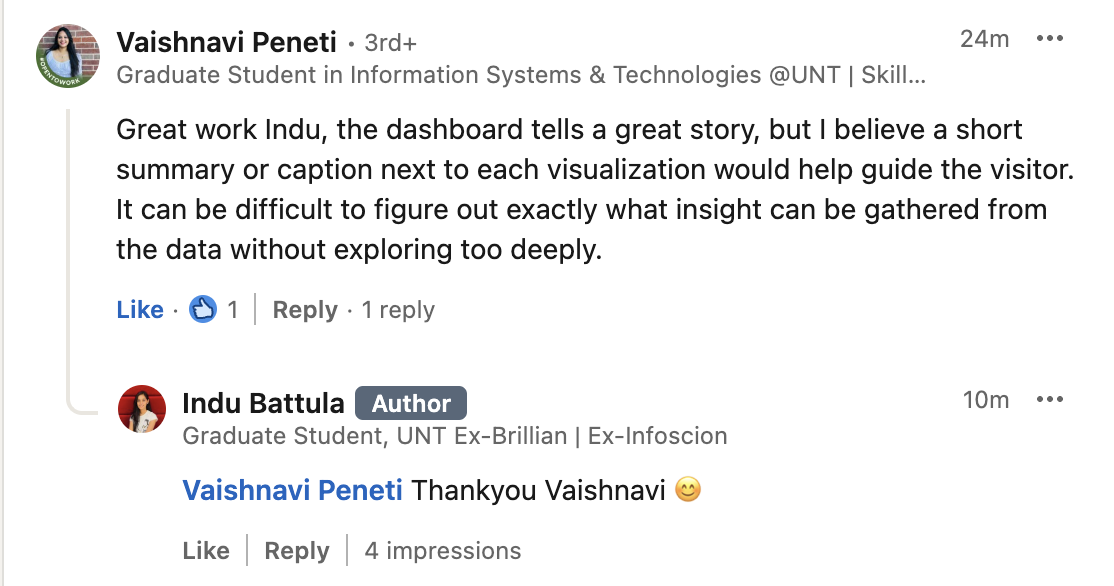
**Screenshots of Social Media Post & Feedback:**

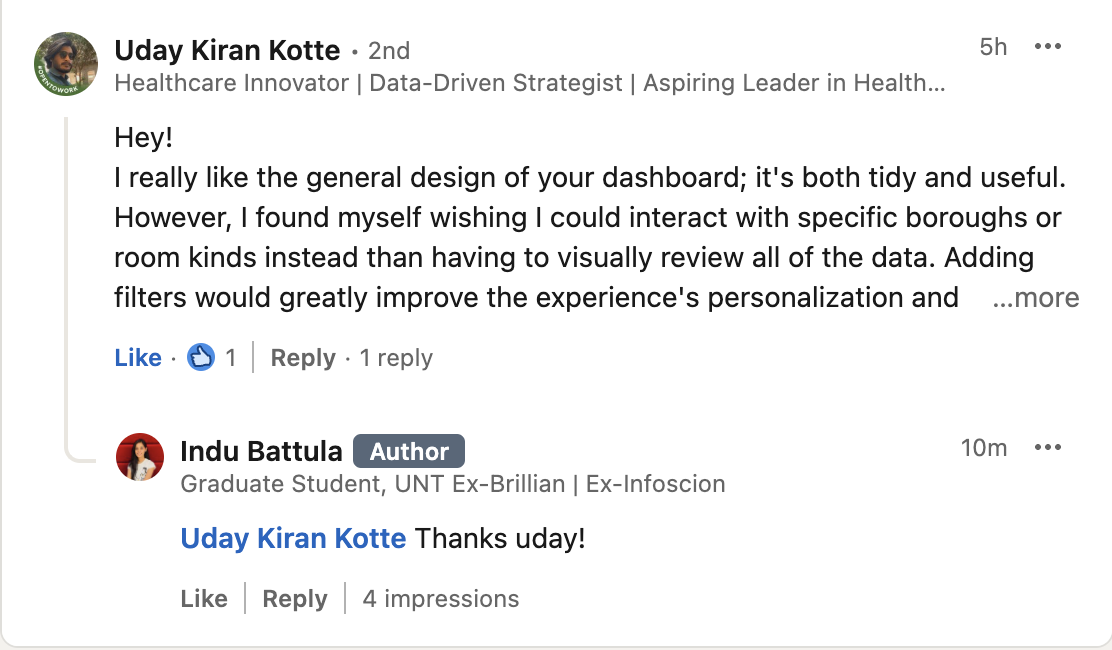
**A screenshot of a social media post

AI-generated content may be incorrect.**

****

****

****

****

**Social Media Feedback and Improvements:**

**Visualization 1: Average Price by Neighbourhood Group (Map)**

* **Feedback Received**:
  + Prices are labelled, but it would be even better if I could hover to see more details like reviews etc.,
* **Improvements Made**:
  + Enhanced the **tooltip content** to include additional context: average reviews per borough, average minimum nights per borough.
  + Made the map more interactive and informative for users who want to dig deeper.

**Visualization 2: Price by Room Type (Box-and-Whisker Plot)**

* **Feedback Received**:
  + The tooltips were too detailed and occasionally hard to read, especially when hovering over tightly packed visuals.
  + Recommended simplifying the content to improve clarity.
* **Improvements Made**:
  + Refined the tooltips to include only the most critical information.
  + Removed unnecessary text and aligned the values.

**Visualization 3: Minimum Nights vs Price (Line Chart with Trend Line)**

* **Feedback Received**:
  + It would be even more effective with interactive filters.
  + Hard to focus on a specific borough or room type without manually scanning through the entire visual.
* **Improvements Made**:
  + To address this, added dynamic filters for minimum nights and price, allowing users to explore data relevant to their interest with just a few clicks. This made the dashboard more user-friendly and personalized.

**Visualization 4: Top 10 Neighbourhoods by Number of Listings (Bar Chart)**

* **Feedback Received**:
  + It’s hard to instantly tell which neighbourhood has the most listings without manually reading through all the bars.
  + Highlight or label the top values to make them stand out.
* **Improvements Made**:
  + Added **data labels** to show the exact number of listings for each neighbourhood.
  + Also sorted the listings in descending order and filtered top 10 listings display.

**Visualization 5: Listings by Room Type and Neighbourhood Group (Treemap)**

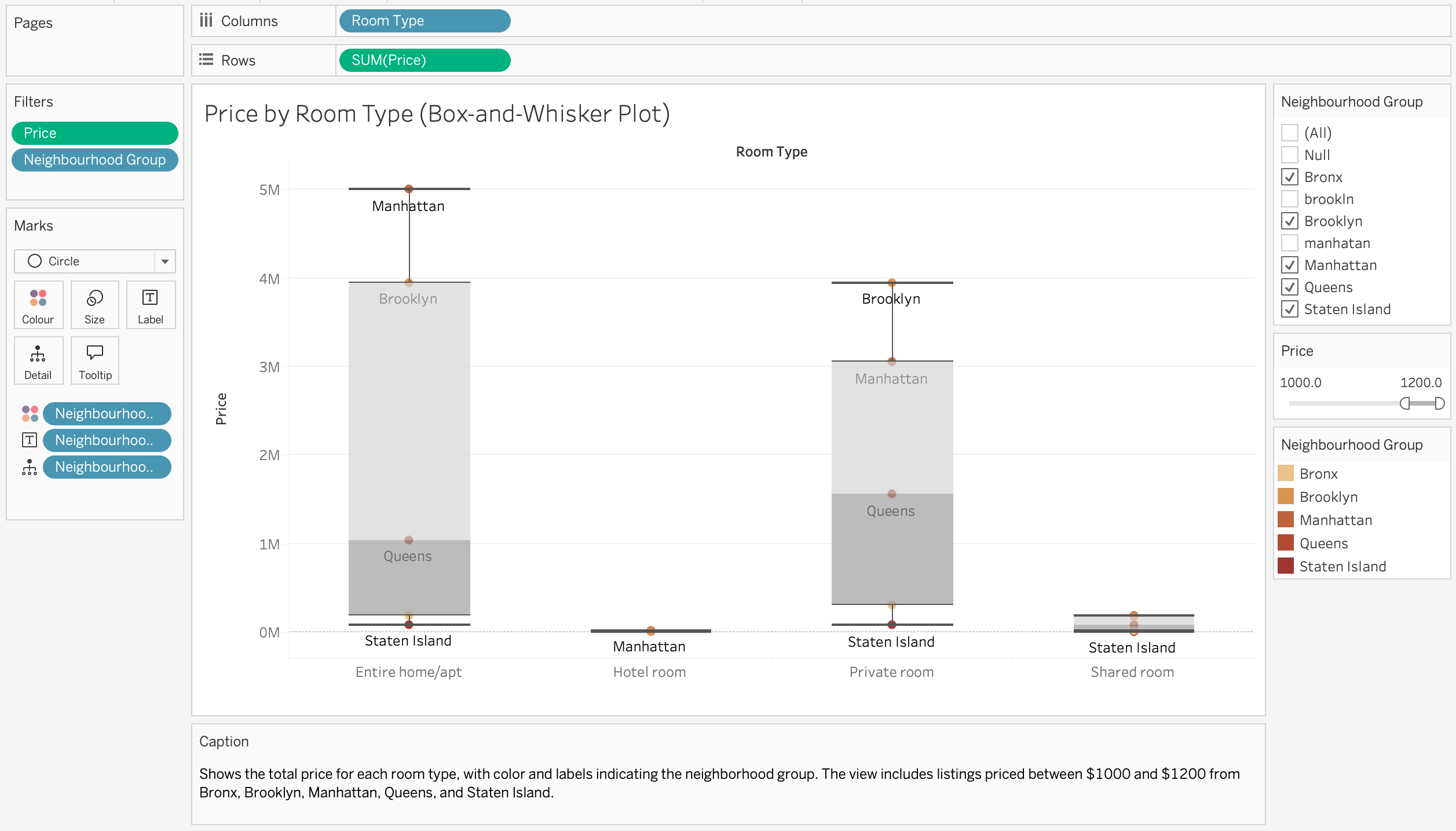
* **Feedback Received**:
  + Colour schemes varied too much between visualizations, making it difficult to track categories like room type or borough across charts.
* **Improvements Made**:
  + Applied consistent colour schemes across all charts e.g., each room type, and neighbourhood group.
  + All sheets share the same colour throughout the dashboard.

**Screenshots of Visualizations:**

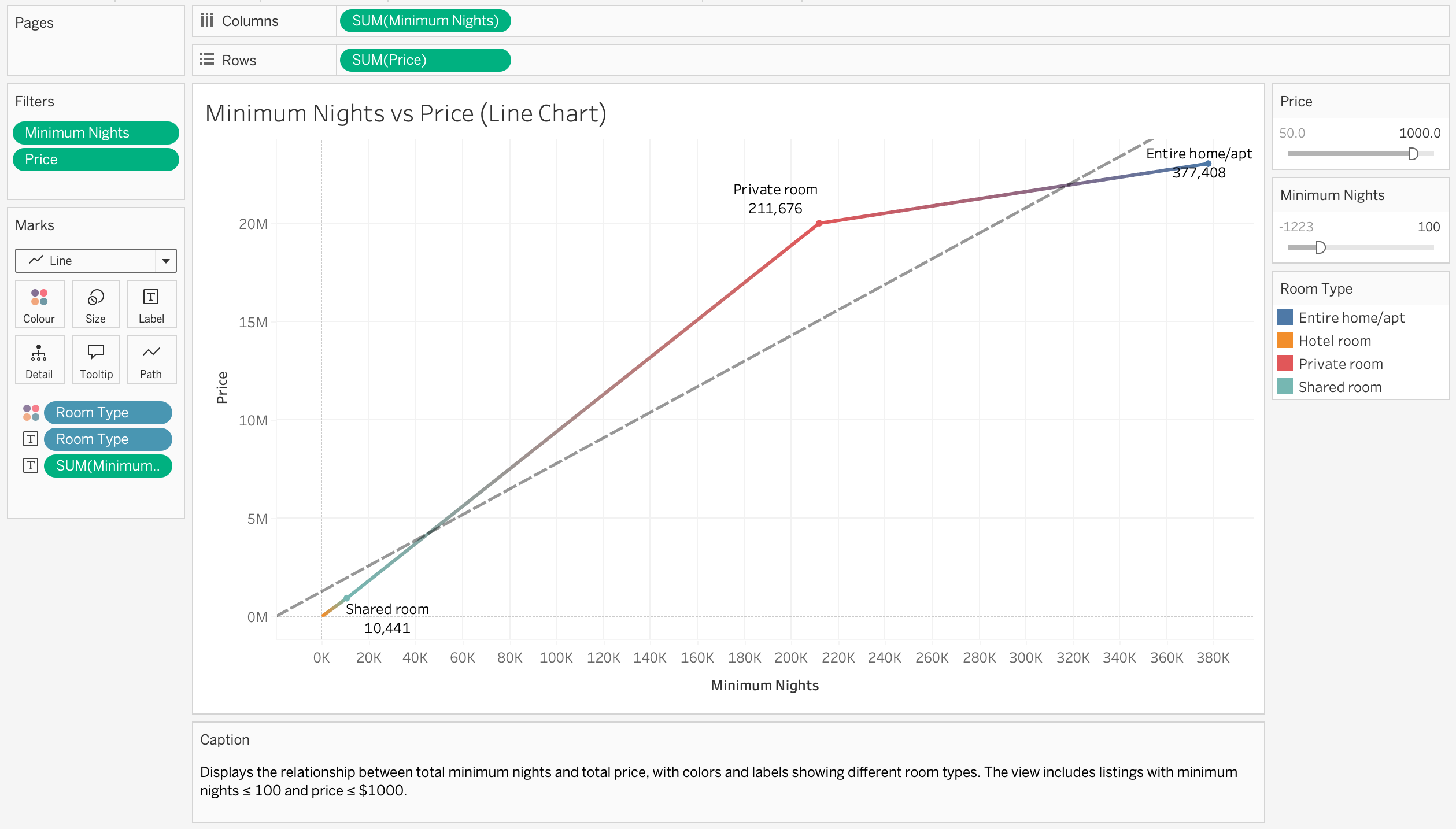
**Visualization 1: Average Price by Neighbourhood Group (Map)**

****

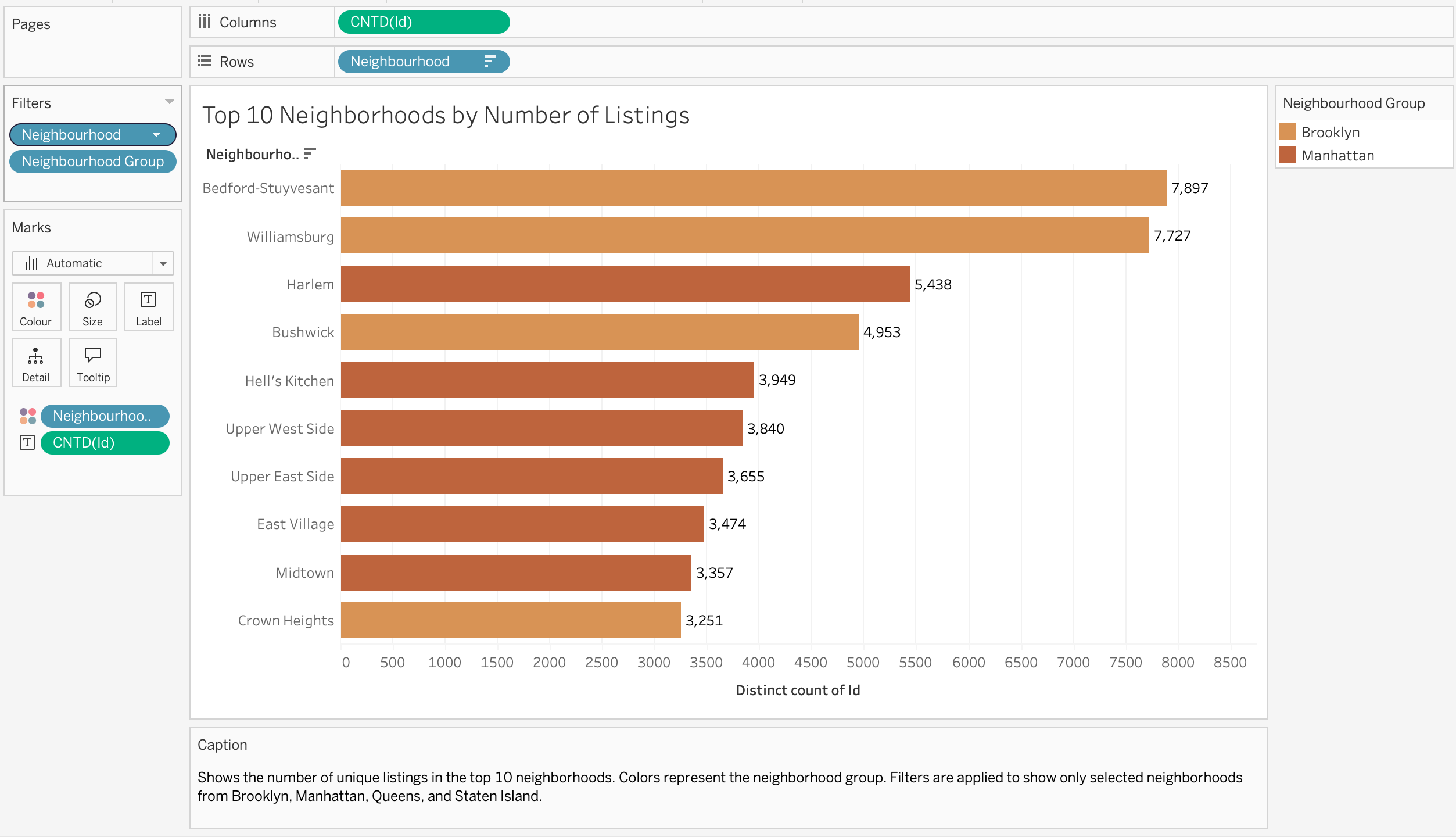
**Visualization 2: Price by Room Type (Box-and-Whisker Plot)**

****

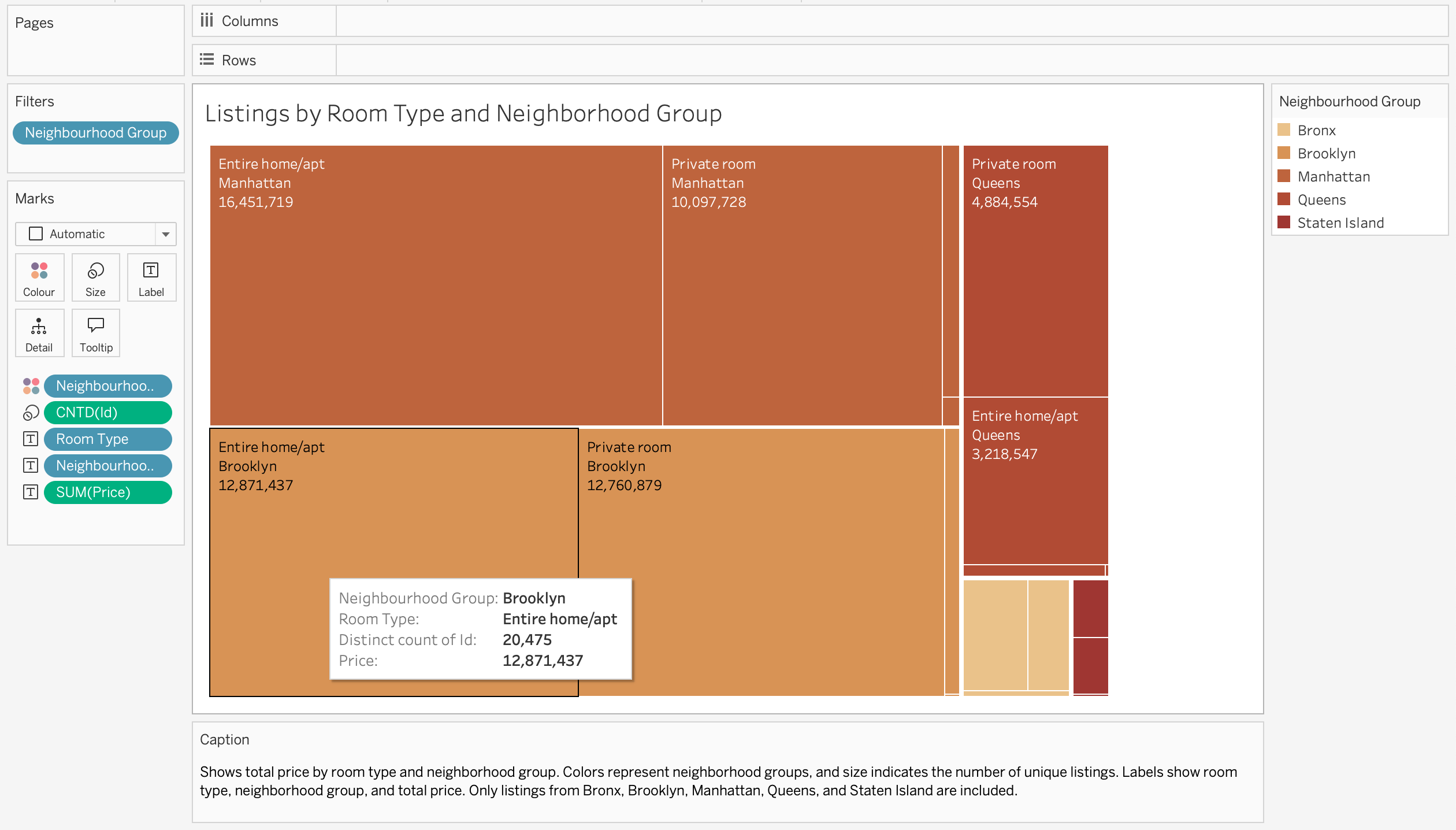
**Visualization 3: Minimum Nights vs Price (Line Chart with Trend Line)**

****

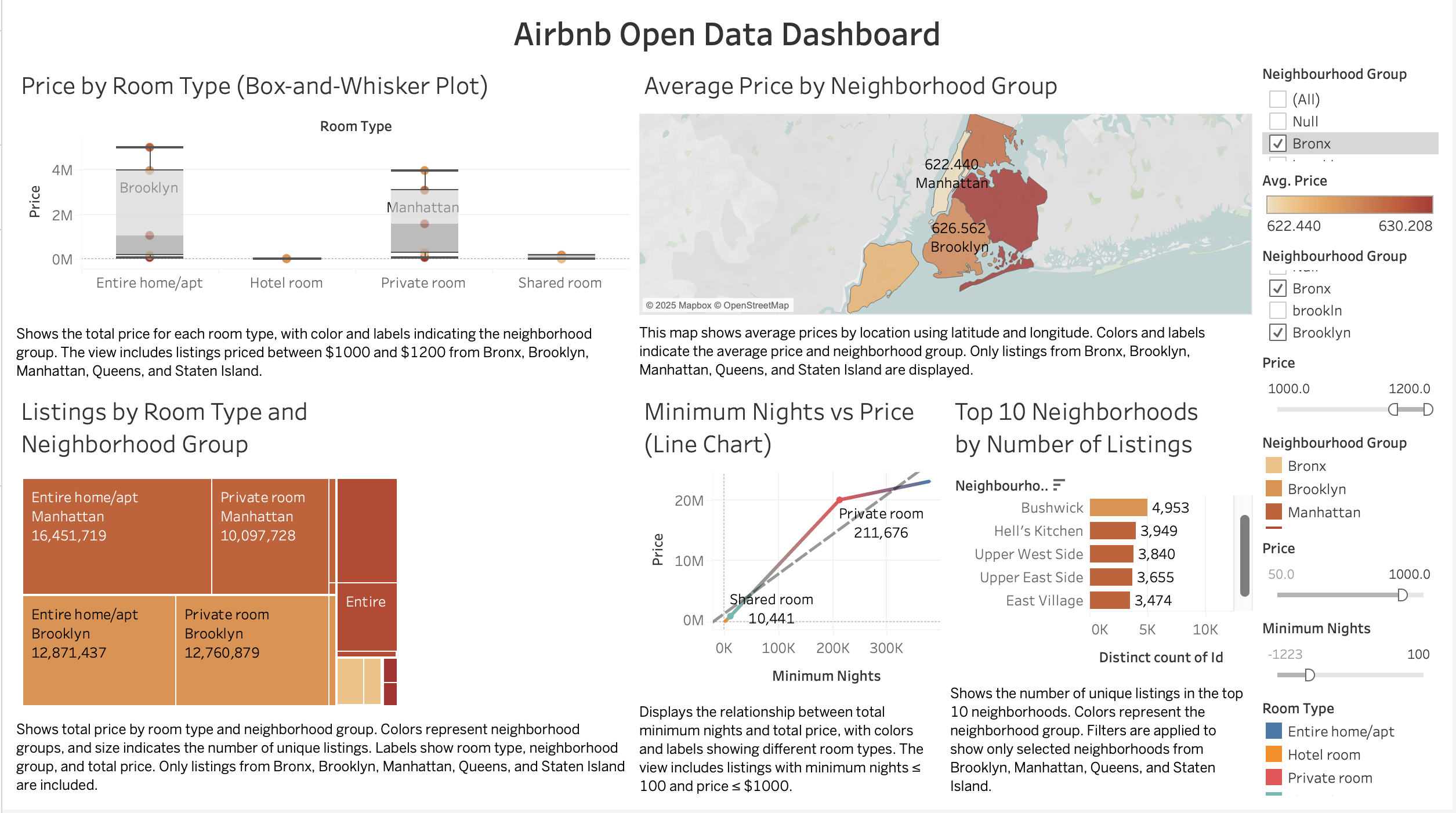
**Visualization 4: Top 10 Neighbourhoods by Number of Listings (Bar Chart)**

****

**Visualization 5: Listings by Room Type and Neighbourhood Group (Treemap)**

****

**Dashboard:**

****

**Design Principles Used:**

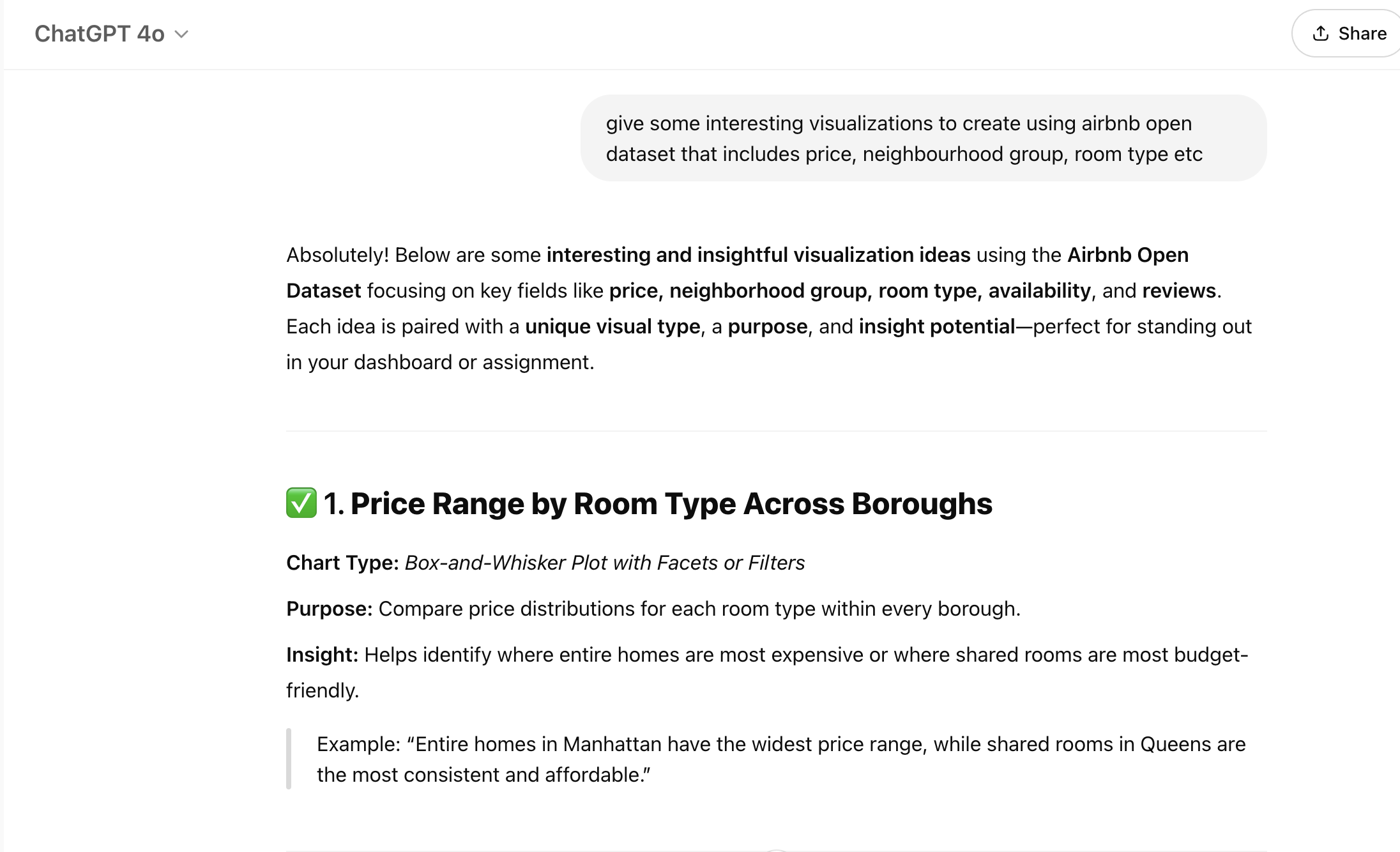
* To enhance the clarity and effectiveness of my dashboard, I applied key visualization principles from **Cole Nussbaumer Knaflic, Edward Tufte**, and **Stephen Few**.
* In line with **Stephen Few’s** (2009) best practices, I maintained a **consistent layout, typography, and formatting** across all visuals. This made the dashboard more visually coherent and allowed users to **quickly extract insights** without distraction.
* These combined principles significantly improved the **usability, storytelling, and visual integrity** of my dashboard, helping viewers focus on what matters most.

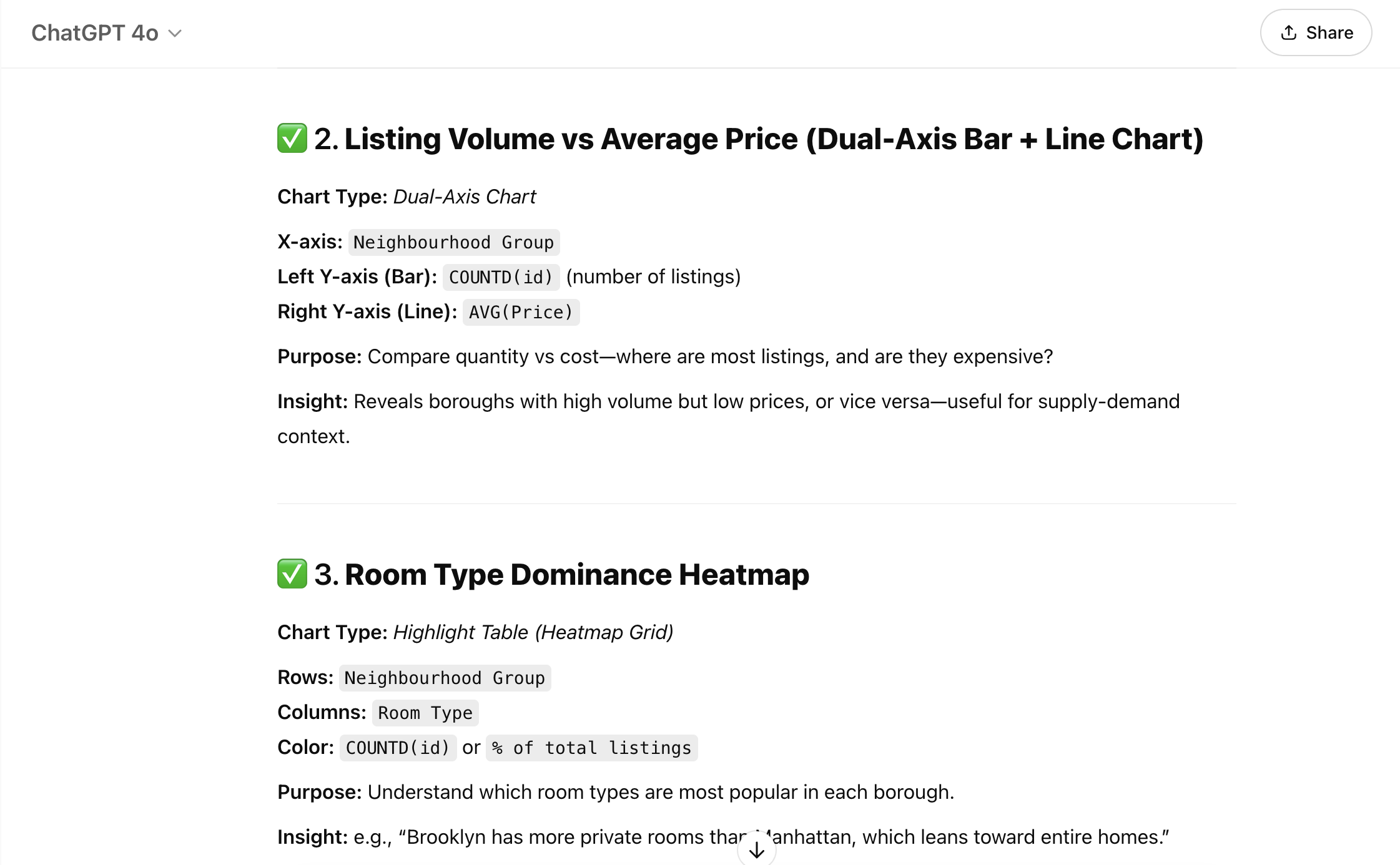
**References**:

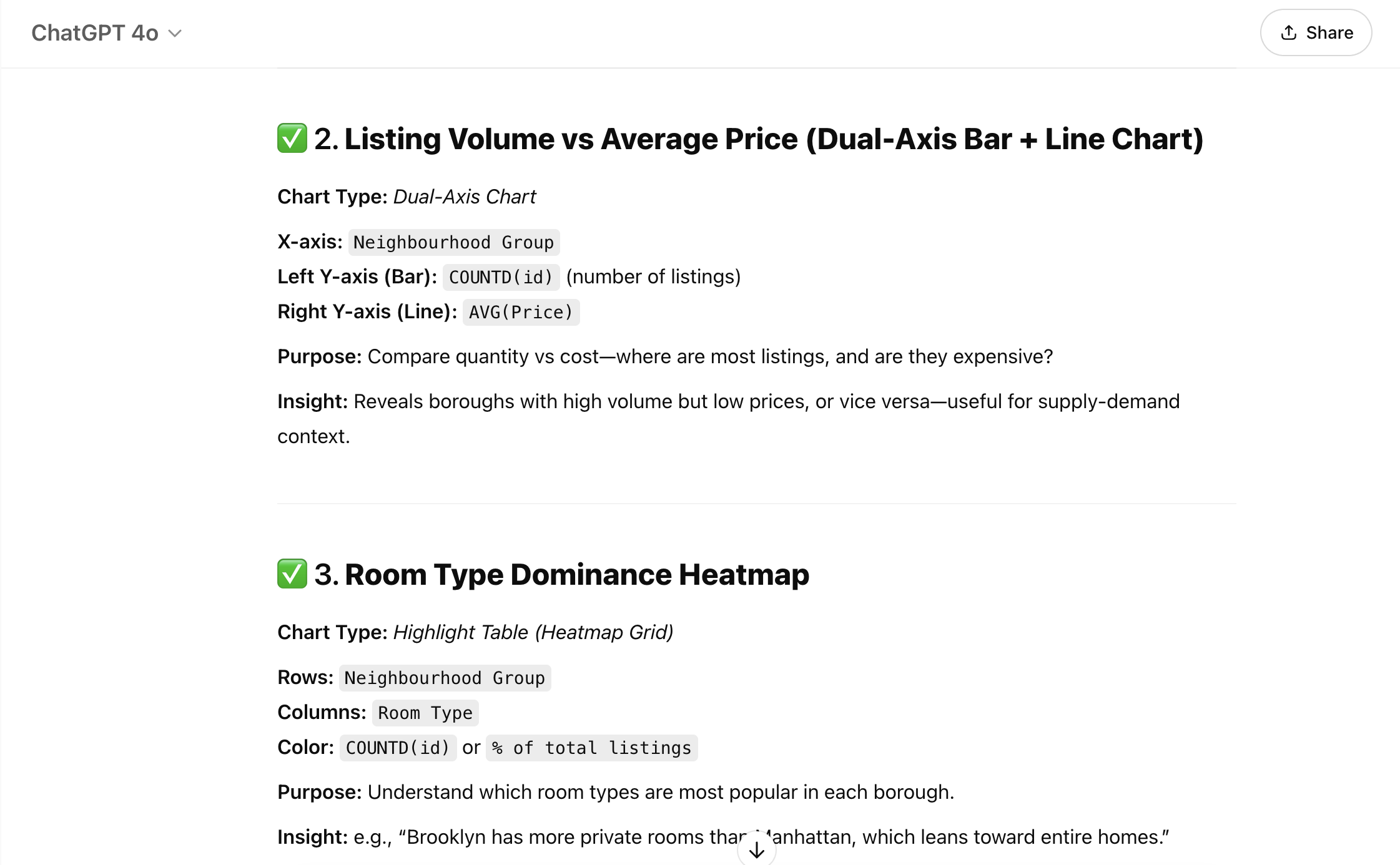
* Tufte, E. R. (2001). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.
* TheNewtonsRing. (2023). Global YouTube Statistics 2023 – Data analysis project [Code repository]. GitHub. <https://github.com/TheNewtonsRing/Global-YouTube-Statistics-2023>
* Nelgiriyewithana, N. (2023). Global YouTube Statistics 2023 [Dataset]. Kaggle. <https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023>
* Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley. <https://www.wiley.com/en-us/Storytelling+with+Data%3A+A+Data+Visualization+Guide+for+Business+Professionals-p-9781119002253>
* Kaggle Dataset: <https://www.kaggle.com/datasets/arianazmoudeh/airbnbopendata>

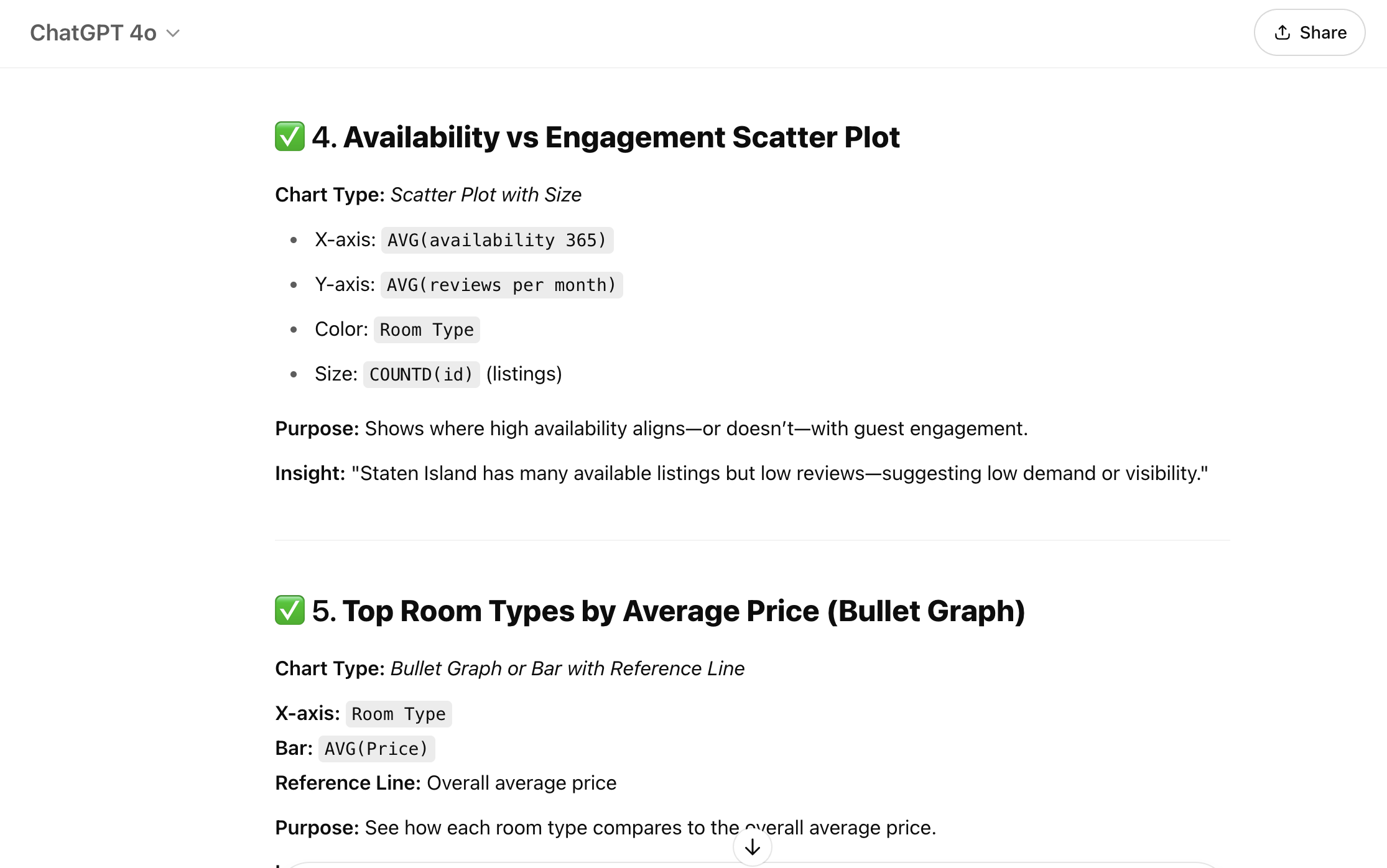
**Use of ChatGPT & Screenshots:**

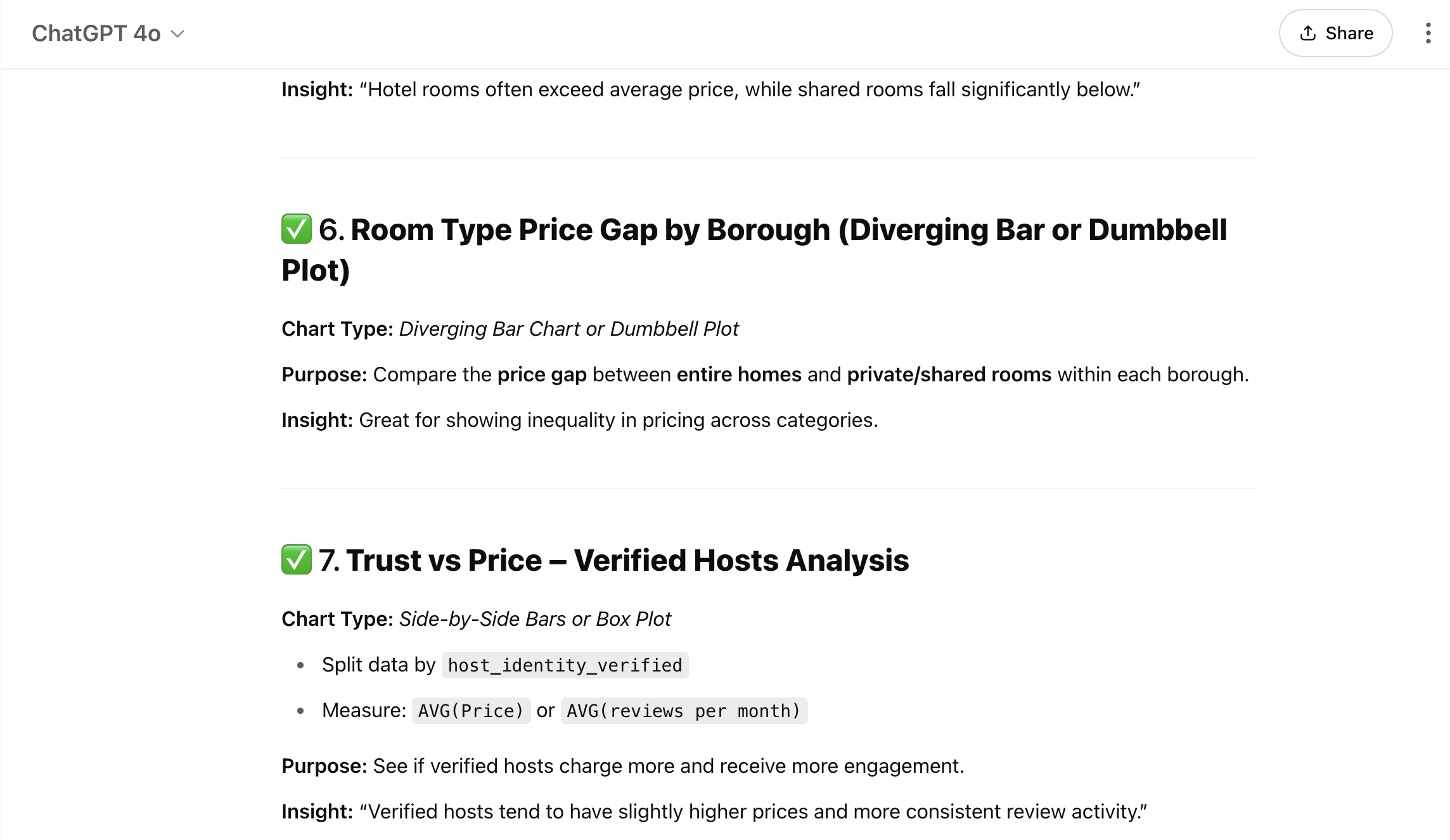
**Prompt 1 & Response:**

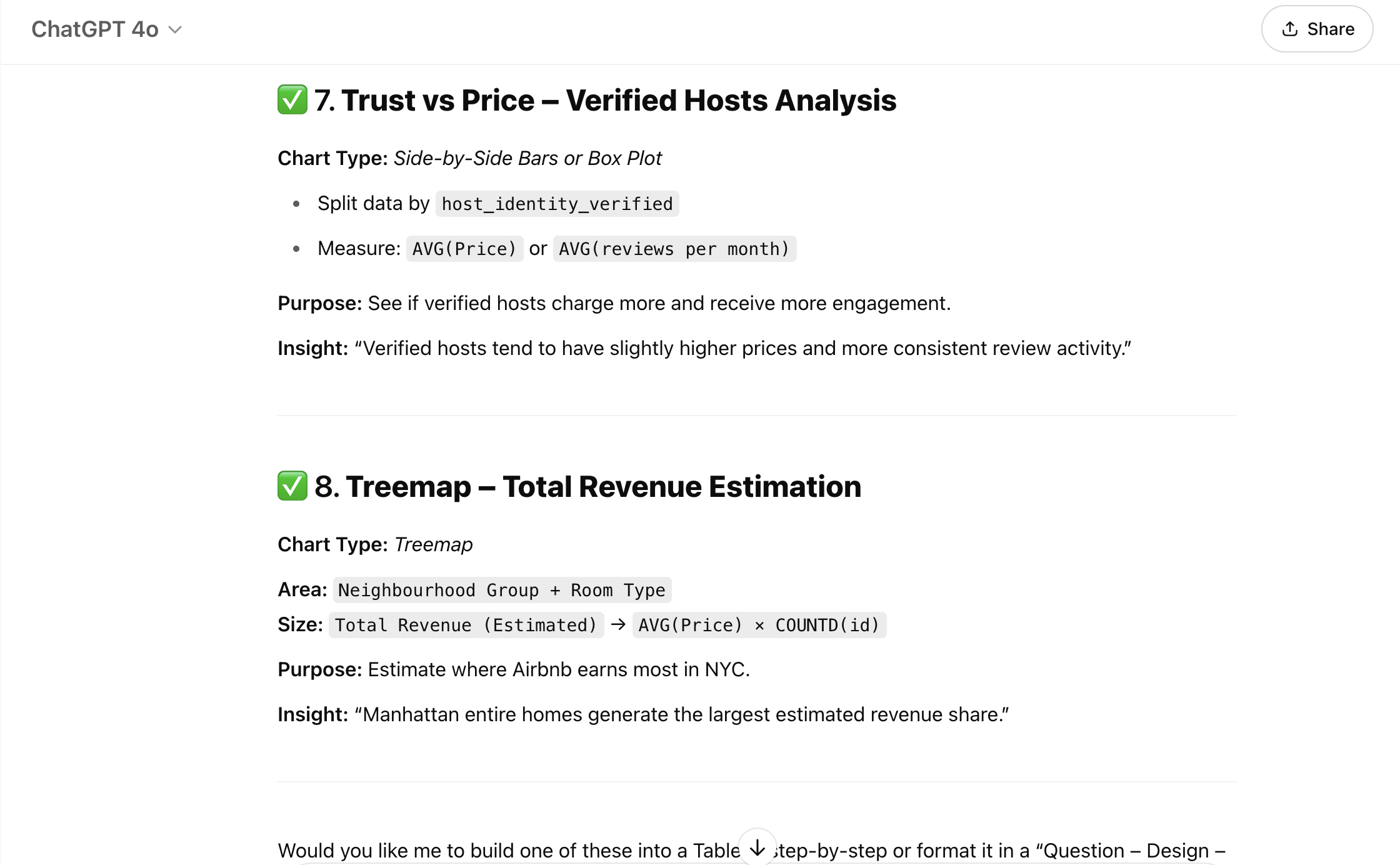
****

****

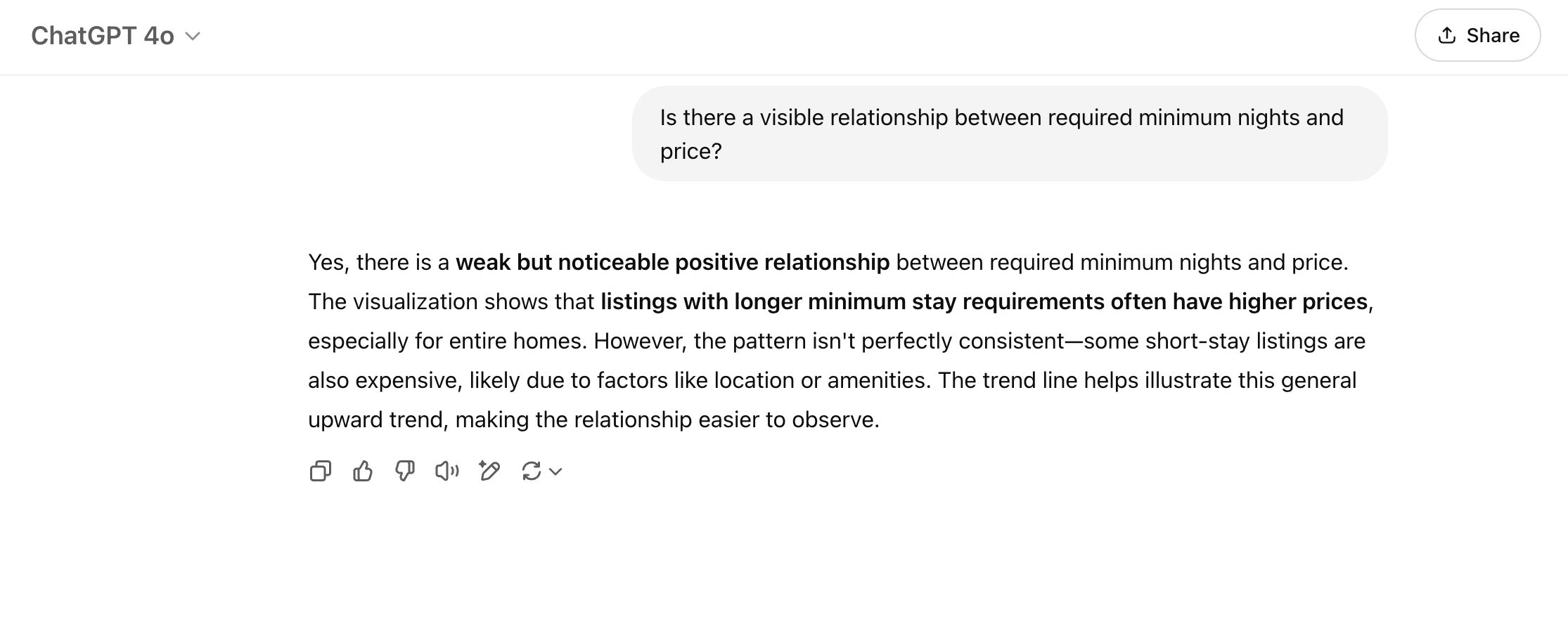
****

****

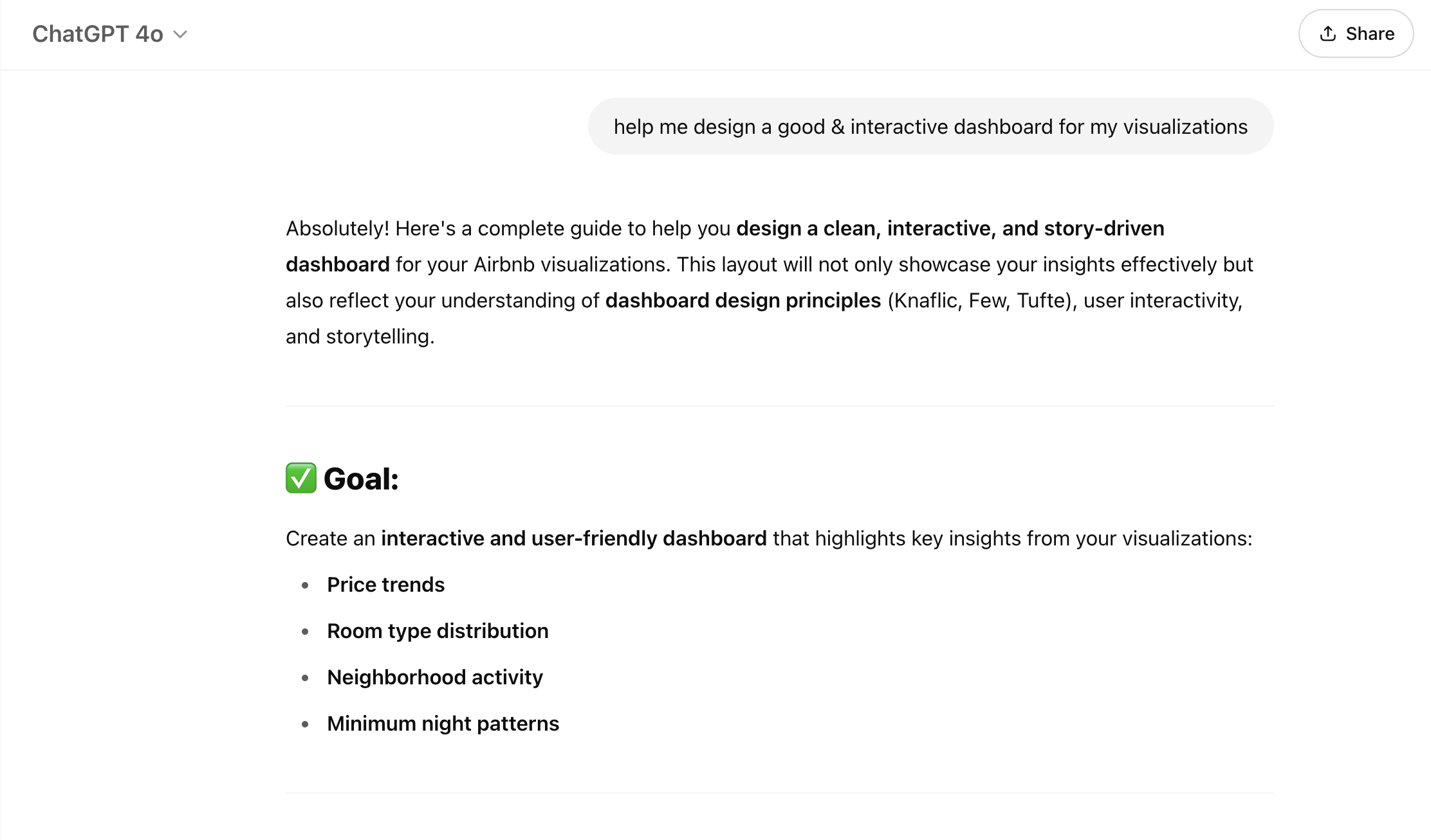
****

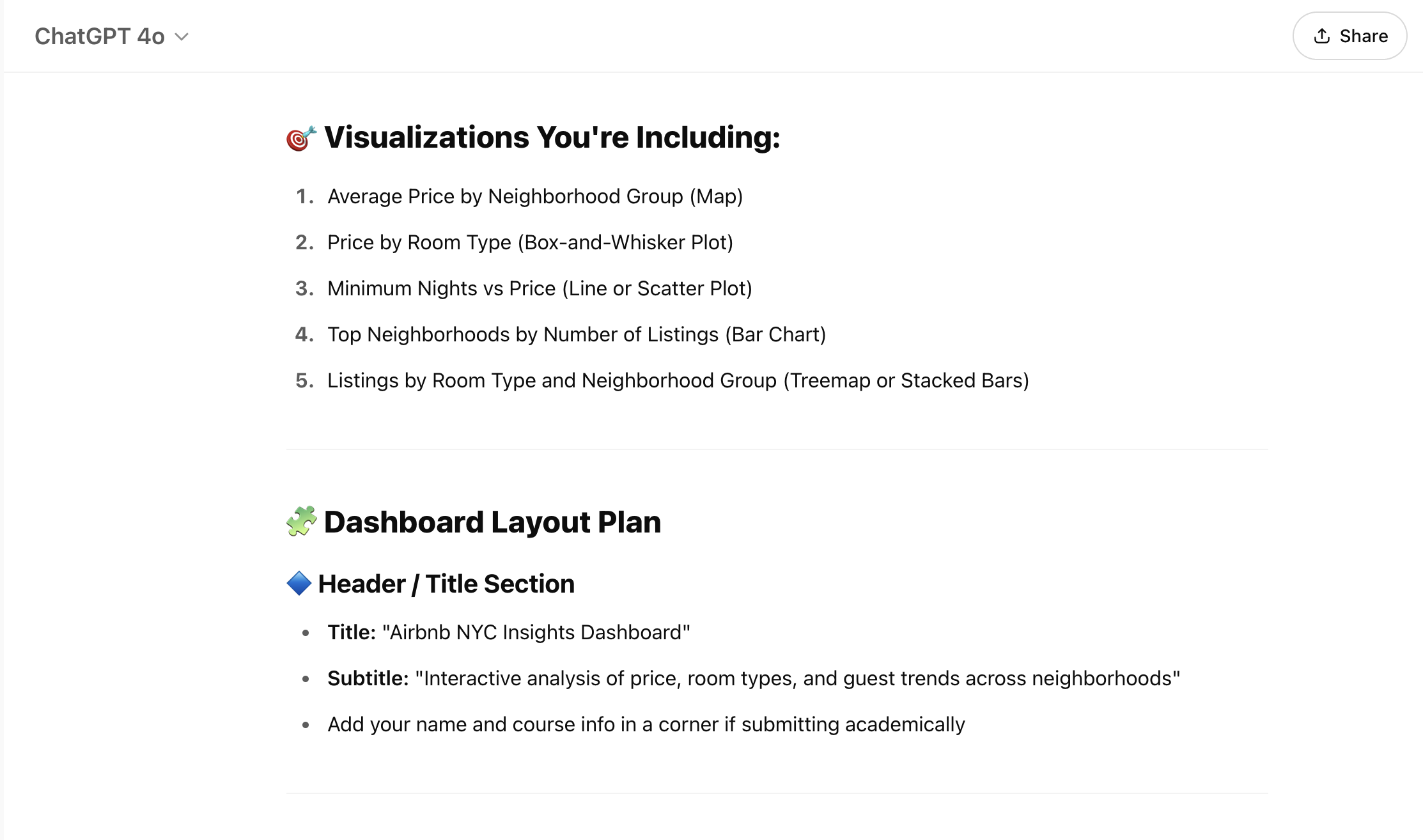
****

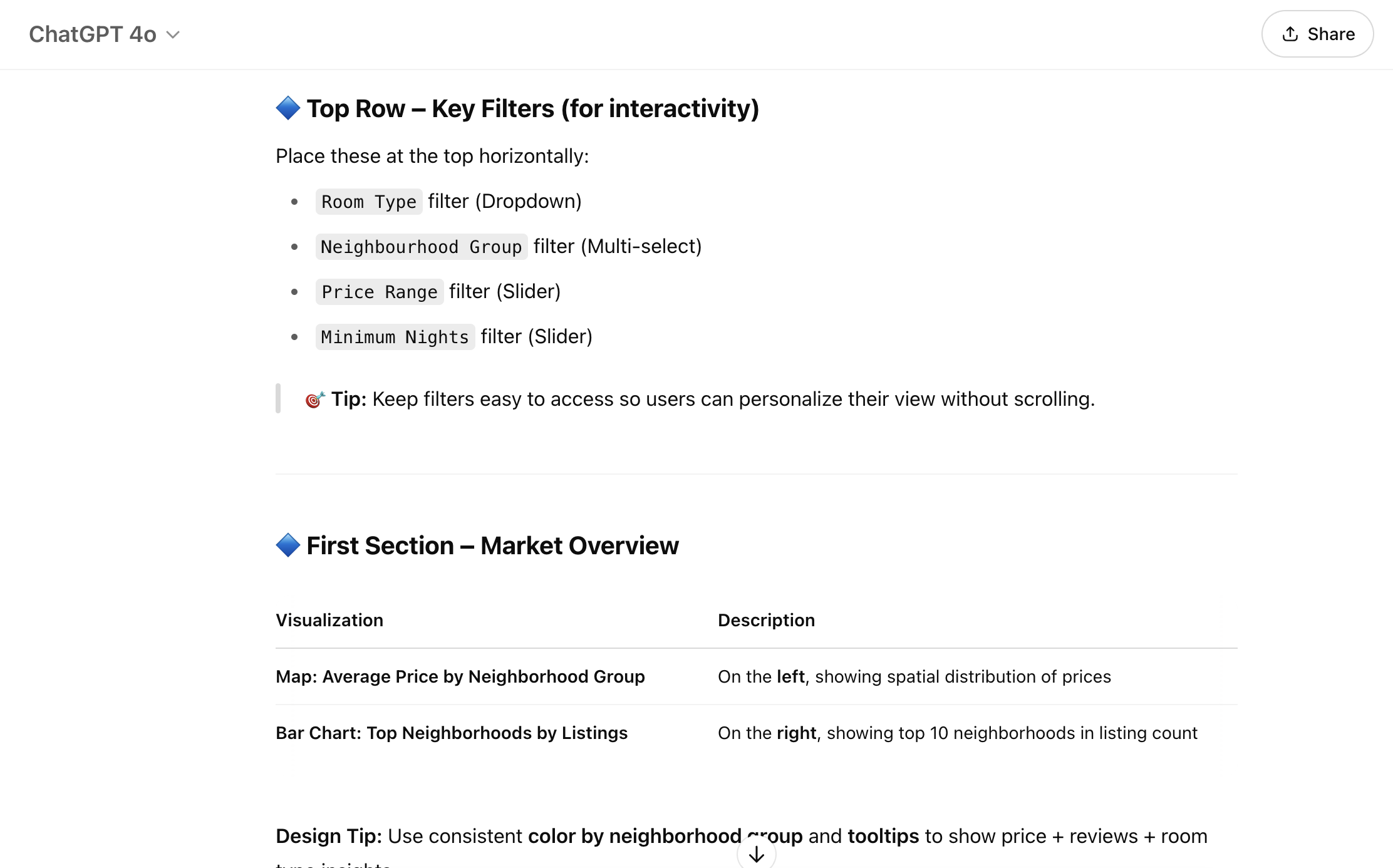
**Prompt 2 & Response:**

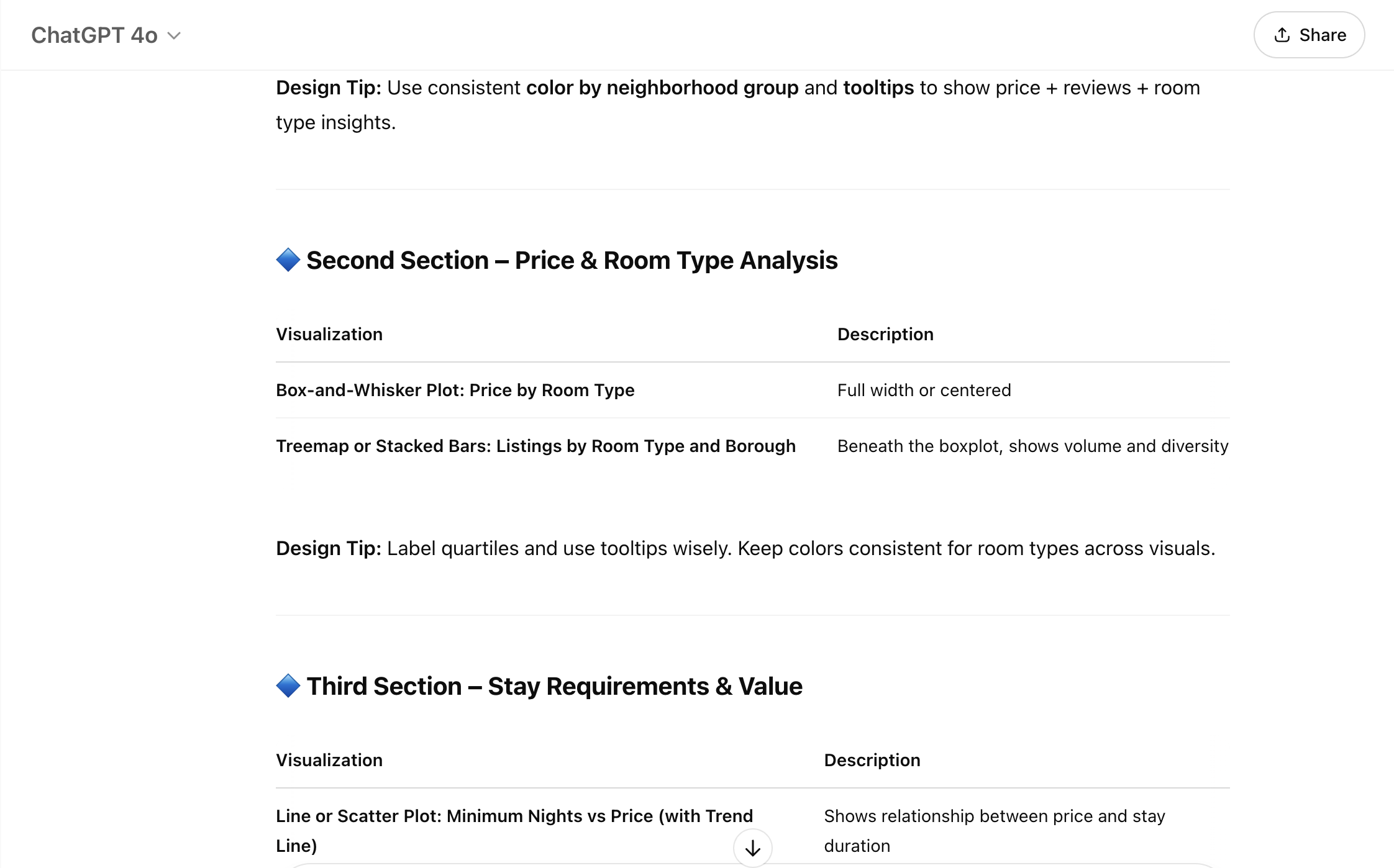
****

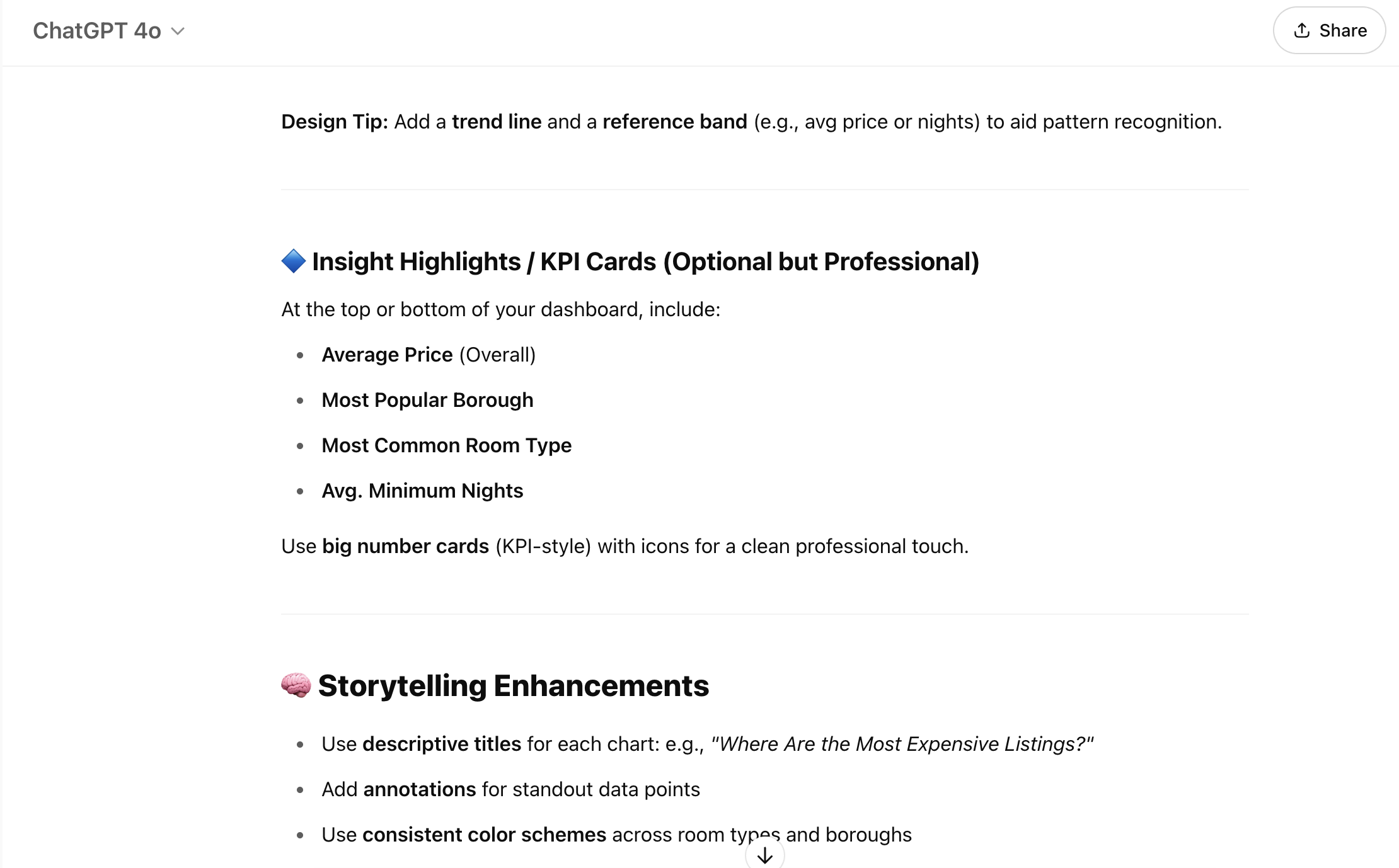
**Prompt 3 & Response:**

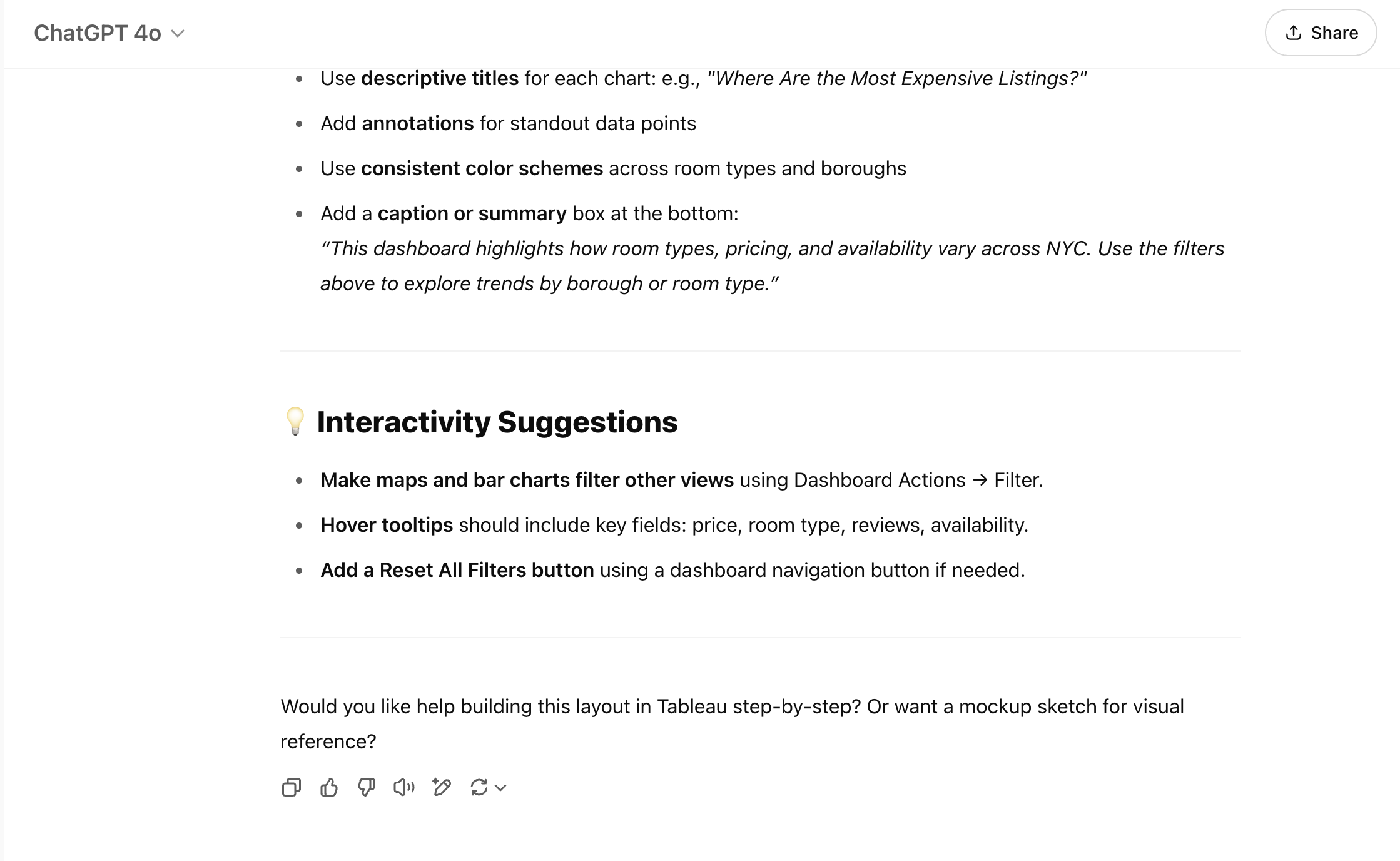
****

****

****

****

****

****

**AI Modification Summary**

* AI-generated text was used to generate ideas & outline initial visualization descriptions.
* I revised and restructured the content to match the tone of my report and personalize insights based on my actual Tableau work.