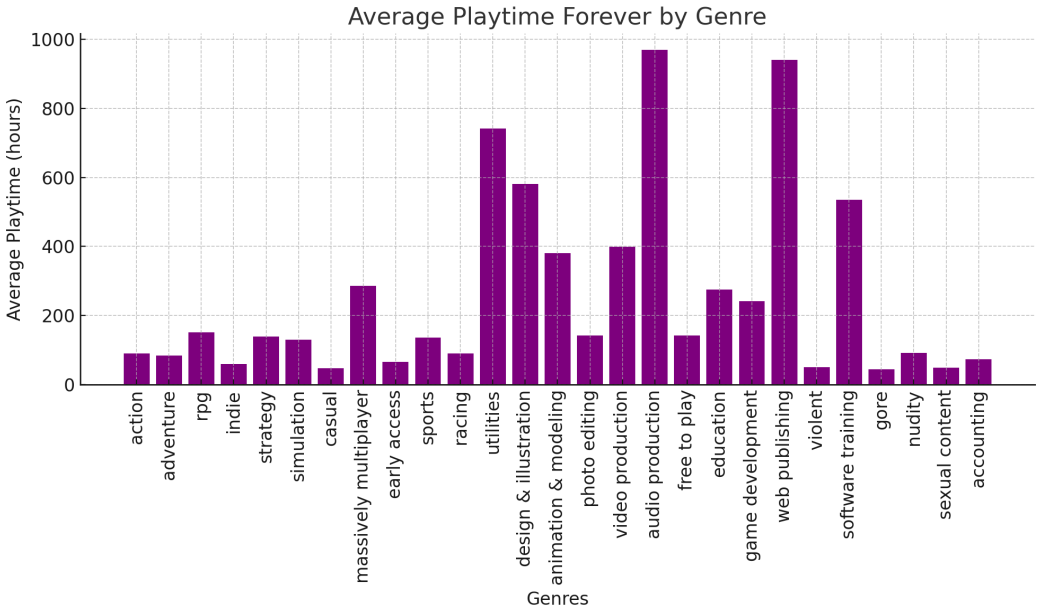
Backup Documentation (for the graphs, in case jupyter is not installed)

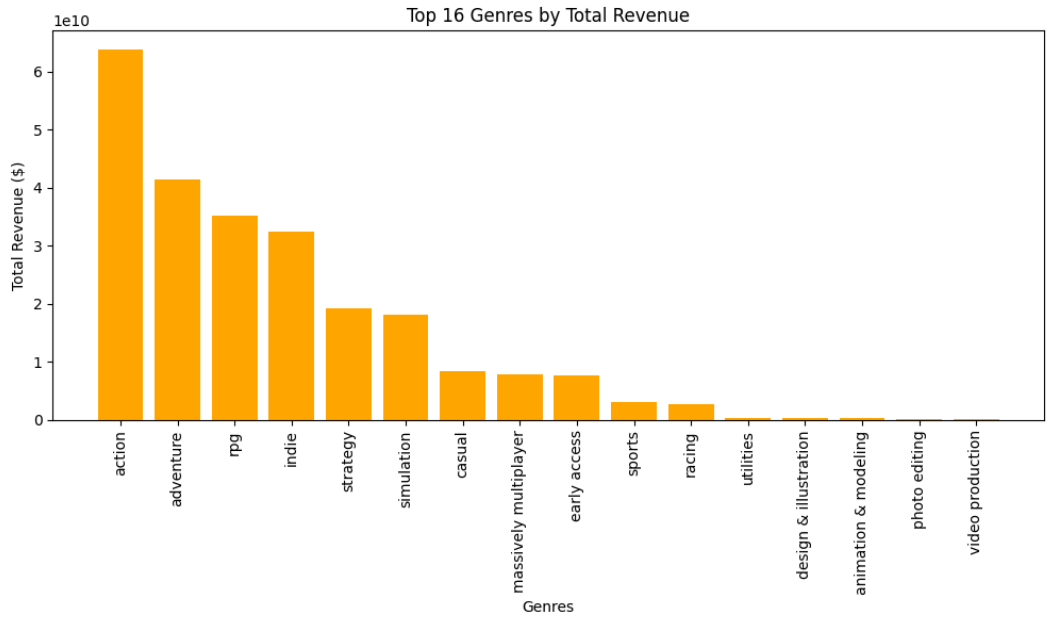
Average Playtime Forever by Genre

This chart identifies which genres tend to have the highest player engagement by comparing the average playtime across genres. The dataset includes categories like “Audio Production” and “Web Publishing”, along with other non-gaming genres like “Video Production”, “Photo Editing”, and “Design & Illustration”. These categories most likely represent software or tools available on platforms like Steam, which offer more than just games. **“Audio Production”** and **“Web Publishing”** show significantly higher average playtime compared to game genres like **“RPG”** and **“Strategy”**, suggesting that users spend a substantial amount of time using these software tools.



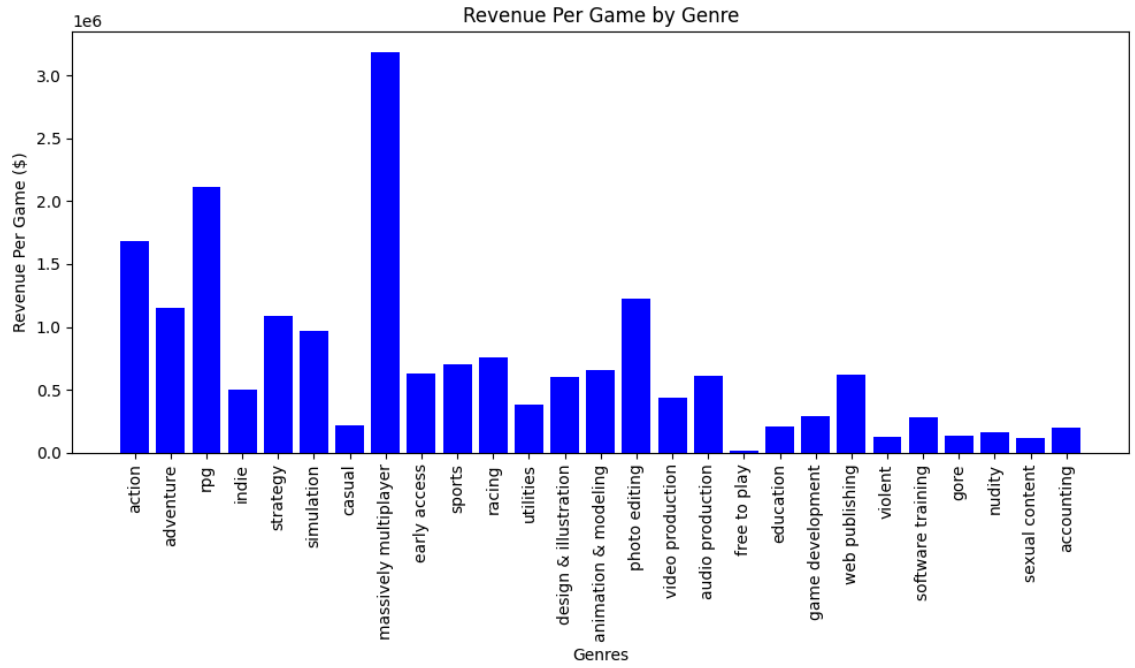
Total Revenue by Genre

This line plot highlights the most financially successful game categories. The **"Action"** genre leads in revenue generation, followed by **"Adventure"** and **"RPG".**



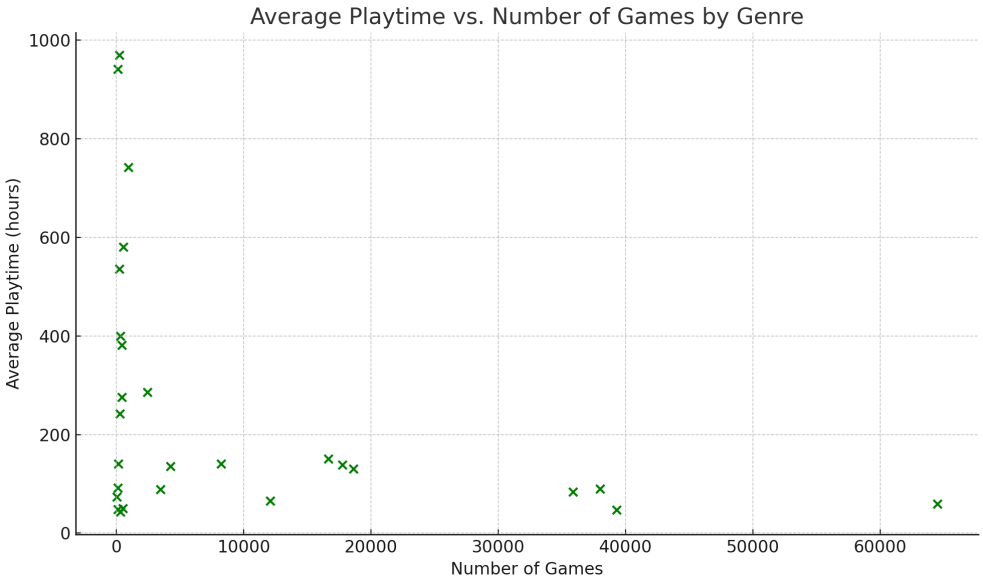
Total Revenue Per Game by Genre

This bar plot helps identify which genres are more profitable on a per-game basis. There is considerable variation in revenue per game across different genres, but genres such as “**Massively Multiplayer”** and “**RPG**” show exceptionally high revenue per game.



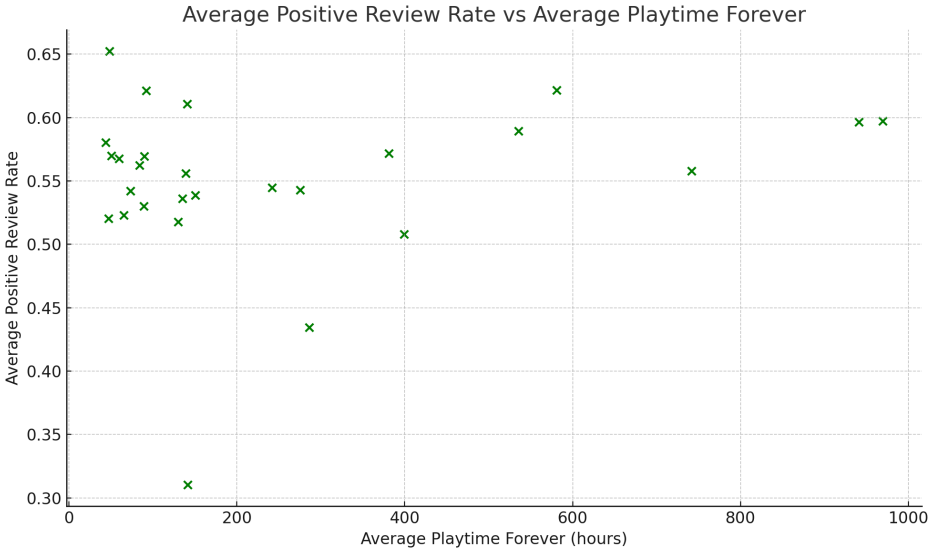
Average Playtime vs. Number of Games by Genre

This plot indicates a **weak inverse trend** where genres with **more games** tend to have **lower average playtimes**, while genres with fewer games (such as RPGs) tend to have higher average playtimes. This suggests that game types with more immersive or in-depth experiences (fewer but longer games) are more time-consuming, while genres with many shorter, less immersive titles may have more games but less player engagement per title.



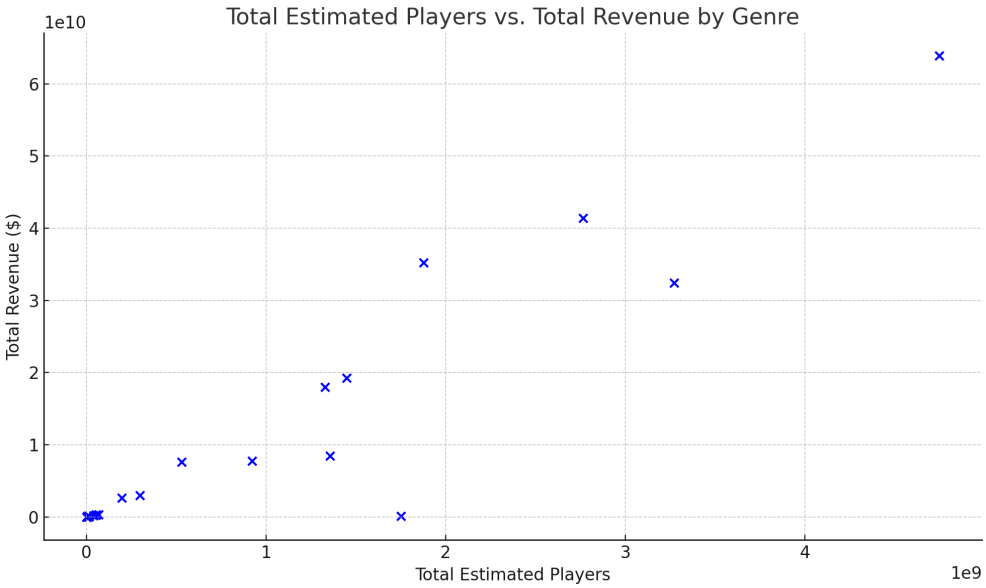
Average Positive Review Rate vs. Average Playtime Forever

This plot suggests that longer playtime does not always equate to higher player satisfaction, and other factors (like game quality or genre preferences) may influence review rates more than just the time spent playing.



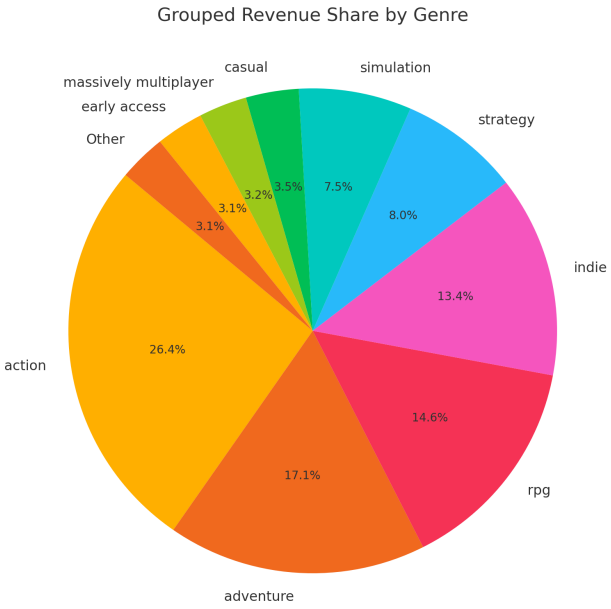
Total Estimated Players vs. Total Revenue by Genre

The plot shows a strong positive correlation between the number of players and total revenue. This indicates that genres with more players tend to generate higher revenue, which is expected since a larger player base often leads to increased sales or in-game purchases.



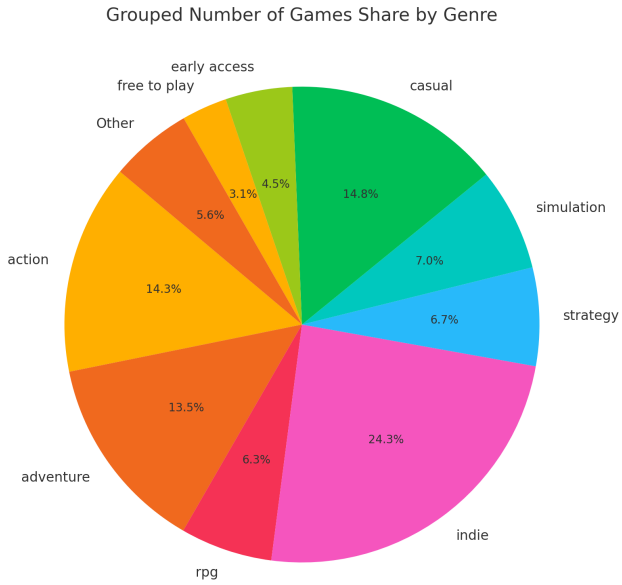
Revenue Share by Genre

This pie chart visualizes how different genres contribute to the total gaming industry revenue. Genres like "Action" and "Adventure" dominate the revenue share, highlighting their financial impact.



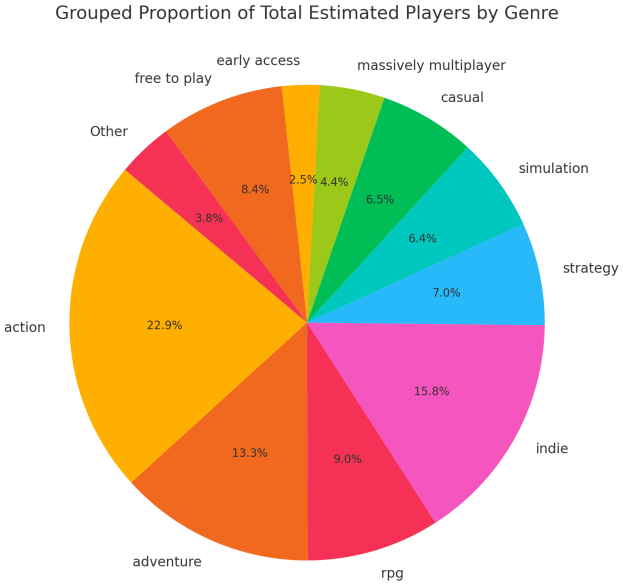
Number of Games Share by Genre

This pie chart reveals that the "Indie" genre leads in the number of games, followed by “Casual”. This highlights the large volume of games developed within these two categories.



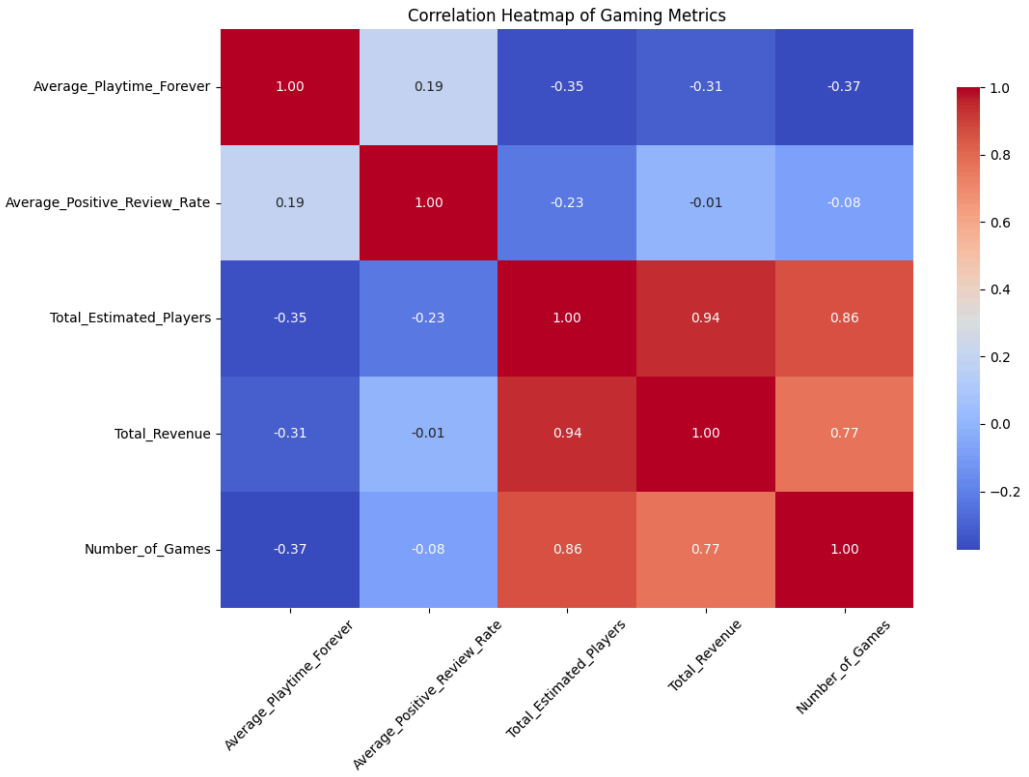
Proportion of Total Estimated Players by Genre

This pie chart reveals that genres like "Action" and "Adventure" attract the largest player base, emphasizing their popularity among gamers.



Heatmap of the correlation matrix

A heatmap is useful for visualizing the relationships and correlations between different metrics like average playtime, total revenue, positive review rate, and total estimated players.



These are some insights gained for the correlation matrix of the *games_genre_metrics* dataset:

- Total Revenue and Total Estimated Players: There's a very strong positive correlation (0.94) between these two variables, indicating that genres with higher player counts generally generate more revenue.

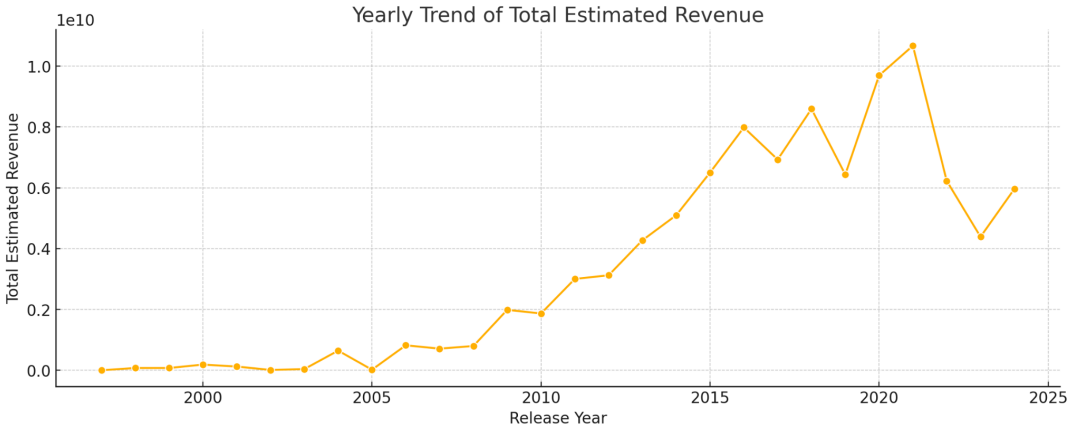
- Total Revenue and Number of Games: There is a positive correlation (0.77), suggesting that genres with more games tend to generate more revenue.

- Average Playtime and Total Revenue/Players: Average Playtime shows negative correlations with Total Revenue (-0.31) and Total Estimated Players (-0.35), suggesting that genres where players spend more time per game do not necessarily have higher revenue or player counts.

- Average Positive Review Rate and Revenue: Surprisingly, these metrics show little to no correlation, indicating that higher player satisfaction does not necessarily translate to higher financial success.

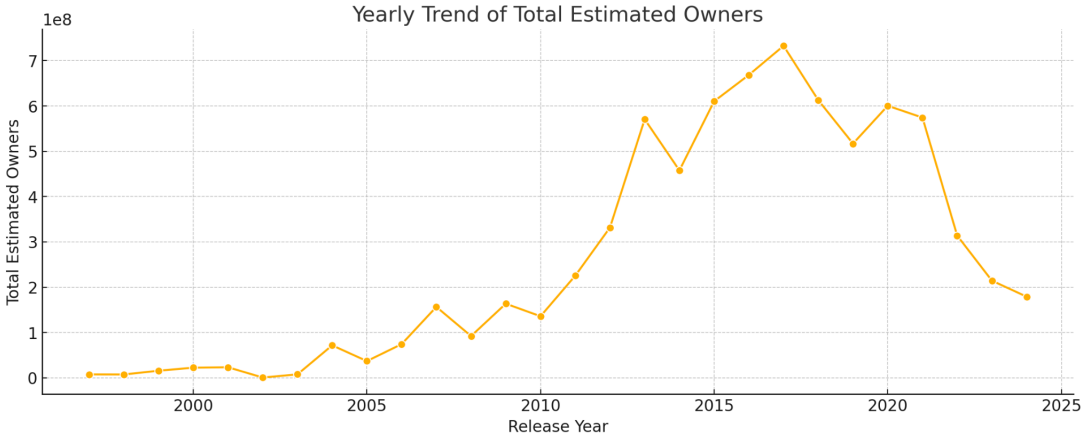
Yearly Trend of Total Estimated Revenue

This plot shows a significant fluctuation in revenue over the years, with certain periods showing sharp increases, indicating successful game releases or industry growth during those times.



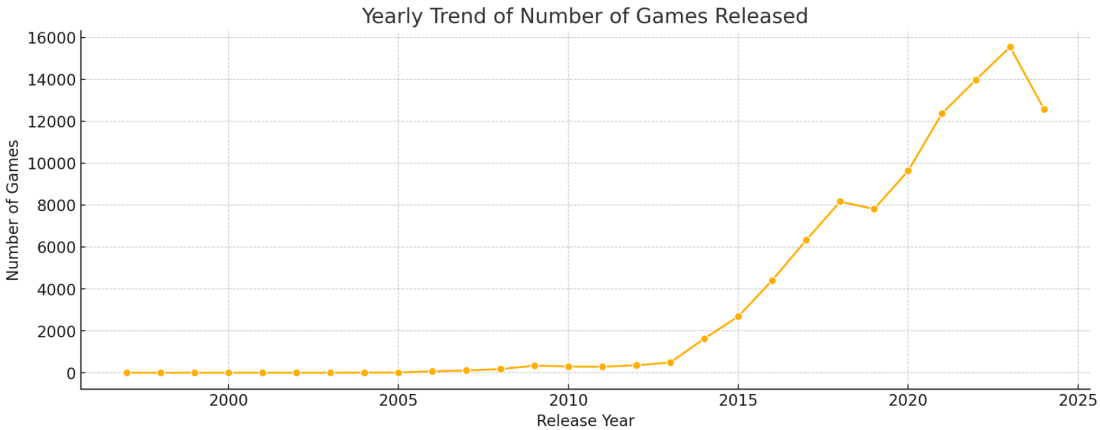
Yearly Trend of Total Estimated Owners

The number of game owners also fluctuates, mirroring the revenue trends, which suggests a strong correlation between user acquisition and revenue growth.



Yearly Trend of Number of Games Released

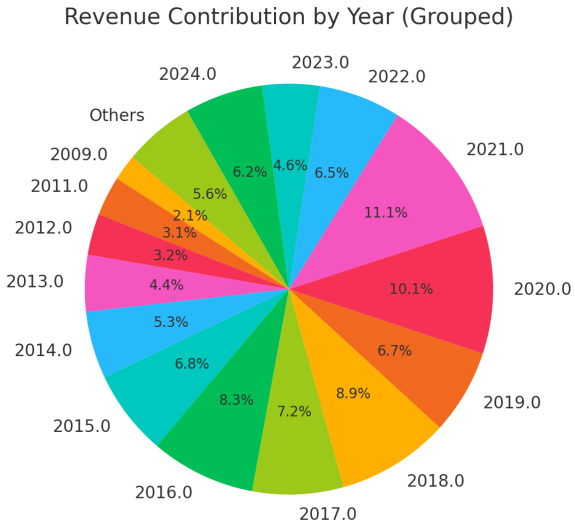
The number of games released annually shows a positive upward trend over time, suggesting a growing gaming industry.

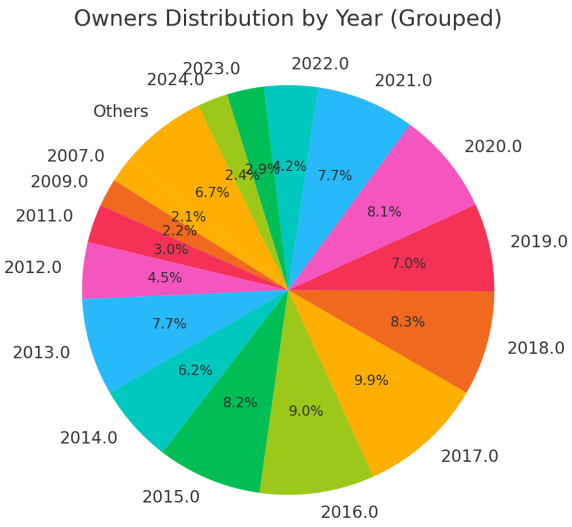


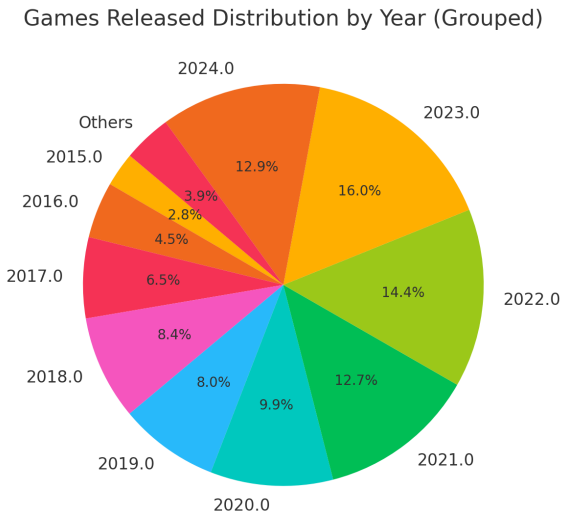
Revenue Contribution by Year: Displays the major contributors to total revenue, with smaller contributions grouped under "Others".

Owners Distribution by Year: Shows the distribution of game owners across years, with minor contributions also grouped as "Others".

Games Released Distribution by Year: Highlights the share of game releases per year, grouping years with smaller numbers under "Others".

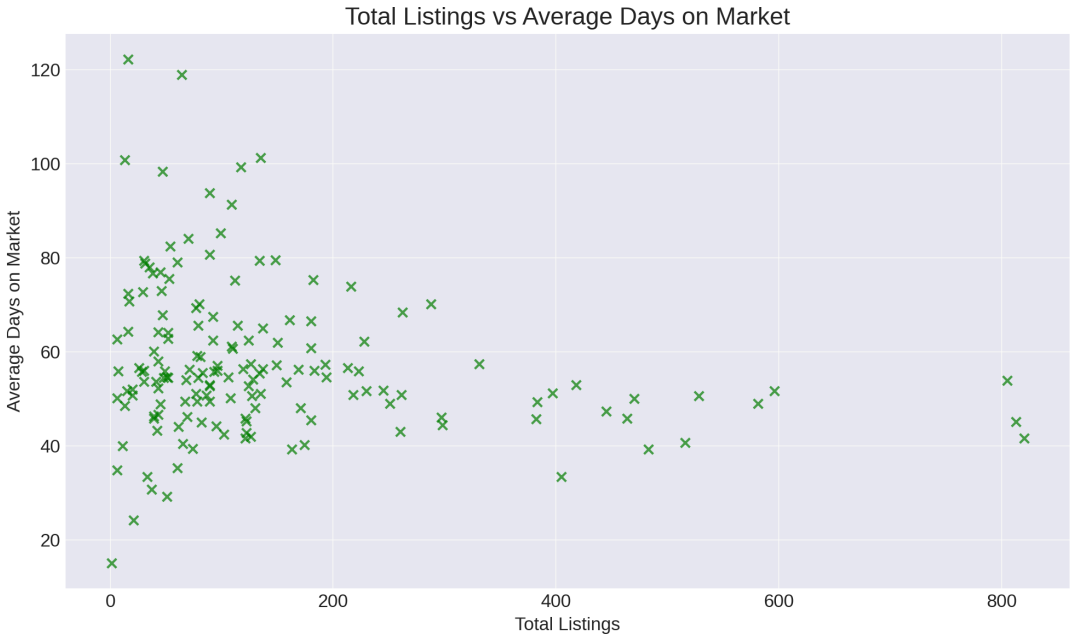






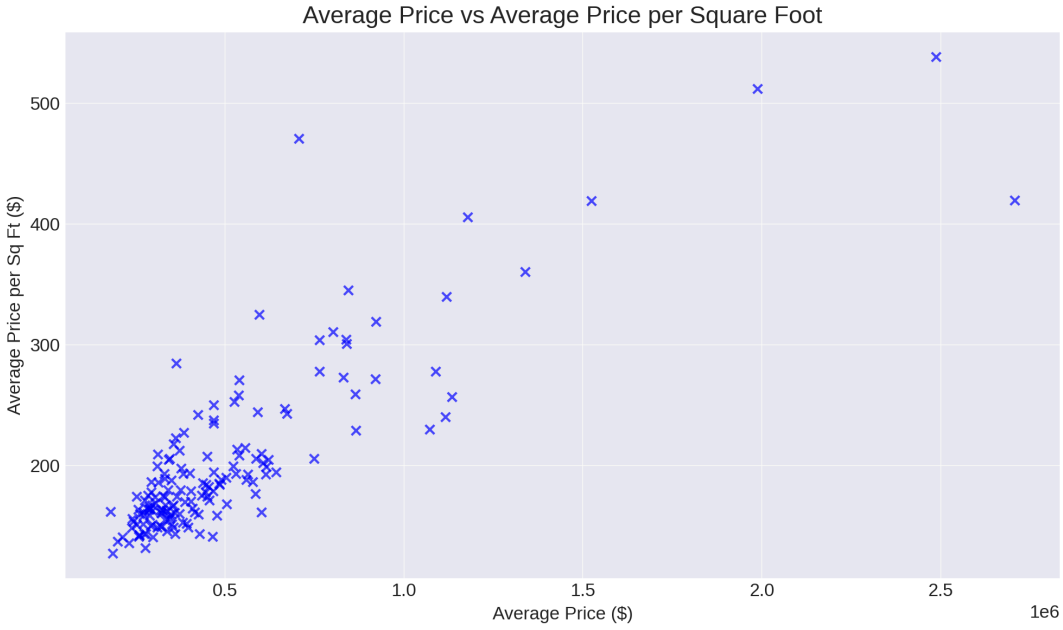
Total Listings vs Average Days on Market

This plot suggests that the quantity of listings alone doesn't determine how quickly homes sell. Other factors, like location, price, or buyer demand, could have a more significant influence on how long homes remain on the market. However, any clusters of points could indicate areas with oversupply (lots of listings and long market times).

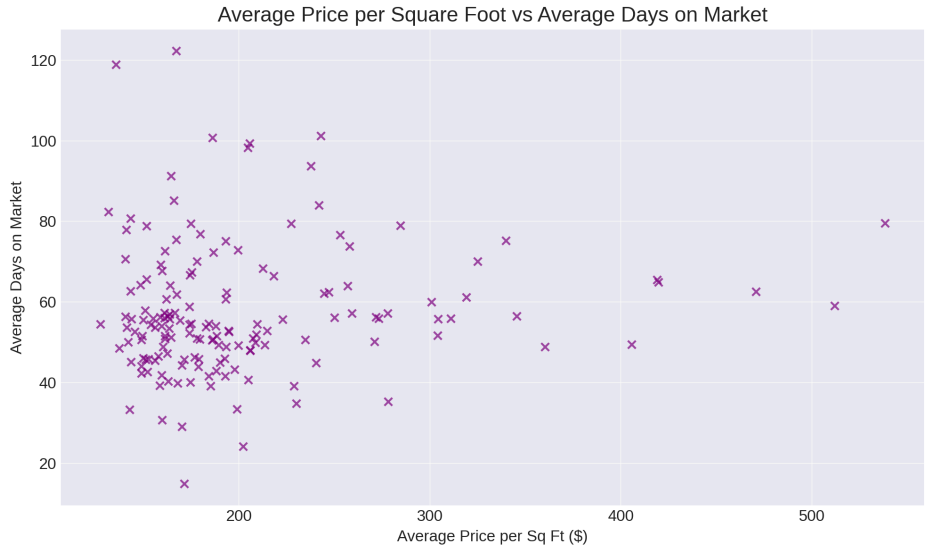


Average Price vs Average Price per Square Foot

There seems to be a positive correlation, meaning that as the overall price of a property increases, the price per square foot also tends to rise. This is typical in real estate markets, where more expensive homes often have higher value per unit area. Outliers (properties with high prices but relatively lower price per square foot or vice versa) might suggest different types of properties, such as luxury homes with more space but less concentrated value.

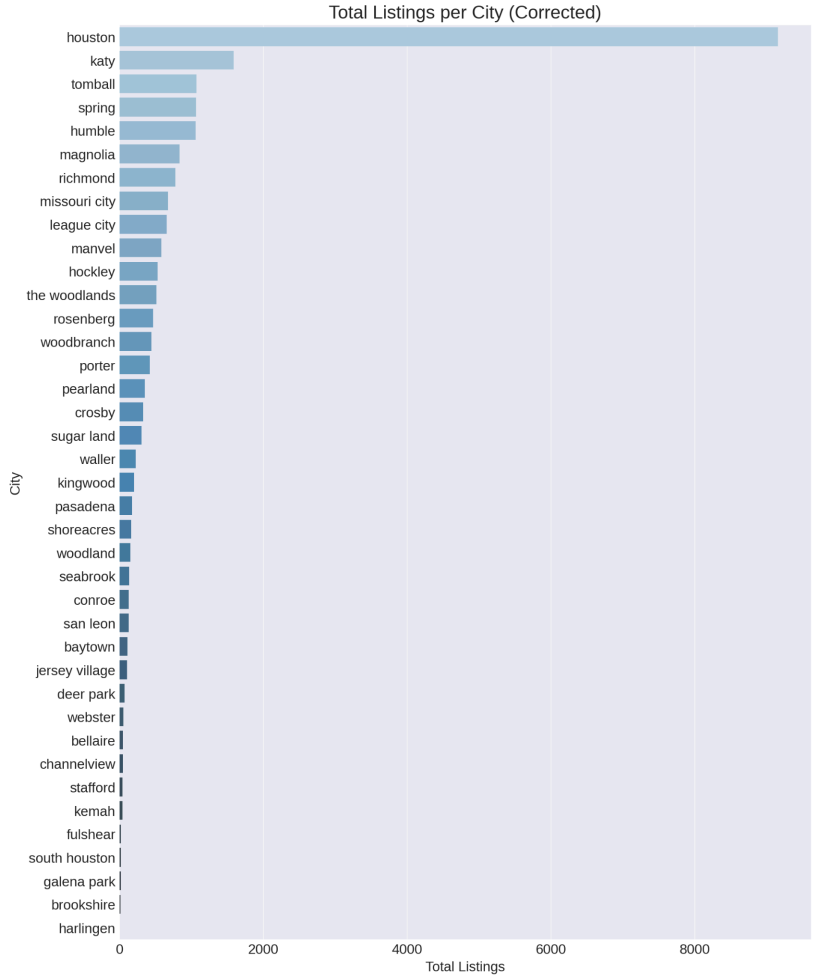


Average Price per Square Foot vs Average Days on Market
In this plot, higher price per square foot seems to be weakly associated with longer market durations, which could reflect that high-end homes take longer to sell.



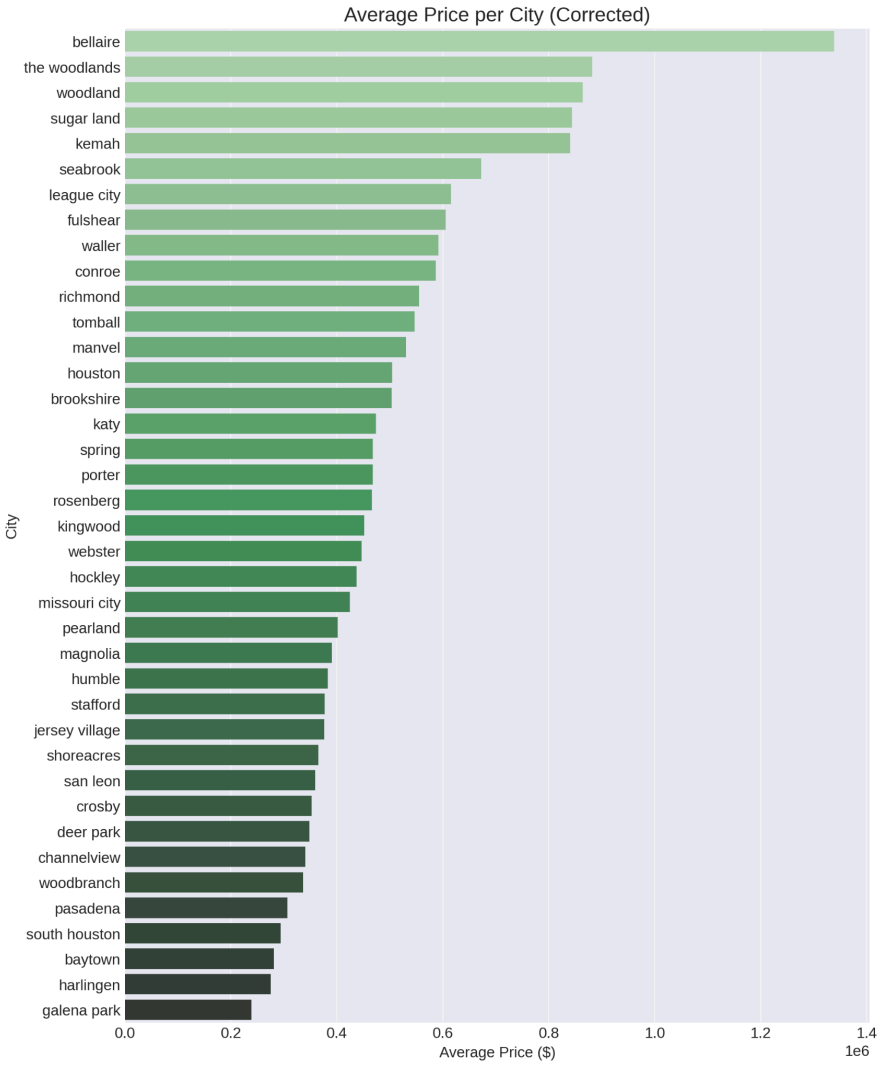
Total Listings per City

Houston dominates the market with the most listings (9,160), far ahead of cities like Katy (1,590) and Tomball (1,071). This highlights Houston as the key player in the region's real estate activity.



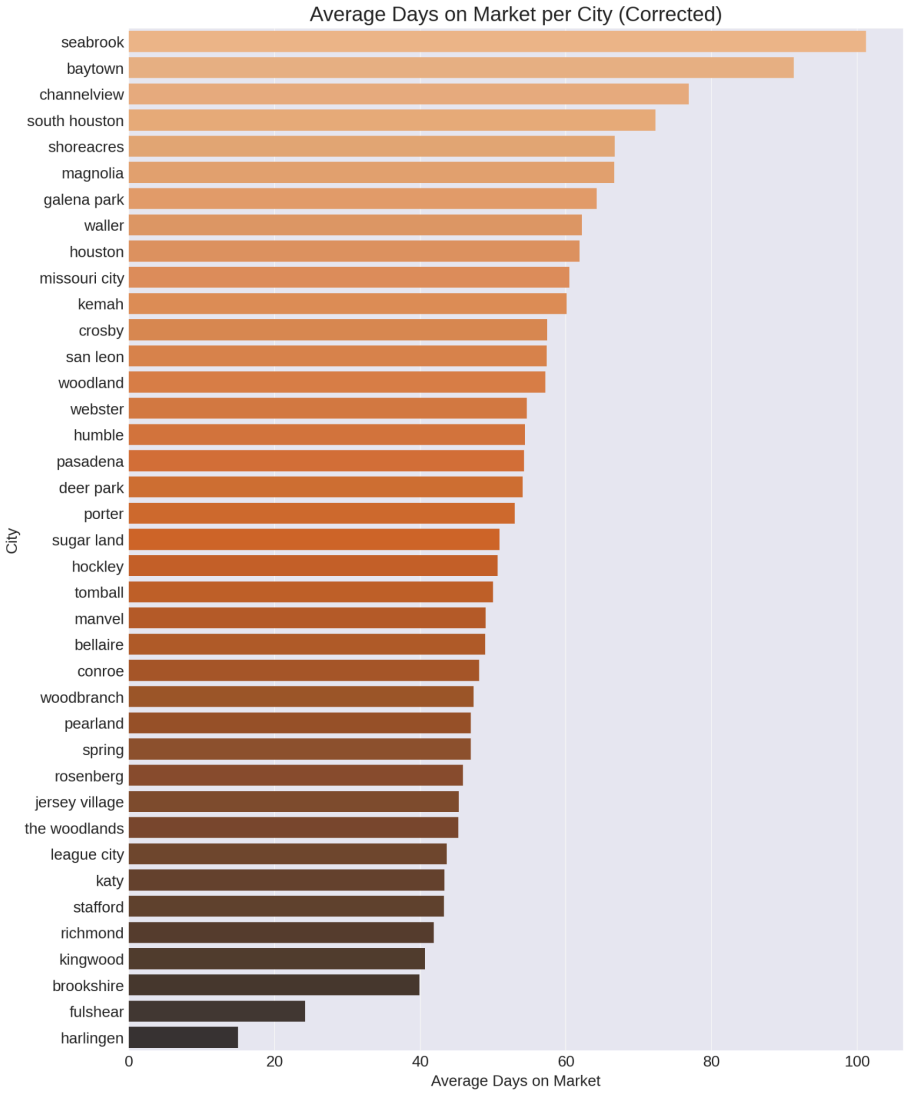
Average Price per City

Cities like Houston and Sugar Land have the highest average prices, suggesting premium markets. More affordable cities include Humble and Webster, which may attract price-sensitive buyers.



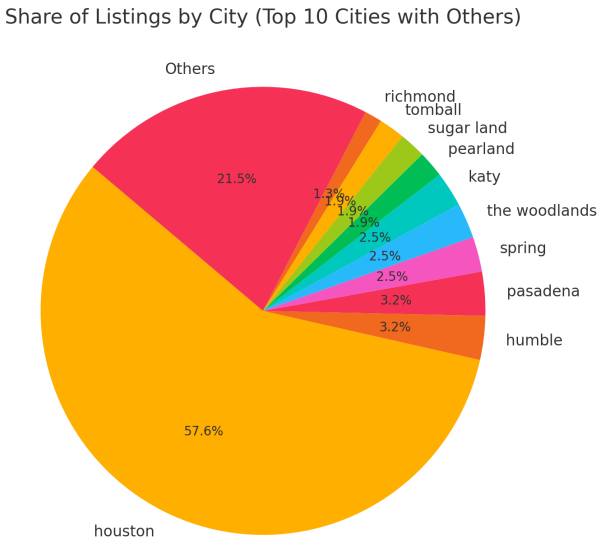
Average Days on Market per City

Prairie View has the highest average days on market, indicating slower sales activity, while Houston has shorter listing durations, reflecting faster-moving real estate and higher demand.



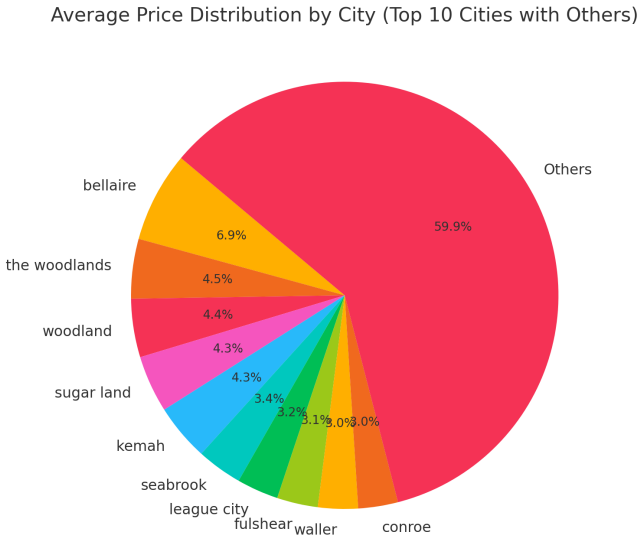
Share of Listings by City (Top 10 Cities with Others)

The top cities, especially Houston, dominate the listings, indicating higher market activity in these areas compared to smaller cities.



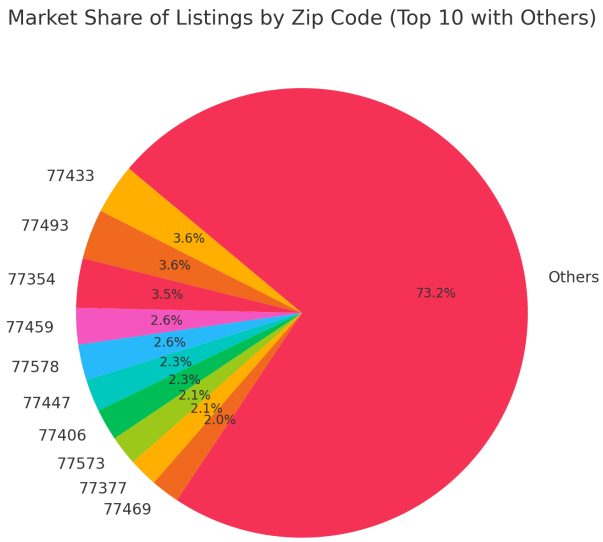
Average Price Distribution by City (Top 10 Cities with Others)

Cities like Bellaire and The Woodlands have the highest average prices, reflecting premium market segments compared to more affordable cities grouped under "Others."



Market Share of Listings by Zip Code (Top 10 with Others)

A few ZIP codes hold most listings, suggesting concentrated demand or larger inventory in specific neighborhoods.

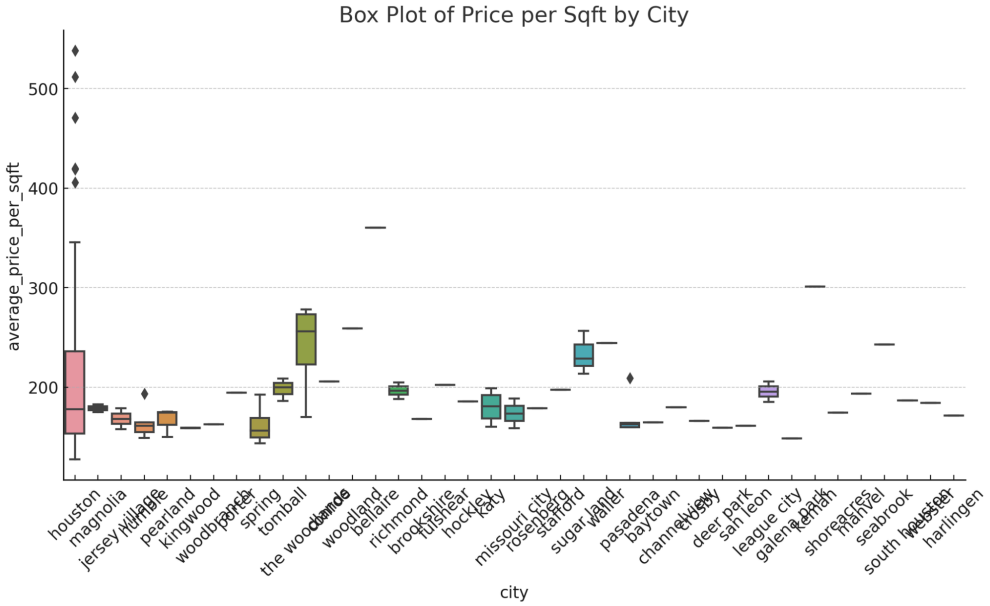


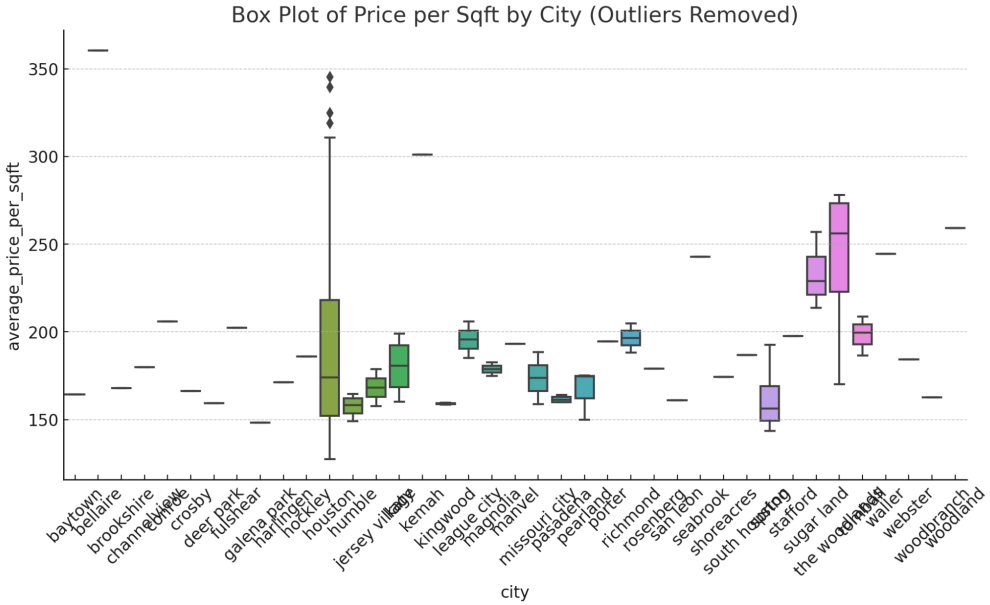
Box Plot of Price per Sqft by City

The box plot reveals how price per square foot varies by city, highlighting median costs and the consistency of prices. Cities with tight ranges and low medians are generally more affordable and have stable pricing, while wider ranges suggest diverse property types and pricing variability. Outliers indicate exceptionally high or low-priced properties, reflecting mixed housing options within a city.

Houston has the widest range (127 to 538), reflecting high diversity in property prices. This may indicate a mix of both high-end and affordable homes.The Woodlands area has one of the highest price ranges (170 to 278), with a median above 256. It suggests a higher-cost market with a few ultra-expensive outliers, likely reflecting premium housing.

Excluding the outliers gives a more refined view of the price per square foot distribution across cities. With outliers removed, it's easier to see the typical price ranges and any remaining differences in price consistency and affordability by city.





Correlation Heatmap of Houston Housing Market Data

The heatmap shows a strong positive correlation between average price and price per sqft, indicating that higher-priced areas tend to have higher costs per square foot. Total listings and average days on market show a moderate positive correlation, suggesting that areas with more listings may experience longer market times. Average price and total listings have little correlation, implying that inventory levels don't directly influence price. Overall, this highlights how premium pricing often aligns with high cost per unit area in the Houston market.