US Rental Market Analysis

Akash Kumar Yadav

Fordham University, Lincoln Center

113 W 60th Street, New York, NY 10023, USA

Data Visualization CISC-6745-L01

Professor Anusha Dandapani

December 8th, 2024

ABSTRACT:

This project analyzes a substantial dataset comprising around 100,000 apartment listings across the United States, sourced from the University of California, Irvine (UCI). The dataset encapsulates a diverse array of attributes, including pricing, amenities, and geographical location, which are pivotal in understanding the dynamics of the apartment rental market. The primary objective of this research is to identify and analyze trends in rental pricing, examine how geographical location influences apartment features and prices, and determine the impact of amenities on rental costs. The study uses multivariate analysis techniques to explore relationships within the data, facilitating the development of predictive models and innovative feature engineering, such as calculating price per square foot and categorizing listings into luxury versus budget segments. The research findings are visualized through a series of static plots, including filled maps, bar charts, treemaps, and line charts, which illustrate the distribution and variability of rental prices across different states and are based on varying apartment features. Additionally, an interactive dashboard and a story created using Tableau will allow users to dynamically filter listings and visualize real-time market trends. This report aims to deliver actionable insights that could benefit renters, real estate professionals, and policymakers, enhancing understanding of the complex factors that drive the US apartment rental market.

INTRODUCTION:

In today's data-driven environment, the ability to analyze and display complex datasets is critical for gaining insights and making sound decisions. This project uses Tableau's advanced visualization features to analyze a real-world dataset of US apartment listings obtained from the University of California, Irvine (UCI). The purpose is to discover rental market factors, such as pricing trends, property availability, and the impact of variables such as pet policies and bedroom counts, across states and localities. By creating an interactive dashboard, users may dynamically explore data using filters such as location, price range, and property categories, resulting in actionable insights. This study emphasizes the transformative significance of data visualization in making complex patterns understandable and fostering successful decision-making.

DATA DESCRIPTION:

The data utilized in this project is sourced from the University of California, Irvine(UCI), and the dataset used in this analysis comprises 99,105 apartment listings across the United States. This dataset is a rich repository of both numerical and categorical data, providing a comprehensive snapshot of the current U.S. apartment rental market. The dataset includes 23 critical features such as pricing, amenities, the number of bedrooms and bathrooms, square footage, and geographical information (latitude, longitude, city, and state).

```
# 1. Display data overview
print(data.head())
print(data.info())
```

✓ 0.1s

```
          id                category                          title  \
0  5668640009  housing/rent/apartment       One BR 507 & 509 Esplanade
1  5668639818  housing/rent/apartment      Three BR 146 Lochview Drive
2  5668639686  housing/rent/apartment  Three BR 3101 Morningside Drive
3  5668639659  housing/rent/apartment            Two BR 209 Aegean Way
4  5668639374  housing/rent/apartment         One BR 4805 Marquette NE

                                        body amenities  bathrooms  \
0  This unit is located at 507 & 509 Esplanade, R...       NaN        1.0
1  This unit is located at 146 Lochview Drive, Ne...       NaN        1.5
2  This unit is located at 3101 Morningside Drive...       NaN        2.0
3  This unit is located at 209 Aegean Way, Vacavi...       NaN        1.0
4  This unit is located at 4805 Marquette NE, Alb...       NaN        1.0

   bedrooms currency fee  has_photo  ... price_display  price_type  \
0       1.0      USD  No  Thumbnail  ...        $2,195     Monthly
1       3.0      USD  No  Thumbnail  ...        $1,250     Monthly
2       3.0      USD  No  Thumbnail  ...        $1,395     Monthly
3       2.0      USD  No  Thumbnail  ...        $1,600     Monthly
4       1.0      USD  No  Thumbnail  ...          $975     Monthly

  square_feet             address       cityname state latitude longitude  \
0         542   507  509 Esplanade  Redondo Beach    CA  33.8520 -118.3759
1        1500      146 Lochview Dr   Newport News    VA  37.0867  -76.4941
2        1650  3101 Morningside Dr        Raleigh    NC  35.8230  -78.6438
...
 21  time          99492 non-null  int64
dtypes: float64(5), int64(3), object(14)
```

Dataset Criteria Satisfaction:

• Multivariate Analysis Capability: The dataset supports complex multivariate analyses due to its diverse range of numerical and categorical variables, making it ideal for predictive modeling and exploratory data analysis.

• Public Accessibility: As a dataset curated by a reputable educational institution (UCI), it is publicly available and meets the project requirements for using nonclassified, accessible data.

• Feature Engineering Potential: The dataset allows for extensive feature engineering, such as calculating price per square foot or categorizing properties into luxury versus budget, enhancing the depth of the analysis.

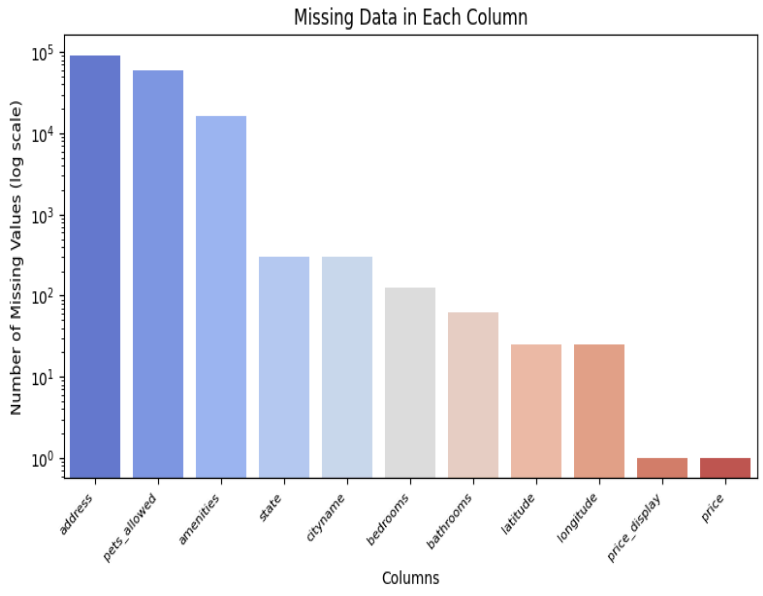
Dependent and Independent Variables:

• Dependent Variable: The primary dependent variable in this dataset is price_display, representing the rental price of each apartment. Most analyses will aim to predict or explain this variable of interest.

• Independent Variables: Independent variables include amenities, bathrooms, bedrooms, square_feet, state, and cityname, among others. These variables are expected to influence the rental pricing and are used to understand variations in apartment costs across different markets.

DATA PROCESSING:

In preparing the dataset for analysis, I implemented several data-cleaning strategies to handle missing values, ensuring both the integrity and usability of the data. Each method was carefully chosen based on the specific characteristics of the data and the analytical requirements of the project:



Missing Data Summary	
Column Name	Missing Values
address	91549
pets_allowed	60424
amenities	16044
state	302
cityname	302
bedrooms	124
bathrooms	63
latitude	25
longitude	25
price_display	1
price	1

1. Filling Missing Values with Default Text:
I filled in missing values for the **amenities** column with 'Not Specified'. This approach allowed me to preserve the dataset's structure while marking the absence of detailed amenities, which might influence the analytical results on how amenities impact rental prices.

In the **pets_allowed** column, I filled in missing entries with 'Unknown'. To maintain consistency within the dataset, I further refined this by ensuring that all entries align with expected categories ("None", "Cats", "Dogs", "Cats, Dogs"). Any deviations were also categorized as 'Unknown'. This standardization is essential for analyzing the impact of pet policies on the rental market.

Missing values in **cityname** and state were filled with 'Unknown', allowing the retention of as many entries as possible for broader analyses while acknowledging the data's incompleteness.

2. Replacing Missing Values with Median:
Missing values in the bedrooms and bathroom columns were replaced with their respective medians. This technique is particularly suitable for handling numerical data where preserving the central tendency is crucial. The median, being robust against outliers, ensures that the replacement does not distort the data distribution, thus maintaining the integrity of the dataset.

3. Dropping Rows with Missing Critical Information:
I removed rows missing critical information such as latitude, longitude, price, and price_display. Given the centrality of these variables to geographic and price analyses, their absence would significantly impair the reliability of subsequent insights.

4. Dropping Unimportant Column:
Due to a high proportion of missing entries, I opted to drop the address column entirely. This decision was based on the judgment that the address details would not significantly impact broad market trend analyses and would simplify the dataset by eliminating a variable with extensive incompleteness.

By employing various cleaning techniques, from median imputation to the elimination of rows or columns, I ensured that the dataset was tailored for specific analytical needs and enhanced for reliability without compromising data validity. This careful preparation facilitates robust and insightful real estate market analysis.

Cleaned Dataset:

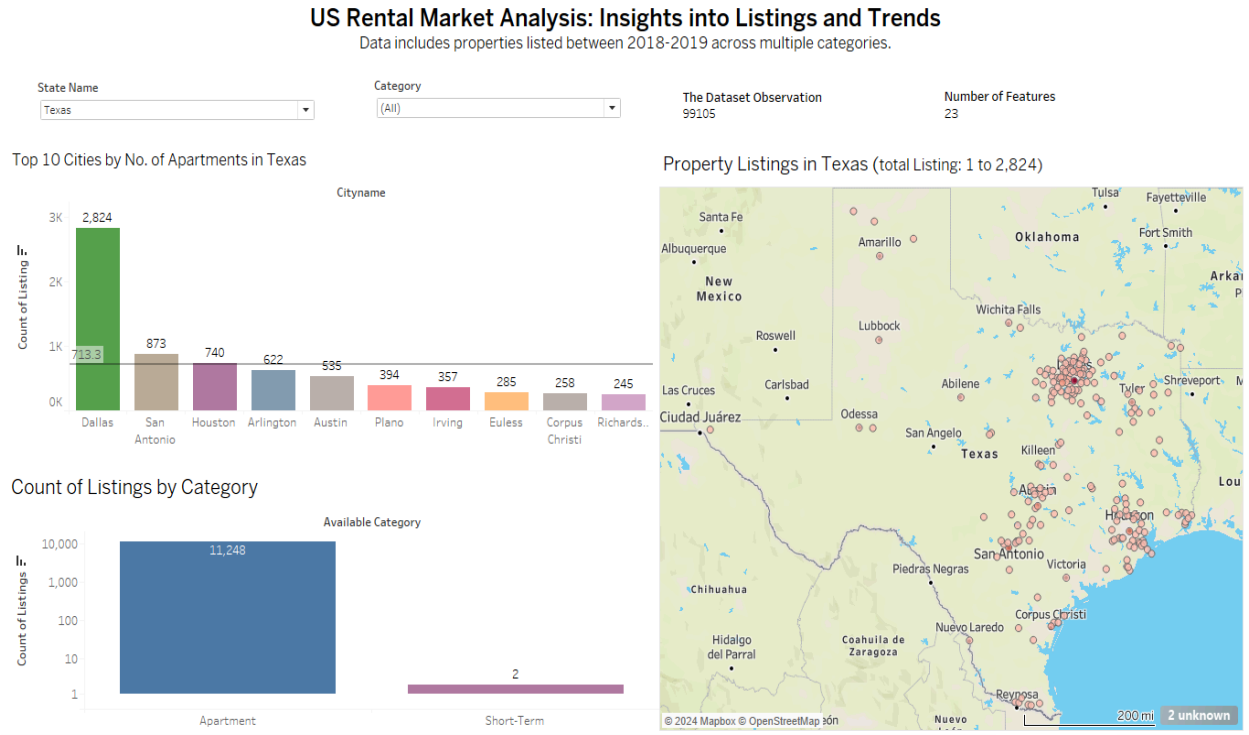
```
          id                category                           title  \
0  5668640009  housing/rent/apartment       One BR 507 & 509 Esplanade
1  5668639818  housing/rent/apartment      Three BR 146 Lochview Drive
2  5668639686  housing/rent/apartment  Three BR 3101 Morningside Drive
3  5668639659  housing/rent/apartment            Two BR 209 Aegean Way
4  5668639374  housing/rent/apartment         One BR 4805 Marquette NE

                                                body      amenities  \
0  This unit is located at 507 & 509 Esplanade, R...  Not Specified
1  This unit is located at 146 Lochview Drive, Ne...  Not Specified
2  This unit is located at 3101 Morningside Drive...  Not Specified
3  This unit is located at 209 Aegean Way, Vacavi...  Not Specified
4  This unit is located at 4805 Marquette NE, Alb...  Not Specified

   bathrooms  bedrooms currency fee  has_photo  ... price_type  square_feet  \
0        1.0       1.0      USD  No  Thumbnail  ...    Monthly          542
1        1.5       3.0      USD  No  Thumbnail  ...    Monthly         1500
2        2.0       3.0      USD  No  Thumbnail  ...    Monthly         1650
3        1.0       2.0      USD  No  Thumbnail  ...    Monthly          820
4        1.0       1.0      USD  No  Thumbnail  ...    Monthly          624

        cityname state  latitude longitude     source        time  \
0  Redondo Beach    CA   33.8520 -118.3759  RentLingo  1577360355
1   Newport News    VA   37.0867  -76.4941  RentLingo  1577360340
2        Raleigh    NC   35.8230  -78.6438  RentLingo  1577360332
3      Vacaville    CA   38.3622 -121.9712  RentLingo  1577360330
4    Albuquerque    NM   35.1038 -106.6110  RentLingo  1577360308

       state_name            datetime
0      California 2019-12-26 11:39:15
1        Virginia 2019-12-26 11:39:00
2  North Carolina 2019-12-26 11:38:52
3      California 2019-12-26 11:38:50
4      New Mexico 2019-12-26 11:38:28

[5 rows x 23 columns]
```

Snapshot of the dataset

	id	bathrooms	bedrooms	price	square_feet	latitude	longitude	time	datetime
count	9.910500e+04	99105.000000	99105.000000	99105.000000	99105.000000	99105.000000	99105.000000	9.910500e+04	99105
mean	5.358114e+09	1.445149	1.728682	1525.507694	955.996045	36.941309	-91.555552	1.559652e+09	2019-06-04 12:39:48.780596736
min	5.121046e+09	1.000000	0.000000	100.000000	101.000000	19.573800	-159.369800	1.544174e+09	2018-12-07 09:20:18
25%	5.197948e+09	1.000000	1.000000	1012.000000	729.000000	33.743600	-104.817100	1.550832e+09	2019-02-22 10:33:41
50%	5.508672e+09	1.000000	2.000000	1350.000000	900.000000	37.213900	-84.549400	1.568745e+09	2019-09-17 18:29:48
75%	5.509006e+09	2.000000	2.000000	1795.000000	1115.000000	39.955900	-77.576700	1.568767e+09	2019-09-18 00:35:46
max	5.669439e+09	9.000000	9.000000	52500.000000	40000.000000	64.833200	-68.778800	1.577391e+09	2019-12-26 20:17:05
std	1.846979e+08	0.547074	0.748731	903.660869	387.732138	4.604559	15.832541	1.105025e+07	NaN

Statistical Summary of Numerical Features

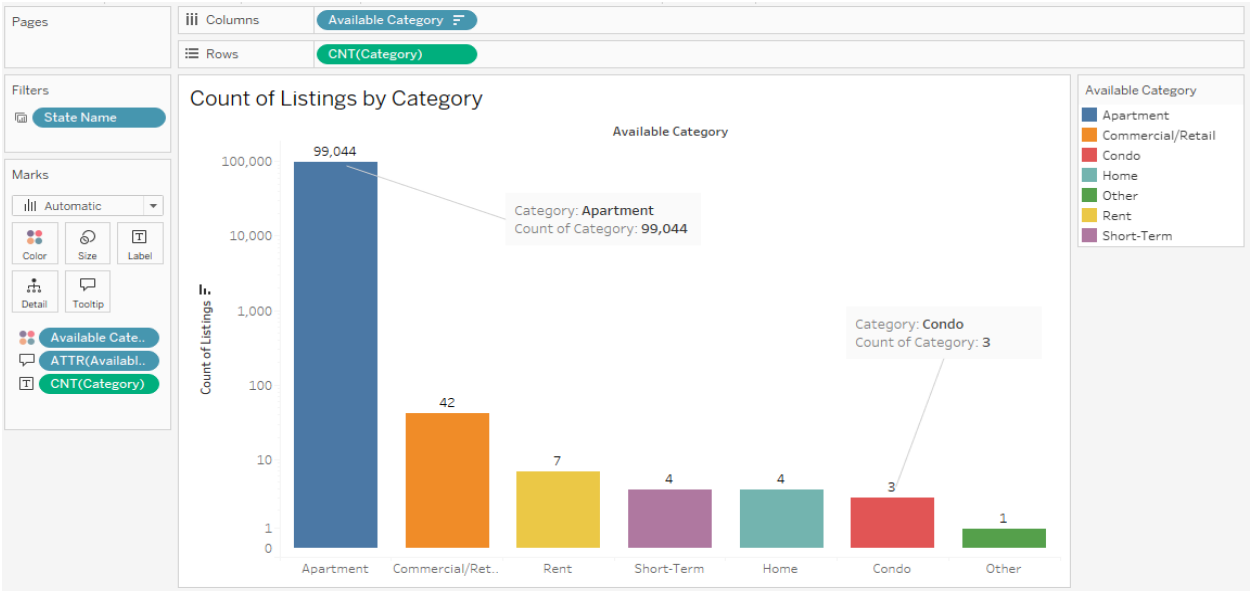
Rental Dashboard



Observations:

The dashboard provides a detailed analysis of the US rental market, focusing on property listings by state, city, and category. In the selected state of Texas, Dallas emerges as the leading city with the highest listings (2,824), followed by San Antonio and Houston, highlighting vigorous rental activity in urban areas. Most listings are apartments, with entries dominating the rental market compared to only two short-term listings, indicating a preference for long-term rentals. The map visualization shows that rental properties are concentrated around major metropolitan regions such as Dallas, Houston, and San Antonio, while rural areas exhibit sparse activity. This dashboard is a powerful tool for understanding rental trends and identifying hotspots in specific states.

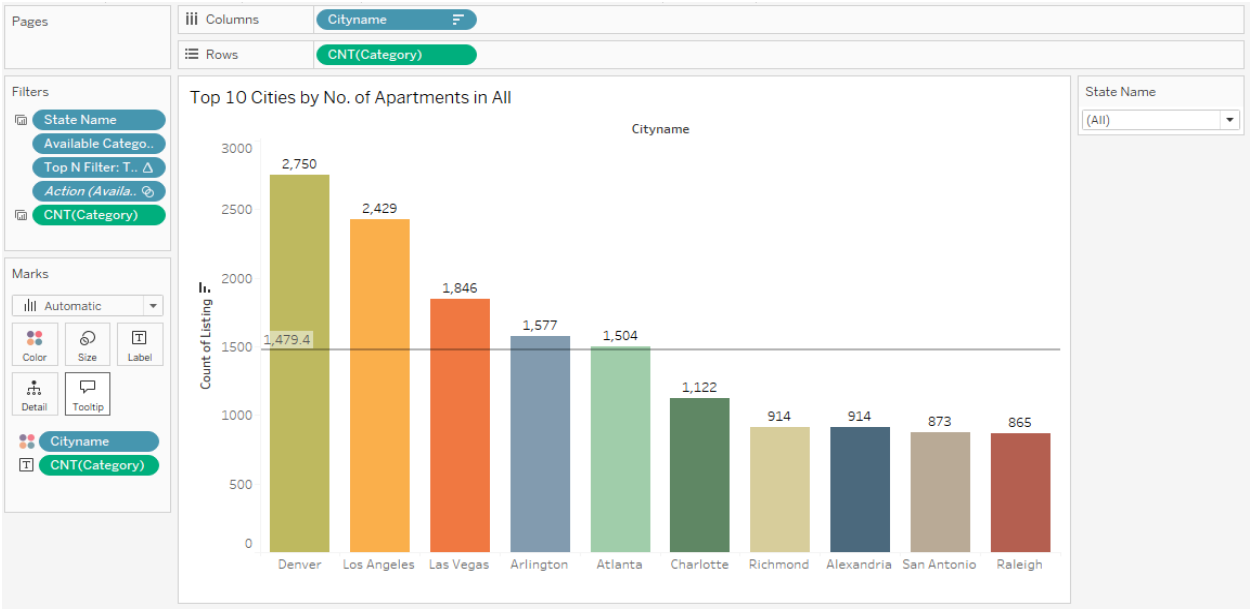
Dashboard-1



Observations:

A dominant number of listings in the 'Apartment' category suggests apartments are the most commonly listed rental property in the dataset. The 'Commercial/Rent' and other residential rental categories like Short-Term, Rent, and Condo are significantly less frequent. The vast difference between the number of apartment listings and other categories indicates a possible focus or higher demand for this type of rental in the market the data represents.

Dashboard-2

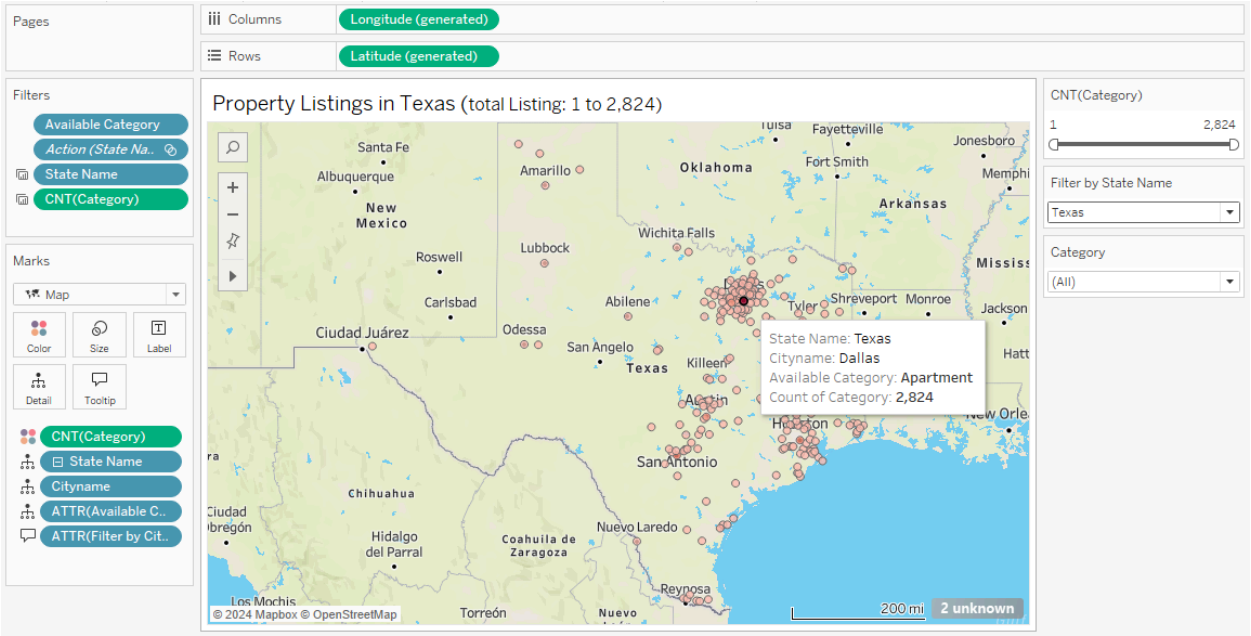


Observations:

Denver leads the listings, indicating a high rental market activity in the city. This could be due to population growth, urbanization, and housing demand. Cities like Los Angeles and Las Vegas also rank highly, reflecting strong urban rental markets. These cities are known for their significant population sizes and diverse housing options.

The bottom four cities (Charlotte, Richmond, Alexandria, San Antonio, and Raleigh) have similar listing numbers, ranging from 865 to 1,122. This suggests a more even distribution of rental activity across these regions compared to the top cities. The high-ranking cities are typically urban centers or metropolitan hubs, emphasizing the strong correlation between city size and rental market activity.

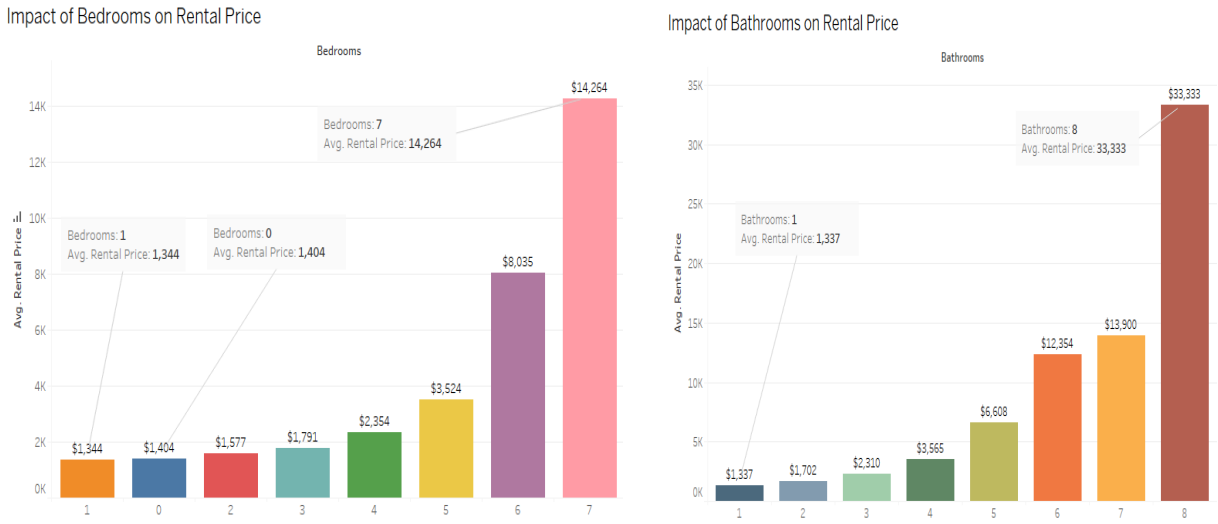
Dashboard -3



Observations:

This map visualization depicts the distribution of property listings across Texas, with a notable concentration in urban regions such as Dallas, which has the most apartment listings. The density of listings reduces dramatically in rural areas, highlighting metropolitan centers' dominance in the rental market. Filters for state names and categories allow users to narrow their view and focus on specific states and property kinds or areas.

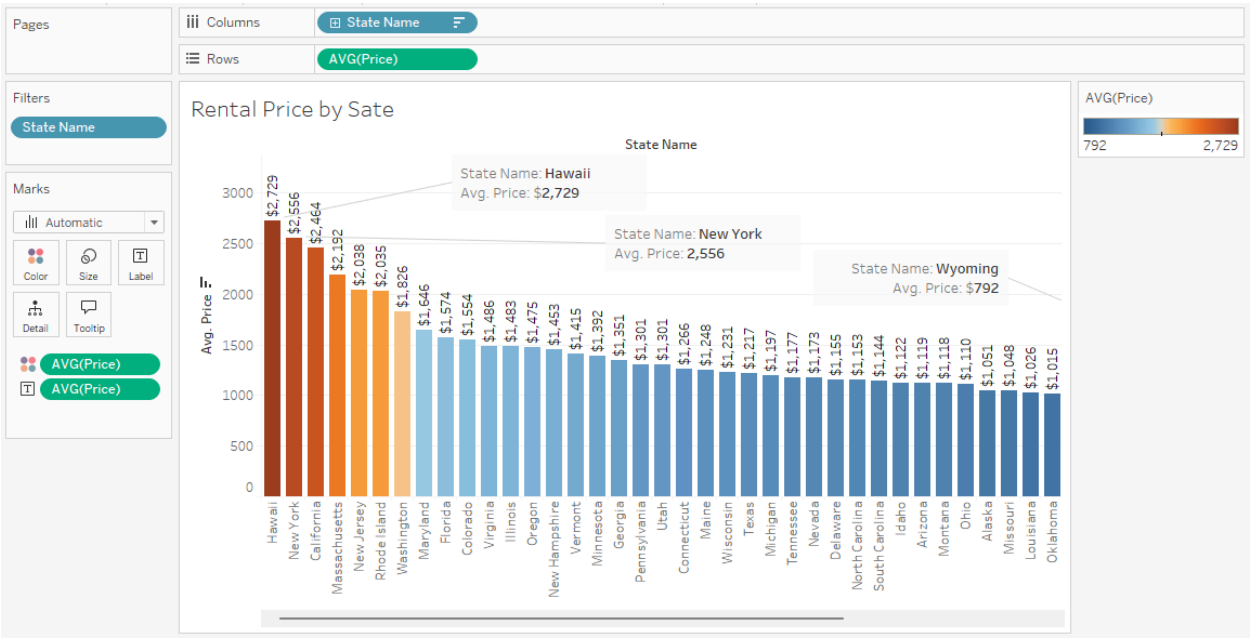
Dashboard -3



Observations:

As their number increases, so does the spread and range of rental prices for both bathrooms and bedrooms, which is evident from the increasing trend. A more significant number of bathrooms and bedrooms correlates with a higher rental price, as shown by the increase in the rental price.

Dashboard -4

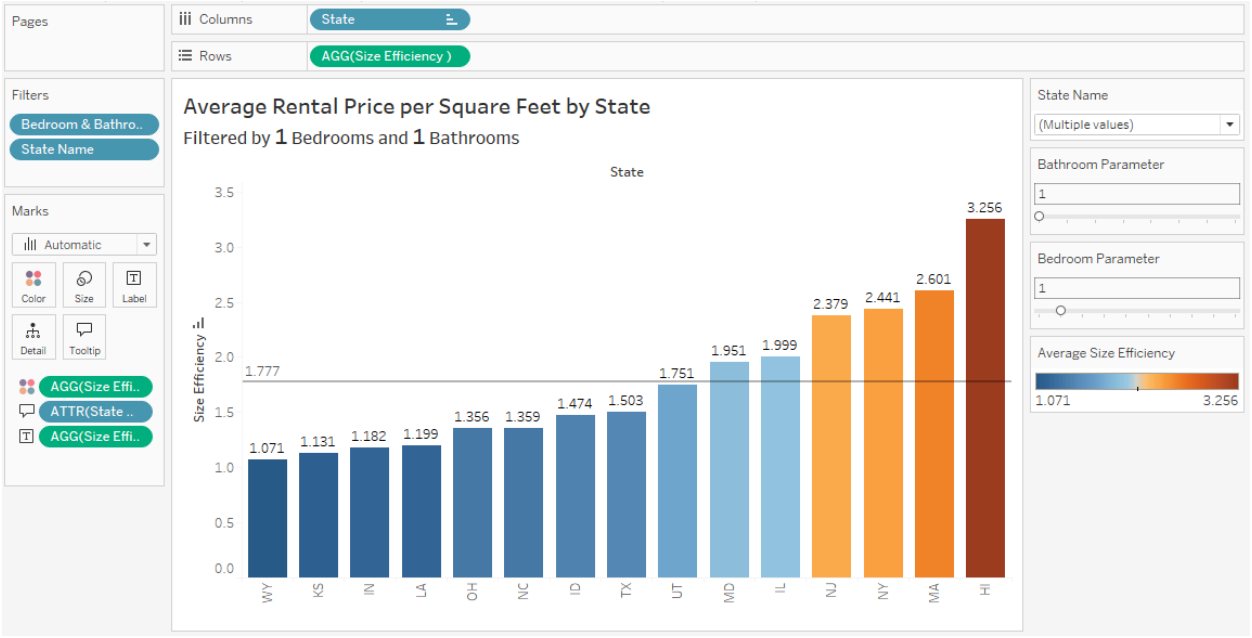


Observations:

The visualization illustrates a gradient of average rental prices among states, with Hawaii (HI) leading, followed by New York (NY) and California (CA), emphasizing the influence of high living costs and metropolitan demand in these areas. The central states have a more moderate range of rental prices, potentially reflecting differences in economic conditions and housing supply and demand.

A general trend suggests a decrease in average rental prices as we move from the West Coast to the East Coast, except for some states like New York (NY), which show high rental averages.

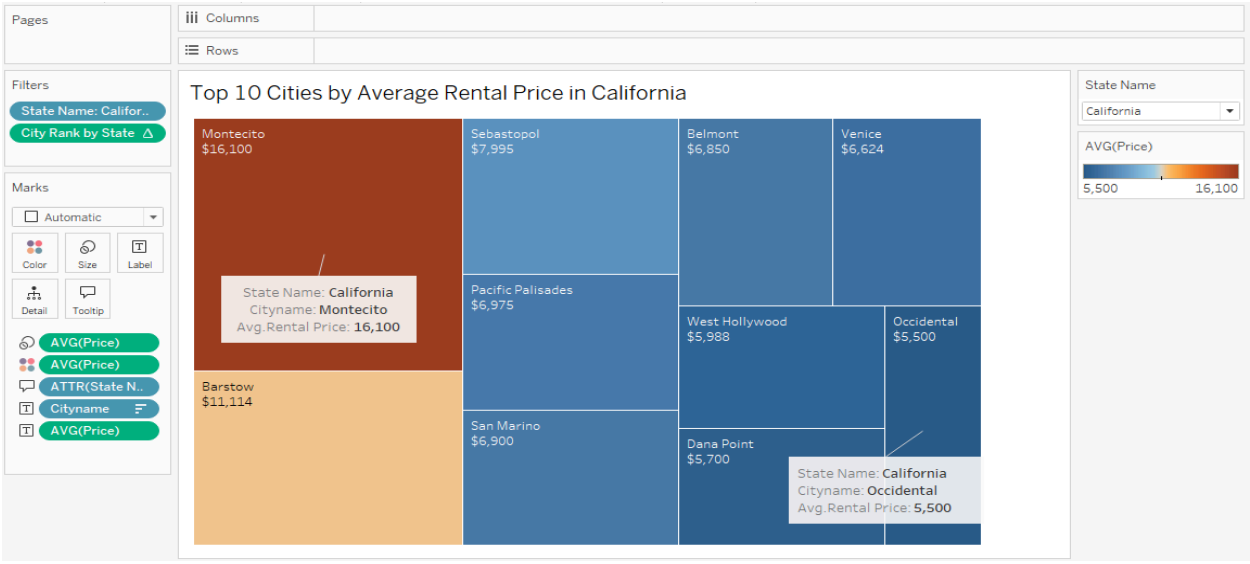
Dashboard-5



Observations:

This graph illustrates the average rental price per square foot in the states for one-bedroom and one-bathroom units. Hawaii has the highest size efficiency at $3.256, followed by Massachusetts and New York, demonstrating the high demand and cost of living in metropolitan, densely populated places. Wyoming and Kansas, on the other hand, have the lowest size efficiency, indicating cheaper rental alternatives in less densely populated areas. The price gradient from low to high illustrates how geography and housing demand affect rental pricing.

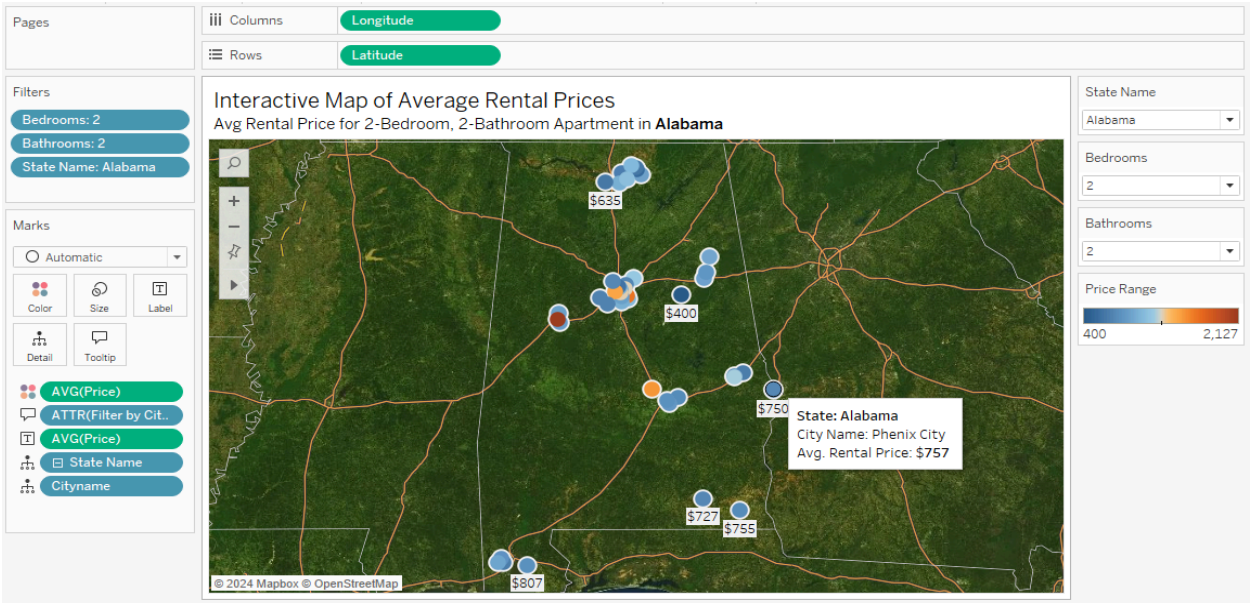
Dashboard-6



Observations:

This chart allows users to filter by any state to see the top ten cities by average rental price per square foot, allowing for state-specific insights. Montecito now has California's highest average rental price, highlighting its exclusivity and luxury market. Occidental and Dana Point, with average rental prices, are more economical choices among the top ten cities. The enormous gap in rental pricing shows the state's distinct housing markets, which range from ultra-luxury to more modestly priced areas, providing vital insights into demand and market segmentation.

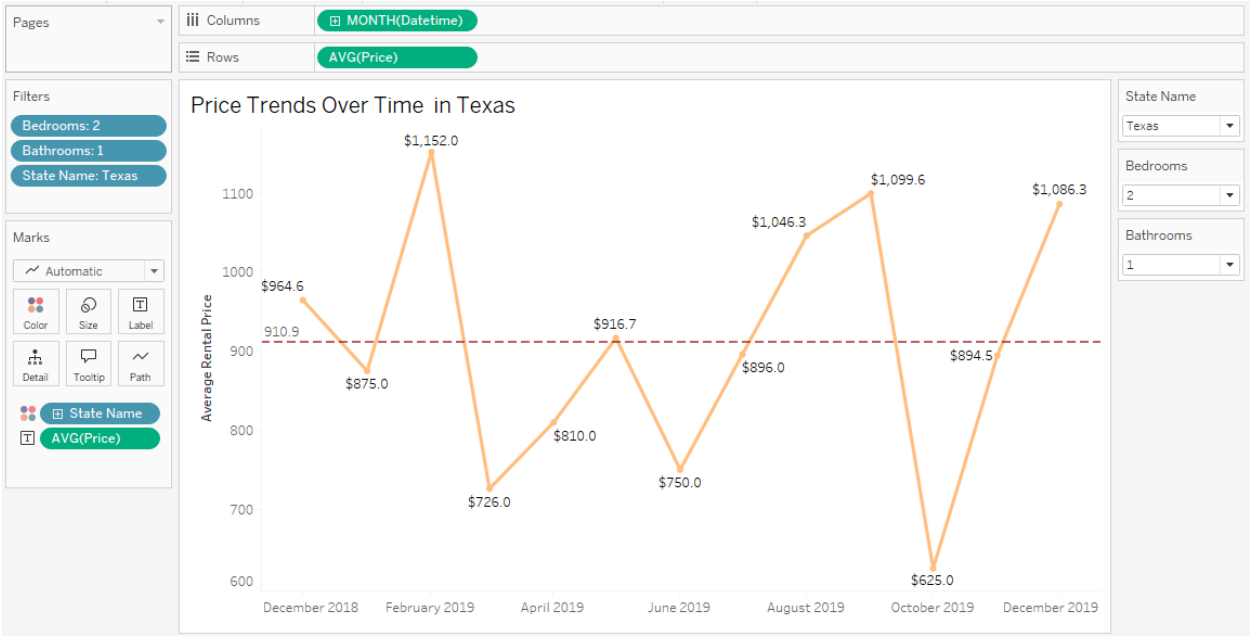
Dashboard- 7



Observations:

This interactive map displays the average rental prices for two-bedroom, two-bathroom apartments in Alabama, which range from $400 to $2,127. Urban places, such as cities near major roads, have higher rental prices, as evidenced by postings for $807 and $750, but more rural areas have much lower prices, such as $400. Users can refine their search using the state and property type filters, making this graphic a useful tool for examining rental trends and affordability throughout Alabama's many regions.

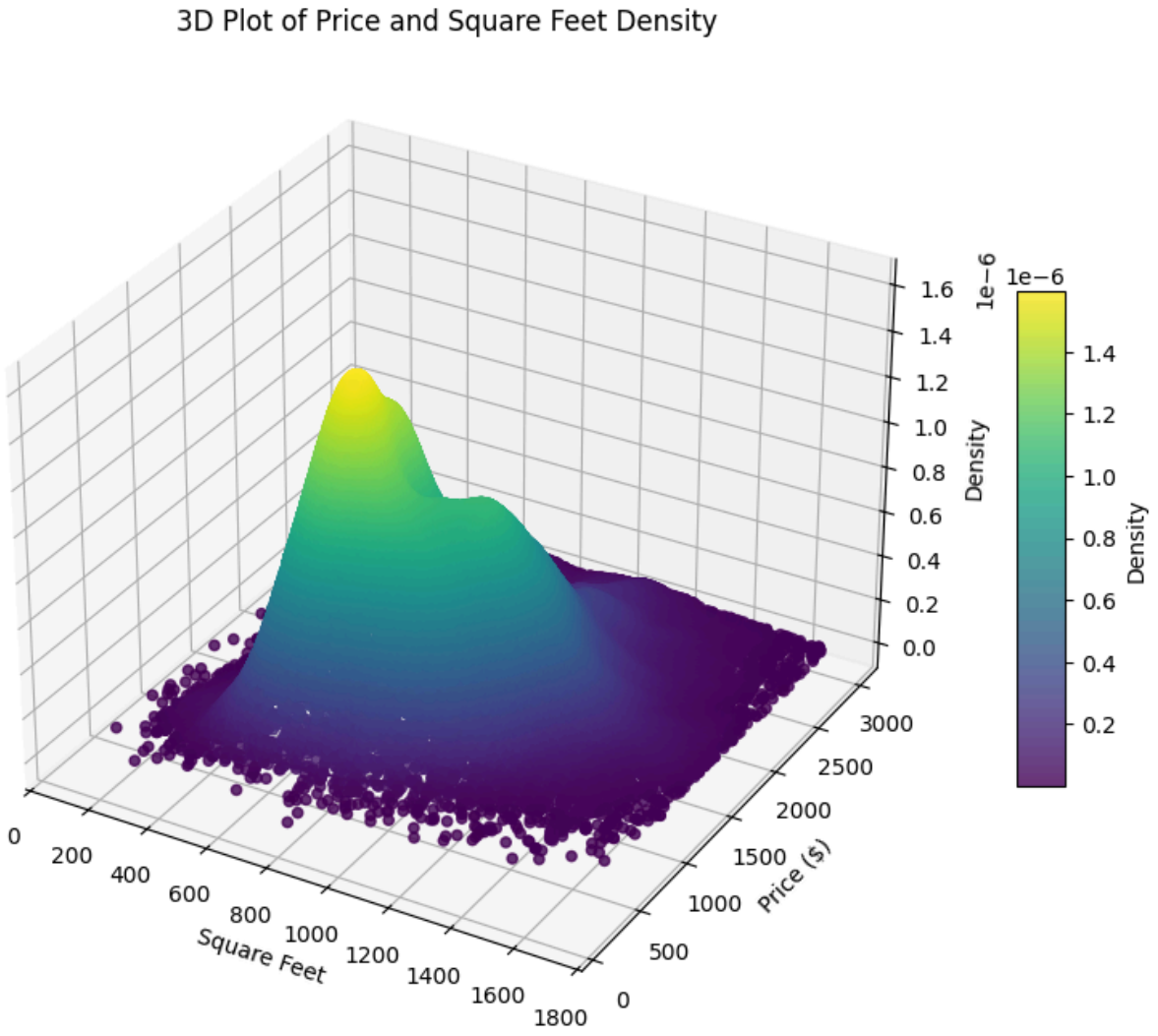
Dashboard- 8



Observations:

This graph depicts the rental pricing trends for two-bedroom, one-bathroom apartments in Texas over time. Prices fluctuate significantly, peaking at $1,152 in February 2019 and plummeting to a low of $625 in October 2019. The average rental price normally ranges between $900 and $1,100, with considerable fluctuation caused by seasonal demand or market shifts. This pattern emphasizes the dynamic character of Texas rental markets, necessitating regular pricing monitoring by renters and investors alike.

Dashboard- 9



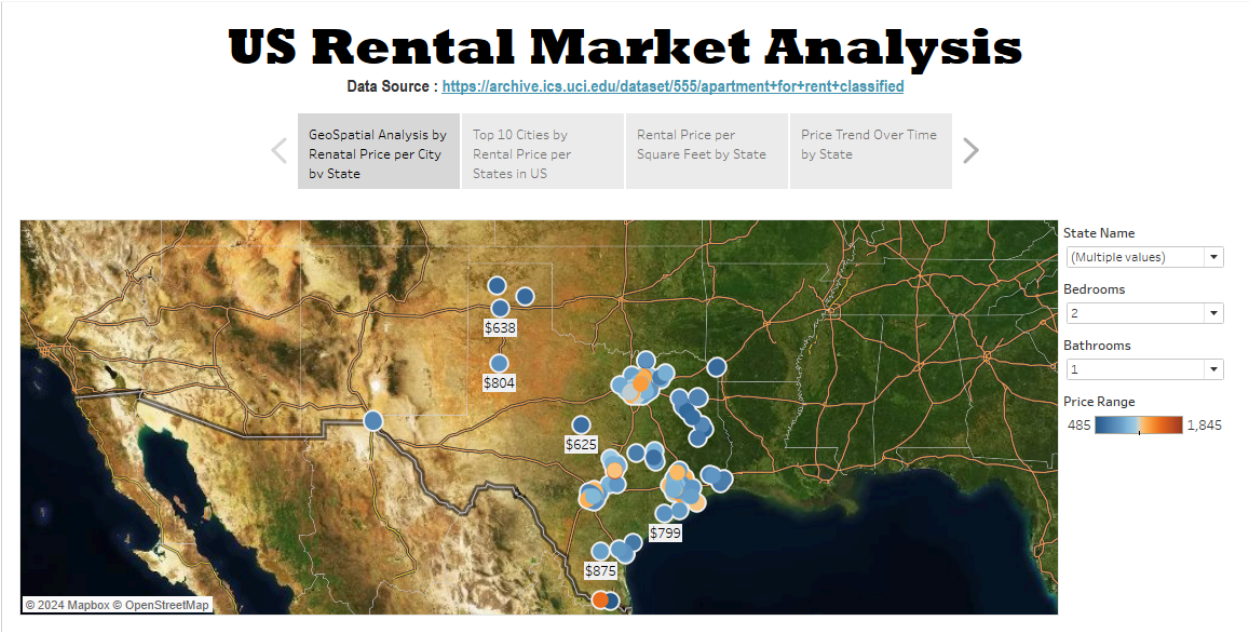
Observations:

The peak density area illustrates the most common combination of price and size at the lower end of both dimensions.

Higher prices show less density, confirming that higher-priced apartments are less common.

The plot also shows that the density does not increase proportionally with size and price, suggesting that other factors may influence the relationship between apartment size and price.

RENTAL STORY:



LIMITATIONS:

1. **Outliers and Skewed Data**:

➔ The extreme values for high bedroom and bathroom counts and luxury rentals (e.g., Montecito in California) heavily influence averages and may not reflect typical market trends. This skews the data and reduces its generalizability.

2. **The granularity of Analysis**:

➔ The visualizations need more detailed breakdowns at micro-levels, such as neighborhood-specific insights within cities or rural-urban disparities within states, which could provide a more nuanced understanding.

3. **Seasonality Impact**:

➔ While seasonal trends are shown, the analysis doesn't delve into the causes of these fluctuations, such as economic changes, holidays, or academic schedules, leaving the trends open to speculation.

4. **Limited Data Context**:

➔ The data doesn't account for factors like property age, amenities, or proximity to key locations (e.g., schools or transportation), which are critical for understanding rental price variations comprehensively.

5. **Geospatial Representations**:

➔ The maps show property density and rental prices but don't account for population density, income levels, or housing demand in different regions, which are crucial to interpreting market conditions.

6. **Static vs. Dynamic Trends**:

➔ The insights provided are static and based on the uploaded data, making them less applicable to future scenarios or evolving rental market trends. Lack of forecasting or predictive modeling limits forward-looking decisions.

Conclusion:

- **Market Profiles**: Most apartments have a modest size and price range, with fewer luxury options.
- **Location Costs**: Coastal and metropolitan areas typically command higher rental prices.
- **Timing Matters**: Rental prices fluctuate seasonally, absent a consistent long-term trend.
- **Feature Premiums**: Additional bathrooms slightly elevate rental prices amidst other influencing factors.
- **Size Matters**: There's a general, though variable, trend of larger apartments attracting higher prices.
- **Market Dynamics**: Visual data analysis reveals intricate patterns of how features and geography drive rental pricing.

REFERENCE:

1. Data Source: https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified
2. Dashboard Idea:
https://public.tableau.com/app/profile/traceywana/viz/ZillowRentDashboard/US_rental