Can clustering, classification, and regression techniques be used to predict rental cost based on square footage, location and amenities in the real estate industry?

Module: Big Data and Business Intelligence

Module Leader: Mansha Nawaz

Student: Olena Moroz
ID: S3063766

Middlesbrough, 2024

Contents

SECTION 1. BUSINESS INTELLIGENCE DESIGN

1. Executive Summary

This report delves into the intricate relationship between property size and pricing, employing advanced visualization techniques to unravel the underlying dynamics. The findings provide a nuanced understanding of how size, location, and various factors interplay to influence real estate pricing.

Moreover, this report presents an insightful spatial analysis of real estate pricing trends, particularly focusing on the contrast between major urban centers and their surrounding suburbs, as well as on a detailed analysis of the correlation between rental prices and the allowance of certain amenities, focusing on pets, gyms, pools, and parking, comprehensive analysis of bathroom, bedroom distribution in the dataset

The data reveals a discernible pattern where major cities like New York or Los Angeles predominantly constitute the category of expensive real estate, contrasting with significantly lower property prices found in nearby suburban areas.

To investigate the correlation, a sophisticated scatter plot approach was adopted, integrating a comprehensive legend to categorize homes by price. Linear regression was applied for visual clarity, incorporating trend and average lines to enhance the interpretability of the data.

Furthermore, diverse histograms were employed to illustrate the trends within various price categories. Leveraging the power of AI-generated charts, these visual aids adeptly elucidate the intricate relationships between price and size, price and location, as well as price and amenities. This sophisticated approach enhances our understanding of the nuanced dynamics shaping the real estate landscape.

Objective:

To understand the factors influencing the rental cost across the US and make the list of recommendations on the base of the insights.

Analysis Goals:

- Determine trends and relationships between determinants of rent cost and square feet, location and additional amenities.
- Examine how additional amenities affect cost of rent for the property of the same square feet.
- Perform clustering analysis to identify distinct groups of properties based on square feet and rent cost.
- Apply classification techniques to predict the category of a property based on its square feet.
- Develop regression models to predict the rent cost based on square feet, bathrooms, bedrooms, and other relevant features.

Key Findings:

A discernible correlation exists between state locations and property prices. California stands out with an impressive average price increase of $823.3, underscoring the state's influence on real estate costs. Similarly, in Massachusetts, properties see a noteworthy uptick in average prices by $765.

States with an average property price exceeding $2,000 include Hawaii, California, Massachusetts, and the District of Columbia.

Major cities, exemplified by New York and Los Angeles, exhibit a trend of higher real estate prices, frequently surpassing the $1750 to $2000 threshold.

Proximity to major cities reveals identifiable suburbs where property prices experience a notable decrease compared to their urban counterparts.

While gyms, pools, and parking facilities are prevalent across various price ranges, their distribution is notably less compared to pet-friendly properties. This indicates that the inclusion of amenities is not as directly correlated with pricing as it is with pet policies.

The dominance of properties with a single bathroom suggests a prevalent preference among buyers or renters for more compact living spaces or a demographic that values economic considerations.

The visualization reveals a discernible relationship between property size and pricing, characterized by a well-defined trend. This provides a visual representation of how these factors interact within the real estate market.

Beyond location and bedroom count, several factors intricately influence property prices. Factors such as the number of bathrooms, overall size of the property, and the inclusion of property images contribute to the cost dynamics.

The number of bedrooms emerges as a significant pricing determinant. Properties boasting more than 3 bedrooms experience an average cost escalation of $399. This insight emphasizes the importance of considering the property's spatial configuration in pricing strategies.

It's crucial to acknowledge the presence of outliers, particularly noticeable in the realm of smaller apartments and houses. This sector exhibits significant variability in pricing, indicating the influence of diverse factors.

Recommendations:

Use the spatial insights to segment the market effectively, developing targeted marketing strategies for urban properties aiming at luxury markets and suburban properties for more price-sensitive segments.

Diversify real estate portfolios by balancing investments in both urban and suburban markets, leveraging the varying risk profiles associated with each.

Align pricing strategies with the demand for specific amenities, considering the observed patterns. Offering pet-friendly options in the mid-price range may prove advantageous.

Property developers and managers can leverage this information to tailor their marketing strategies, highlighting pet-friendly features in mid-range properties to capture a niche market segment.

Continuously monitor bathroom preferences in the real estate market to adapt strategies based on evolving consumer demands.

For property developers, diversifying offerings to include a mix of single and double-bathroom configurations may cater to a broader market and enhance overall portfolio attractiveness.

To conclude, by incorporating these findings into decision-making processes, businesses gain a competitive edge. Tailoring offerings based on the identified factors ensures a more accurate reflection of market dynamics, fostering customer satisfaction and enhancing the overall competitiveness of real estate ventures.

In essence, these key findings not only shed light on the immediate influences on property prices but also provide a strategic roadmap for businesses to navigate the complex and dynamic real estate landscape.

2. Body

2.1. Introduction

To begin with, the rental real estate of the USA is a subject of interest of this research as it is a valuable dataset of the real-world, it provides us with well-structured and well-collected data and, thus, it can be used to analyse the key features of property affecting the rental cost. It could be useful for real estate agents and investors to tale into account the results of this survey when making decisions.

The purpose and motivation behind choosing this ICA is to highlight Business Intelligence for Real Estate to be presented to the stakeholders of: Real Estate Agents, Property Management Companies and BI Exhibition for **TU PowerBI for Women by Women**.

To illustrate the capacity to apply data analysis and visualisation methods to offer insights and make defensible judgements in the healthcare industry. The proposal for this project will incorporate cutting-edge technology into real estate industry. Furthermore, by applying cutting-edge technologies to tackle pressing real estate concerns, this project is in line with the goals of Industry 4.0.

There is some of the interesting abstracts about Big Data in Real Estate we found, which underline the specificity of this sphere and the problems of collecting and analysing real estate data.

Determinants of Rental Rates in Major Cities in the United States (2022): "On average, the rental rate in the United States fluctuates around $750-800…Cities with higher incomes are associated with higher rental rates, which imply that there is a positive relationship between the two variables. The same goes for increasing housing costs and population densities, which relate to increasing rates. Poverty rates also influenced the rental market with increasing percentages equaling increasing rates**…**The availability of the data is a factor that must be noted with dealing with this type of research. Variables such as distances from central areas in cities require more consideration and can probably be analyzed on a local level and not on a national one. Common practices such as rent control or free housing units clearly play a role in manipulating the study. It is difficult to remove these outliers when looking at the data from a national scale. Also, when dealing with data on a national scale, it is important to note that different cities will respond to variables at different intensities, though they will generally follow the same trends overall". (URL)

America's Rental Housing (2022): "Conditions in the rental market are increasingly polarized, reinforcing the stark divide between higher- and lower-income households. While renters with financial resources can choose from a variety of rental options in desirable locations, millions of cash-strapped households struggle simply to find housing anywhere they can afford… Rent levels also vary by region, reflecting differences in household incomes, land values, and age of the housing stock….The rental stock differs across urban, suburban, and rural housing markets. Fully 49 percent of the rental stock was located in urban areas in 2019, a far larger share than the 36 percent in suburban areas and 15 percent in rural areas". (URL)

Real Estate Market Analysis System Based on Big Data (2023): "The real estate industry is a non-linear economic system with a large amount of data and complex structure. The monitoring and analysis of the real estate market is a normalization work that needs continuous improvement of Systems thinking. The construction of a dynamic database for the life cycle of houses enhances the application value of market analysis results in government decision-making". (URL)

Research on Real Estate Information System of the Real Estate Market Based on Big Data Technology (2021): This article examines the significance of creating a real estate information system in the market. It focuses on enhancing data efficiency, supporting marketing strategies, and the ongoing trend of marketing system reform. The author explores

technical aspects such as system analysis, module design, structure planning, database design, data collection methods, and key technology application. **"**The main work content of the establishment of the real estate information system is to collect statistics on all kinds of application data such as real estate area, real estate layout, real estate floor, real estate location, real estate price, etc., so as to lay a good application foundation for the smooth progress of the system". (URL)

2.2. Data Source

Dataset Source(s): **Apartments for rent**
UC Irvine Machine Learning Repository. URL: https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified
Dataset: **Apartments for rent (2019)**
Industry: Real Estate
Name of the Dataset: Apartment for rent classified
This is a dataset of apartments for rent in the USA.
Instances: 10,000.
Features: 22.
Dataset contains diverse attributes such as location, size, amenities, associated costs and originates from the real-world dataset.
Reason for selection this dataset: with background in International Project Management and a strong interest in real estate, my purpose is to conduct an analysis of real estate market trends and patterns, and analyse the factors of forming the rental price
Real-World Relevance: Understanding the cost of rent in different U.S. cities is a practical and real-world concern for many people, including students. It can be a relatable and relevant dataset for analysis and modelling.
Data Exploration and Visualization: Rental data provide opportunities to practice data exploration and visualisation techniques. One can create visualisations to compare rental prices in different locations.
Predictive Modelling: Rental data can also serve as a foundation for predictive modelling. One can build regression models to predict future rent prices based on various features like location, square footage and different amenities.
Other reasons: **Publicly Available, Comprehensive Variables**

2.3. Dataset Source:

Apartment for rent classified // *UC Irvine Machine Learning Repository* // (Donated on 12/25/2019) / URL: https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified (Accessed: 13 October 2023).

2.4. Description

S/N	Column Name	Description	Data type
1.	id	unique identifier of apartment	numeric
2.	category	category of classified	character
3.	title	title text of apartment	character
4.	body	body text of apartment	character
5.	amenities	like AC, basketball,cable, gym, internet access, pool, refrigerator etc.	character
6.	bathrooms	number of bathrooms	character
7.	bedrooms	number of bedrooms	character
8.	currency	price in current	character
9.	fee	fee	character
10.	has_photo	photo of apartment	character
11.	pets_allowed	what pets are allowed dogs/cats etc.	character
12.	price	rental price of apartment	integer
13.	price_display	price converted into display for reader	character
14.	price_type	price in USD	character
15.	square_feet	size of the apartment	integer
16.	address	where the apartment is located	character
17.	cityname	where the apartment is located	character
18.	state	where the apartment is located	character
19.	latitude	where the apartment is located	character
20.	longitude	where the apartment is located	character
21.	source	origin of data	character
22.	time	when classified was created bout each attribute in the data set	integer

Table1 1: Rental Data Columns and Descripitons

S/N	Column	Description
1.	Code	2-Letter abbreviation of States in the US
2.	State	Name of State in the US

Table 2: US-State Codes Table Columns and Descriptions

Dataset Characteristics: Multivariate.
Subject Area: Business.
Associated Tasks: Classification, Regression, Clustering, Prediction.

Screenshots of the dataset is presented below:

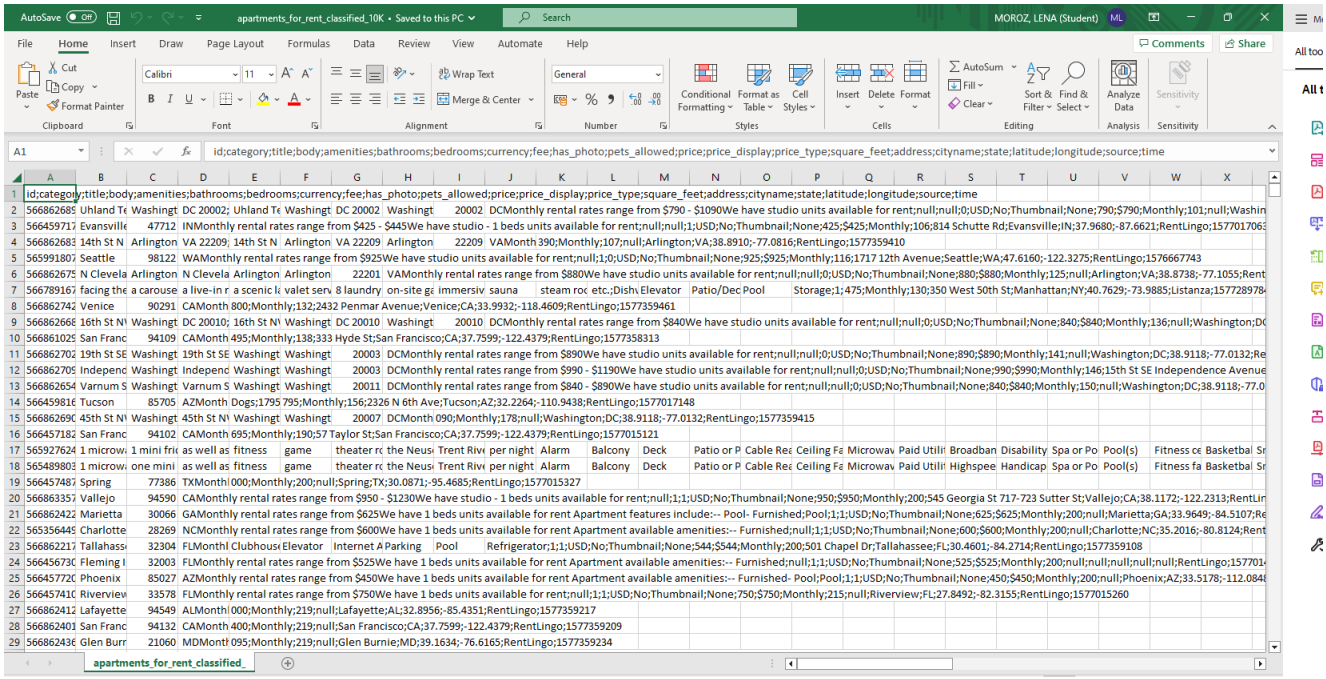


Fig. 1: Rental Data

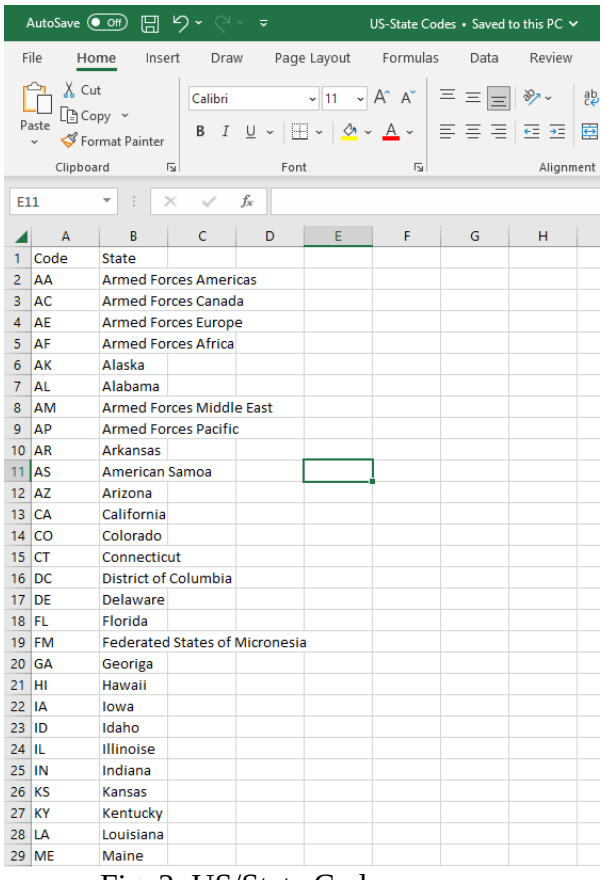


Fig. 2: US/State Codes

2.5. BI Questions

The purpose and motivation behind choosing this ICA is to highlight Business Intelligence for Real Estate to be presented to the stakeholders of: Real Estate Agents, Property Management Companies and BI Exhibition for **TU PowerBI for Women by Women**.

To illustrate the capacity to apply data analysis and visualisation methods to offer insights and make defensible judgements in the healthcare industry. The proposal for this project will incorporate cutting-edge technology into real estate industry. Furthermore, by applying cutting-edge technologies to tackle pressing real estate concerns, this project is in line with the goals of Industry 4.0.

Stakeholders:

- Real estate agents
- Property buyers and sellers
- Property management companies
- Investors in the real estate market
- Potential renters

Objective: To understand the factors influencing the rental cost across the US.

Analysis Goals:

- Determine trends and relationships between determinants of rental cost and square feet, location and additional amenities.
- Examine how additional amenities affect cost of rent for the property of the same square feet.
- Perform clustering analysis to identify distinct groups of properties based on square feet and rental cost.
- Apply classification techniques to predict the category of a property based on its square feet.
- Develop regression models to predict the rent cost based on square feet.

Ramifications:

Data analysis can provide real estate investors and developers with valuable insights into market trends, demand, and pricing. This information can guide investment decisions, helping stakeholders identify lucrative opportunities and avoid potential risks.

Property owners and managers can use data analysis to set competitive and dynamic rent and pricing strategies. By analyzing historical data, current market conditions, and tenant preferences, they can adjust rents to maximize occupancy and revenue.

Local governments can utilize data analysis to inform urban planning decisions. This includes determining where affordable housing initiatives are needed, optimizing land use, and identifying areas where rent control policies may be necessary.

Data analysis can help forecast future real estate trends, allowing stakeholders to adapt their strategies accordingly. For instance, it can predict areas likely to experience rising rents or declining property values, helping individuals make informed decisions about buying or renting.

Property managers can use data analysis to gain insights into tenant behavior and preferences. This information can inform property improvements, amenities, and services to attract and retain tenants.

Business Intelligence Questions and KPIs

Comparison Analysis

1. What is the distribution of property sizes?
2. What is the distribution of property prices by size?
3. Which cities have the highest average price per square feet?
4. What is the distribution of property sizes by number of bedrooms?

5. What is the distribution of property sizes by number of bathrooms?
6. How many properties are listed from each source?
7. What is the average price of properties with photos versus properties without photos?
8. Geographical Evaluation
9. Which states have the highest average property prices?
10. What are the most common amenities in each city?
11. How many properties have a certain price range in each city?
12. What is the average size of properties in each city?
13. How many properties have more than a certain number of bathrooms in each city?
14. What is the average price per square feet for each city?
15. What is the average number of bedrooms and bathrooms for each city?

Social and Economic Divides

1. Which cities have the highest total price for all properties?
2. How does the number of bathrooms affect the price?
3. How many properties are available in each price range?
4. Is there a relationship between rent prices and the presence of a pool, parking, refrigerator, dishwasher and other amenities in the property?

Clustering, Classification and Predictive Modelling

1. How does the size of properties impact the price?
2. Can distinct clusters of properties be identified based on square feet and rent cost?
3. Can a property's category be accurately predicted based on its square feet?
4. How accurately can the rent cost be predicted based on square feet, bathrooms, and bedrooms?
5. Are there specific amenities that significantly affect the rent cost of properties with similar square feet?
6. How can the predictive models be used to optimize pricing strategies for rental properties?

3. Findings based on analysis and evaluation

The dashboard below shows the combination of charts used to answer for the following questions.

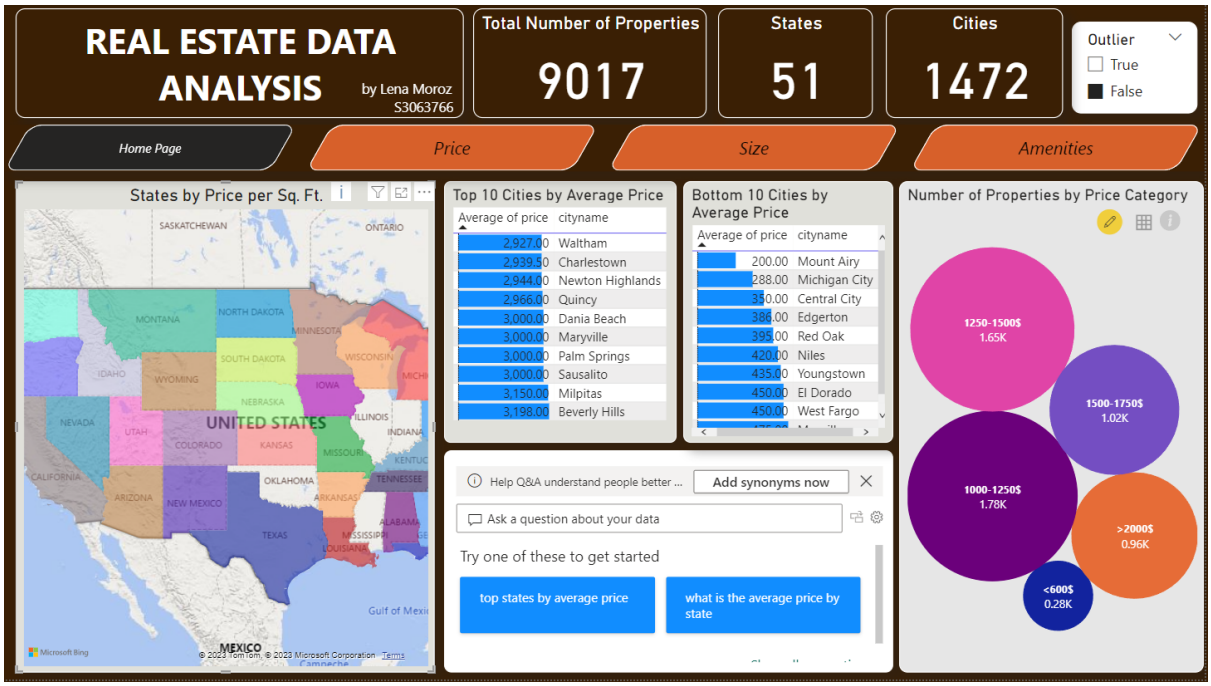


Fig. 1: Page 1 of the BI Report

Question 1: What is the distribution of property by rental cost?

The Bubble Chart was used to visualise the number of properties by price category. In the graph, we observe that the majority of rental properties in our dataset fall within the $1000-$1250 price range. Following with a slight margin, properties in the $1250-$1500 range come next. Rentals priced above $1500 and especially those exceeding $2000 are comparatively less frequent. Similarly, properties priced below $1000 are relatively uncommon. It can be stated that the median rental market value for houses is in the range of $1000-$1250. This insight underscores the sweet spot in pricing for optimal market engagement.

Question 2: What is the average price per squaree feet in each state?

Using 2 of the presented charts – the filled map and the Q&A – we can answer this business question.

Question 3: What are the top 10 cities by the highest average price?

The table reveals that, on average, the top 10 cities include places like Waltham and so forth. This suggests the possibility that these cities may be generally expensive to live in. However, it's also plausible that only one or a few properties from these cities have made it into our dataset, leading to an inflated average cost. To validate this hypothesis, we can employ supplementary charts and tables, as well as utilize the Q&A section provided in the box below. This multi-faceted approach will offer a more nuanced understanding of the cost dynamics in

these cities, allowing us to make informed and strategic decisions based on a comprehensive analysis.

Question 4: What are the top 10 cities by the lowest average price?

The ten cities with the lowest prices also reflect metrics based on the average price presented in this dataset. Furthermore, by clicking on any of these cities, we can visualize their location on the map, identify the corresponding state, and explore the average price per square foot for that state. This interactive feature enhances the depth of our analysis, providing a geospatial perspective on pricing trends. Additionally, for further inquiries and a more dynamic engagement with the data, we can pose additional questions through the Q&A box, ensuring a comprehensive and insightful exploration of the factors influencing real estate prices across different regions. This interactive approach not only enriches our understanding but also empowers strategic decision-making in the realm of real estate business.

Question 5: What is the total number of the properties analysed?

The total number of the properties analysed after the cleansing, processing and pre-processing stages is 9017, which was shown by the card visual using the total function to count the total number of the id.

Question 6: What is the total number of the states analysed?

The total number of the states analysed is 51, which was shown by the card visual using the total function to count the total number of the states.

Question 7: What is the total number of the cities analysed?

The total number of the cities analysed is 1472, which was shown by the card visual using the total function to count the total number of the cities.

The Q&A AI chart can be used to answer the basic questions about the data as ‘top states by average price’, ‘what is the average price by state’ and many others.

The slicer was used to filter the outliers.

There also were used the buttons for navigation in the Dashboard.

The second page of the report is shown below.

It mostly covers the questions about the propery prices.

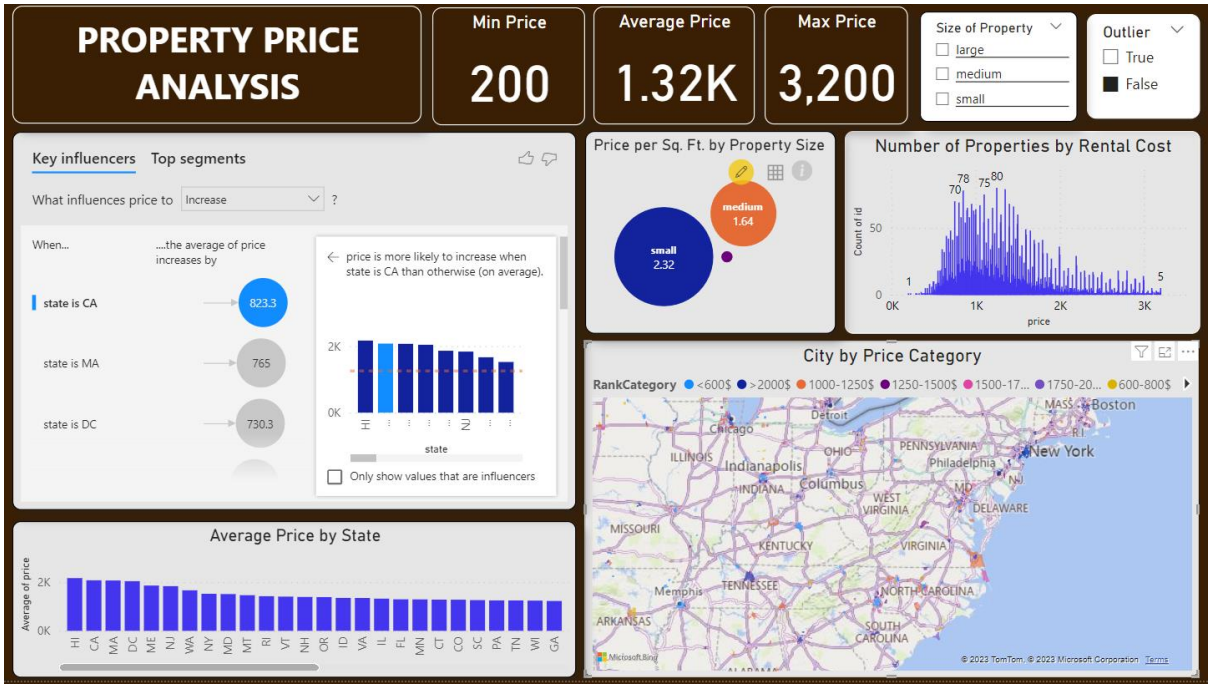


Fig. 2: Page 2 of the BI Report

Question 8: What is the minimum price of the property analysed?

The card visual was chosen to show the minimum price of the property analysed. It is 200$.

Question 9: What is the average price of the property analysed?

The card visual was chosen to show the average price of the property analysed. It is 1,320$.

Question 10: What is the maximum price of the property analysed?

The card visual was chosen to show the maximum price of the property analysed. It is 3,200$.

Question 11: What are the key influencers for increasing of the rental cost?

There is an AI tool which was chosen to answer this question. The Key Influencers chart shows the key feautures of the dataset which impact the price to increase or to decrease.

For example, when the state is California, the average price increases by 823.3$. When the location of the property is Massachusetts, the average price increases by 765$. If there are more than 3 bedrooms in the property the cost increases by 399$. Other features that impact the price are not only the location, but also the number of bathrooms, the size and the availability of the picture of the property.

Question 12: What is the average price by State?

The average price by state is above 2,000$ for Hawaii, California, Massachusetts, District of Columbia. The Key Influencers for these 4 states are:

Fig. 3: Key Influencers for Hawaii

Fig. 4: Key Influencers for California

Fig. 5: Key Influencers for Massachusetts

Fig. 6: Key Influencers for District of Columbia

Question 13: How does the price per squaree feet differ by property size?

To construct the bubble chart, we leveraged precalculated square footage prices for each property in the dataset and correlated them with the property size, categorized as small, medium, and large. Our approach involved exploring additional functionalities to identify the most suitable chart for this visualization. Consequently, we observe that smaller-sized properties exhibit a higher cost per square foot in rentals, ranking second are medium-sized properties, while large properties represent the most cost-effective option per square foot.

This insightful bubble chart not only visually represents the relationship between property size and rental costs but also underscores strategic implications for real estate stakeholders. The nuanced understanding gained from this visualization allows for informed decision-making in the dynamic landscape of property investment and management, offering a valuable tool for optimizing pricing strategies and enhancing overall business performance.

Question 14: What is the number of properties listed in each price category?

To address this inquiry, we opted for a clustered chart, where rental price categories are horizontally arranged, and the vertical axis represents the quantity of homes within each category. In doing so, we observe that the smallest number of homes is rented at prices below $600. The maximum count falls within the $1000 to $1250 range, signifying both a diverse selection and heightened competition for tenants in this price category. Apartments priced between $800 and $1000, as well as those in the $1250 to $1500 range, enjoy nearly equal popularity, reflecting a balanced demand across these segments.

Interestingly, higher-end homes priced at $1750 and above are less common but more prevalent than their counterparts in the sub-$600 category. This nuanced insight into the distribution of rental properties across different price brackets unveils strategic opportunities for property management and investment decisions. It not only highlights market dynamics but also provides a foundation for devising competitive pricing strategies and optimizing portfolio management in the ever-evolving real estate landscape.

Question 15: What is the geographical distribution of the properties in each price category?

On the map, a discernible trend emerges where major cities like New York or Los Angeles typically fall into the category of expensive real estate, with prices often exceeding $1750 or $2000. However, in close proximity to each of these cities, suburbs can be identified where property prices are significantly lower. The map conveniently illustrates the contrast in pricing between suburbs and urban centers, offering a valuable tool for conducting more localized analyses.

This spatial insight not only underscores the typical cost differentials between urban hubs and their surrounding areas but also presents a strategic opportunity for real estate stakeholders. By leveraging this visual representation, businesses can tailor their market strategies based on the nuanced dynamics of specific regions. It allows for a more granular understanding of the real estate landscape, enabling targeted decision-making for property investment, development, and management in both urban and suburban markets.

The third page of the report is shown below.

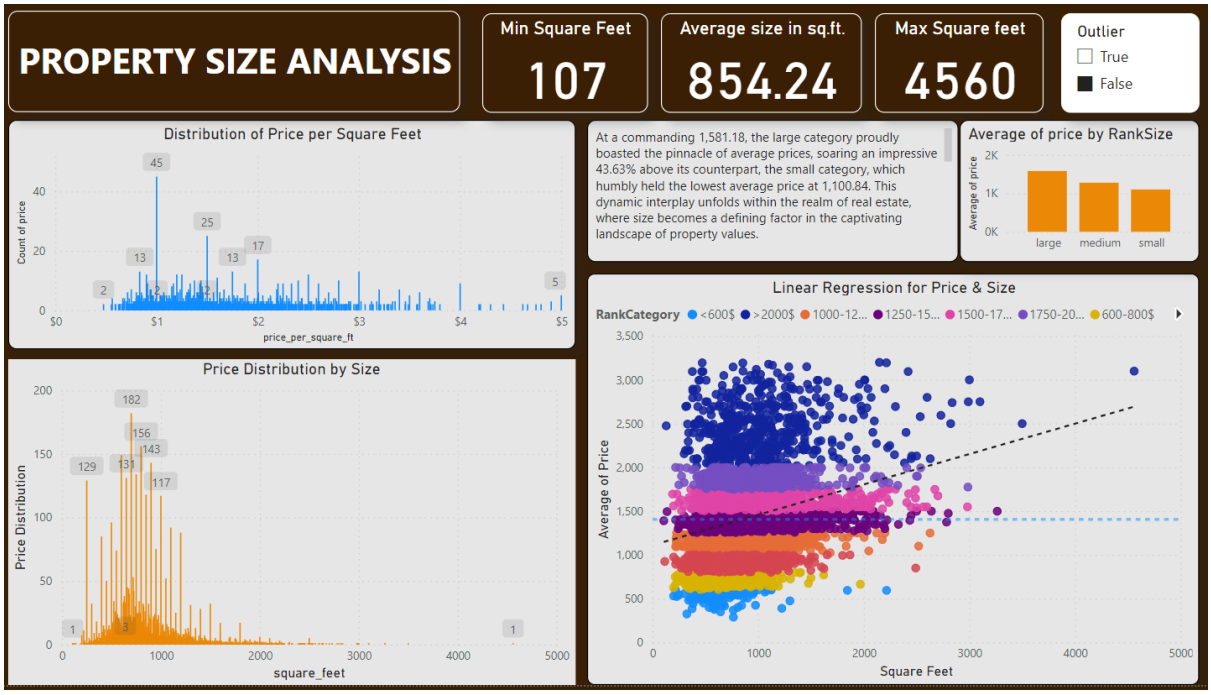


Fig. 7: Page 3 of the BI Report

Question 16: What is the minimum size of the property analysed?

The minimum size of the property analysed is 107 squaree feet. There is a visual card used to show this.

Question 17: What is the average size of the property analysed?

The average size of the property analysed is 854 squaree feet. There is a visual card used to show this.

Question 18: What is the maximun size of the property analysed?

The maximum size of the property analysed is 4560 squaree feet. There is a visual card used to show this.

Question 19: What is the price distribution by size of the property?

The graph reveals a non-normal distribution of prices based on square footage. Notably, there are outliers for homes significantly larger in size than the majority. Additionally, the majority of apartments fall within the range of 300 to 1100 square feet, with a peak centered around 750 square feet.

This deviation from a typical distribution pattern presents an intriguing insight into the diverse landscape of property sizes within our dataset. The outliers suggest the presence of unique, larger properties that could potentially cater to a specific market segment. Meanwhile, the concentration of apartments in the 300 to 1100 square feet range indicates a prevalent market trend, with the peak at 750 square feet highlighting a potential sweet spot in terms of demand or popularity. Understanding these size-related nuances can inform strategic decisions

for property developers, investors, and real estate professionals aiming to align their offerings with market preferences and optimize returns on investment.

Question 20: What is the average price of the property by its rank size?

The data illustrates that the average price is higher for homes falling into the large category, which is quite intuitively expected. The average price for these large homes hovers just above $1500. In comparison, the average price for medium-sized homes is approximately $1200, while for small homes, it stands at $1050.

This pricing trend aligns with the conventional understanding that larger properties often command higher prices. However, these insights provide a quantitative basis for understanding the nuanced relationship between home size and average pricing. Businesses in the real estate sector can leverage this information strategically, tailoring their marketing and pricing strategies based on the size categories that resonate most with their target audience. Such data-driven decision-making ensures a more informed and competitive approach in the dynamic landscape of the real estate market.

Question 21: What is the correlation between price and property size?

Exploring the correlation between property size and pricing, we opted for a sophisticated scatter plot approach with a legend distinguishing price categories for homes. Employing linear regression for visual clarity, we introduced a trend line and an average line into the mix. The visualization effectively unveils a discernible relationship between price and size, marked by a well-defined trend. However, it's crucial to note the presence of outliers, particularly notable in the realm of smaller apartments and houses where the price range exhibits significant variability.

This phenomenon can be primarily attributed to the property's location. In urban settings, smaller apartments are more prevalent, while larger homes are often found outside city limits. Nevertheless, pricing for these properties is influenced not only by size and location but by other factors as well. The robust visual representation not only provides insights into the pricing dynamics but also lays the groundwork for a nuanced understanding of the diverse factors influencing the real estate market. It sets the stage for strategic decision-making, enabling businesses to navigate the complexities of property pricing with a more informed and adaptable approach.

Question 22: What is the distribution of price per square feet?

The predominant pricing paradigm in the real estate market revolves around the one-dollar-per-square-foot mark, representing the largest share of properties. Despite this prevalent trend, the overall distribution exhibits a normal pattern, encompassing diverse properties with varying price points per square foot.

It's noteworthy that a majority of properties conform to the conventional pricing norm, while a select few stand out as premium offerings with prices exceeding $4 per square foot. This nuanced distribution underscores the market's capacity to accommodate a range of pricing structures, providing both affordability and premium options for potential buyers or investors.

The following dashboard provides a detailed analysis of the correlation between rental prices and the allowance of certain amenities, focusing on pets, gyms, pools, and parking. The findings reveal distinct patterns in the property market, offering valuable insights for property developers, investors, and real estate professionals.

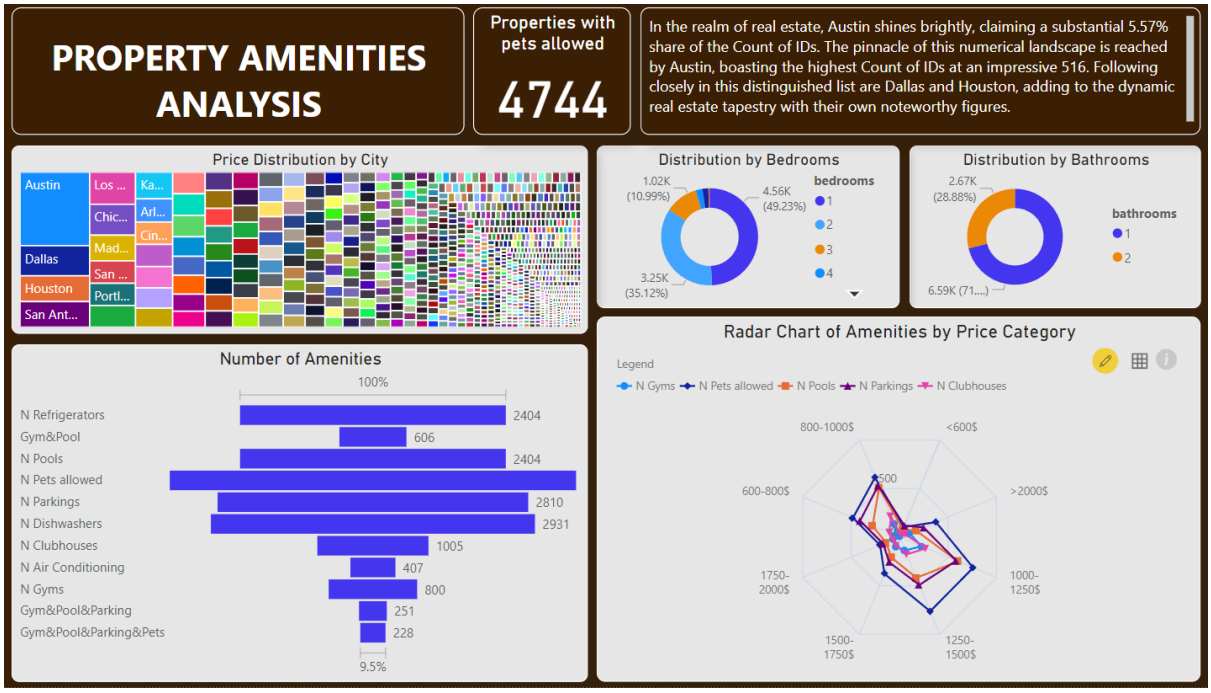


Fig. 8: Page 4 of the BI Report

Question 23: What is the price distribution by city?

We can see that there is a big distribtuion of prices in Austin, then Dallas, Houston and so on. This shows the variety of prices of the analysed dataset.

Question 25: What is the number of amenities available for all properties in the dataset?

The chart reveals that among the over 9,000 homes analyzed in the dataset, more than 3,000 permit living with pets. Additionally, over 2,900 homes are equipped with dishwashers, while more than 2,800 provide parking facilities. Approximately 2,404 homes feature a pool, and an equivalent number come with refrigerators. A noteworthy 1,005 homes boast clubhouses, 800 offer access to a gym, and 407 are air-conditioned. Furthermore, 606 homes provide both a gym and a pool. Interestingly, a mere 9.5 percent have the combined amenities of a gym, pool, parking, and pet-friendly policies. This comprehensive snapshot not only highlights the diverse range of amenities available in the housing market but also lays the groundwork for strategic insights into property preferences and market demands.

Question 26: What is the distribution of properties by bedrooms?

Bedroom Distribution:
49% of the properties feature 1 bedroom.
35% of the properties offer 2 bedrooms.
11% of the properties provide 3 bedrooms.

The donut chart shows the dominance of 1-bedroom properties suggests a significant demand for smaller living spaces, potentially driven by factors such as urbanization, changing demographics, and lifestyle preferences. Developers and investors should consider this trend when planning new constructions or property acquisitions.

While 1-bedroom properties dominate, there is a substantial market share for 2-bedroom and 3-bedroom options. Recognizing the demand for larger living spaces, developers can explore opportunities to diversify their portfolios and cater to a broader range of potential buyers or renters.

Question 27: What is the distribution of properties by bathrooms?

Bathroom Distribution:

- 71% of the properties in the dataset feature 1 bathroom.
- The remaining 29% of properties are equipped with 2 bathrooms.

The dominance of properties with a single bathroom suggests a prevalent preference among buyers or renters for more compact living spaces or a demographic that values economic considerations.

Question 28: How does the availability of different amenities reflect on the rental cost of the property?

The radar chart was used to visualise the amenities available by price category of the property.

A majority of pet-friendly properties fall within the rental price ranges of $600 to $1000 and $1000 to $1500.

Similar to pet-friendly properties, the highest concentration of amenities like gyms, pools, and parking facilities is observed in the $600 to $1500 rental price range.

The concentration of pet-friendly properties in the $600 to $1000 and $1000 to $1500 price ranges suggests a demand for affordable yet accommodating living spaces for pet owners.

While gyms, pools, and parking facilities are prevalent across various price ranges, their distribution is notably less compared to pet-friendly properties. This indicates that the inclusion of amenities is not as directly correlated with pricing as it is with pet policies.

Property developers and managers can leverage this information to tailor their marketing strategies, highlighting pet-friendly features in mid-range properties to capture a niche market segment.

4. Conclusions and Recommendations

Implications and Insights

The dominance of 1-bedroom properties suggests a significant demand for smaller living spaces, potentially driven by factors such as urbanization, changing demographics, and lifestyle preferences. Developers and investors should consider this trend when planning new constructions or property acquisitions.

While 1-bedroom properties dominate, there is a substantial market share for 2-bedroom and 3-bedroom options. Recognizing the demand for larger living spaces, developers can explore opportunities to diversify their portfolios and cater to a broader range of potential buyers or renters.

The distribution of bedrooms can influence pricing strategies. Understanding the preferences for different bedroom configurations allows for more targeted and competitive pricing, optimizing returns on investment.

Real estate professionals should remain agile and responsive to changing market dynamics. As bedroom preferences evolve, staying attuned to these shifts will be crucial for maintaining a competitive edge in the industry.

The dominance of properties with a single bathroom suggests a prevalent preference among buyers or renters for more compact living spaces or a demographic that values economic considerations.

Developers and investors may find this data useful for tailoring their property offerings to align with the demand for single-bathroom residences. This information is crucial for optimizing construction plans and marketing strategies.

Understanding the bathroom distribution allows for a strategic pricing approach. Single-bathroom properties may be positioned differently in the market compared to those with multiple bathrooms, potentially impacting their perceived value and pricing strategies.

The presence of outliers, especially in smaller properties, is primarily attributed to the property's location. Urban settings are characterized by a prevalence of smaller apartments, whereas larger homes are typically situated outside city limits.

Pricing for these properties is influenced not only by size and location but also by a myriad of other factors. The visual representation showcases the complexity of these interactions within the real estate market.

A discernible correlation exists between state locations and property prices. California stands out with an impressive average price increase of $823.3, underscoring the state's influence on real estate costs. Similarly, in Massachusetts, properties see a noteworthy uptick in average prices by $765.

The number of bedrooms emerges as a significant pricing determinant. Properties boasting more than 3 bedrooms experience an average cost escalation of $399. This insight emphasizes the importance of considering the property's spatial configuration in pricing strategies.

Beyond location and bedroom count, several factors intricately influence property prices. Factors such as the number of bathrooms, overall size of the property, and the inclusion of property images contribute to the cost dynamics. A comprehensive understanding of these variables is crucial for nuanced pricing strategies.

States surpassing an average property price of $2,000 include Hawaii, California, Massachusetts, and the District of Columbia. Recognizing the regional variations in pricing provides strategic direction for market positioning and resource allocation.

Recommendations

Customized Marketing:

Develop targeted marketing campaigns that emphasize the advantages of single-bathroom properties, such as cost-effectiveness and simplicity.

Diversification:

For property developers, diversifying offerings to include a mix of single and double-bathroom configurations may cater to a broader market and enhance overall portfolio attractiveness.

Market Research:

Continuously monitor bathroom preferences in the real estate market to adapt strategies based

Pricing and Amenities Alignment:

Align pricing strategies with the demand for specific amenities, considering the observed patterns. Offering pet-friendly options in the mid-price range may prove advantageous.

Targeted Marketing:

Develop targeted marketing campaigns emphasizing pet-friendly features within the identified price ranges to attract potential renters or buyers.

Diversification of Amenities:

Consider diversifying amenity offerings to include a balance between pet-friendly features and other amenities to cater to a broader market spectrum.

Leveraging these insights, businesses can tailor their approach to pricing strategies based on regional nuances and property features. For instance, understanding the premium associated with specific states allows for targeted marketing efforts and optimized pricing structures.

Conclusion

This analysis underscores the importance of understanding the interplay between rental prices and amenities. While pet-friendly properties exhibit distinct price-related trends, other amenities are distributed across various price ranges. Strategic decision-making based on these insights will empower real estate professionals to meet market demands effectively and enhance the overall competitiveness of their property portfolios.

In conclusion, the analysis of real estate data highlights the predominant demand for 1-bedroom properties, with substantial opportunities for diversification into 2-bedroom and 3-bedroom offerings. This strategic insight can guide stakeholders in making informed decisions to capitalize on market trends and enhance their position in the ever-evolving real estate landscape.

This analysis sheds light on the significance of bathroom distribution within the real estate market, offering actionable insights for stakeholders to make informed decisions. The data presented serves as a valuable resource for adapting to changing market dynamics and maximizing returns on real estate investments.

This analysis, integrating advanced visualization techniques, sheds light on the multifaceted relationship between property size and pricing. The findings empower real estate professionals to make strategic decisions based on a comprehensive understanding of the market dynamics. As the real estate landscape continues to evolve, this approach provides a solid foundation for businesses to adapt and thrive in a dynamic and competitive market.

By incorporating these findings into decision-making processes, businesses gain a competitive edge. Tailoring offerings based on the identified factors ensures a more accurate

reflection of market dynamics, fostering customer satisfaction and enhancing the overall competitiveness of real estate ventures.

In essence, these key findings not only shed light on the immediate influences on property prices but also provide a strategic roadmap for businesses to navigate the complex and dynamic real estate landscape.

SECTION 2: BUSINESS INTELLIGENCE SOLUTION
ICA – Appendix: BI Design

---

1. Data Pre-processing Roadmap for Apartments for Rent Dataset

1.1. Loading the Data

The first step in this analysis is loading the dataset. This is done via the 'Get Data' button on the home tab. The data for this project is in CSV format so the Text/CSV option was selected. First, the
Aviation Data dataset was chosen. This opened a dialog box showing the top 20 rows of the data (based on the top 200 rows) with an option to either load, transform data or cancel as shown in Fig. 3

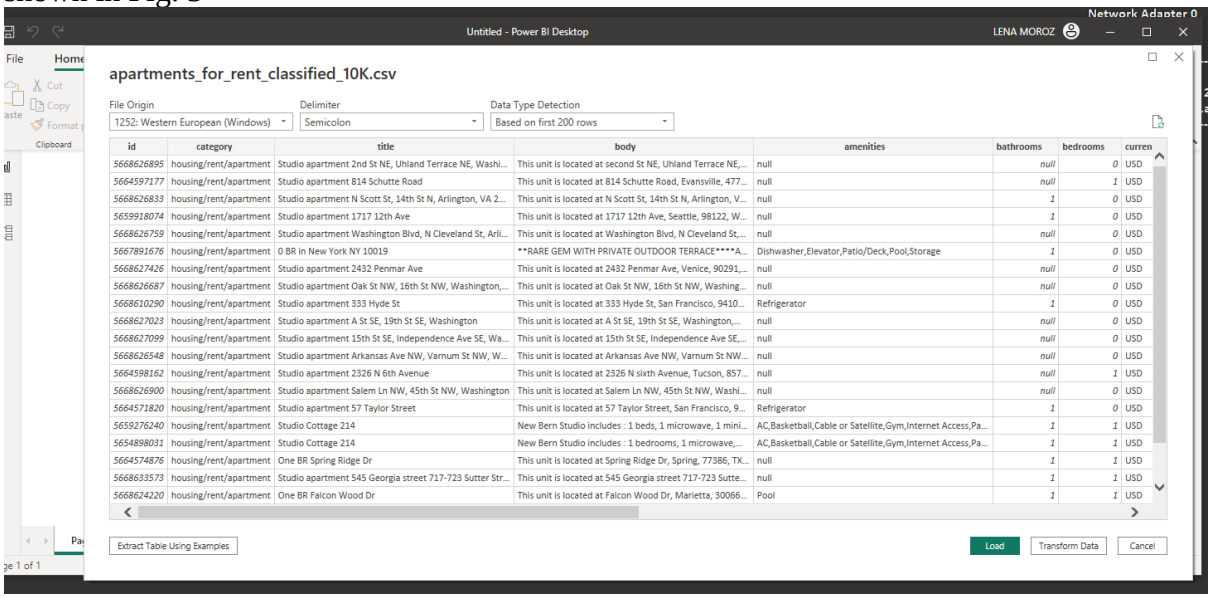


Fig. 3: Dialog Box showing the top 20 rows of the data and options for importation.

After choosing the option to 'Transform Data', the table was successfully loaded in Power BI.

1.2. Data Cleaning Tasks

The next step is the data cleaning and pre-processing which will all be done in Power Query.
Remove Duplicate Entries

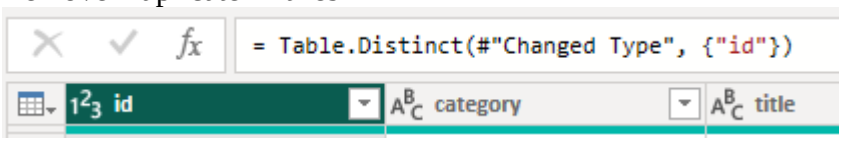


Fig. 4: Formula for removing duplicates

Remove rows with invalid or missing price values.

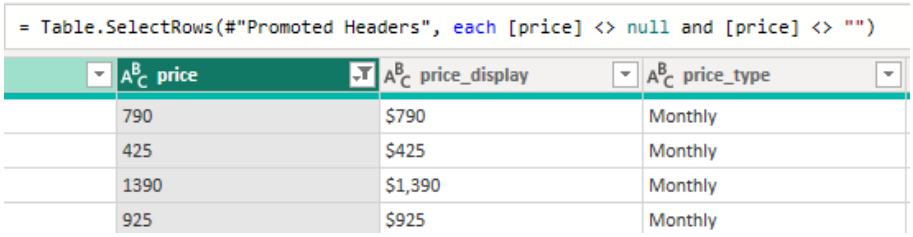


Fig. 5: Formula for removing rows with invalid or missing price values.

Clean Invalid Bathroom Values
Clean Invalid Bedroom Values

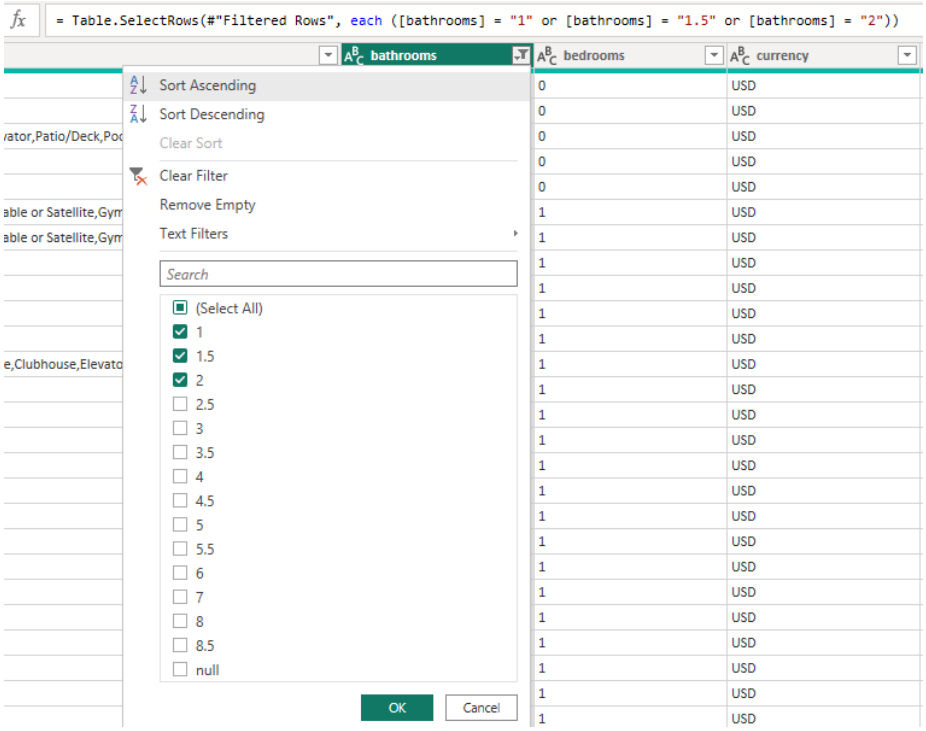


Fig. 6: Cleaning invalid bathroom values

Remove rows with invalid or missing bedroom values.

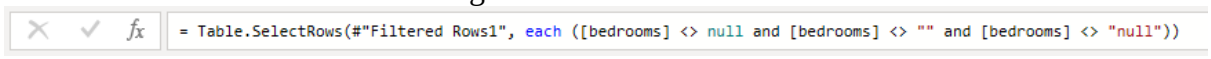


Fig. 7: Removing rows with invalid or missing bedroom values

1.3. Data Transformation Tasks

Change the column ‘category’ into 3 types: ‘apartment’, ‘home’ and ‘short_term’ with the function of replacing values.

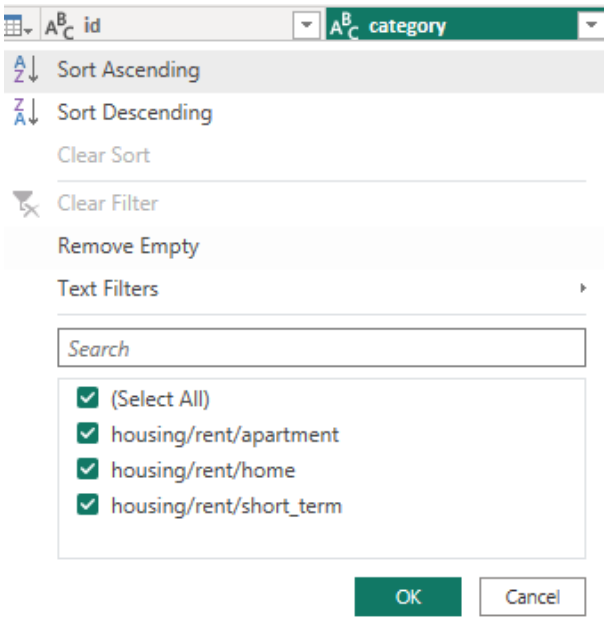


Fig. 8: Types of categories in the dataset

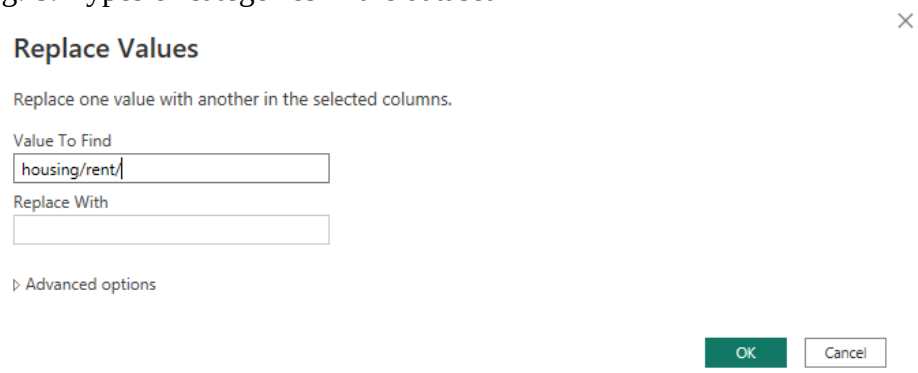


Fig. 9: Replacing values

body
This unit is located at N Scott St, 14th St N, Arlington, VA 22209, Arling...
This unit is located at 1717 12th Ave, Seattle, 98122, WAMonthly rent...
RARE GEM WITH PRIVATE OUTDOOR TERRACE**AVAILABLE IMM...
This unit is located at 333 Hyde St, San Francisco, 94109, CAMonthly r...
This unit is located at 57 Taylor Street, San Francisco, 94102, CAMonth...
New Bern Studio includes : 1 beds, 1 microwave, 1 mini fridge and a b...
New Bern Studio includes : 1 bedrooms, 1 microwave, one mini fridge ...
This unit is located at Spring Ridge Dr, Spring, 77386, TXMonthly rental...
This unit is located at 545 Georgia street 717-723 Sutter Street, Vallejo...
This unit is located at Falcon Wood Dr, Marietta, 30066, GAMonthly re...
This unit is located at Branthurst Dr, Charlotte, 28269, NCMonthly rent...
This unit is located at 501 Chapel Drive, Tallahassee, 32304, FLMonthly...
This unit is located at Leeward Ct, Fleming Island, 32003, FLMonthly re...
This unit is located at Union Hills, Phoenix, 85027, AZMonthly rental ra...
This unit is located at Herons Path Pl, Riverview, 33578, FLMonthly ren...
This unit is located at S. Pond Court, Lafayette, 94549, ALMonthly rent...
This unit is located at Arballo Dr 12d, San Francisco, 94132, CAMonthly...
This unit is located at Elias Way, Glen Burnie, 21060, MDMonthly rent...
This unit is located at W. Ken Caryl Place #d, Littleton, 80128, COMont...
This unit is located at Skipjack Ct, Waldorf, 20603, MDMonthly rental r...
This unit is located at Windmill Cove Dr, Riverview, 33569, FLMonthly ...

Fig. 10: ‘Body’ Column

Remove the column ‘body’ as it has the text description of the property which sums up each whole row.

Remove empty rows in ‘square_feet’.

```
= Table.SelectRows(#"Removed Columns2", each [square_feet] <> null and [square_feet] <> "")
```

Fig. 11: DAX Formula for removing empry rows in Column ‘square_feet’.

Remove the column ‘address’ which has the names of streets and numbers of houses, and where are many empty values.

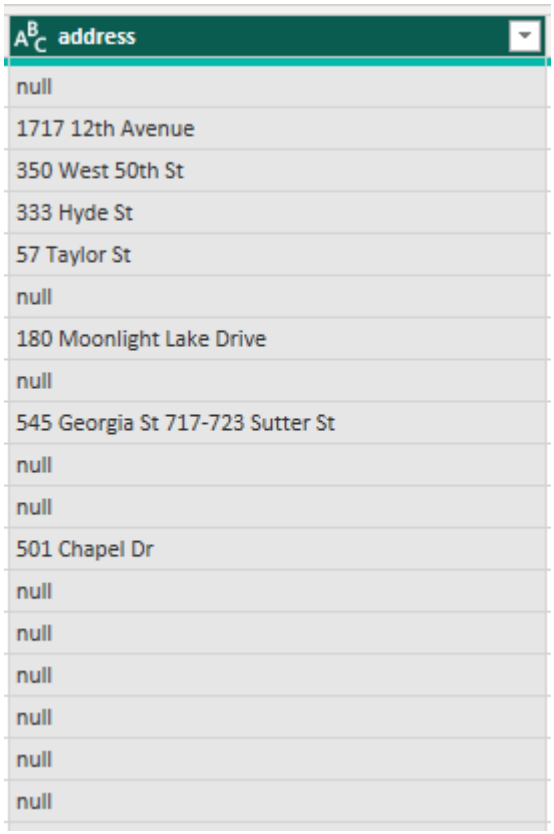


Fig. 12: Screenshot of column ‘address’

Remove empty in ‘cityname’ and ‘state’ columns.

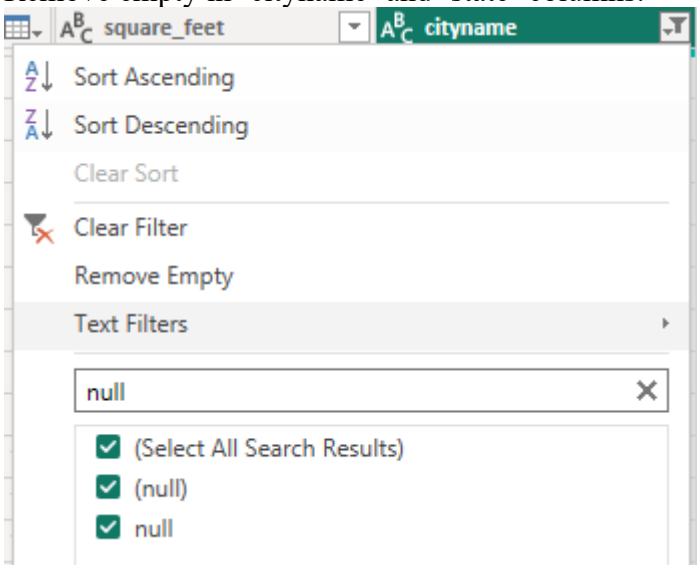


Fig. 13: Removing nulls and empty values in ‘cityname’ and ‘state’ columns

Remove the column ‘time’ as it shows nonvaluable information.

time
1576618076
1576406273
1577015327
1577359918
1577359224
1576322215
1577359108
1577015539
1577015260
1577359217
1577359209

Fig. 14: Screenshot of the column ‘time’

Remove ‘title’ column.

title
Studio apartment N Scott St, 14th St N, Arlington, VA 22209
Studio apartment 1717 12th Ave
0 BR in New York NY 10019
Studio apartment 333 Hyde St
Studio apartment 57 Taylor Street
Studio Cottage 214
Studio Cottage 214
One BR Spring Ridge Dr
Studio apartment 545 Georgia street 717-723 Sutter Street
One BR Falcon Wood Dr
One BR Branthurst Dr

Fig. 15: Screenshot of the column ‘title’

By this time we have applied the following steps, shown on the following screenshot:

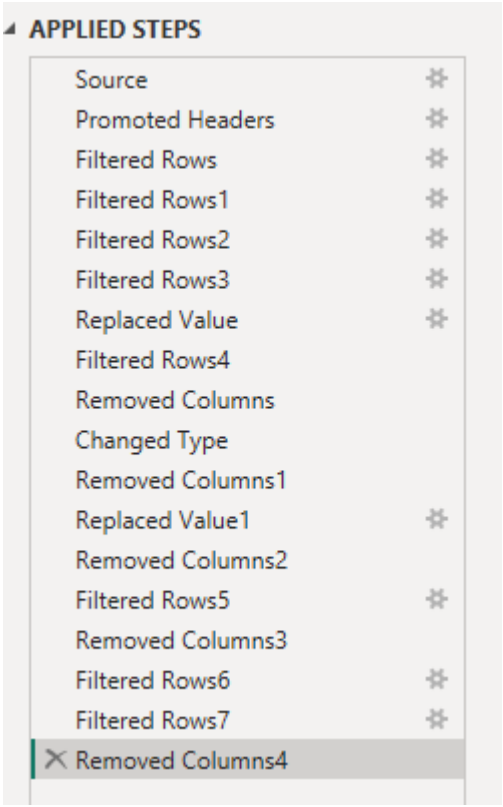


Fig. 16: Applied pre-processing steps

We are now duplicate our main table and create new ones called ‘Location’ and ‘Background’.

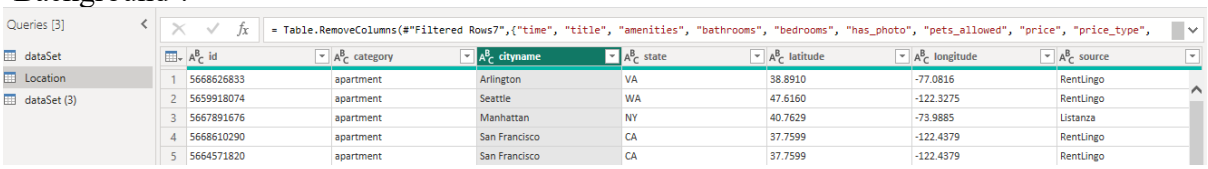


Fig. 17: Duplication of the table ‘dataset’

Close and apply after creating 3 tables.
We have the following autamatical relationships between ctreated tables as is shown on the picture.

Change data types for the column from text to whole numbers.

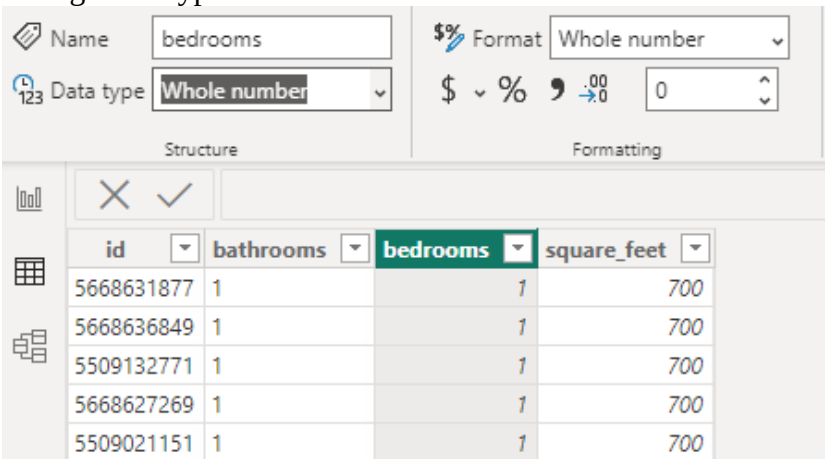
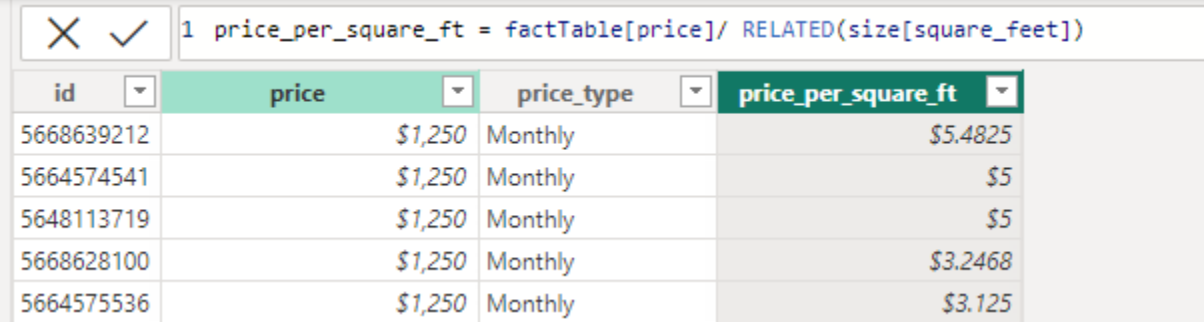


Fig. 18: Changing data type to whole numbers

New Measures:

Price per Square Feet

Calculate the price per square feet by dividing the price by the square feet.



```
1  price_per_square_ft = factTable[price]/ RELATED(size[square_feet])
```

id	price	price_type	price_per_square_ft
5668639212	$1,250	Monthly	$5.4825
5664574541	$1,250	Monthly	$5
5648113719	$1,250	Monthly	$5
5668628100	$1,250	Monthly	$3.2468
5664575536	$1,250	Monthly	$3.125

Fig. 19: DAX Formula for Price per Square Feet

Adjusted Price with Fee

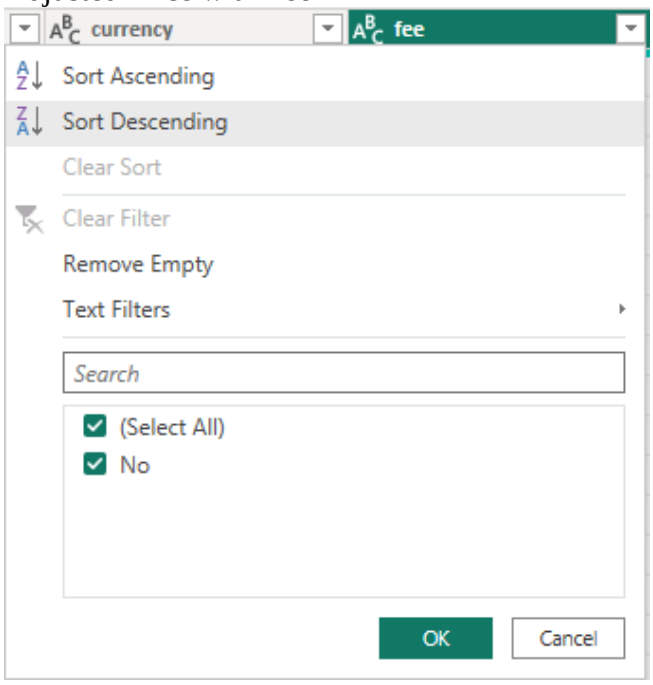


Fig. 20: Screenshot of the column ‘fee’

As no property has a fee, we remove this column.

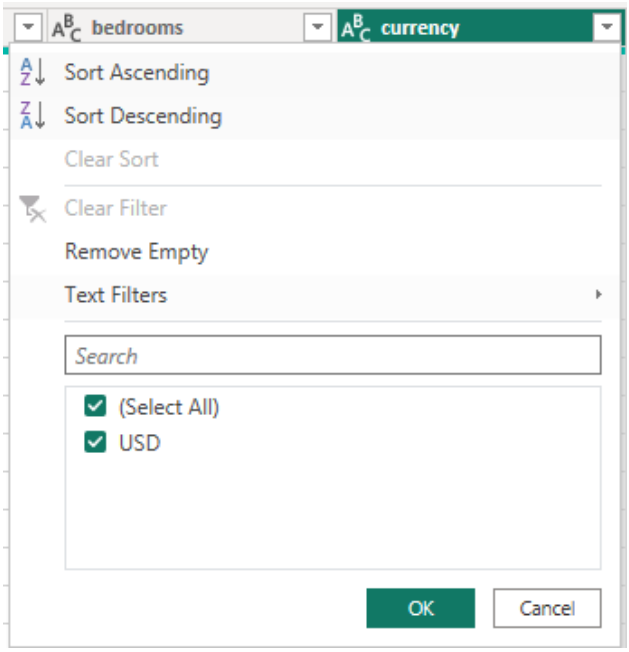


Fig. 21: Screenshot of the column ‘currency’

Remove ‘currency’ column as it is the same for each property and does not make sense.

```
= Table.TransformColumnTypes(#"Removed Columns",{{"price", Int64.Type}})
```

Fig. 22: DAX Formula of removing the ‘currency’ column

Change type of the data from text to integer numbers.
Then remove the column ‘price_display’.
Remove category as it has only 2 rows in category ‘home’, others are ‘apartments’.

```
= Table.SelectRows(#"Removed Columns1", each ([category] = "home"))
```

	id	category	price	price_type
1	5663106093	home	475	Monthly
2	5663111837	home	625	Monthly

Fig. 23: Screenshot of ‘home’ value in ‘category’ column

Create a new column 'has_photo_indicator' to indicate if a property has a photo or not.

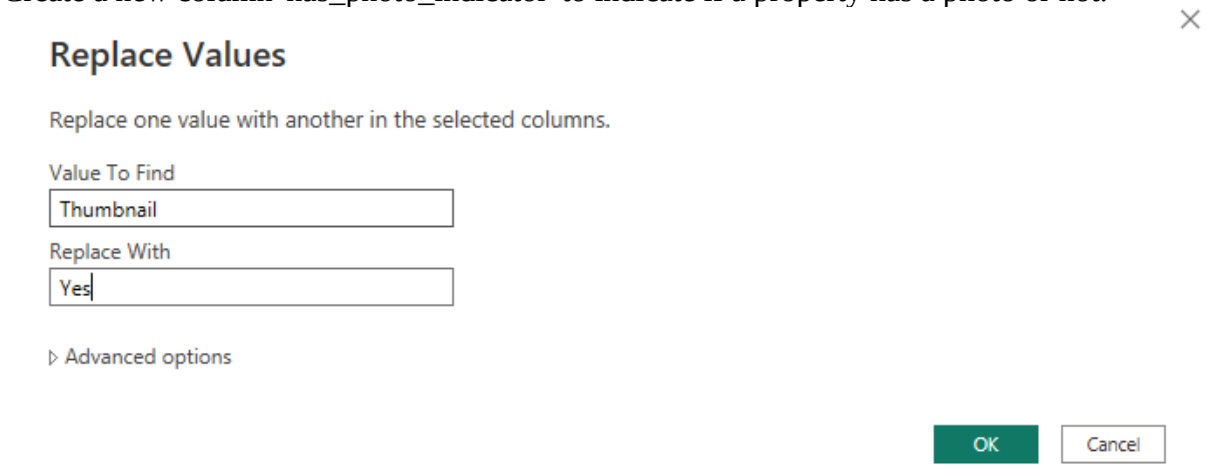


Fig. 24: Replacing values in clumn ‘has_photo’

Remove the only row in the filtered column ‘price_type’ by Weekly:

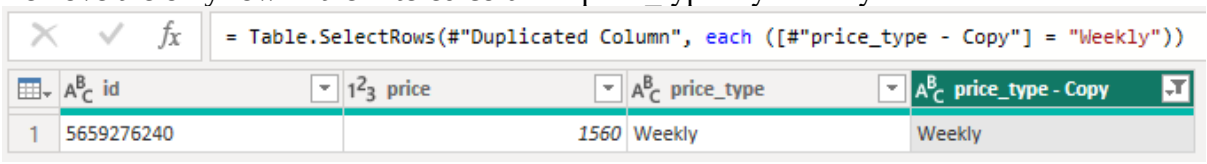


Fig. 25: Screenshot of ‘wekly’ price type in ‘price_type’ column

Splitting the column ‘Amenities’, which has a lot of different varitions, into multiple logical columns through adding conditional columns

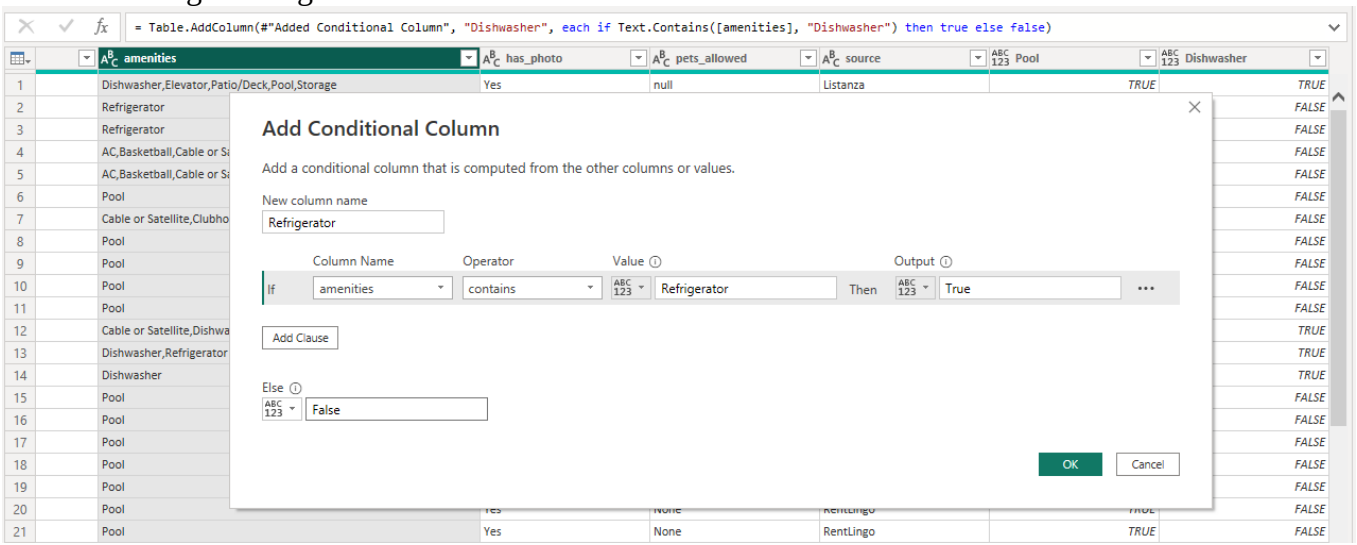


Fig. 26: Adding the additional column for amenities

Adding custom conditional column, where True for Dogs and/or Cats allowed, and False for ‘None’ allowed

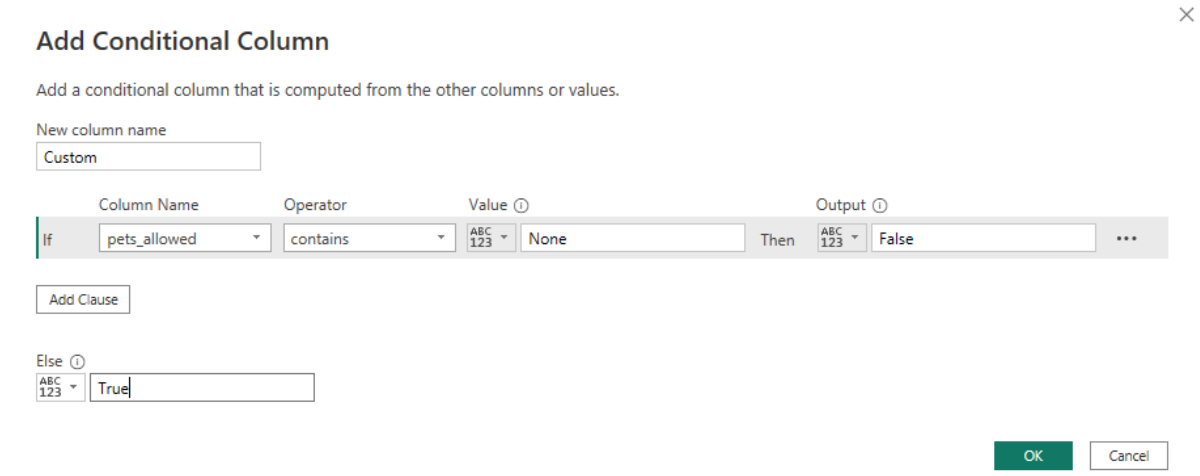


Fig. 27: Adding the additional column for ‘pets_allowed’

Change data type to logical

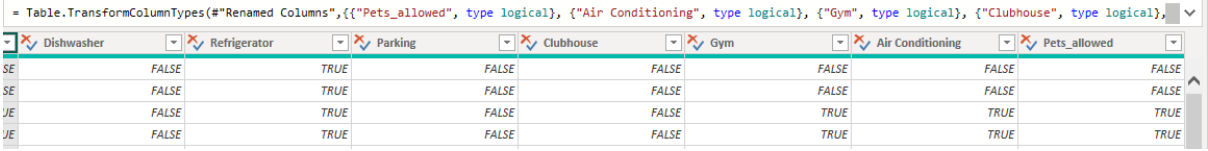


Fig. 28: Screenshot after changing data type to logical

Data Cleansing from outliers

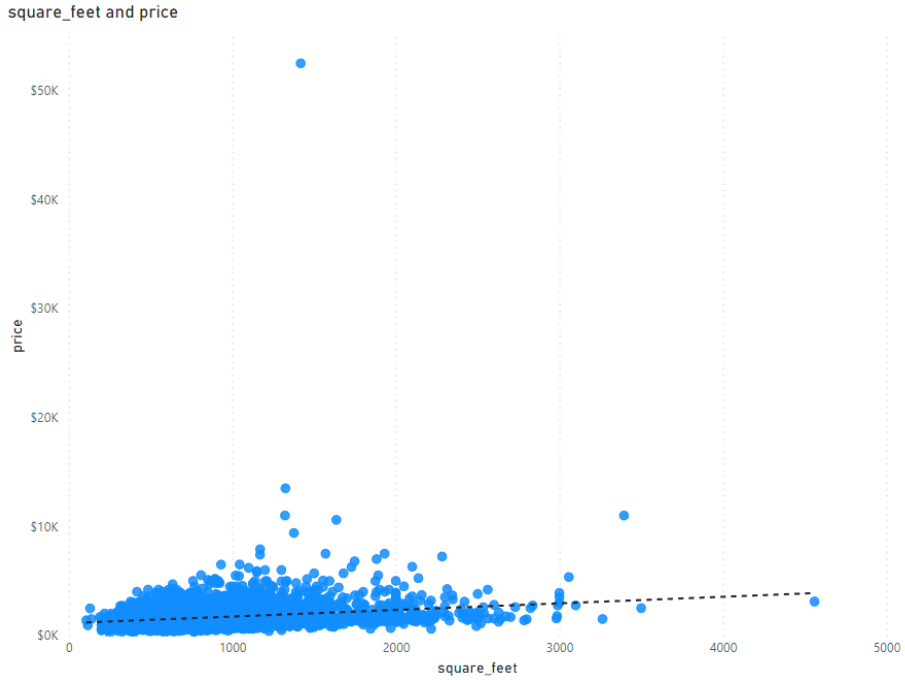


Fig. 29: Scatterplot for visualisation of the outliers in the dataset

Removed outliars:

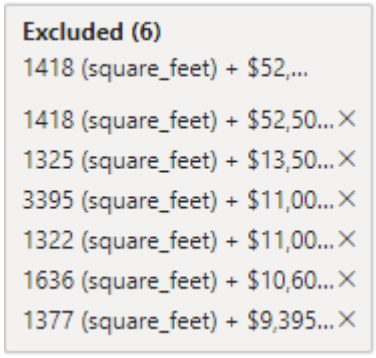


Fig. 30: Manual excluding of the outliers

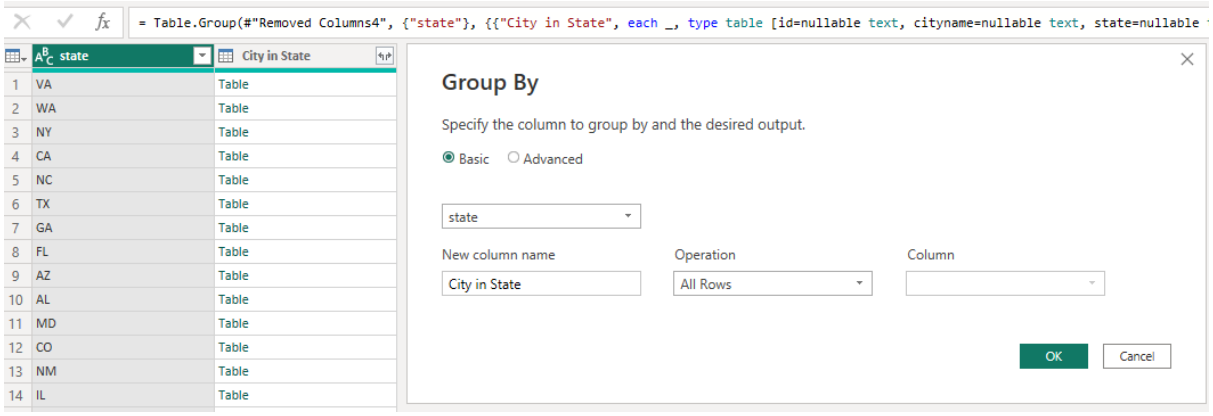


Fig. 31: Grouping ‘Cities’ in the Table ‘Location’ by ‘State’

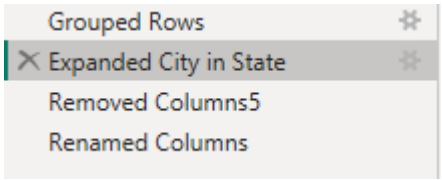


Fig. 32: Expanding the created table

2. Data Modelling

BI Data Modelling via Star Schema - Facts and Dimensions

Firstly, after data cleansing and pre-processing stage data model looked like:

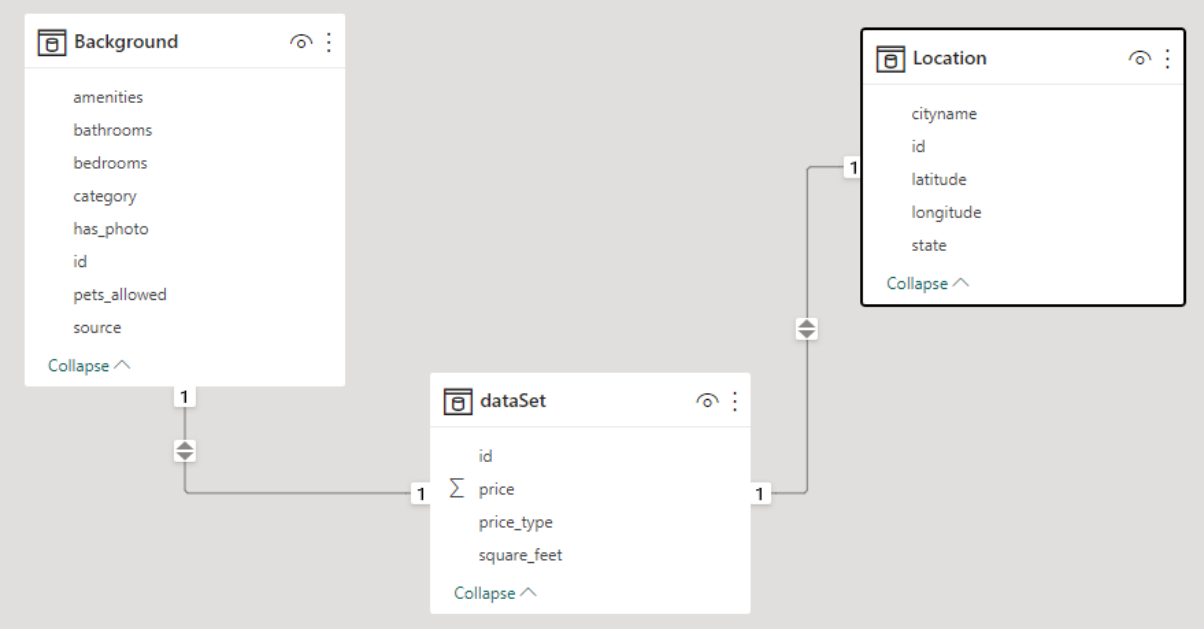


Fig. 33: Automatically created Data Model

We need to change the relationships to ‘one-to-many’ type.

We habe chosen the ‘ID’ column as the key connecting the fact table with dimension tables.

The decision to change relationships to a 'one-to-many' type and selecting the 'ID' column as the key connecting the fact table with dimension tables is typically driven by the need for a more efficient and organized database structure. Here are a few reasons why this might be done:

By establishing a 'one-to-many' relationship, you ensure data integrity. Each unique 'ID' in the fact table corresponds to one or more entries in the dimension table. This structure prevents data redundancy and inconsistencies.

This approach aligns with the principles of database normalization. Breaking down data into smaller, related tables reduces redundancy and dependency, making the database more flexible and easier to maintain.

'One-to-many' relationships are well-suited for efficient querying. They allow for faster retrieval of specific data points, as the database can easily navigate through the relationships using the shared 'ID' key.

Connecting the fact table with dimension tables through a common 'ID' column provides a structured and organized way to store and retrieve information. This facilitates better management of data and supports more complex analysis.

Using the 'ID' column as a key ensures consistency across tables. It establishes a clear and standardized link between the fact and dimension tables, making it easier to maintain and understand the database schema.

'One-to-many' relationships enable the use of joins in queries. This allows for combining data from different tables based on the common 'ID,' providing a comprehensive view of information that spans multiple dimensions.

This structure is scalable as it accommodates the growth of data. New entries in the dimension tables can be linked to existing entries in the fact table, or vice versa, without significant modifications to the database structure.

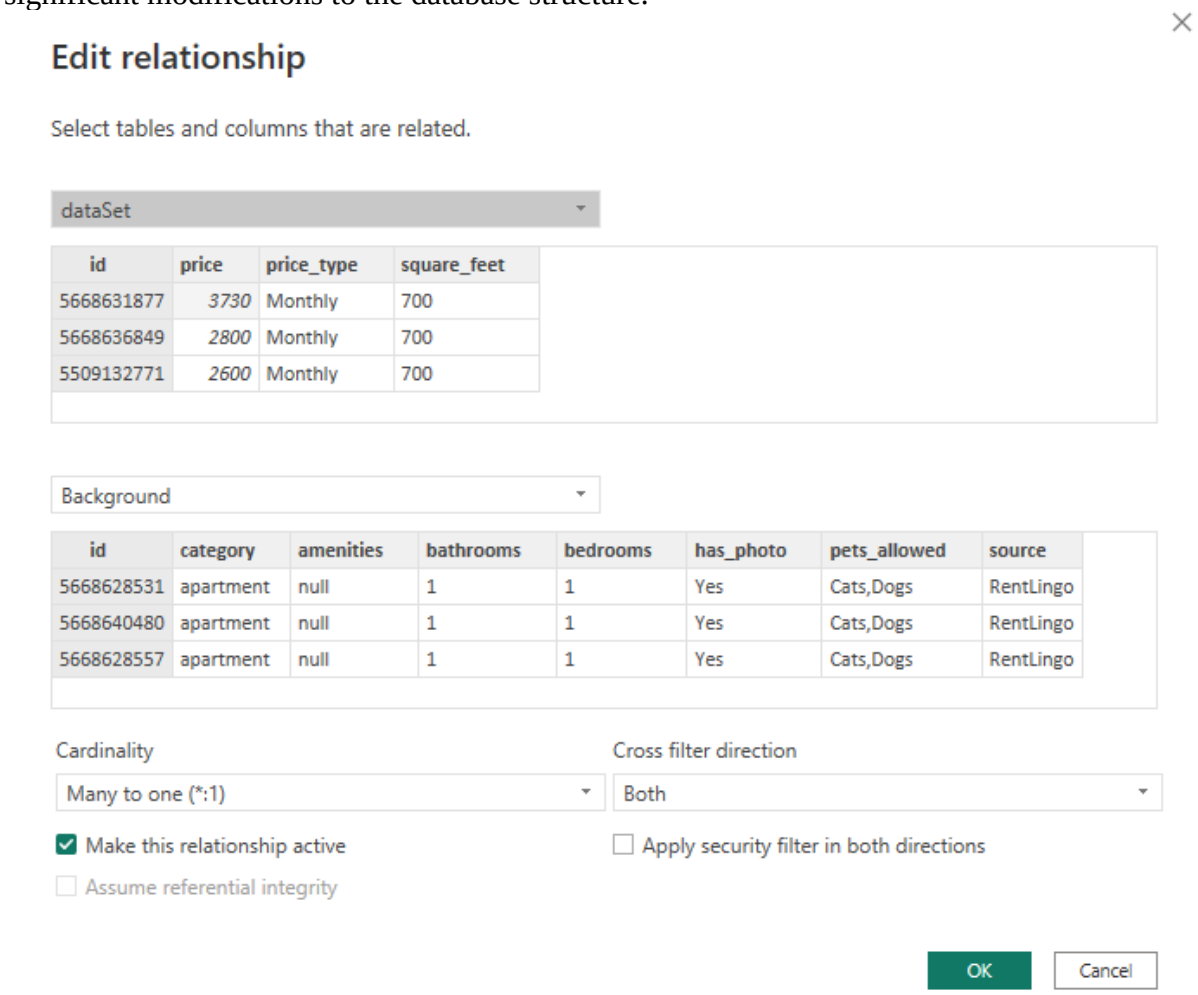


Fig. 34: Editing the relationship to many-to-one relationship

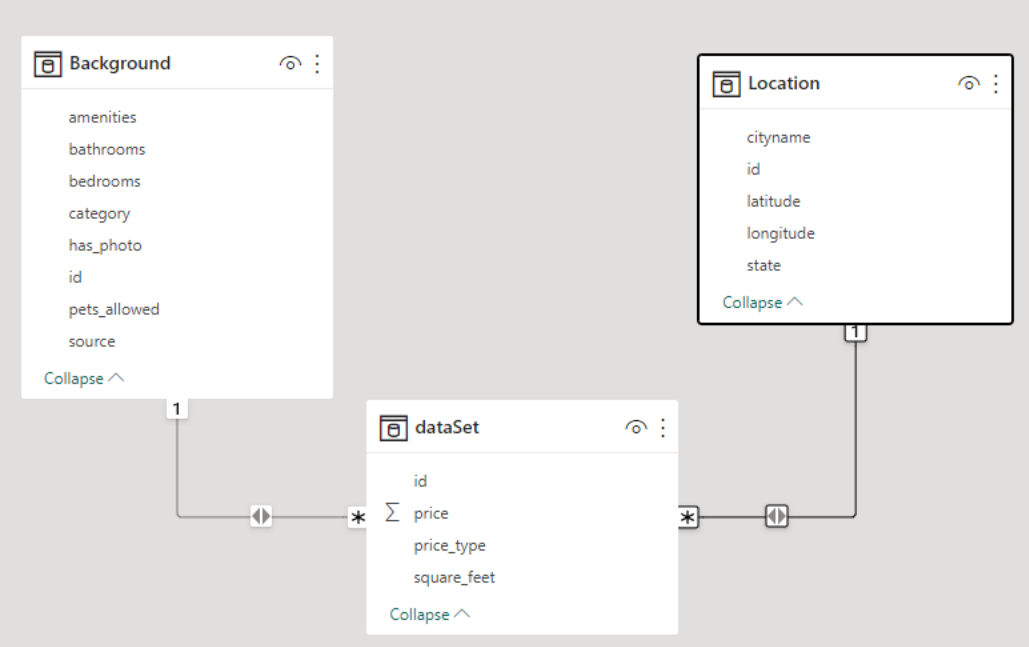


Fig. 35: Model with elationship many-to-one

We now have one big table (Background) that needs to be normalised into smaller tables (Dimension Tables), and one small Fact Table (dataset) and table Location. It is necessary to reduce complexity, avoid data duplication and for ease of analysing and creating visuals. We split our Background table into two table: ‘Background’ and ‘Size’. Our relationships model now looks like shown in the picture.

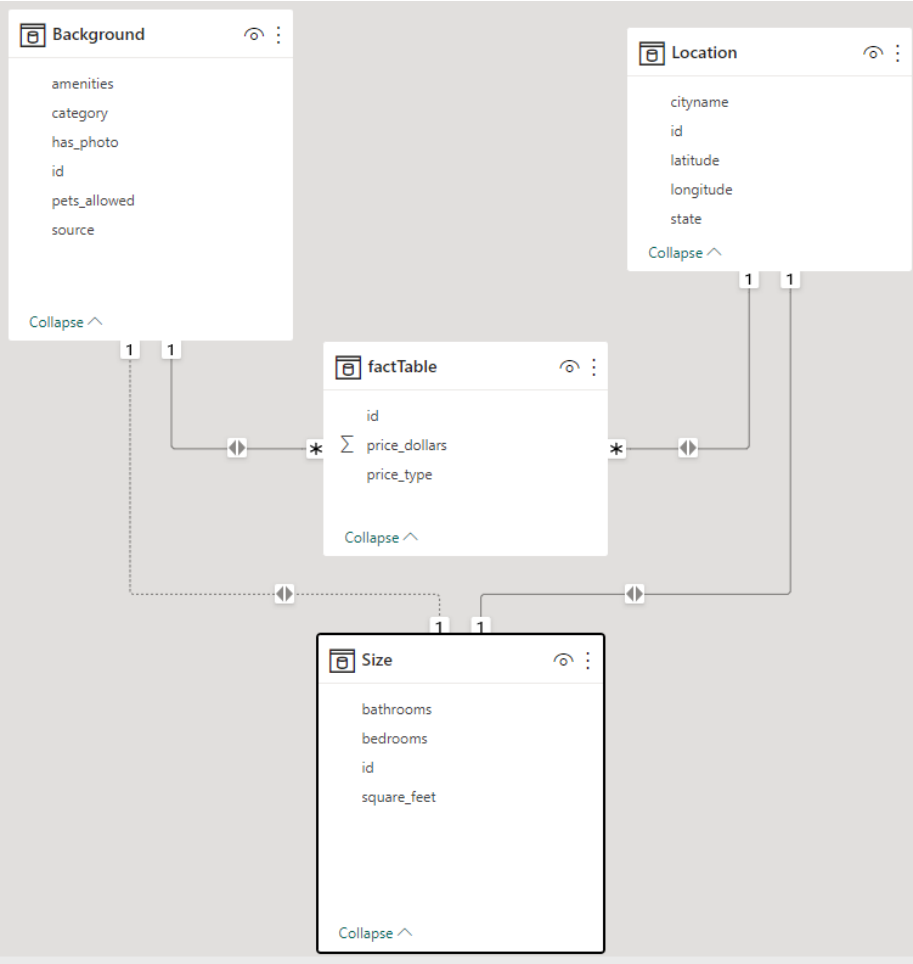


Fig. 36: Relationship model after splitting the dimension table into 2 tables

Change into the following structure, which is the final relationship model.

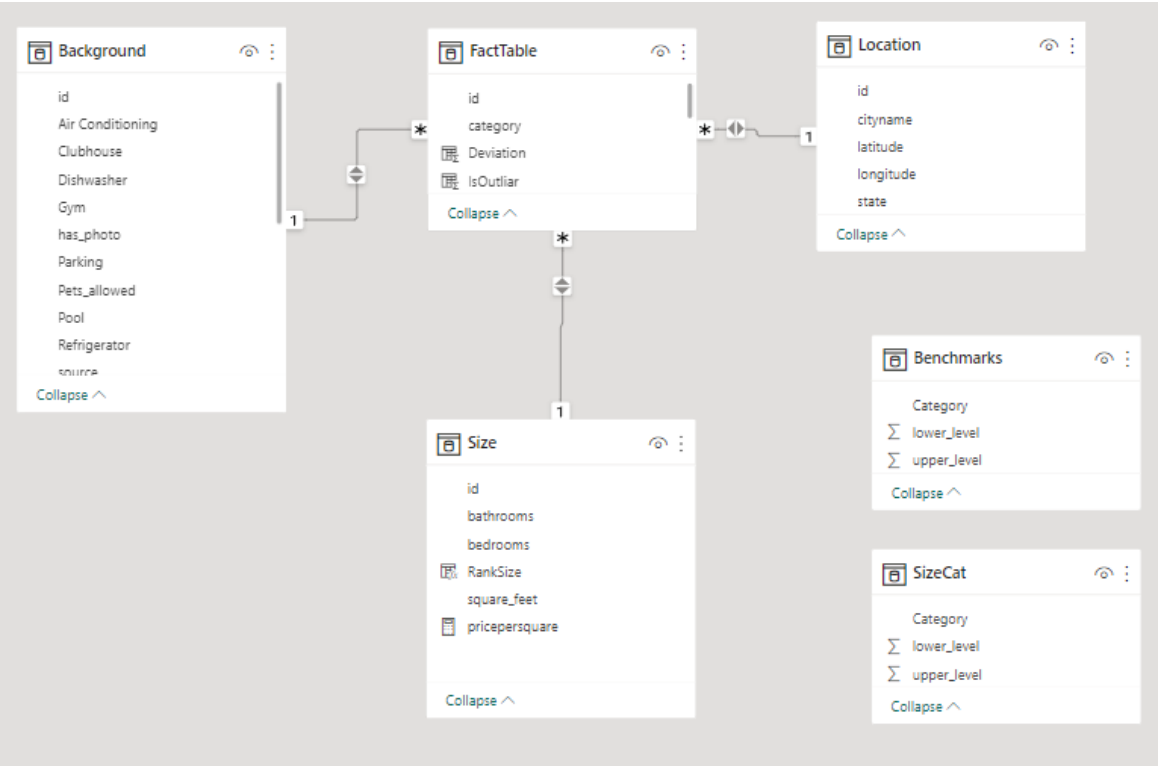


Fig. 37: Final relationship model

3. DAX and M Language

To calculate the Z score and identify outliers:
Calculate the Mean and Standard Deviation

```
1 Deviation = STDEV.P(FactTable[price])
```

id	price	price_type	price_per_square_ft	price_average	category	Deviation
5668639212	$1,250	Monthly	$5.4825	1400	apartment	911.785176263598
5664574541	$1,250	Monthly	$5	1400	apartment	911.785176263598
5648113719	$1,250	Monthly	$5	1400	apartment	911.785176263598
5668628100	$1,250	Monthly	$3.2468	1400	apartment	911.785176263598

Fig. 38: DAX code for the Mean and Standard Deviation

Calculate Z Score

```
1 Zscore = (FactTable[price] - FactTable[price_average])/FactTable[Deviation]
```

id	price	price_type	price_per_square_ft	price_average	category	Deviation	Zscore
5668639212	$1,250	Monthly	$5.4825	1400	apartment	911.785176263598	-0.164512435500086
5664574541	$1,250	Monthly	$5	1400	apartment	911.785176263598	-0.164512435500086

Fig. 39: DAX code for Z Score

Create a binary column (1 for outliers, 0 for non-outliers) based on the chosen threshold

```
1 IsOutliar = IF(ABS(FactTable[Zscore])>2,1,0)
```

id	price	price_type	price_per_square_ft	price_average	category	Deviation	Zscore	IsOutliar
5668639212	$1,250	Monthly	$5.4825	1400	apartment	911.785176263598	-0.164512435500086	0
5664574541	$1,250	Monthly	$5	1400	apartment	911.785176263598	-0.164512435500086	0

Fig. 40: DAX code for the binary column

We've got 242 outliars for 10k dataset
Create the Benchmarks table using M language:

Advanced Editor

Query1

```
let
    Source = Table.FromRecords({
        [Category = "up to 600$", lower_level = 0, upper_level = 600],
        [Category = "600-800$", lower_level = 600, upper_level = 800],
        [Category = "800-1000$", lower_level = 800, upper_level = 1000],
        [Category = "1000-1200$", lower_level = 1000, upper_level = 1200],
        [Category = "1200-1500$", lower_level = 1200, upper_level = 1500],
        [Category = "above 1500$", lower_level = 1500, upper_level = 1000000]
    })
in
    Source
```

Fig. 41: M language to create the benchmarks of rental cost

Mapping the cost of rent data into different benchmarks is a common practice in real estate and financial analysis. It will allow us to compare the cost of rent for a specific property or set of properties with similar properties in the market.

```
1 RankCategory =
2 VAR CatVar=[price]
3 RETURN
4 CALCULATE (VALUES (Benchmarks[Category]), CatVar> Benchmarks[lower_level],CatVar<=Benchmarks[upper_level])
```

Fig. 42: Mapping the Cost of Rent Data into the different Benchmarks
DAX for Price Per square Feet using ‘Related’ feature

```
price_per_square_ft = FactTable[price]/ RELATED(Size[square_feet])
```

id	price	price_type	price_per_square_ft	price_average	category
5668639212	1,250	Monthly	$5.4825	1400	apartment
5664574541	1,250	Monthly	$5	1400	apartment
5648113719	1,250	Monthly	$5	1400	apartment
5668628100	1,250	Monthly	$3.2468	1400	apartment
5664575536	1,250	Monthly	$3.125	1400	apartment

Fig. 43: DAX for Price Per square Feet using ‘Related’ feature

When calculating Price Per Square Feet (PPSF) using DAX and the 'Related' function, it typically involves working with two tables that are related to each other. The RELATED function is used to follow the relationship between the two tables and retrieve the square footage for the corresponding property in the PropertyTable. The formula then divides the total price by the square footage to calculate the PPSF.

Using DAX and the 'Related' function allows us to leverage the relationships between tables in your data model, making it easier to perform calculations that involve data from multiple tables. It helps ensure that the calculations are based on the correct relationships between the tables, providing more accurate and reliable results.

DAX Formula for Price Distribution:

```
Price Distribution = COUNTROWS(FILTER('FactTable', 'FactTable'[price] >= 0 && 'FactTable'[price] <= 100000))
```

Fig. 44: DAX Formula for Price Distribution

Create a new column or measure to check if a property has certain amenities (such as a pool, refrigerator, etc.) and then evaluate whether its price is above the average.
Create new table with the categories of property by Size using M language.

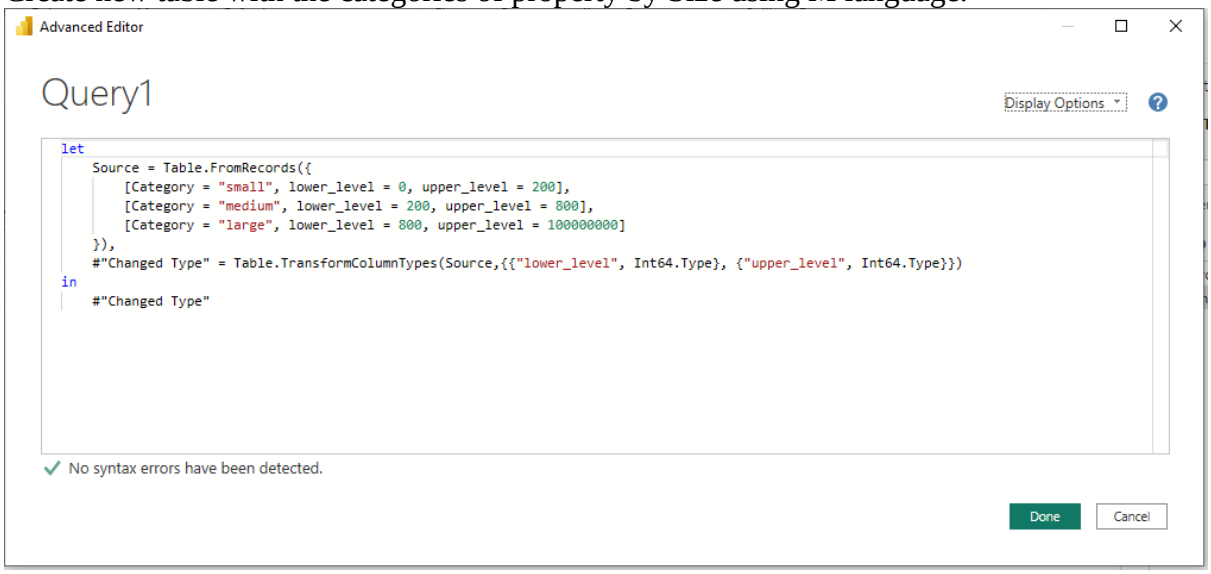


Fig. 45: M code for creating a new table ‘SizeCat’ with categories of properties by size

```
1 RankSize =
2 VAR CatVar=[square_feet]
3 RETURN
4 CALCULATE (VALUES (SizeCat[Category]), CatVar> SizeCat[lower_level],CatVar<=SizeCat[upper_level])
```

id	bathrooms	bedrooms	square_feet	RankSize
5509127621	1	1	575	medium
5508817650	1	1	575	medium
5664568520	1	1	575	medium
5668618260	1	1	575	medium
5668631595	1	1	575	medium
5664579369	1	1	575	medium

Fig. 46: DAX code to adjust RankSize with Table SizeCat

We count the number of of rows where the logical column has a true value to count the presence of amenities.

```
1 numgym = SUMX(Background, IF(Background[Gym] = TRUE(), 1, 0))
```

Fig. 47: DAX code for number of gyms

We want to analyze the co-occurrence of having both a pool and a gym in the property, so we will create a measure to identify instances where both conditions are met. This measure counts the rows where both the Pool and Gym conditions are true.

```
1 CoOccurrenceMeasure = CALCULATE(COUNTROWS(Background), Background[Pool] && Background[Gym])
```

Fig. 48: DAX code for cooccurancies

Then we create new measures for the variables of interest. For example, to count the number of amenities which have a gym, a pool, a parking and allow pets.

```
1 Gym&Pool&Parking&Pets = CALCULATE(COUNTROWS(Background), Background[Pool] && Background[Gym] && Background[Parking] && Background[Pets_allowed])
```

Fig. 49: DAX code for adjusting new column ‘RankSize’ with the Table ‘SizeCat’

KPI and BI Questions:

BI Question: What are top 10 Cities by average price?
KPI: Average of Price by City Name
Field: Location & Price
Visual: Table showing the average of price by city name

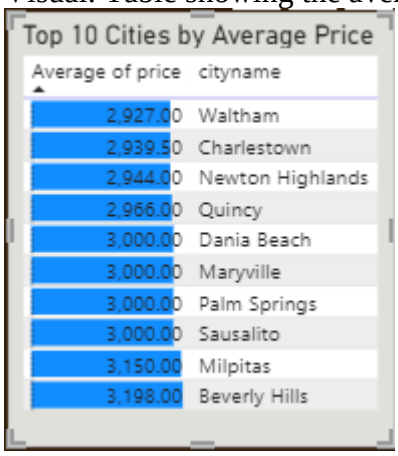


Fig. 50: Table of top 10 Cities by average price

BI Question: What are bottom 10 Cities by average price?
KPI: Average of Price by City Name
Field: Location & Price
Visual: Table showing the average of price by city name

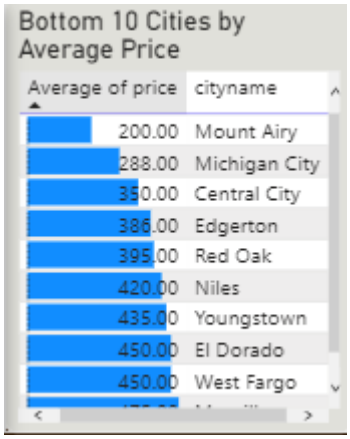


Fig. 51: Table of bottom 10 Cities by average price

BI Question: What is the distribution of property prices by size?
KPI: Price distribution by size
Field: Size, Price
Calculation: Price Distribution = COUNTROWS(FILTER(Size, Price >= 0 && Price <= 100000)) // Example distribution range
Recommended Visual: Histogram showing the distribution of property prices by size

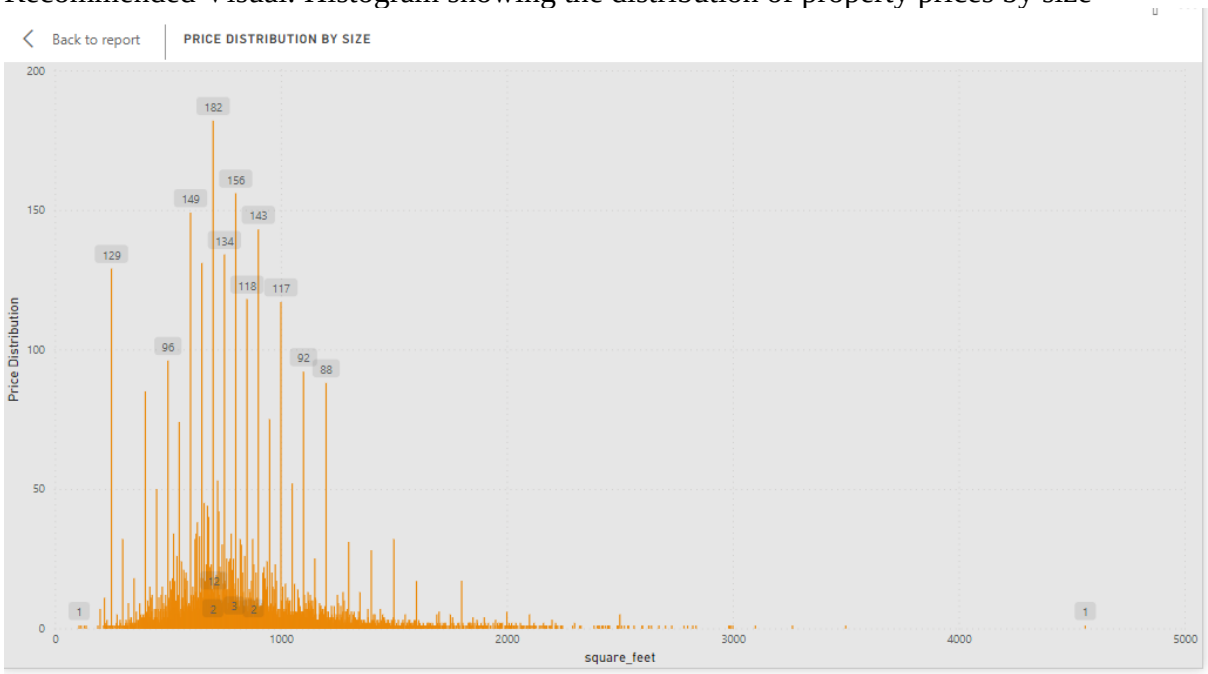


Fig. 52: Histogram of Price distribution by size

BI Question: What is the distribution of price per square feet?
KPI: Price distribution by square feet
Field: Price per square feet
Recommended Visual: Histogram showing the distribution of property prices per square feet

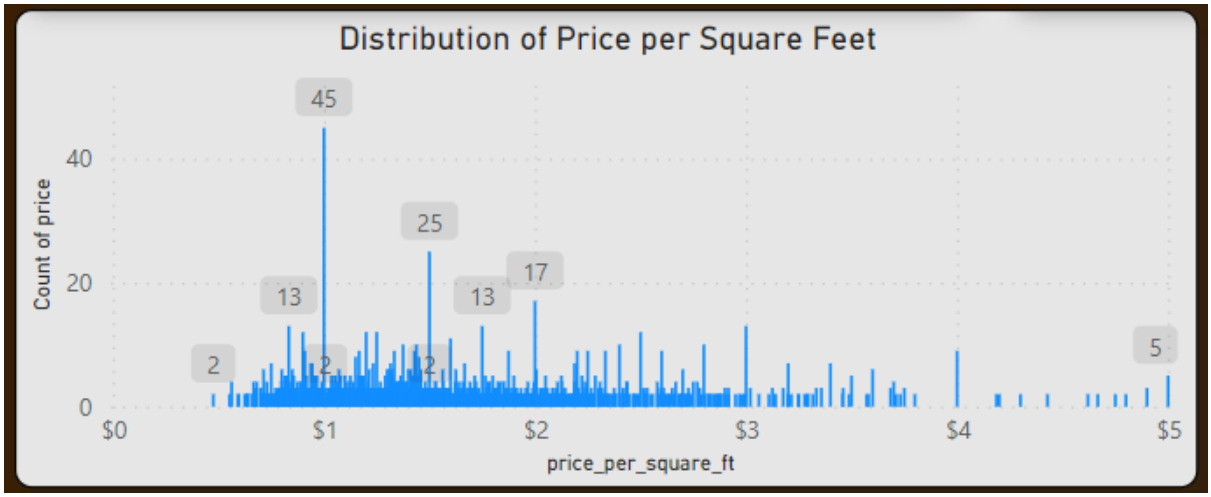


Fig. 52: Histogram of Distribution of Price per Square Feet

BI Question: How many properties by Price Categpory have amenities available?
KPI: Properties with amenities count by Price Category of Property
Field: Amenities
Calculation: Properties with Amenities = COUNTROWS(FILTER(Additional, Additional[amenities] <> ""))
Visual: Radar Chart

Fig. 53: Radar Chart of Properties with amenities count by Price Category of Property

BI Question: How many properties allow pets?
KPI: Pet-friendly properties count
Field: Amenities
Calculation: Pet-friendly Properties = COUNTROWS(FILTER(Additional, Additional[pets_allowed] = "Yes"))
Recommended Visual: Card showing the count of pet-friendly properties

Fig. 54: Visual Card of Number of Pet-friendly properties

BI Question: What is the distribution of bedrooms in properties?
KPI: Average number of bedrooms
Field: Size
Calculation: Average Bedrooms = AVERAGE(Size[bedrooms])
Visual: Donut chart

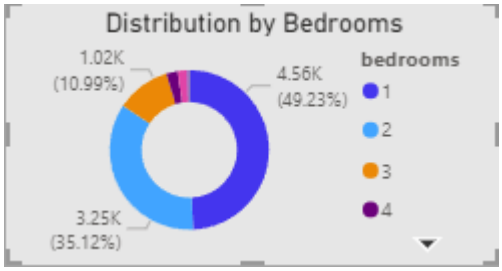


Fig. 55: Donut chart number of bedrooms

BI Question: What is the distribution of bathrooms in properties?
KPI: Average number of bathrooms
Field: Size
Visual: Donut chart

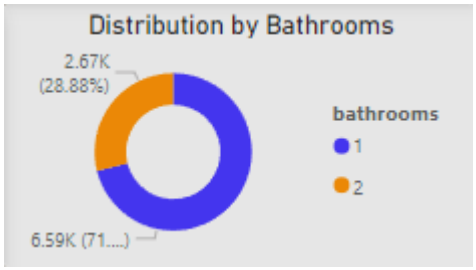


Fig. 56: Donut chart of number of bathrooms

BI Question: What is the distribution of property prices by city?
KPI: Price distribution by city
Field: Location, Price
Calculation: Price Distribution = COUNTROWS(FILTER(Location, Price >= 0 && Price <= 100000)) // Example distribution range
Recommended Visual: Treemap

Fig. 57: Treemap of Price distribution by city

BI Question: What is the number of properties by different amenities?
KPI: Number of Properties with certain Amenities
Field: Amenities
Recommended Visual: Funnel

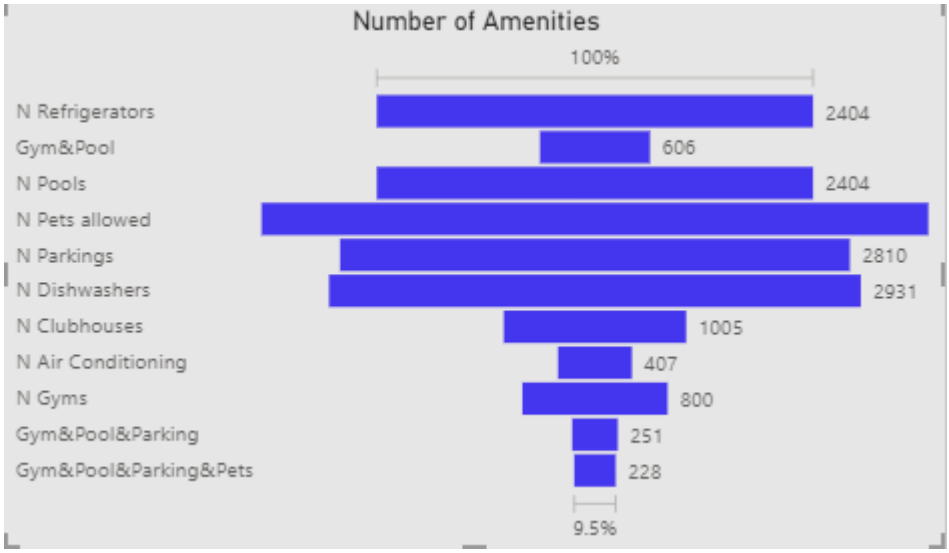


Fig. 58: Funnel of Number of Properties with certain Amenities

BI Question: How many properties are there in each price Category?
KPI: Properties count by Rank category
Field: Price
Visual: Packed Bubble

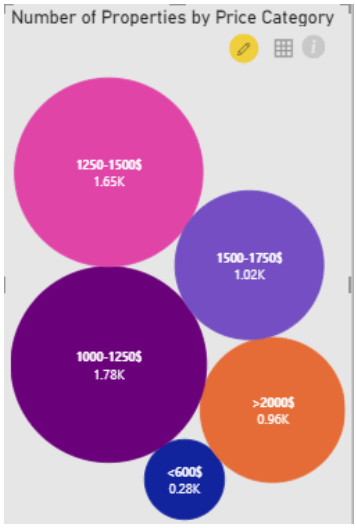


Fig. 59: Bubble chart for Properties count by Rank category

BI Question: What is the Total Number of Properties
KPI: Count of id
Field: FactTable
Visual: Card

Fig. 60: Visual Card of tTotal Number of Properties

BI Question: What is the Total Number of States in the analysed data?
KPI: Count of States
Field: FactTable
Visual: Card

Fig. 61: Visual Card of Number of States

BI Question: What is the Total Number of Cities in the analysed data?
KPI: Count of Cities
Field: FactTable
Visual: Card

Fig. 62: Visual Card of Number of Cities

BI Question: What is the average price of property in each state?
KPI: Average price by State
Field: Location
Calculation: Properties Count = COUNT(Location[state])
Visual: Clustered column chart showing average price by state

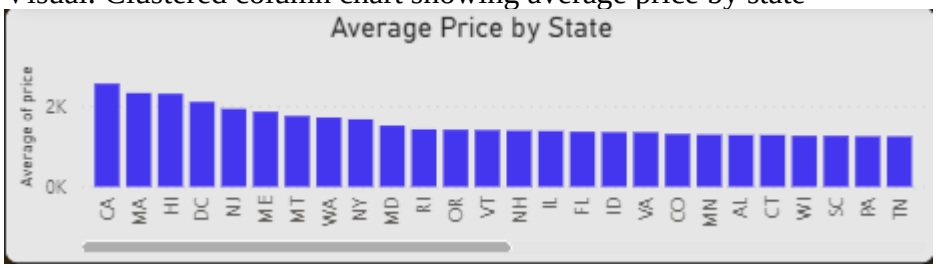


Fig. 63: Chart of average price by state

BI Question: What is the distribution of properties by rental cost?
KPI: Number of properties by rental cost
Field: Price
Recommended Visual: Histogram with data labels

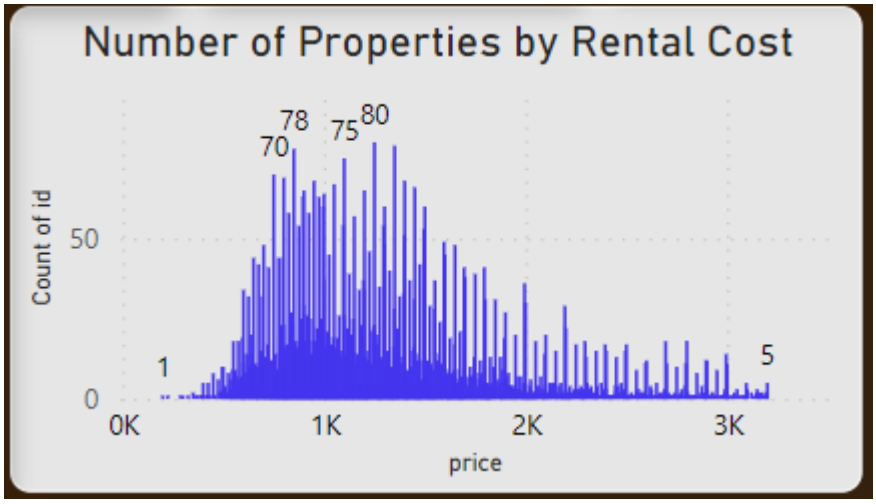


Fig. 63: Histogram of number of properties by rental cost

BI Question: What is the geographical distribution of cities by price category?
KPI: City by Price Category
Field: Location, Price
Recommended Visual: Filled map

Fig. 64: Filled map of City by Price Category

BI Question: What is the average price per square feet by property size?
KPI: Price per square feet by Size
Field: Price, Size
Recommended Visual: Packed Bubble

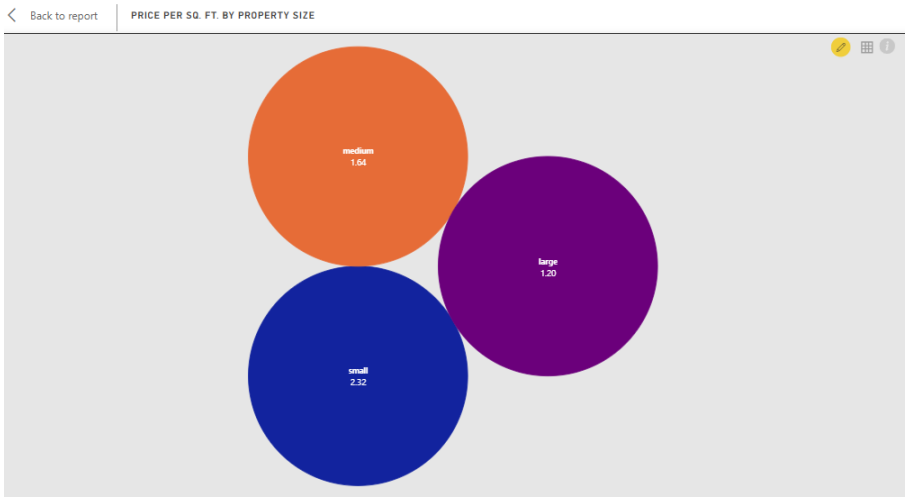


Fig. 65: Bubble chart for Price per square feet by Size

BI Question: What is the average price per square feet by property size?
KPI: Number of Properties by Rental Cost
Field: Price
Recommended Visual: Clustered Column Chart

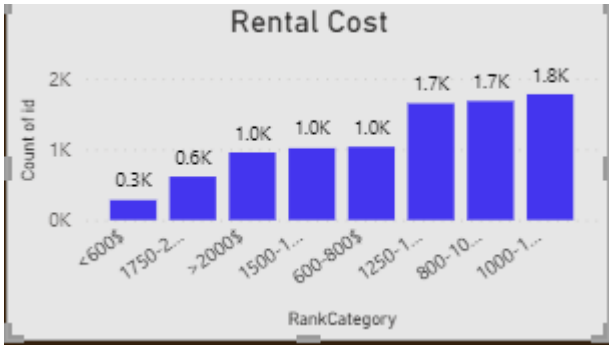


Fig. 66: Chart of Number of Properties by Rental Cost

BI Question: What is the minimum rental price among all properties?
KPI: Min Price
Field: Price
Visual: Card

Fig. 67: Visual Card for Min Price

BI Question: What is the average rental price among all properties?
KPI: Average Price
Field: Price
Visual: Card

Fig. 68: Visual Card for Average Price

BI Question: What is the maximum rental price among all properties?
KPI: Max Price
Field: Price
Visual: Card

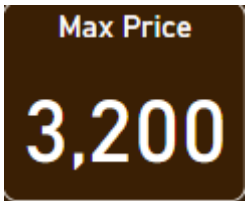


Fig. 69: Visual Card for Max Price

BI Question: What is the minimum size of the property among all?
KPI: Min Size
Field: Size
Visual: Card

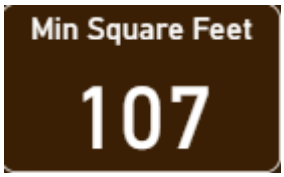


Fig. 70: Visual Card for Min Size

BI Question: What is the average size of the property among all?
KPI: Average Size
Field: Size
Visual: Card

Fig. 71: Visual Card for Average Size

BI Question: What is the maximum size of the property among all?
KPI: Max Size
Field: Size
Visual: Card

Fig. 72: Visual Card for Max Size

Charts using AI:

G&A chart:

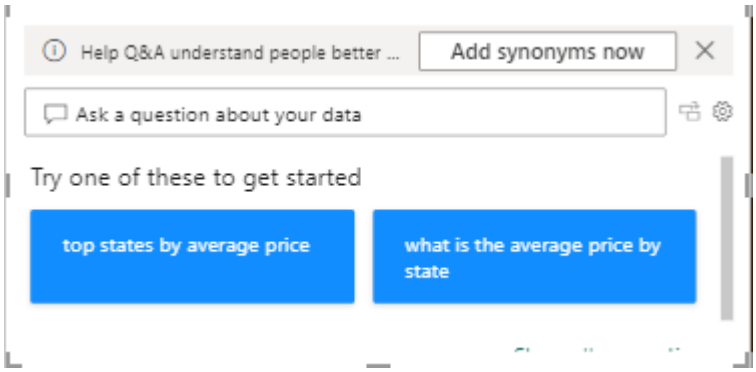


Fig. 73: AI Visual of Q&A

Key influencers:

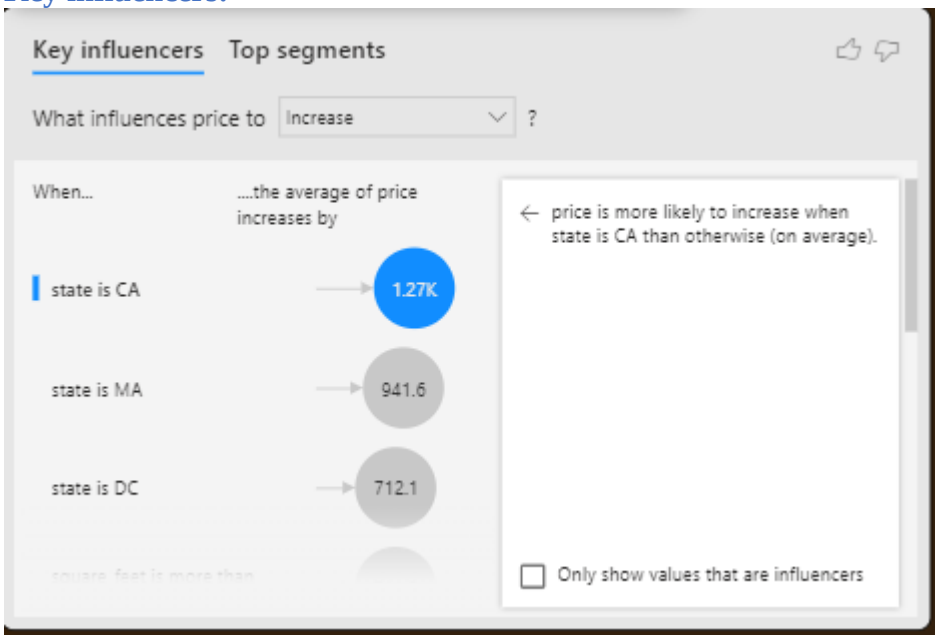


Fig. 74: AI Visual of Key Influencers

Slicer for outliers:

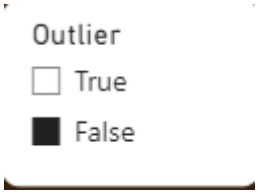


Fig. 75: Slicer to filter outliers

Slicer for Size of Property:

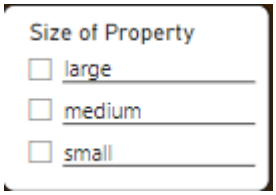


Fig. 76: Slicer to filter properties by size

4. Dashboard

There are 4 card visuals, 1 slicer, 1 map, 4 buttons for navigation, 2 tables, 1 bubble chart and a Q&A chart on the first page of the Dashboard. They mostly show the main information about the dataset, for example, the number of properties analysed, number of states, number of cities. Also there is an option to filter outiars using the slicer.

The map shows the sttes with different colour and the average price per squaree feet for each of them. There are two tables with top 10 and bottom 10 cities by Average Price. Moreover we can use the Q&A chart to see the visuals for the main KPI questions as top states by average price, average price by state and many others.

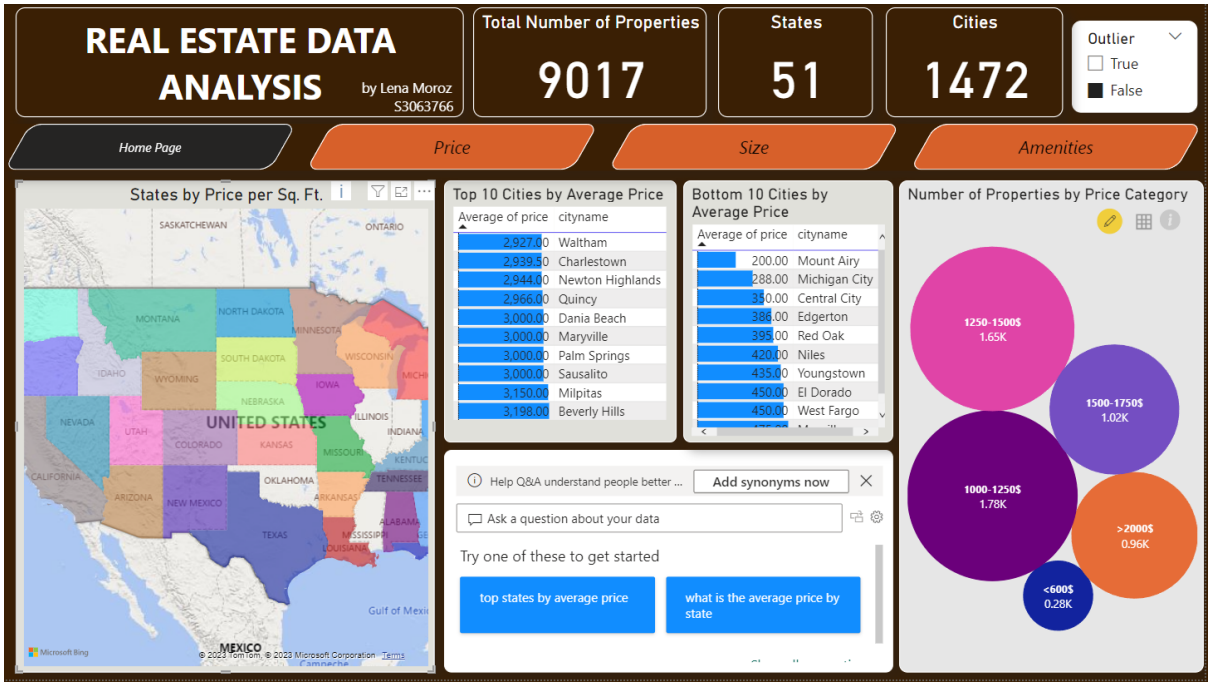


Fig. 77: Page 1

On the second page of the Dashboard we have 4 visual cards of minimum, average and maximum price, two slicers, one for outliers, one for size of the property, and 5 more charts.

First chart was made using the AI tools – it is the chart of key Influencers.

Second chart is the clustered column chart. It shows the average price by State.

Third chart is the Bubble chart showing the price per fquaree feet by property size.

Fourth chart is the clustered column chart. It shows the number of properties by rental cost.

Fifth visual is a filled map to show the city ranked by price category.

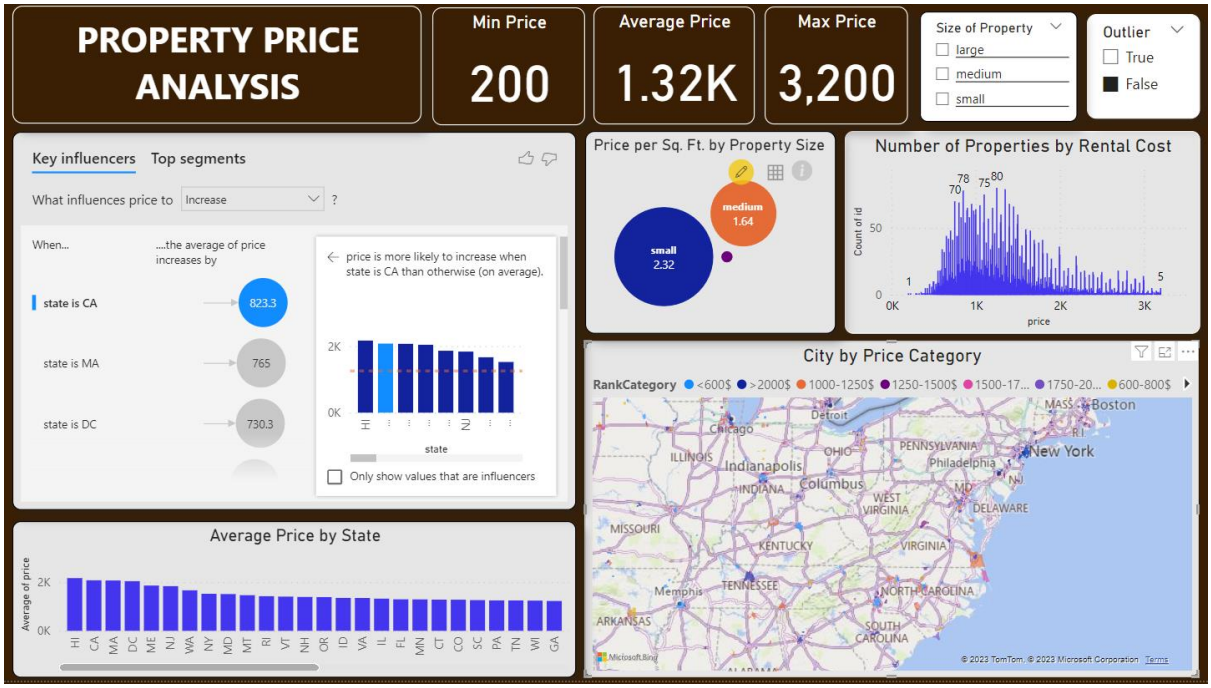


Fig. 78: Page 2

On the third page of the dashboard we can see 4 cards, 1 slicer and 5 charts. Cards are used to show the KPIs for minimum squaree feet, average size in squaree feet, maximum squaree feet of the Dataset. There is a slicer to filter the outliers.

Infographic Visual was used to show the number of properties by Rank Category. Stacked column chart was used to show the price distribution by size. There are data labels to show the most important numbers. There was also used smart narrative as an AI visual.

Another stacked column chart shows the average price by Rank Size. It has three categories of properties by size: small, medium and large. The scatter plot was used to show the linear regression for price and xize. There was used the average line and the trend line. Moreover, the scatter plot was divided by colour by the rank category of rental price.

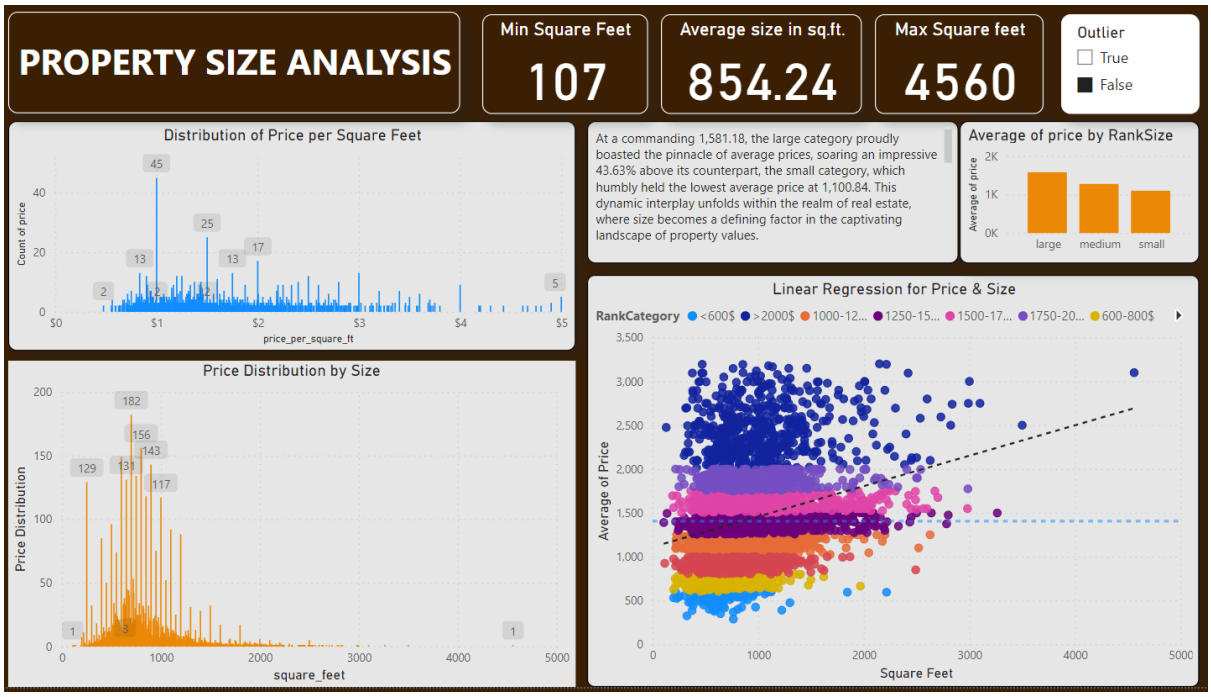


Fig. 79: Page 3

There is 1 visual card, 1 text card, 1 AI smart narrative and 5 charts on the fourth page of the Dashboar.

First chart is the treeap showing the price distribution by city. Second chart is the funnel showing the number of amenities available for all properties in the dataset.

There are two donut charts to show the distribution of properties by bedroons and bathrooms.

The radar chart was used to visualise the amenities available by price category of the property.

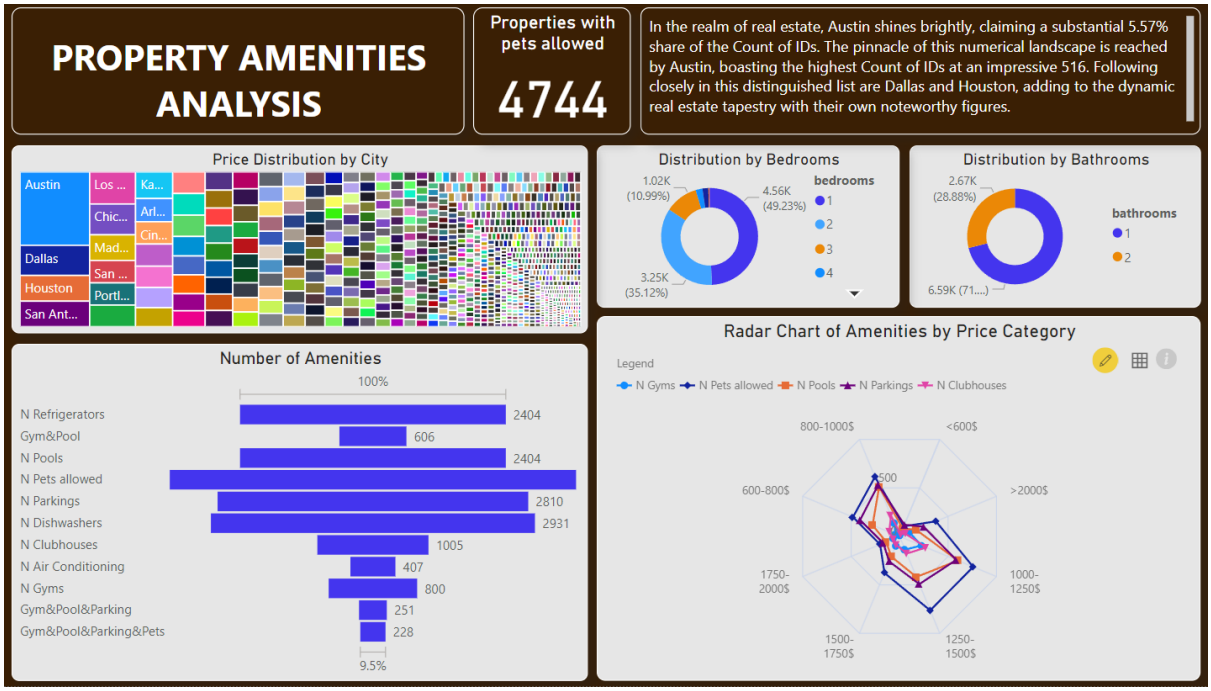


Fig. 80: Page 4

Use the table below to **self-assess** your work. This will help reflect on your work. You must keep this table in your report.

Report Section	**Description**	**Grade your work from 0 to 100**
Report Structure	The report is well-written, and it contains all the relevant sections	100
Data Pre-processing and Data Modelling	Many pre-processing steps have been applied. The data model is well-structured	100
Dax and M language	Both DAX and M Language have been **extensively** used in the report	100
Dashboard Design	The dashboard contains a variety of charts, including advanced ones not covered in the module.	90
Average		**Add below the average of the four cells above:** **97.5**

References

1. Apartment for rent classified. UC Irvine Machine Learning Repository. 2019. URL: https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified

2. **Business Intelligence System Design Methodology:**

- Agile methodology with iterative development cycles
- Use of project management tools such as Trello, Monday or Jira for task tracking and collaboration

3. **Business Intelligence Software Development:**

- PowerBI for building BI KPI-Dashboards

Google Scholar Research Papers:

4. Google Scholar: Real Estate Rent Cost USA
5. Factors of Influence on the Real Estate Market: Sergiy Kobzan & Olena Pomortseva (2023)
6. Optimization of United States Residential Real Estate Investment through Geospatial Analysis and Market Timing: Harrison Gale & Shouraseni Sen Roy - Applied Spatial Analysis and Policy (2023)
7. *TU-AI-projectMansh: Ask me to recommend a Project Proposal for Academia or Industry?* Mansha Nawaz, WNR AI. (2023). (https://wnr.ai/templates/tu-ai-projectmansh-ask-me-to-recommend-a- project-proposal-for-for-academia-or-industry)
8. Whismer. (2023). (https://whismer.com/)
9. Apartment for Rent Classified (no date) UCI Machine Learning Repository. Available at: https://archive.ics.uci.edu/dataset/555/apartment+for+rent+classified (Accessed: 21 November 2023).
10. Available Cran packages by name (no date) CRAN Packages By Name. Available at: https://cran.r-project.org/web/packages/available_packages_by_name.html#available-packages-E (Accessed: 21 November 2023).
11. Grolemund, H.W. and G. (no date) R for data science, Welcome. Available at: https://r4ds.had.co.nz/index.html (Accessed: 21 November 2023).
12. Kobzan, S. and Pomortseva, O. (1970) Factors of influence on the real estate market, SpringerLink. Available at: https://link.springer.com/chapter/10.1007/978-3-031-31248-9_5 (Accessed: 21 November 2023).
13. RStudio open-source packages (no date) RStudio. Available at: https://www.rstudio.com/about/rstudio-open-source-packages/ (Accessed: 21 November 2023).
14. 'Real estate market analysis system based on Big Data' (2023) Academic Journal of Computing; *Information Science*, 6(7). doi:10.25236/ajcis.2023.060723. (Accessed: 21 November 2023).
15. Li, B. (2021) 'Research on real estate information system of the real estate market based on Big Data Technology', *E3S Web of Conferences*, 257, p. 02037. doi:10.1051/e3sconf/202125702037. (Accessed: 21 November 2023).
16. Ruivo, R. (2010). Determinants of Rental Rates in Major Cities in the United States. Empirical Economic Bulletin, An Undergraduate Journal, [online] 3(1). Available at: https://digitalcommons.bryant.edu/eeb/vol3/iss1/6/. (Accessed: 21 November 2023).
17. America's Rental Housing 2022 | Joint Center for Housing Studies. [online] Available at: https://www.jchs.harvard.edu/americas-rental-housing-2022. (Accessed: 21 November 2023).

Agile methodology with iterative development cycles:

18. Bhuvan, U. (2017a) 'Business process modeling (BPM), agile practices, and big data adoption process', *Big Data Strategies for Agile Business*, pp. 125–162. doi:10.4324/9781315120232-4.

19. Bhuvan, U. (2017b) 'Functioning agile organization with Big Data Adoption', *Big Data Strategies for Agile Business*, pp. 385–413. doi:10.4324/9781315120232-12.

20. Sohail, F. et al. (2021) 'Impact of agile methodology on software development life cycle', *Pakistan Journal of Engineering and Technology*, 4(2), pp. 153–158. doi:10.51846/vol4iss2pp153-158.

21. Taylor, P. and Sharrock, E. (2021) 'Agile Project Management', *Make Your Business Agile*, pp. 56–71. doi:10.4324/9781003031826-10.

Star Scheme Data Modelling:

22. Understand star schema and the importance for Power BI | *Microsoft Learn*. Available at: https://learn.microsoft.com/en-us/power-bi/guidance/star-schema (Accessed: 13 November 2023).

23. 'Conceptual model for Star Schema Data Warehouse' (2023) *International Research Journal of Modernization in Engineering Technology and Science* [Preprint]. doi:10.56726/irjmets41621.

Power BI for data visualization and dashboard creation:

24. Create a power BI Dashboard from a report - power bi, Power BI | Microsoft Learn. Available at: https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboard-create (Accessed: 13 November 2023).

25. Intro to dashboards for power bi designers - Power BI | *Microsoft Learn*. Available at: https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboards (Accessed: 13 November 2023).

M programming language:

26. Understanding power query m functions - powerquery m, PowerQuery M | Microsoft Learn. Available at: https://learn.microsoft.com/en-us/powerquery-m/understanding-power-query-m-functions (Accessed: 13 November 2023).

Jupyter Notebook for data analysis and modelling:

27. Rovnik, V. (2020) Data Analysis and visualization with Jupyter Notebook, Medium. Available at: https://towardsdatascience.com/data-analysis-and-visualization-with-jupyter-notebook-22f6dcd25cc5 (Accessed: 13 November 2023).

28. Wickramasinghe, S. (2021) Jupyter notebooks for Data Analytics: A beginner's guide, BMC Blogs. Available at: https://www.bmc.com/blogs/installing-jupyter-for-big-data-and-analytics/ (Accessed: 13 November 2023).

Google Scholar Research Papers for Business Intelligence in real estate:

29. Greene, D. (2022) 'Landlords of the internet: Big data and big real estate', *Social Studies of Science*, 52(6), pp. 904–927. doi:10.1177/03063127221124943.

30. 'Real estate market analysis system based on Big Data' (2023) Academic Journal of Computing; *Information Science*, 6(7). doi:10.25236/ajcis.2023.060723.

31. Li, B. (2021) 'Research on real estate information system of the real estate market based on Big Data Technology', *E3S Web of Conferences*, 257, p. 02037. doi:10.1051/e3sconf/202125702037.

32. Mourouzi-Sivitanidou, R. and Sivitanides, P. (2020) ‘Data sources for real estate market analysis’, *Market Analysis for Real Estate*, pp. 427–432. doi:10.1201/9780429279409-26.