

Data Analysis Project: Australian Real Estate

Team Members:

Khushbu Kothari (Khushi)

Napat

Tsz Hin (Raymond) Tang &

Kajal Jain



TABLE OF CONTENTS

01

Overview

Purpose and scope of this EDA

02

Change of Dataset

why we decided to change our primary dataset.

03

Key Questions

Our focused questions which we are trying to analyze

04

Price analysis

Price distribution and average prices.

05

Property Type Analysis

Property type distribution and Dominating property type

06

Property Feature Analysis

How does different features influence the price

07

Avg Listing Price of Agencies

Top 10 agencies and their average listing price

08

Summary of Analysis

Overall summary of the findings and insights

PURPOSE AND SCOPE

Australian Real Estate Dataset (Primary Dataset)

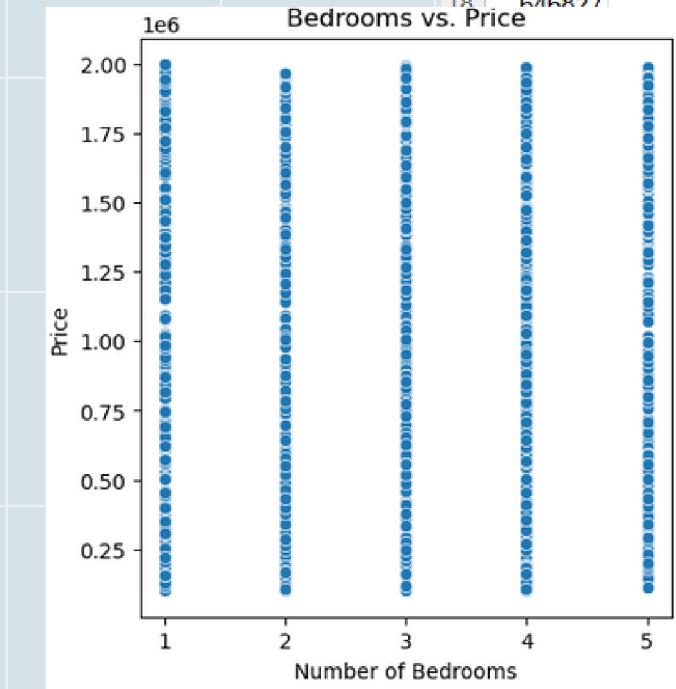
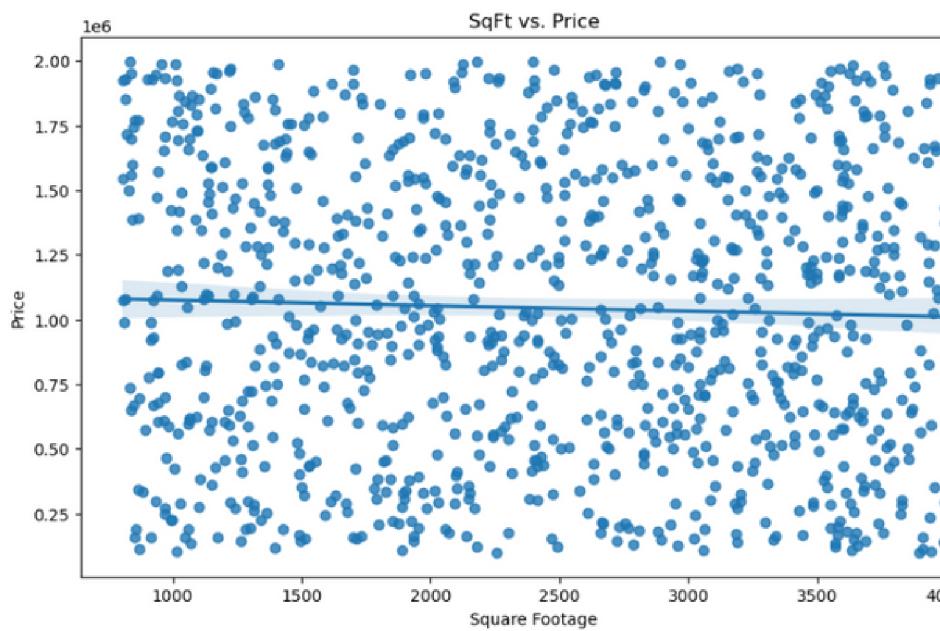
The purpose of this EDA is to explore and analyse a real estate dataset to understand underlying patterns, relationships and trends within the data.

Key questions we tried to analyze.

- What is the average price of properties in different cities?
- How does the price distribution look, and are there any outliers?
- Which features, such as the number of bedrooms, bathrooms, and square footage, are most strongly correlated with the price?

Change of Primary Dataset

- First attempt of choosing the data
- significant inconsistencies
- no-correlation between the columns
- no outliers
- Data distribution



| A | B | C | D | E | F | G | H | I | J | |
|----|---------|----------|-----------|------|-----------|-------|------------|-----------|--------|----------|
| 1 | Price | Bedrooms | Bathrooms | SqFt | City | State | Year_Built | Type | Garage | Lot_Area |
| 2 | 982112 | 4 | 1 | 1561 | Adelaide | QLD | 2021 | Townhouse | 0 | 1357 |
| 3 | 571388 | 3 | 2 | 3735 | Melbourne | VIC | 1999 | House | 1 | 8397 |
| 4 | 866821 | 5 | 1 | 2032 | Sydney | VIC | 1976 | Townhouse | 0 | 3478 |
| 5 | 1230977 | 4 | 3 | 3861 | Brisbane | SA | 1978 | House | 0 | 7619 |
| 6 | 241787 | 4 | 2 | 3150 | Perth | SA | 1992 | Apartment | 0 | 8324 |
| 7 | 631871 | 4 | 2 | 1923 | Brisbane | SA | 2002 | House | 1 | 6638 |
| 8 | 550777 | 4 | 1 | 829 | Adelaide | QLD | 2022 | Townhouse | 0 | 2752 |
| 9 | 1509971 | 1 | 1 | 2447 | Brisbane | QLD | 1967 | Apartment | 1 | 2792 |
| 10 | 935088 | 4 | 2 | 3652 | Melbourne | QLD | 1985 | House | 0 | 1645 |
| 11 | 999053 | 1 | 1 | 3114 | Sydney | NSW | 1962 | House | 0 | 4195 |
| 12 | 679189 | 4 | 3 | 2891 | Sydney | WA | 1999 | Apartment | 0 | 2827 |
| 13 | 977320 | 1 | 1 | 2218 | Adelaide | QLD | 1972 | Apartment | 0 | 8425 |
| 14 | 901416 | 4 | 3 | 1771 | Brisbane | QLD | 1995 | Apartment | 1 | 8587 |
| 15 | 911184 | 4 | 3 | 3855 | Sydney | WA | 1955 | House | 0 | 5090 |
| 16 | 945381 | 3 | 2 | 1672 | Brisbane | QLD | 1989 | House | 0 | 9937 |
| 17 | 1453282 | 5 | 3 | 1842 | Melbourne | NSW | 1993 | Townhouse | 0 | 1298 |
| 18 | 616827 | 5 | 2 | 2211 | Melbourne | NSW | 1988 | House | 1 | 5629 |
| | | | | 810 | Melbourne | NSW | 2013 | Townhouse | 1 | 7482 |
| | | | | 1359 | Adelaide | QLD | 1991 | House | 0 | 9376 |
| | | | | 1233 | Adelaide | NSW | 1964 | Townhouse | 0 | 7669 |
| | | | | 1465 | Brisbane | SA | 1966 | House | 0 | 7456 |
| | | | | 2231 | Brisbane | QLD | 1995 | House | 0 | 8726 |
| | | | | 3578 | Brisbane | QLD | 1991 | House | 0 | 9936 |

Redefined Purpose And Scope

Northern Territory region -Property price (New Dataset)

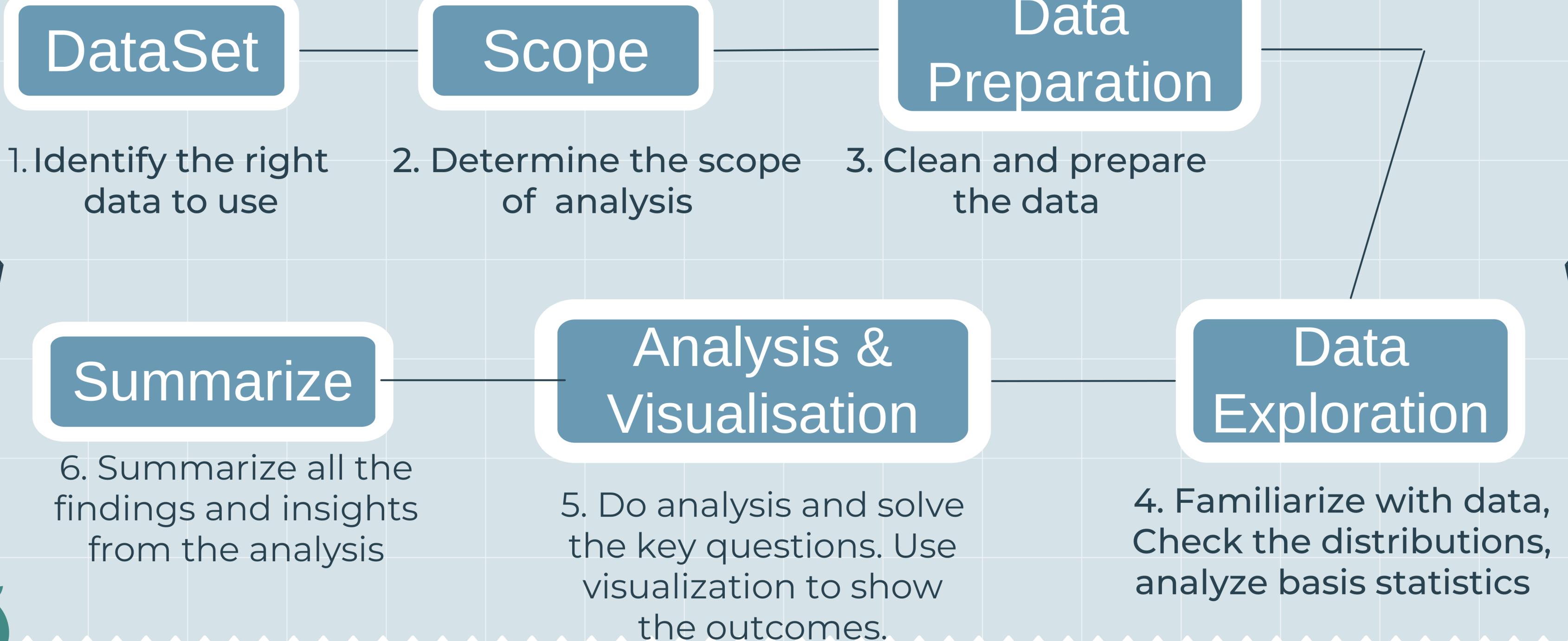
we decided to pivot our focus to the Northern Territory (NT) region of Australia.

we redefined our goal and aimed to explore and analyze NT real estate market to understand underlying patterns, relationships and trends within the data.

Key questions we tried to analyze.

- What is the distribution shape of real estate prices in the northern territories ?
- Which property type has the highest demand in the Northern Territory market.
- How does different features (bedrooms, bathrooms, carpark) influence the price of the properties ?
And how it correlated ?
- Compare the market share of different real estate agencies
- what is the Average Listing Price by Top 10 Agencies?

OUR FRAMEWORK



Data cleaning

| index | TID | property_type | price | location_type | location_name | city | state | latitude | longitude | product_depth |
|-------|---------|---------------|-------------------------------------|---------------|-------------------------------------|-------------|-------|----------|-----------|---------------|
| 8 | 1350996 | Unit | \$439,000 | Buy | \$439,000 | Darwin City | NT | | | premiere |
| 9 | 1350997 | Apartment | \$455,000 | Buy | \$455,000 | Darwin City | NT | | | premiere |
| 10 | 1350998 | Apartment | \$280,000 | Buy | \$280,000 | Darwin City | NT | | | premiere |
| | | | Openn Negotiation PRICE GUIDE | Buy | Openn Negotiation PRICE GUIDE | Darwin City | NT | | | premiere |
| 11 | 1350999 | Apartment | \$439,000 | Buy | \$439,000 | Darwin City | NT | | | premiere |
| 12 | 1351000 | Unit | \$775,000 | Buy | \$775,000 | Darwin City | NT | | | premiere |
| 13 | 1351001 | House | \$625,000 | Buy | \$625,000 | Darwin City | NT | | | premiere |
| 14 | 1351002 | Apartment | | | | Darwin City | NT | | | premiere |

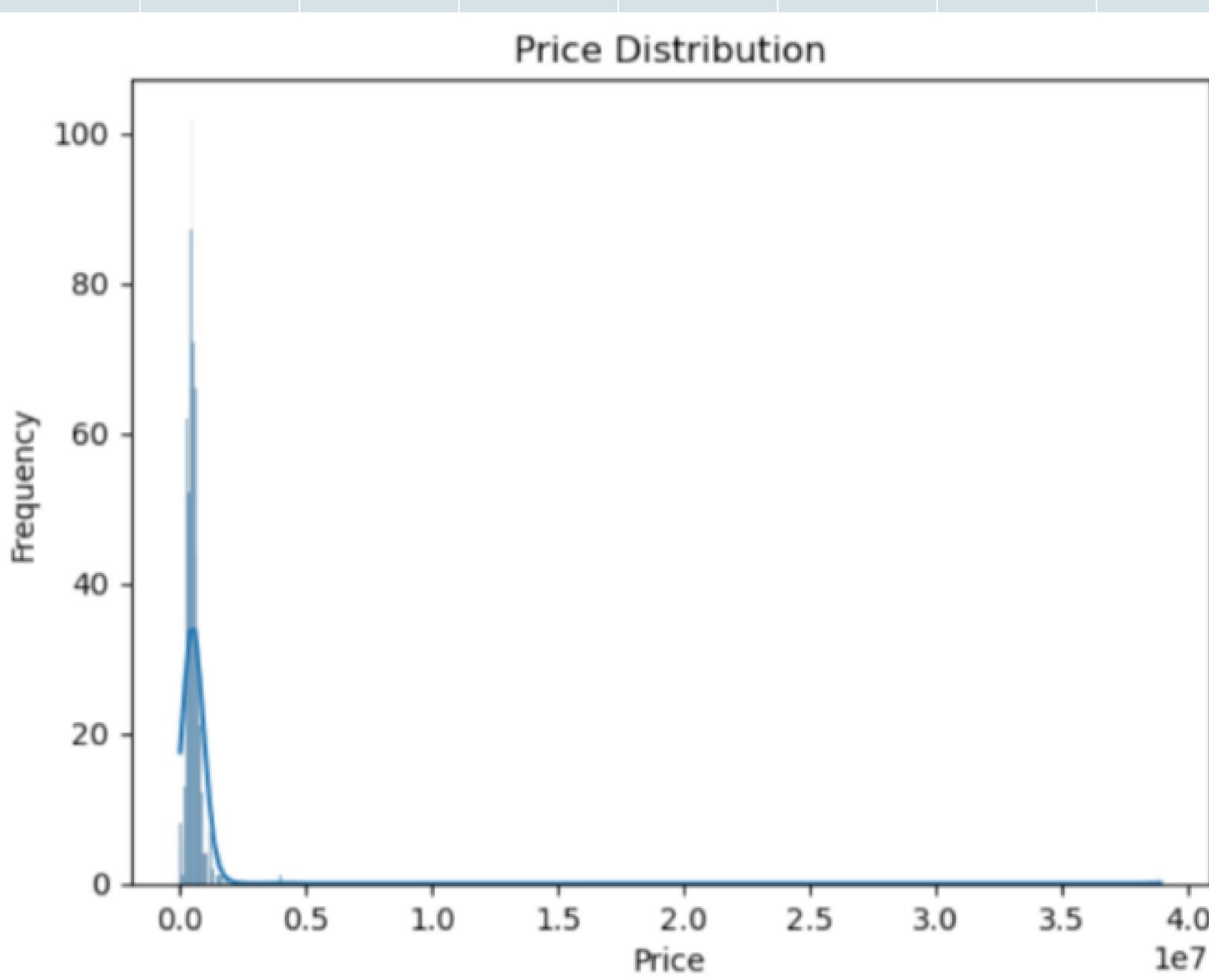
Shape of Dataset : 1000 rows : 27 Columns

- Dropped irrelevant columns ,
columns with constant values ('state', 'RunDate') , and
those with a high similarity to other columns ('price', 'location_name')
- Dropped rows with missing values
- Formatted data type of price column (Overs Over \$599,000 Considered) to 599000
- Identified number of outliers: 42 , removed those records
- Post-cleaning, we were left with 702 rows and 10 columns

Statistical Analysis Techniques Used

- Hypothesis Test
- Outlier Detection and Removal (IQR Method)
- Linear Regression Analysis
- Visualization through Scatter Plots and Bar Charts

Price distribution in Northern Territory



Question:

What is the distribution shape of real estate prices in the Northern Territory.

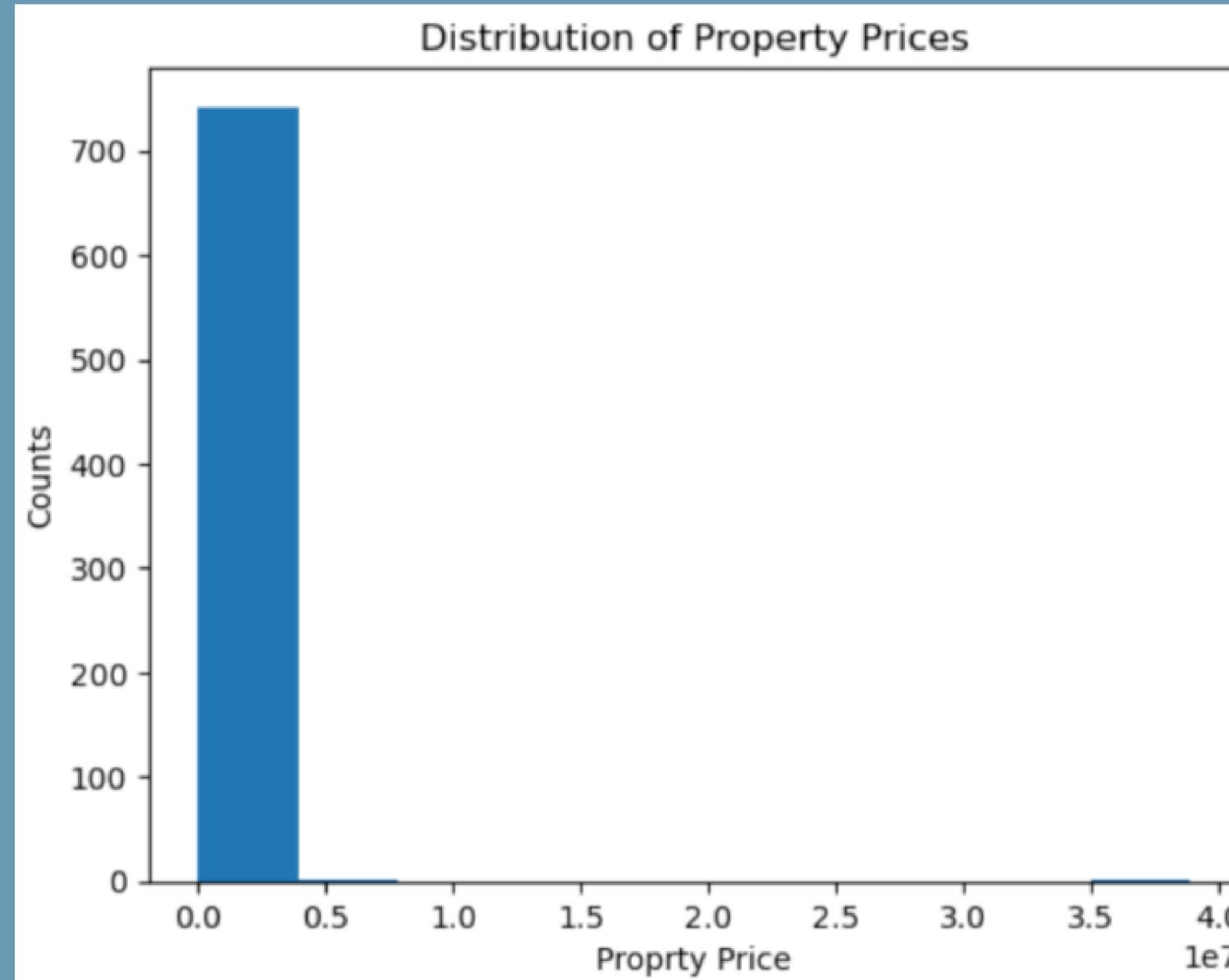
Null hypothesis:

Property prices are normally distributed.

Alternative hypothesis:

Property prices are not normally distributed

Price distribution in Northern Territory

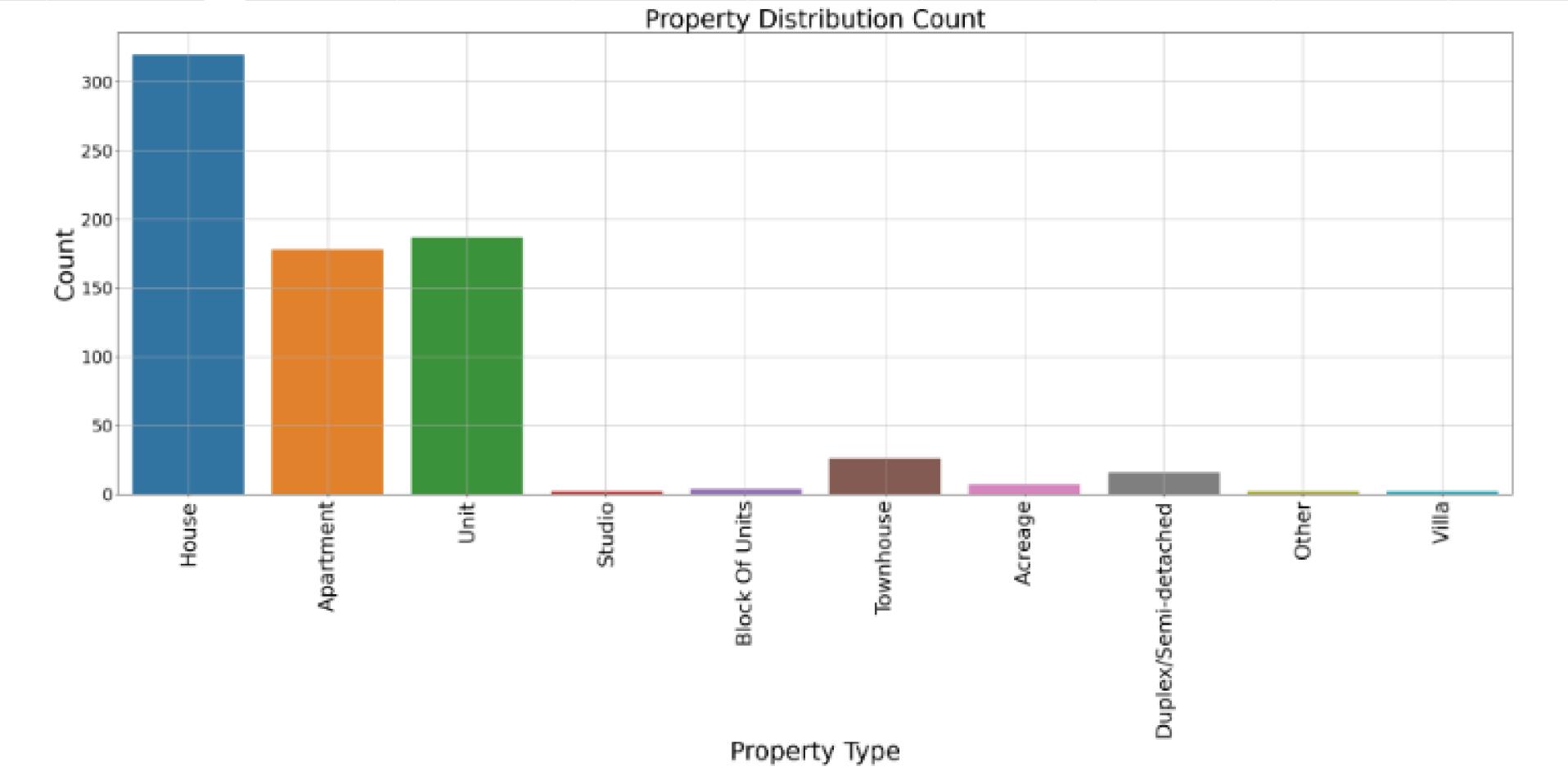


- Reject null hypothesis:
 - P-value of 0.000566
 - As the data is not normally distributed.
- Right-Skewed Distribution:
 - Most properties are priced at the lower end
- Most Common Price Range:
 - It lies in between 0 and 0.5 million AUD.
- Presence of Outliers:
 - A few high-priced properties extends up to 38.9 million AUD.

| | price | location_number | zip_code | bedroom_count | bathroom_count | parking_count |
|-------|-------------|-----------------|----------|---------------|----------------|---------------|
| count | 736.00 | 7.360000e+02 | 736.00 | 736.00 | 736.00 | 736.00 |
| mean | 581085.26 | 1.415119e+08 | 816.22 | 2.76 | 1.69 | 2.07 |
| std | 1437865.65 | 4.161221e+07 | 13.67 | 1.09 | 0.60 | 1.47 |
| min | 99950.00 | 1.085305e+08 | 800.00 | 0.00 | 1.00 | 0.00 |
| 25% | 383750.00 | 1.386080e+08 | 800.00 | 2.00 | 1.00 | 1.00 |
| 50% | 490000.00 | 1.389919e+08 | 812.00 | 3.00 | 2.00 | 2.00 |
| 75% | 612725.00 | 1.392750e+08 | 830.00 | 3.00 | 2.00 | 2.00 |
| max | 38900000.00 | 7.001996e+08 | 839.00 | 8.00 | 5.00 | 12.00 |

*1. Price Statistics: -Price Range: The minimum price is AUD 99,950 , maximum price is 38900000.00

Which property type has the highest demand in the Northern Territory market



- Under the sample data, house is the most common property type (> 300 counts).
- Apartment and Unit have similar counts (slightly below 200).
- Townhouse has a notable count but is significantly less than House, Apartment, and Unit.
- Other property types (Studio, Acreage, Duplex/Semi-detached, Other, Villa) have very low counts, with Villa being the least common.

Conclusion:

- Houses dominate the market, indicating higher availability and preference for this property type.



Count of the property type

House

House is the highest
number in NT

Apartment

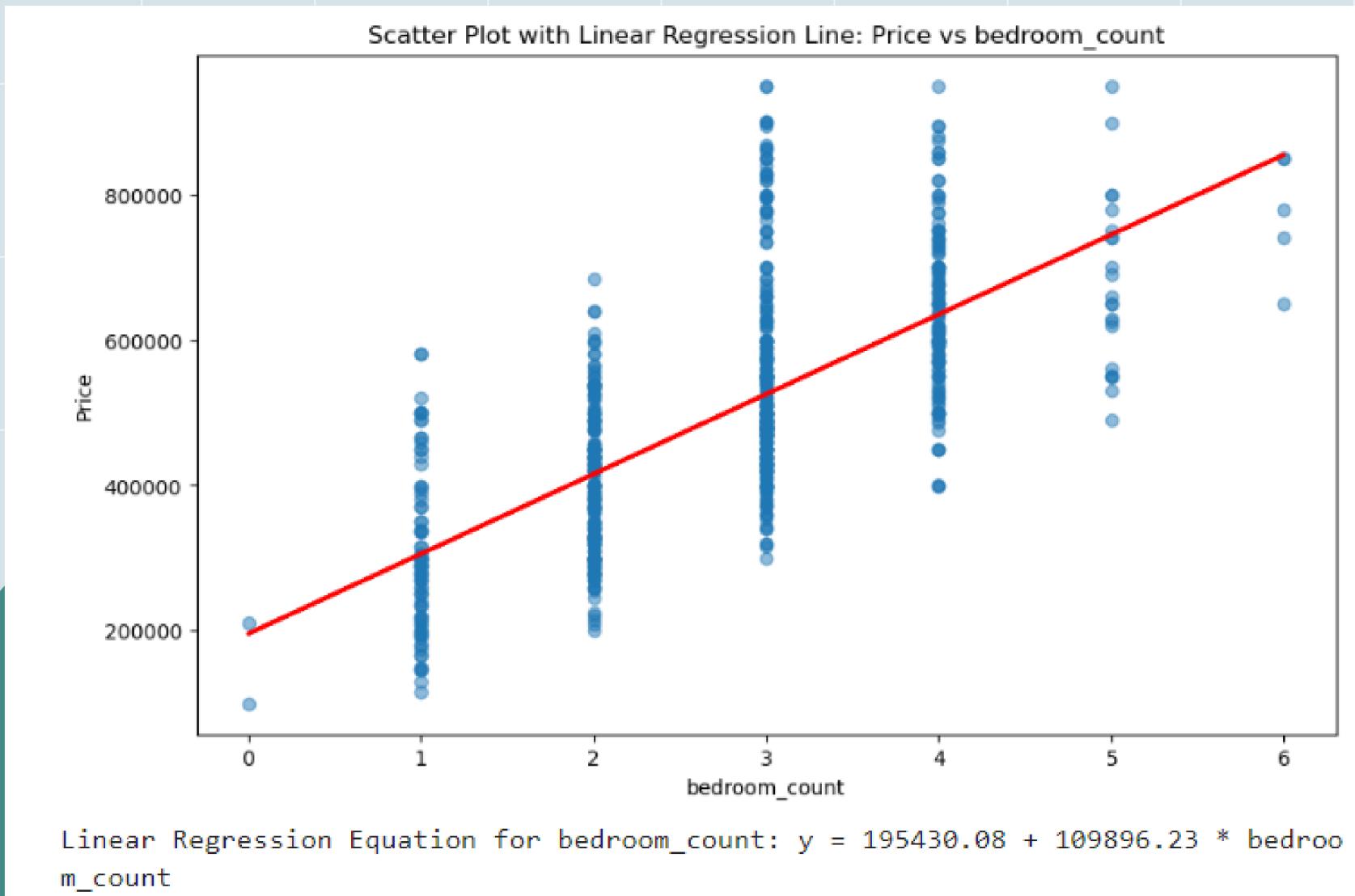
Unit

Townhouse
and others

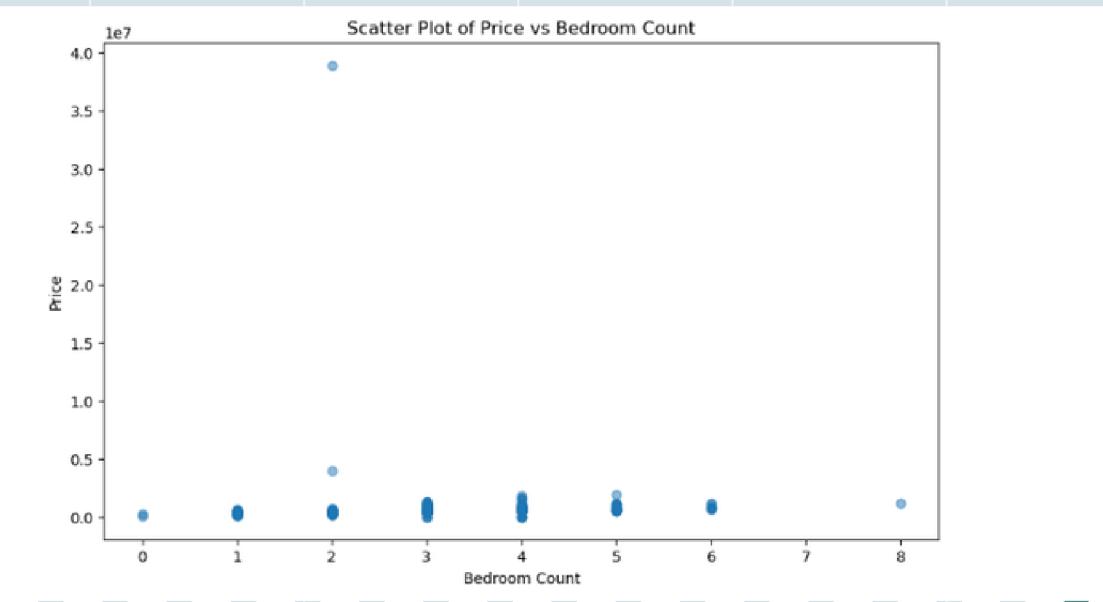


How does different features influence the price of the property?

Is there any correlation between number of bedrooms and property price?



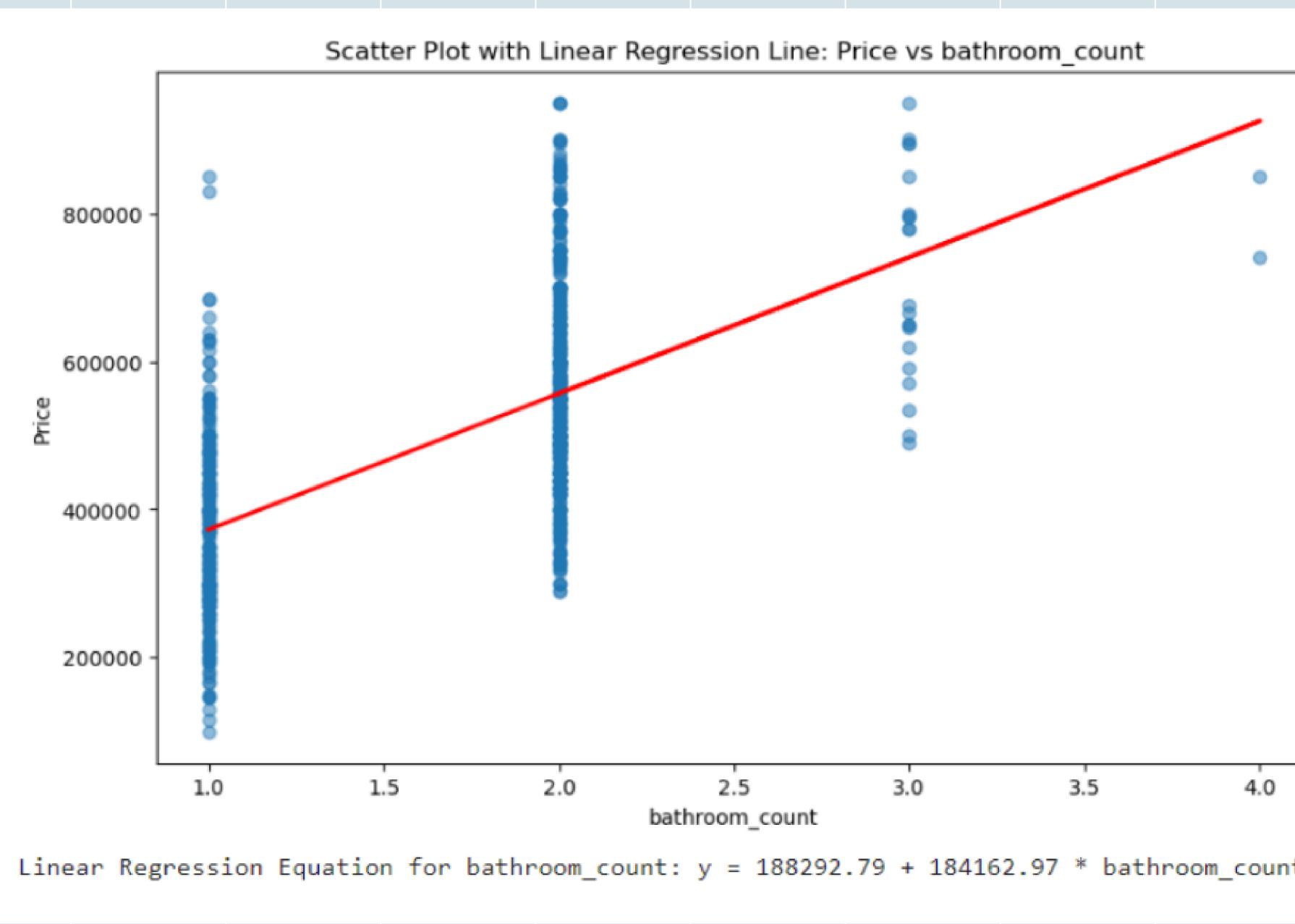
- There are extreme values that deviate significantly from the rest of the data; we will remove those outliers.
- Removing Outliers: Helps in improving the model accuracy (below scatterplot)
- formula: $y = 195430.08 + 109896.23 * \text{bedroom count}$
1 bedroom more, \$109k more on the property price
- Clear positive correlation observed.
- Conclusion: More bedrooms generally lead to higher property prices.



How does different features influence the price of the properties ? And how it correlated ?

Bathroom vs Price

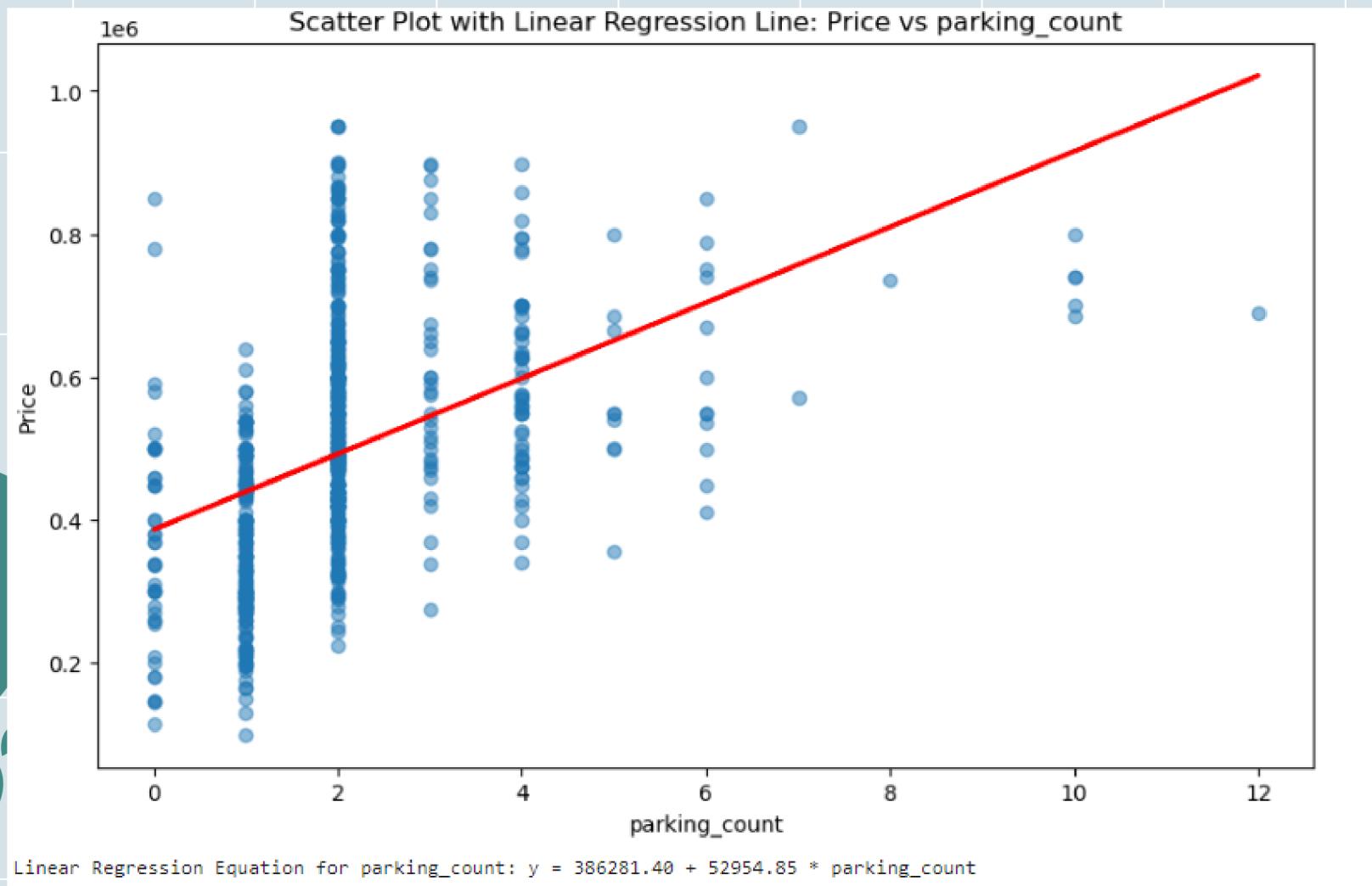
Is there any correlation between number of bathrooms and property price?



- Positive correlation observed.
- $y = 188292.79 + 184162.97 * \text{bathroom count}$
- 1 bathroom more, \$184k more on the property price
- Conclusion: More bathrooms correlate with higher property prices.

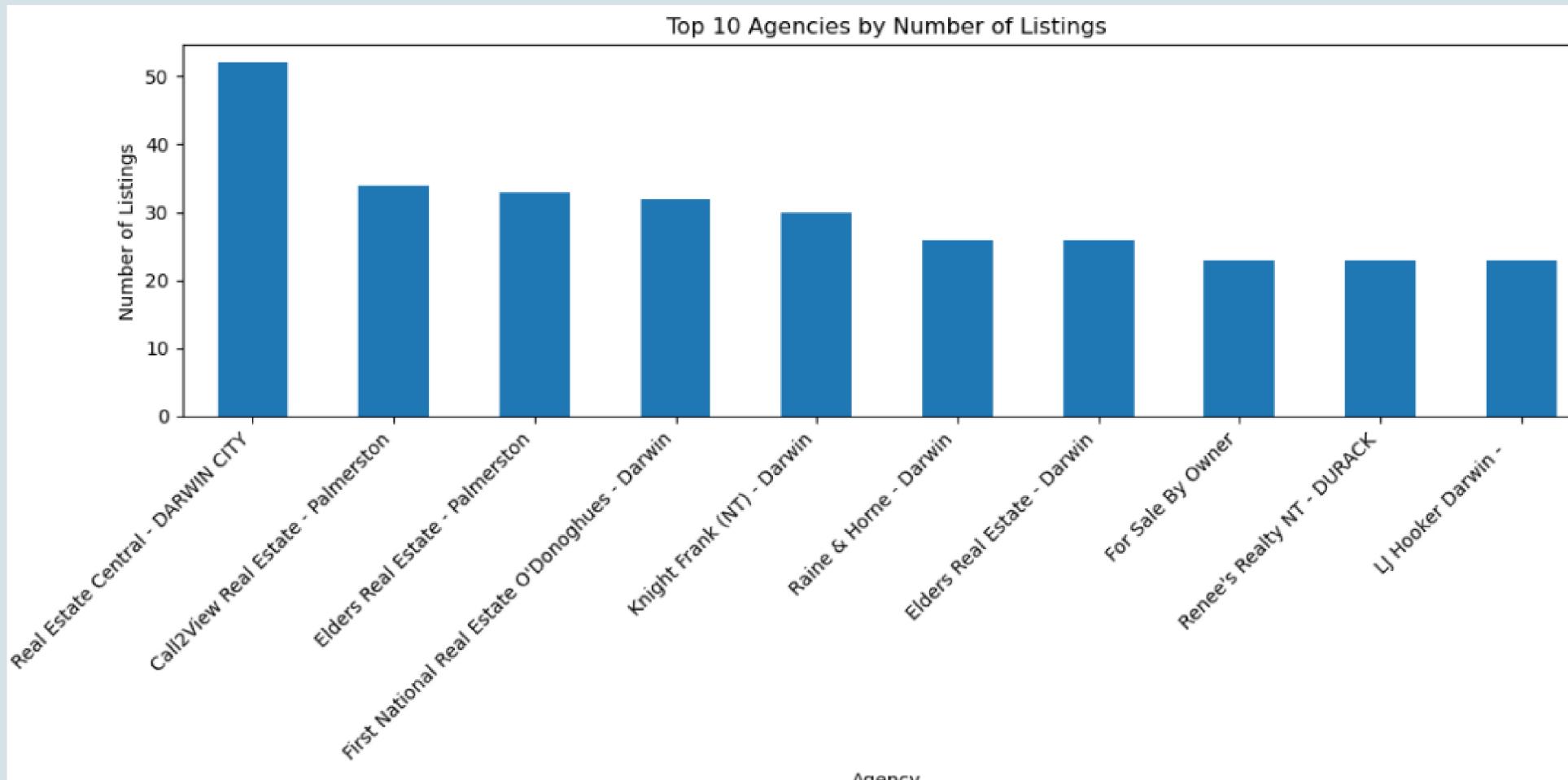
How does different features influence the price of the properties ?And how it correlated ? Parking vs Price

Is there any correlation between number of parking and property price?

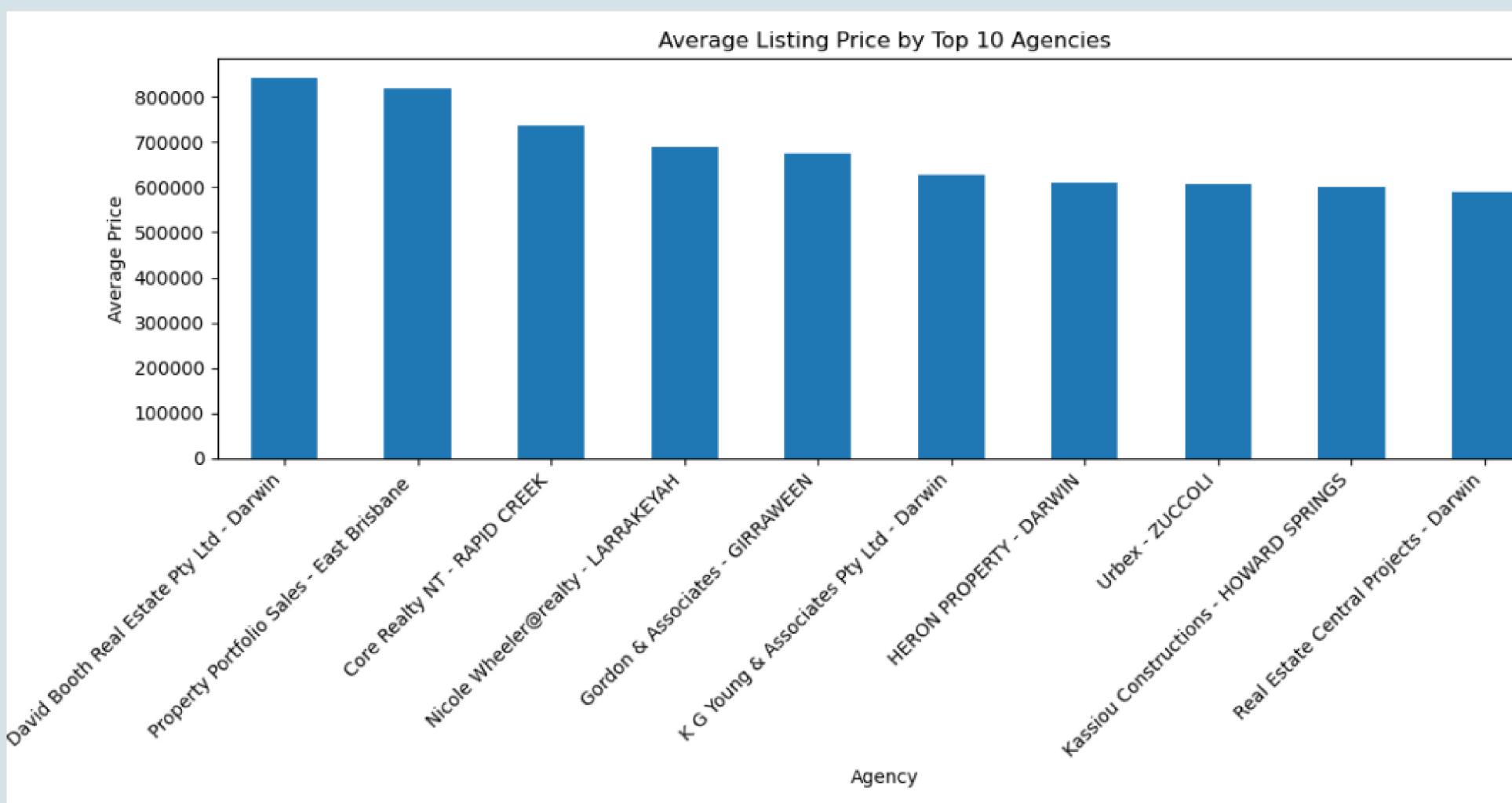


- Positive correlation observed.
- $y = 386281.40 + 52954.85 * \text{parking count}$
- 1 parking more, \$52k more on the property price
- Conclusion: More parking spaces correlate with higher property prices.

Kajal



Here is the quick comparison of the market share of top 10 real estate agencies in terms of active property listings.



Average listing price of top 10 real estate agencies in the area, with some agencies seemingly specializing in higher-end properties while others focus on more moderately priced listings.

Summary

- As part of the project scope, the project team has identified dataset that has real estate information of Northern Territory region. The data was then cleansed, explored and analyzed to understand the patterns, relationships and trends within this region.
- Three focus areas of our analysis Price Distribution (selected properties ranging from 99,950 to 38million), *Property Distribution, Agency distribution*
- ***The data was presented visually by various graphs as you have seen in the previous slides,***
 - From the **Property distribution** graph, it is clear that houses are the highest in demand followed by units and apartments. Indicating that units are more expensive than apartments (Alternative hypothesis --> true)
 - **Price distribution** graph shows that prices are highly concentrated in lower end and few properties prices are in the high range, proving that the price distribution is not normally distributed (alternative hypothesis --> true)



Summary

- Various linear regression plots are created with price in comparison to various features like bedroom count, bathrooms and they are strongly correlated with the price
 - Positive slope: The red regression line has a clear upward slope, indicating that as the number of bed/bathrooms increases, the price also increases
- Finally, some analysis was carried out and presented in terms of agencies operating in NT along with their market share in each segment.

Overall, the data presented in this slide pack can be effectively used by real investors, which helps them make an informative decision.





Thank you



SALE

