Melbourne Housing Market Analysis: Data-Driven Insights on Property Prices

Exploratory Data Analysis and Visualization of Key Market Drivers

Team Information:

- Elvin MAHMUDZADA
- Kamal MUSTAFAYEV
- Huseyn BABAYEV

Abstract

- Analyzed 23,547 Melbourne housing sales records
- Research question: Which property features most influence sale prices, and how do regional variations and seller performance impact the market?
- Methods: Data cleaning, exploratory analysis, and visualization
- Key factors identified: Location, number of rooms, property type, distance from CBD, and seller performance.
- Findings provide insights for buyers, sellers, investors, and real estate professionals

Motivation

- Melbourne housing market is dynamic and challenging for buyers/investors
- Understanding price determinants reduces financial risks
- Aim: Identify critical price drivers for diverse stakeholders
- Insights help homebuyers, investors, and policymakers make data-driven decisions

Dataset

- 23,547 Melbourne property transactions analyzed
- 21 attributes including suburb, address, rooms, price, sale method, year built
- Data sourced from public real estate records
- Dataset provides a comprehensive snapshot for studying market price influences

Melbourne Housing Market Dataset

Dataset

Column	Short Description
Suburb	Property neighborhood
Address	Street address
Rooms	Bedrooms (primary)
Type	h=house, t=townhouse, u=unit
Price	Sale price (AUD)
Method	Sale method (S=sold, etc.)
SellerG	Real estate agent
Date	Sale date
Distance	Km from CBD
Postcode	Postal code
Bedroom2	Bedrooms (alternative)
Bathroom	Bathrooms
Car	Car spaces
Landsize	Land area (sqm)
BuildingArea	Building size (sqm)
YearBuilt	Build year
CouncilArea	Local government area
Regionname	Regional classification
Propertycount	Properties in suburb

Data preparation and cleaning

Data Loading & Initial Inspection

- Which functions are used?

Standardizing Missing Values with NaN

- How?

Data Loading and Initial Cleaning

- Imported pandas, numpy, datetime
- Loaded 'Melbourne.csv' into DataFrame
- Replaced empty strings with NaN using replace()
- Counted and dropped duplicate rows with drop_duplicates()

Checking for Missing Values:

- Defined function **check_missing_columns()** to identify columns with NaN
- Printed list of columns with missing values

```python
# Function to check for missing values
def check_missing_columns(df):
    missing_cols = [col for col in df.columns if df[col].isnull().any()]
    return missing_cols

# Initial check for missing values
cols_with_missing = check_missing_columns(df)
print("Columns with missing values:", cols_with_missing)
```

Filling Missing Values

- Filled Suburb, Regionname, and CouncilArea using grouped mode fallbacks
- Imputed Price (mean by Suburb → Regionname) and Propertycount (mode/median with fallbacks)
- Filled YearBuilt, Bathroom, and Car with global statistics
- Replaced zeros in Distance, Landsize, BuildingArea, Postcode with grouped medians (Suburb/Regionname)
- Set Bedroom2 missing or zero values from Rooms; replaced invalid bathroom zeros by grouped media

```
# Price: mean by Suburb, fallback to Regionname
df['Price'] = df['Price'].fillna(df.groupby('Suburb')['Price'].transform(lambda x: round(x.mean(), 1)))
df['Price'] = df['Price'].fillna(df.groupby('Regionname')['Price'].transform(lambda x: round(x.mean(), 1)))
```

Final Adjustments & Validation

- Dropped **Lattitude** and **Longtitude**
- Renamed **Rooms** → **Bedroom** and corrected invalid YearBuilt values
- Validated and cleaned Date, YearBuilt, Price, and Postcode
- Saved cleaned dataset as **Melbourne_cleaned.csv** and displayed final preview

```
# Step 4: Drop unnecessary columns
df.drop(columns=['Lattitude', 'Longtitude'], inplace=True)

# Rename Rooms to Bedroom (after all Rooms-based fills)
df = df.rename(columns={'Rooms': 'Bedroom'})
```

```
# Save cleaned data
df.to_csv('Melbourne_cleaned.csv', index=False)
print(df.head())
```

Research Questions

- What key property features and location factors most influence housing prices in Melbourne?
- How do market dynamics vary across different suburbs and regions within Melbourne?
- Can we identify top-performing properties at both local (suburb) and broader (region) levels for better investor and buyer guidance?

Insight 1: Properties with Larger Land Size

- Filtered properties with land sizes above average for deeper focus
- Sorted these properties by **region, suburb, distance from city center, year built, and price**
- Aim: Understand patterns in premium properties with larger land sizes

This helps highlight key features influencing property value beyond size alone

[26]:

	Suburb	Address	Bedroom	Type	Price	Method	SellerG	Date	Distance	Postcode	Bedroom2	Bathroom	Car	Landsize	BuildingArea	YearBuilt
0	Bayswater	2/1 Orchard Rd	2	h	661000.0	S	Barry	26-08-2017	23.2	3153	2	2	1	735.0	103.0	2007
1	Bayswater	3 Grieve St	3	h	824000.0	S	One	03-09-2017	23.2	3153	3	1	2	725.0	112.0	1980
2	Bayswater	15 Clifford St	4	h	705000.0	PI	Biggin	27-05-2017	23.2	3153	4	2	4	765.0	141.0	1975
3	Bayswater	85 Farnham Rd	4	h	709690.5	SP	Barry	08-07-2017	23.2	3153	3	1	7	773.0	120.0	1975
4	Bayswater	3/23 Begonia Av	2	h	425000.0	VB	McGrath	19-08-2017	23.2	3153	2	1	1	735.0	63.0	1970
5	Bayswater	1/33 Begonia Av	2	u	430000.0	SP	Stockdale	17-06-2017	23.2	3153	3	1	1	735.0	126.5	1970
6	Bayswater	1/8 Tracey St	3	h	610000.0	S	McGrath	27-05-2017	23.2	3153	3	1	1	735.0	126.5	1970
7	Bayswater	6 Larne Av	2	u	620000.0	S	First	22-07-2017	23.2	3153	3	1	1	735.0	126.5	1970
8	Bayswater	16 Wiltshire Av	4	h	640000.0	VB	iTRAK	03-06-2017	23.2	3153	3	1	1	735.0	126.5	1970
9	Bayswater	5 Susan St	4	h	709690.5	PI	Philip	24-06-2017	23.2	3153	4	2	3	736.0	162.0	1970

Insight 2 : Top Properties Recommendations

- Extracted **the top ten properties** overall from the filtered and sorted list
- Selected a random property from the top ten for demonstration/recommendation
- Allows showcasing best housing options from an analytical standpoint
- Connects data insights to practical uses like buyer recommendations

Key words: Bias, Buyer's perspective

Randomly recommended house

```
top10 = df_sorted[0:10]
random_index = rd.randint(0, 9)
random_house = top10.iloc[random_index]
print(random_house)
```

```
Suburb                        Bayswater
Address                    2/1 Orchard Rd
Bedroom                               2
Type                                  h
Price                          661000.0
Method                                S
SellerG                           Barry
Date                         26-08-2017
Distance                           23.2
Postcode                           3153
Bedroom2                              2
Bathroom                              2
Car                                   1
Landsize                          735.0
BuildingArea                      103.0
YearBuilt                          2007
CouncilArea                        Knox
Regionname         Eastern Metropolitan
Propertycount                      5030
```

Insight 3 — Top 10 Houses per Suburb

- Grouped the filtered data by suburb and took the top ten houses per suburb – location-specific
- Allows focused analysis on high-value properties in each neighborhood
- Supports localized market insights, crucial for buyers targeting specific suburbs

Real estate values change a lot from neighborhood to neighborhood, so looking by suburb gives us a local view.

[28]:

```
<bound method NDFrame.head of            Suburb          Address  Bedroom Type     Price Method   SellerG  \
0        Bayswater   2/1 Orchard Rd        2    h  661000.0      S     Barry
1        Bayswater      3 Grieve St        3    h  824000.0      S       One
2        Bayswater    15 Clifford St       4    h  705000.0     PI    Biggin
3        Bayswater    85 Farnham Rd        4    h  709690.5     SP     Barry
4        Bayswater  3/23 Begonia Av        2    h  425000.0     VB   McGrath
...            ...              ...      ...  ...       ...    ...       ...
2285  Melton South     5 Leggatt St        3    h  370000.0      S  Reliance
2286  Melton South     48 Manson Dr        3    h  385000.0      S       YPA
2287  Melton South     97 Exford Rd        3    h  392500.0      S     Raine
2288  Melton South      1 Fraser St        3    h  395000.0      S       YPA
2289  Melton South     23 Neerim St        4    h  426000.0      S     Raine

            Date  Distance  Postcode  Bedroom2  Bathroom  Car  Landsize  \
0     26-08-2017      23.2      3153         2         2    1     735.0
1     03-09-2017      23.2      3153         3         1    2     725.0
2     27-05-2017      23.2      3153         4         2    4     765.0
3     08-07-2017      23.2      3153         3         1    7     773.0
4     19-08-2017      23.2      3153         2         1    1     735.0
...          ...       ...       ...       ...       ...  ...       ...
2285  16-09-2017      29.8      3338         3         1    2     599.0
2286  29-07-2017      29.8      3338         3         1    2     670.0
2287  16-09-2017      29.8      3338         3         1    2     599.0
2288  26-08-2017      29.8      3338         3         1    1     649.0
2289  15-07-2017      29.8      3338         4         2    3     655.0

      BuildingArea  YearBuilt CouncilArea           Regionname  Propertycount
0            103.0       2007        Knox  Eastern Metropolitan           5030
1            112.0       1980        Knox  Eastern Metropolitan           5030
2            141.0       1975        Knox  Eastern Metropolitan           5030
3            120.0       1975        Knox  Eastern Metropolitan           5030
4             63.0       1970        Knox  Eastern Metropolitan           5030
...            ...        ...         ...                  ...            ...
2285         119.0       1975      Melton      Western Victoria           4718
2286         124.0       1975      Melton      Western Victoria           4718
2287         171.0       1975      Melton      Western Victoria           4718
2288         105.0       1975      Melton      Western Victoria           4718
2289         142.0       1975      Melton      Western Victoria           4718

[2290 rows x 19 columns]>
```

Insight 4 — Top 10 Houses per Region

- Grouped filtered data **by regional zones** and selected top ten per region
- Provides aggregated market snapshot over larger geographic areas than suburbs
- Useful for broader regional market trends and strategic planning

Key point: Suburbs give very local insights, but regions help investors or developers see bigger market trends.

[29]:

```
<bound method NDFrame.head of           Suburb          Address  Bedroom Type     Price Method  \
0      Bayswater   2/1 Orchard Rd        2    h  661000.0      S
1      Bayswater      3 Grieve St        3    h  824000.0      S
2      Bayswater   15 Clifford St        4    h  705000.0     PI
3      Bayswater    85 Farnham Rd        4    h  709690.5     SP
4      Bayswater  3/23 Begonia Av        2    h  425000.0     VB
..           ...              ...      ...  ...       ...    ...
75        Melton       14 Musk Ct        3    h  295000.0     SP
76        Melton      1 Irving Rd        4    h  400000.0      S
77        Melton    53 Barries Rd        3    h  420718.8     SN
78        Melton      28 Atkin St        4    h  550000.0     PI
79  Melton South     69 Andrew St        3    h  375467.4     PI

          SellerG        Date  Distance  Postcode  Bedroom2  Bathroom  Car  \
0           Barry  26-08-2017      23.2      3153         2         2    1
1             One  03-09-2017      23.2      3153         3         1    2
2          Biggin  27-05-2017      23.2      3153         4         2    4
3           Barry  08-07-2017      23.2      3153         3         1    7
4         McGrath  19-08-2017      23.2      3153         2         1    1
..            ...         ...       ...       ...       ...       ...  ...
75  PRDNationwide  27-05-2017      31.7      3337         3         1    4
76            YPA  03-06-2017      31.7      3337         4         2    2
77          Barry  22-07-2017      31.7      3337         3         1    2
78          Ryder  08-07-2017      31.7      3337         4         2    2
79          Raine  08-07-2017      29.8      3338         3         2    2

    Landsize  BuildingArea  YearBuilt CouncilArea           Regionname  \
0      735.0        103.00       2007        Knox  Eastern Metropolitan
1      725.0        112.00       1980        Knox  Eastern Metropolitan
2      765.0        141.00       1975        Knox  Eastern Metropolitan
3      773.0        120.00       1975        Knox  Eastern Metropolitan
4      735.0         63.00       1970        Knox  Eastern Metropolitan
..       ...           ...        ...         ...                   ...
75     600.0        122.86       1975      Melton      Western Victoria
76     643.0         30.00       1970      Melton      Western Victoria
77     780.0        122.86       1970      Melton      Western Victoria
78    1241.0        194.00       1970      Melton      Western Victoria
79     699.0         96.00       2000      Melton      Western Victoria
```

Which methods were used?

1. **Data Cleaning & Imputation:** Addressed missing and inconsistent values using targeted imputation (mean, median, mode) based on logical grouping (suburb, region, postcode) to restore data integrity without losing local context.

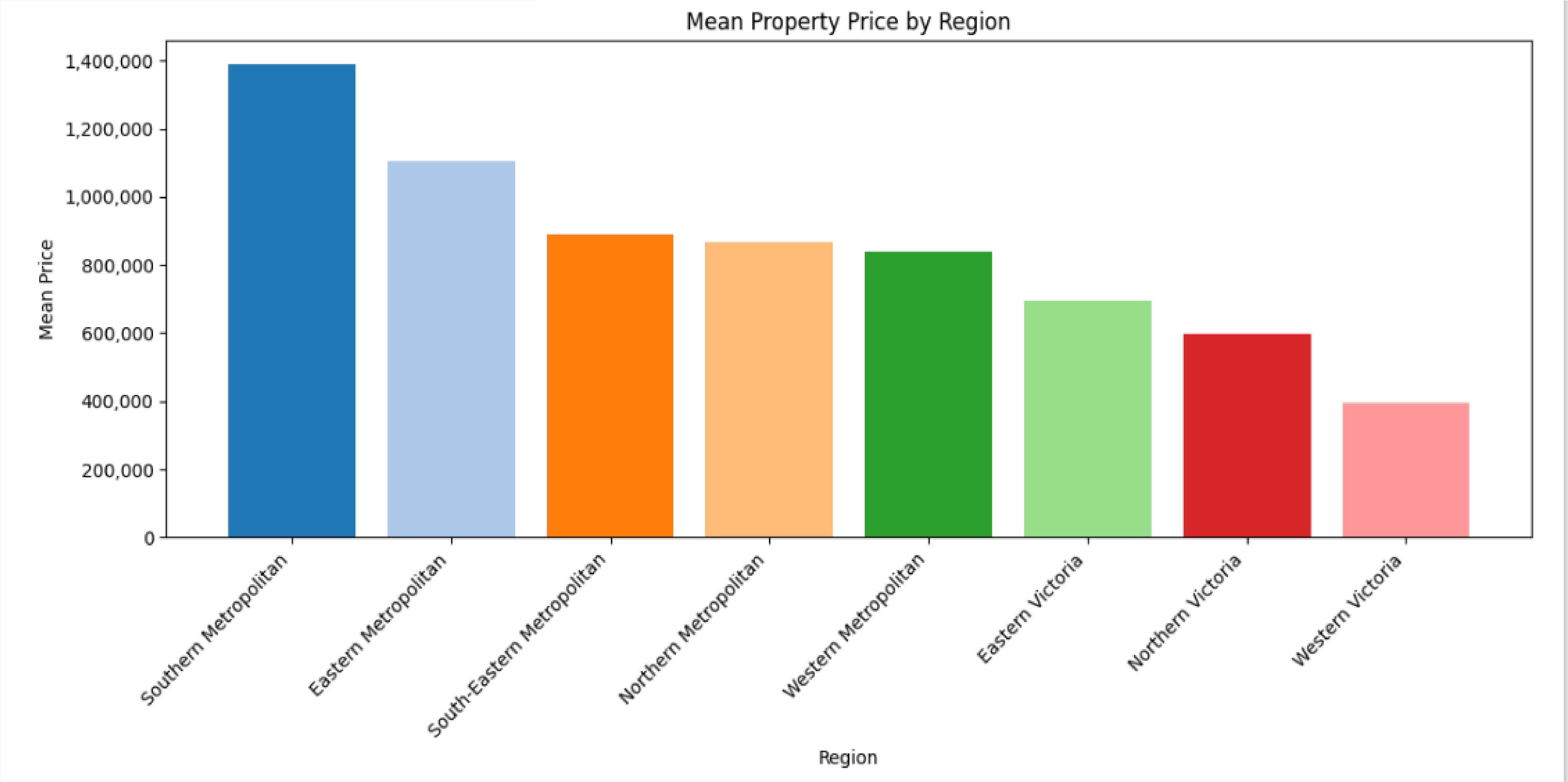
2. **Exploratory Data Analysis (EDA):** Performed granular grouping, sorting, and aggregation by location and key features (land size, year built, price) to reveal underlying patterns and market segments.

3. **Ranking & Filtering:** Focused on properties with above-average land size to isolate premium market segments. Sorted by multiple factors to prioritize important drivers of price and desirability.

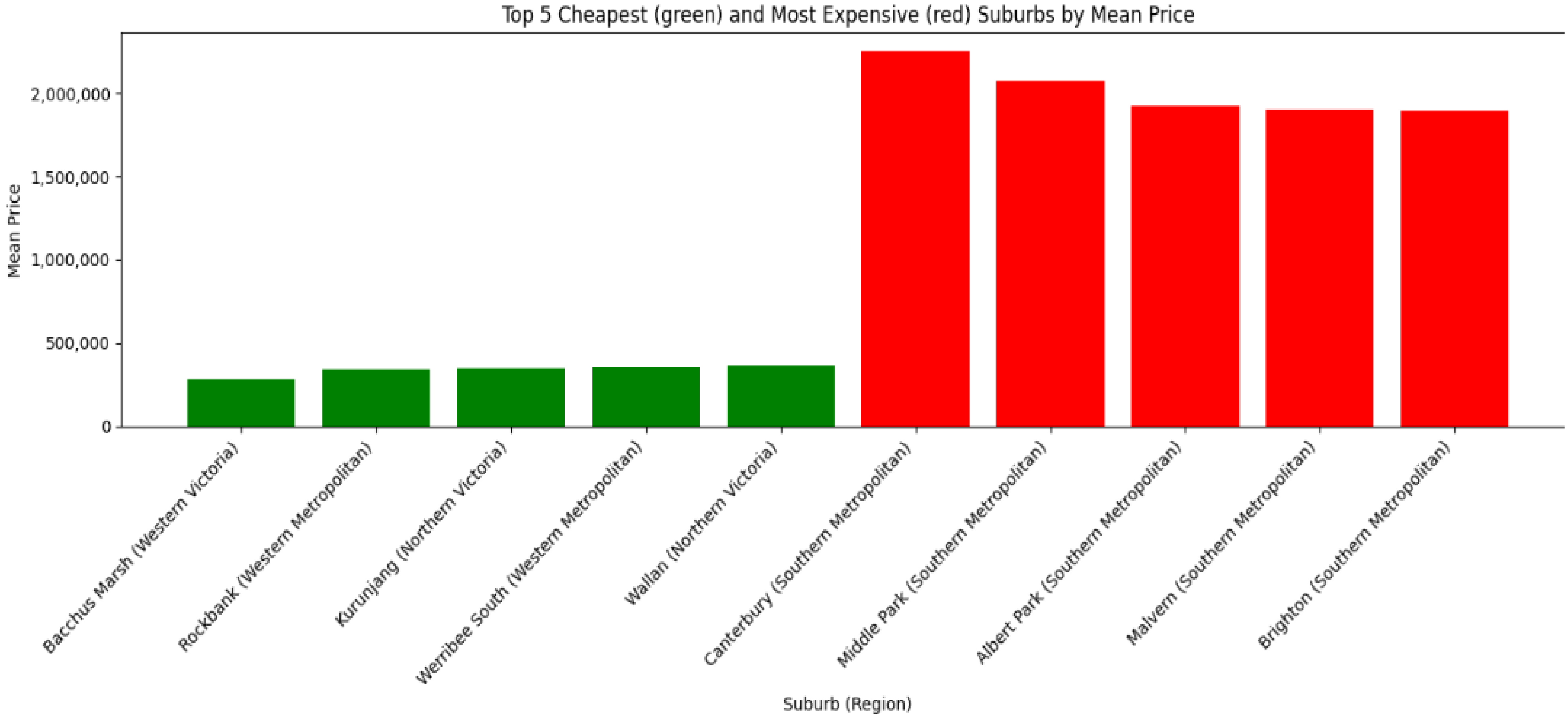
4. **Top Property Identification:** Grouped by suburb and region to identify localized and broader top-performing houses, balancing micro and macro market views for actionable insights.

5. **Business Alignment:** Methods chosen reflect market realities, enabling insights meaningful for investors, buyers, and strategic planners in the Melbourne housing market context.

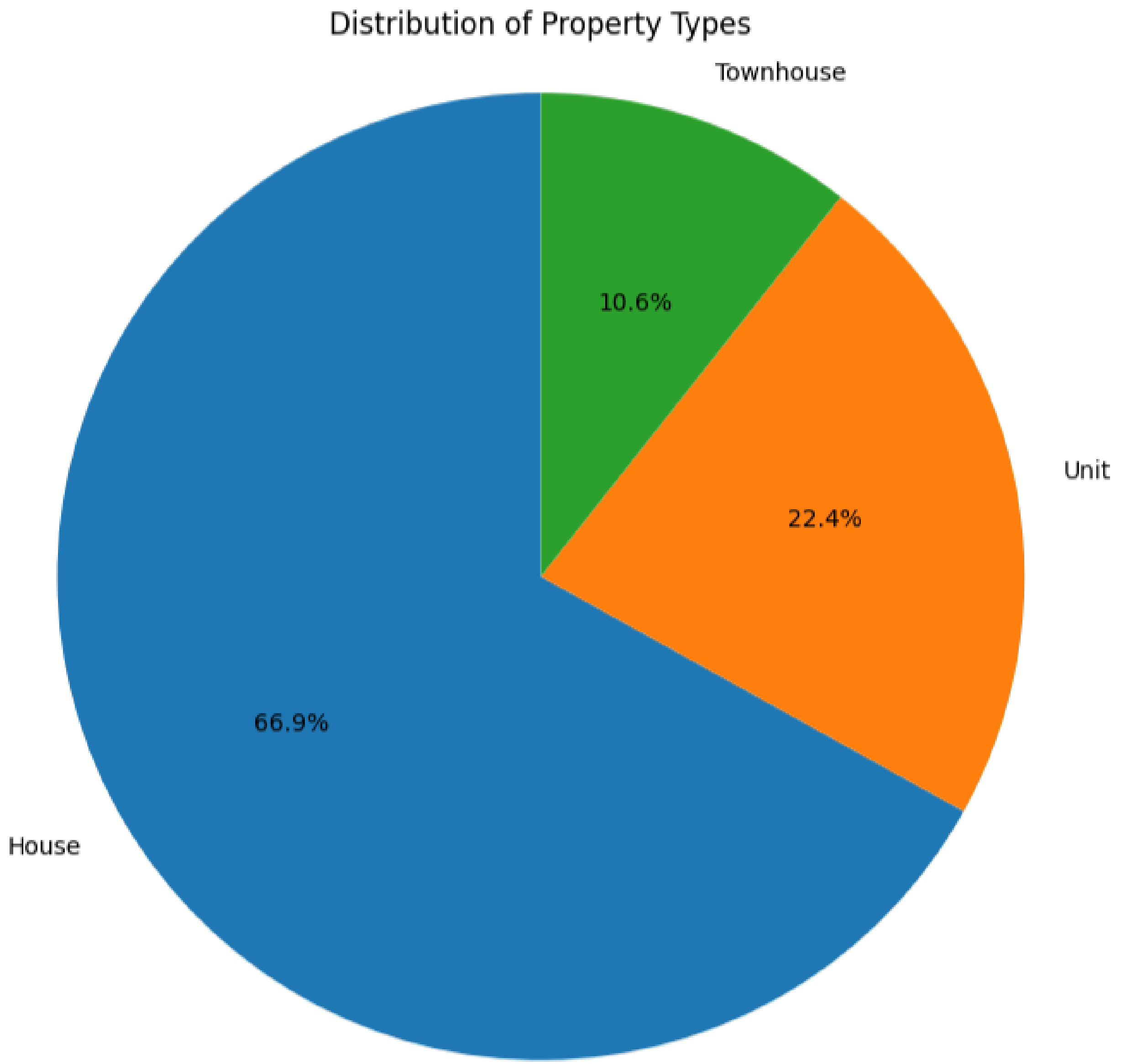
Finding 1: Mean Property Prices by Region



Finding 2: Cheapest and Most Expensive Suburbs

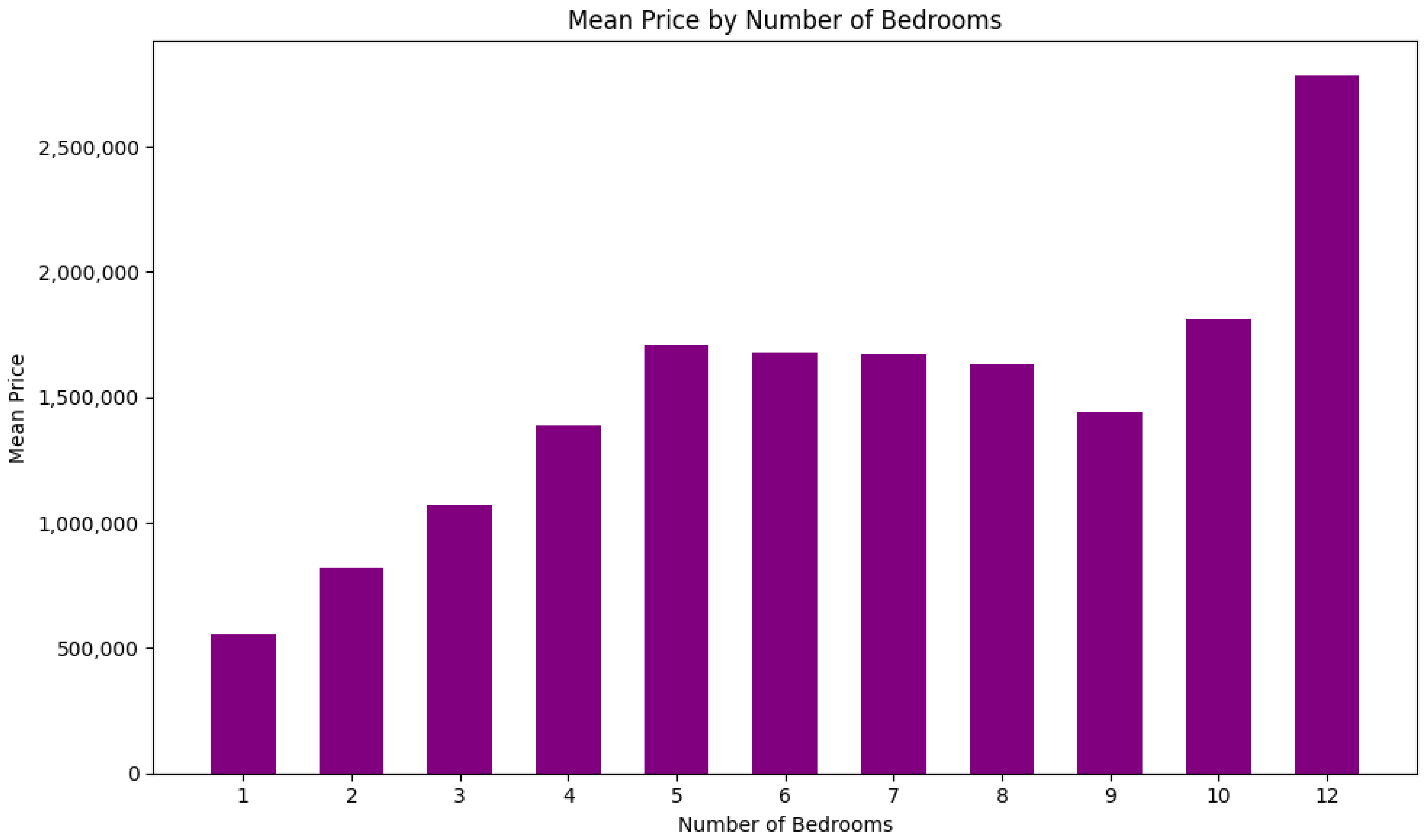


Finding 3: Distribution of Property Types

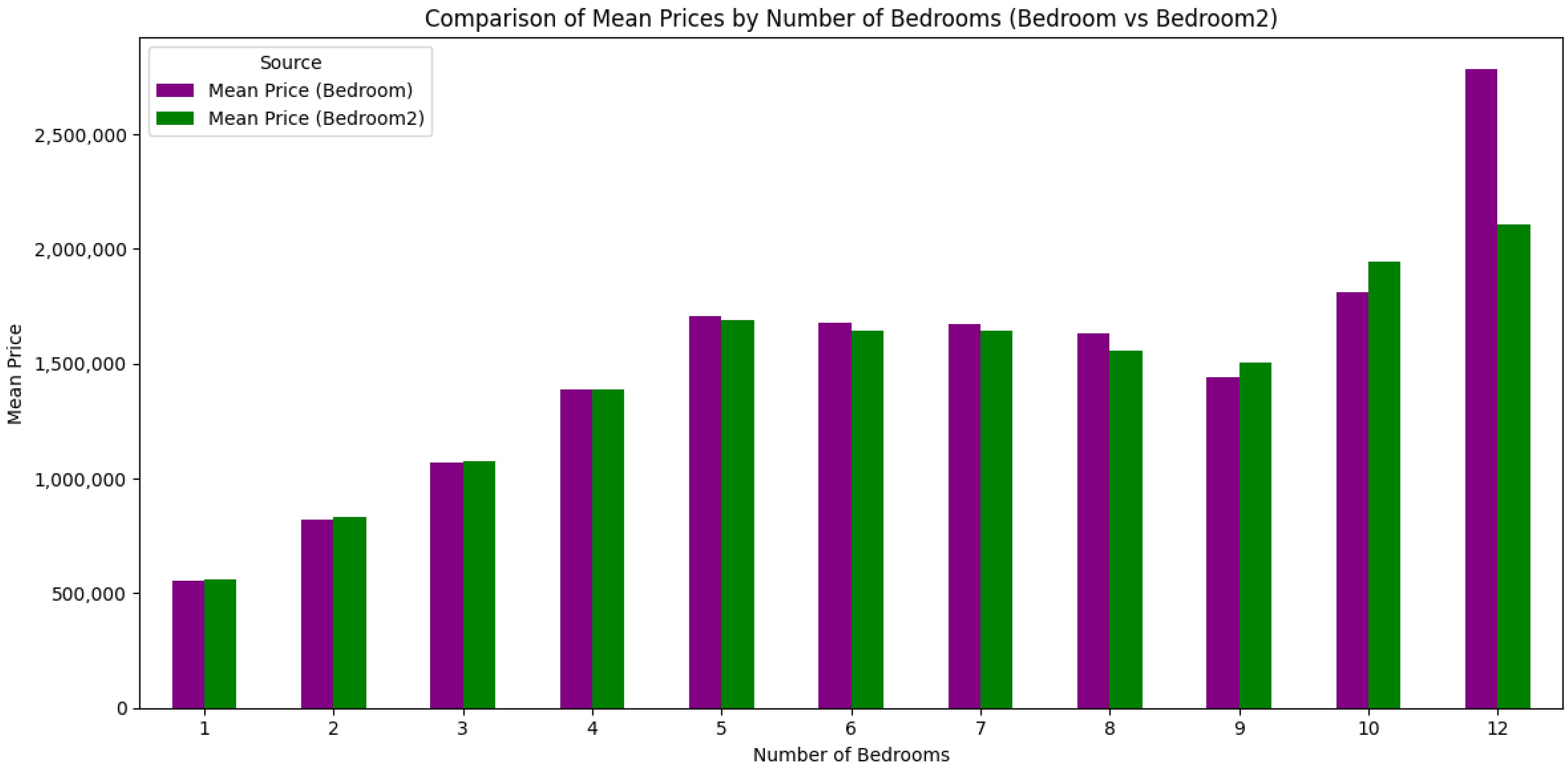


Finding 4: Mean Price by Property Type

Finding 5: Price/Bedrooms relationship

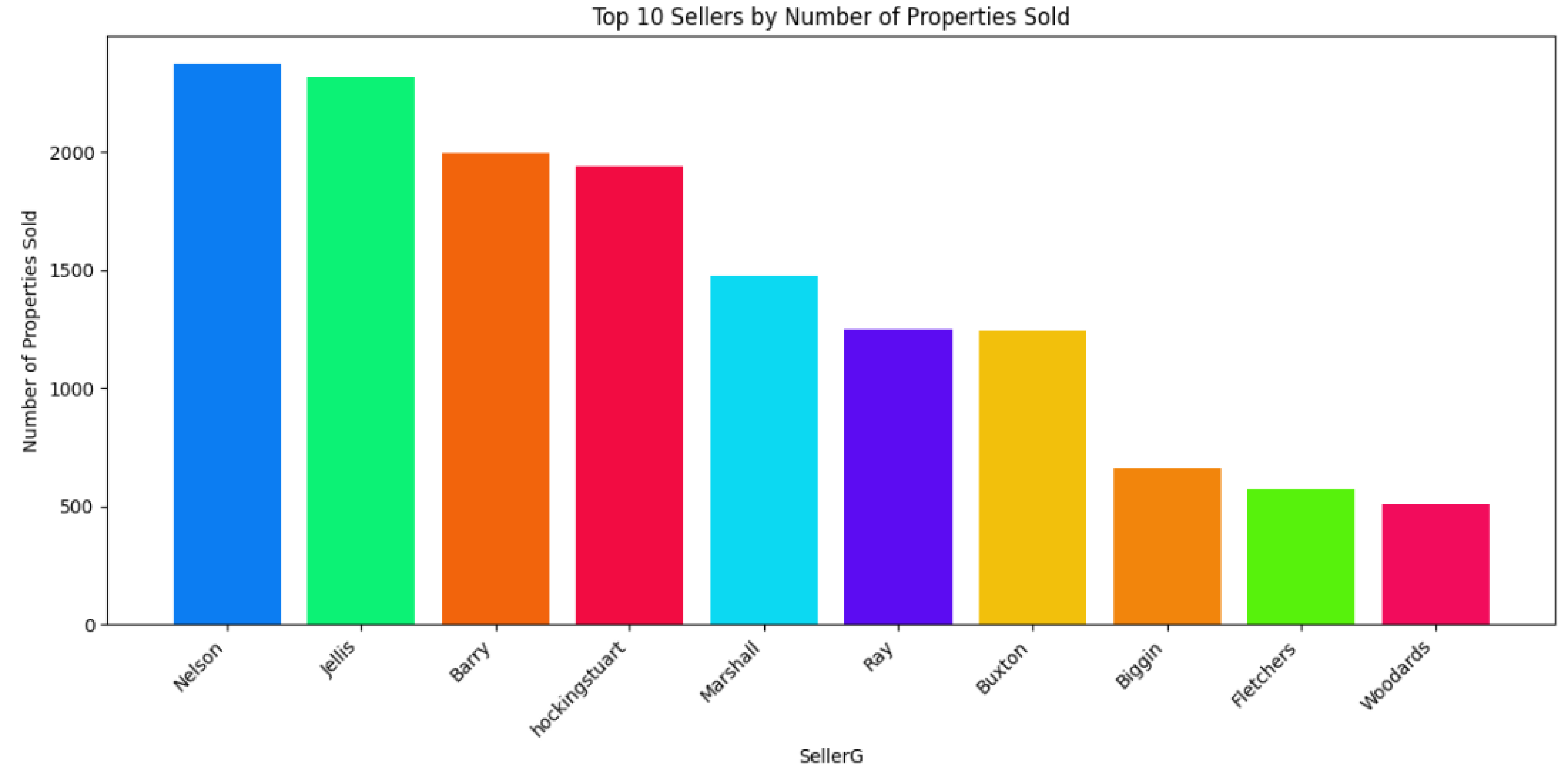


Finding 6: Source “Bedroom” vs. source “Bedroom2”

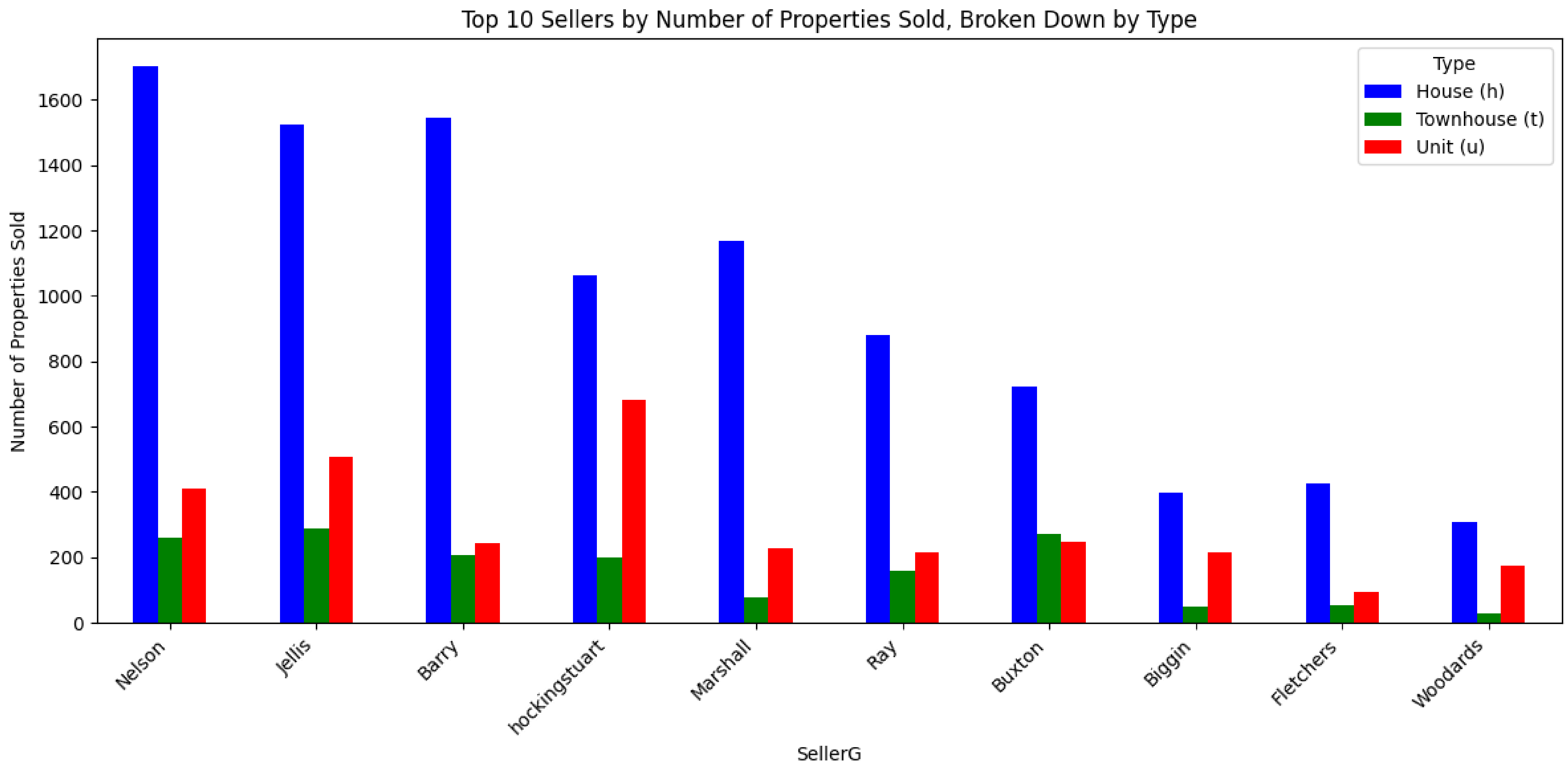


Finding 7: Monthly Price Trends

What about sellers?

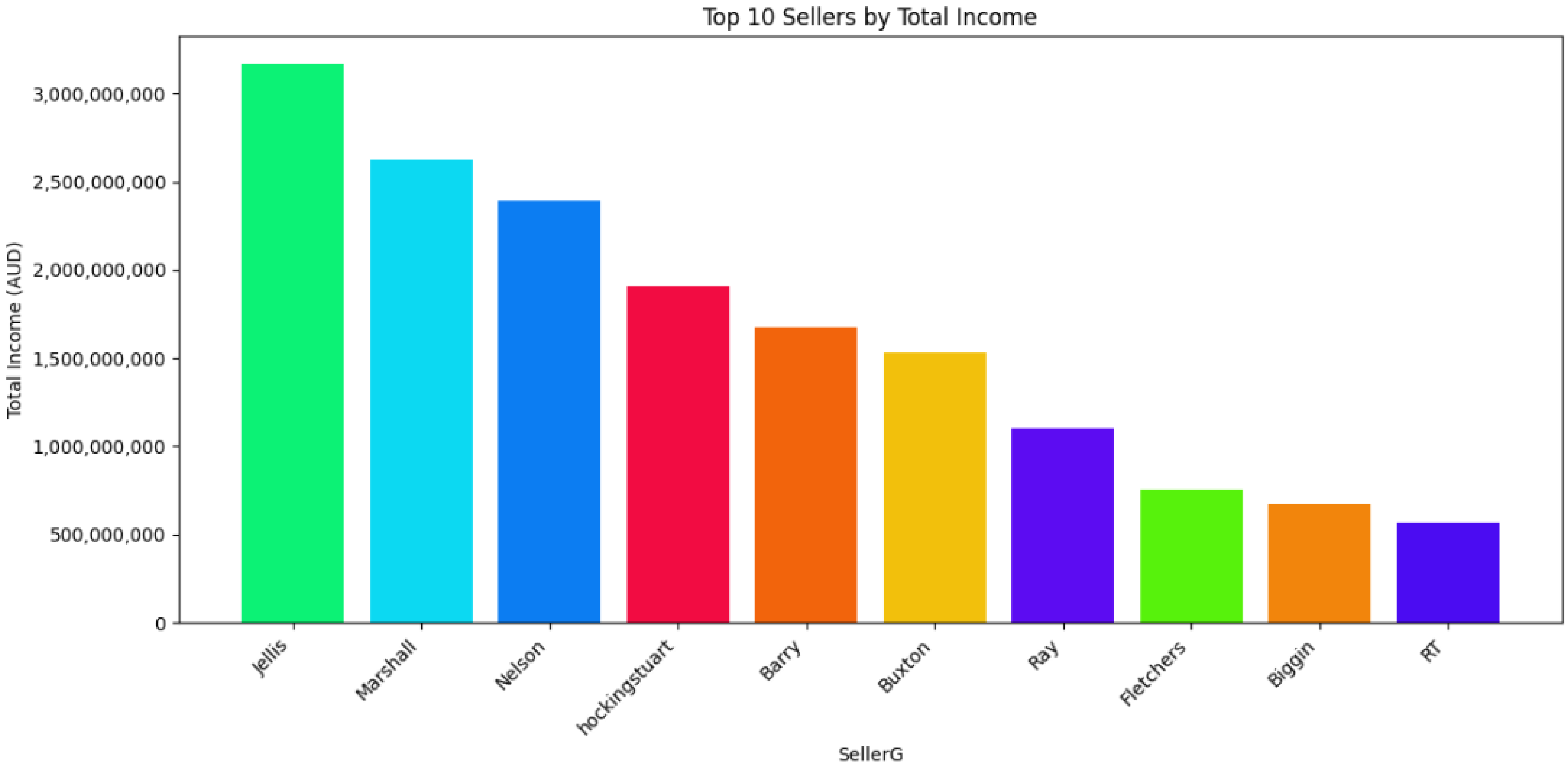


Finding 8: Most Popular Sellers described by the Number of Properties Sold



Finding 8: Most Popular Sellers described by the Number of Properties Sold

Finding 9: The most successful sellers



Limitations of Findings

- Dataset is restricted to Melbourne; results may not generalize to other cities or housing markets.
- Potential data gaps: off-market, private, or unreported sales are absent, limiting market completeness.
- Data quality depends on source accuracy; errors or outdated entries affect findings.
- Missing values imputed statistically may not reflect actual property characteristics precisely.
- External market drivers (e.g., economic shifts, policy changes, development projects) are not captured.
- Real estate market dynamics evolve rapidly; past data may lose predictive power over time.

Conclusions

- Property prices in Melbourne are strongly influenced by land size, property age, bedrooms.
- Market behavior varies significantly across suburbs and broader regions, highlighting localized demand and growth areas.
- Identification of premium properties at both suburb and region levels provides *actionable insights for investors and buyers targeting high-value opportunities.*