Exploratory Data Analysis(EDA)

In the exploratory data analysis, I will be looking at the data and try to undersatnd the data. I will looking at the distribution of data across the dataset, followed by visualizing the data to understa relationship between the features and the target variable.

Modules Required

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading the Cleaned Dataset

In [2]:
```
df = pd.read_csv("../data/interim/cleaned_data.csv")
#Test with printing the first five rows of dataset
df.head()
```

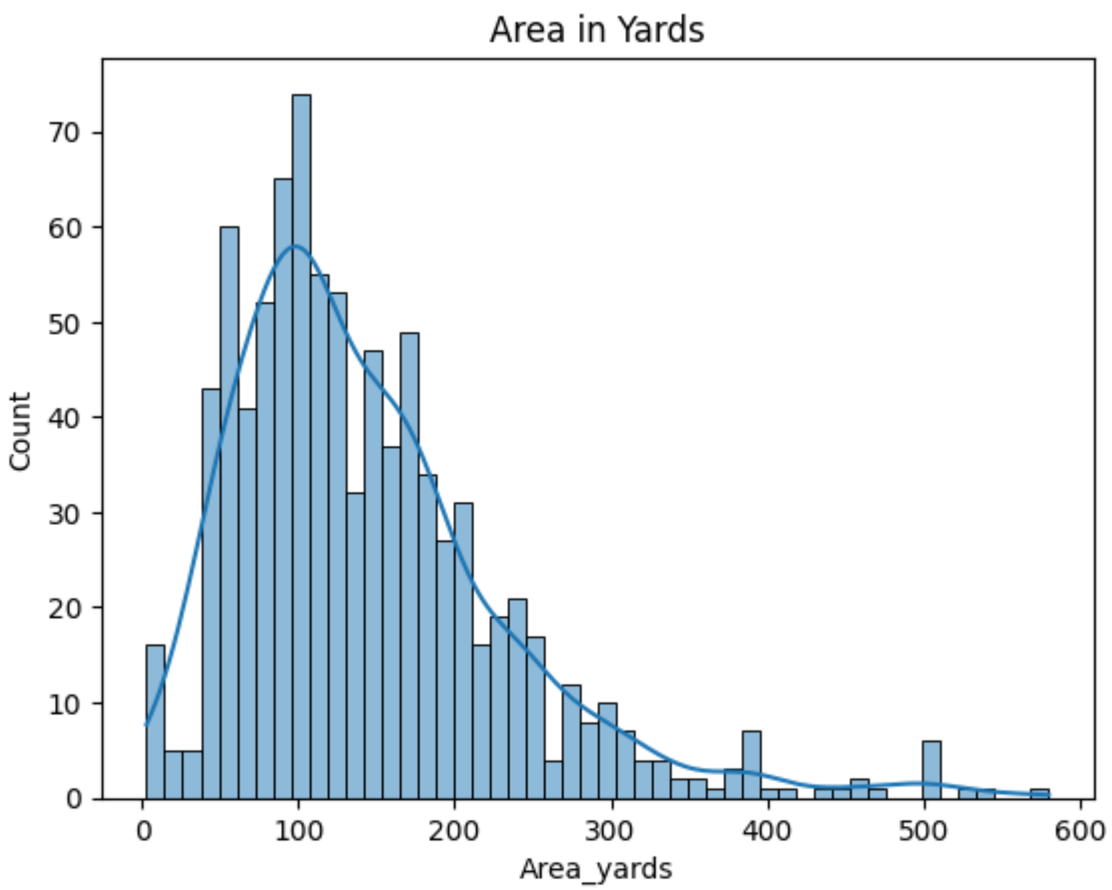
Out[2]:

	Area	BHK	Bathroom	Furnishing	Locality	Parking	Price	Status	T
0	750.0	2	2	Semi-Furnished	Rohini Sector	1	5000000.0	Ready_to_move	Ne
1	950.0	2	2	Furnished	Rohini Sector	1	15500000.0	Ready_to_move	
2	600.0	2	2	Semi-Furnished	Rohini Sector	1	4200000.0	Ready_to_move	
3	650.0	2	2	Semi-Furnished	Rohini Sector	1	6200000.0	Ready_to_move	Ne
4	1300.0	4	3	Semi-Furnished	Rohini Sector	1	15500000.0	Ready_to_move	Ne

Area of houses

In [3]:
```
sns.histplot(x=df['Area_yards'], kde=True, bins=50).set_title('Area in Ya
```

Out[3]:
```
Text(0.5, 1.0, 'Area in Yards')
```



Area (in Square Yards) Analysis

- **Distribution Shape:**
  The `Area_yards` feature displays a **right-skewed (positively skewed)** distribution, indic
  that most houses have smaller area sizes, with a few larger properties extending the tail to t

- **Common Range:**
  The majority of properties are clustered in the **50 to 150 square yards** range, suggesting th
  typical residential size in Delhi.

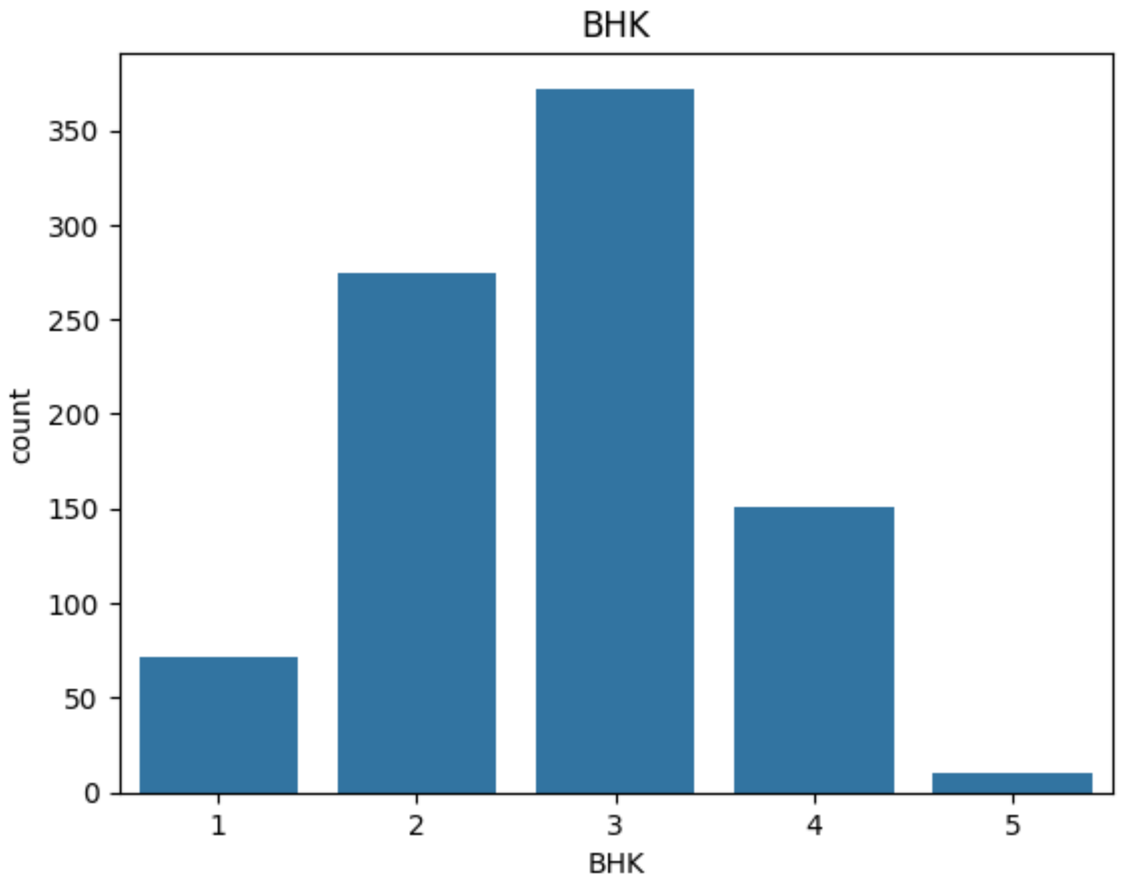
- **Outliers:**
  Properties with area values exceeding **300 yards** are relatively rare and may represent **luxu
  homes or builder floors**. These outliers could affect model performance if not handled appropriately.

- **Insights:**
  The area distribution is **unimodal**, with one clear peak, reinforcing that there's a dominant h
  size trend in the dataset.

BHK

```
In [4]: sns.countplot(x = 'BHK', data = df).set_title('BHK')
```

```
Out[4]: Text(0.5, 1.0, 'BHK')
```



BHK and Area Distribution Relationship

BHK stands for **Bedroom, Hall, and Kitchen** — a standard way of describing house configuratio
India.

From our visualizations:

- The majority of houses are **3 BHK**, indicating it's the most common design among Delhi hon
- This is followed by **2 BHK**, **4 BHK**, **1 BHK**, and a small number of **5 BHK** properties.
- In the area distribution chart, most houses fall between **80–200 sq. yards**.

By combining these insights:

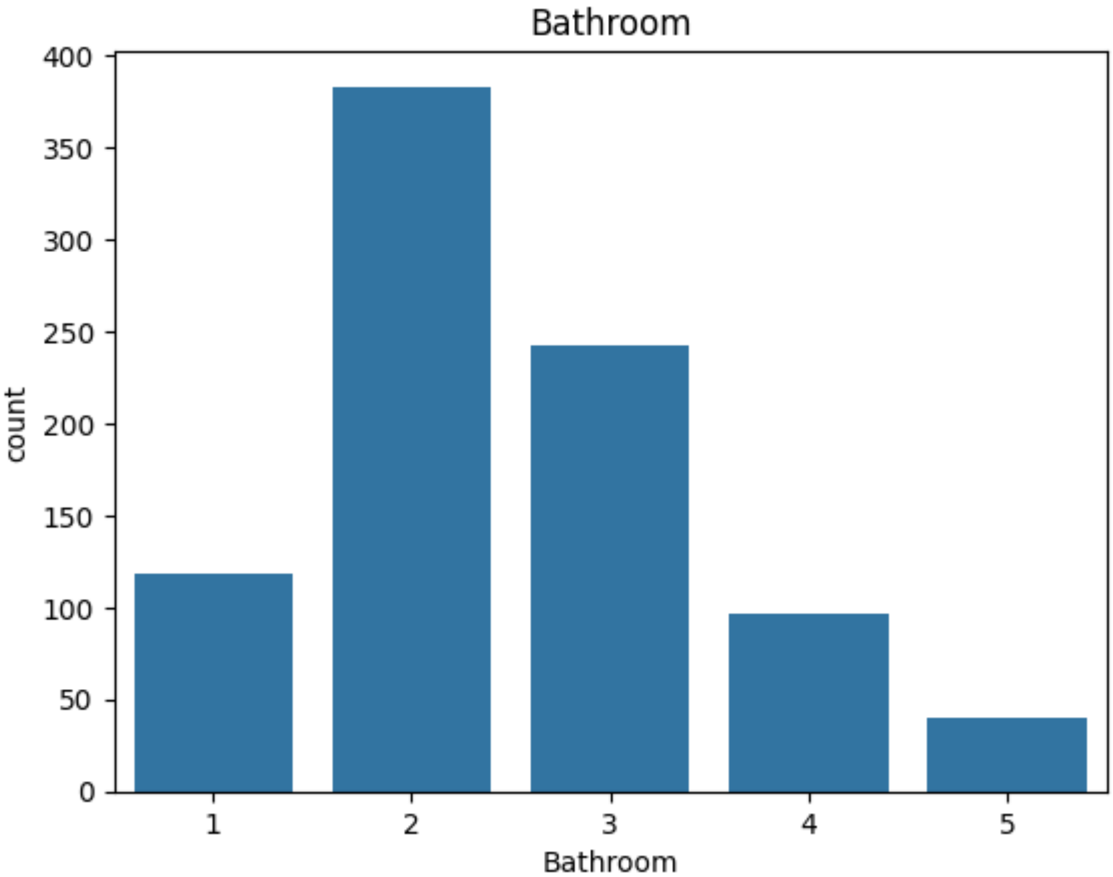
- **1 BHK homes** are typically found in smaller areas, usually **below 80 sq. yards**, suggesting living spaces or studio-type apartments.
- **2 BHK homes** are often associated with areas around **100 sq. yards**, providing a practical b of space and affordability.
- **3 BHK homes**, the most common, tend to cluster around **150–200 sq. yards**, indicating a st family-sized configuration.
- **4 BHK and 5 BHK homes** require larger areas (often **250+ sq. yards**), but such spacious p are relatively rare in the dataset.

This shows a clear structural pattern: **as BHK count increases, the typical house area also increases**. This correlation helps us understand how space is utilized and allocated in Delhi's ho market.

Bathroom Count

In [5]:
```
sns.countplot(x = 'Bathroom', data = df).set_title('Bathroom')
```

Out[5]:
```
Text(0.5, 1.0, 'Bathroom')
```



Brief Analysis of Bathroom Count Data

Overview

- The bar chart illustrates the distribution of the number of bathrooms in a dataset.
- The x-axis represents the number of bathrooms (from 1 to 5), while the y-axis indicates the c occurrences.

Key Observations

Peak Count at 2 Bathrooms :

- The highest frequency is for properties with 2 bathrooms, with a count of approximately 350.

Moderate Count for 3 Bathrooms :

- There is a noticeable drop in frequency for 3 bathrooms, which has a count of roughly 150.

Lower Counts for 1, 4, and 5 Bathrooms :

- Properties with 1 bathroom show a count of about 100.
- Counts for 4 bathrooms and 5 bathrooms are significantly lower, around 50 and 25 respectiv

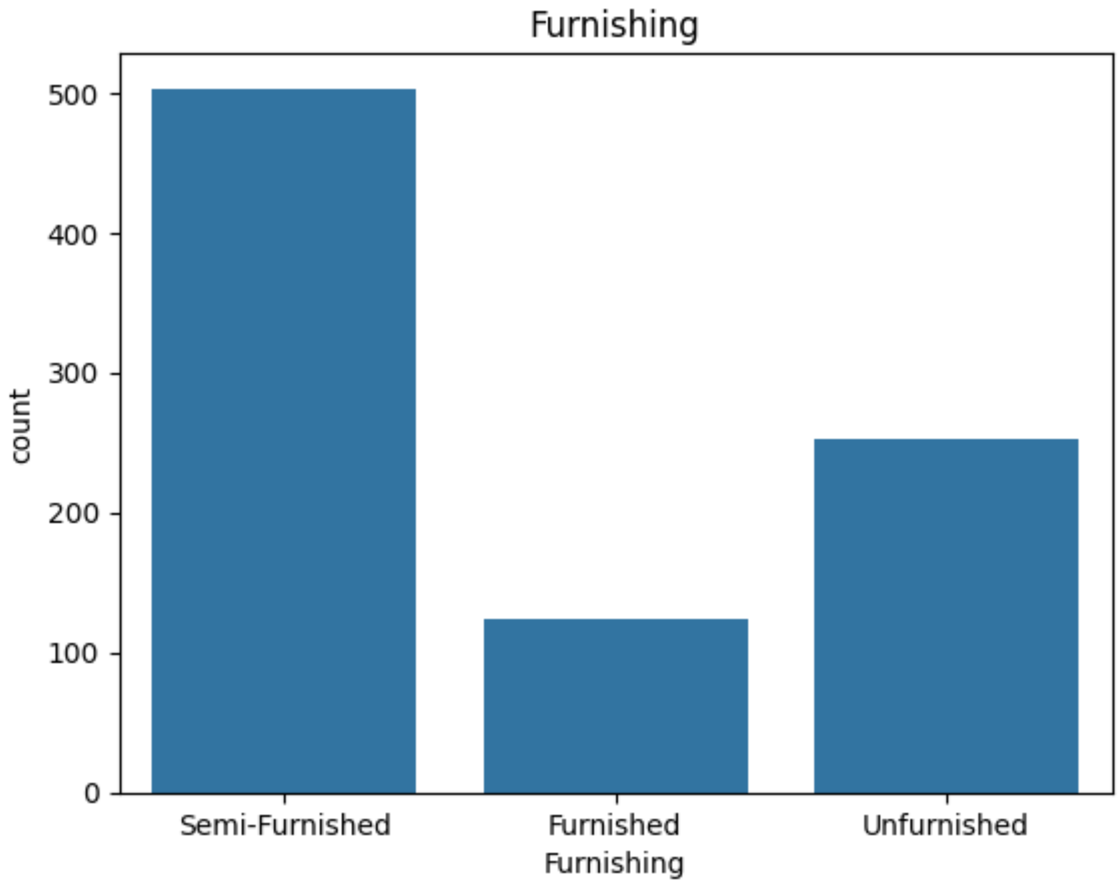
Conclusion

- The data indicates a preference or higher prevalence of properties with 2 bathrooms, sugges configuration may be more common or desirable in the dataset analyzed.
- Properties with fewer or more bathrooms appear to be less frequent. This trend might inform potential market characteristics or buyer preferences.

Furnishing

In [6]:
```
sns.countplot(x='Furnishing',data=df).set_title('Furnishing')
```

Out[6]:
```
Text(0.5, 1.0, 'Furnishing')
```



Overview of the Chart

- The chart categorizes the furnishing types into three groups: **Semi-Furnished**, **Furnished**, a **Unfurnished**.
- The y-axis represents the **count** of occurrences for each category, while the x-axis labels th furnishing types.

Data Insights

- **Semi-Furnished**: Has the highest count, close to **500**. This indicates a strong preference or availability for semi-furnished options among the surveyed population.
- **Furnished**: Shows a significantly lower count, approximately **100**. This category is less popu available compared to semi-furnished items.
- **Unfurnished**: Also has a count around **300**, placing it between the semi-furnished and furnis categories. This suggests a reasonable demand for un-furnished spaces, but not as high as semi-furnished ones.

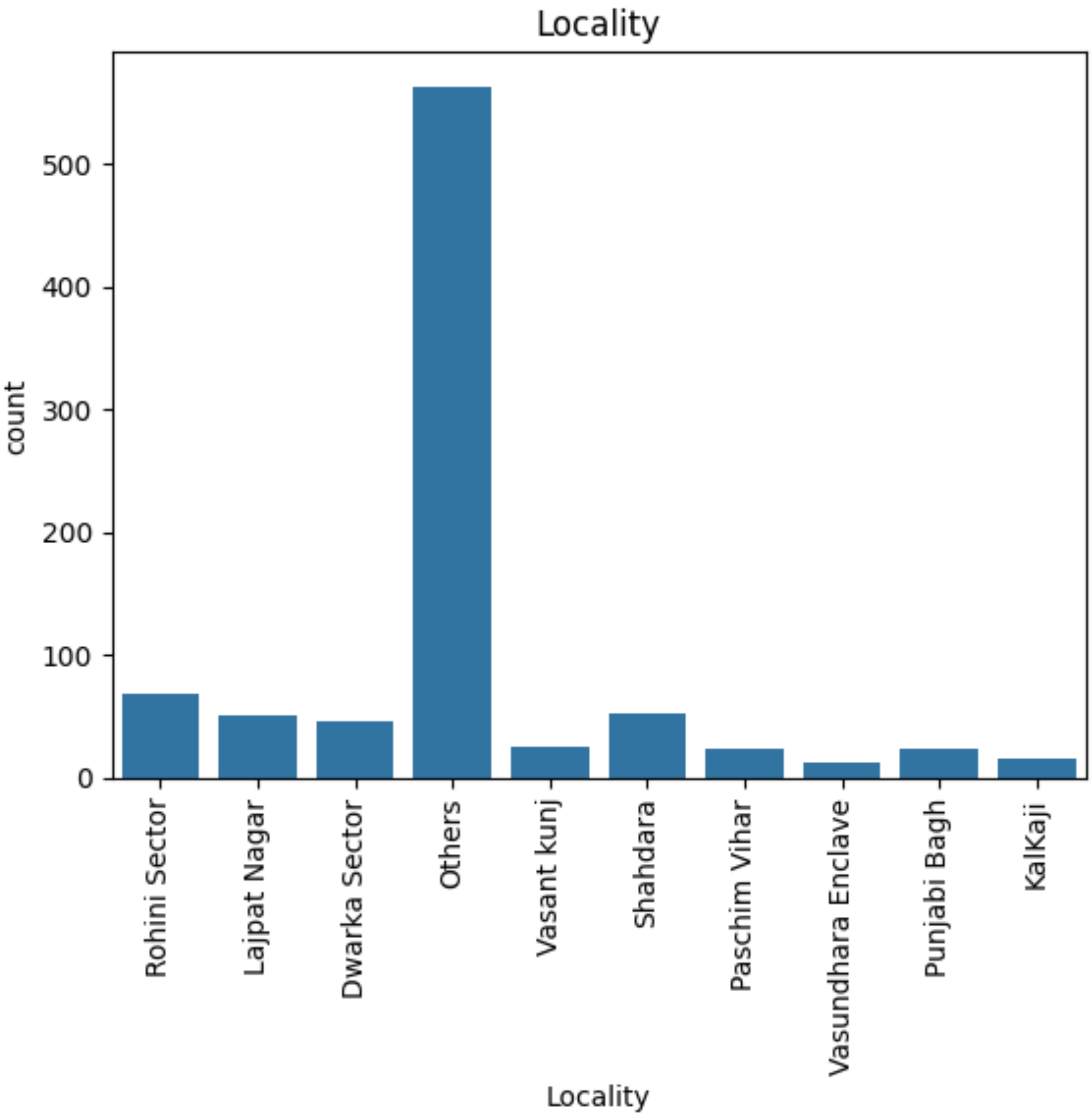
Conclusion

- The bar chart clearly indicates that **semi-furnished** is the most preferred option, while **furni** items are the least sought after. The demand for **unfurnished** options is moderate, indicatin diverse market for different furnishing preferences.

Locality

```
In [7]: sns.countplot(x = 'Locality', data = df).set_title('Locality')
        plt.xticks(rotation = 90)
```

```
Out[7]: ([0, 1, 2, 3, 4, 5, 6, 7, 8, 9],
         [Text(0, 0, 'Rohini Sector'),
          Text(1, 0, 'Lajpat Nagar'),
          Text(2, 0, 'Dwarka Sector'),
          Text(3, 0, 'Others'),
          Text(4, 0, 'Vasant kunj'),
          Text(5, 0, 'Shahdara'),
          Text(6, 0, 'Paschim Vihar'),
          Text(7, 0, 'Vasundhara Enclave'),
          Text(8, 0, 'Punjabi Bagh'),
          Text(9, 0, 'KalKaji')])
```



Overall Trend:

- The chart represents the distribution of counts across various localities.

Key Observations:

- **Dominance of "Others":**
  - The category labeled "Others" has a significantly higher count (over 500) compared to c
    localities. This suggests that many addresses or entries fall outside the specific localitie
- **Lower Counts for Specific Localities:**
  - The counts for localities such as Rohini Sector, Lajpat Nagar, Dwarka Sector, Vasant Ku
    Shahdara, Paschim Vihar, Vasundhara Enclave, Punjabi Bagh, and Kalkaji are all relativ
    generally under 100.
- **Rohini Sector and Lajpat Nagar:**
  - Among the specific localities, Rohini Sector and Lajpat Nagar show modest counts, but
    pale in comparison to the "Others" category.

Implications:

- The data indicates a concentrated population or activity in the "Others" category, which may encompass numerous smaller or less defined areas.
- This could suggest limited visibility or reporting for the localities listed, implying potential gap collection or representation.

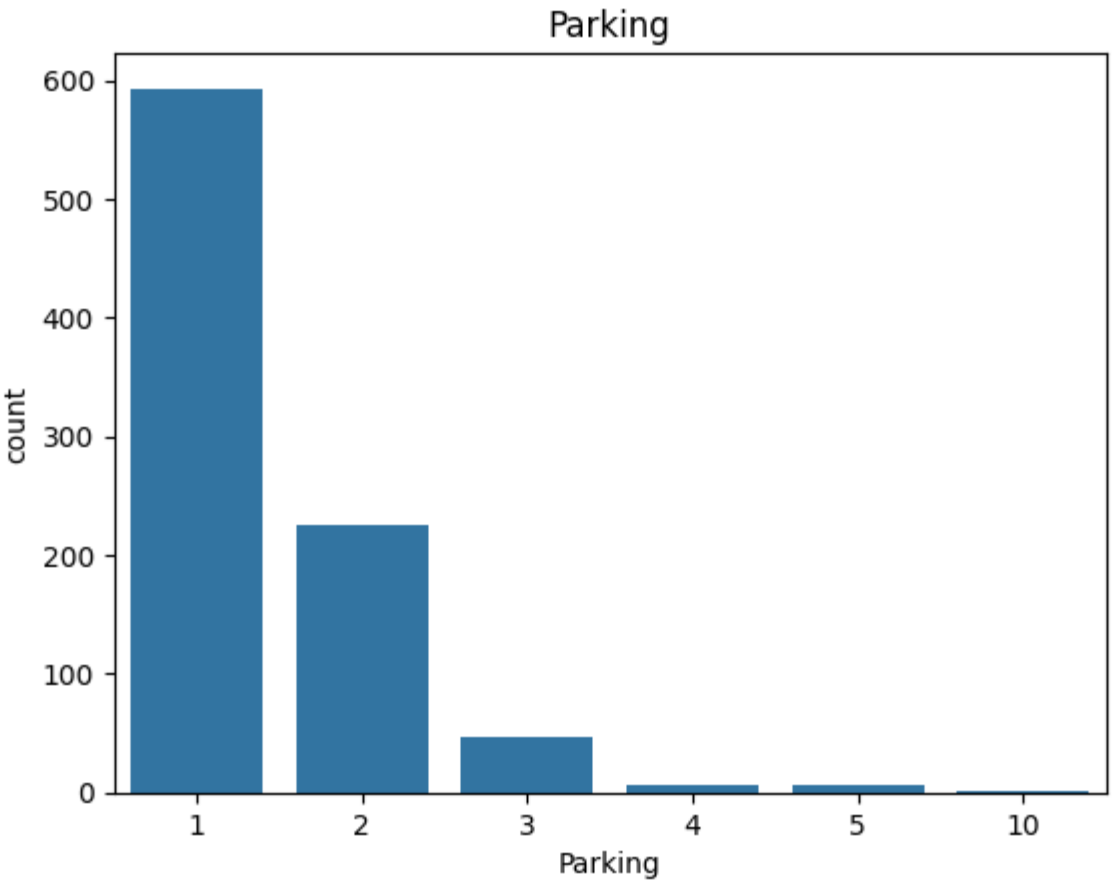
Conclusion:

- The bar chart effectively highlights the disparity in counts among the localities, showcasing t overwhelming presence of a broader category versus specific identified areas. Addressing th imbalance could provide more insight into the local dynamics.

Parking

In [8]:
```
sns.countplot(x = 'Parking', data = df).set_title('Parking')
```

Out[8]:
```
Text(0.5, 1.0, 'Parking')
```


Parking
600
500
400
300
200
100
0
count
1
2
3
4
5
10
Parking

Analysis of Parking Spaces in Delhi Houses

- **Common Trends:**
  - **One Car Parking:**
    - The majority of houses (over 500) have **one car parking space**.
  - **Two Car Parking:**
    - A significant number of houses (around 200) offer **two car parking spaces**.
  - **Three or More Car Parking:**
    - Very few houses have **three or more** parking spaces, with counts dropping signific spaces greater than three.
- **Relation to House Area:**
  - **House Size:**
    - Most houses fall within the **100 - 200 sq. yards** area range, which likely correlates predominance of one car parking space.
  - **Larger Houses:**
    - Houses that exceed **200 sq. yards** are more likely to provide **two or more car par spaces**.

Graph Interpretation:

- **X-Axis (Parking Spaces):**
  - Represents the number of car parking spaces available in the houses.
- **Y-Axis (Count):**
  - Displays the number of houses corresponding to the number of parking spaces.

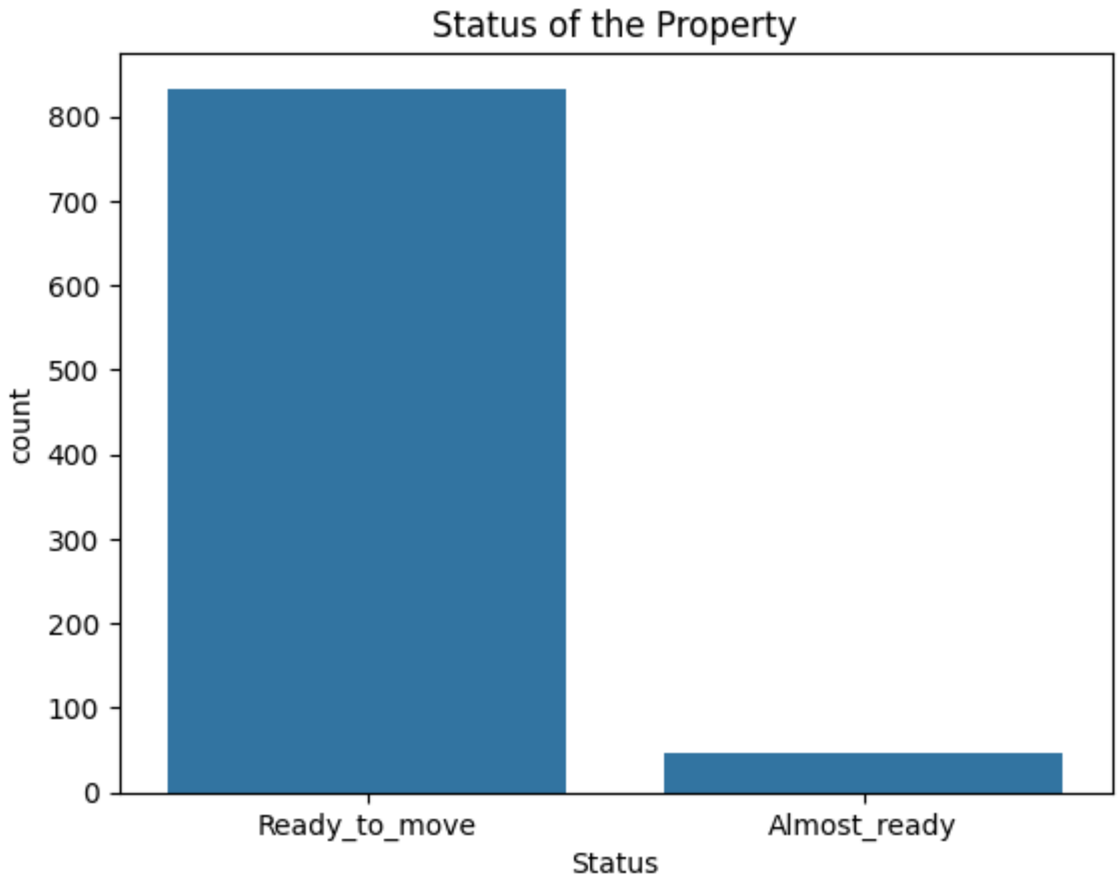
Conclusion:

The data illustrates a clear relationship between the number of car parking spaces and the area houses in Delhi. The trend of having predominantly one car parking space aligns with the typical size of 100 - 200 sq. yards, while larger houses accommodate additional parking.

Status

In [9]:
```
sns.countplot(x = 'Status', data= df).set_title('Status of the Property')
```

Out[9]:
```
Text(0.5, 1.0, 'Status of the Property')
```


Status of the Property
count
800
700
600
500
400
300
200
100
0
Ready_to_move
Almost_ready
Status

Analysis of Property Status Bar Chart

Graph Overview

The chart presents a comparison of the number of properties based on their readiness status, categorized as:

- **Ready_to_move**
- **Almost_ready**

Key Observations

1. **Dominance of Ready-to-Move Properties**
   - The majority of the properties fall under the **"Ready_to_move"** category.
   - Count exceeds **800**, indicating a high availability of properties that are fully constructed and immediately available for possession.
2. **Scarcity of Almost-Ready Properties**
   - The **"Almost_ready"** category has a significantly lower count, approximately around **50**.
   - This suggests a much smaller segment of the market is in the near-completion phase.
3. **Market Implications**
   - **Buyer Preference:** Buyers may prefer ready-to-move-in options due to immediate usability avoidance of construction delays.
   - **Developer Strategy:** Developers may be focusing more on completing and offering ready-tc projects, possibly due to demand trends.
   - **Reduced Speculative Buying:** The low number of almost-ready properties could indicate le speculative investment in under-construction homes.
4. **Potential Buyer Advantage**
   - High availability of ready-to-move options could give buyers more choices and potentially mc negotiating power.

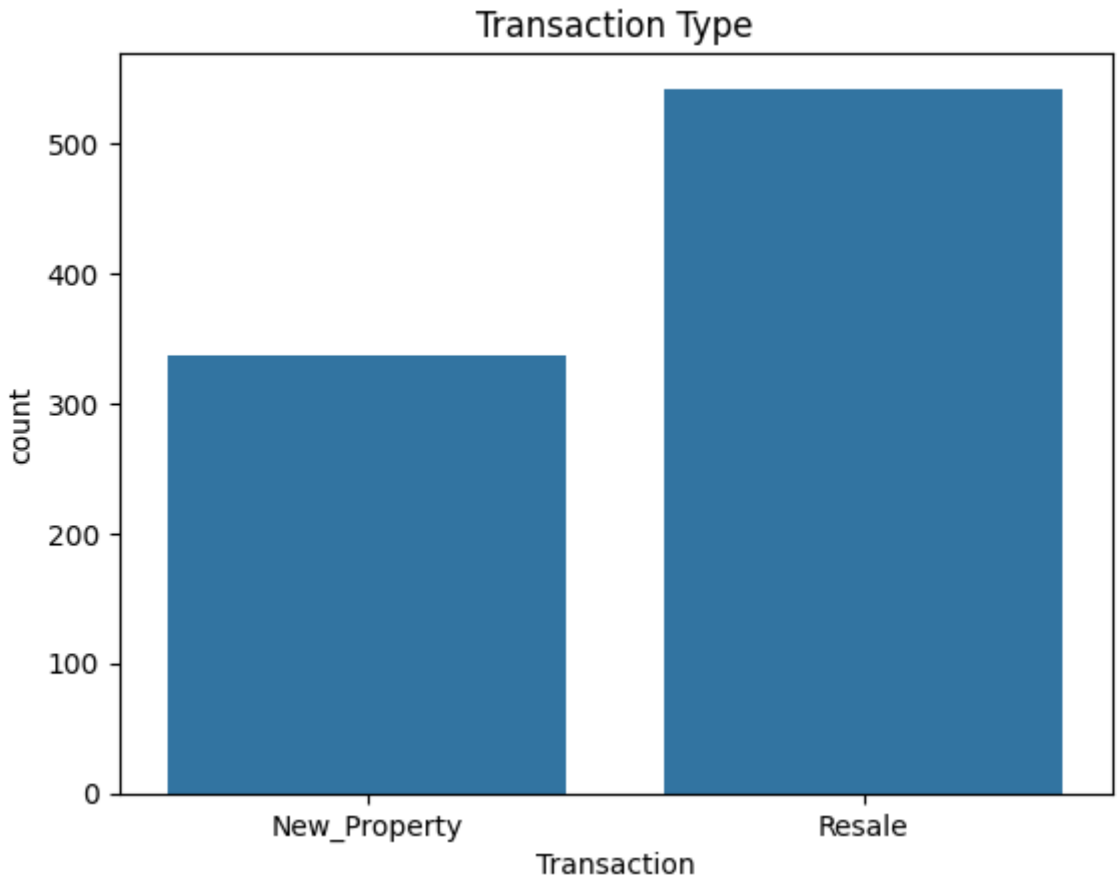
Conclusion

The graph reveals a clear skew in the real estate inventory toward **ready-to-move** properties, su that the current market is geared toward immediate occupancy rather than near-future possessio could be reflective of buyer confidence, urgency, or a strategic shift by developers in response to demand.

Transaction Type

In [10]:
```
sns.countplot(x='Transaction', data=df).set_title('Transaction Type')
```

Out[10]:
```
Text(0.5, 1.0, 'Transaction Type')
```


Transaction Type
count
0
100
200
300
400
500
New_Property
Resale
Transaction

Analysis of Transaction Type Bar Chart

Graph Overview

The chart displays the distribution of properties based on the **type of transaction**, categorized a

- **New_Property**
- **Resale**

Key Observations

1. **Higher Number of Resale Transactions**

- **Resale** properties dominate the market with a count of over **500**.
- This suggests a strong trend in secondary sales or previously owned property exchanges.

2. **Fewer New Property Transactions**

- **New_Property** transactions are fewer, with a count of around **340**.
- This indicates that newly launched or never-owned properties make up a smaller share of th market.

3. **Market Insights**

- **Buyer Behavior:** Buyers may be gravitating toward resale properties due to factors such as
  - Lower cost compared to new launches.
  - Established locations and amenities.
  - Immediate possession availability.
- **Supply Factors:** The resale market might be more active due to owners relocating, upgradi monetizing their investments.

4. **Implication for Developers**

- Real estate developers might face stiffer competition from the resale segment.
- To attract more buyers, they may need to offer added value like flexible payment plans, read move inventory, or location advantages.

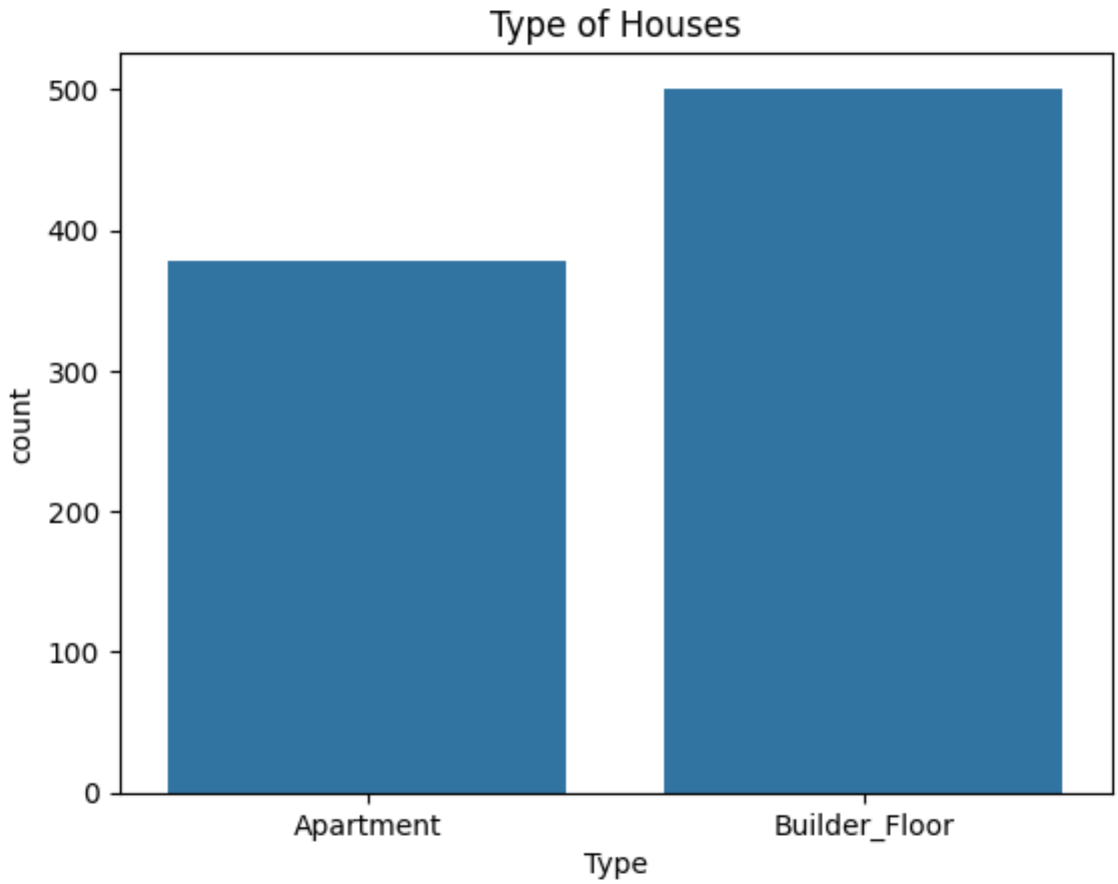
Conclusion

The graph indicates a notable preference for **resale** transactions over **new properties**. This tren reflects both market maturity and buyer preferences for affordability and immediacy. Developers need to adapt strategies to stay competitive against the growing resale market.

House Type

In [11]:
```
sns.countplot(x='Type', data=df).set_title('Type of Houses')
```

Out[11]:
```
Text(0.5, 1.0, 'Type of Houses')
```



Type of Houses
count
500
400
300
200
100
0
Apartment
Builder_Floor
Type

Analysis of Type of Houses Bar Chart

Graph Overview

The chart illustrates the distribution of properties based on the **type of house**, categorized as:

- **Apartment**
- **Builder_Floor**

Key Observations

1. **Builder_Floor is the Most Common House Type**

   - **Builder_Floor** properties have the highest count, approximately **500**.
   - This indicates a strong presence or preference for low-rise, independent-floor living spaces.

2. **Apartments Have a Slightly Lower Count**

   - **Apartment** properties account for around **370–380** listings.
   - While still significant, they fall behind builder floors in terms of quantity.

3. **Market Insights**

   - **Preference Shift:** The higher number of builder floors may indicate a preference for:
     - More privacy and independence.
     - Lower density living compared to high-rise apartments.
   - **Urban Spread:** Builder floors are often common in suburban or plotted developments, reflec
     expansion beyond high-rise zones.
   - **Investment Appeal:** Builder floors may appeal to investors due to land ownership share anc
     customization potential.

4. **Implication for Real Estate Strategy**

   - Developers and agents may focus more on builder floor projects in regions where there's de
     semi-independent housing.
   - Apartment developers may need to emphasize amenities, security, and community living to s
     competitive.

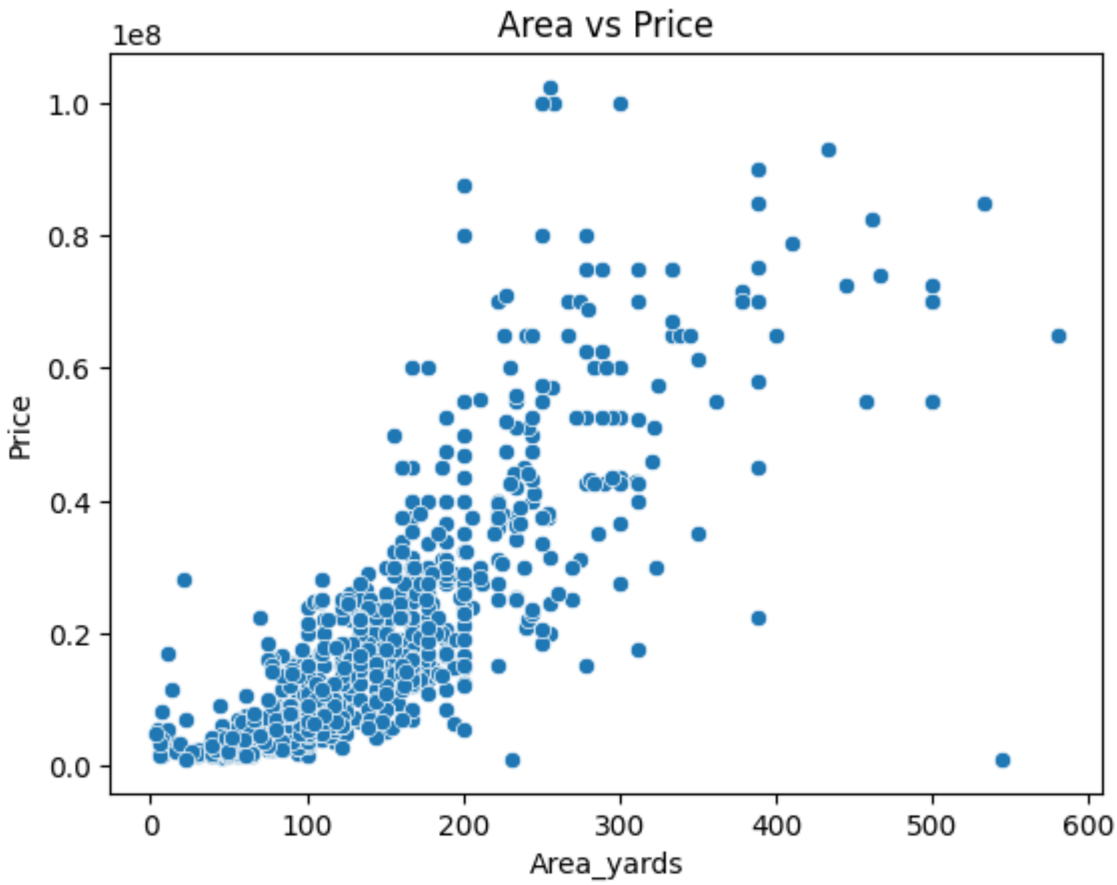
Conclusion

The bar chart suggests a market trend that leans slightly more toward **builder floors** than **apartı**
This could reflect buyer preferences for more space, privacy, or ownership advantages, especiall
expanding urban areas. Both categories, however, maintain strong representation in the market.

Area and Price

In [12]:
```
sns.scatterplot(x= 'Area_yards', y='Price', data=df).set_title('Area vs P
```

Out[12]:
```
Text(0.5, 1.0, 'Area vs Price')
```


Area vs Price
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Price
0
100
200
300
400
500
600
Area_yards

Analysis of Area vs Price Scatter Plot

Graph Overview

The scatter plot visualizes the relationship between:

- **Area (in square yards)** on the X-axis
- **Price (in rupees)** on the Y-axis

Each point represents a property.

Key Observations

1. **Positive Correlation**

- There is a clear **positive correlation** between area and price.
- As the **area of a property increases**, the **price generally increases** as well.

2. **Wider Price Range at Larger Areas**

- Properties with **larger area sizes (200–500+ yards)** show a **wider spread in pricing**, sugg
  - Varied property features
  - Differences in location, construction quality, or amenities

3. **Price Outliers**

- Some properties, especially around 250–300 yards, have **extremely high prices**, acting as
- These may represent luxury listings or prime-location properties.

4. **Concentration of Listings**

- Most properties are clustered in the **50–250 yard** range, with prices between **0 to 4 crore IN (approx.)**
- This suggests that mid-sized plots dominate the market.

Market Implications

- **Investor Insight:** Investing in mid-sized properties (100–250 yards) may offer stable pricing
- **Developer Insight:** The wide spread at larger areas suggests potential for premium develop or varied pricing strategies.
- **Buyer Insight:** Buyers may find consistent price-to-area value in smaller-to-mid-sized prope but need to scrutinize high-area listings for fair valuation.

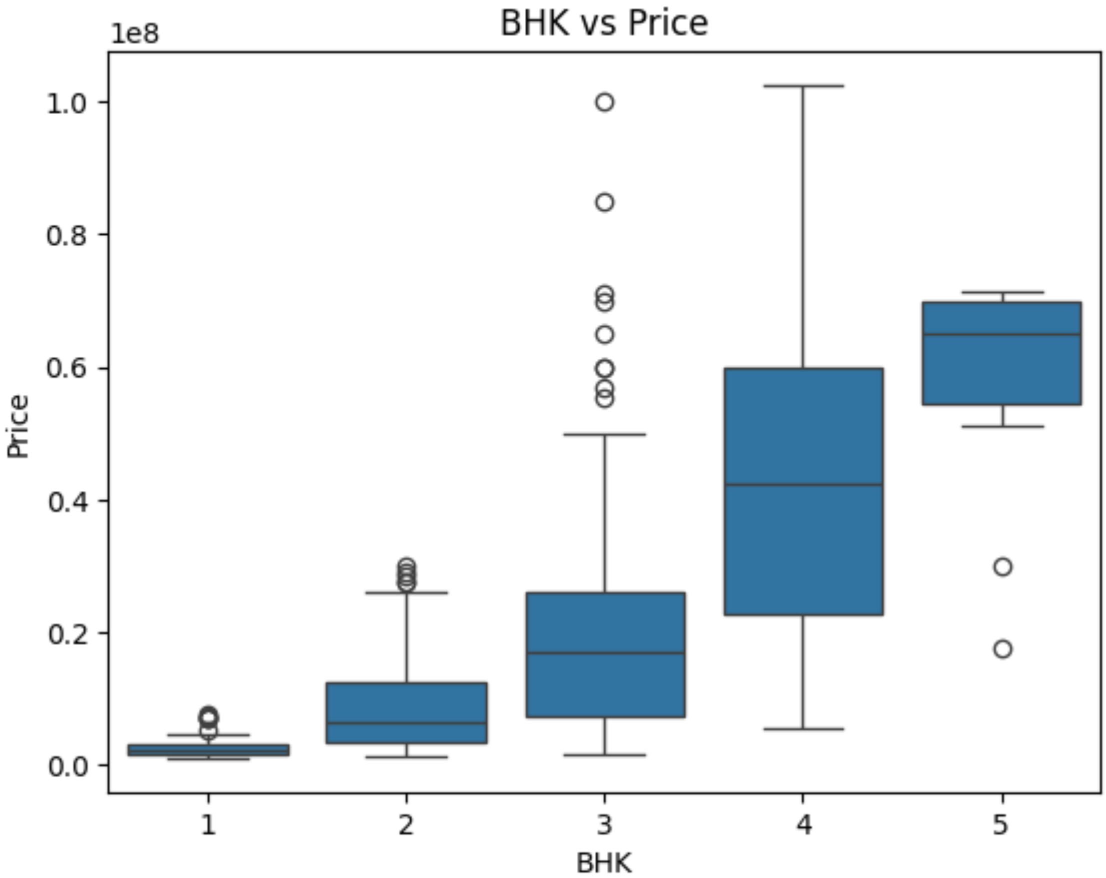
Conclusion

The scatter plot shows a **strong area-price relationship** with expected positive correlation. How pricing becomes more **variable at higher areas**, likely influenced by additional factors like locati amenities, and construction quality. Understanding these dynamics is essential for pricing, invest and purchasing decisions.

BHK and Price

In [13]:
```
sns.boxplot(x = 'BHK', y= 'Price', data = df).set_title('BHK vs Price')
```

Out[13]:
```
Text(0.5, 1.0, 'BHK vs Price')
```



Analysis of BHK vs Price Box Plot

Graph Overview

The box plot compares the **price distribution** of properties based on the number of **BHKs (Bed Hall, Kitchen)**:

- X-axis: Number of BHKs (1 to 5)
- Y-axis: Price (in rupees)

Key Observations

1. **General Trend: Higher BHK → Higher Price**

- Median price increases as the number of BHKs increases.
- This reflects the intuitive trend that larger homes (with more bedrooms) cost more.

2. **Price Range and Spread**

- **1 BHK:** Lowest price range, tightly grouped with minimal variation.
- **2 BHK:** Moderate increase in both price and spread.
- **3 BHK:** Larger spread; prices vary widely, with multiple high outliers.
- **4 BHK:** Very high variability in price, including the **widest interquartile range** and some **ext outliers**.
- **5 BHK:** High median price and tight IQR, suggesting luxury positioning with consistent prem pricing.

3. **Outliers**

- Significant price outliers are seen in the **3 BHK and 4 BHK** categories.
- These likely represent **luxury homes** or **premium locations** driving prices above the typica

---

Market Insights

- **Buyer Perspective:** Price rises sharply beyond 2 BHKs. Buyers should balance space neec affordability.
- **Developer Strategy:** 3 BHK and 4 BHK segments show strong demand across a wide price —ideal for both mid-range and luxury targeting.
- **Investor Insight:** Wide price spread in higher BHKs suggests investment opportunities in bo affordable and premium segments.

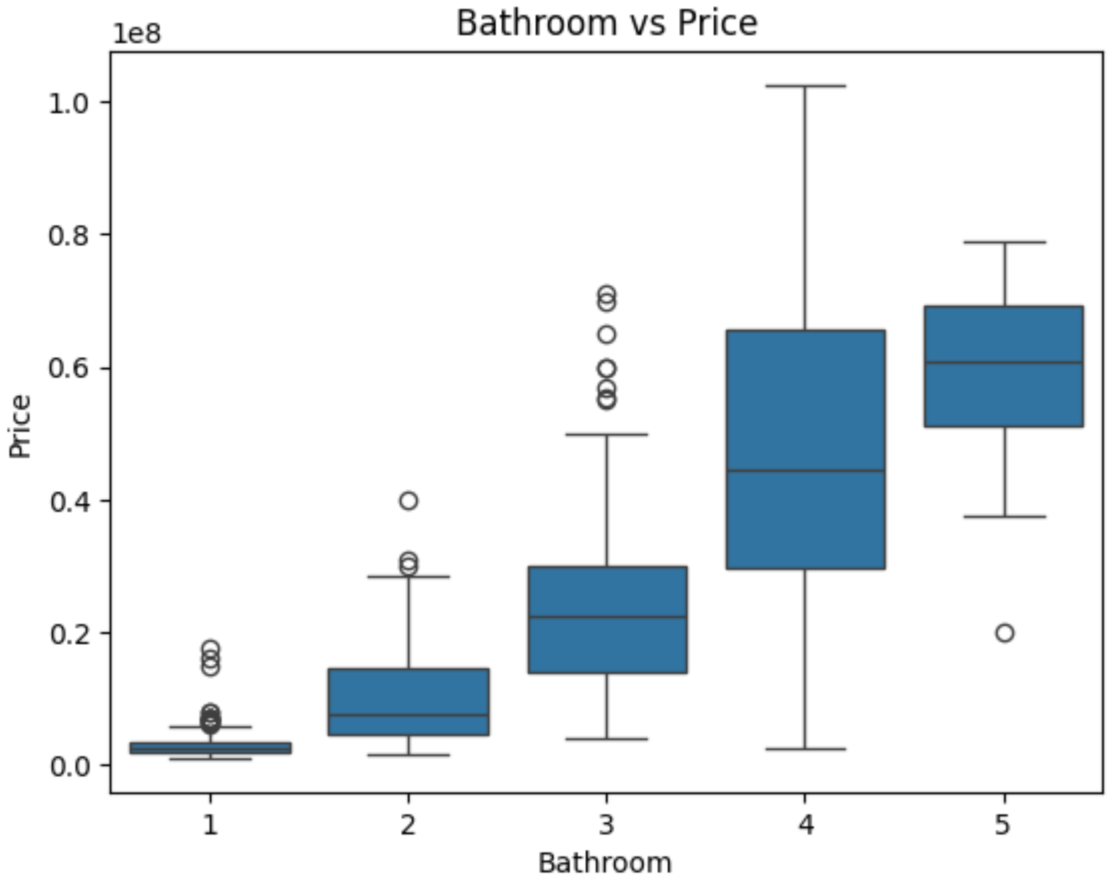
Conclusion

The plot shows a **positive relationship between BHK count and price**, with increasing price va as BHKs increase. While 1 and 2 BHK properties are relatively affordable and stable in pricing, 3 BHK units show **significant variance and high-end market potential**, especially in the 4 and 5 categories.

Bathroom count and Price

In [14]:
```
sns.boxplot(x= 'Bathroom', y = 'Price', data = df).set_title('Bathroom vs
```

Out[14]:
```
Text(0.5, 1.0, 'Bathroom vs Price')
```



Analysis of Bathroom vs Price Box Plot

Graph Overview

This box plot illustrates the distribution of **property prices** based on the **number of bathrooms**:

- **X-axis:** Number of Bathrooms (1 to 5)
- **Y-axis:** Price (in rupees)

Key Observations

1. **Positive Relationship**

- As the **number of bathrooms increases**, the **median property price** also increases.
- This aligns with market expectations, as more bathrooms generally indicate a larger or more luxurious property.

2. **Price Distribution**

- **1 Bathroom:** Lowest price range with minimal variation; tightly clustered around the lower e
- **2 Bathrooms:** Slightly higher prices and increased variability.
- **3 Bathrooms:** Noticeable jump in price; wider interquartile range and multiple outliers.
- **4 Bathrooms:** Broad price range with high median and many premium-priced properties.
- **5 Bathrooms:** High and relatively consistent prices, indicating luxury segment properties.

3. **Outliers**

- Higher bathroom counts (especially 3, 4, and 5) include several **outliers** with **very high pric** pointing to **high-end or exclusive listings**.

Market Insights

- **Buyer Perspective:** Properties with 1–2 bathrooms are generally more affordable and fall w tighter price band.
- **Luxury Indicator:** Properties with 4–5 bathrooms are likely to be luxury homes, showing hig median prices and greater consistency at the top end.
- **Investment Insight:** Bathroom count can be a strong indicator of property tier—useful for segmentation and targeting in marketing or investment strategies.

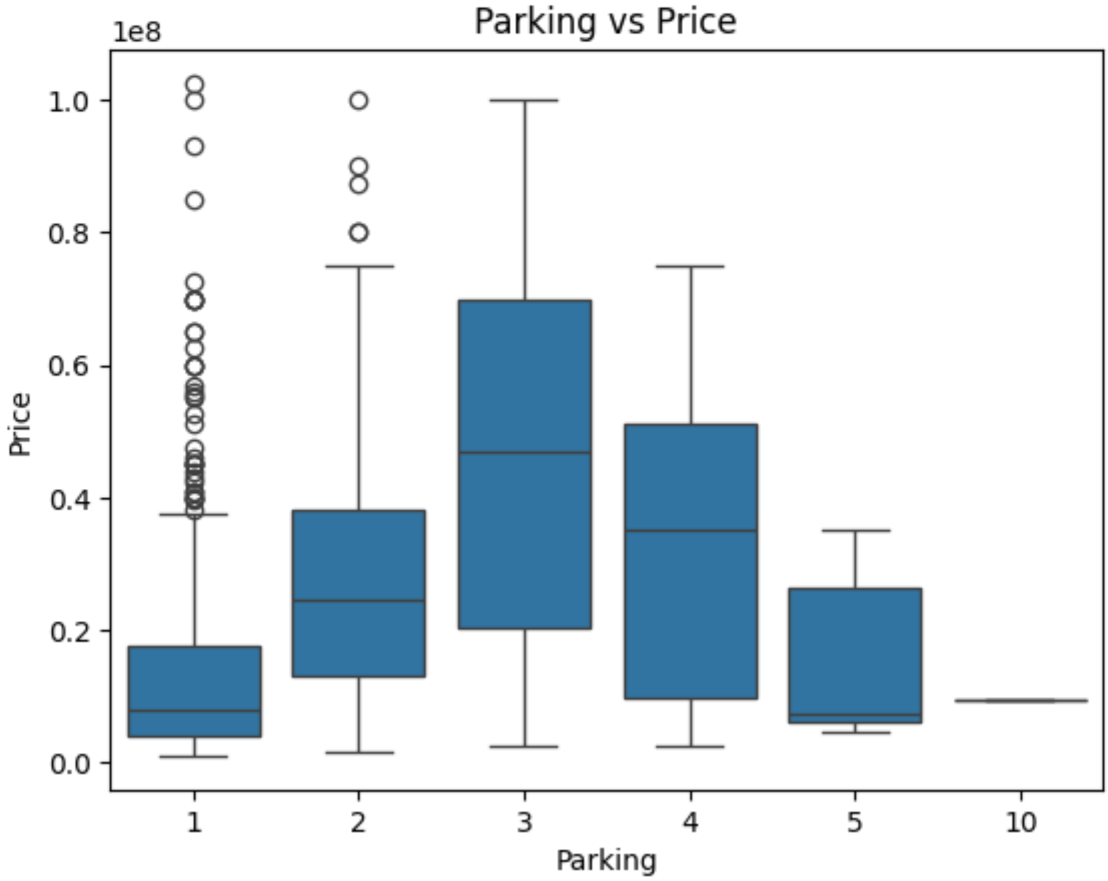
Conclusion

The plot reveals a **clear upward trend** in property price with increasing bathroom count. While lo bathroom counts are common in more affordable homes, higher bathroom numbers correlate wit **property value**, making bathroom count a useful proxy for assessing luxury or premium status ir housing market.

Parking and Price

```
In [15]: sns.boxplot(x = 'Parking', y = 'Price', data = df).set_title('Parking vs I
```

```
Out[15]: Text(0.5, 1.0, 'Parking vs Price')
```



Analysis of Parking vs Price Box Plot

Graph Overview

This box plot shows how **property prices** vary based on the **number of parking spaces**:

- **X-axis:** Number of Parking Spaces (1, 2, 3, 4, 5, 10)
- **Y-axis:** Price (in rupees)

Key Observations

1. **General Trend**

- Property prices tend to **increase with more parking spaces** up to a point, especially betwe 3.
- The median price peaks around **3 parking spaces**, indicating a potential sweet spot for prer properties.

2. **Variability**

- **1 Parking Space:** Most frequent category with many outliers. Large spread in prices, but ge lower median compared to higher parking counts.
- **2–3 Parking Spaces:** Median price rises with greater spread. Properties in this range are lik to-high-end homes.
- **4 Parking Spaces:** Slight dip in median, but still shows wide distribution, indicating variabilit category.
- **5 Parking Spaces:** Lower median and less spread compared to 3 and 4. Possibly fewer lux listings or more standardized pricing.
- **10 Parking Spaces:** Flat distribution at a specific price level; this could indicate a **rare case** outlier dataset entry.

3. **Outliers**

- Particularly noticeable in the **1 and 2 parking space** categories.
- These could represent smaller properties priced higher due to other amenities or location advantages.

Market Insights

- **Luxury Indicator:** 3 parking spaces appear as a strong indicator of luxury or premium pricir
- **Saturation Point:** More than 3–4 parking spots may not significantly raise prices and might diminishing returns.
- **Investor Focus:** Optimal pricing visibility in 2–3 parking range—valuable for segmenting pre buyers.

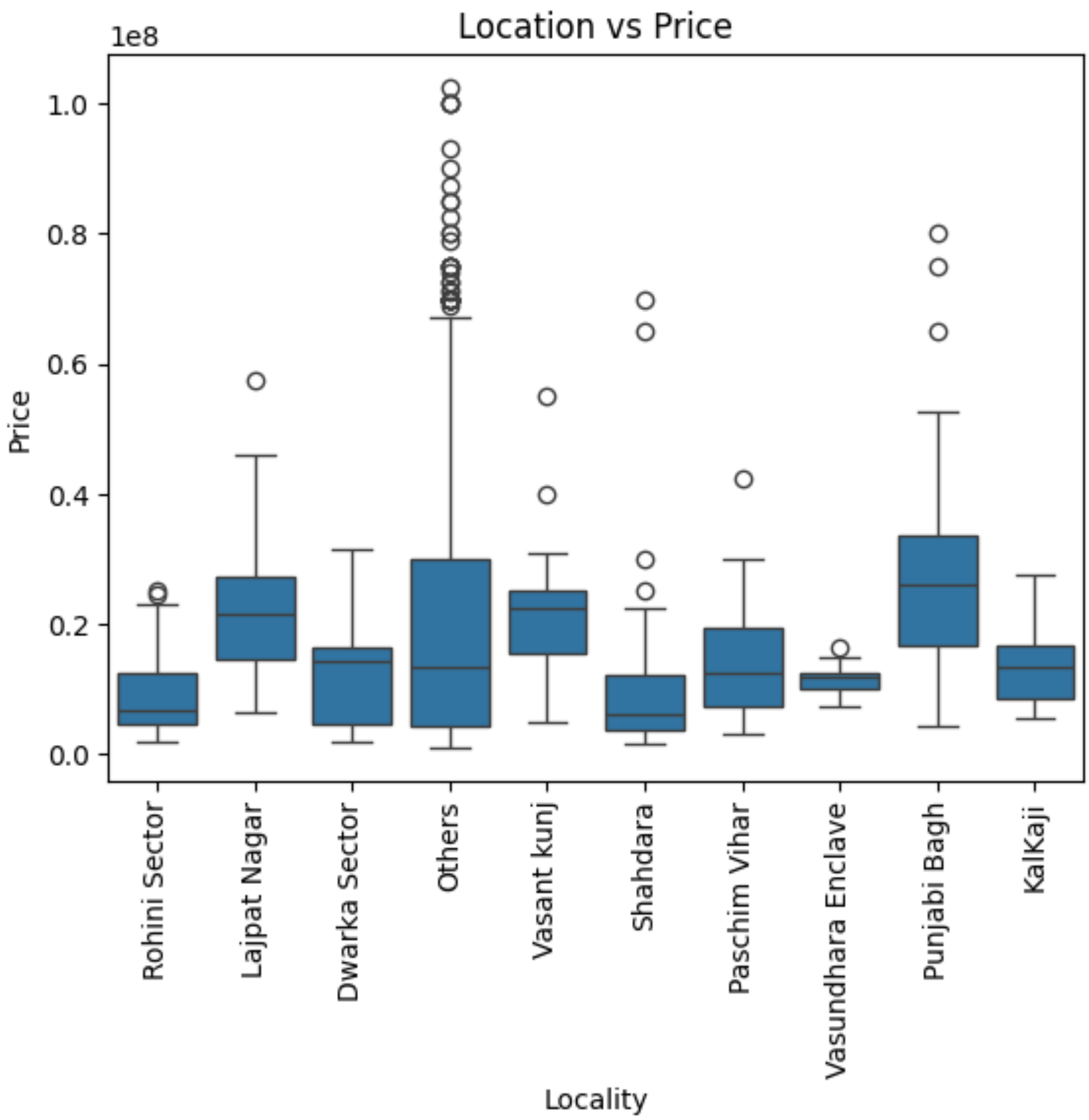
Conclusion

The box plot suggests a **positive correlation** between the number of parking spaces and proper
especially up to 3 spots. Beyond that, the trend becomes less predictable, potentially due to low

Locality and Price

In [16]:
```
sns.boxplot(x = 'Locality', y = 'Price', data = df).set_title('Location vs
plt.xticks(rotation = 90)
```

Out[16]:
```
([0, 1, 2, 3, 4, 5, 6, 7, 8, 9],
 [Text(0, 0, 'Rohini Sector'),
  Text(1, 0, 'Lajpat Nagar'),
  Text(2, 0, 'Dwarka Sector'),
  Text(3, 0, 'Others'),
  Text(4, 0, 'Vasant kunj'),
  Text(5, 0, 'Shahdara'),
  Text(6, 0, 'Paschim Vihar'),
  Text(7, 0, 'Vasundhara Enclave'),
  Text(8, 0, 'Punjabi Bagh'),
  Text(9, 0, 'KalKaji')])
```



Location vs Price Box Plot Analysis

Graph Overview

This box plot displays the distribution of **property prices** across various **localities**:

- **X-axis:** Locality (e.g., Rohini Sector, Lajpat Nagar, Punjabi Bagh, etc.)
- **Y-axis:** Price (in rupees)

Key Observations

1. **Top Expensive Locations**
   - **Punjabi Bagh** and **Others** have the **highest median prices**, with Punjabi Bagh also showir wide price spread.
   - **Lajpat Nagar** and **Vasant Kunj** also have relatively high medians, indicating affluent or high demand neighborhoods.

2. **Price Distribution**
   - **Others** has a **significant number of outliers**, suggesting the inclusion of very high-priced properties that don't belong to predefined localities.
   - **Dwarka Sector** and **Kalkaji** have moderate price ranges and medians, potentially represent tier real estate zones.

3. **More Affordable Locations**
   - **Rohini Sector**, **Shahdara**, and **Vasundhara Enclave** show **lower medians**, indicating thes be more affordable or have smaller properties.
   - **Paschim Vihar** shows moderate prices with a tighter distribution.

4. **Outliers**
   - Outliers are present in nearly all locations, particularly in **Others**, indicating luxury properties significantly exceed the average price range.

Market Insights

- **Luxury Cluster:** Punjabi Bagh, Lajpat Nagar, and Vasant Kunj show pricing characteristics premium markets.
- **Investment Consideration:** Dwarka Sector and Kalkaji might provide a **balanced price-to- ratio**.
- **Hidden Gems:** Shahdara and Vasundhara Enclave may be attractive for **budget-consciou buyers** or **investors seeking high ROI** in emerging areas.

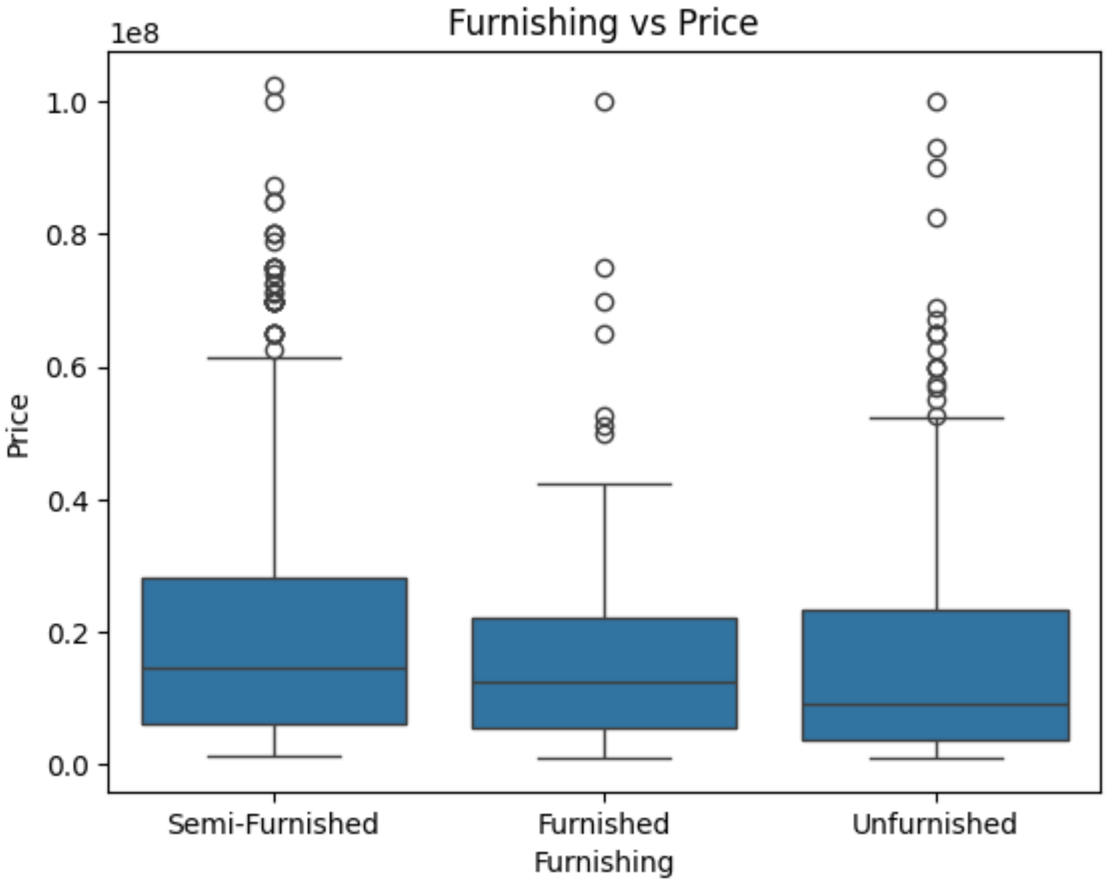
Conclusion

This visualization highlights how **locality greatly influences property pricing**. High-priced area luxury developments, while emerging or outer sectors offer affordability. Strategic selection base locality can drive **better investment decisions** or **targeted marketing** for sellers.

Furnishing and Price

In [17]:
```
sns.boxplot(x = 'Furnishing', y = 'Price', data = df).set_title('Furnishi
```

Out[17]:
```
Text(0.5, 1.0, 'Furnishing vs Price')
```



Furnishing vs Price Box Plot Analysis

Graph Overview

This box plot shows how **property prices** vary based on **furnishing status**:

- **X-axis:** Furnishing type (Semi-Furnished, Furnished, Unfurnished)
- **Y-axis:** Price (in rupees)

Key Observations

1. **Semi-Furnished Properties**

- These have the **highest median price** among all furnishing categories.
- Also show a **wider interquartile range (IQR)**, indicating greater variability.
- Many **high-value outliers** are present, suggesting inclusion of premium semi-furnished listi

2. **Furnished Properties**

- The median price is slightly lower than semi-furnished.
- Price distribution is tighter, implying more **consistent pricing** for fully furnished homes.
- Fewer extreme outliers compared to semi-furnished and unfurnished.

3. **Unfurnished Properties**

- Median price is similar to furnished but with **more high-end outliers**, possibly indicating bu willing to customize interiors.
- Range of prices is broad, suggesting variability in location, area, or amenities.

Market Insights

- **Semi-furnished properties** may offer a **sweet spot** for buyers—cost-effective with basic se done.
- **Fully furnished homes** appeal to a specific segment (e.g., renters, expats) looking for conv but may not fetch the highest prices.
- **Unfurnished listings** might attract **investors or long-term buyers** who prefer personalizat

---

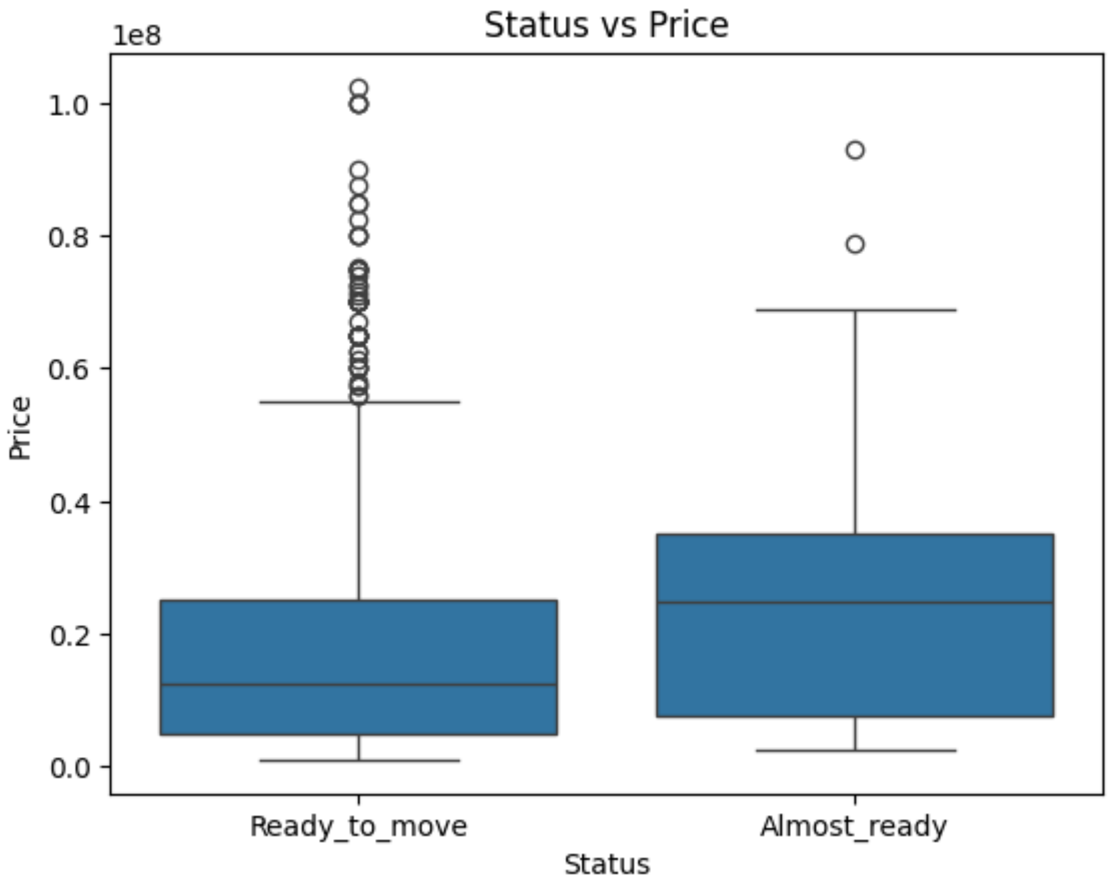
Conclusion

Furnishing has a notable impact on pricing, but **semi-furnished homes** dominate in value. For p sellers, adding basic furnishing could enhance market appeal. Buyers should weigh cost vs conv when evaluating furnishing options.

Status and Price

In [18]:
```
sns.boxplot(x = 'Status', y = 'Price', data = df).set_title('Status vs Pr
```

Out[18]:
```
Text(0.5, 1.0, 'Status vs Price')
```



Status vs Price
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Price
Ready_to_move
Almost_ready
Status

Property Status vs Price Box Plot Analysis

Graph Overview

This box plot displays how **property prices** vary based on their **construction status**:

- **X-axis:** Property status (Ready_to_move vs Almost_ready)
- **Y-axis:** Price (in rupees)

Key Observations

1. **Ready_to_move Properties**

- Lower **median price** compared to almost_ready properties.
- Numerous **outliers** on the higher end suggest many premium listings in this category.
- Moderate interquartile range (IQR), indicating a broad but stable pricing distribution.

2. **Almost_ready Properties**

- **Higher median price** than ready-to-move properties.
- Wider price spread and several **high-end listings**, including fewer outliers.
- May indicate these are newer or part of upscale projects nearing completion.

Market Insights

- **Almost_ready properties** may be priced higher due to **new construction** or **modern ame**
- **Ready-to-move homes**, while more abundant, might include older properties or urgent sale contributing to the price spread and outliers.
- Buyers may pay a premium for **newer, near-complete projects** in anticipation of future app or better features.

Conclusion

The construction **status significantly influences pricing**. While ready-to-move options are curr more available and varied in price, **almost-ready properties trend higher in value**, possibly ref newer development and investor interest.

Transaction Type and Price

In [19]:
```
sns.boxplot(x = 'Transaction', y = 'Price', data = df).set_title('Transac
```

Out[19]:
```
Text(0.5, 1.0, 'Transaction vs Price')
```


Transaction vs Price
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Price
New_Property
Resale
Transaction

Transaction Type vs Price Box Plot Analysis

Graph Overview

This box plot compares property **prices** across two types of transactions:

- **X-axis:** Type of transaction (New_Property vs Resale)
- **Y-axis:** Price (in rupees)

Key Observations

1. **New_Property**

- **Higher median price** compared to resale properties.
- Wider **interquartile range**, suggesting a more diverse price spectrum.
- Numerous **outliers** at the high end, indicating the presence of premium or luxury new listing
- Buyers may be paying a premium for newer construction and modern facilities.

2. **Resale**

- Lower median price overall.
- Tighter distribution with fewer extreme outliers than new properties.
- Could indicate more price predictability or older properties being sold at depreciated values.

Market Insights

- **New properties tend to cost more**, likely due to being recently built, offering updated amer and sometimes including builder premiums.
- **Resale properties** may appeal to budget-conscious buyers or those seeking quicker posse:
- **Higher outliers** in both categories suggest a range of premium listings, but they're more prc in the new property segment.

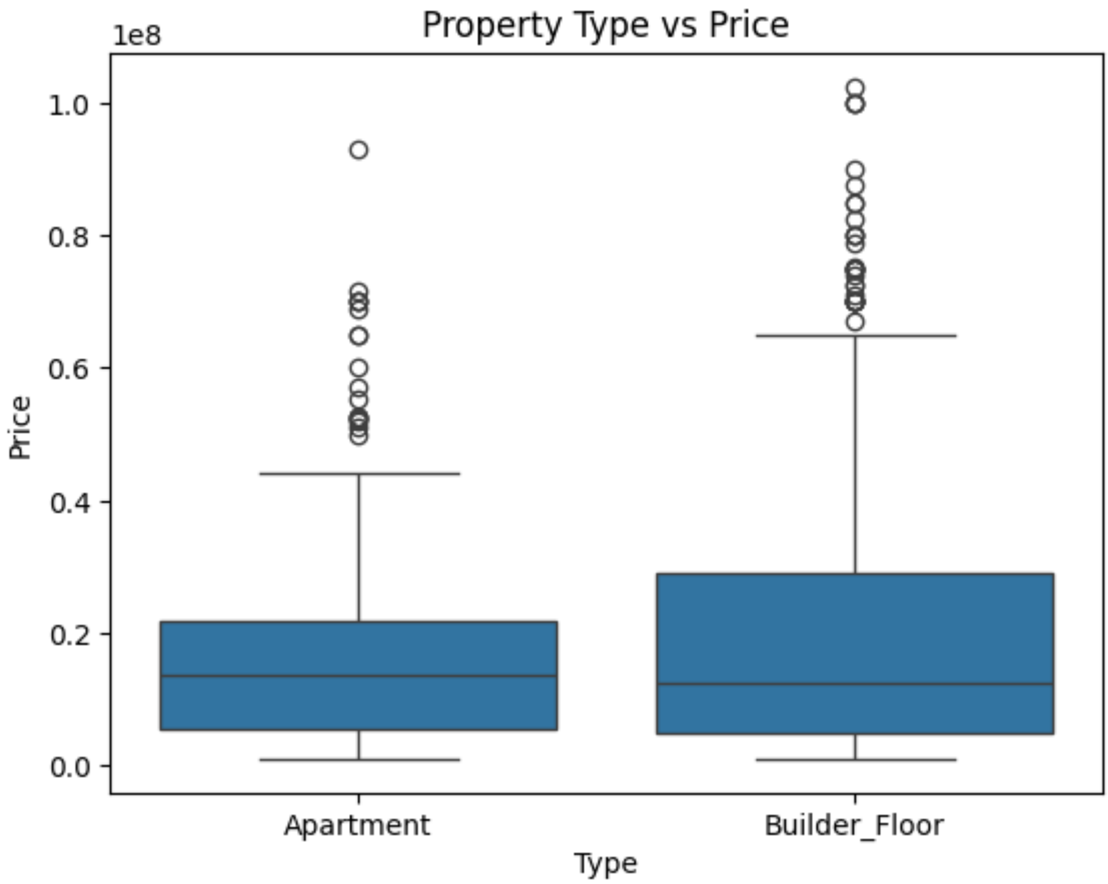
Conclusion

The type of **transaction significantly influences price** trends. New properties generally comma **higher prices and wider variance**, while resale units are **more moderately priced and consis** Buyers should weigh the cost-benefit between **new construction perks** and **resale affordabilit**

Property Type and Price

In [20]:
```
sns.boxplot(x = 'Type', y = 'Price', data = df).set_title('Property Type '
```

Out[20]:
```
Text(0.5, 1.0, 'Property Type vs Price')
```



Property Type vs Price
1e8
1.0
0.8
0.6
0.4
0.2
0.0
Price
Apartment
Builder_Floor
Type

Property Type vs Price Box Plot Analysis

Graph Overview

This box plot compares **property prices** between two types of real estate:

- **X-axis:** Property Type (Apartment vs Builder_Floor)
- **Y-axis:** Price (in rupees)

Key Observations

1. **Apartment**

- **Higher median price** compared to Builder_Floor.
- Tight **interquartile range**, suggesting consistent pricing across most listings.
- Presence of **moderate outliers**, indicating a few premium apartments.

2. **Builder_Floor**

- Slightly **lower median price**, but a **wider range** in price distribution.
- **More high-end outliers** compared to apartments, which could be custom-built or in premiur localities.
- The broader spread may reflect variability in build quality, location, or floor exclusivity.

Market Insights

- **Apartments** show more pricing consistency, possibly due to standardized construction and amenities by builders.
- **Builder floors** offer a mix—some are affordable, while others can be ultra-premium depend floor size, locality, and construction style.
- Buyers seeking **predictable pricing** might prefer apartments, while those looking for **flexibi exclusivity** may lean toward builder floors.

Conclusion

Property **type has a distinct impact on pricing**. While both types overlap in pricing range, build tend to **skew higher with more premium outliers**, whereas apartments provide a **more stable profile**. Your choice depends on your priorities—**stability vs. uniqueness**.