Introduction:

In this report, I will analyze a dataset containing real estate data including property details, pricing, bedrooms, bathrooms, and more. The dataset is loaded using the pandas library and consists of various features such as location, price, property type, area, and seller type. Our objective is to uncover insights and answer specific questions regarding the properties in the dataset.

In [1]:
```
#import Libraries:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

About Sales Data

Data Loading and Initial Exploration:

In [2]:
```
# Read the dataset:
data = pd.read_excel(r"data test Sales.xlsx")
```

In [3]:
```
data
```

Out[3]:

	Property_ID	Location	Property_Type	Listing_Type	Price	Bedrooms	Bathrooms	Area	Seller_Type
0	1	Cairo	Villa	Sale	5000000	4	3	250	Individual
1	2	Giza	Apartment	Sale	1500000	2	1	100	Developer
2	5	Giza	Villa	Sale	3000000	3	2	200	Individual
3	6	Alexandria	Apartment	Sale	1200000	2	1	90	Developer
4	7	Cairo	Villa	Sale	4500000	5	4	300	Individual
5	10	Cairo	Apartment	Sale	2000000	3	2	150	Developer
6	11	Giza	Villa	Sale	3800000	4	3	230	Individual
7	14	Giza	Apartment	Sale	1600000	3	2	130	Developer
8	15	Alexandria	Villa	Sale	5500000	5	4	320	Individual
9	18	Alexandria	Apartment	Sale	1800000	2	1	110	Individual
10	19	Cairo	Villa	Sale	4200000	4	3	280	Developer

In [4]:
```
# Display the first 5 rows of the dataset:
data.head()
```

Out[4]:

	Property_ID	Location	Property_Type	Listing_Type	Price	Bedrooms	Bathrooms	Area	Seller_Type
0	1	Cairo	Villa	Sale	5000000	4	3	250	Individual
1	2	Giza	Apartment	Sale	1500000	2	1	100	Developer
2	5	Giza	Villa	Sale	3000000	3	2	200	Individual
3	6	Alexandria	Apartment	Sale	1200000	2	1	90	Developer

4	7	Cairo	Villa	Sale	4500000	5	4	300	Individual

In [5]:
```
# Calculate summary statistics:
data.describe()
```

Out[5]:

	Property_ID	Price	Bedrooms	Bathrooms	Area
count	11.000000	1.100000e+01	11.000000	11.000000	11.000000
mean	9.818182	3.100000e+06	3.363636	2.363636	196.363636
std	6.177672	1.560769e+06	1.120065	1.120065	84.649008
min	1.000000	1.200000e+06	2.000000	1.000000	90.000000
25%	5.500000	1.700000e+06	2.500000	1.500000	120.000000
50%	10.000000	3.000000e+06	3.000000	2.000000	200.000000
75%	14.500000	4.350000e+06	4.000000	3.000000	265.000000
max	19.000000	5.500000e+06	5.000000	4.000000	320.000000

In [6]:
```
# Checking for null values:
pd.isna(data).sum()
```

Out[6]:
```
Property_ID      0
Location         0
Property_Type    0
Listing_Type     0
Price            0
Bedrooms         0
Bathrooms        0
Area             0
Seller_Type      0
dtype: int64
```

There is no null values

In [7]:
```
#The shape of the data:
data.shape
```

Out[7]:
```
(11, 9)
```

In [8]:
```
#Check for the duplication of the rows:
data.duplicated().sum()
```

Out[8]:
```
0
```

No duplication of the rows

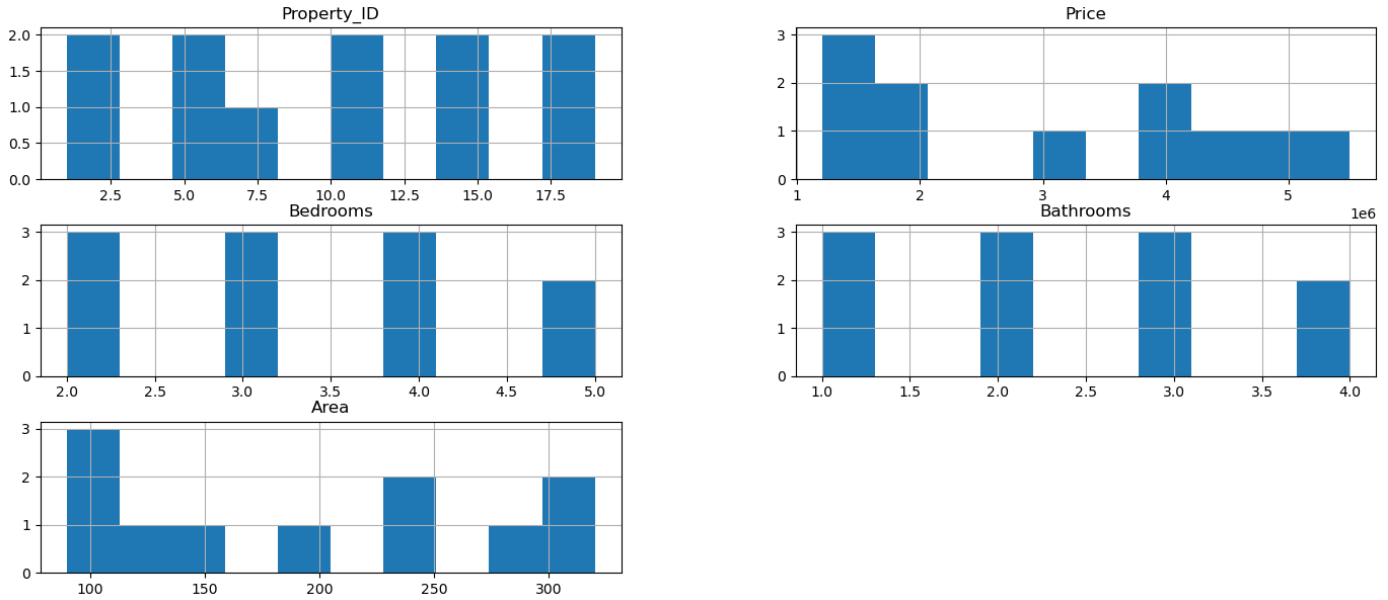
In [9]:
```
# checking the info of the data:
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11 entries, 0 to 10
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   Property_ID    11 non-null     int64 
 1   Location       11 non-null     object
 2   Property_Type  11 non-null     object
 3   Listing_Type   11 non-null     object
 4   Price          11 non-null     int64 
```

```
 5   Bedrooms       11 non-null     int64
 6   Bathrooms      11 non-null     int64
 7   Area           11 non-null     int64
 8   Seller_Type    11 non-null     object
dtypes: int64(5), object(4)
memory usage: 920.0+ bytes
```

Data Exploration and Analysis and Visualization:

In [10]:
```
# Identify trends and patterns:
data.hist(figsize=(17,7));
```



In [11]:
```
#Average Bedrooms and Bath rooms by Location and Property type
# Calculate the average bedrooms and bathrooms by location and property type
average_bed_bath_by_location_property = data.groupby(['Location', 'Property_Type'])[['Be

average_bed_bath_by_location_property
```

Out[11]:

		Bedrooms	Bathrooms
Location	**Property_Type**		
Alexandria	**Apartment**	2.000000	1.000000
	Villa	5.000000	4.000000
Cairo	**Apartment**	3.000000	2.000000
	Villa	4.333333	3.333333
Giza	**Apartment**	2.500000	1.500000
	Villa	3.500000	2.500000

The questions I need to answer from the data are:

1- What is the average price of properties in each location?

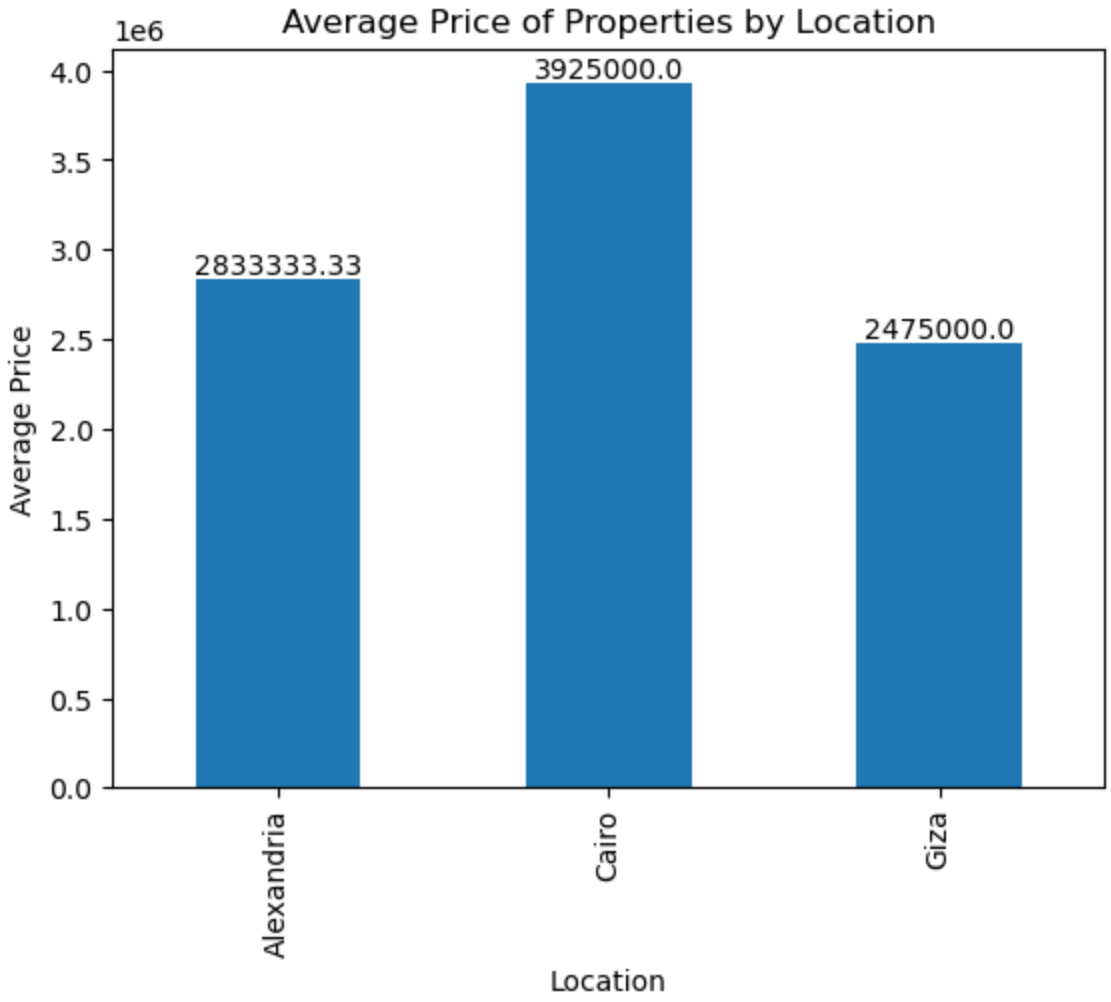
2- What is the average price of properties and Property Type in each location?

3- Which property type (Villa or Apartment) is more common?

4- What is the distribution of properties based on the seller type?

In [25]:
```
# Average price by location
average_price_by_location = data.groupby('Location')['Price'].mean()
print(average_price_by_location)

# Visualize the average price by location
ax = average_price_by_location.plot(kind='bar')
plt.xlabel('Location')
plt.ylabel('Average Price')
plt.title('Average Price of Properties by Location')

# Add data labels above each bar
for p in ax.patches:
    ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_he

plt.show()
```

```
Location
Alexandria    2.833333e+06
Cairo         3.925000e+06
Giza          2.475000e+06
Name: Price, dtype: float64
```



Cairo has the highest average price among all locations.

In [29]:
```
# Average price and Property type by location
average_price_by_location_property_type = data.groupby(['Location', 'Property_Type'])['P
print(average_price_by_location_property_type)

# Reshape the data to have Property_Type as columns and Location as index
average_price_by_location_property_type = average_price_by_location_property_type.unstac

# Plotting the grouped bar chart
ax = average_price_by_location_property_type.plot(kind='bar', stacked=False)
```

```
plt.xlabel('Location')
plt.ylabel('Average Price')
plt.title('Average Price by Location and Property Type')
plt.legend(title='Property Type')

for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() + p.get_width() / 2., p.get_height()), h

plt.show()
```

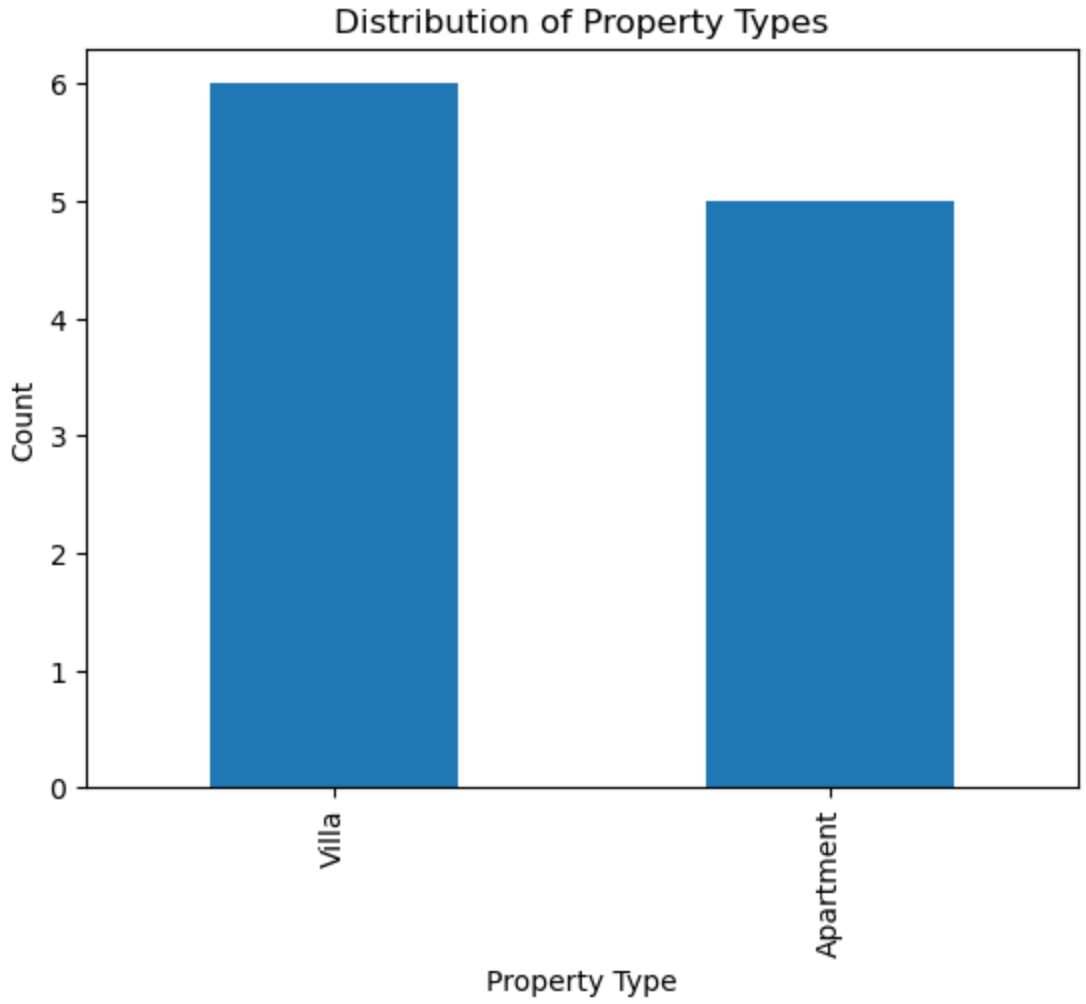
```
Location    Property_Type
Alexandria  Apartment        1.500000e+06
            Villa            5.500000e+06
Cairo       Apartment        2.000000e+06
            Villa            4.566667e+06
Giza        Apartment        1.550000e+06
            Villa            3.400000e+06
Name: Price, dtype: float64
```

The average price of villas is the highest among all property types and locations, and the average price of apartments in Cairo is the highest among all locations.

In [30]:

```
# Count the occurrences of each property type
property_type_counts = data['Property_Type'].value_counts()
print(property_type_counts)

# Plotting the bar chart
property_type_counts.plot(kind='bar')
plt.xlabel('Property Type')
plt.ylabel('Count')
plt.title('Distribution of Property Types')
plt.show()
```

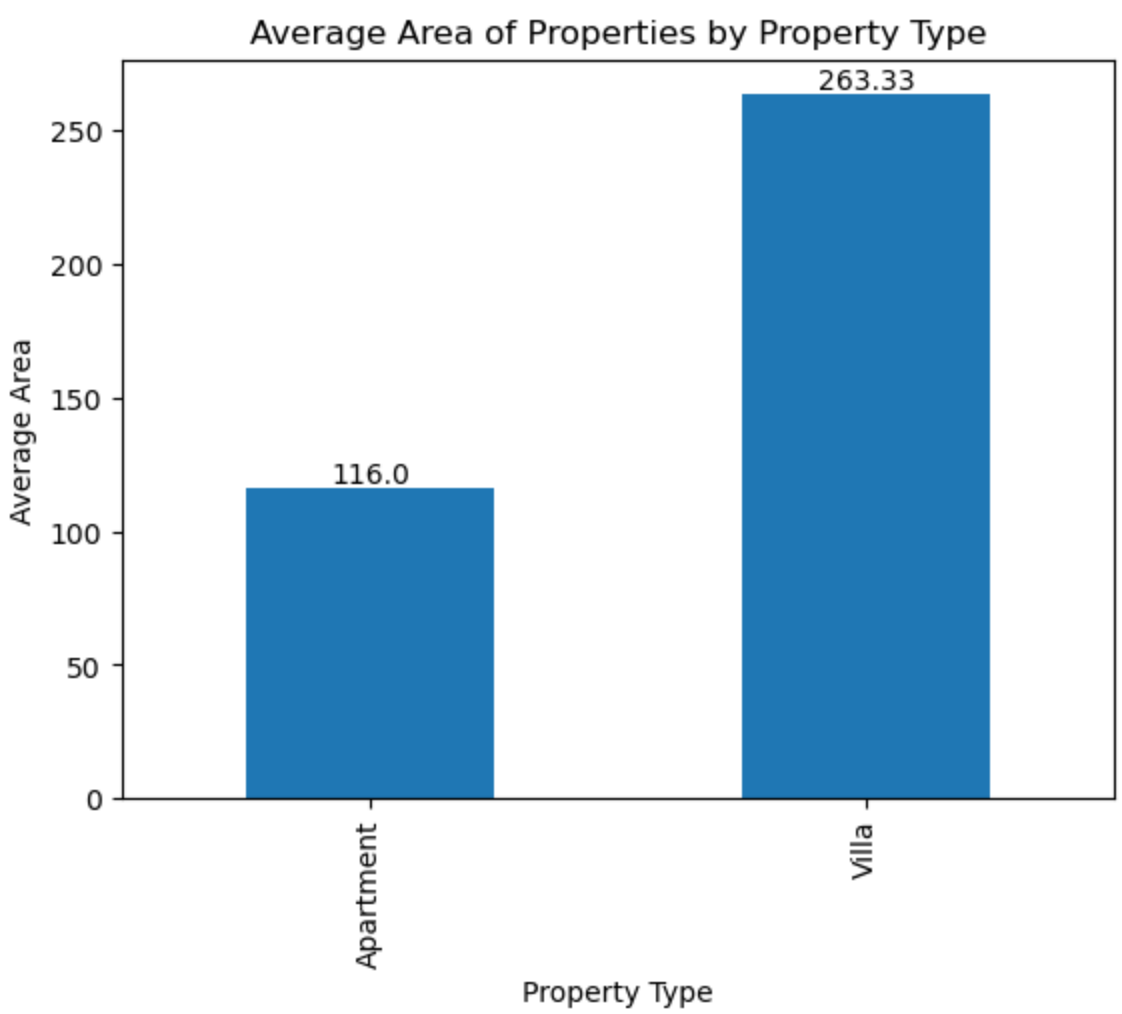
```
Villa        6
Apartment    5
Name: Property_Type, dtype: int64
```



The count of Villas in the data is greater than the count of Apartments in the data.

In [31]:
```
# Average area of properties by property type
average_area_by_property_type = data.groupby('Property_Type')['Area'].mean()
print(average_area_by_property_type)

# Plotting the bar chart
ax = average_area_by_property_type.plot(kind='bar')
plt.xlabel('Property Type')
plt.ylabel('Average Area')
plt.title('Average Area of Properties by Property Type')

for p in ax.patches:
    ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_he

plt.show()
```

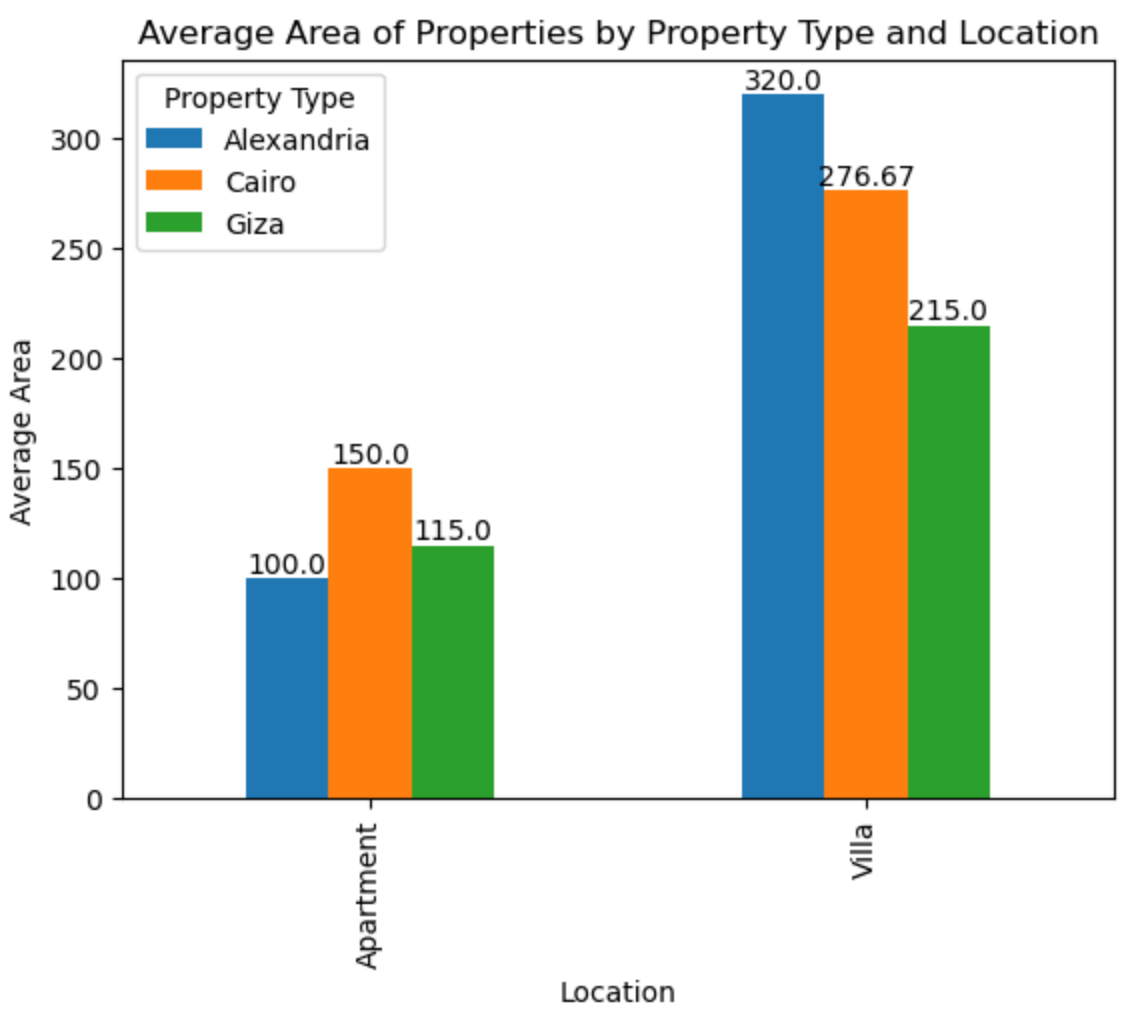
```
Property_Type
Apartment    116.000000
Villa        263.333333
Name: Area, dtype: float64
```



The average area of apartments is 116 square meters, and the average area of villas is 263.33 square meters.

In [32]:
```
#the average area of properties by property type and location:
average_area_by_property_type_location = data.groupby(['Property_Type', 'Location'])['Ar
print(average_area_by_property_type_location)

# Reshape the data to have Property_Type as columns and Location as index
average_area_by_property_type_location = average_area_by_property_type_location.unstack(

# Plotting the grouped bar chart
ax = average_area_by_property_type_location.plot(kind='bar')
plt.xlabel('Location')
plt.ylabel('Average Area')
plt.title('Average Area of Properties by Property Type and Location')
plt.legend(title='Property Type')

# Add data labels above each bar
for p in ax.patches:
    ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_he

plt.show()
```

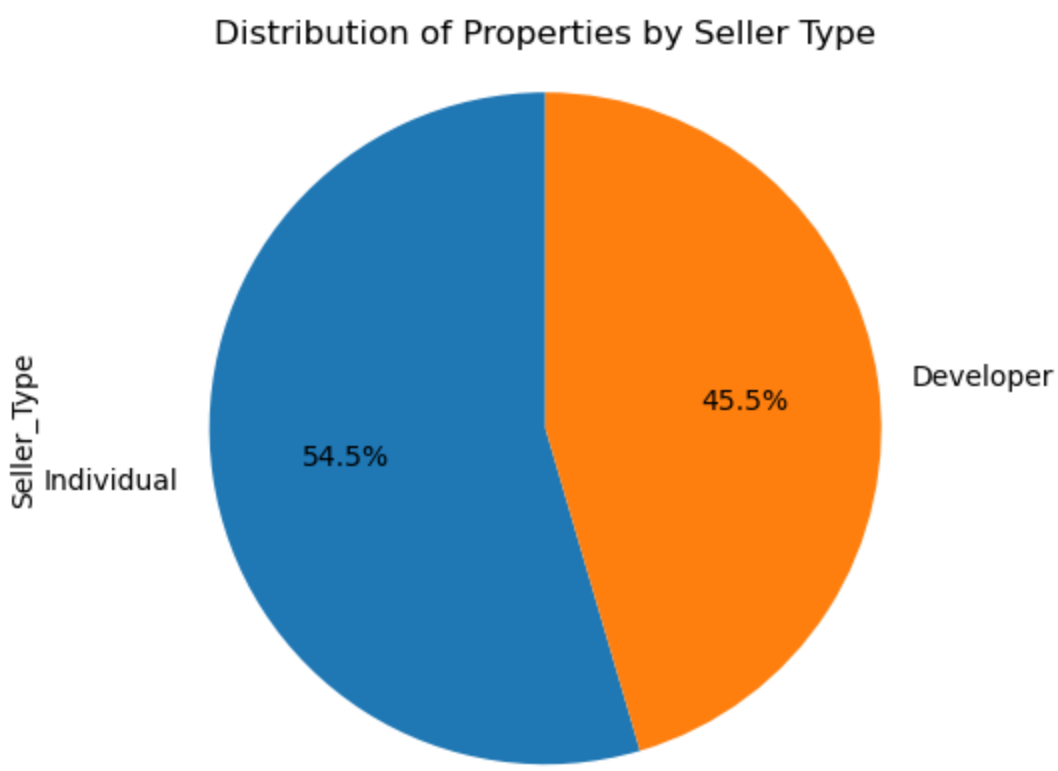
```
Property_Type  Location
Apartment      Alexandria    100.000000
               Cairo         150.000000
               Giza          115.000000
Villa          Alexandria    320.000000
               Cairo         276.666667
               Giza          215.000000
Name: Area, dtype: float64
```



The highest average area is observed for villas in Alexandria, and the highest average area is observed for apartments in Cairo.

In [33]:
```
# Count the properties by seller type
seller_type_counts = data['Seller_Type'].value_counts()
print(seller_type_counts)

# Visualize the distribution of properties by seller type
seller_type_counts.plot(kind='pie', autopct='%1.1f%%', startangle=90)
plt.title('Distribution of Properties by Seller Type')
plt.axis('equal')
plt.show()
```

```
Individual    6
Developer     5
Name: Seller_Type, dtype: int64
```



The distribution of seller types shows that there is a higher proportion of individual sellers compared to developers.

About Rent Data :

In [18]:
```
# Read the dataset
Data = pd.read_excel(r"data test Rent.xlsx")
Data
```

Out[18]:

	Property_ID	Location	Property_Type	Listing_Type	Price	Bedrooms	Bathrooms	Area	Seller_Type
0	3	Alexandria	Apartment	Rent	8000	1	1	80	Individual
1	4	Cairo	Apartment	Rent	10000	2	2	120	Developer
2	8	Giza	Apartment	Rent	7000	1	1	70	Individual
3	9	Alexandria	Apartment	Rent	9000	2	1	100	Developer
4	12	Alexandria	Apartment	Rent	7500	1	1	85	Individual
5	13	Cairo	Apartment	Rent	8500	2	2	110	Developer
6	16	Cairo	Apartment	Rent	9500	2	1	95	Individual
7	17	Giza	Apartment	Rent	7500	1	1	75	Developer
8	20	Giza	Apartment	Rent	6500	1	1	60	Individual

In [19]:
```
# Calculate summary statistics
Data.describe()
```

Out[19]:

	Property_ID	Price	Bedrooms	Bathrooms	Area
count	9.000000	9.000000	9.000000	9.000000	9.000000
mean	11.333333	8166.666667	1.444444	1.222222	88.333333

std	5.830952	1172.603940	0.527046	0.440959	19.525624
min	3.000000	6500.000000	1.000000	1.000000	60.000000
25%	8.000000	7500.000000	1.000000	1.000000	75.000000
50%	12.000000	8000.000000	1.000000	1.000000	85.000000
75%	16.000000	9000.000000	2.000000	1.000000	100.000000
max	20.000000	10000.000000	2.000000	2.000000	120.000000

In [20]:
```
#Average Bedrooms and Bath rooms by Location and Property type
# Calculate the average bedrooms and bathrooms by location and property type
average_bed_bath_by_location_property = Data.groupby(['Location'])[['Bedrooms', 'Bathroo

average_bed_bath_by_location_property
```

Out[20]:

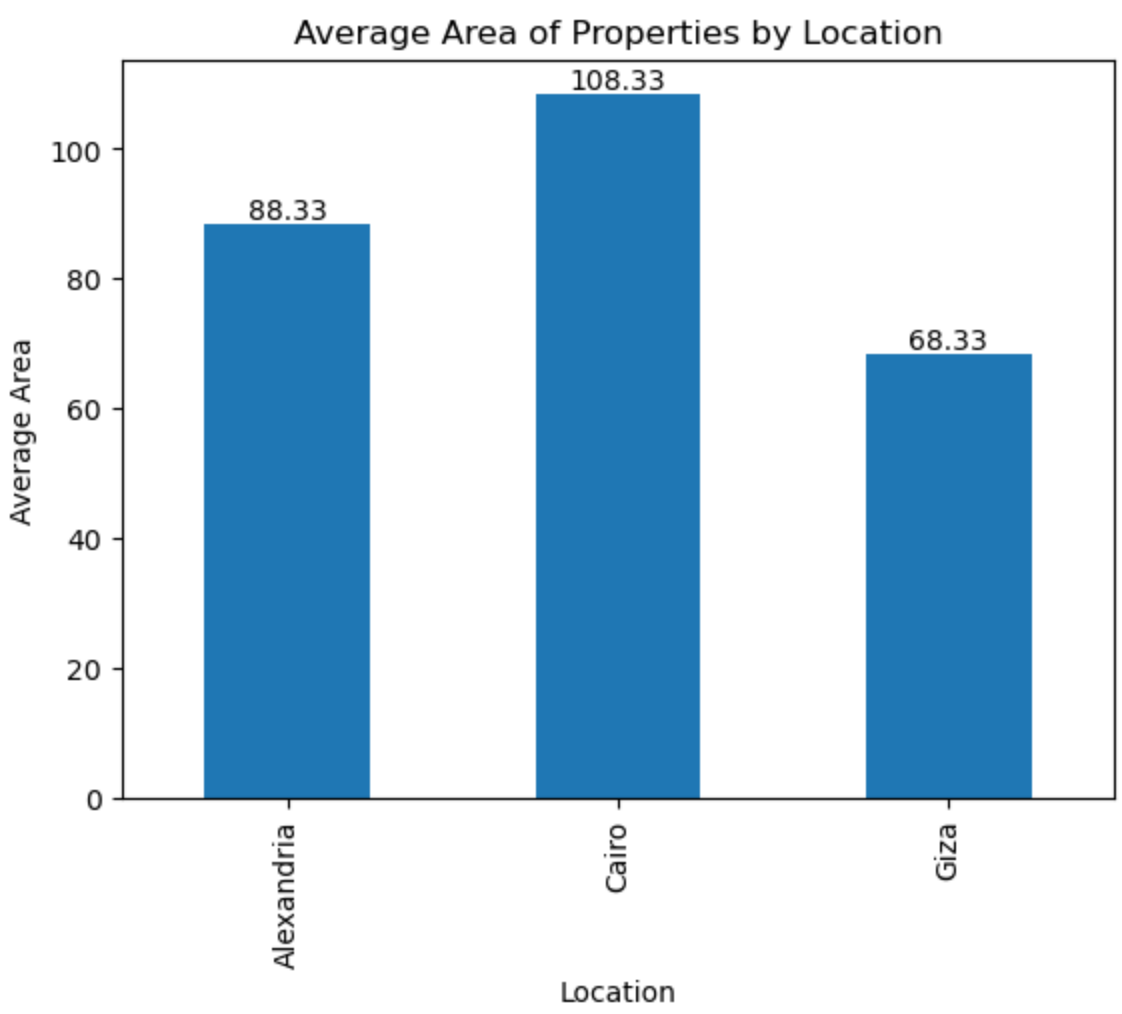
Location	Bedrooms	Bathrooms
Alexandria	1.333333	1.000000
Cairo	2.000000	1.666667
Giza	1.000000	1.000000

In [21]:
```
# Calculate the average price by location
average_price_by_location = Data.groupby('Location')['Price'].mean()

# Plotting the bar chart
ax = average_price_by_location.plot(kind='bar')
plt.xlabel('Location')
plt.ylabel('Average Price')
plt.title('Average Price of Properties by Location')

for p in ax.patches:
    ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_he

plt.show()
```

The average Rent Price in Cairo is the highest

In [34]:
```
# Calculate the average area by location
average_area_by_location = Data.groupby('Location')['Area'].mean()

print(average_area_by_location)

# Plotting the bar chart
ax = average_area_by_location.plot(kind='bar')
plt.xlabel('Location')
plt.ylabel('Average Area')
plt.title('Average Area of Properties by Location')

for p in ax.patches:
    ax.annotate(str(round(p.get_height(), 2)), (p.get_x() + p.get_width() / 2., p.get_he

plt.show()
```

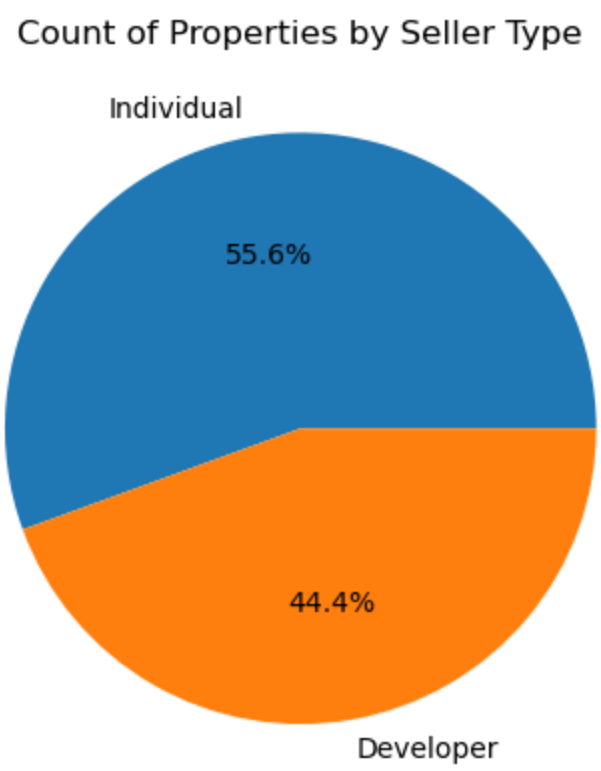
```
Location
Alexandria     88.333333
Cairo         108.333333
Giza           68.333333
Name: Area, dtype: float64
```



the average area of apartments for rent in Cairo is the highest compared to other locations.

In [35]:
```
# Count the properties by seller type
property_count_by_seller_type = Data['Seller_Type'].value_counts()

print(property_count_by_seller_type)

# Plotting the pie chart
plt.pie(property_count_by_seller_type, labels=property_count_by_seller_type.index, autop
plt.title('Count of Properties by Seller Type')

plt.show()
```

```
Individual    5
Developer     4
Name: Seller_Type, dtype: int64
```



The distribution of seller types shows that there is a higher proportion of individual sellers compared to developers.

Real Estate Analysis Insights

Real Estate Data Analysis Dashboard

Developer | Apartment | Rent
Individual | Villa | Sale

1.71M
Average Price

147.75
Average Area

4
Max Bathrooms

5
Max Bedrooms

1
Min Bathrooms

1
Min Bedrooms

320
Max Area

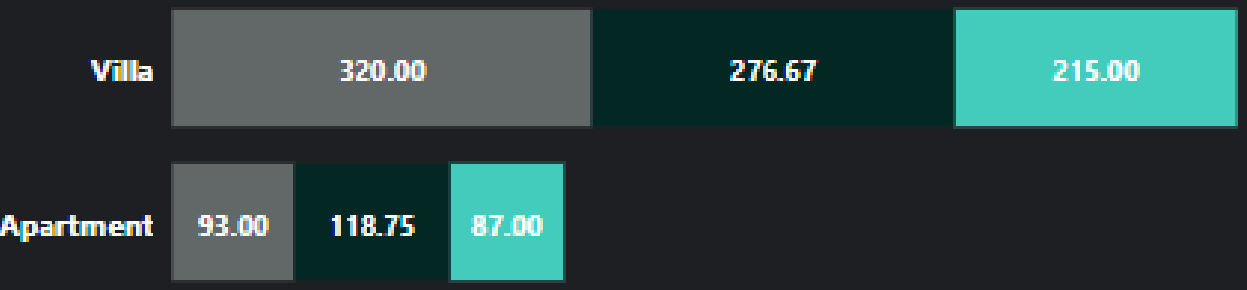
60
Min Area

34M
Total Price

Location
All

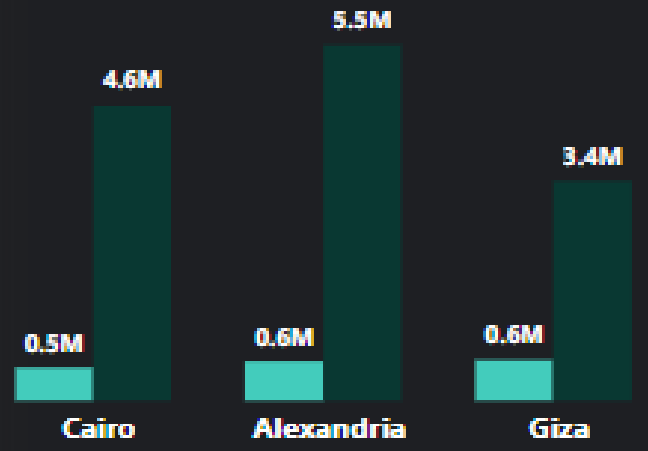
Average Area / Location , Property Type

Alexandria Cairo Giza



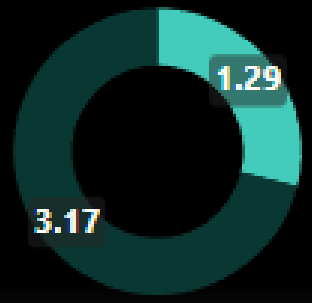
Average Price / Location , Property Type

Apartment Villa



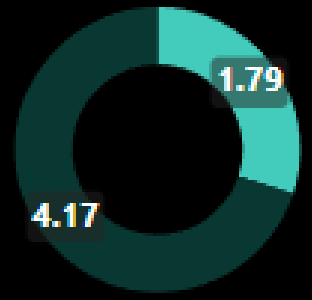
Average Bathrooms / Property Type

Apartment Villa



Average Bedrooms / Property Type

Apartment Villa



Insights For Sales Data:

1- Cairo has the highest prices among all locations and the average price of apartments in Cairo is the highest among all locations and The selling price in Cairo from individuals is the highest, regardless of the area.

In my opinion, the reasons are as follows:

1-Economic Importance: Cairo is the capital city of Egypt and serves as the country's political, administrative, and economic - center. Being the heart of economic activities, Cairo attracts significant investments, businesses, and job opportunities. This economic prominence can drive up demand for properties in Cairo, leading to higher sale prices.

2-Population Density: Cairo has a much larger population compared to Alexandria and Giza. With a population of over 20 million people, Cairo is densely populated, which creates a high demand for housing and limited availability of land. The scarcity of land and high population density can result in higher property prices.

3-Infrastructure and Amenities: Cairo generally offers better infrastructure and a wider range of amenities compared to other cities. It has a well-developed transportation system, including a metro network, which makes commuting easier. Cairo also boasts a variety of educational institutions, healthcare facilities, shopping centers, and entertainment options. The availability of these amenities and infrastructure can contribute to higher property prices.

4-Employment Opportunities: As the capital city, Cairo offers numerous job opportunities across various industries, including government, finance, media, and tourism. Many individuals migrate to Cairo in search of better employment prospects, leading to increased demand for housing. The higher demand for properties can drive up sale prices.

5-Real Estate Market Dynamics: The dynamics of the real estate market, including supply and demand, investor preferences, and market speculation, can influence property prices. Cairo's real estate market may have experienced significant growth and investor interest, leading to higher sale prices. Additionally, factors such as urbanization, gentrification, and development projects can also impact property values.

2- The average price of villas is the highest among all property types and locations.

In my opinion, the reasons are as follows:

1- Demand and Prestige: Villas are often associated with luxury, spaciousness, and exclusivity. They tend to offer more privacy, larger living spaces, and amenities such as gardens, pools, and private parking. The higher average price of villas may be due to the high demand for such prestigious and upscale properties.

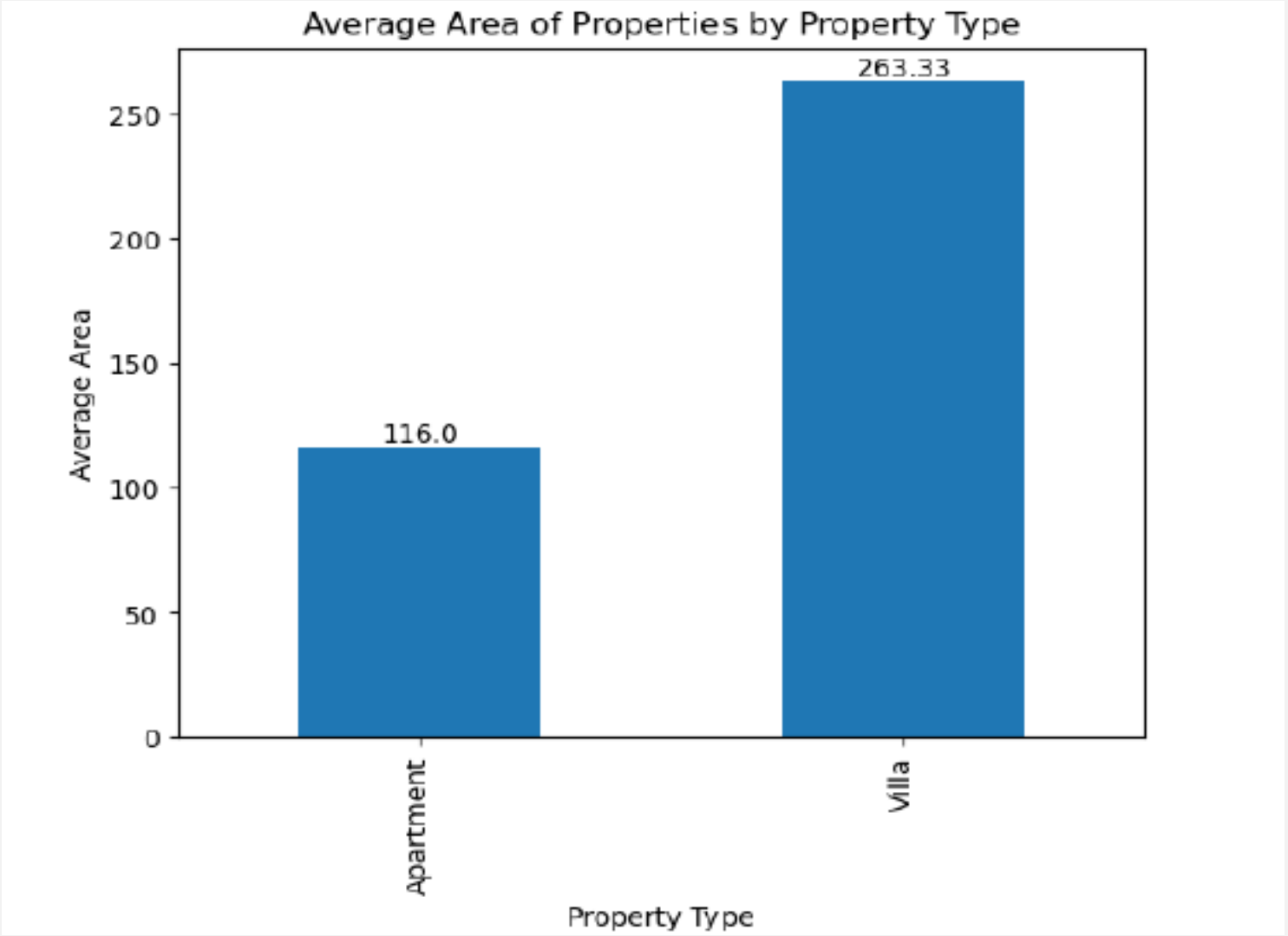
2- Scarcity of Land: Villas typically require larger plots of land compared to other property types. In densely populated areas like cities, finding available land for villas can be challenging. The limited supply of land suitable for villas can drive up their average prices.

3- Location and Desirability: The location of villas can significantly influence their prices. Villas situated in prime or sought-after locations, such as exclusive neighborhoods, waterfronts, or areas with scenic views, tend to command higher prices. Factors like proximity to amenities, good infrastructure, and safety can also contribute to the desirability and higher prices of villas.

4- Quality and Customization: Villas often offer high-quality construction materials, finishes, and customization options. Buyers may be willing to pay a premium for the superior craftsmanship, unique architectural designs, and personalized features that villas offer.

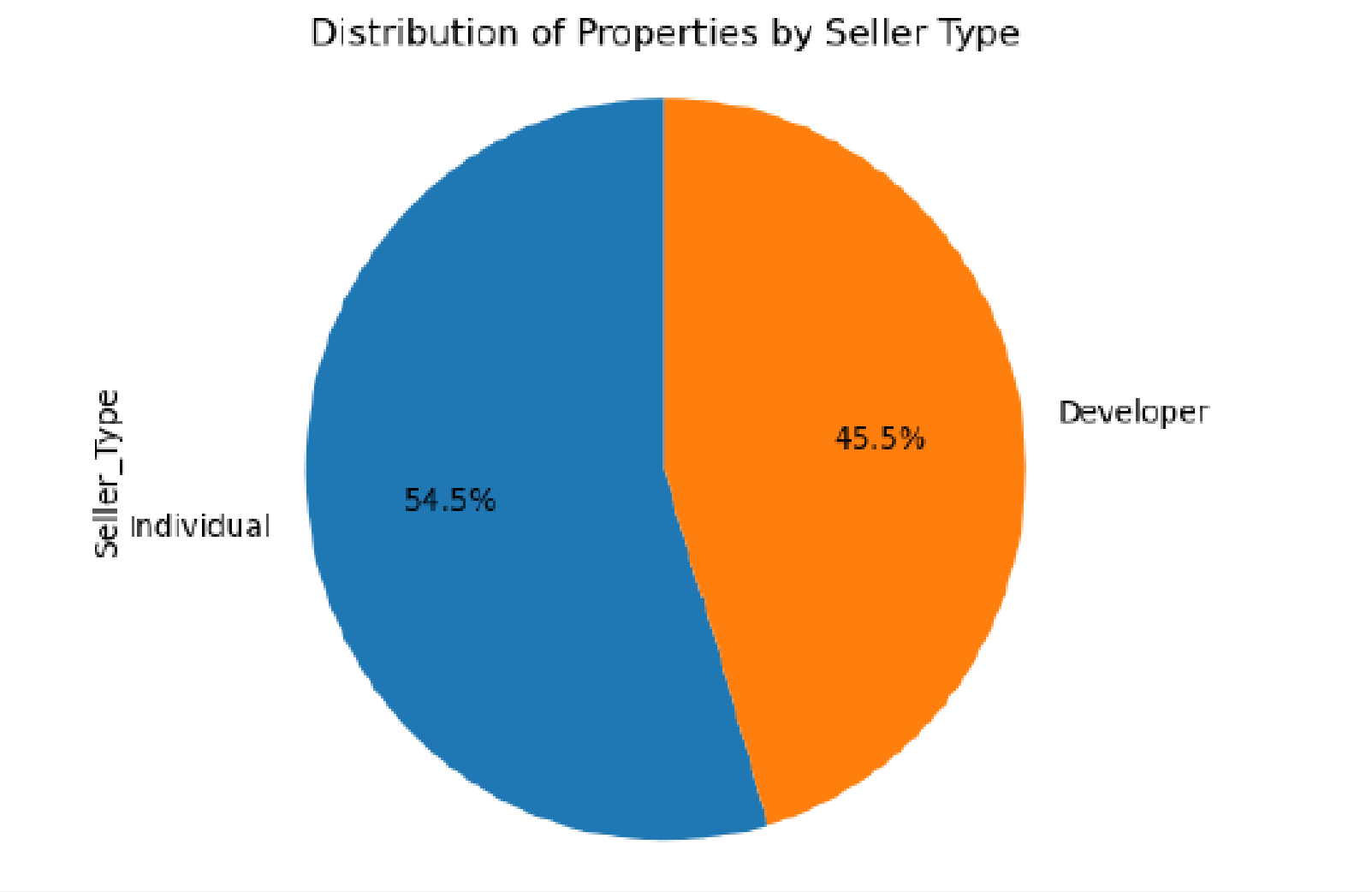
3- The count of villas in the data is greater than the count of apartments.

4- The average area of apartments is 116 square meters, and the average area of villas is 263.33 square meters.



5- The highest average area is observed for villas in Alexandria, and the highest average area is observed for apartments in Cairo.

6- The distribution of seller types shows a higher proportion of individual sellers compared to developers.



7- There is a positive correlation between the selling price in Alexandria and Giza and the area , regardless of the seller type.

In my opinion, the reasons are as follows:

1- Supply and Demand: Larger properties, typically associated with larger areas, are often in higher demand due to the benefits they offer, such as more living space, potential for expansion, and increased privacy. The limited supply of larger properties compared to smaller ones can drive up their prices.

2- Increased Utility: Larger properties provide more usable space, both indoors and outdoors. This can be particularly appealing to buyers who have specific needs, such as families requiring more bedrooms, space for home offices, or outdoor areas for recreational activities.

3- Land Value: The value of land typically increases with its size, especially in desirable areas. As the area of a property increases, so does the land area associated with it. If the property is located in a sought-after neighborhood or has desirable features, the land value can significantly impact the sale price.

4- Construction Costs: Generally, constructing a larger property requires more resources, materials, and labor, resulting in higher construction costs. The sale price of a property often reflects these costs, along with a markup for profit. Therefore, larger properties tend to have higher sale prices to account for the higher construction expenses.

5- Market Preferences: Buyer preferences and trends can influence property prices. In some markets, there may be a preference for larger properties, which can drive up their prices. This could be due to cultural factors, lifestyle preferences, or changing market dynamics.

8- The selling prices in Cairo are generally higher from individuals than from developers.

In my opinion, the reasons are as follows:

Location: Individual sellers or landlords might own properties in prime locations within Cairo, such as in the city center or in popular neighborhoods. These locations often command higher prices due to their proximity to amenities, transportation, and commercial centers.

Flexibility: Individual sellers or landlords may have more flexibility in negotiating prices and rental terms compared to developers who often have fixed pricing structures. This flexibility allows individuals to set higher prices based on factors such as demand, property condition, or unique features.

Supply and demand: In some cases, the supply of properties from developers might outstrip the demand, leading to more competitive pricing and lower rental rates. On the other hand, individual sellers may have a limited number of properties available, creating a situation where demand exceeds supply and drives up prices.

Property condition and customization: Individual sellers may have invested in renovations, upgrades, or customization of their properties, which can justify higher selling or rental prices. Buyers may be willing to pay more for properties that offer desirable features, modern amenities, or unique design elements.

Middleman fees: Individuals often involve real estate agencies or brokers in the sale or rental process, which can add additional fees and commissions to the overall cost. Dealing directly with developers can sometimes eliminate these middleman fees, resulting in lower overall costs for buyers.

Insights For Rent Data:

1- Cairo has the highest prices among all locations.

2- The average area of apartments in Cairo is the highest among all locations.

3- The distribution of seller types shows a higher proportion of individual sellers compared to developers.

4- The rental price in Cairo from individuals is the highest, regardless of the area.

5- There is a direct correlation between the selling price in Alexandria and Giza and the area , regardless of the seller type.

6- The rental prices in Cairo are generally higher from individuals than from developers.


THANK YOU