In [ ]:
```
#Required Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

# Enable inline plotting
%matplotlib inline
```

In [35]:
```
#Load the dataset
file_path = "resale-flat-prices-based-on-registration-date-from-jan-2017-onwards
df = pd.read_csv(file_path)
```

In [37]:
```
#Initial Data Exploration
print("\nFirst 5 rows of the dataset:")
print(df.head())

print("\nData types and missing values:")
print(df.info())

print("\nSummary statistics:")
print(df.describe())
```

```
First 5 rows of the dataset:
     month        town flat_type block        street_name storey_range  \
0  2017-01  ANG MO KIO    2 ROOM   406  ANG MO KIO AVE 10     10 TO 12
1  2017-01  ANG MO KIO    3 ROOM   108   ANG MO KIO AVE 4     01 TO 03
2  2017-01  ANG MO KIO    3 ROOM   602   ANG MO KIO AVE 5     01 TO 03
3  2017-01  ANG MO KIO    3 ROOM   465  ANG MO KIO AVE 10     04 TO 06
4  2017-01  ANG MO KIO    3 ROOM   601   ANG MO KIO AVE 5     01 TO 03

   floor_area_sqm      flat_model  lease_commence_date     remaining_lease  \
0            44.0        Improved                 1979  61 years 04 months
1            67.0  New Generation                 1978  60 years 07 months
2            67.0  New Generation                 1980  62 years 05 months
3            68.0  New Generation                 1980   62 years 01 month
4            67.0  New Generation                 1980  62 years 05 months

   resale_price
0      232000.0
1      250000.0
2      262000.0
3      265000.0
4      265000.0

Data types and missing values:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204452 entries, 0 to 204451
Data columns (total 11 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   month                204452 non-null  object
 1   town                 204452 non-null  object
 2   flat_type            204452 non-null  object
 3   block                204452 non-null  object
 4   street_name          204452 non-null  object
 5   storey_range         204452 non-null  object
 6   floor_area_sqm       204452 non-null  float64
 7   flat_model           204452 non-null  object
 8   lease_commence_date  204452 non-null  int64
 9   remaining_lease      204452 non-null  object
 10  resale_price         204452 non-null  float64
dtypes: float64(2), int64(1), object(8)
memory usage: 17.2+ MB
None

Summary statistics:
       floor_area_sqm  lease_commence_date  resale_price
count   204452.000000        204452.000000  2.044520e+05
mean        96.912628          1996.267833  5.130761e+05
std         24.031798            14.201654  1.803265e+05
min         31.000000          1966.000000  1.400000e+05
25%         82.000000          1985.000000  3.800000e+05
50%         93.000000          1996.000000  4.800000e+05
75%        112.000000          2011.000000  6.150000e+05
max        366.700000          2021.000000  1.600000e+06
```

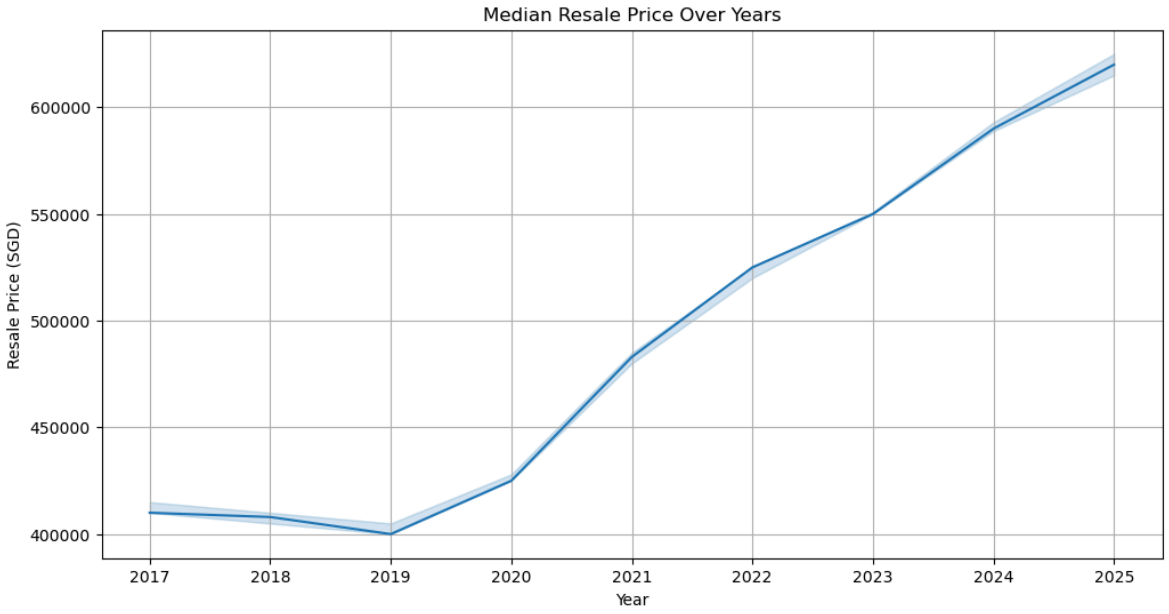
In [39]:

```
# Data Cleaning
# Convert 'month' to datetime
print("\nConverting 'month' column to datetime format...")
df['month'] = pd.to_datetime(df['month'])

# Extract 'year' and 'quarter' for time series analysis
```

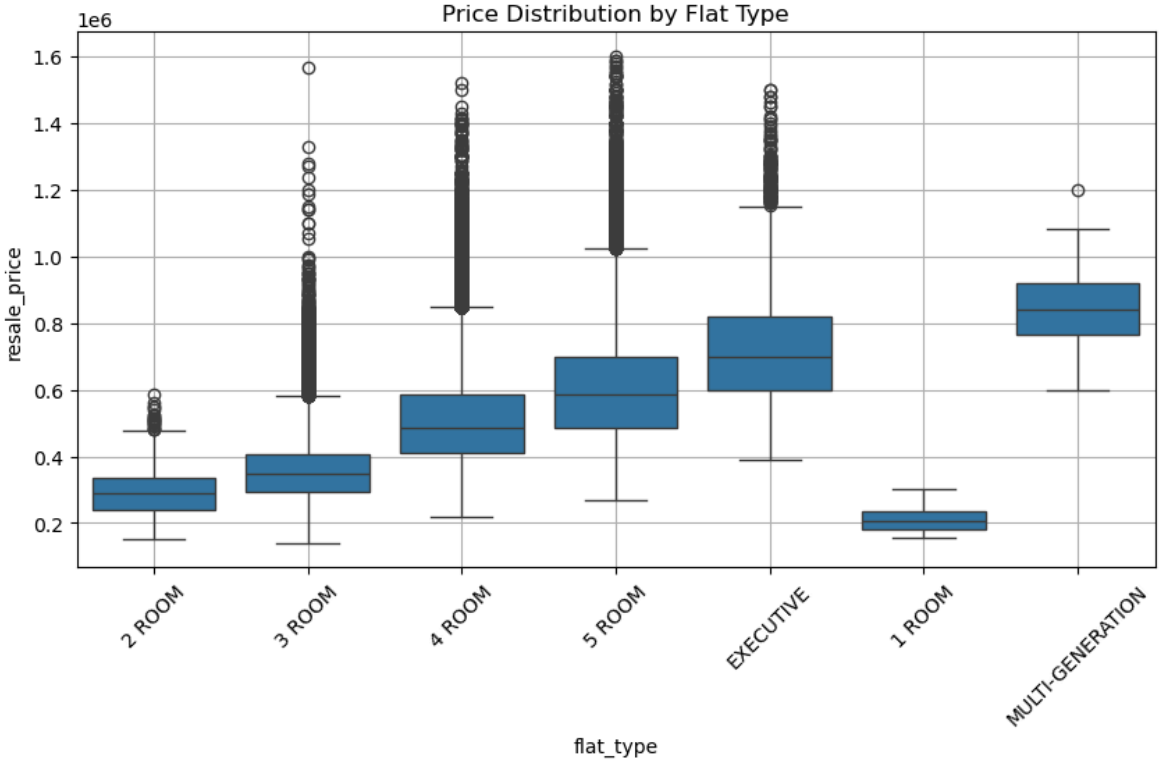
```
df['year'] = df['month'].dt.year
df['quarter'] = df['month'].dt.to_period("Q")

# Create a new feature: price per square meter
df['price_per_sqm'] = df['resale_price'] / df['floor_area_sqm']
```

```
Converting 'month' column to datetime format...
```

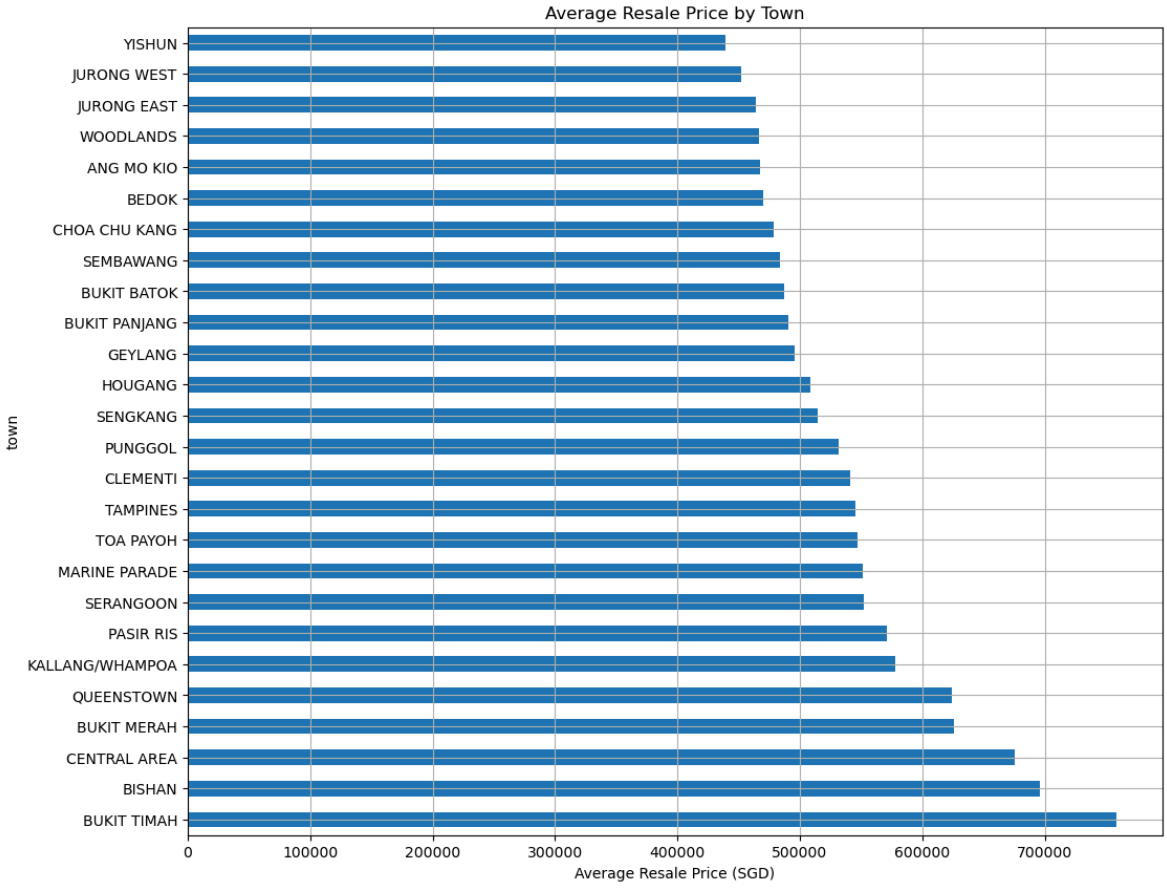
In [41]:
```
# Exploratory Data Analysis (EDA)

## Price Trends Over Time
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='year', y='resale_price', estimator='median')
plt.title("Median Resale Price Over Years")
plt.ylabel("Resale Price (SGD)")
plt.xlabel("Year")
plt.grid(True)
plt.show()
```



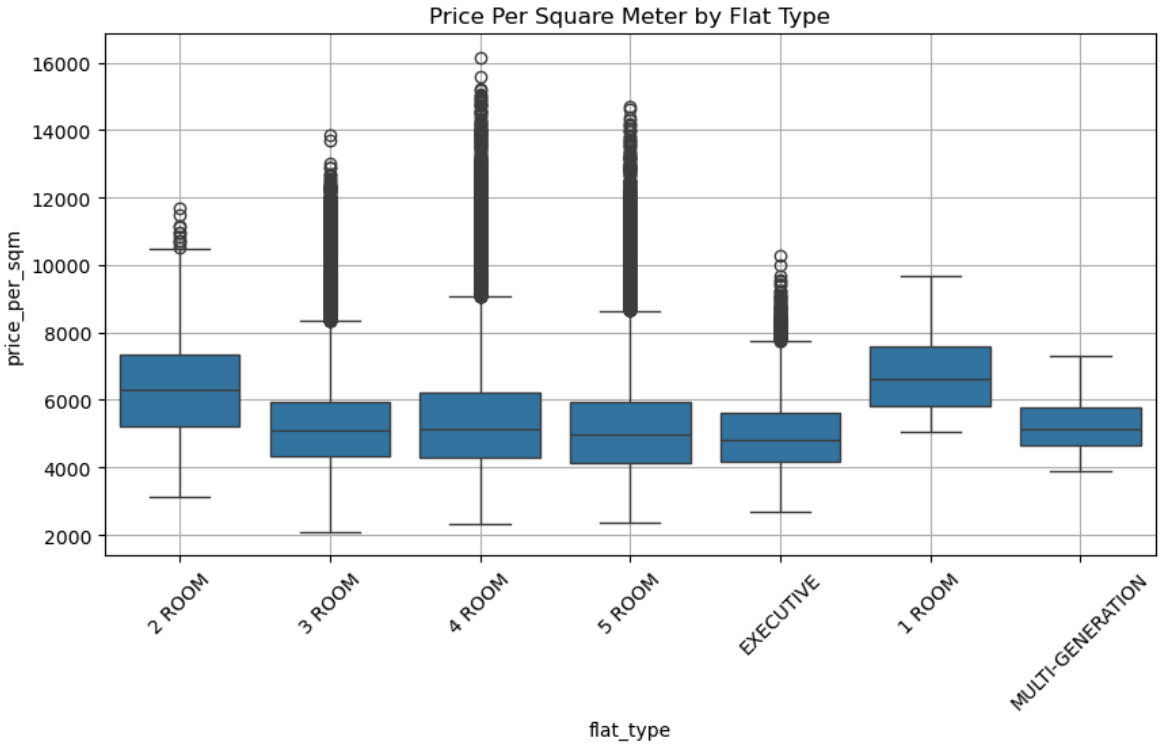
In [42]:
```
## Price Distribution by Flat Type
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='flat_type', y='resale_price')
plt.title("Price Distribution by Flat Type")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



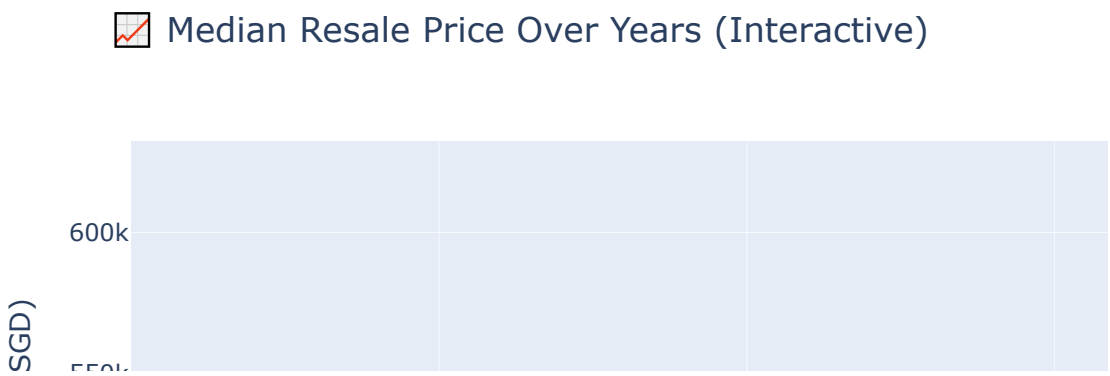
In [43]:
```
## Average Price by Town
avg_price_town = df.groupby('town')['resale_price'].mean().sort_values(ascending
plt.figure(figsize=(12, 10))
avg_price_town.plot(kind='barh')
plt.title("Average Resale Price by Town")
plt.xlabel("Average Resale Price (SGD)")
plt.grid(True)
plt.show()
```



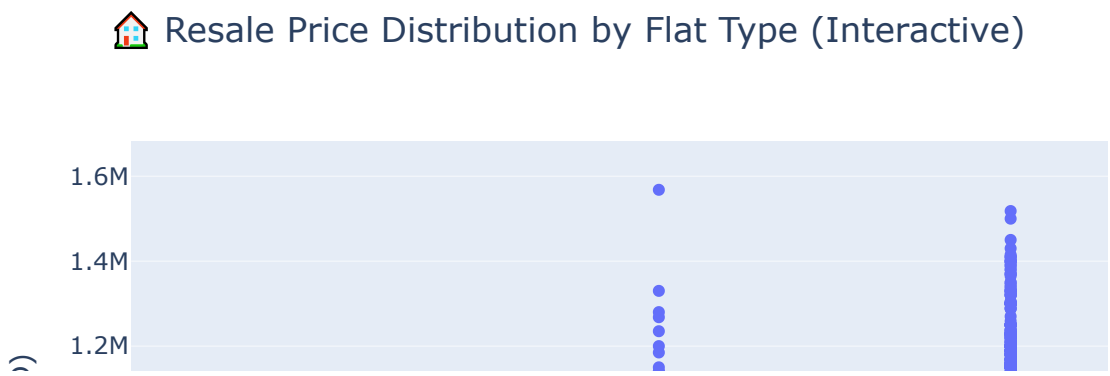
In [50]:
```
## Price per Square Meter by Flat Type
plt.figure(figsize=(10, 5))
sns.boxplot(data=df, x='flat_type', y='price_per_sqm')
plt.title("Price Per Square Meter by Flat Type")
plt.xticks(rotation=45)
plt.grid(True)
plt.show()
```



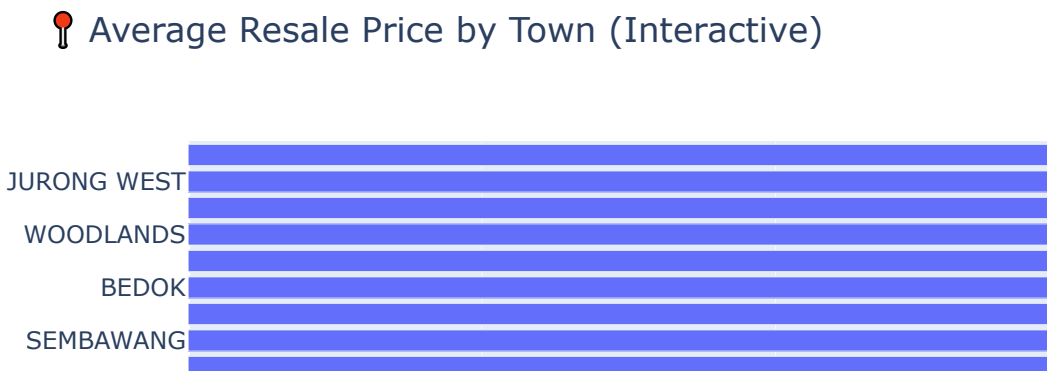
In [52]:
```
# Business Insights (with Interactive Charts)

# 🔍 1. Median Resale Price Over Time (Interactive)
fig1 = px.line(
    df.groupby('year')['resale_price'].median().reset_index(),
    x='year', y='resale_price',
    title='📈 Median Resale Price Over Years (Interactive)',
    labels={'resale_price': 'Median Resale Price (SGD)', 'year': 'Year'}
)
fig1.show()
```



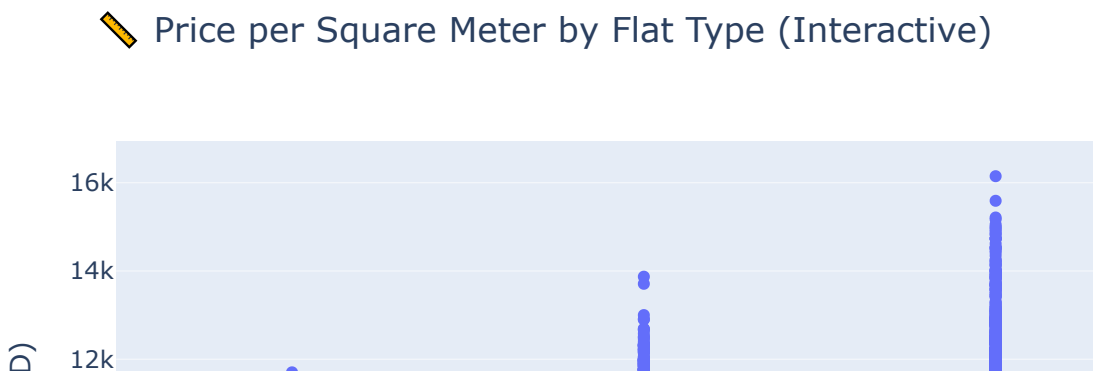
In [54]:
```
# Average Resale Price by Flat Type (Interactive)
fig2 = px.box(
    df, x='flat_type', y='resale_price',
    title='🏠 Resale Price Distribution by Flat Type (Interactive)',
    labels={'resale_price': 'Resale Price (SGD)', 'flat_type': 'Flat Type'}
)
fig2.show()
```



In [56]:
```
#  Average Price by Town (Interactive)
avg_price_town_df = df.groupby('town', as_index=False)['resale_price'].mean().so
fig3 = px.bar(
    avg_price_town_df, x='resale_price', y='town', orientation='h',
    title='📍 Average Resale Price by Town (Interactive)',
    labels={'resale_price': 'Average Resale Price (SGD)', 'town': 'Town'}
)
fig3.show()
```



In [58]:
```
# Price per Square Meter by Flat Type (Interactive)
fig4 = px.box(
    df, x='flat_type', y='price_per_sqm',
    title='📏 Price per Square Meter by Flat Type (Interactive)',
    labels={'price_per_sqm': 'Price per sqm (SGD)', 'flat_type': 'Flat Type'}
)
fig4.show()
```



In [60]:
```
# Save Cleaned Dataset
df.to_csv("cleaned_resale_prices.csv", index=False)
```

📊 Business Insights

This analysis explores resale flat prices in Singapore from 2017 onwards. Below are the key business insights derived from the dataset.

🔍 1. Resale Prices Show a Steady Upward Trend Since 2017

- **Observation**: Median resale prices have consistently increased from 2017 through 2023.
- **Interpretation**: Demand for HDB resale flats remains strong, possibly due to population growth and stable economic recovery post-COVID.
- **Business Implication**: Property investors and developers can expect continued appreciation in property values, especially in mature estates.

🏠 2. Flat Type Strongly Influences Resale Price

- **Observation**: Executive and 5-room flats command significantly higher resale prices than smaller units (e.g., 2-room or 3-room).
- **Interpretation**: Larger flats are more desirable among growing families and multi-generational households.
- **Business Implication**: Developers and agents can focus on marketing 5-room and executive flats to families seeking long-term housing.

📍 3. Towns Like Queenstown and Bukit Timah Have the Highest Average Prices

- **Observation**: Central and mature towns (e.g., Queenstown, Bukit Merah, Bishan) show higher average resale prices.
- **Interpretation**: These areas are well-connected, developed, and attractive for working professionals.
- **Business Implication**: Real estate companies can target high-income buyers and investors in these premium locations.

📐 4. Price per Square Meter Is Highest for Smaller Flats

- **Observation**: 2-room flats have the highest price per sqm, even though their total price is lower.
- **Interpretation**: Buyers are paying a premium for smaller units, likely due to limited supply or demand from singles/elderly.
- **Business Implication**: There's an opportunity to cater to niche segments (e.g., singles or retirees) looking for affordable, compact living spaces.

✅ Summary & Recommendations

- 📈 **Monitor year-on-year trends to anticipate price shifts.**
- 🏘️ **Target larger flat types for families and mid-range buyers.**
- 📍 **Focus on high-demand towns like Queenstown, Bukit Merah, and Bishan.**
- 🎯 **Cater to niche segments (e.g., elderly buyers of 2-room flats).**