Exploring Property Sale Prices

Case Study

DS-01

Mardanova Liana
Elisei Smirnov

Agenda

1. Dataset Used
2. Motivation
3. Statistical Techniques Used
4. Differences in Sale Prices per sqm: Residential vs Commercial Real Estate
5. Impact of Subprime Mortgage Crisis: Changes in Sales Volume and Sales Prices per sqm
6. Variation in Sale Prices per sqm across different Exterior Wall Materials
7. Our mistakes

Motivation:
Which propery to buy?

Motivation

Researching real estate prices offers valuable insights for various stakeholders. Examining the crisis's impact on sales volume and prices per square meter enables the identification of trends and facilitates preparation for future occurrences. Furthermore, delving into price variations based on exterior finishing materials aids in discerning market preferences and determining which material holds greater appeal for potential buyers.

Data Sample Used

Sales data from Milwaukee city from 2002 to 2018

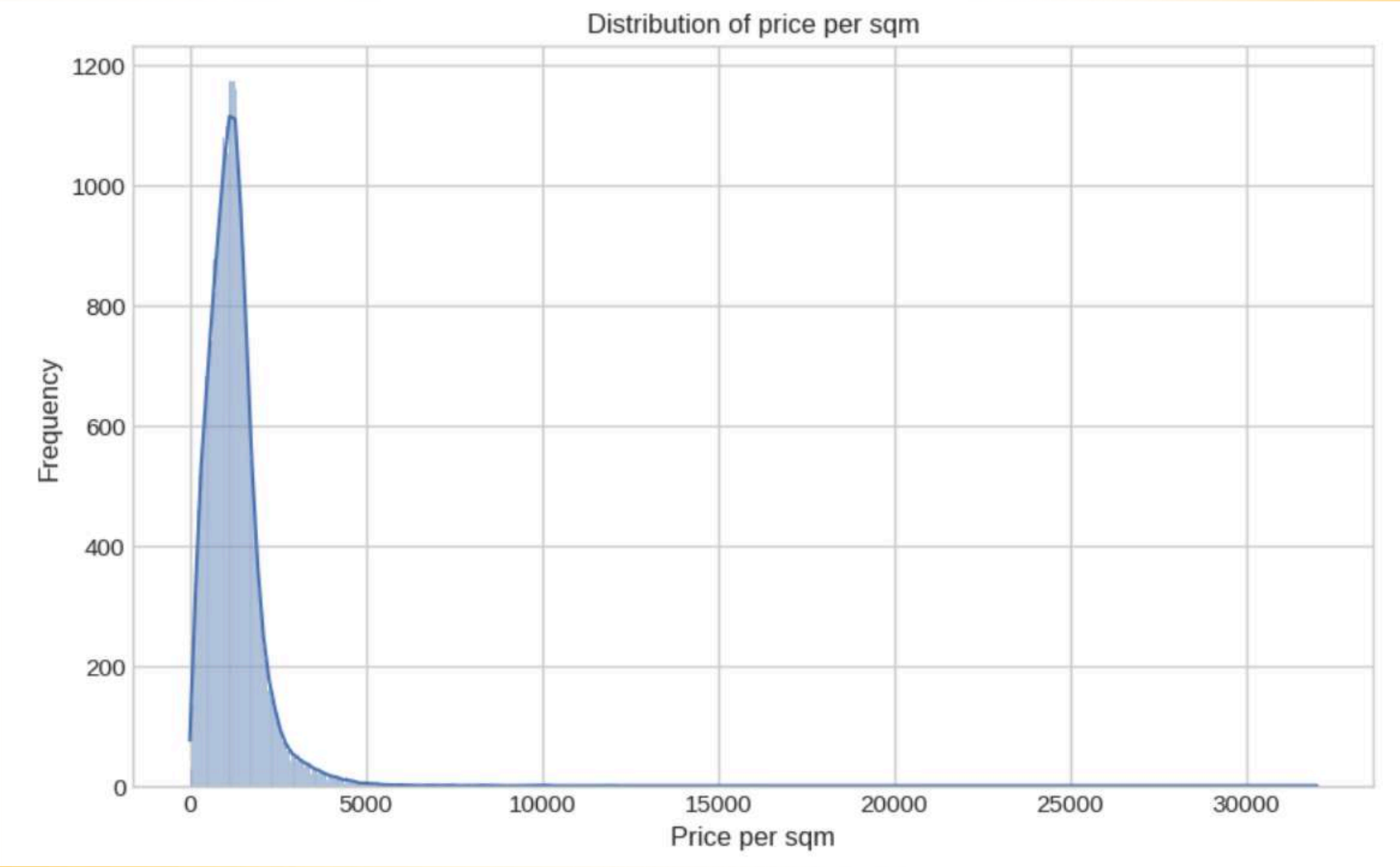
Mostly used fields

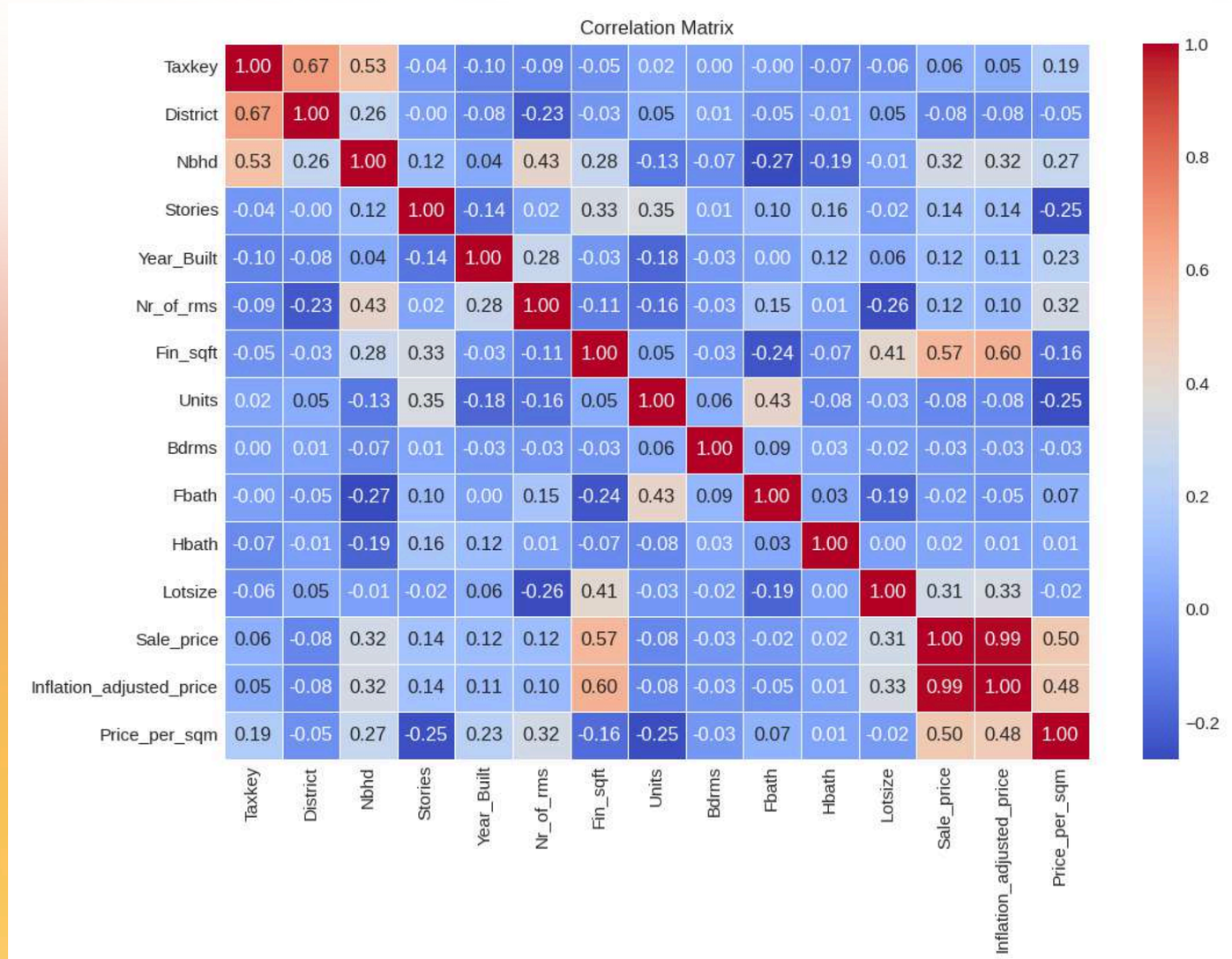
Field Name	Description
PropType	The type of property (e.g., Commercial or Residential).
Extwall	The type of exterior wall material used.
FinishedSqft	The total square footage of finished space in the property.
Sale_date	The date when the property was sold.
Sale_price	The sale price of the property.

Added fields

Inflation_adjusted_price	The inflation-adjusted sale price of the property.
Price_per_sqm	The price per square meter calculated based on the inflation-adjusted price and finished square footage.

Histogram constructed using frequency distribution, accompanied by its KDE plot.





- Property tax is closely correlated with the location of the property.
- The total finished square footage of a property strongly correlates with its sale price.

Statistical Techniques Used

in our case study

Kolmogorov-Smirnov Test	Assesses whether samples come from a specified distribution
Kruskal-Wallis Test	Determines if multiple samples come from the same distribution
Mann-Whitney U Test	Compares two sample means to determine their equality
Cohen's d	Measures the effect size of differences between two means
Rank-biserial correlation	Quantifies the association between a binary and a continuous variable
Chi-square test of independence	Assesses association between two categorical variables in a contingency table.

Differences in Sale Prices per sqm:

Residential vs Commercial Real Estate

01

KS Test for Normality of Price Data

H0: sale prices per sqm follows normal distribution

Residential Property

```
mean = np.mean(log_residential_prices)
std = np.std(log_residential_prices)

_, p_value = kstest(log_residential_prices,
                    'norm',
                    args=(mean, std))
```

```
p_value
```

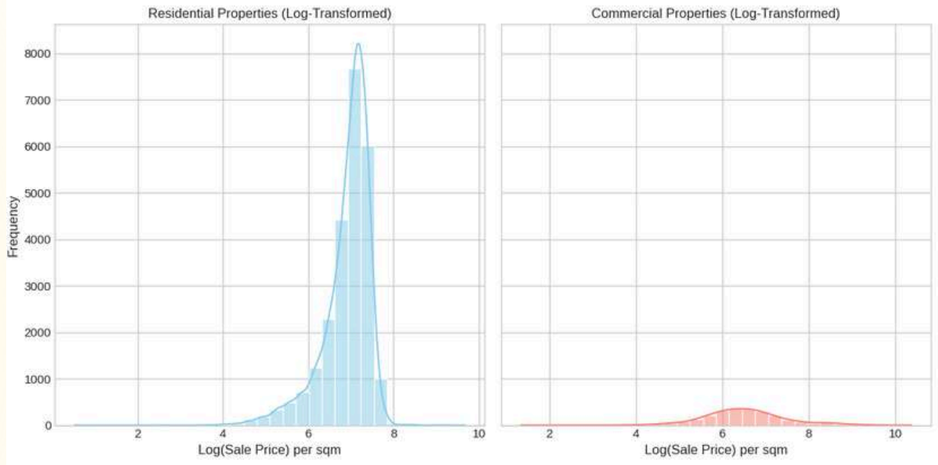
```
0.0
```

Commercial Property

```
mean = np.mean(log_commercial_prices)
std = np.std(log_commercial_prices)

_, p_value = kstest(log_commercial_prices,
                    'norm',
                    args=(mean, std))
```

```
p_value
```

```
2.6118370306104404e-05
```



Differences in Sale Prices per sqm :

Residential vs Commercial Real Estate

01

KS Test for Normality of Price Data

H0: sale prices per sqm follows normal distribution

Residential Property

```
mean = np.mean(log_residential_prices)
std = np.std(log_residential_prices)

_, p_value = kstest(log_residential_prices,
                    'norm',
                    args=(mean, std))
```

```
p_value
```

```
0.0
```

Commercial Property

```
mean = np.mean(log_commercial_prices)
std = np.std(log_commercial_prices)

_, p_value = kstest(log_commercial_prices,
                    'norm',
                    args=(mean, std))
```

```
p_value
```

```
2.6118370306104404e-05
```

H0 Rejected.

Sale prices per sqm do not follow normal distribution.

Differences in Sale Prices per sqm:

Residential vs Commercial Real Estate

02

Mann-Whitney U test*

H0: no significant difference in the sale prices per sqm between residential and commercial properties.

```
from scipy.stats import mannwhitneyu

result = mannwhitneyu(log_residential_prices,
                      log_commercial_prices)

U_test = result.statistic
result.pvalue
```

```
1.8219566653432748e-187
```

*non-parametric test since data is not normally distributed

Differences in Sale Prices per sqm:

Residential vs Commercial Real Estate

02

Mann-Whitney U test*

H0: no significant difference in the sale prices between residential and commercial properties.

```
from scipy.stats import mannwhitneyu

result = mannwhitneyu(log_residential_prices,
                      log_commercial_prices)

U_test = result.statistic
result.pvalue
1.8219566653432748e-187
```

H0 Rejected.

There is a significant difference in the sale prices per sqm between residential and commercial properties.

*non-parametric test since data is not normally distributed

Differences in Sale Prices per sqm:

Residential vs Commercial Real Estate

03

Effect size

Rank-biserial correlation

```
n_r = len(df_filtered.loc[df_filtered['PropType'] == 'Residential'])
n_c = len(df_filtered.loc[df_filtered['PropType'] == 'Commercial'])
```

```
effect_size = 1 - (2*U_test)/(n_r*n_c)

effect_size
```

```
-0.3577868105636728
```

a moderate negative effect size, implying that commercial properties tend to have lower sale prices per sqm than residential properties

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

Subprime mortgage crisis

The American subprime mortgage crisis was a multinational financial crisis that occurred between 2007 and 2010 that contributed to the 2007–2008 global financial crisis. The crisis led to a severe economic recession, with millions of people losing their jobs and many businesses going bankrupt. Wikipedia

End date: 2010

Start date: 2007

Location: United States

We partitioned the data into two distinct epochs:

Economic Crisis Period

(2007 to 2010)

Stable Economic Period

(2010 to 2018)

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

	Stable	Crisis
Residential Properties Volume	22191	2363
Commercial Properties Volume	956	253

01

Chi-square test of independence

H0: there is no significant difference in the change in sales volume between the stable and crisis economic periods for residential and commercial properties.

	Stable	Crisis
Residential Properties Expected Freq.	22060.76	22493.24
Commercial Properties Expected Freq.	1086.24	122.76

```
from scipy.stats import chi2_contingency

observed = [[crisis_residential_volume, crisis_commercial_volume],
            [stable_residential_volume, stable_commercial_volume]]

chi2_contingency(observed).pvalue
```

```
1.0671343662715973e-36
```

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

	Stable	Crisis
Residential Properties Volume	22191	2361
Commercial Properties Volume	956	100

01

Chi-square test of independence

	Stable	Crisis
Residential Properties Expected Freq.	22192.48	2359.52
Commercial Properties Expected Freq.	954.52	101.48

H0: there is no significant difference in the change in sales volume between the stable and crisis economic periods for residential and commercial properties.

```
from scipy.stats import chi2_contingency

observed = [[crisis_residential_volume, crisis_commercial_volume],
            [stable_residential_volume, stable_commercial_volume]]
```

```
0.9163897603790214
```

H0 Rejected.

There is a significant difference in the change in sales volume between the stable and crisis economic periods for residential and commercial properties.

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

02

KS Test for Normality of Price Data

H0: Prices per sqm for stable and crisis periods are normally distributed

```
_, p_value = kstest(crisis_residential_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
0.00011102861499953041
```

Crisis Residential

```
mean = np.mean(stable_residential_prices['Price_per_sqm'])
std = np.std(stable_residential_prices['Price_per_sqm'])

_, p_value = kstest(stable_residential_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
7.467503471727292e-15
```

Stable Residential

```
mean = np.mean(crisis_commercial_prices['Price_per_sqm'])
std = np.std(crisis_commercial_prices['Price_per_sqm'])

_, p_value = kstest(crisis_commercial_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
1.259098292478997e-12
```

Crisis Commercial

```
_, p_value = kstest(stable_commercial_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
3.976354117276576e-62
```

Stable Commercial

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

02

KS Test for Normality of Price Data

H0: Prices per sqm for stable and crisis periods are normally distributed

Crisis Residential

```
_, p_value = kstest(crisis_residential_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
0.00011102861499953041
```

Stable Residential

```
mean = np.mean(stable_residential_prices['Price_per_sqm'])
std = np.std(stable_residential_prices['Price_per_sqm'])

_, p_value = kstest(stable_residential_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
7.467503471727292e-15
```

Crisis Commercial

```
mean = np.mean(crisis_commercial_prices['Price_per_sqm'])
std = np.std(crisis_commercial_prices['Price_per_sqm'])

_, p_value = kstest(crisis_commercial_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
1.2590982924789976-12
```

H0 Rejected.

Sale prices do not follow normal distribution.

Stable Commercial

```
_, p_value = kstest(stable_commercial_prices['Price_per_sqm'],
                    'norm',
                    args=(mean, std))
p_value
```

```
3.976354117276576e-62
```

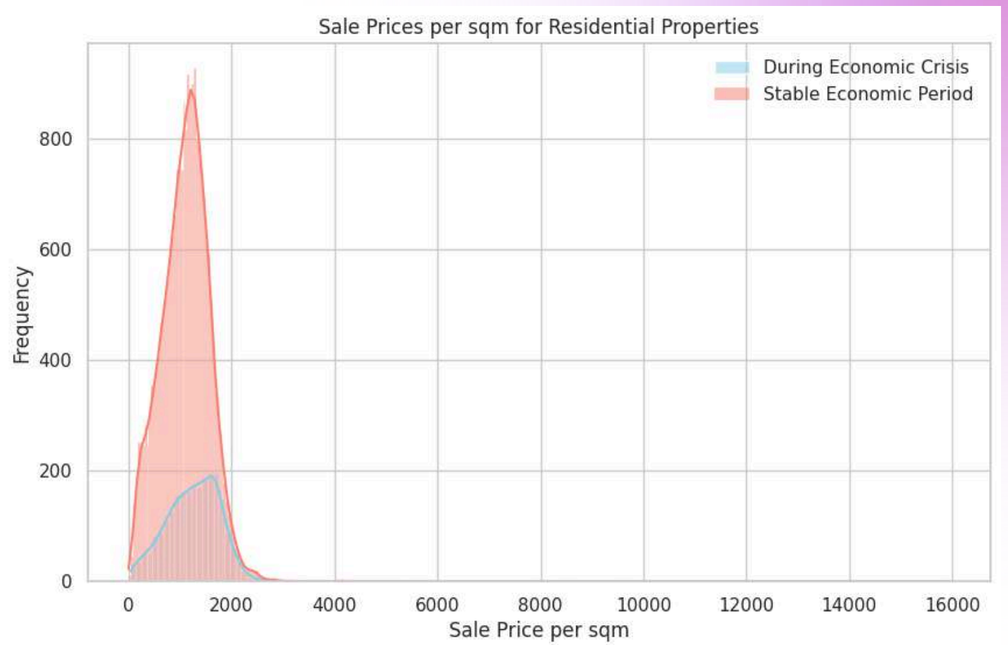
Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

03

Mann-Whitney U test

H0: There is no significant difference in the prices per sqm between residential properties during crisis periods and stable periods.




```
mannwhitneyu(crisis_residential_prices['Price_per_sqm'],
             stable_residential_prices['Price_per_sqm']
             ).pvalue
```

```
5.789641655420015e-47
```

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

03

Mann-Whitney U test

H0: There is no significant difference in the prices per sqm between residential properties during crisis periods and stable periods.


```
mannwhitneyu(crisis_residential_prices['Price_per_sqm'],
                                        ce_per_sqm']
                ).pvalue
2.7398327273647 09e-47
```

H0 Rejected.

There is a significant difference between the price per sqm of residential properties during crisis periods compared to stable periods.

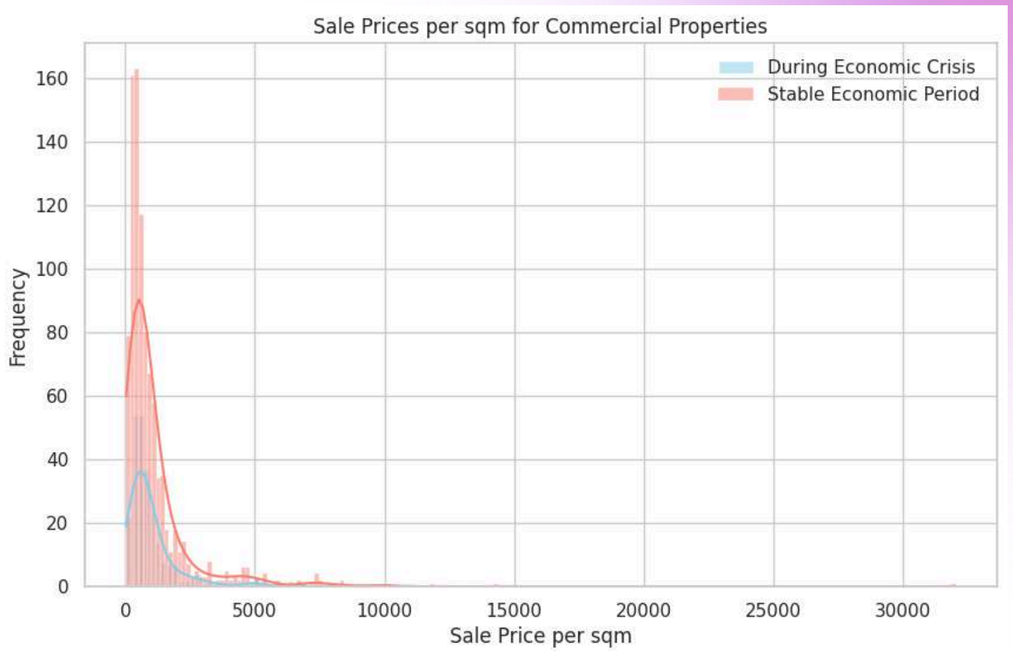
Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

04

Mann-Whitney U test

H0: There is no significant difference in the prices per sqm between commercial properties during crisis periods and stable periods.




```
mannwhitneyu(crisis_commercial_prices['Price_per_sqm'],
             stable_commercial_prices['Price_per_sqm']
             ).pvalue
```

```
0.24771065920971602
```

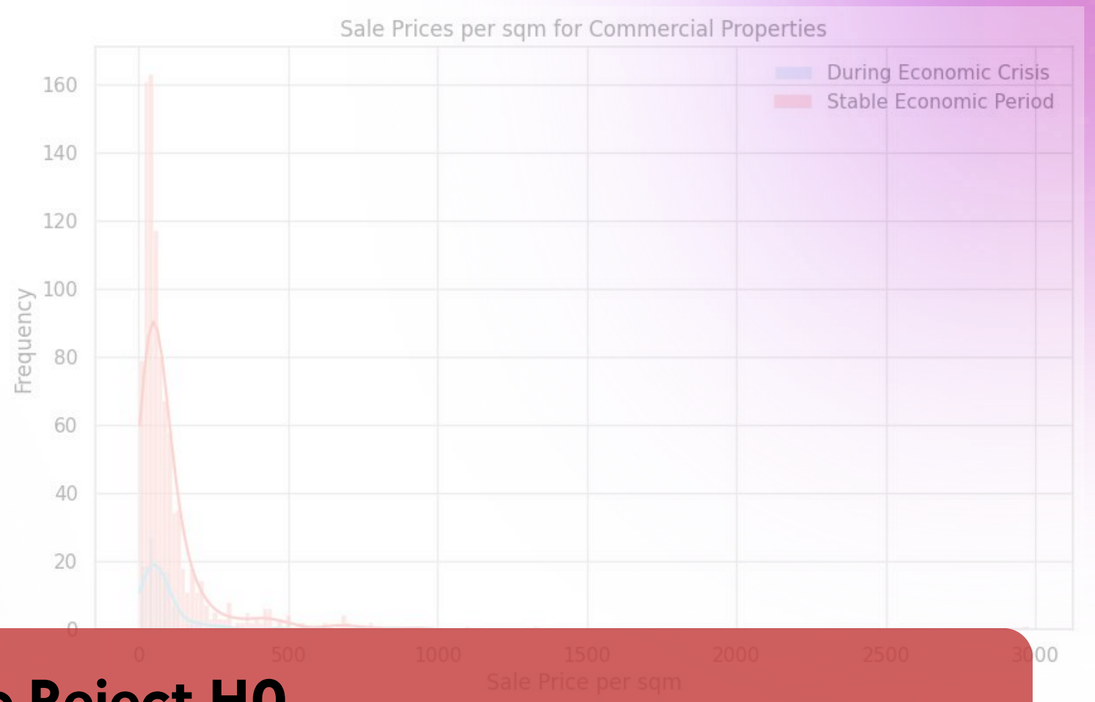
Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

04

Mann-Whitney U test

H0: There is no significant difference in the prices per sqm between commercial properties during crisis periods and stable periods.



Fail to Reject H0.

We cannot conclude tat there is a significant difference between the price per sqm of commercial properties during crisis periods compared to stable periods.

Impact of Subprime Mortgage Crisis:

Changes in Sales Volume and Sales Prices per sqm

05

Effect size

Cohen’s d

```
Cohen's d for Residential Properties: 0.2854714532260657
```

The mean price per sqm for residential properties during the crisis is significantly higher than during the stable period. The effect size is moderate, indicating a noticeable difference in prices.

```
Cohen's d for Commercial Properties: -0.0635925472972412 7
```

The mean price per sqm for commercial properties during the stable period is slightly higher than during the crisis. The effect size is small, suggesting a minimal difference in prices.

Variation in Sale Prices per sqm

across different Exterior Wall Materials

01

KS Test for Normality of Price Data

H0: sale prices per sqm across different exterior wall materials follows normal distribution

Aluminum / Vinyl	KS Test p-value: 1.6799463268698312e-08
Block	p-value: 0.18116323767763753
Brick	p-value: 3.5826771821146805e-13
Frame	p-value: 3.387265329637e-05
Masonry / Frame	p-value: 0.1293190394459751 1
Prem Wood	p-value: 0.3862091550098079
Stone	p-value: 0.51283999349352
Stucco	p-value: 0.366874820386936
Fiber-Cement	p-value: 0.9600050635721683

Reject H0: Does not follow normal distribution	**Aluminum / Vinyl**	KS Test p-value: 1.6799463268698312e-08
Fail to Reject H0: Follows normal distribution	**Block**	p-value: 0.18116323767763753
Reject H0: Does not follow normal distribution	**Brick**	p-value: 3.5826771821146805e-13
Reject H0: Does not follow normal distribution	**Frame**	p-value: 3.387265329637e-05
Fail to Reject H0: Follows normal distribution	**Masonry / Frame**	p-value: 0.1293190394459751 1
Fail to Reject H0: Follows normal distribution	**Prem Wood**	p-value: 0.3862091550098079
Fail to Reject H0: Follows normal distribution	**Stone**	p-value: 0.51283999349352
Fail to Reject H0: Follows normal distribution	**Stucco**	p-value: 0.366874820386936
Fail to Reject H0: Follows normal distribution	**Fiber-Cement**	p-value: 0.9600050635721683

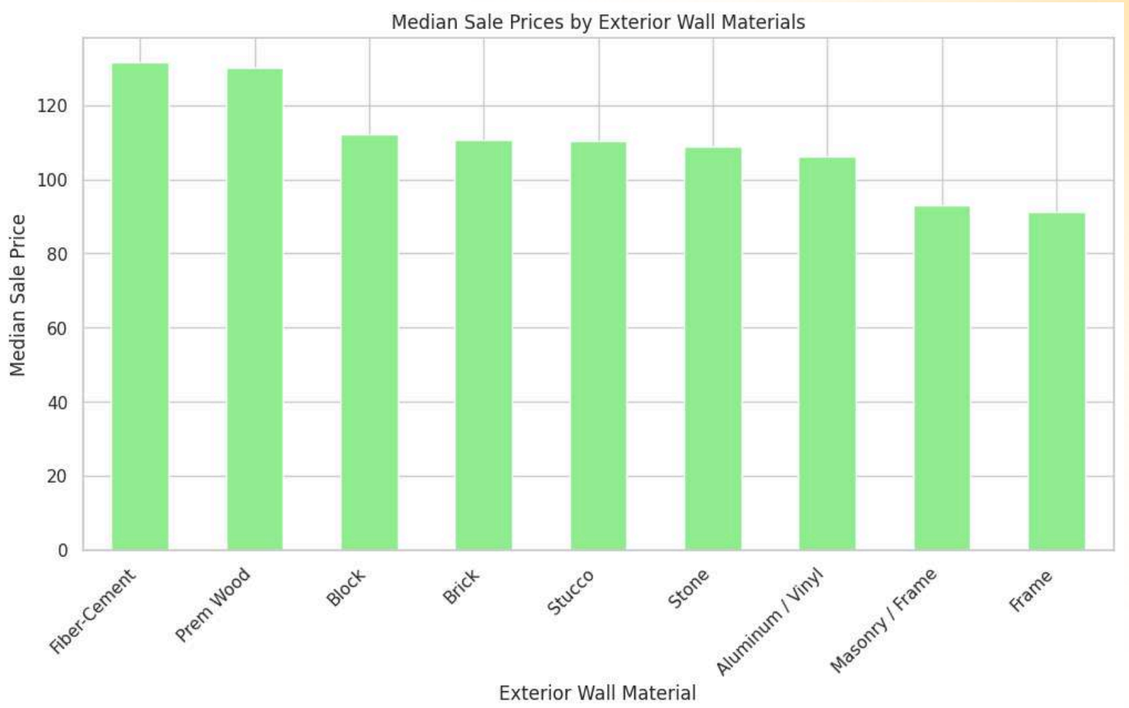
Variation in Sale Prices per sqm

across different Exterior Wall Materials

02

Kruskal-Wallis Test

H0: There is no significant difference in the median sale prices per sqm across different exterior wall materials

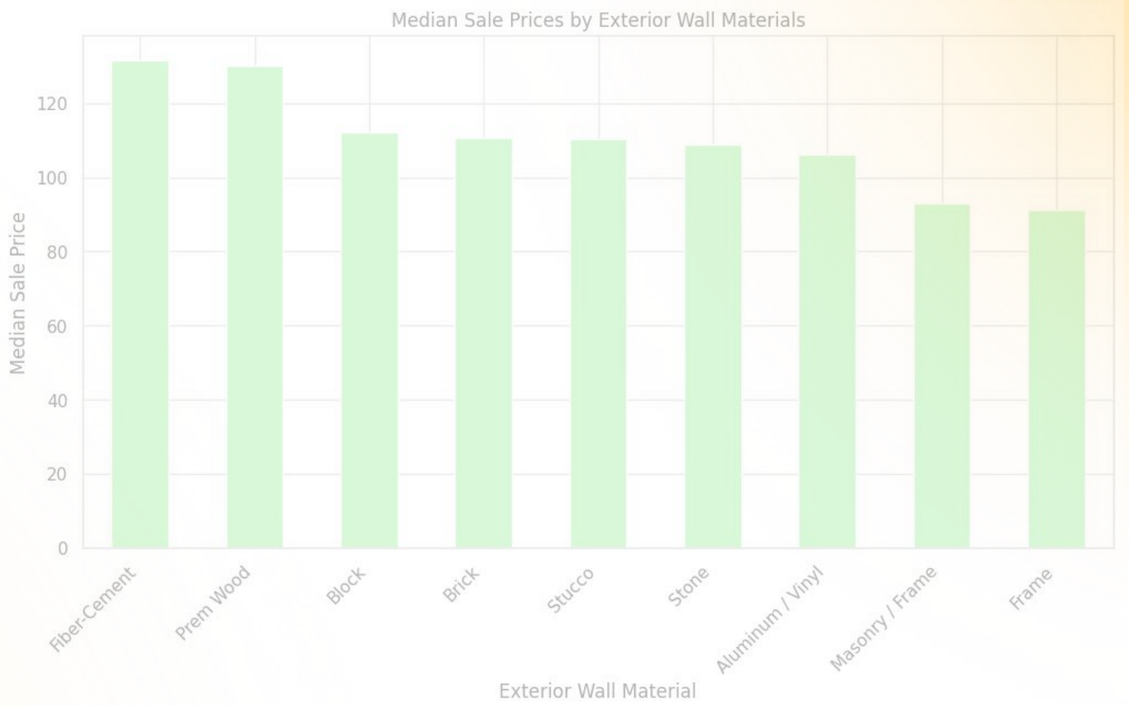



```
from scipy.stats import kruskal

H_extwall, p_value_extwall = kruskal(*[group[1]['Price_per_sqm']
                                       for group in df.groupby('Extwall')])

print(f"H = {H_extwall}, p-value = {p_value_extwall}")
```

```
H = 435.8475925783783, p-value = 3.97788958012855e-89
```

Variation in Sale Prices per sqm

across different Exterior Wall Materials



02

Kruskal-Wallis Test

H0: There is no significant difference in the median sale prices per sqm across different exterior wall materials

H0 Rejected.

There is a significant difference in the median sale prices per sqm across the different ex. wall materials

Variation in Sale Prices per sqm

across different Exterior Wall Materials

04

Pairwise Mann-Whitney U tests with Bonferroni correction

Material 1	Material 2	p-value
Frame	Stone	4.33e-21
Frame	Brick	5.87e-68
Frame	Aluminum / Vinyl	4.35e-36
Frame	Stucco	2.42e-12
Frame	Block	0.000256
Frame	Prem Wood	2.60e-10
Frame	Fiber-Cement	1.32e-24
Stone	Masonry / Frame	9.55e-11
Stone	Prem Wood	1.04e-05
Stone	Fiber-Cement	8.51e-14
Brick	Aluminum / Vinyl	1.69e-18
Brick	Masonry / Frame	6.91e-23
Brick	Prem Wood	8.87e-05
Brick	Fiber-Cement	2.16e-12
Aluminum / Vinyl	Masonry / Frame	1.15e-08
Aluminum / Vinyl	Prem Wood	2.49e-06
Aluminum / Vinyl	Fiber-Cement	3.16e-16
Stucco	Masonry / Frame	7.08e-07
Stucco	Prem Wood	0.001415
Stucco	Fiber-Cement	8.73e-08
Block	Masonry / Frame	0.003463
Block	Prem Wood	0.000593
Block	Fiber-Cement	3.07e-07
Masonry / Frame	Prem Wood	1.60e-10
Masonry / Frame	Fiber-Cement	2.55e-23

p-values indicate diverse distributions in price per square meter, suggesting no uniformity in pricing based on material type.

Variation in Sale Prices per sqm

across different Exterior Wall Materials

05

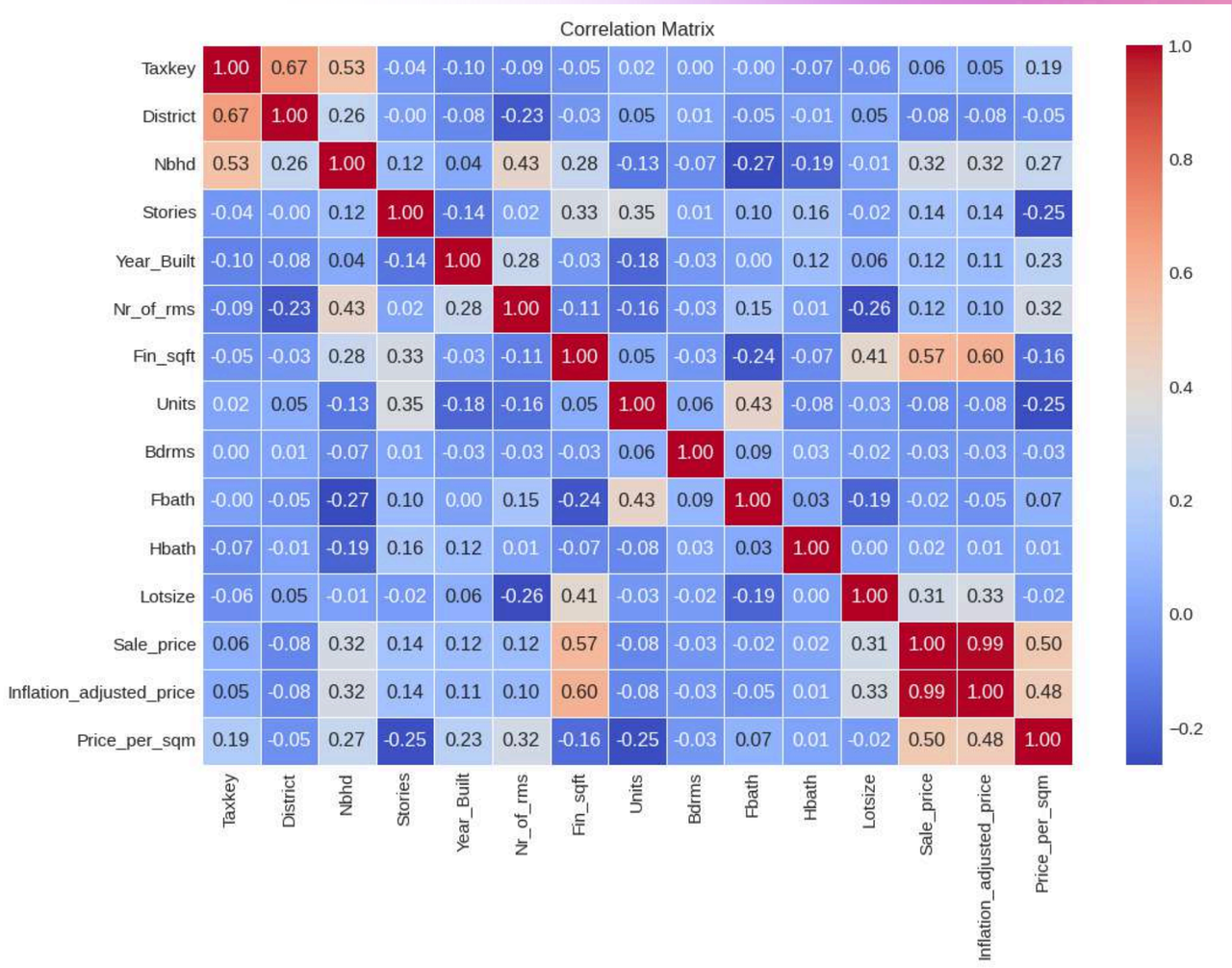
Effect Size:

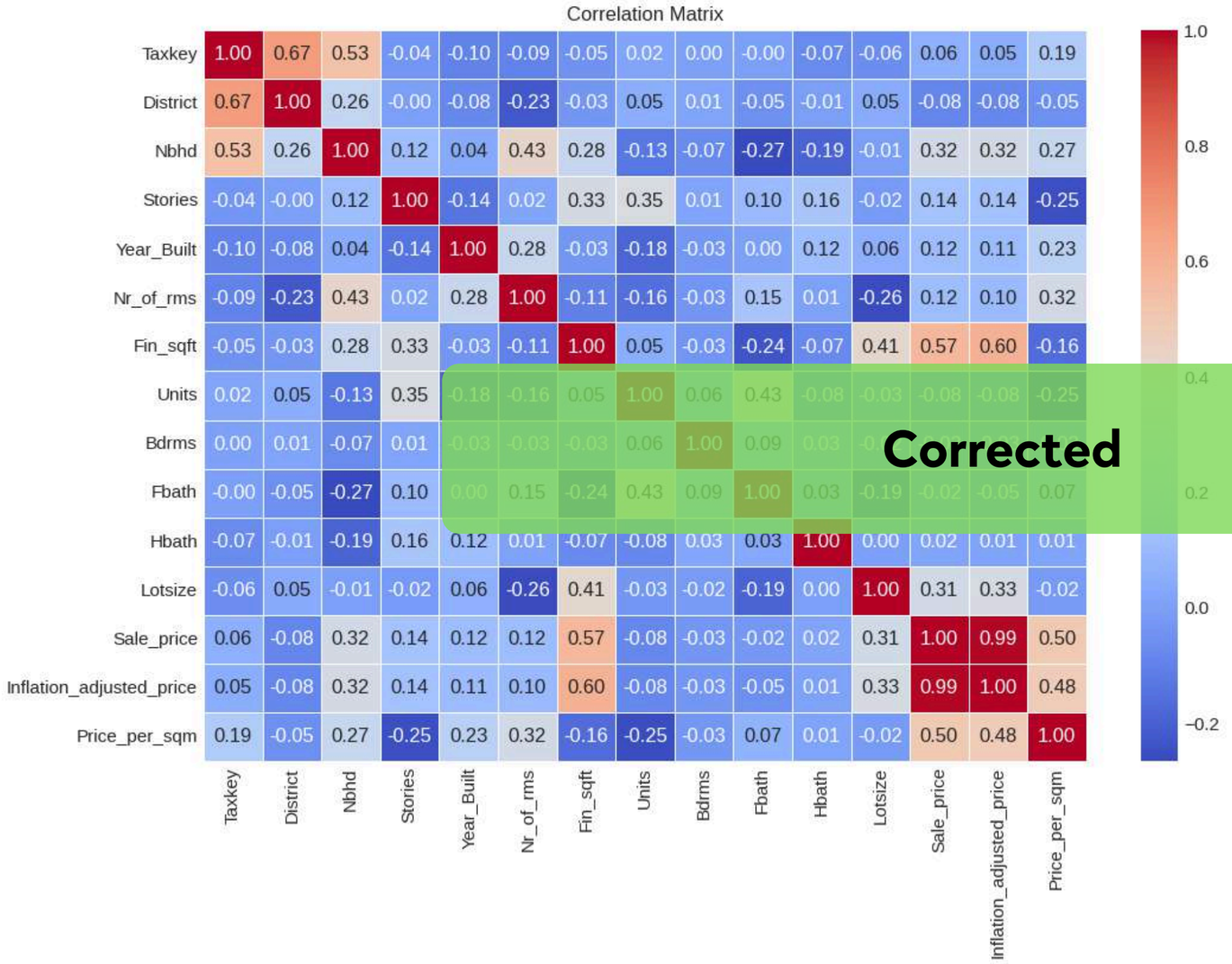
Pairwise Cohen’s d

Material 1	Material 2	Effect Size (Cohen's d)
Frame	Stone	-0.32
Frame	Brick	-0.40
Frame	Aluminum / Vinyl	-0.26
Frame	Stucco	-0.33
Frame	Block	-0.31
Frame	Masonry / Frame	-0.10
Frame	Prem Wood	-0.75
Frame	Fiber-Cement	-0.82
Stone	Brick	-0.08
Stone	Aluminum / Vinyl	0.08
Stone	Stucco	-0.04
Stone	Block	-0.02
Stone	Masonry / Frame	0.27
Stone	Prem Wood	-0.64
Stone	Fiber-Cement	-0.72
Brick	Aluminum / Vinyl	0.16
Brick	Stucco	0.04
Brick	Block	0.06
Brick	Masonry / Frame	0.31
Brick	Prem Wood	-0.44
Brick	Fiber-Cement	-0.52
Aluminum / Vinyl	Stucco	-0.12
Aluminum / Vinyl	Block	-0.10
Aluminum / Vinyl	Masonry / Frame	0.16
Aluminum / Vinyl	Prem Wood	-0.61
Aluminum / Vinyl	Fiber-Cement	-0.69
Stucco	Block	0.02
Stucco	Masonry / Frame	0.28
Stucco	Prem Wood	-0.45
Stucco	Fiber-Cement	-0.53
Block	Masonry / Frame	0.26
Block	Prem Wood	-0.45
Block	Fiber-Cement	-0.54
Masonry / Frame	Prem Wood	-0.83
Masonry / Frame	Fiber-Cement	-0.91
Prem Wood	Fiber-Cement	-0.09

Our mistakes

No correlation matrix analysis -1pt;



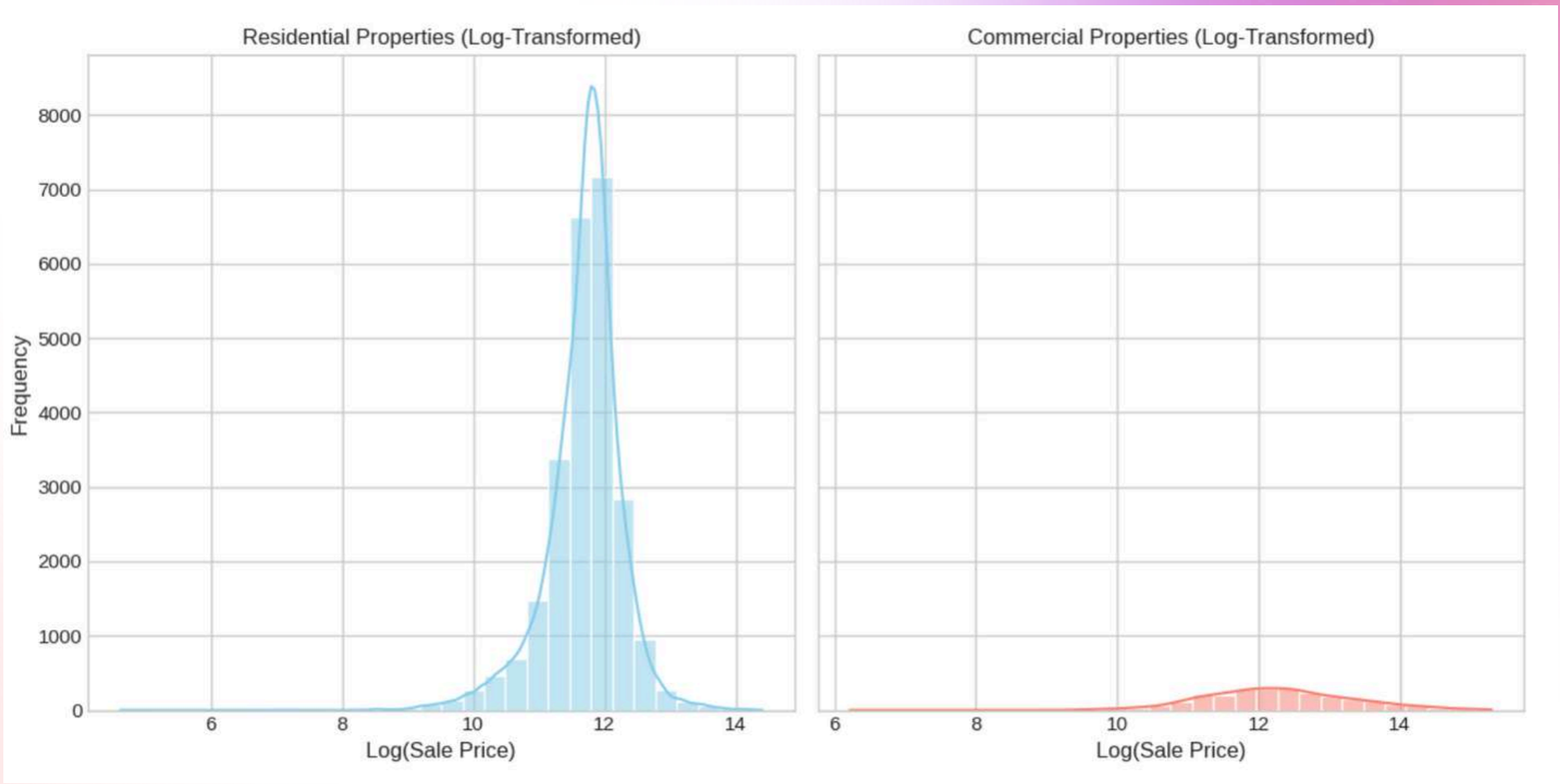


Corrected

- Property tax is closely correlated with the location of the property.
- The total finished square footage of a property strongly correlates with its sale price.

Our mistakes

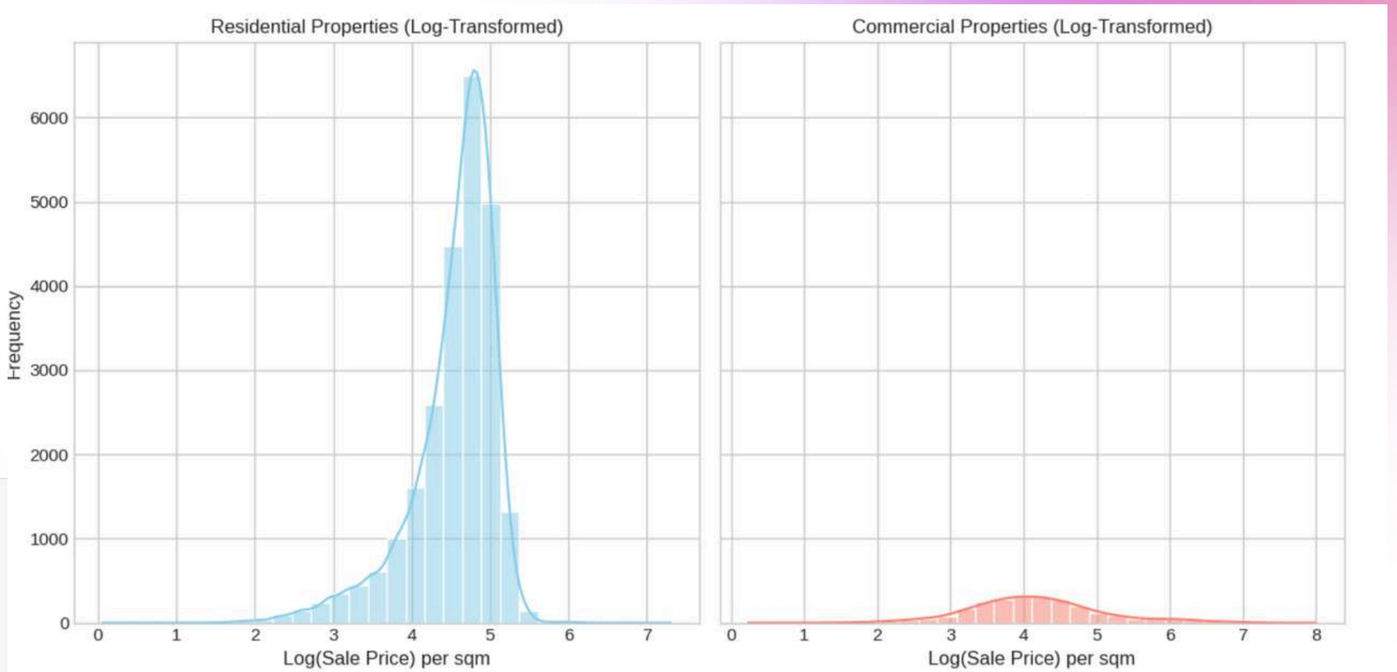
KS test by default compares with a standard normal distributioin but the mean of distribution of log prices is closer to 12 (and variance is likely different from 1) -1pt




```
# perform Kolmogorov-Smirnov test
_, p_value = kstest(crisis_residential_prices['Sale_price'], 'norm')
p_value
```

Our mistakes

Corrected




```python
from scipy.stats import kstest

# Calculate mean and standart deviation
mean = np.mean(log_residential_prices)
std = np.std(log_residential_prices)
print(mean, std)

# perform Kolmogorov-Smirnov test
_, p_value = kstest(log_residential_prices, 'norm', args=(mean, std))

if p_value < 0.05 or p_value is None:
  print("p-value is equal to {}, so we can reject the null hypothesis".format(p_value))
else:
  print("p-value is equal to {}, so we cannot reject the null hypothesis".format(p_value))
```

```
4.536816999498866 0.5516978394577793
p-value is equal to 2.610505e-318, so we can reject the null hypothesis
```

Our mistakes

2002

100 000 $

2018

110 000 $

No correction for inflation -1pt;

Our mistakes

```python
# Adjust prices by taking into account inflation 2002-2018 (https://www.officialdata.org/us/inflation/2002?endYear=2018&amount=100)
inflations = [1.58, 2.28, 2.66, 3.39, 3.23, 2.85, 3.84, -0.36, 1.64, 3.16, 2.07, 1.46, 1.62, 0.12, 1.26, 2.13, 2.49]
def calculate_adjustment(row):
    year = row['Sale_date'].year
    inflation_rates = inflations[year - 2002:2019 - 2002]  # Adjust for the range from the sale year to 2018
    price = copy.copy(row['Sale_price'])
    for inf_rate in inflation_rates:
        price = price * (1 + inf_rate / 100)
    return price

# Apply the function to each row to calculate the inflation-adjusted price
df['Inflation_adjusted_price'] = df.apply(calculate_adjustment, axis=1)
```

2002

140 000 $

Corrected

2018

110 000 $

Our mistakes

250 sqm

140 000 $

200 sqm

110 000 $

Differences in prices might be caused by the area of property -4pt;

Our mistakes

250 sqm

140 000 $

Price per sqm: 560 $

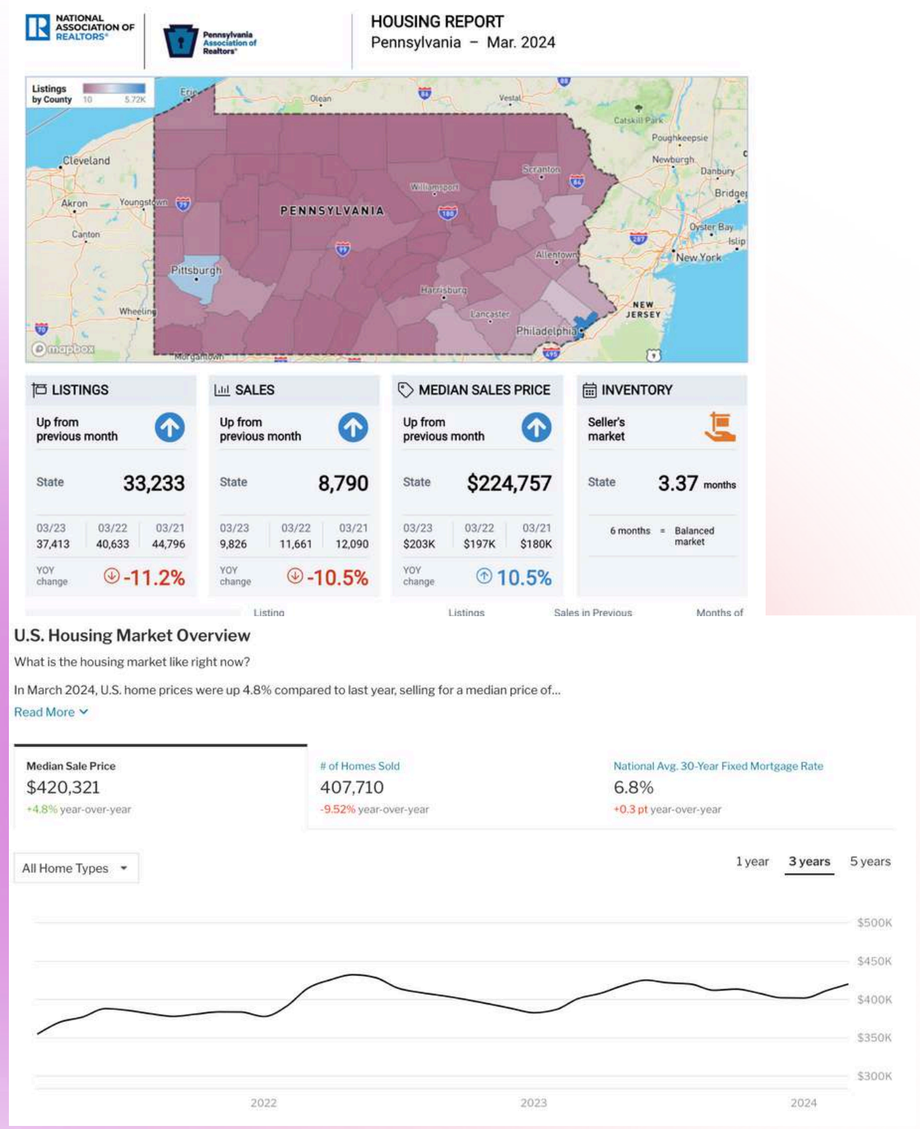
200 sqm

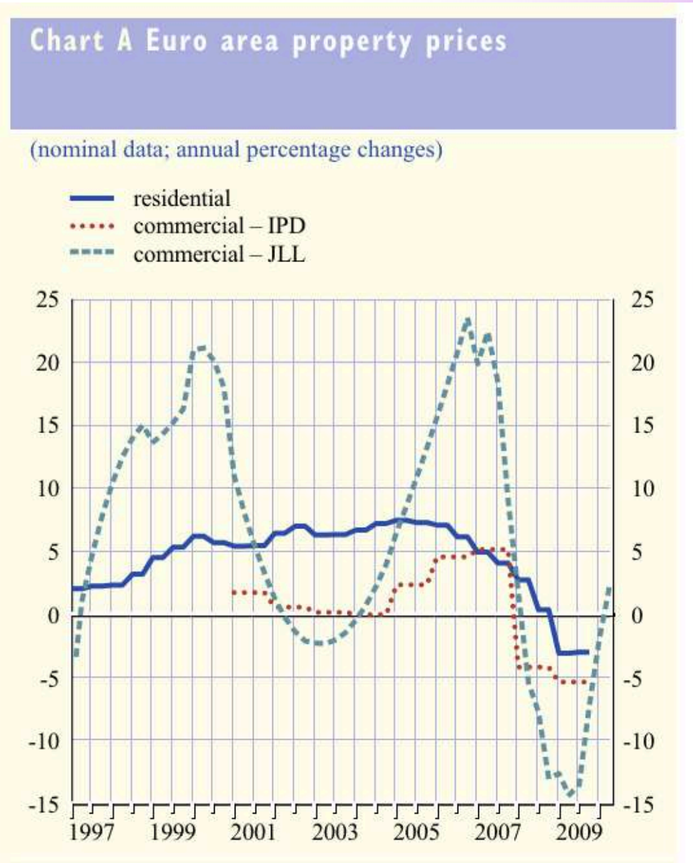
110 000 $

Price per sqm: 550 $

Corrected

Our mistakes

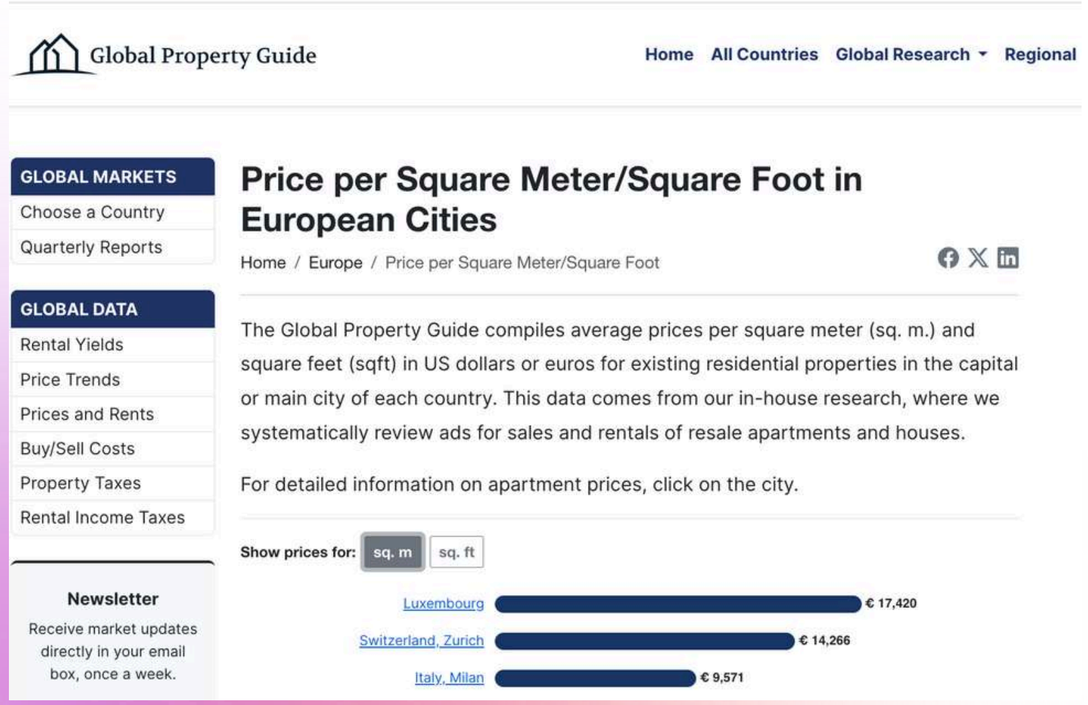


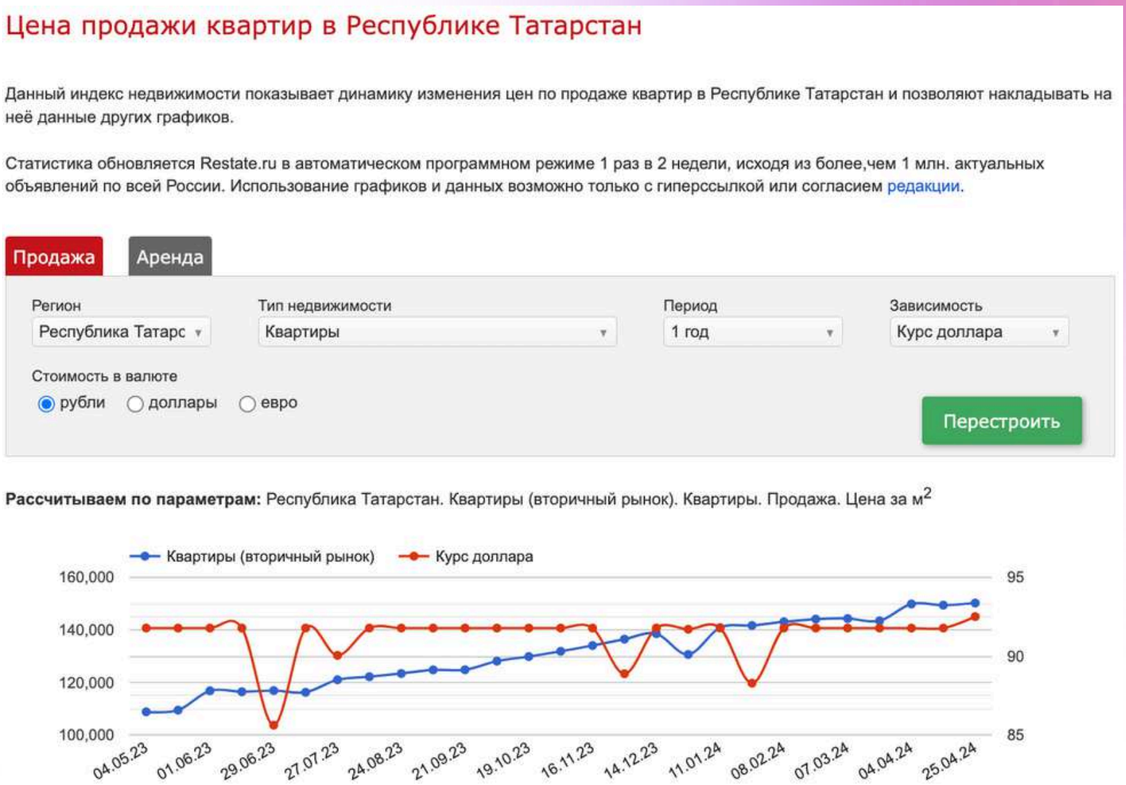


EUROPEAN CENTRAL BANK:

A COMPARISON OF TRENDS IN EURO AREA COMMERCIAL AND RESIDENTIAL PROPERTY PRICES

Our mistakes





Not clear

Thank you for your attention. Wishing you all the wealth to never have to ponder over choosing commercial or residencial property, but to simply buy everything without a second thought 🙏🙏🙏