

Dissecting the Drivers of Housing Prices in Egypt: A Statistical Investigation of Property Features and Location Effects

Yahia Abdelaziz, Seif Waleed, Magdy Omar

Abstract:

This study presents a comprehensive statistical investigation into the factors that influence housing prices across key urban markets in Egypt. Leveraging a dataset of approximately 27,000 property listings, the research explores the relative importance of intrinsic property features such as bedrooms, bathrooms, and area versus extrinsic locational variables, including city and compound-specific identifiers. Through a combination of ANOVA, Kruskal-Wallis tests, regression modeling, pairwise comparisons, and bootstrapping, the analysis identifies significant intra-city and inter-city pricing variations. The findings reveal that while physical features contribute to price formation, location factors both geographic and developer-related play a dominant role in determining market value.

1. Introduction

Housing prices are influenced by both physical features and location. While attributes like bedrooms, bathrooms, and area play a key role, city-level and developer-specific factors can significantly shift market value. In Egypt's diverse real estate landscape, this study analyzes how these variables compare in shaping price outcomes across major cities.

1.1 Objective

The objective of this study is to evaluate the relative impact of property features and location on housing prices in Egypt. It aims to determine whether physical attributes or geographic and developer-based factors serve as stronger predictors of price variation across and within major cities.

2. Methodology

2.1 Data Collection

The dataset comprises approximately 27,000 housing listings collected from public shared datasets in Egypt. Each entry includes details on property price, number of bedrooms, number of bathrooms, area in square meters, city, and compound name. Listings span both compound-based and non-compound properties across multiple urban regions.

2.2 Variables and measures

The analysis focuses on the following variables

I. Dependent Variable:

- A. Price (in Egyptian Pounds)

II. Independent Variables:

- A. Bedrooms (numeric)
Bathrooms (numeric)
- B. Area (square meters)
- C. City (categorical)
- D. Compound (categorical; includes "not in compound")

3. Data Processing

1. Standardized city and compound names.
2. Removed duplicate listings.
3. Encoded categorical variables.
4. Rounded area values and binned prices.
5. Separated "not in compound" listings as a distinct group.
6. Filtered missing and extreme values.

4. Hypothesis Question 1: Intra-City Analysis

This part examines whether prices vary between compounds within the same city when property features remain constant. Matched listings with identical features were compared across multiple compounds to isolate the effect of location.

4.1 Hypothesis Formulation

Null Hypothesis (H_0): Compound location does not significantly affect prices within a city when features are constant.

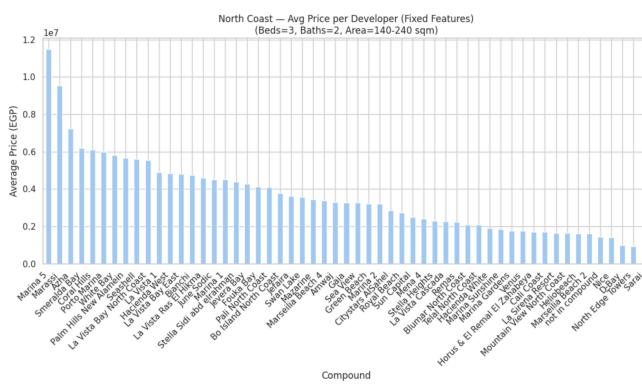
Alternative Hypothesis (H_1): Compound location significantly affects prices within a city when features are constant.

4.2.1 Test 1: Price Variation Across Developers in North Coast

A subset of 499 listings in North Coast, all with the same property specifications (3 bedrooms, 2 bathrooms, and an area between 140–240 sqm), was analyzed to assess whether average prices differ across compounds. The dataset included 55 developers, each with at least three listings, ensuring statistical reliability. Two statistical tests were applied:

- **ANOVA:** F-statistic = **11.8373**, p-value < **0.001**
- **Kruskal-Wallis:** Test statistic = **241.9820**, p-value < **0.001**

Both tests indicate a statistically significant difference in both mean and median prices across developers, even though the structural features were held constant. This confirms that location, specifically the compound—has a strong influence on pricing in the North Coast.

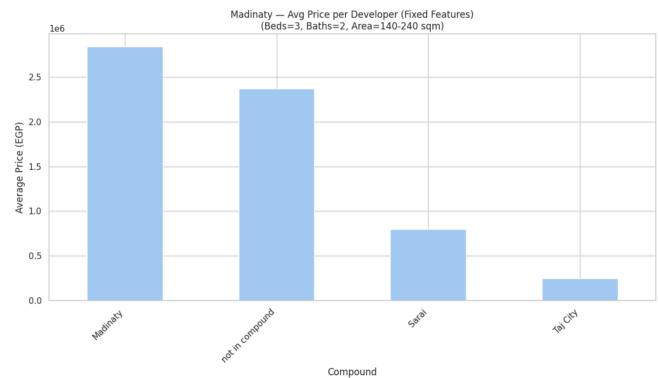


The figure 1 Above illustrates these differences, showing average prices per developer. Some compounds (e.g., **Marina 5**, **Marassi**, **Smeralda Bay**) consistently command higher prices, while others—such as **Nice**, **Sarai**, and **North Edge Towers**—fall significantly below the mean. Listings marked as “**not in compound**” are positioned toward the lower end, highlighting the price premium associated with branded or gated developments.

4.2.2 Test 1: Price Variation Across Developers in Madinaty City

A total of 90 listings in Madinaty with identical features (3 bedrooms, 2 bathrooms, area between 140–240 sqm) were analyzed across four developer groups. Each group included a minimum of three listings, ensuring statistical validity.

- **ANOVA:** F-statistic = **5.9291**, p-value = **0.0010**
- **Kruskal-Wallis:** Test statistic = **12.5897**, p-value = **0.0056**



The chart Above, Figure 2, displays average prices per developer. **Madinaty** listings show the highest mean price, followed by **non-compound** listings. In contrast, compounds such as **Sarai** and **Taj City** fall notably below the area's average, reinforcing the role of brand and location prestige in price formation.

4.2.3 Test 1: Price Variation Across Developers in New Cairo - El tagamo

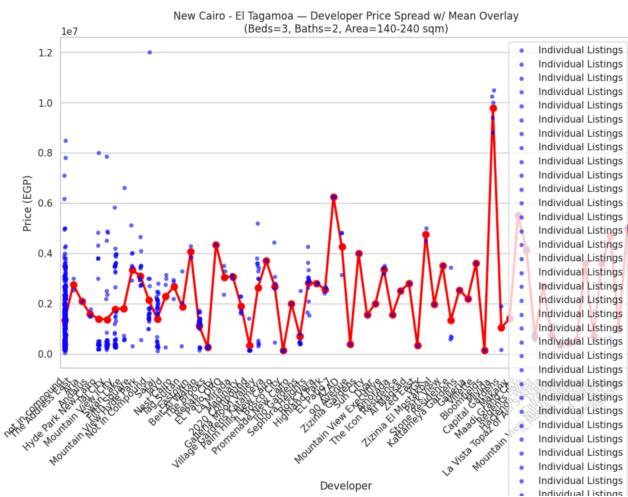
A total of 793 listings in New Cairo – El Tagamo with consistent property specifications (3 bedrooms, 2 bathrooms, and area between 140–240 sqm) were analyzed across 68 developer groups, each containing at least three listings.

- **ANOVA:** F-statistic = **6.6766**, p-value < **0.001**
- **Kruskal-Wallis:** Test statistic = **208.3086**, p-value < **0.001**

Both tests confirm highly significant price differences across developers, indicating that even with fixed property features, compound-specific location strongly influences pricing in this area.

Figure 3 below shows the distribution of individual listing prices (blue points) per developer, with the red line representing the average price for each.

Developers such as **ZED East**, **The Icon Residence**, and **La Vista Topaz** show substantial deviations from the general price level, while others—including **not in compound** listings—consistently appear in the lower range. The spread of values, especially within some developers, highlights market inconsistency and reinforces the weight of location branding and perceived value.

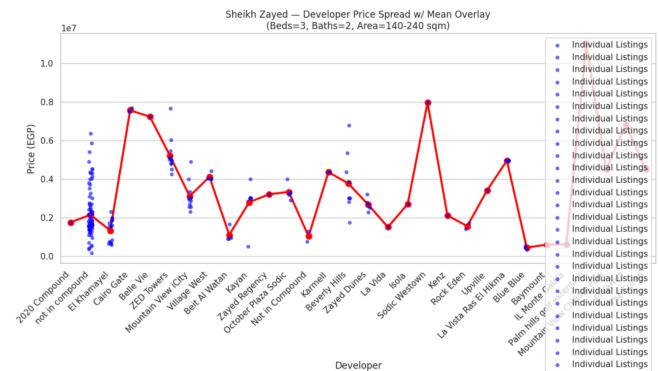


4.2.4 Test 1: Price Variation Across Developers in New Cairo - El tagamo

A total of 198 listings from Sheikh Zayed were selected, all sharing the same specifications of 3 bedrooms, 2 bathrooms, and an area between 140–240 sqm. Listings spanned 30 different developers, each represented by at least three entries.

- **ANOVA:** F-statistic = **14.4429**, p-value < **0.001**
- **Kruskal-Wallis:** Test statistic = **120.4305**, p-value < **0.001**

Figure 4 below displays the individual property listings (blue points) alongside the average price per developer (red line). Prices in compounds such as **Belle Vie**, **ZED Towers**, and **Sodic Westown** exceed 7 million EGP on average, while others like **Taj City**, **Baymount**, and **Blue Blue**—fall significantly below the overall trend. Properties categorized as “**not in compound**” also tend to cluster toward the lower end of the price spectrum.



The pronounced price disparity between developers confirms that location-specific branding and perceived value contribute substantially to pricing in Sheikh Zayed.

5. Hypothesis 2: Inter-City Analysis

5.1 Hypothesis Formulation

Null hypothesis (H_0): There is no significant price variation between cities when controlling for property features.

Alternative hypothesis (H_1): Significant price variations exist between cities, independent of property features.

5.2 Analysis Method

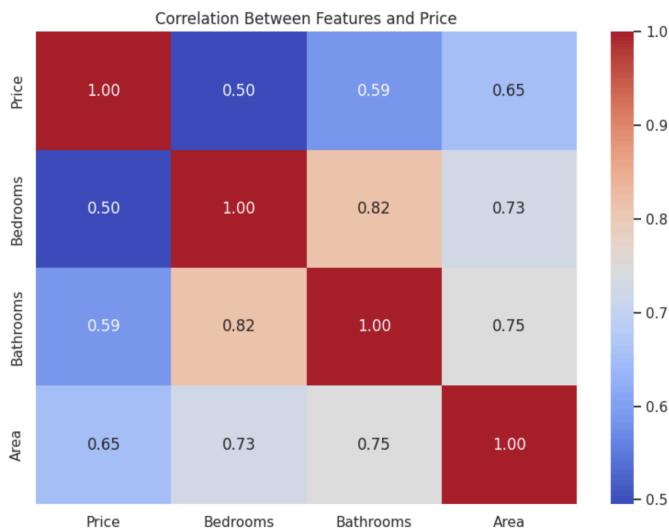
To test for significant price differences between cities while controlling for property features, several statistical methods were applied:

1. **Linear regression** was used to model the overall relationship between average prices and city location.
2. **Pearson correlation** assessed the relationship strength between price and key features
3. **Two-sample t-tests and z-tests** were used to compare average prices between city pairs.
4. **Confidence intervals** were calculated around group means to estimate the range within which true price differences are likely to fall.
5. **Pairwise Cohen's d** measured the effect size of price differences between cities.

5.3 Pearson Correlation Between Features and Price

Pearson correlations were computed between key property features and price using non-compound listings from major cities. Results showed consistent positive correlations, with **areas** having the strongest relationship to price, followed by **bathrooms** and then **bedrooms**.

The heatmap below (Figure 5) presents the overall correlation matrix, confirming that features do influence price, but none show a perfect or overwhelming correlation. These results reinforce that while structure matters, **location remains a key independent driver** of price differences.



Correlation Summary (Selected Cities):

Madinaty: Area (0.82), Bathrooms (0.72), Bedrooms (0.64)

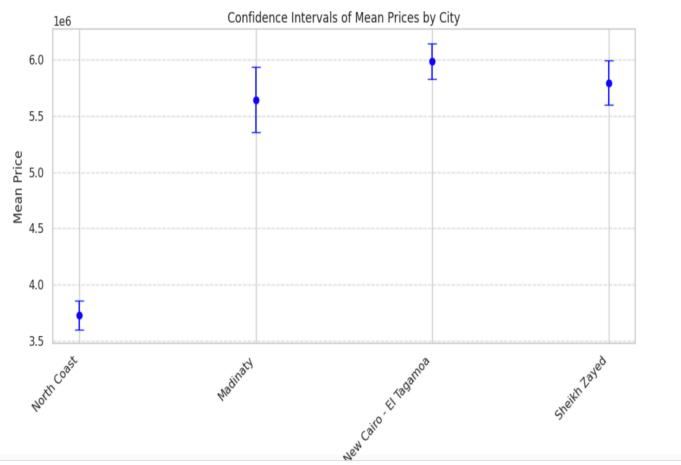
New Cairo – El Tagamoa: Area (0.73), Bathrooms (0.62), Bedrooms (0.53)

North Coast and **Sheikh Zayed** also showed moderate to strong correlations across all features.

Across all cities, area consistently shows the highest correlation with price, followed by bathrooms and bedrooms. Correlation strengths range from moderate to strong, with Madinaty showing the strongest overall relationships. However, despite these positive associations, the presence of substantial inter-city price variation as shown in regression and hypothesis tests—indicates that location remains a dominant factor independent of these features.

5.4 Confidence Intervals - Pairwise Price Comparison

Confidence intervals were calculated for the mean prices in four major cities using non-compound listings. As shown in the figure below (Figure 6), the intervals for **Madinaty**, **New Cairo – El Tagamoa**, and **Sheikh Zayed** overlap slightly, while the **North Coast** remains clearly lower and separate from the others. This suggests that North Coast's lower pricing is statistically distinct from the other cities.



T-Test Results:

North Coast vs Madinaty: $t = -11.71, p < 0.001$

North Coast vs New Cairo: $t = -21.68, p < 0.001$

North Coast vs Sheikh Zayed: $t = -17.23, p < 0.001$

Madinaty vs New Cairo: $t = -2.00, p = 0.0454$

Madinaty vs Sheikh Zayed: $t = -0.83, p = 0.4087$

New Cairo vs Sheikh Zayed: $t = 1.49, p = 0.1360$

These results show that the **North Coast** is significantly different from all three other cities. In contrast, **Madinaty**, **New Cairo – El Tagamoa**, and **Sheikh Zayed** are statistically similar in price when controlling for features, as their p-values exceed 0.05.

Together, the confidence intervals and t-tests confirm meaningful **inter-city price variation**, especially between North Coast and higher-priced urban centers.

5.5 Pairwise Z-Tests for Inter-City Price Comparison

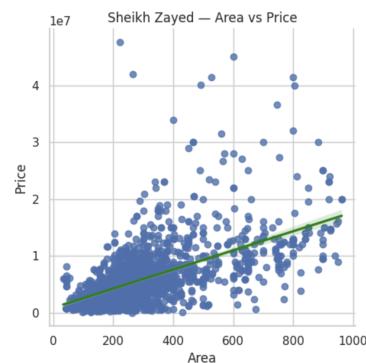
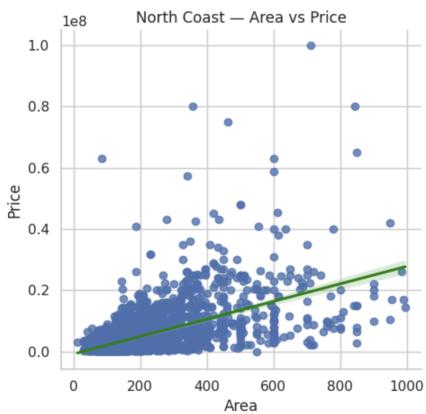
Pairwise z-tests were used to assess differences in mean housing prices between cities while controlling for property features. These tests are well-suited for large sample sizes, as they rely on approximated population variances and are more stable under high observation counts.

The z-test results closely mirrored those of the t-tests, confirming that most statistically significant differences particularly between North Coast and the other cities are robust to test selection. This alignment is expected in large datasets, where the margin of error is small and the two test types tend to converge.

One notable and consistent finding was that the North Coast exhibits significantly lower mean prices compared to urban areas like Madinaty, New Cairo - El Tagamo, and Sheikh Zayed. This can be attributed to the region's seasonal nature, where many listings represent vacation or secondary homes rather than primary residences. These properties often trade at different price dynamics due to less consistent demand, fewer year-round services, and reduced proximity to economic centers.

5.6 Linear Regression Analysis: Price Prediction by Features

Linear regression models were fitted for four major sub-cities to assess how well structural features (area, bedrooms, and bathrooms) explain variations in property price. The regression included only listings not in compounds to isolate the effect of location and structure without developer influence.



Key Results:

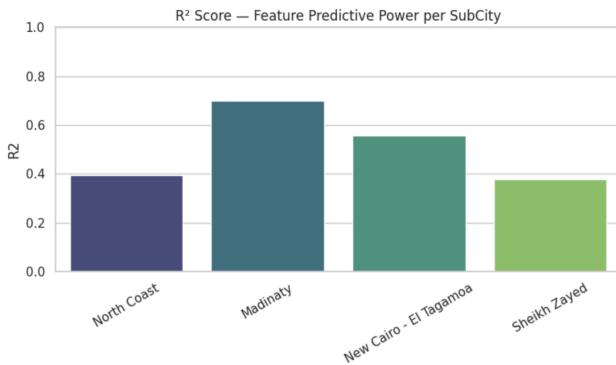
Madinaty showed the strongest linear relationship between features and price, with an **R² of 0.698**, indicating that nearly 70% of the variance in price can be explained by area, bedrooms, and bathrooms. **New Cairo – El Tagamo** followed with an **R² of 0.557**, showing a moderately strong fit. **North Coast** and **Sheikh Zayed** exhibited weaker fits, with **R² scores of 0.393** and **0.378**. This suggests that price variation in these cities is less predictable from features alone.

5.6.1 Interpretation of Coefficients:

- **Area** had a consistently strong positive influence across all cities, with the highest price-per-square-meter effect in **New Cairo – El Tagamoa** (EGP 24,028) and **North Coast** (EGP 22,794).
- **Bathrooms** also contributed positively in all cases, often with a stronger effect than bedrooms.
- Interestingly, **bedrooms showed a negative coefficient** in all models, likely due to multicollinearity or their weaker influence once area is accounted for.

The relatively low R^2 values in **North Coast** and **Sheikh Zayed** suggest that factors beyond physical features such as proximity to the beach, seasonal usage, or neighborhood desirability play a greater role in price formation. In contrast, more structured urban developments like **Madinaty** exhibit pricing that aligns more directly with measurable features.

The bar chart below (Figure 11) visualizes the **R^2 scores** of the linear regression models across the four sub-cities. These scores represent the proportion of price variance explained by structural features (area, bedrooms, and bathrooms):



6. Bootstrapping Analysis

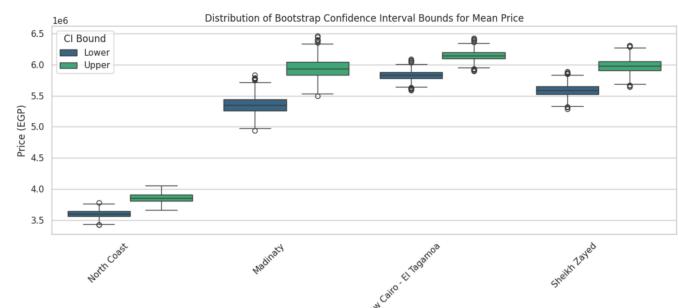
6.1 Methodology

Bootstrapping is a resampling technique employed to validate the stability and reliability of statistical results. It involves repeatedly drawing samples (with replacement) from the original data to create multiple simulated datasets.

original data to create multiple simulated datasets. In this study, bootstrapping was conducted with 1,000 iterations to estimate the variability and robustness of the z-tests, confidence intervals, and Pearson correlations. The mean and standard deviation of each statistic across bootstrap samples were calculated to assess their consistency and reliability.

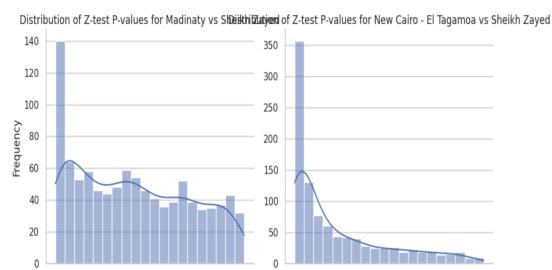
6.2 Bootstrap Analysis and Key Results

The boxplot below (Figure 12) presents the distribution of bootstrapped **confidence interval bounds** for the mean prices in each city.



North Coast stands out with its lower mean price range (~3.6M EGP), and tight CI distribution, reflecting **high confidence in its lower market tier**. The intervals for **Madinaty**, **New Cairo – El Tagamoa**, and **Sheikh Zayed** largely overlap, suggesting these cities operate within the same pricing zone. **Sheikh Zayed** displays slightly more variation in CI width, reflecting marginally more price dispersion across resamples.

The **right-skewed distributions** in (Figure 13) below show high variability and a wide spread of p-values. While some resamples approach significance, most values remain above the 0.05 threshold, confirming that the observed differences between these city pairs are **not statistically robust**. This aligns with earlier t- and z-test results indicating **no consistent price gap** between these high-demand urban areas.



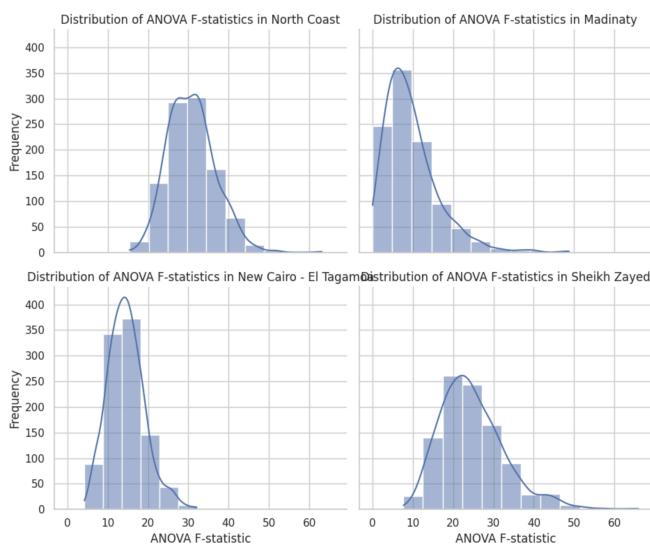
6.3 Interpretation of Results

The bootstrap analysis reinforces prior findings, confirming that location significantly influences housing prices beyond structural features. North Coast emerged consistently distinct from other cities, with stable and statistically robust lower pricing. Conversely, the overlapping confidence intervals and higher variability in z-test p-values among Madinaty, New Cairo – El Tagamoa, and Sheikh Zayed demonstrate weaker and less certain location-driven price differences.

Pearson correlations highlighted that property area consistently holds strong explanatory power across cities. These findings confirm the importance of both structural features and location attributes in price determination, with location playing a particularly strong differentiating role between coastal and inland urban centers.

6.4 Bootstrap ANOVA Testing by Developer Across Cities

To further validate intra-city pricing variability, **bootstrap based ANOVA tests** were applied across 1,000 resampled datasets in four major sub-cities. The tests were limited to properties with **fixed features** (3 bedrooms, 2 bathrooms, and 140–240 sqm area) to isolate the impact of developer and compound-level location on price.



The histograms above show the distribution of **F-statistics** across the 1,000 bootstrap samples for each city:

1. **North Coast** and **Sheikh Zayed** display broad, symmetrical distributions centered on high F-values, indicating **strong and consistent price variation across compounds**.
2. **New Cairo – El Tagamoa** exhibits a tighter distribution, further reinforcing the **reliability** of its developer-driven pricing variance.
3. **Madinaty**, however, has a right-skewed and compressed F-distribution, reflecting **less pronounced and more volatile differences** among developers.

6.4.1 Interpretation

Across all cities, the results affirm that **developer-level location effects significantly influence price**, even when physical property characteristics are held constant. This effect is most stable and pronounced in North Coast, Sheikh Zayed, and New Cairo – El Tagamoa, while Madinaty's results suggest a more uniform developer pricing strategy, potentially driven by centralized planning or regulation.

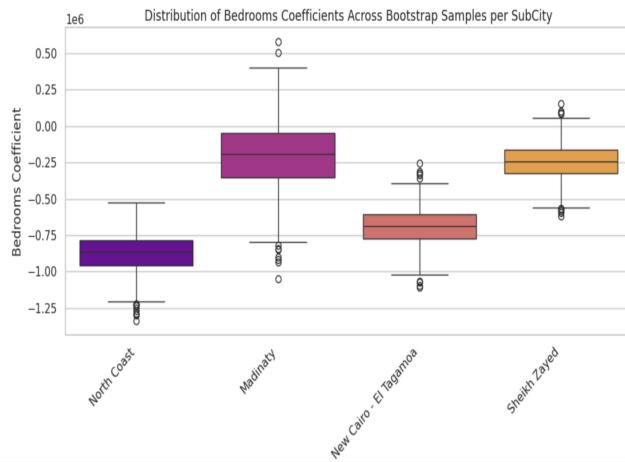
These findings extend the conclusions of Hypothesis 1 by demonstrating that **location-based pricing is not only city-driven but also strongly shaped by micro-location (compound/developer) factors**.

6.5 Bootstrap Linear Regression Analysis by Sub-City

To reinforce the reliability of linear regression findings, a **bootstrap regression analysis** was conducted across 1,000 resampled datasets for each sub-city. The models evaluated how well the area, **number of bedrooms, and number of bathrooms** predicted price. This approach helps assess the **stability** of regression outputs, such as R² scores and feature coefficients.

R² Scores – Predictive Power:

- **Madinaty**: 0.699 ± 0.019 → Highest and most consistent predictive strength.
- **New Cairo – El Tagamoa**: 0.558 ± 0.010 → Moderate predictive power.
- **North Coast**: 0.394 ± 0.019 → Weak feature-driven predictability.
- **Sheikh Zayed**: 0.380 ± 0.024 → Weakest overall predictive fit.



This boxplot illustrates the **distribution of area coefficients** across bootstrap samples:

- **New Cairo – El Tagamoa** consistently assigned the **highest price value per sqm**, reinforcing its premium urban character.
- **North Coast** displayed **wider variance**, reflecting its less predictable price structure—potentially influenced by seasonal demand, varying proximity to beaches, or lifestyle-driven fluctuations.
- **Sheikh Zayed** had the **lowest and tightest range** of area coefficients, suggesting that price is less dependent on size and more influenced by location-specific amenities or neighborhood quality.
- **Madinaty** maintained moderate values with **strong consistency**, in line with its urban planning and pricing discipline.

6.5.1 Interpretation

The bootstrap regression results strongly align with prior findings, which is that **Structural features explain prices more effectively in planned urban centers** like Madinaty and New Cairo – El Tagamoa. **Area** remains the most influential feature, while **bedrooms consistently exhibit a negative coefficient**, likely due to their overlap with area and weaker individual contribution once other variables are included. The low R^2 in **North Coast and Sheikh Zayed** again underscores the role of

non-structural factors, such as lifestyle appeal, seasonal pricing, or developer branding, in shaping property values..

These robust and consistent trends reinforce the broader conclusion that **while features matter, location-specific characteristics (both macro and micro)** significantly shape housing price behavior.

7. Discussion

The intra-city analysis revealed that **developer-level location differences significantly affect housing prices**, even when property features are held constant. ANOVA tests across cities like New Cairo, Sheikh Zayed, and the North Coast consistently confirmed strong price variability among developers, while Madinaty showed a more regulated pricing pattern. These results demonstrate that **micro-location (compound or developer)** can be as important—if not more important—than the number of bedrooms or square footage.

In the inter-city analysis, **city-level location effects remained highly significant**. Pairwise t-tests and z-tests confirmed strong and consistent price differences between cities such as the North Coast and New Cairo, even when controlling for structural characteristics. However, comparisons among cities like Madinaty, New Cairo, and Sheikh Zayed were less stable, indicating **a pricing convergence among premium urban zones**.

Bootstrapping added statistical robustness to these insights. Confidence intervals, z-test distributions, and repeated regressions confirmed that the core findings were stable across resampled subsets, with low standard deviations and consistent coefficient patterns. Additionally, F-statistic distributions across 1,000 ANOVA tests underscored the strength of compound-based pricing differentiation, especially in Sheikh Zayed and the North Coast.

8. Conclusion

The findings of this study support the conclusion that location exerts a stronger and more variable influence on housing prices than property features alone. While features like area and bathroom count contribute meaningfully to price determination,

they cannot fully explain pricing behavior across Egypt's real estate landscape. Instead, macro-location (city) and micro-location (developer/compound) effects dominate, especially in less regulated or lifestyle-driven markets such as the North Coast.

These results highlight the need for stakeholders—buyers, developers, and policymakers—to consider both structural attributes and locational context in pricing strategy, investment decisions, and market planning. Future work could expand this analysis by incorporating temporal trends, rental markets, or predictive models to further unravel the complex dynamics of real estate pricing in Egypt.