**Business Problem: Real Estate Evaluation in Miami**

This model estimates the value for residential real estate in the city of Miami. It is a major benefit to both buyers and sellers of real estate. First if you are a real estate agent or a home seller, you can use this model to help price your listing. This model will also help home buyers evaluate whether a home is overvalued and should not purchase or if a house is significantly undervalued and make an offer.

**Data Source and Preprocessing Steps**

* **Data Source**: [Miami Housing Dataset](https://www.kaggle.com/datasets/deepcontractor/miami-housing-dataset) - from Kaggle, contains 13,932 rows, 17 columns
* Checked for missing values - 0 values missing
* Target: Sales Price, written as SALE\_PRC
* Features: Everything besides SALE\_PRC
* Created Visualizations:
  + Bar Chart & Box Plot - Displaying Distribution of Housing Prices
  + Correlation Matrix - to evaluate features for multicollinearity & feature selection
  + Scatter Plots - to show relationships between top features and Sales Price
* Train/Test Split: 80/20 split,
* Performed Feature Scaling

**Model Selection Process & Results**

Used Linear Regression and Random Forrest Model. Random Forrest Model performed better on every metric but r-squared especially, 89% on Testing data vs. 69% for linear regression.

**Key Insights and Recommendations**

**Size Matters** - Square Footage/Land Area has a direct relationship with Sale Price

**LOCATION! LOCATION! LOCATION!** - this famous phrase is often used for justifying the price of real estate and is proven to be true as location does have a strong correlation to the sales price of a house, a home with similar features tends to always be more expensive based on its distance to the city center.

**Limitations**

This data covers a static period & is not updated on recent trends in the market. For example, lower/higher interest rates will affect the price of a home. Home values can go up or down significantly in just a year based on external macro-economic factors.

**Future Improvements**

The data was trained on just 2 models, due to Random Forest being so accurate, no other regression models were used.