

# Global Housing Market Analysis (2015–2024)

## Leveraging Data Science to Decode Housing Market Trends

### Dataset & Methodology

Data Sources Primary Dataset: global\_housing\_market\_extended.csv (200+ rows, 11 features).

Variables:

Target: House Price Index

Predictors: GDP Growth, Mortgage Rates, Inflation, Urbanization, etc.

### Tech Stack

```
# Core Libraries
import pandas as pd # Data manipulation
import seaborn as sns # Visualization
from sklearn.ensemble import RandomForestRegressor # Predictive modeling
```

### Key Findings

```
data = {
    "": ["House Price Index", "GDP Growth", "Urbanization Rate"],
    "House Price Index": [1.00, 0.82, -0.76],
    "GDP Growth": [0.82, 1.00, -0.45],
    "Urbanization Rate": [-0.76, -0.45, 1.00]
}

# Membuat DataFrame
df_corr = pd.DataFrame(data)
df_corr.set_index("", inplace=True)

# Menampilkan tabel korelasi
print(df_corr)
```

	House Price Index	GDP Growth	Urbanization Rate
House Price Index	1.00	0.82	-0.76
GDP Growth	0.82	1.00	-0.45
Urbanization Rate	-0.76	-0.45	1.00

# Machine Learning Insights

## Technical Deep Dive

Data Preprocessing Handled missing data (<1% values) via median imputation.

Removed outliers using IQR (e.g., Construction Index >150 dropped).

Encoded categorical variables (Country → One-Hot).

Feature Engineering Created compound indices (e.g., Affordability Stress Score = Rent Index / GDP per Capita).

Normalized features using StandardScaler.

Model Validation Train-Test Split: 80-20 stratified by country.

Hyperparameter Tuning: GridSearchCV optimized n\_estimators=200.

## □ How to Reproduce

### Clone Repository

```
git clone https://github.com/yourusername/housing-market-analysis.git cd housing-market-analysis
```

### Install Dependencies

```
pip install -r requirements.txt # pandas, scikit-learn, seaborn
```

### Run Analysis

```
jupyter notebook Global_Housing_Analysis.ipynb
```

```
jupyter notebook Global_Housing_Analysis.ipynb
```

## □ Business Implications

Investors: Target countries with GDP growth >3% and urbanization <75% for optimal ROI.

Policymakers: Cap mortgage rates at 5% to stabilize affordability in high-risk regions.

## □ Future Enhancements

Geospatial Analysis: Map price hotspots using geopandas.

Live Data Pipeline: Automate updates via APIs (Zillow, World Bank).

Deep Learning: LSTM for price forecasting at city-level granularity.

## □ Contact

Haeril Wahyudi

LinkedIn: <https://www.linkedin.com/in/haeril-wahyudi-b89695124/>

Email: [haerilwahyudi12@gmail.com](mailto:haerilwahyudi12@gmail.com)

Repository: [https://haerilwahyudi.github.io/Porto\\_me/](https://haerilwahyudi.github.io/Porto_me/)

## □ Why This Stands Out

Result-Driven: Quantified economic impacts with statistical rigor.

Actionable Insights: Direct applications for investors/policymakers.

Reproducibility: Clear steps to replicate analysis.