Real Estate Pricing - EDA and Predictive Modeling Report

Objective

To explore housing data and uncover factors that significantly influence home prices using Exploratory Data Analysis (EDA) and Machine Learning models.

Exploratory Data Analysis Highlights

Top Correlated Features with Sale Price:

- **OverallQual** (Quality of material and finish): 0.79
- **GrLivArea** (Above ground living area): 0.71
- **GarageCars** (Garage car capacity): 0.64
- **GarageArea**, **TotalBsmtSF**, **1stFlrSF**: ~0.60
- **Engineered Features** like `PricePerSqft` and `PropertyAge` show meaningful correlation.

Size Impact:

- Homes with more **bedrooms**, **bathrooms**, and **living area** tend to be priced higher.
- **Newer properties** or those with **larger garages** also command higher prices.

Market Trends:

- An upward trend is observed in sale prices from 2006 to 2009.
- Seasonal fluctuations and annual shifts suggest market volatility.

Machine Learning Models

Models Trained:

1. **Linear Regression**
2. **Random Forest Regressor**

Performance Metrics:

| Model | RMSE | R^2 Score | |---|---|---| | Linear Regression | 29,573 | 0.8957 | | Random Forest Regressor | 23,462 | 0.9344 |

- **Random Forest** performs better with **lower RMSE** and **higher R^2**, making it the preferred model.

Recommendations

- Focus marketing and pricing strategy around properties with **high quality (OverallQual)**, **larger area (GrLivArea)**, and **good garage space**.
- Maintain detailed records of **construction year** and **lot attributes** to better understand property depreciation.
- Leverage advanced models like **Random Forest** for future price predictions.

Next Steps

- Improve model accuracy using additional engineered features and hyperparameter tuning.
- Apply clustering techniques to segment customer preferences and optimize amenity pricing.
- Deploy predictive model into a real-time pricing recommendation system.