**Trends in Real Estate Market –USA 🏡**

**1. Introduction**

Understanding what drives property values is essential in real estate. This analysis leverages a housing dataset with rich property-level features to explore how various factors relate to sale prices. Through systematic exploration and statistical assessment using Python, aim to generate insights into the housing market that benefit stakeholders including homebuyers, sellers, and real estate investors.

**2. Objective**

The core objective of this study is to investigate how home characteristics—such as living area, quality ratings, basement size, and neighborhood—affect house prices. We employ descriptive statistics, data visualization, and basic inferential techniques to uncover underlying trends and patterns.

**3. Problem Context**

Accurately estimating home values based on their attributes is challenging. Real estate professionals must understand how individual factors, like house condition, space, and location, influence market prices. This project addresses the problem of isolating and quantifying the effect of these key variables on sale prices, providing clarity to a typically complex process.

**4. Methodology**

The analysis follows a structured approach:

* **Data Import & Inspection:** The dataset was loaded and previewed for structure.
* **Cleaning:** Missing values, incorrect data types, and outliers were treated.
* **Feature Engineering:** New features such as total square footage and house age were created.
* **EDA:** Patterns and relationships were visualized through histograms, boxplots, scatterplots, and heatmaps.
* **Insights:** Descriptive insights were drawn and supported with summary statistics.

**5. Data Overview**

The dataset contains variables describing home features including size, quality, construction type, and sale price. It consists of both numeric fields (e.g., Ground Living Area, Sale Price) and categorical fields (e.g., Neighborhood, House Style). Preliminary review showed notable variance in sale price and strong patterns in some features.

**6. Data Preparation**

* **Handling Nulls:** Imputed numerical missing values with median; categorical with mode or 'None'.
* **Data Type Correction:** Converted relevant columns to appropriate data types (e.g., categories).
* **Outlier Handling:** Used IQR technique to filter out extreme price and area values.
* **Normalization:** Skewed features such as Sale Price and Ground Living Area were log-transformed for visualization.

**7. Feature Engineering**

To enhance interpretability:

* Total.Sf was created as the sum of basement and floor areas.
* House Age was computed using the year sold and the year built.
* Simplified quality scores into categories (e.g., Low, Medium, High) for group comparison.

**8. Focused Data Slices**

Data was filtered to investigate:

* Homes grouped by neighborhood
* Property quality segments
* Size-based categorization (Small, Medium, Large)  
  This helped isolate the effect of specific features on price variation.

**9. Descriptive Analysis**

* **Central Tendencies:** Median and mean provided for key variables.
* **Variation:** Standard deviations highlighted the spread in price and area.
* **Distribution:** Sale Price was heavily right-skewed, justifying transformation.

**10. EDA - Univariate Analysis**

* **Price:** Distribution showed a long right tail with most properties under the $300k range.
* **Lot Area and Living Area:** Both showed skewness and required log transformation.
* **Quality Scores:** Majority of homes scored medium-to-high on quality metrics.

**11. Bivariate Insights**

* **SalePrice vs Ground Living Area:** Strong linear relationship visible in scatterplot.
* **SalePrice vs OverallQual:** Boxplot showed rising median prices with higher quality ratings.
* **SalePrice by Neighborhood:** Significant variation observed across different localities.

**12. Multivariate Relationships**

* A correlation matrix revealed:
  + High positive correlation of SalePrice with OverallQual, Ground Living Area, GarageCars
  + Strong multicollinearity among size-related variables
* Pair plots were used to visualize interactions between top variables

**13. Major Findings**

* **Overall Quality:** Strongest predictor of price; better quality consistently leads to higher prices.
* **Size Metrics:** More square footage (both above and below ground) directly correlates with higher values.
* **Garage and Basement:** Homes with more garage spaces and finished basements fetched higher prices.
* **Location:** Neighborhoods differ significantly in value, emphasizing the importance of location.

**14. Conclusion**

The housing analysis project successfully revealed key trends and statistically validated relationships in the real estate dataset. Significant variables like living area, renovation status, and waterfront view were shown to affect property prices. These insights can support data-driven pricing strategies and better property investment decisions. Future work can extend this foundation to build predictive models for home pricing.