Understanding The Concept of Variance : 01

Problem Statement (As Defined By The Client)

Client Requirement

- The Client, A Real Estate Firm, Has Been Using A Prediction Model To Forecast House Prices Based on Features Such As The Number of Bedrooms And Size in Square Feet
- The Client Has Noticed A Discrepancy Between Actual House Prices And The Predicted Prices, Which Has Resulted in Inconsistent Pricing Strategies
- The Client Has Provided Historical Data, Including Actual House Prices And Predicted Prices For Different House Features
- The Client Wishes To Understand The Variance Between The Predicted Prices And Actual Prices, Identify The Extent of Variance, And Visualize it To Determine if The Prediction Model Needs Adjustments

Specific Goals

1. Identify And Visualize The Variance Trend Between Actual And Predicted Prices For Different House Properties
2. Quantify The Degree of Variance Using Appropriate Statistical Measures (e.g., Variance, Standard Deviation) And Graphical Representations
3. Suggest Improvements To Reduce Variance And Enhance The Prediction Model's Accuracy For Better Real Estate Pricing Strategies

Problem Statement (As Defined By The Data Scientist)

Title

Analyzing Variance in House Price Predictions Using Historical Data

Objective

- To Evaluate And Visualize The Variance Between Actual House Prices And The Predicted Prices Generated By The Existing Prediction Model
- The Project Aims To Understand The Distribution And Deviation of Predictions From Actual Values, Quantify The Variance, And Suggest Methods To Minimize Variance For Improved Accuracy

Problem Description

- The Client’s Existing Prediction Model Outputs House Price Predictions Based on Property Features Like The Number of Bedrooms And House Size
- The Predicted Prices Have Shown A Significant Variance When Compared To Actual Prices, Indicating Potential Issues With The Model
- There is A Need To Evaluate And Visualize The Variance To Understand The Model’s Performance And Identify if The Variance is Consistent OR Sporadic

Remarks on The Current Model

- The Variance is Represented By The Difference Between Actual Price And Predicted Price
- If The Variance is High, it Suggests That The Model is Not Accurately Capturing The Relationship Between Property Features And House Prices
- The Model’s Accuracy Needs To Be Visualized And Quantified Using Statistical And Graphical Methods To Clearly Identify The Presence And Extent of Variance

Understanding The Concept of Variance : 01

Approach

1. **Data Analysis**
   - Use The Provided House Price Data, Including The Number of Bedrooms, Size in Square Feet, Actual Price, And Predicted Price
   - Calculate The Differences And Squared Differences Between Actual And Predicted Prices To Evaluate The Variance
2. **Data Visualization**
   - Create Scatter Plots, Bar Charts, And Line Graphs To Visualize The Variance Between Actual And Predicted House Prices
   - Use Color Coding To Highlight Houses With The Highest Variance
3. **Variance Analysis**
   - Calculate The Variance And Standard Deviation To Measure The Spread of Prediction Errors
   - Visualize The Squared Differences Using A Separate Line Chart To Clearly Show How Predictions Deviate From Actual Prices
4. **Model Improvement Suggestions**
   - Provide Insights on Why The Current Model May Be Showing Variance (e.g., Lack of Feature Considerations, Non-Linear Relationships)
   - Suggest Modifications To The Existing Model, Such As Incorporating More Features OR Using Advanced Regression Models

Expected Outcome

- A Clear Visualization of Variance in The Prediction Model, Showing How Far The Predictions Deviate From Actual House Prices
- Quantification of Variance Through Statistical Measures Such As Variance And Standard Deviation
- Recommendations on How To Refine The Model To Reduce Variance And Achieve More Accurate House Price Predictions

Sample Data

House ID	Bedrooms	Size (sq ft)	Price ($)	Predicted Price
1	3	1,500.00	300,001.00	320,889.00
2	4	2,000.00	400,000.00	404,926.00
3	2	1,200.00	251,000.00	236,852.00
4	3	1,600.00	320,000.00	320,889.00
5	5	2,500.00	500,000.00	488,963.00
6	4	1,800.00	360,000.00	404,926.00
7	3	1,400.00	290,000.00	320,889.00
8	2	1,100.00	240,000.00	236,852.00
9	4	2,200.00	450,000.00	404,926.00
10	3	1,750.00	350,000.00	320,889.00