### **Updated Ideas for Current Status**

1. **Exploratory Data Analysis (EDA) - Deeper Insights:**
   * We’ve expanded the EDA to identify more nuanced patterns in the data, such as correlations between location-specific features (e.g., neighborhood and sale price) or temporal trends (e.g., year built vs. sale price).
   * Conducted **distribution checks** on key features like square footage, number of rooms, and house age to understand their effect on the target variable.
2. **Advanced Feature Engineering:**
   * We are currently developing **new engineered features** like price per square foot, house age, and interaction features (e.g., number of rooms \* square footage) to provide additional information to our models.
   * We are also **transforming skewed variables**, particularly **SalePrice**, using **log transformations** to stabilize variance and reduce skewness.
3. **Handling Missing Data and Imputation Strategies:**
   * Missing values are being handled using advanced imputation techniques like **KNN imputation** for numerical features and **frequency encoding** for categorical features, ensuring more reliable handling of missing data.
   * We’ve also explored multivariate imputation techniques like **MICE** (Multiple Imputation by Chained Equations) to enhance data integrity.
4. **Outlier Detection and Treatment:**
   * A comprehensive **outlier analysis** was conducted using boxplots and z-scores. We identified outliers in key variables like LotFrontage and GrLivArea and tested different strategies for handling them (e.g., capping or transformation).
5. **Model Building - Progress on Initial Models:**
   * After testing baseline models such as **Linear Regression**, **Ridge**, and **Lasso**, we are moving forward with more complex models, including **Gradient Boosting** and **XGBoost** to improve accuracy and reduce overfitting.
   * We have started tuning hyperparameters for these models using **RandomizedSearchCV** and **GridSearchCV** for better performance.
6. **Cross-validation and Robust Evaluation:**
   * To avoid overfitting and improve generalization, we’ve implemented **k-fold cross-validation** (k=5) for model validation, ensuring that the results are consistent across multiple subsets of the data.
   * Performance metrics like **RMSE**, **MAE**, and **R-squared** are being calculated to evaluate model performance comprehensively.
7. **Feature Selection and Dimensionality Reduction:**
   * We are incorporating **feature selection techniques** like **Recursive Feature Elimination (RFE)** to reduce the number of features without losing valuable predictive information.
   * Additionally, we’re testing **PCA (Principal Component Analysis)** to reduce dimensionality while maintaining model accuracy.
8. **Model Interpretability:**
   * We are investigating **model interpretability tools** such as **SHAP** (Shapley Additive Explanations) to understand the influence of individual features on the predicted house prices and ensure transparency.
9. **Ensemble Techniques and Model Stacking:**
   * We are testing **ensemble methods** like **Bagging** and **Stacking** to combine the predictions of multiple models, improving overall accuracy and robustness.
   * Initial experiments with **Stacking Classifiers** are showing promising results in reducing prediction error.
10. **Next Steps - Finalizing Model and Preparing Presentation:**
    * The next phase involves **hyperparameter tuning** for the final model, improving performance based on cross-validation results.
    * We’re also preparing the final **project report and presentation**, where we will discuss insights, model performance, and limitations