Project Summary:

This project focused on implementing a hybrid approach combining K-Nearest Neighbors (KNN) and Linear Regression to predict housing prices. The complete machine learning pipeline included spatial analysis, feature engineering, and a custom implementation of both algorithms without relying on scikit-learn. The model leverages geographical data and housing characteristics to create accurate price predictions while maintaining interpretability.

Specific Actions and Technologies:

- Utilized Python with Pandas, NumPy, and Matplotlib libraries for data analysis and visualization
- Performed initial data processing including:
    - Calculating price per square foot as a derived feature
    - Creating 20 distinct price categories using quantile binning
    - Engineering a combined 'rooms' feature from bedrooms and bathrooms
- Implemented a custom KNN algorithm for spatial price category prediction:
    - Used latitude and longitude as features and Built from scratch without scikit-learn dependencies
    - Visualized predictions against actual values
- Developed a custom Linear Regression model:
    - Features: KNN price category, square footage, lot acres, fireplaces, garage, rooms
    - Implemented using the Normal Equation method and added bias terms and performed feature scaling.
- Model Evaluation and Analysis:
    - Calculated and reported R^2 scores for both training and test sets
    - Generated detailed regression equation with coefficients
    - Analysed feature importance based on standardized coefficients

This summary captures the sophisticated approach taken in the project, highlighting both the technical implementation details and the practical aspects of the housing price prediction system. The hybrid model combines the spatial awareness of KNN with the interpretability of Linear Regression, creating a robust foundation for accurate price predictions.