**Classification using the labels given. Use Naive Bayes, KNN, SVM, and Decision Trees. Compare the Precision, Recall, and F1-score for all these approaches:**

**1. Naive Bayes:**

* **Precision:** 0.70 (70% of positive predictions were correct)
* **Recall:** 0.56 (56% of actual positive cases were correctly identified)
* **F1-score:** 0.62 (a balanced measure of precision and recall)

Naive Bayes seems to be making a fair number of mistakes. While it might have a decent precision, the low recall suggests it misses a significant portion of the actual positive cases/

**2. K-Nearest Neighbours (KNN):**

* **Precision:** 0.76 (76% of positive predictions were correct)
* **Recall:** 0.58 (58% of actual positive cases were correctly identified)
* **F1-score:** 0.66 (slightly better balance than Naive Bayes)

KNN shows a slight improvement over Naive Bayes, particularly in recall. It identifies a higher proportion of actual positive cases but still makes some mistakes in both directions.

**3. Support Vector Classification (SVC):**

* **Precision:** 0.82 (82% of positive predictions were correct)
* **Recall:** 0.29 (only 29% of actual positive cases were correctly identified)
* **F1-score**: 0.42(much worse than both)

SVC has the highest precision, suggesting a low false positive rate (correctly identifying most positive predictions). However, its recall is very low, meaning it misses a large portion of the actual positive cases. This creates a high imbalance, resulting in a low F1 score.

**4. Decision Tree:**

* **Precision:** 0.82 (82% of positive predictions were correct)
* **Recall:** 0.83 (83% of actual positive cases were correctly identified)
* **F1-score:** 0.82 (highest balanced score)

Decision Tree emerges as the best performer among the four models. It achieves a high precision similar to SVC but also boasts a good recall, indicating it correctly identifies a significant portion of the actual positive cases. This balance leads to the highest F1 score.

**Analysis of the data:**

**1. Potential Complexity:**

* The fact that different models perform differently suggests some level of complexity in the data. A simple dataset might have yielded more similar performance across all models.

**2. Class Imbalance (Possibly):**

* The very low recall for SVC (Support Vector Classification) hints at a potential class imbalance. If there are significantly fewer instances of some property types compared to others, SVC might struggle to learn those underrepresented classes effectively.

**3. Feature Relevance (Likely):**

* The success of the Decision Tree model indicates that the features we were using (postcode, price, bedrooms, year, month) are likely relevant for classifying property types. The decision tree can capture these relationships and make accurate predictions.