Features Description:

1. MedInc - Median income in the district (in tens of thousands of dollars)
2. HouseAge - Median age of houses in the district (in years)
3. AveRooms - Average number of rooms per household
4. AveBedrms - Average number of bedrooms per household
5. Population - Total population in the district
6. AveOccup - Average number of household members
7. Latitude - Geographic latitude of district center
8. Longitude - Geographic longitude of district center

Target Variable:

- MEDV - Median house value for the district (in hundreds of thousands of dollars)

---

1) Needed Libraries :

```
sklearn.datasets ,pandas , numpy , matplotlib.pyplot , seaborn
```

2) Basic Data Exploration

- Display the first 10 rows of the dataset
- Check the dataset shape and basic statistics
- Check for missing values
- Display the correlation matrix with the target variable

3) Train & Test Split

Split the data into training (80%) and testing (20%) sets.
(Use train_test_split with random_state=42 for reproducible results.)

4) Feature Selection

Based on your correlation analysis, select the top 3 most correlated features with the target for initial modeling.

5) Building the Regression Tree (25 points)

1. Create a DecisionTreeRegressor with max_depth=3 and random_state=42
2. Train the model on the training data with your selected features
3. Make predictions on both training and test sets

6) Model Evaluation

1. Calculate Mean Squared Error (MSE) and R^2 score for both training and test sets
2. Print the results in a clear format

- Is the model overfitting or underfitting?
- What does the R^2 score tell you about model performance?

7) Tree Visualization

Visualize the decision tree using `plot_tree`.

Set `filled=True`, `rounded=True`, and specify `feature_names` for better readability.