

Tasks

1. Data Exploration

- Load the [black Friday dataset from kaggle](#) and examine the first few rows.
- Explore summary statistics (mean, median, missing values) for key features.
- Visualize relationships between features and sale price using scatter plots or boxplots.

2. Data Preparation

- Handle any missing values appropriately.
- Encode categorical variables (e.g., neighborhood).
- Split data into training and testing sets (e.g., 80% train, 20% test).

3. Model Building

- Train a decision tree regression model on the training data.
- Examine the tree structure and feature importance.

4. Model Evaluation

- Predict sale prices on the test set.
- Calculate evaluation metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.
- Plot predicted vs actual sale prices.

5. Model Tuning

- Experiment with hyperparameters like max depth or min samples split to improve performance.
- Discuss overfitting and underfitting in the context of your results.

6. Reflection

- Which features are most influential in predicting housing prices?
- What improvements could be made to the model or data collection?

Deliverables

- A Jupyter notebook or script with all steps and outputs clearly commented.
- Visualizations supporting your analysis.
- A short summary paragraph highlighting key findings and model performance.