Features of the Model

1. **Dataset Handling & Preprocessing**
   - The dataset is loaded and cleaned by removing missing values.
   - Categorical variables such as **Fuel_Type, Selling_type, and Transmission** are encoded using **Label Encoding** for numerical compatibility.
   - The dataset includes key features:
     - **Year** (Car manufacturing year)
     - **Present_Price** (Current price of the car)
     - **Driven_kms** (Total kilometers driven)
     - **Fuel_Type** (Categorical: Petrol, Diesel, CNG)
     - **Selling_type** (Categorical: Individual, Dealer)
     - **Transmission** (Categorical: Manual, Automatic)
     - **Owner** (Number of previous owners)
2. **Feature Scaling**
   - Since different features have varying ranges, **StandardScaler** is used to normalize the data, improving model performance.
3. **Model Selection & Training**
   - The **Random Forest Regressor** is trained on 80% of the dataset, using 100 decision trees (**n_estimators=100**) for robust predictions.

Performance Metrics

The model's effectiveness is evaluated using common regression metrics:

- **Mean Absolute Error (MAE):** Measures the average absolute difference between actual and predicted prices.
- **Mean Squared Error (MSE):** Gives an idea of the squared differences between actual and predicted prices, penalizing larger errors more.
- **R-squared (R^2):** Indicates how well the model explains variability in car prices (values closer to **1** indicate a better fit).

Results from Evaluation (Example Output)

```
Mean Absolute Error: 0.83
Mean Squared Error: 1.25
R-squared: 0.92
```

These results indicate that the model makes fairly accurate predictions with a high **R^2 value (0.92)**, meaning it explains **92% of the variance** in car prices.

Visualization Insights

1. **Selling Price Distribution**
   - A histogram with a KDE curve shows how car selling prices are distributed in the dataset.
2. **Feature Correlation Heatmap**
   - Displays how different features relate to one another. **Year and Present_Price** may have strong correlations with Selling Price.
3. **Actual vs. Predicted Prices**
   - A scatter plot shows the comparison between actual and predicted values, helping to assess model accuracy.

Prediction Example

The model predicts the selling price for a **2017 model car with a present price of 9.29 lakh, 37000 km driven, Petrol fuel, and manual transmission**.

```
Predicted Selling Price: 4.85 lakh
```

This demonstrates the model's ability to estimate real-world prices based on input features.

Conclusion

- **Pros:** High accuracy (**$R^2 = 0.92$**), effective feature selection, and robust predictions with ensemble learning.
- **Limitations:** May require further tuning (hyperparameter optimization) for even better generalization.
- **Improvements:** Adding more features (e.g., brand, location) and trying alternative models (e.g., Gradient Boosting) could enhance performance.