In [1]:
```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split,cross_val_score
from sklearn.metrics import mean_squared_error, r2_score,mean_absolute_error,mean_absolu
import matplotlib.pyplot as plt
data=pd.read_csv('CarPrice_Assignment.csv')
```

In [2]:
```
data=data.drop(['car_ID','CarName'],axis=1)#removes unnecessary columns
data = pd.get_dummies(data, drop_first=True)# Handle categorical variables
```

In [3]:
```
X = data.drop('price', axis=1)
Y = data['price']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=42
```

In [4]:
```
model = LinearRegression()
model.fit(X_train, Y_train)
```

Out[4]:
```
LinearRegression
LinearRegression()
```

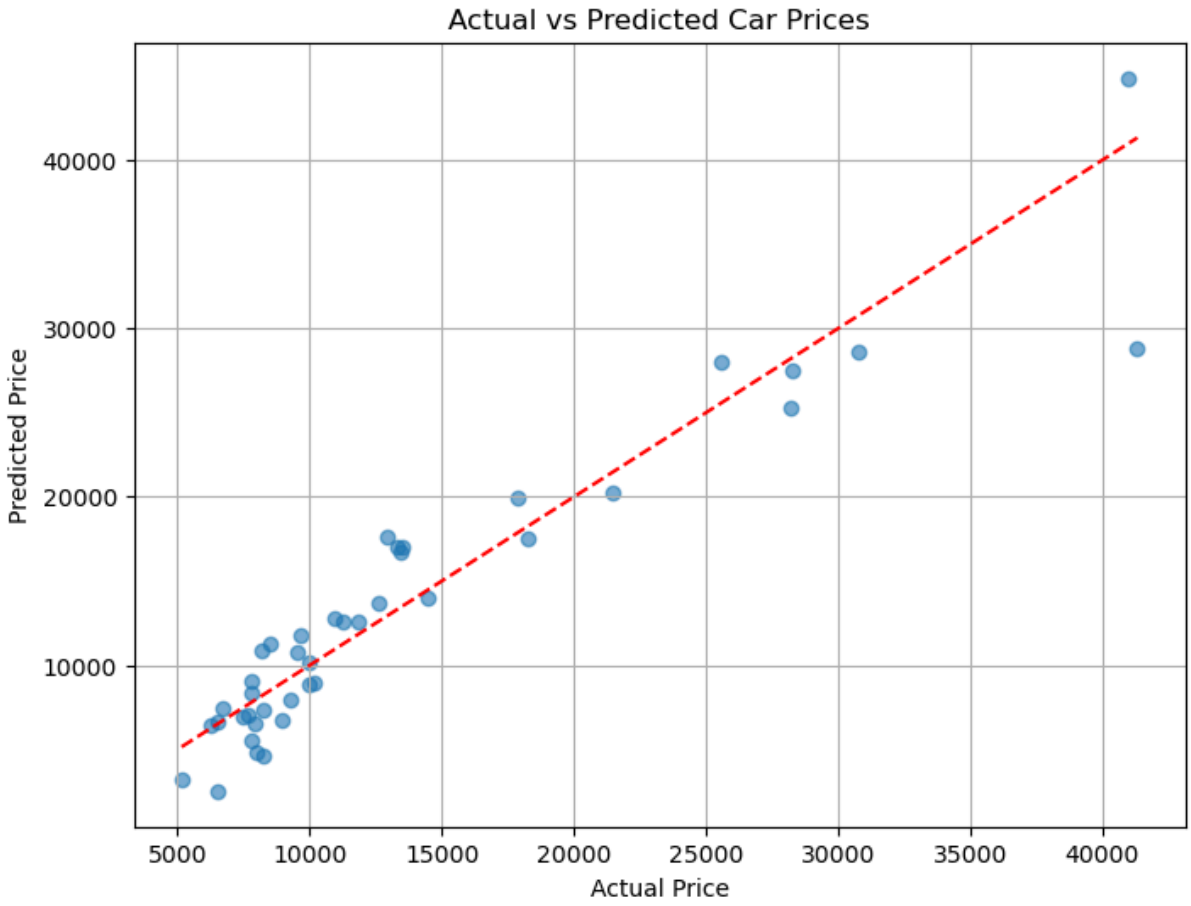
In [5]:
```
print("Name: RAGHUL.S")
print("Reg. No: 212225040325")
print("\n=== Cross-validation ===")
cv_scores=cross_val_score(model,X,Y,cv=5)
print("Fold R2 scores:",[f"{score:.4f}" for score in cv_scores])
print(f"Average R2: {cv_scores.mean():.4f}")
```

```
Name: RAGHUL.S
Reg. No: 212225040325

=== Cross-validation ===
Fold R2 scores: ['0.6238', '0.6316', '0.3132', '0.3643', '-0.4944']
Average R2: 0.2877
```

In [6]:
```
Y_pred = model.predict(X_test)
print("\n=== Test Set Performance ===")
print(f"MSE: {mean_squared_error(Y_test, Y_pred):.2f}")
print(f"MAE: {mean_absolute_error(Y_test, Y_pred):.2f}")
print(f"R²: {r2_score(Y_test, Y_pred):.4f}")
plt.figure(figsize=(8, 6))
plt.scatter(Y_test, Y_pred, alpha=0.6)
plt.plot([Y_test.min(), Y_test.max()],[Y_test.min(), Y_test.max()],'r--')
plt.xlabel("Actual Price")
plt.ylabel("Predicted Price")
plt.title("Actual vs Predicted Car Prices")
plt.grid(True)
plt.show()
```

```
=== Test Set Performance ===
MSE: 8482008.48
MAE: 2089.38
R²: 0.8926
```



In [ ]: