TASK 2: CAR PRICE PREDICTION

```python
import pandas as pd

import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_squared_error, r2_score

from sklearn.preprocessing import StandardScaler

df = pd.read_csv("S:\SEM2\Machine learning\\car data.csv")

print("First 5 rows of the dataset:")

print(df.head())

print("\nMissing values in each column:")

print(df.isnull().sum())

for col in df.columns:

    if df[col].dtype == 'object':

        df[col].fillna(df[col].mode()[0], inplace=True)

    else:

        df[col].fillna(df[col].mean(), inplace=True)

df = pd.get_dummies(df, drop_first=True)

X = df.drop("Price", axis=1)

y = df["Price"]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()

X_train_scaled = scaler.fit_transform(X_train)

X_test_scaled = scaler.transform(X_test)

model = RandomForestRegressor(n_estimators=100, random_state=42)

model.fit(X_train_scaled, y_train)

y_pred = model.predict(X_test_scaled)
```

```
mse = mean_squared_error(y_test, y_pred)

rmse = np.sqrt(mse)

r2 = r2_score(y_test, y_pred)

print("\nEvaluation Metrics:")

print(f"Mean Squared Error: {mse:.2f}")

print(f"Root Mean Squared Error: {rmse:.2f}")

print(f"R-squared: {r2:.2f}")
```

OUTPUT:

```
First 5 rows of the dataset:
  Car_Name  Year  Selling_Price  Present_Price  Driven_kms Fuel_Type  \
0     ritz  2014           3.35           5.59       27000    Petrol
1      sx4  2013           4.75           9.54       43000    Diesel
2     ciaz  2017           7.25           9.85        6900    Petrol
3  wagon r  2011           2.85           4.15        5200    Petrol
4    swift  2014           4.60           6.87       42450    Diesel

  Selling_type Transmission  Owner
0       Dealer       Manual      0
1       Dealer       Manual      0
2       Dealer       Manual      0
3       Dealer       Manual      0
4       Dealer       Manual      0

Missing values in each column:
Car_Name         0
Year             0
Selling_Price    0
Present_Price    0
Driven_kms       0
Fuel_Type        0
Selling_type     0
Transmission     0
Owner            0
dtype: int64

Evaluation Metrics:
Mean Squared Error: 0.78
Root Mean Squared Error: 0.88
R-squared: 0.97
```