DT

```
#import sys
#print(sys.executable)
```

```
#import sys
#!{sys.executable} -m pip install pandas
```

```
#import sys
#!{sys.executable} -m pip install matplotlib
```

```
#import sys
#!{sys.executable} -m pip install seaborn
```

```
#import sys
#!{sys.executable} -m pip install scikit-learn
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import mean_squared_error, confusion_matrix, classification_report
from sklearn.tree import DecisionTreeClassifier, DecisionTreeRegressor, plot_tree
```

```
cs = pd.read_csv('https://raw.githubusercontent.com/PratheepaJ/datasets/master/Carseats.csv')

cs.info()
cs.head(5)
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 400 entries, 0 to 399
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Sales        400 non-null    float64
 1   CompPrice    400 non-null    int64
 2   Income       400 non-null    int64
 3   Advertising  400 non-null    int64
 4   Population   400 non-null    int64
 5   Price        400 non-null    int64
 6   ShelveLoc    400 non-null    object
 7   Age          400 non-null    int64
 8   Education    400 non-null    int64
 9   Urban        400 non-null    object
 10  US           400 non-null    object
 11  High         400 non-null    object
dtypes: float64(1), int64(7), object(4)
memory usage: 37.6+ KB
```

	Sales	CompPrice	Income	Advertising	Population	Price	ShelveLoc	Age	Education	Urban	U
0	9.50	138	73	11	276	120	Bad	42	17	Yes	Y
1	11.22	111	48	16	260	83	Good	65	10	Yes	Y
2	10.06	113	35	10	269	80	Medium	59	12	Yes	Y
3	7.40	117	100	4	466	97	Medium	55	14	Yes	Y
4	4.15	141	64	3	340	128	Bad	38	13	Yes	N

Convert categories onto dummy vatiables

```
cat = ['ShelveLoc', 'Urban', 'US']
cs = pd.get_dummies(cs,columns=cat)
```

```
cs.head(5)
```

	Sales	CompPrice	Income	Advertising	Population	Price	Age	Education	High	ShelveLoc_Bad
0	9.50	138	73	11	276	120	42	17	Yes	True
1	11.22	111	48	16	260	83	65	10	Yes	False
2	10.06	113	35	10	269	80	59	12	Yes	False
3	7.40	117	100	4	466	97	55	14	No	False
4	4.15	141	64	3	340	128	38	13	No	True

	Sales	CompPrice	Income	Advertising	Population	Price	Age	Education	High	ShelveLoc_Bad

terminal

Github link

Github link