

main

September 15, 2025

```
[4]: import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import f1_score

df = pd.read_csv('car.data', header=None)
# Rename columns according to attribute names
df.columns = ['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class']

# One-hot encode categorical features (excluding the target 'class')
df_encoded = pd.get_dummies(df, columns=['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety'])

# Display the first few rows of the encoded DataFrame
df_encoded.head()

# Prepare features and target
X = df_encoded.drop('class', axis=1)
y = df_encoded['class']
X.describe()
```

```
[4]:      buying_high buying_low buying_med buying_vhigh maint_high maint_low \
count      1728      1728      1728      1728      1728      1728
unique         2         2         2         2         2         2
top          False      False      False      False      False      False
freq         1296      1296      1296      1296      1296      1296

      maint_med maint_vhigh doors_2 doors_3 ... doors_5more persons_2 \
count      1728      1728      1728      1728 ...      1728      1728
unique         2         2         2         2 ...         2         2
top          False      False      False      False ...      False      False
freq         1296      1296      1296      1296 ...      1296      1152

      persons_4 persons_more lug_boot_big lug_boot_med lug_boot_small \
count      1728      1728      1728      1728      1728
unique         2         2         2         2         2
```

top	False	False	False	False	False
freq	1152	1152	1152	1152	1152

	safety_high	safety_low	safety_med
count	1728	1728	1728
unique	2	2	2
top	False	False	False
freq	1152	1152	1152

[4 rows x 21 columns]

```
[5]: y.describe()
```

```
[5]: count      1728
unique         4
top      unacc
freq      1210
Name: class, dtype: object
```