In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df = pd.read_csv('USA_cars_datasets.csv')
df.head()
```

Out[2]:

	Unnamed: 0	price	brand	model	year	title_status	mileage	color	vin	lot	state	country	condition
0	0	6300	toyota	cruiser	2008	clean vehicle	274117.0	black	jtezu11f88k007763	159348797	new jersey	usa	10 days left
1	1	2899	ford	se	2011	clean vehicle	190552.0	silver	2fmdk3gc4bbb02217	166951262	tennessee	usa	6 days left
2	2	5350	dodge	mpv	2018	clean vehicle	39590.0	silver	3c4pdcgg5jt346413	167655728	georgia	usa	2 days left
3	3	25000	ford	door	2014	clean vehicle	64146.0	blue	1ftfw1et4efc23745	167753855	virginia	usa	22 hours left
4	4	27700	chevrolet	1500	2018	clean vehicle	6654.0	red	3gcpcrec2jg473991	167763266	florida	usa	22 hours left

Top car brand in terms of number of cars.

In [3]:
```
print(df.isnull().sum())
```

```
Unnamed: 0      0
price           0
brand           0
model           0
year            0
title_status    0
mileage         0
color           0
vin             0
lot             0
state           0
country         0
condition       0
dtype: int64
```

In [4]:
```
gmc = df[df.brand == 'gmc']['lot'].sum()
nissan = df[df.brand == 'nissan']['lot'].sum()
dodge = df[df.brand == 'dodge']['lot'].sum()
ford = df[df.brand == 'ford']['lot'].sum()
list1 = [gmc, nissan, dodge, ford]
print(max(list1), ford)
```

```
207115996784 207115996784
```

Which of the following car brand is not there in the dataset?

In [5]:
```
list2 = ['ford', 'dodge', 'nissan', 'gmc', 'ferrari']
df.brand.unique()
for i in list2:
    print(i,i in df.brand.unique())
```

```
ford True
dodge True
nissan True
gmc True
ferrari False
```

Average Price of cars in 2018? (Neglect the decimal values)

In [6]:
```
sums = df[df.year == 2018]["price"].sum()
count = df[df.year == 2018]["price"].count()
print(sums//count)
```

```
19058
```

Most expensive car brand with maximum average car price?

In [7]:
```
list3 = ['ford', 'lexus', 'nissan', 'harley-davidson']
for i in list3:
    sums = 0
    count = 0
    count = df[df.brand == i]['price'].count()
    sums = df[df.brand == i]['price'].sum()
    print(i, sums//count)
```

```
ford 21666
lexus 33220
nissan 12065
harley-davidson 54680
```

Which state is having most expensive cars (maximum average car price)?

In [8]:
```
list4 = ['kentucky','ontario','new mexico','california']
for i in list4:
    sums = 0
    count = 0
    count = df[df.state == i]['price'].count()
    sums = df[df.state == i]['price'].sum()
    print(i, sums//count)
```

```
kentucky 44969
ontario 30357
new mexico 27375
california 17836
```

In [ ]:
```

```