```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
```

Writing function to see null values, describing data, shape of data

```
def read_check(X):
    df = pd.read_csv(X)
    print(df.isnull().sum())
    print(df.describe())
    print(df.shape)
    return df


df = read_check('/content/car.data')
```

```
vhigh      0
vhigh.1    0
2          0
2.1        0
small      0
low        0
unacc      0
dtype: int64
       vhigh vhigh.1     2   2.1 small   low  unacc
count   1727    1727  1727  1727  1727  1727   1727
unique     4       4     4     3     3     3      4
top     high    high     3     4   med   med  unacc
freq     432     432   432   576   576   576   1209
(1727, 7)
```

look at the data

```
df.head
```

```
<bound method NDFrame.head of       vhigh vhigh.1      2   2.1  small   low  unacc
0     vhigh   vhigh      2     2  small   med  unacc
1     vhigh   vhigh      2     2  small  high  unacc
2     vhigh   vhigh      2     2    med   low  unacc
3     vhigh   vhigh      2     2    med   med  unacc
4     vhigh   vhigh      2     2    med  high  unacc
...     ...     ...    ...   ...    ...   ...    ...
1722    low     low  5more  more    med   med   good
```

```
1723    low     low  5more  more    med  high  vgood
1724    low     low  5more  more    big   low  unacc
1725    low     low  5more  more    big   med   good
1726    low     low  5more  more    big  high  vgood

[1727 rows x 7 columns]>
```

```
col_names = ['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class_val']
```

```
df.columns = col_names
```

```
col_names
```

```
['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety', 'class_val']
```

```
df.head()
```

	buying	maint	doors	persons	lug_boot	safety	class_val
0	vhigh	vhigh	2	2	small	med	unacc
1	vhigh	vhigh	2	2	small	high	unacc
2	vhigh	vhigh	2	2	med	low	unacc
3	vhigh	vhigh	2	2	med	med	unacc
4	vhigh	vhigh	2	2	med	high	unacc

```
new_df = df.query('maint == "vhigh" & doors == "4" & safety == "high" & lug_boot == "big" & c
```

```
new_df.head()
```

	buying	maint	doors	persons	lug_boot	safety	class
61	vhigh	vhigh	4	2	big	high	unacc
70	vhigh	vhigh	4	4	big	high	unacc
79	vhigh	vhigh	4	more	big	high	unacc
493	high	vhigh	4	2	big	high	unacc
502	high	vhigh	4	4	big	high	unacc

Colab paid products - Cancel contracts here

✓ 0s completed at 4:37 PM