

In Exercise 6.4.csv (right-click on the file name to download the file) there is market survey data collected by a car manufacturing firm. The company want to determine if the model appeal is dependent on the age group. What can the manufacturer conclude based on the data? Write 2-3 sentences describing your findings.

```
In [21]: import pandas as pd      #importing pandas package

from scipy import stats #importing stats module from scipy library

df = pd.read_csv (r'Exercise 6.4.csv')
df
```

Out[21]:

	id	age	like
0	0	<20	n
1	1	20-40	n
2	2	40-50	y
3	3	40-50	n
4	4	20-40	y
...	...	...	...
495	495	20-40	y
496	496	>50	y
497	497	<20	y
498	498	<20	n
499	499	<20	y

500 rows × 3 columns

For solving this problem we will use the Pearson's Chi-Square statistical hypothesis which is a test for checking independence between categorical variables

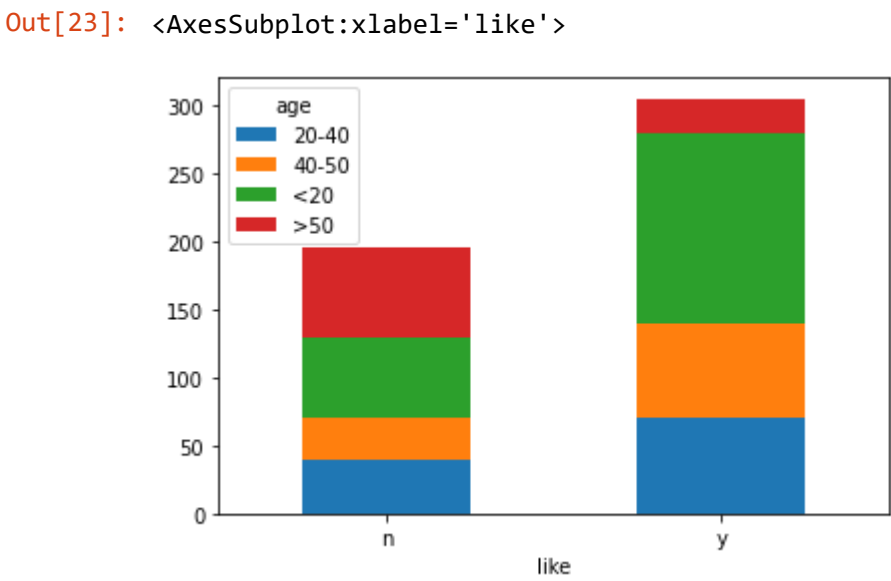
```
In [22]: cont_table = pd.crosstab(df['like'], df['age']) #A Contingency table (also called crosstab) is used in statistics to summarise the relationship between several categorical variables.

cont_table # print the contingency table
```

Out[22]:

age	20-40	40-50	<20	>50
like				
n	40	30	60	65
y	70	70	140	25

```
In [23]: cont_table.plot(kind="bar", stacked=True, rot=0) #The `crosstab` function returns a dataframe, so we can easily visualize the contingency table as stacked bars
```



The figure clearly depicts the people who belong to age group - less than 20years old (<20) hold a greater share of people who are appealed by the car model as compared to all other age groups

```
In [24]: chi2, p_val, dof, expected = stats.chi2_contingency(cont_table) #To run a chi-square test on the contingency table, we input it directly into the `chi2_contingency` function.

#The function returns four values : the chi-square value, p-value, degrees of freedom and a table with values expected under the independence assumption.

print(f"p-value: {p_val}") #printing the p-value

p-value: 2.5971843881195768e-11
```

The p-value is greater than 0.05 which clearly illustrates that we fail to reject null hypothesis and therefore there is no dependency of model appeal with age