


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

dataframe=pd.read_csv('marketing_data.csv',index_col=0)
dataframe
```



	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWi
ID									
1826	1970	Graduation	Divorced	\$84,835.00	0	0	6/16/14	0	
1	1961	Graduation	Single	\$57,091.00	0	0	6/15/14	0	
10476	1958	Graduation	Married	\$67,267.00	0	1	5/13/14	0	
1386	1967	Graduation	Together	\$32,474.00	1	1	5/11/14	0	
5371	1989	Graduation	Single	\$21,474.00	1	0	4/8/14	0	
...
10142	1976	PhD	Divorced	\$66,476.00	0	1	3/7/13	99	
5263	1977	2n Cycle	Married	\$31,056.00	1	0	1/22/13	99	
22	1976	Graduation	Divorced	\$46,310.00	1	0	12/3/12	99	
528	1978	Graduation	Married	\$65,819.00	0	0	11/29/12	99	
4070	1969	PhD	Married	\$94,871.00	0	2	9/1/12	99	

2240 rows × 27 columns

```
dataframe.head(5)
```

	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWi
ID									
1826	1970	Graduation	Divorced	\$84,835.00	0	0	6/16/14	0	
1	1961	Graduation	Single	\$57,091.00	0	0	6/15/14	0	
10476	1958	Graduation	Married	\$67,267.00	0	1	5/13/14	0	
1386	1967	Graduation	Together	\$32,474.00	1	1	5/11/14	0	
5371	1989	Graduation	Single	\$21,474.00	1	0	4/8/14	0	

5 rows × 27 columns

```
dataframe.tail(5)
```

	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWi
ID									
10142	1976	PhD	Divorced	\$66,476.00	0	1	3/7/13	99	
5263	1977	2n Cycle	Married	\$31,056.00	1	0	1/22/13	99	
22	1976	Graduation	Divorced	\$46,310.00	1	0	12/3/12	99	
528	1978	Graduation	Married	\$65,819.00	0	0	11/29/12	99	
4070	1969	PhD	Married	\$94,871.00	0	2	9/1/12	99	

5 rows × 27 columns

```
dataframe.shape
```

(2240, 27)

```
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2240 entries, 1826 to 4070
Data columns (total 27 columns):
#   Column                Non-Null Count  Dtype
#   ...
```

```

---
0 Year_Birth 2240 non-null int64
1 Education 2240 non-null object
2 Marital_Status 2240 non-null object
3 Income 2216 non-null object
4 Kidhome 2240 non-null int64
5 Teenhome 2240 non-null int64
6 Dt_Customer 2240 non-null object
7 Recency 2240 non-null int64
8 MntWines 2240 non-null int64
9 MntFruits 2240 non-null int64
10 MntMeatProducts 2240 non-null int64
11 MntFishProducts 2240 non-null int64
12 MntSweetProducts 2240 non-null int64
13 MntGoldProds 2240 non-null int64
14 NumDealsPurchases 2240 non-null int64
15 NumWebPurchases 2240 non-null int64
16 NumCatalogPurchases 2240 non-null int64
17 NumStorePurchases 2240 non-null int64
18 NumWebVisitsMonth 2240 non-null int64
19 AcceptedCmp3 2240 non-null int64
20 AcceptedCmp4 2240 non-null int64
21 AcceptedCmp5 2240 non-null int64
22 AcceptedCmp1 2240 non-null int64
23 AcceptedCmp2 2240 non-null int64
24 Response 2240 non-null int64
25 Complain 2240 non-null int64
26 Country 2240 non-null object
dtypes: int64(22), object(5)
memory usage: 490.0+ KB

```

```
dataframe.describe()
```

	Year_Birth	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts
count	2240.000000	2240.000000	2240.000000	2240.000000	2240.000000	2240.000000	2240.000000	2240.000000
mean	1968.805804	0.444196	0.506250	49.109375	303.935714	26.302232	166.950000	166.950000
std	11.984069	0.538398	0.544538	28.962453	336.597393	39.773434	225.715373	225.715373
min	1893.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1959.000000	0.000000	0.000000	24.000000	23.750000	1.000000	16.000000	16.000000
50%	1970.000000	0.000000	0.000000	49.000000	173.500000	8.000000	67.000000	67.000000
75%	1977.000000	1.000000	1.000000	74.000000	504.250000	33.000000	232.000000	232.000000
max	1996.000000	2.000000	2.000000	99.000000	1493.000000	199.000000	1725.000000	1725.000000

8 rows × 22 columns

```
numerical_feature_column=list(dataframe._get_numeric_data().columns)
numerical_feature_column
```

```

['Year_Birth',
 'Kidhome',
 'Teenhome',
 'Recency',
 'MntWines',
 'MntFruits',
 'MntMeatProducts',
 'MntFishProducts',
 'MntSweetProducts',
 'MntGoldProds',
 'NumDealsPurchases',
 'NumWebPurchases',
 'NumCatalogPurchases',
 'NumStorePurchases',
 'NumWebVisitsMonth',
 'AcceptedCmp3',
 'AcceptedCmp4',
 'AcceptedCmp5',
 'AcceptedCmp1',
 'AcceptedCmp2',
 'Response',
 'Complain']

```

```
dataframe.isna().sum()
```

```

Year_Birth      0
Education        0
Marital_Status  0
Income          24
Kidhome          0
Teenhome         0

```

Dt_Customer	0
Recency	0
MntWines	0
MntFruits	0
MntMeatProducts	0
MntFishProducts	0
MntSweetProducts	0
MntGoldProds	0
NumDealsPurchases	0
NumWebPurchases	0
NumCatalogPurchases	0
NumStorePurchases	0
NumWebVisitsMonth	0
AcceptedCmp3	0
AcceptedCmp4	0
AcceptedCmp5	0
AcceptedCmp1	0
AcceptedCmp2	0
Response	0
Complain	0
Country	0
dtype:	int64

```
categorical_feature_columns=list(set(dataframe.columns)-set(dataframe._get_numeric_data().columns))
```

```
categorical_feature_columns
```

```
[' Income ', 'Education', 'Dt_Customer', 'Marital_Status', 'Country']
```

```
import seaborn as sns
```

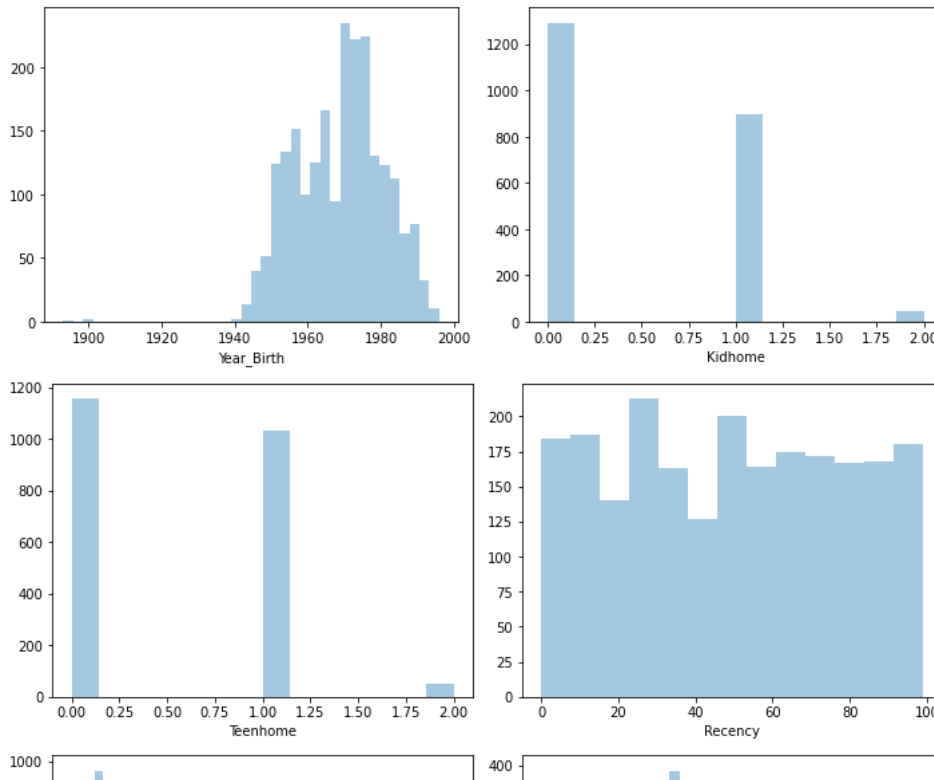
```
num_cols = ['Year_Birth','Kidhome', 'Teenhome', 'Recency', 'NumDealsPurchases','NumWebVisitsMonth','Response']
```

```
for i in range(0,len(num_cols),2):
    plt.figure(figsize=(10,4))
    plt.subplot(121)
    sns.distplot(dataframe[num_cols[i]], kde=False)
    plt.subplot(122)
    sns.distplot(dataframe[num_cols[i+1]], kde=False)
    plt.tight_layout()
    plt.show()
```

```

/usr/local/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
warnings.warn(msg, FutureWarning)

```



```

num_cols = ['Year_Birth', 'Kidhome', 'Teenhome', 'Recency', 'NumDealsPurchases', 'NumWebVisitsMonth', 'Response']

```

```

for i in range(0, len(num_cols), 2):
    plt.figure(figsize=(10, 4))
    plt.subplot(121)
    sns.distplot(dataframe[num_cols[i]], hist=True, kde=True)
    plt.subplot(122)
    sns.distplot(dataframe[num_cols[i+1]], hist=True, kde=True)
    plt.tight_layout()
    plt.show()

```



```

num_cols = ['Year_Birth', 'Kidhome', 'Teenhome', 'Recency', 'NumDealsPurchases', 'NumWebVisitsMonth', 'Response']

```

```

facet = None

```

```

for i in range(0, len(num_cols), 2):
    plt.figure(figsize=(10, 4))
    plt.subplot(121)
    sns.boxplot(facet, num_cols[i], data = dataframe)
    plt.subplot(122)
    sns.boxplot(facet, num_cols[i+1], data = dataframe)
    plt.tight_layout()
    plt.show()

```



```

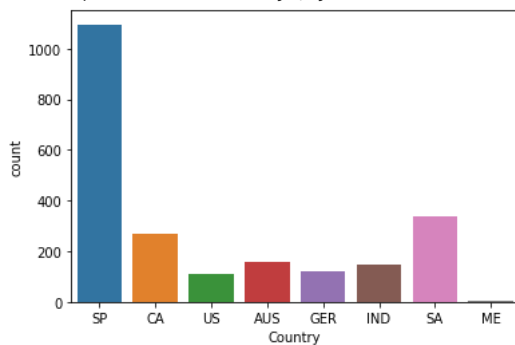
sns.countplot('Country', data=dataframe)

```

```

/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
<AxesSubplot: xlabel='Country', ylabel='count'>

```



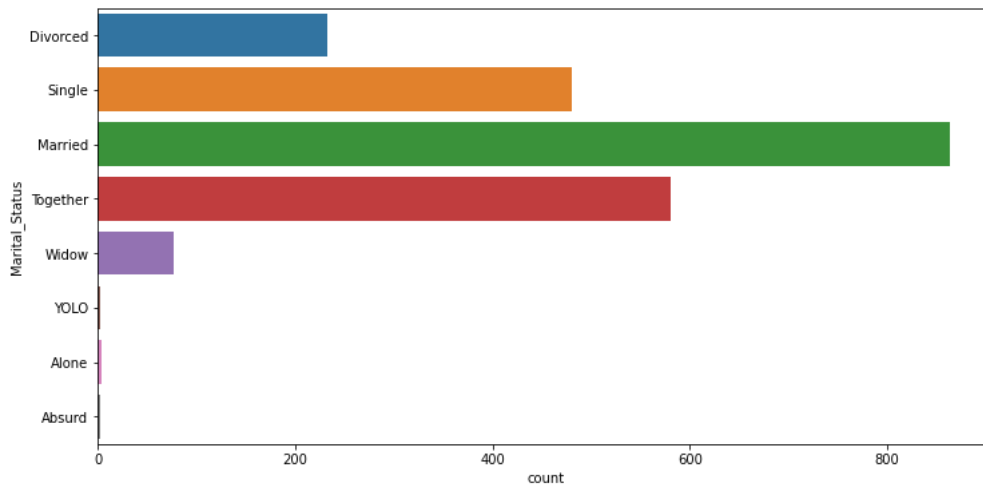
```

plt.figure(figsize=(12, 6))

```

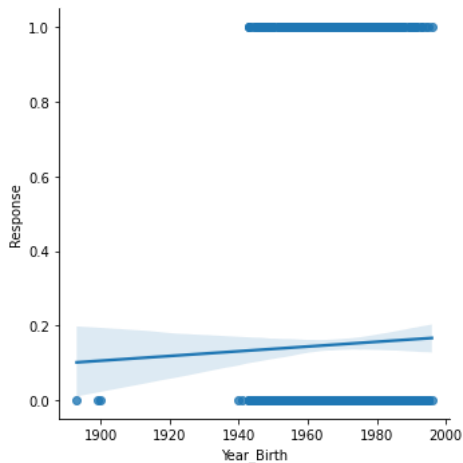
```
sns.countplot(y='Marital_Status', data=dataframe)
```

```
<AxesSubplot:xlabel='count', ylabel='Marital_Status'>
```



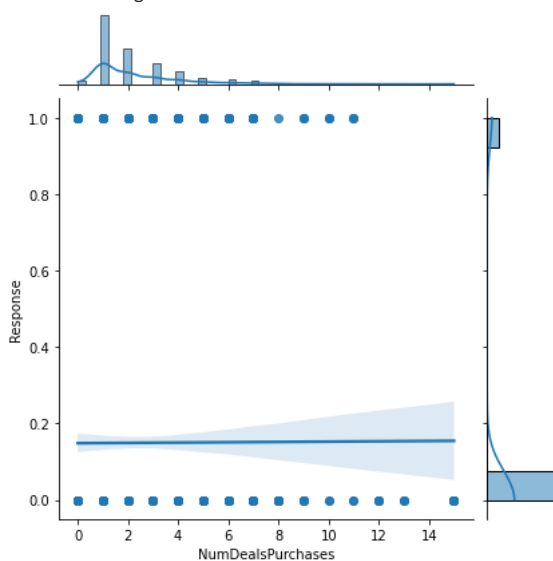
```
sns.lmplot('Year_Birth', 'Response', data=dataframe, fit_reg=True)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
<seaborn.axisgrid.FacetGrid at 0x7f17fc4020d0>
```



```
sns.jointplot('NumDealsPurchases', 'Response', data=dataframe, kind='reg')
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f17fc48f510>
```

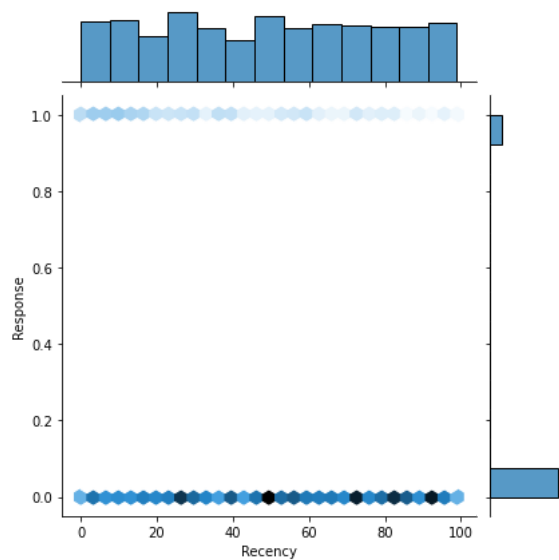


```
sns.jointplot('Recency', 'Response', data=dataframe, kind='hex')
```

```

/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f17fc38c050>

```

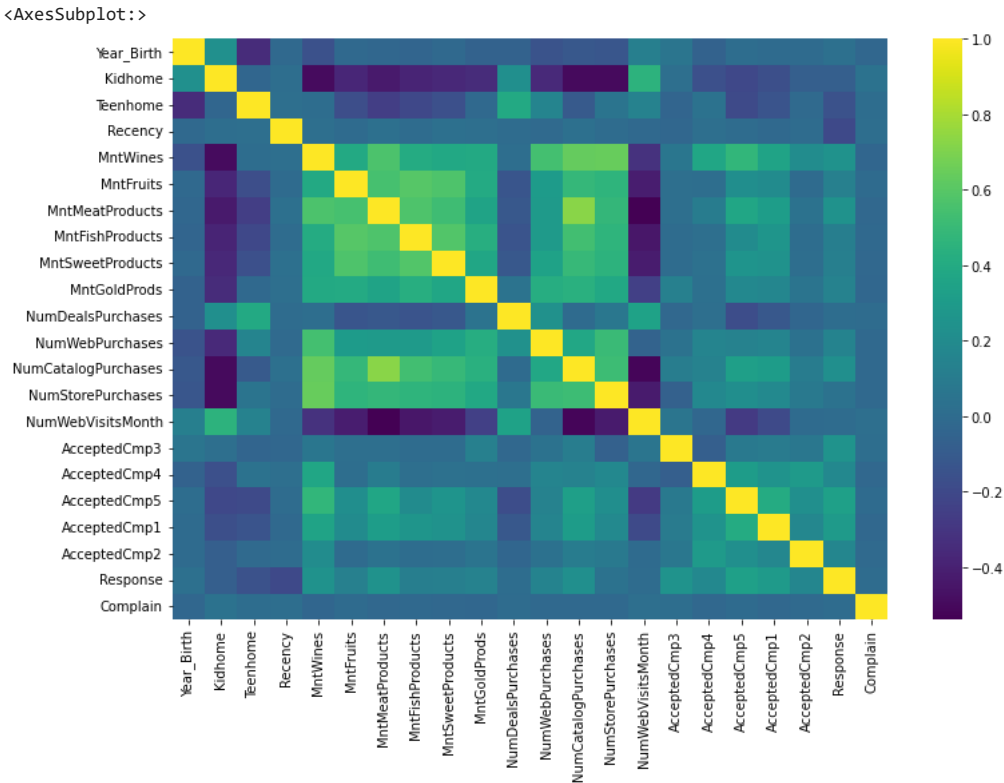


##Its necessary to remove correlated variables to improve our model. We must find correlations can visualize the correlation matrix using

```

plt.figure(figsize=(12,8))
sns.heatmap(dataframe.corr(), cmap='viridis')

```

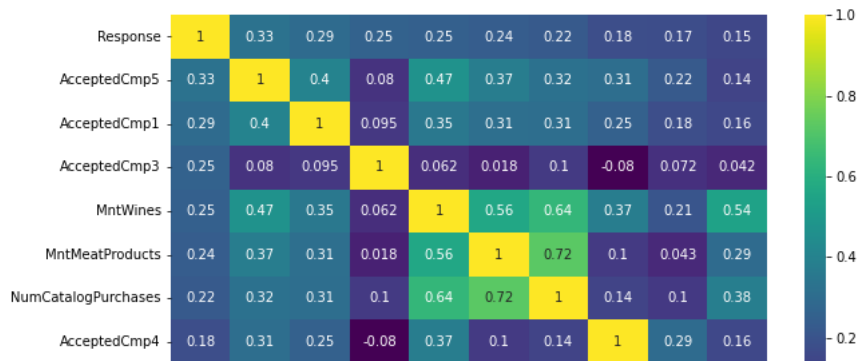


```

#saleprice correlation matrix
k = 10 #number of variables for heatmap
cols = dataframe.corr().nlargest(k, 'Response')['Response'].index
cm = dataframe[cols].corr()
plt.figure(figsize=(10,6))
sns.heatmap(cm, annot=True, cmap = 'viridis')

```

<AxesSubplot:>

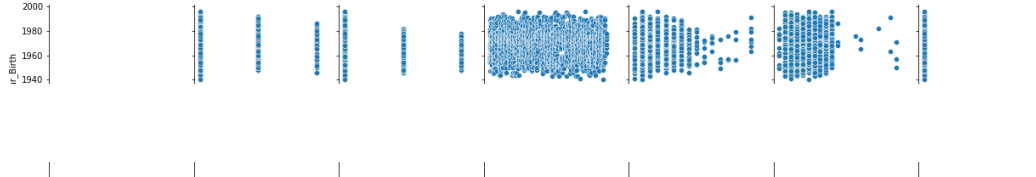


```
# Visualizing relations between all major variables
```

```
cols=['Year_Birth','Kidhome', 'Teenhome', 'Recency', 'NumDealsPurchases','NumWebVisitsMonth','Response']
```

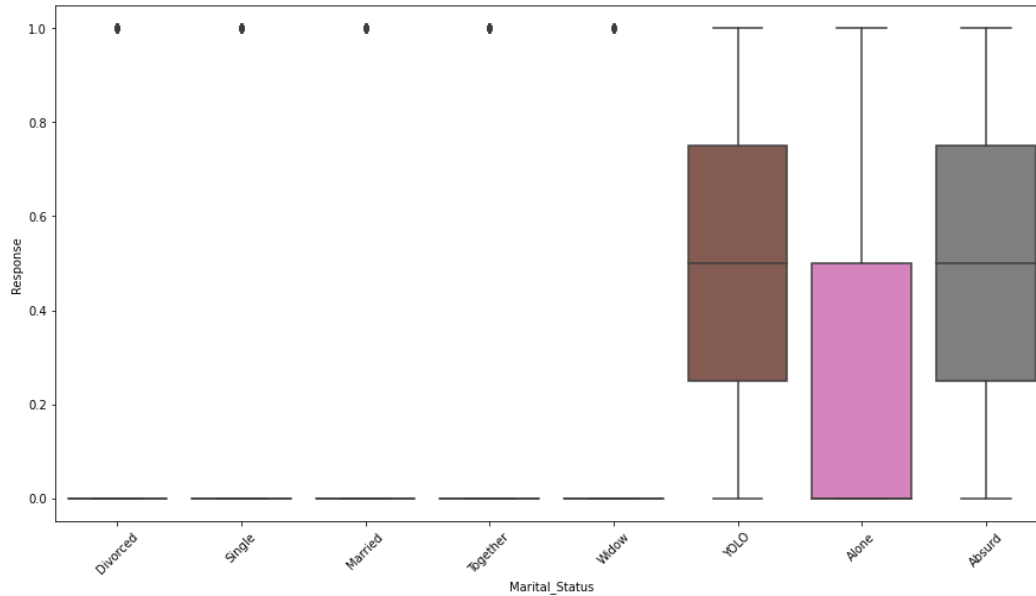
```
sns.pairplot(dataframe[cols])
```

```
<seaborn.axisgrid.PairGrid at 0x7f17fc82a3d0>
```



```
plt.figure(figsize=(15,8))
plt.xticks(rotation=45)
sns.boxplot('Marital_Status', 'Response', data=dataframe)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
<AxesSubplot: xlabel='Marital_Status', ylabel='Response'>
```



```
plt.figure(figsize=(12,6))
sns.swarmplot('AcceptedCmp5', 'Response', data=dataframe)
```

```
/usr/local/lib/python3.7/site-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following var
FutureWarning
/usr/local/lib/python3.7/site-packages/seaborn/categorical.py:1296: UserWarning: 95.1% of the points c
warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/site-packages/seaborn/categorical.py:1296: UserWarning: 37.4% of the points c
warnings.warn(msg, UserWarning)
<AxesSubplot: xlabel='AcceptedCmp5', ylabel='Response'>
```

