In [1]:

```
#Step 1: Import Libraries
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
import missingno as msno
from scipy import stats
from scipy.stats import chi2_contingency
from scipy.stats.contingency import expected_freq
from scipy.stats import chi2
```

In [2]:

```
#Step 2: load dataset
data = pd.read_csv('C:/Users/ericy/Desktop/medical_clean.csv')
```

In [3]:

```
data.info()
```

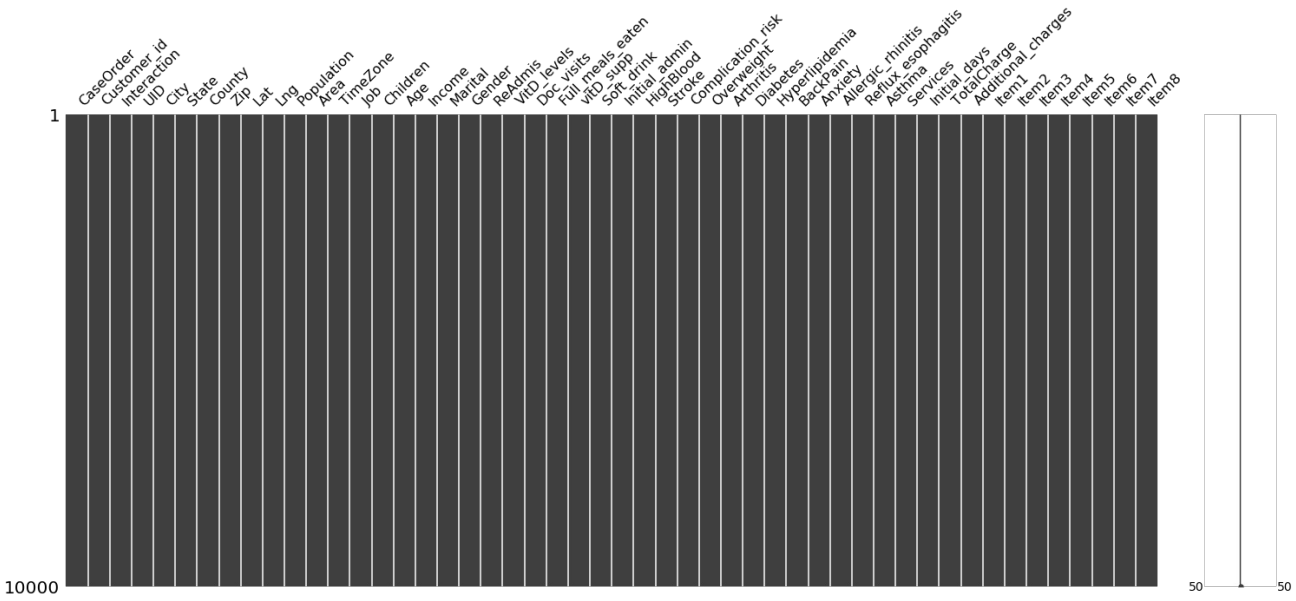
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 50 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   CaseOrder           10000 non-null  int64
 1   Customer_id         10000 non-null  object
 2   Interaction         10000 non-null  object
 3   UID                 10000 non-null  object
 4   City                10000 non-null  object
 5   State               10000 non-null  object
 6   County              10000 non-null  object
 7   Zip                 10000 non-null  int64
 8   Lat                 10000 non-null  float64
 9   Lng                 10000 non-null  float64
 10  Population          10000 non-null  int64
 11  Area                10000 non-null  object
 12  TimeZone            10000 non-null  object
 13  Job                 10000 non-null  object
 14  Children            10000 non-null  int64
 15  Age                 10000 non-null  int64
 16  Income              10000 non-null  float64
 17  Marital             10000 non-null  object
 18  Gender              10000 non-null  object
 19  ReAdmis             10000 non-null  object
 20  VitD_levels         10000 non-null  float64
 21  Doc_visits          10000 non-null  int64
 22  Full_meals_eaten    10000 non-null  int64
 23  vitD_supp           10000 non-null  int64
 24  Soft_drink          10000 non-null  object
 25  Initial_admin       10000 non-null  object
 26  HighBlood           10000 non-null  object
 27  Stroke              10000 non-null  object
 28  Complication_risk   10000 non-null  object
 29  Overweight          10000 non-null  object
 30  Arthritis           10000 non-null  object
 31  Diabetes            10000 non-null  object
 32  Hyperlipidemia      10000 non-null  object
 33  BackPain            10000 non-null  object
 34  Anxiety             10000 non-null  object
 35  Allergic_rhinitis   10000 non-null  object
 36  Reflux_esophagitis  10000 non-null  object
 37  Asthma              10000 non-null  object
 38  Services            10000 non-null  object
 39  Initial_days        10000 non-null  float64
 40  TotalCharge         10000 non-null  float64
 41  Additional_charges  10000 non-null  float64
 42  Item1               10000 non-null  int64
 43  Item2               10000 non-null  int64
 44  Item3               10000 non-null  int64
 45  Item4               10000 non-null  int64
 46  Item5               10000 non-null  int64
 47  Item6               10000 non-null  int64
 48  Item7               10000 non-null  int64
 49  Item8               10000 non-null  int64
dtypes: float64(7), int64(16), object(27)
memory usage: 3.8+ MB
```

In [4]:

```
msno.matrix(data)
```

Out[4]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bf851ec8>
```



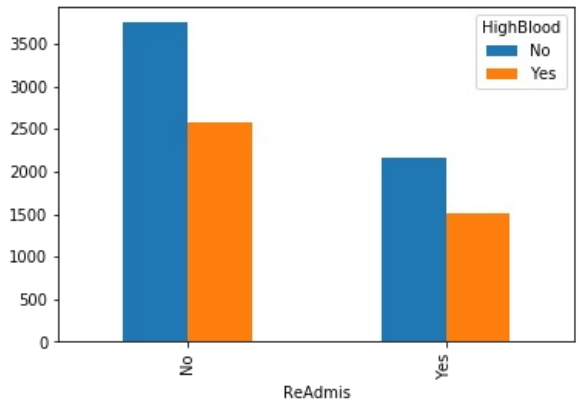
In [5]:

```
#Step B1 - pandas.crosstab ReAdmis by test variables.
# Computes a simple crosstabulation of the two factors. citation: (pandas.crosstab.  n.d.)
#HighBlood
ReHi = pd.crosstab(data['ReAdmis'], data['HighBlood'])
print(ReHi)
ReHi.plot(kind='bar')
```

```
HighBlood    No   Yes
ReAdmis
No         3747  2584
Yes        2163  1506
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bfd8ffc8>
```



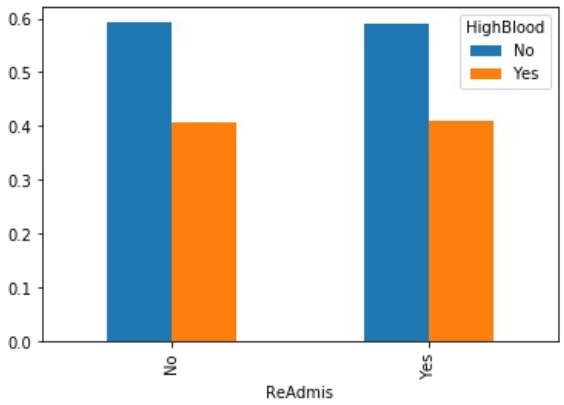
In [6]:

```
ReHiN = pd.crosstab(data['ReAdmis'], data['HighBlood'], normalize='index')
print(ReHiN)
ReHiN.plot(kind='bar')
```

```
HighBlood        No       Yes
ReAdmis
No         0.591850  0.408150
Yes        0.589534  0.410466
```

Out[6]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bfe38748>
```



In [7]:

```
# chi2_contingency with interpretation code reference (Sewell, n.d.)
stat, p, dof, expected = chi2_contingency(ReHi)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.04239657973011679
p-value = 0.8368656684578771
dof = 1
expected distribution
 [[3741.621 2589.379]
 [2168.379 1500.621]]
probability=0.90, critical=2.71, stat=0.04
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.837
Independent.  Fail to reject null hypothesis
```

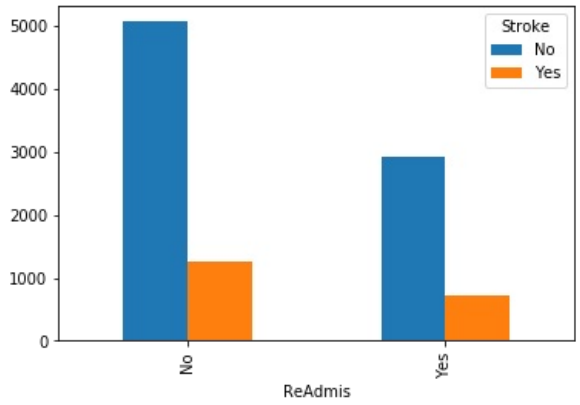
In [8]:

```
#Stroke
ReSt = pd.crosstab(data['ReAdmis'], data['Stroke'])
print(ReSt)
ReSt.plot(kind='bar')
```

```
Stroke     No   Yes
ReAdmis
No       5071  1260
Yes      2936   733
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bfeb8d08>
```



In [ ]:

```
```

In [9]:

```
stat, p, dof, expected = chi2_contingency(ReSt)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.004339571341125097
p-value = 0.9474770077616069
dof = 1
expected distribution
 [[5069.2317 1261.7683]
 [2937.7683  731.2317]]
probability=0.90, critical=2.71, stat=0.00
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.947
Independent.  Fail to reject null hypothesis
```

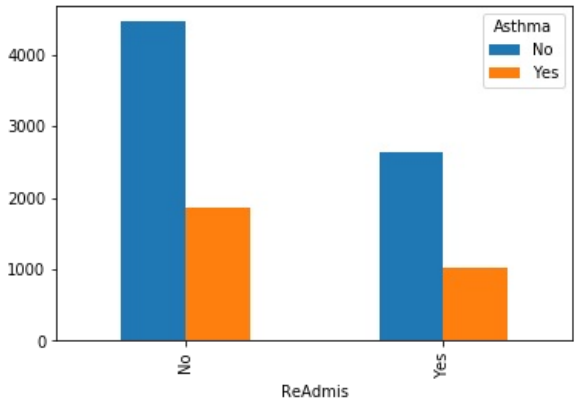
In [10]:

```
#Asthma
ReAs = pd.crosstab(data['ReAdmis'], data['Asthma'])
print(ReAs)
ReAs.plot(kind='bar')
```

```
Asthma     No   Yes
ReAdmis
No       4462  1869
Yes      2645  1024
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bff43e88>
```



In [11]:

```
stat, p, dof, expected = chi2_contingency(ReAs)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 2.857452237605144
p-value = 0.090951450679407
dof = 1
expected distribution
 [[4499.4417 1831.5583]
 [2607.5583 1061.4417]]
probability=0.90, critical=2.71, stat=2.86
Dependent.  Reject null hypothesis
significance=0.100, p=0.091
Dependent.  Reject null hypothesis
```

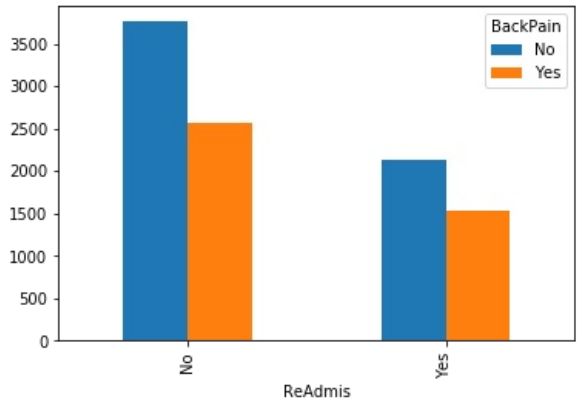
In [12]:

```
#BackPain
ReBa = pd.crosstab(data['ReAdmis'], data['BackPain'])
print(ReBa)
ReBa.plot(kind='bar')
```

```
BackPain    No   Yes
ReAdmis
No        3758  2573
Yes       2128  1541
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bffbf7c8>
```



In [13]:

```
stat, p, dof, expected = chi2_contingency(ReBa)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 1.7166153073233095
p-value = 0.1901293128457823
dof = 1
expected distribution
 [[3726.4266 2604.5734]
 [2159.5734 1509.4266]]
probability=0.90, critical=2.71, stat=1.72
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.190
Independent.  Fail to reject null hypothesis
```

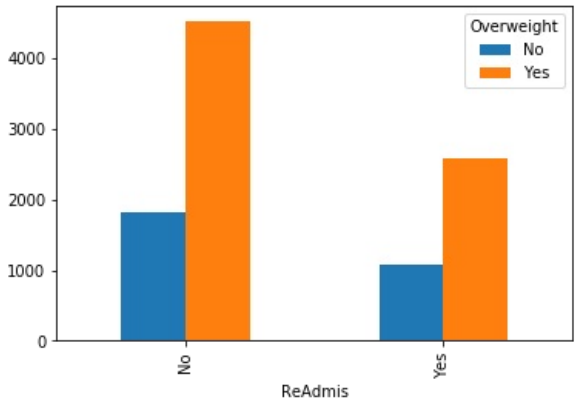
In [14]:

```
#Overweight
ReOv = pd.crosstab(data['ReAdmis'], data['Overweight'])
print(ReOv)
ReOv.plot(kind='bar')
```

```
Overweight    No   Yes
ReAdmis
No          1821  4510
Yes         1085  2584
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c0036dc8>
```



In [15]:

```
stat, p, dof, expected = chi2_contingency(ReOv)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.6984802059617877
p-value = 0.4032948387365497
dof = 1
expected distribution
 [[1839.7886 4491.2114]
 [1066.2114 2602.7886]]
probability=0.90, critical=2.71, stat=0.70
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.403
Independent.  Fail to reject null hypothesis
```

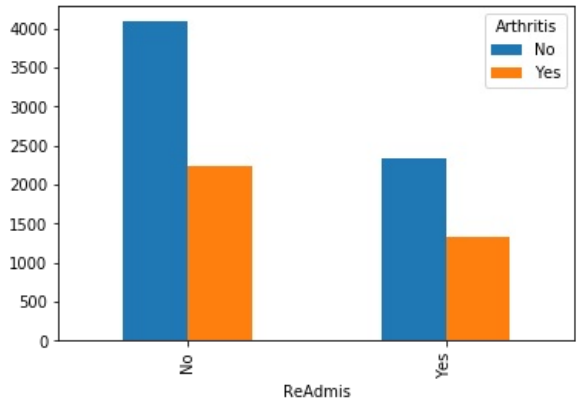
In [16]:

```
#Arthritis
ReAr = pd.crosstab(data['ReAdmis'], data['Arthritis'])
print(ReAr)
ReAr.plot(kind='bar')
```

```
Arthritis    No   Yes
ReAdmis
No         4086  2245
Yes        2340  1329
```

Out[16]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c00aa0c8>
```



In [17]:

```
stat, p, dof, expected = chi2_contingency(ReAr)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.5545124468934712
p-value = 0.4564797501244029
dof = 1
expected distribution
 [[4068.3006 2262.6994]
 [2357.6994 1311.3006]]
probability=0.90, critical=2.71, stat=0.55
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.456
Independent.  Fail to reject null hypothesis
```

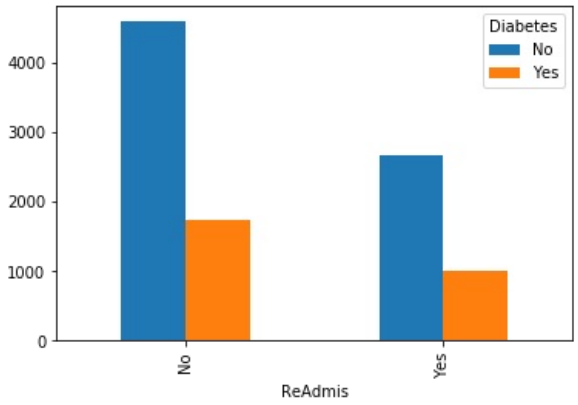
In [18]:

```
#Diabetes
ReDi = pd.crosstab(data['ReAdmis'], data['Diabetes'])
print(ReDi)
ReDi.plot(kind='bar')
```

```
Diabetes    No   Yes
ReAdmis
No        4591  1740
Yes       2671   998
```

Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c0128c08>
```



In [19]:

```
stat, p, dof, expected = chi2_contingency(ReDi)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.07983299478166737
p-value = 0.7775238586920562
dof = 1
expected distribution
 [[4597.5722 1733.4278]
 [2664.4278 1004.5722]]
probability=0.90, critical=2.71, stat=0.08
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.778
Independent.  Fail to reject null hypothesis
```

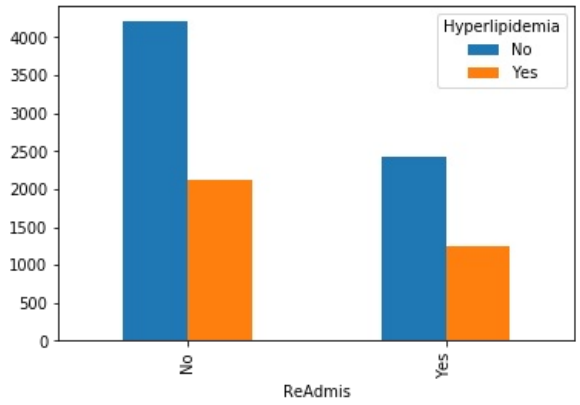
In [20]:

```
#Hyperlipidemia
ReH = pd.crosstab(data['ReAdmis'], data['Hyperlipidemia'])
print(ReH)
ReH.plot(kind='bar')
```

```
Hyperlipidemia    No   Yes
ReAdmis
No              4206  2125
Yes             2422  1247
```

Out[20]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c01a0248>
```



In [21]:

```
stat, p, dof, expected = chi2_contingency(ReH)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.16707357590519195
p-value = 0.6827258178039075
dof = 1
expected distribution
 [[4196.1868 2134.8132]
 [2431.8132 1237.1868]]
probability=0.90, critical=2.71, stat=0.17
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.683
Independent.  Fail to reject null hypothesis
```

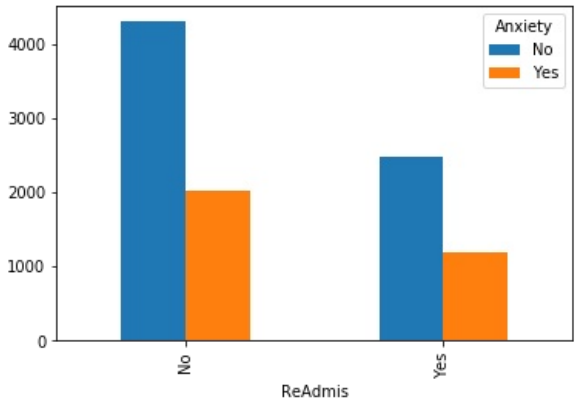
In [22]:

```
#Anxiety
ReA = pd.crosstab(data['ReAdmis'], data['Anxiety'])
print(ReA)
ReA.plot(kind='bar')
```

```
Anxiety    No   Yes
ReAdmis
No       4301  2030
Yes      2484  1185
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c02176c8>
```



In [23]:

```
stat, p, dof, expected = chi2_contingency(ReA)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.047704706965444724
p-value = 0.8271065021209385
dof = 1
expected distribution
 [[4295.5835 2035.4165]
 [2489.4165 1179.5835]]
probability=0.90, critical=2.71, stat=0.05
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.827
Independent.  Fail to reject null hypothesis
```

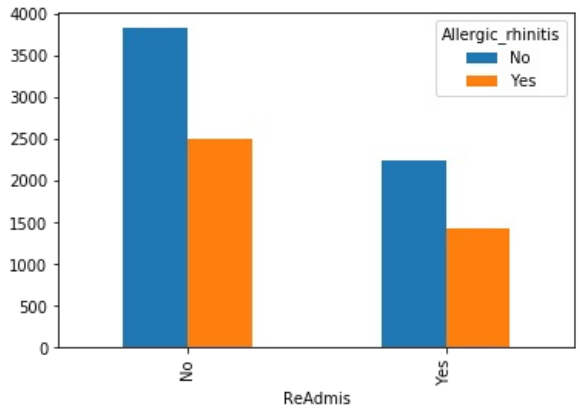
In [24]:

```
#Allergic_rhinitis
ReR = pd.crosstab(data['ReAdmis'], data['Allergic_rhinitis'])
print(ReR)
ReR.plot(kind='bar')
```

```
Allergic_rhinitis    No   Yes
ReAdmis
No                 3825  2506
Yes                2234  1435
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3c02c9488>
```



In [25]:

```
stat, p, dof, expected = chi2_contingency(ReR)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.19699090646919878
p-value = 0.6571607419560208
dof = 1
expected distribution
 [[3835.9529 2495.0471]
 [2223.0471 1445.9529]]
probability=0.90, critical=2.71, stat=0.20
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.657
Independent.  Fail to reject null hypothesis
```

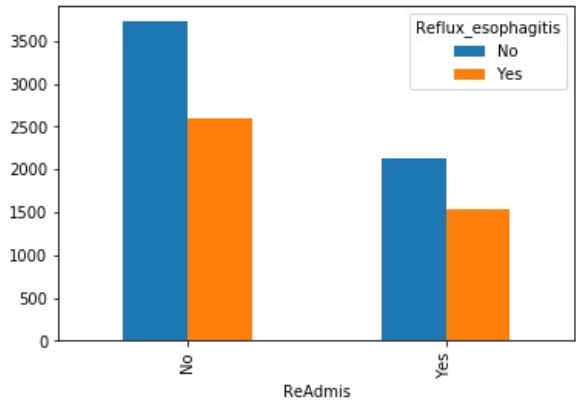
In [26]:

```
#Reflux_esophagitis
ReRe = pd.crosstab(data['ReAdmis'], data['Reflux_esophagitis'])
print(ReRe)
ReRe.plot(kind='bar')
```

```
Reflux_esophagitis    No   Yes
ReAdmis
No                  3726  2605
Yes                 2139  1530
```

Out[26]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c3bff3bac8>
```



In [27]:

```
stat, p, dof, expected = chi2_contingency(ReRe)
print('chi-square statistic =', stat,'\np-value =', p, '\ndof =', dof, '\nexpected distribution', '\n',expected)
#Interpret the chi squre test results
#Analyze test statistic
prob = .9
critical = chi2.ppf(prob, dof)
print('probability=%.2f, critical=%.2f, stat=%.2f' % (prob, critical, stat))
if abs(stat) >= critical:
    print ('Dependent.  Reject null hypothesis')
else:
    print ('Independent.  Fail to reject null hypothesis')
#Analyze p-value
alpha = 1 - prob
print('significance=%.3f, p=%.3f' % (alpha, p))
if p < alpha:
    print('Dependent.  Reject null hypothesis')
else:
    print('Independent.  Fail to reject null hypothesis')
```

```
chi-square statistic = 0.2715629458729544
p-value = 0.6022852568865315
dof = 1
expected distribution
 [[3713.1315 2617.8685]
 [2151.8685 1517.1315]]
probability=0.90, critical=2.71, stat=0.27
Independent.  Fail to reject null hypothesis
significance=0.100, p=0.602
Independent.  Fail to reject null hypothesis
```