Exploratory Data Analysis - DrugsTypes VS Ratings

```python
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

df = pd.read_excel("/content/DrugData.xlsx")

df.head(5)
```

```
                   drugName                      condition  \
0  206461           Valsartan  Left Ventricular Dysfunction
1   95260          Guanfacine                          ADHD
2   92703              Lybrel                  Birth Control
3  138000           Ortho Evra                 Birth Control
4   35696  Buprenorphine / naloxone          Opiate Dependence


                                            review  rating  \
0  "It has no side effect, I take it in combinati...       9
1  "My son is halfway through his fourth week of ...       8
2  "I used to take another oral contraceptive, wh...       5
3  "This is my first time using any form of birth...       8
4  "Suboxone has completely turned my life around...       9


                date  usefulCount
0       May 20, 2012           27
1     April 27, 2010          192
2  December 14, 2009           17
3   November 3, 2015           10
4  November 27, 2016           37
```

```python
df.shape
```

```
(161297, 7)
```

```python
df.columns
```

```
Index([' ', 'drugName', 'condition', 'review', 'rating', 'date',
       'usefulCount'],
      dtype='object')
```

```python
df1 = df.drop(df.columns[[0]],axis = 1)

df1.columns
```

```
Index(['drugName', 'condition', 'review', 'rating', 'date',
'usefulCount'], dtype='object')
```

```python
df1.isnull().sum()
```

```
drugName          0
condition       899
review            0
rating            0
date              0
usefulCount       0
dtype: int64

df1

                        drugName                     condition  \
0                      Valsartan  Left Ventricular Dysfunction
1                     Guanfacine                          ADHD
2                         Lybrel                 Birth Control
3                      Ortho Evra                 Birth Control
4          Buprenorphine / naloxone          Opiate Dependence
...                          ...                           ...
20150                 Etonogestrel                 Birth Control
20151              Junel Fe 1 / 20                 Birth Control
20152                        Xanax                       Anxiety
20153                     Orlistat                       Obesity
20154                   Alprazolam                           NaN

                                                   review  rating  \
0       "It has no side effect, I take it in combinati...     9.0
1       "My son is halfway through his fourth week of ...     8.0
2       "I used to take another oral contraceptive, wh...     5.0
3       "This is my first time using any form of birth...     8.0
4       "Suboxone has completely turned my life around...     9.0
...                                                   ...     ...
20150   "So far, I love the implanon! I read so many h...     9.0
20151   "I have been on this birth control for over 3 ...    10.0
20152   "Hi everyone!\n\nI live in Australia...\n\nI h...    10.0
20153   "I think most people expect changes in body to...     8.0
20154                                                 NaN     NaN

                   date  usefulCount
0          May 20, 2012         27.0
1        April 27, 2010        192.0
2     December 14, 2009         17.0
3      November 3, 2015         10.0
4     November 27, 2016         37.0
...                 ...          ...
20150     April 12, 2015         34.0
20151   October 17, 2017          4.0
20152   October 19, 2016         19.0
20153      July 12, 2016         43.0
20154                NaN          NaN

[20155 rows x 6 columns]
```

```
df1['drugName'].value_counts()

Levonorgestrel                       3657
Etonogestrel                         3336
Ethinyl estradiol / norethindrone    2850
Nexplanon                            2156
Ethinyl estradiol / norgestimate     2117
                                      ...
Omnipaque 350                           1
Vontrol                                 1
Ivabradine                              1
Neo-Poly-Dex                            1
Grifulvin V                             1
Name: drugName, Length: 3436, dtype: int64

df['drugName'].value_counts().nlargest(20)

Levonorgestrel                       3657
Etonogestrel                         3336
Ethinyl estradiol / norethindrone    2850
Nexplanon                            2156
Ethinyl estradiol / norgestimate     2117
Ethinyl estradiol / levonorgestrel   1888
Phentermine                          1543
Sertraline                           1360
Escitalopram                         1292
Mirena                               1242
Implanon                             1102
Gabapentin                           1047
Bupropion                            1022
Venlafaxine                          1016
Miconazole                           1000
Medroxyprogesterone                   995
Citalopram                            995
Lexapro                               952
Bupropion / naltrexone                950
Duloxetine                            934
Name: drugName, dtype: int64

df['drugName'].value_counts().nsmallest(20)

Clemastine                       1
Topicort LP                      1
Smoothie Readi-Cat 2             1
Ipratropium Inhalation Solution  1
Olux-E                           1
Hexachlorophene                  1
Pedi-Dri                         1
Ginseng                          1
Lincocin                         1
EnLyte                           1
```
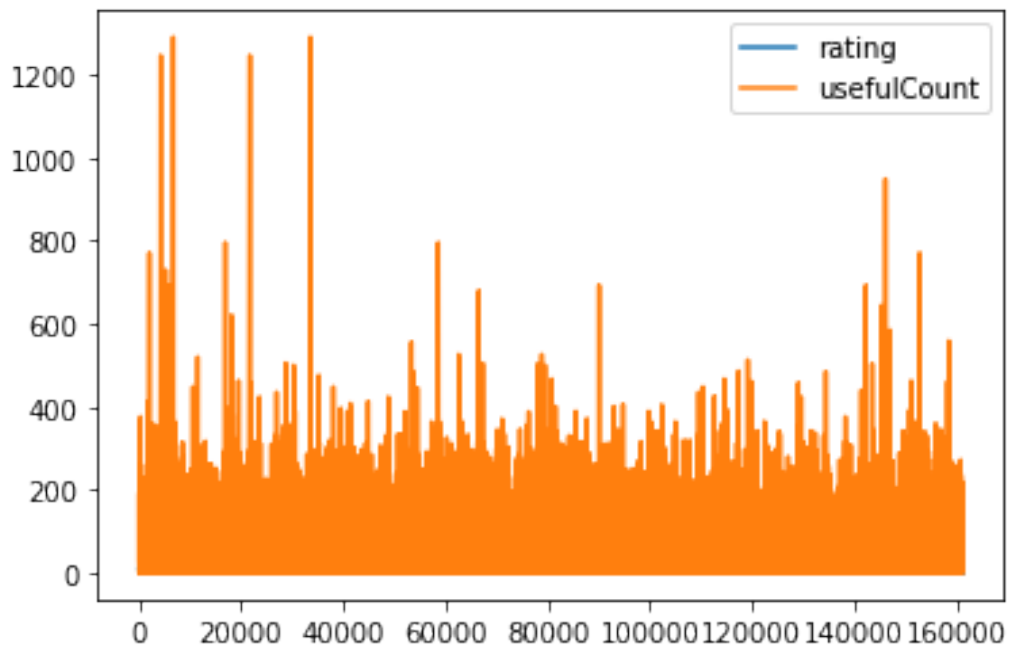
```
Tums Smoothies                  1
Purinethol                      1
Aldomet                         1
Midol Extended Relief           1
Travel-Eze                      1
Ponatinib                       1
Wal-finate                      1
Ciclodan                        1
Niacinamide                     1
Fragmin                         1
Name: drugName, dtype: int64
```
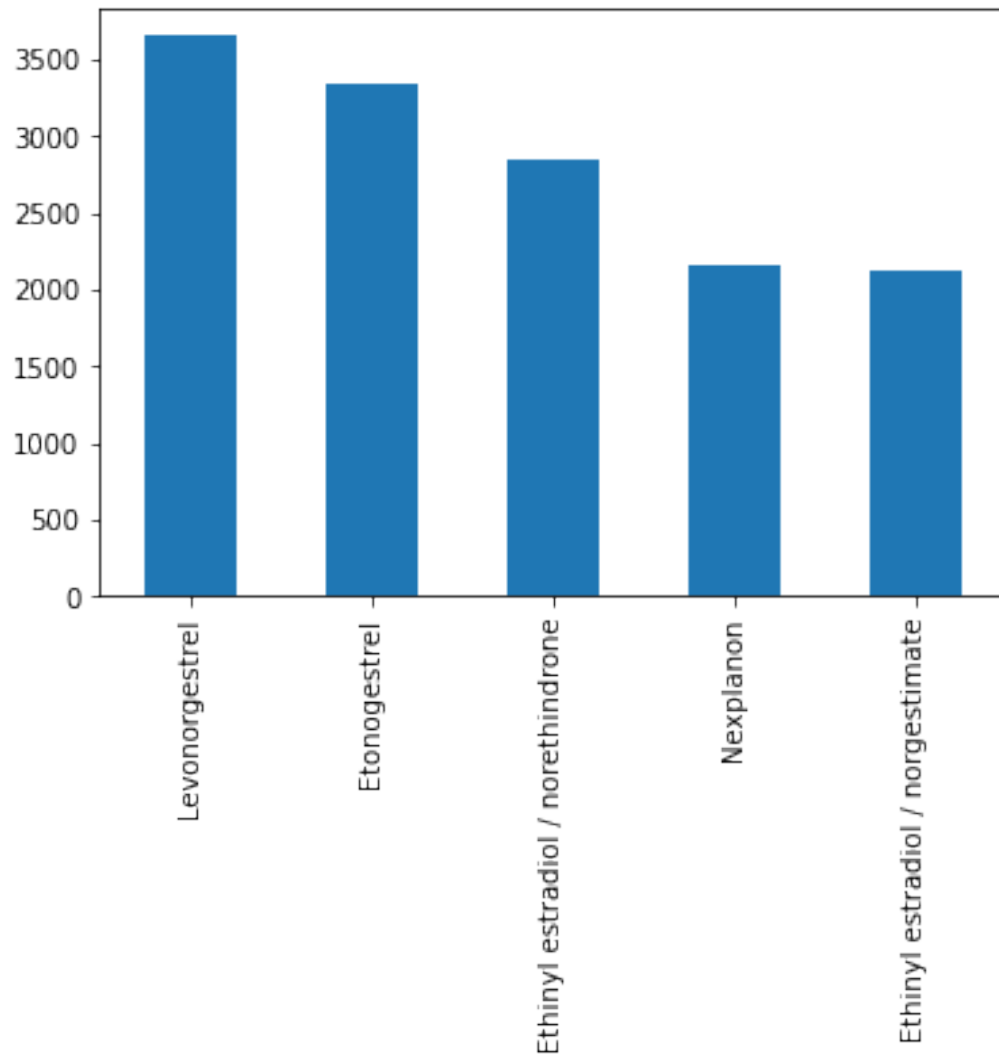
```python
df1.plot()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8c26b71250>
```



```python
df1['drugName'].value_counts().nlargest(5).plot(kind='bar')
```
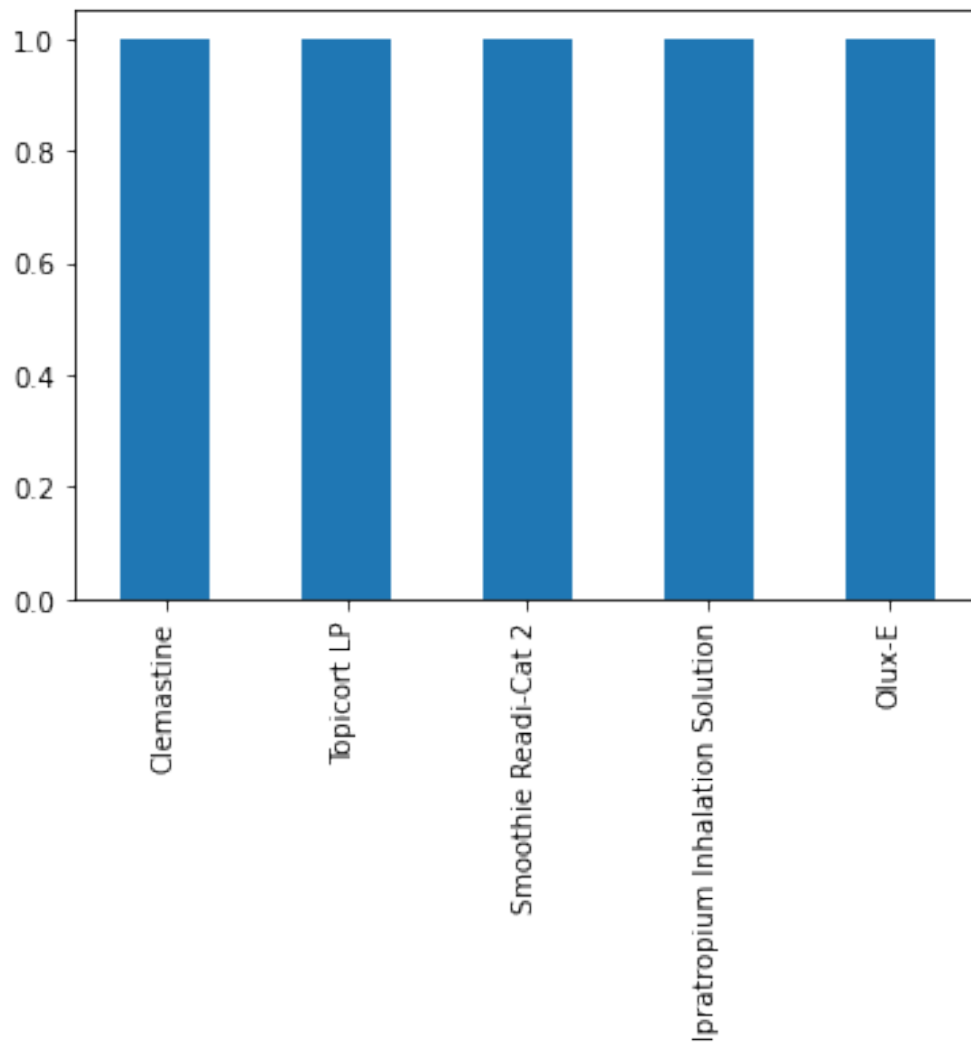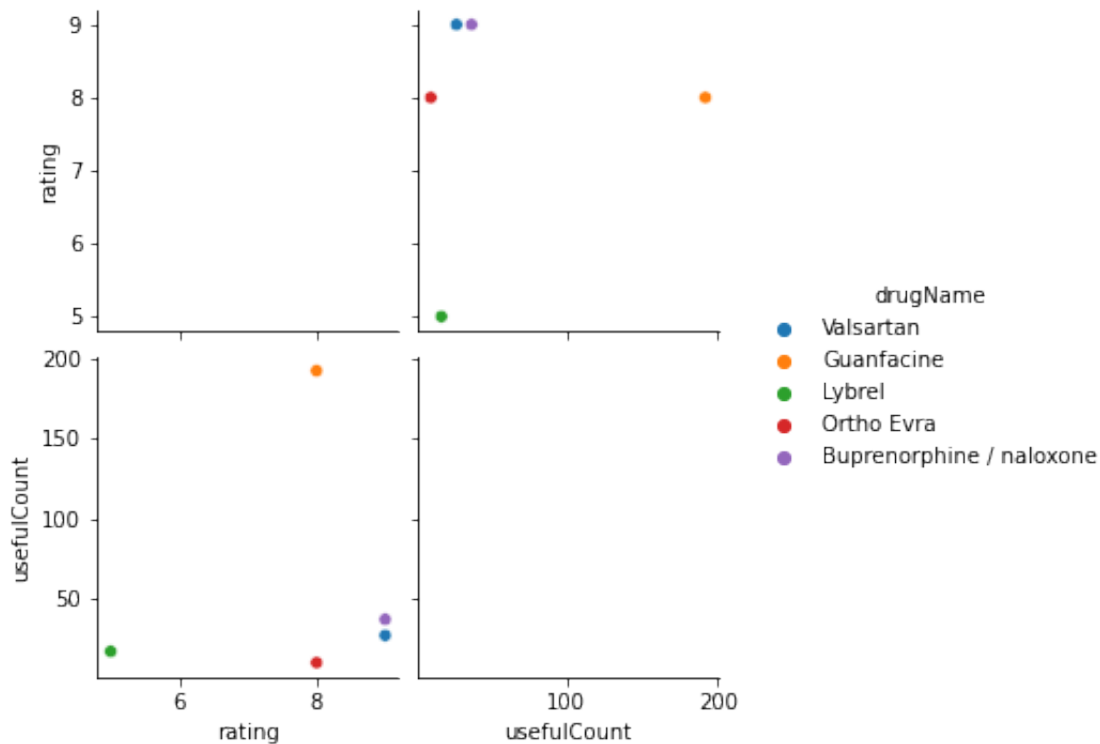
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8c255cecd0>
```

```
df1['drugName'].value_counts().nsmallest(5).plot(kind='bar')
```
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8c25a18e90>
```

```
sns.pairplot(df1.head(), hue = 'drugName')
```

`<seaborn.axisgrid.PairGrid at 0x7f8c25661090>`

Narrative : We have 885 different conditions

```
len(df['condition'].unique().tolist())
```

885

```
df1['condition'].value_counts()
```

```
Birth Control                                             28788
Depression                                                 9069
Pain                                                       6145
Anxiety                                                    5904
Acne                                                       5588
                                                           ...
Dissociative Identity Disorde                                 1
Hydrocephalus                                                 1
Hyperlipoproteinemia Type III, Elevated beta-VLDL    IDL      1
Q Feve                                                        1
Neutropenia                                                   1
Name: condition, Length: 884, dtype: int64
```
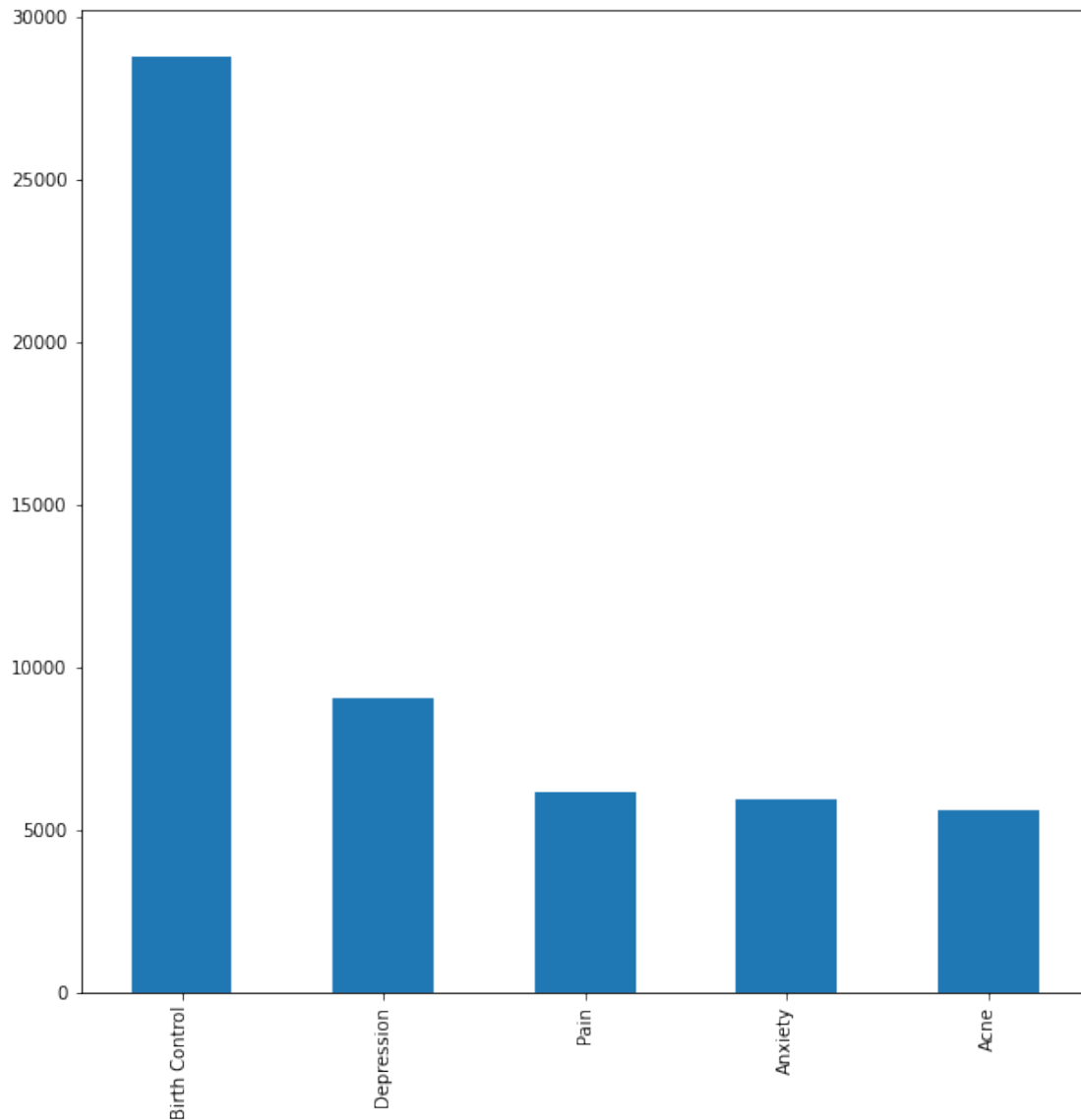
```
df['condition'].value_counts().nlargest(5)
```

```
Birth Control    28788
Depression        9069
Pain              6145
Anxiety           5904
```

```
Acne                  5588
Name: condition, dtype: int64
```

```python
df['condition'].value_counts().nlargest(5).plot(kind='bar',figsize=(10
,10))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8c1a7e2e10>
```



```python
df1.groupby('condition')['drugName'].nunique().nlargest(30)
```

```
condition
Not Listed / Othe                    214
Pain                                 200
Birth Control                        172
High Blood Pressure                  140
Acne                                 117
Depression                           105
```
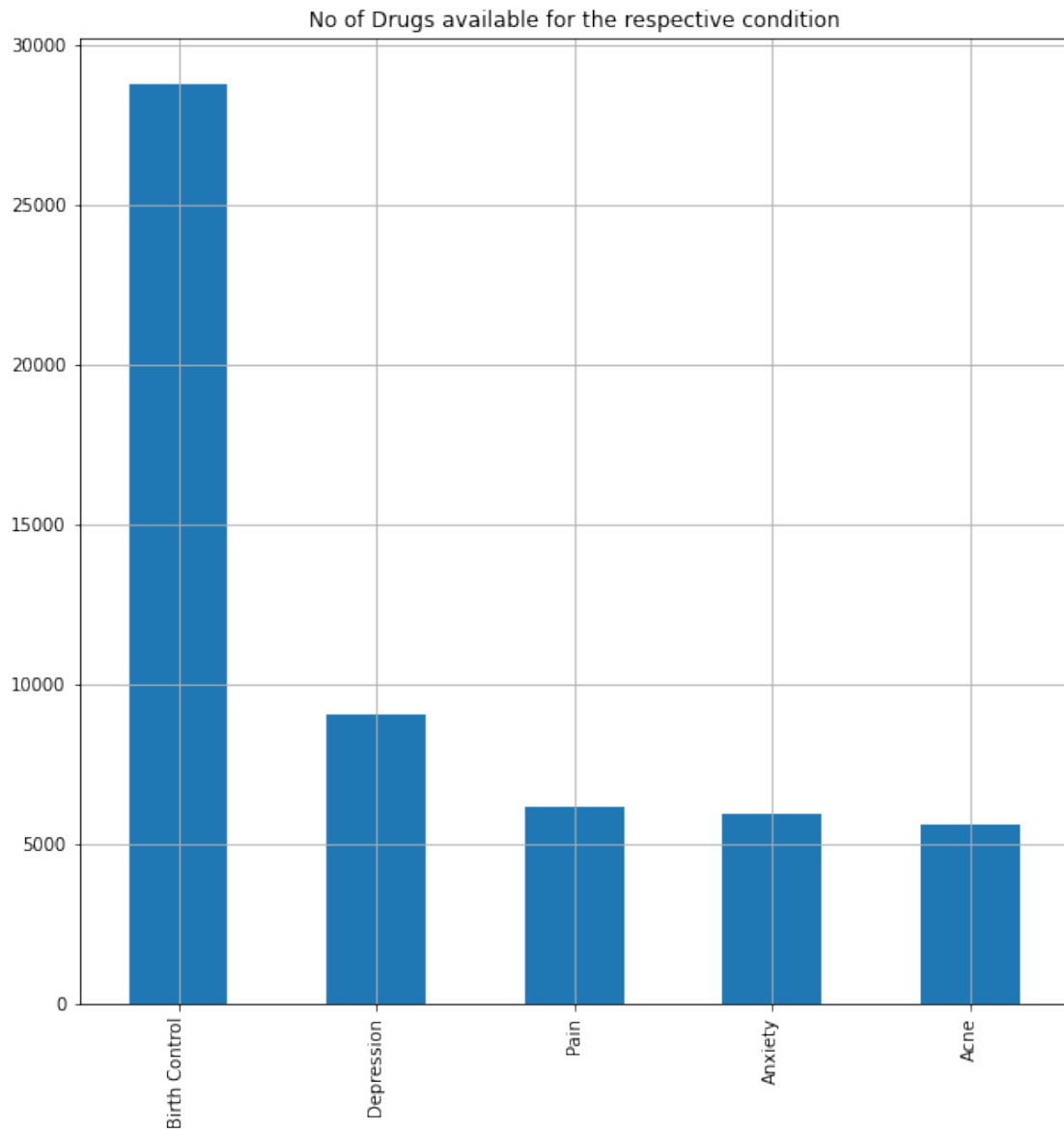
```
Rheumatoid Arthritis                            98
Diabetes, Type 2                                89
Allergic Rhinitis                               88
Bipolar Disorde                                 80
Osteoarthritis                                  80
Anxiety                                         78
Insomnia                                        78
Abnormal Uterine Bleeding                       74
Migraine                                        59
Psoriasis                                       58
3</span> users found this comment helpful.     57
Endometriosis                                   57
ADHD                                            55
Asthma, Maintenance                             54
Chronic Pain                                    53
Migraine Prevention                             50
Irritable Bowel Syndrome                        49
Major Depressive Disorde                        49
Urinary Tract Infection                         47
4</span> users found this comment helpful.     45
ibromyalgia                                     45
Bronchitis                                      44
Postmenopausal Symptoms                         44
2</span> users found this comment helpful.     43
Name: drugName, dtype: int64
```

```python
df['condition'].value_counts().nlargest(5).plot(kind='bar',figsize=(10
,10))

plt.figure(figsize=(10,10))
df['condition'].value_counts().nlargest(5).plot(kind='bar',figsize=(10
,10))
plt.title("No of Drugs available for the respective condition")
plt.grid()
plt.show()
```

No of Drugs available for the respective condition

```
df1['rating']

0            9
1            8
2            5
3            8
4            9
          ..
161292      10
161293       1
161294       2
161295      10
161296       9
Name: rating, Length: 161297, dtype: int64
```
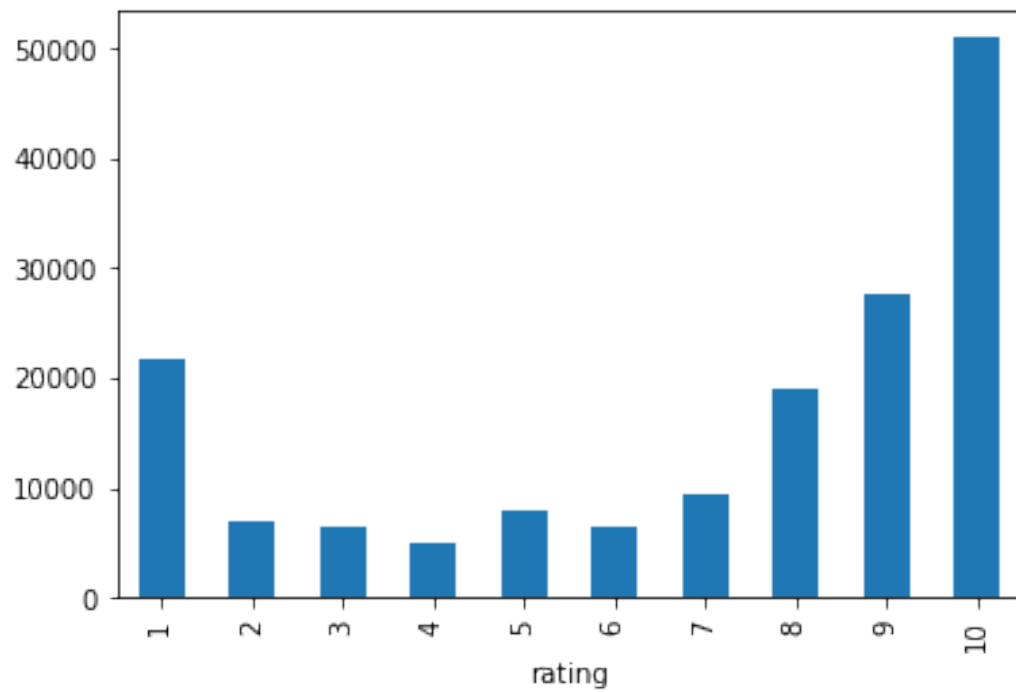
```
df1.groupby('rating').size().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8c1b409b50>
```


Thanks
Rakshit G B