

# Exploratory Data Analysis on Drugs, Side Effects, and Medical Conditions

November 16, 2024

## 1 About Me

Mohammed Imam Uddin Riad  
Data Analyst

Connect With Me - [Portfolio](#) - [LinkedIn](#) - [Email Me](#)

## 2 Project:Exploratory Data Analysis on Drugs, Side Effects, and Medical Conditions

```
[1]: import pandas as pd
```

```
[2]: raw_data = pd.read_csv("drugs_side_effects_drugs_com.csv")
df = raw_data.copy()
df.head()
```

```
[2]:      drug_name medical_condition \
0      doxycycline                Acne
1  spironolactone                Acne
2    minocycline                Acne
3      Accutane                Acne
4   clindamycin                Acne

                                side_effects      generic_name \
0  (hives, difficult breathing, swelling in your ...      doxycycline
1  hives ; difficulty breathing; swelling of your...  spironolactone
2  skin rash, fever, swollen glands, flu-like sym...    minocycline
3  problems with your vision or hearing; muscle o... isotretinoin (oral)
4  hives ; difficult breathing; swelling of your ... clindamycin topical

                                drug_classes \
0      Miscellaneous antimalarials, Tetracyclines
1  Aldosterone receptor antagonists, Potassium-sp...
2                                Tetracyclines
3  Miscellaneous antineoplastics, Miscellaneous u...
4      Topical acne agents, Vaginal anti-infectives
```

```

                                brand_names activity rx_otc  \
0  Acticlate, Adoxa CK, Adoxa Pak, Adoxa TT, Alod...      87%    Rx
1                                Aldactone, CaroSpir      82%    Rx
2  Dynacin, Minocin, Minolira, Solodyn, Ximino, V...      48%    Rx
3                                NaN      41%    Rx
4  Cleocin T, Clindacin ETZ, Clindacin P, Clindag...      39%    Rx

pregnancy_category csa alcohol  \
0                D    N      X
1                C    N      X
2                D    N    NaN
3                X    N      X
4                B    N    NaN

                                related_drugs  \
0  amoxicillin: https://www.drugs.com/amoxicillin...
1  amlodipine: https://www.drugs.com/amlodipine.h...
2  amoxicillin: https://www.drugs.com/amoxicillin...
3  doxycycline: https://www.drugs.com/doxycycline...
4  doxycycline: https://www.drugs.com/doxycycline...

                                medical_condition_description rating no_of_reviews  \
0  Acne Other names: Acne Vulgaris; Blackheads; B...      6.8      760.0
1  Acne Other names: Acne Vulgaris; Blackheads; B...      7.2      449.0
2  Acne Other names: Acne Vulgaris; Blackheads; B...      5.7      482.0
3  Acne Other names: Acne Vulgaris; Blackheads; B...      7.9      623.0
4  Acne Other names: Acne Vulgaris; Blackheads; B...      7.4      146.0

                                drug_link  \
0                https://www.drugs.com/doxycycline.html
1                https://www.drugs.com/spironolactone.html
2                https://www.drugs.com/minocycline.html
3                https://www.drugs.com/accutane.html
4  https://www.drugs.com/mtm/clindamycin-topical...

                                medical_condition_url
0  https://www.drugs.com/condition/acne.html
1  https://www.drugs.com/condition/acne.html
2  https://www.drugs.com/condition/acne.html
3  https://www.drugs.com/condition/acne.html
4  https://www.drugs.com/condition/acne.html

```

```
[3]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2931 entries, 0 to 2930
Data columns (total 17 columns):

```

#	Column	Non-Null Count	Dtype
0	drug_name	2931 non-null	object
1	medical_condition	2931 non-null	object
2	side_effects	2807 non-null	object
3	generic_name	2888 non-null	object
4	drug_classes	2849 non-null	object
5	brand_names	1718 non-null	object
6	activity	2931 non-null	object
7	rx_otc	2930 non-null	object
8	pregnancy_category	2702 non-null	object
9	csa	2931 non-null	object
10	alcohol	1377 non-null	object
11	related_drugs	1462 non-null	object
12	medical_condition_description	2931 non-null	object
13	rating	1586 non-null	float64
14	no_of_reviews	1586 non-null	float64
15	drug_link	2931 non-null	object
16	medical_condition_url	2931 non-null	object

dtypes: float64(2), object(15)  
memory usage: 389.4+ KB

```
[4]: df.describe()
```

```
[4]:
```

	rating	no_of_reviews
count	1586.000000	1586.000000
mean	6.812926	75.064943
std	2.313044	184.900847
min	0.000000	1.000000
25%	5.600000	2.000000
50%	7.000000	12.000000
75%	8.500000	58.000000
max	10.000000	2934.000000

```
[5]: df.describe(include= 'object').T
```

```
[5]:
```

	count	unique	\
drug_name	2931	2912	
medical_condition	2931	47	
side_effects	2807	2759	
generic_name	2888	1392	
drug_classes	2849	274	
brand_names	1718	1552	
activity	2931	93	
rx_otc	2930	3	
pregnancy_category	2702	6	
csa	2931	7	
alcohol	1377	1	

related_drugs	1462	535
medical_condition_description	2931	47
drug_link	2931	2931
medical_condition_url	2931	47

top

\	
drug_name	triamcinolone
medical_condition	Pain
side_effects	hives ; difficult breathing; swelling of your ...
generic_name	diphenhydramine
drug_classes	Upper respiratory combinations
brand_names	Acne-Clear, Benzac AC, BenzePrO, Benziq, Brevo...
activity	0%
rx_otc	Rx
pregnancy_category	C
csa	N
alcohol	X
related_drugs	doxycycline: <a href="https://www.drugs.com/doxycycline...">https://www.drugs.com/doxycycline...</a>
medical_condition_description	Pain An unpleasant sensory and emotional exper...
drug_link	<a href="https://www.drugs.com/doxycycline.html">https://www.drugs.com/doxycycline.html</a>
medical_condition_url	<a href="https://www.drugs.com/condition/pain.html">https://www.drugs.com/condition/pain.html</a>

	freq
drug_name	3
medical_condition	264
side_effects	10
generic_name	17
drug_classes	245
brand_names	10
activity	895
rx_otc	1998
pregnancy_category	1382
csa	2688
alcohol	1377
related_drugs	41
medical_condition_description	264
drug_link	1
medical_condition_url	264

```
[6]: df.duplicated().any()
```

```
[6]: False
```

```
[7]: df.shape
```

```
[7]: (2931, 17)
```

```
[8]: df.isnull().sum()
```

```
[8]: drug_name          0
     medical_condition  0
     side_effects      124
     generic_name       43
     drug_classes       82
     brand_names      1213
     activity           0
     rx_otc             1
     pregnancy_category 229
     csa                0
     alcohol          1554
     related_drugs     1469
     medical_condition_description 0
     rating            1345
     no_of_reviews     1345
     drug_link          0
     medical_condition_url 0
     dtype: int64
```

### 2.0.1 Handling Missing Values in the rating Column by Replacing with Mean

```
[9]: df['rating'] = df['rating'].fillna(df['rating'].mean())
```

```
[10]: df['rating']
```

```
[10]: 0      6.800000
     1      7.200000
     2      5.700000
     3      7.900000
     4      7.400000
     ...
     2926   7.500000
     2927   6.812926
     2928   6.812926
     2929   6.812926
     2930   6.812926
     Name: rating, Length: 2931, dtype: float64
```

### Handling Missing Values in the no\_of\_rating Column by Replacing with Min

```
[11]: df['no_of_reviews'] = df['no_of_reviews'].fillna(df['no_of_reviews'].min())
```

### Replacing Missing Values with 'Unknown' in the DataFrame

```
[12]: df.fillna('unknown',inplace = True)
```

```
[13]: df.isnull().sum()
```

```
[13]: drug_name          0
      medical_condition  0
      side_effects      0
      generic_name       0
      drug_classes       0
      brand_names        0
      activity           0
      rx_otc             0
      pregnancy_category 0
      csa                0
      alcohol            0
      related_drugs       0
      medical_condition_description 0
      rating             0
      no_of_reviews       0
      drug_link           0
      medical_condition_url 0
      dtype: int64
```

## 2.0.2 Chaning Datatype of Activity Column

```
[14]: df['activity'].dtype
```

```
[14]: dtype('O')
```

```
[15]: df['activity'] = df['activity'].replace('%', '', regex=True).astype(int)
```

```
[16]: df['activity'].dtype
```

```
[16]: dtype('int32')
```

## 2.0.3 create a checkpoint

```
[17]: df_cleaned = df.copy()
```

## 2.1 EDA

Count the most common drugs for each medical condition

```
[18]: df_cleaned.head(2)
```

```
[18]:      drug_name medical_condition \
0      doxycycline          Acne
1  spironolactone          Acne

      side_effects      generic_name \
0  (hives, difficult breathing, swelling in your ...  doxycycline
1  hives ; difficulty breathing; swelling of your...  spironolactone
```

```

                                drug_classes \
0      Miscellaneous antimalarials, Tetracyclines
1  Aldosterone receptor antagonists, Potassium-sp...

                                brand_names  activity rx_otc \
0  Acticlate, Adoxa CK, Adoxa Pak, Adoxa TT, Alod...      87    Rx
1                                Aldactone, CaroSpir      82    Rx

pregnancy_category csa alcohol \
0                D    N      X
1                C    N      X

                                related_drugs \
0  amoxicillin: https://www.drugs.com/amoxicillin...
1  amlodipine: https://www.drugs.com/amlodipine.h...

                                medical_condition_description  rating  no_of_reviews \
0  Acne Other names: Acne Vulgaris; Blackheads; B...      6.8        760.0
1  Acne Other names: Acne Vulgaris; Blackheads; B...      7.2        449.0

                                drug_link \
0      https://www.drugs.com/doxycycline.html
1  https://www.drugs.com/spironolactone.html

                                medical_condition_url
0  https://www.drugs.com/condition/acne.html
1  https://www.drugs.com/condition/acne.html

```

### 2.1.1 Top 10 Most Frequently Prescribed Drugs Across Medical Conditions

```

[19]: top_drugs = df.groupby('medical_condition')['drug_name'].value_counts().
      ↪nlargest(10)
      top_drugs

```

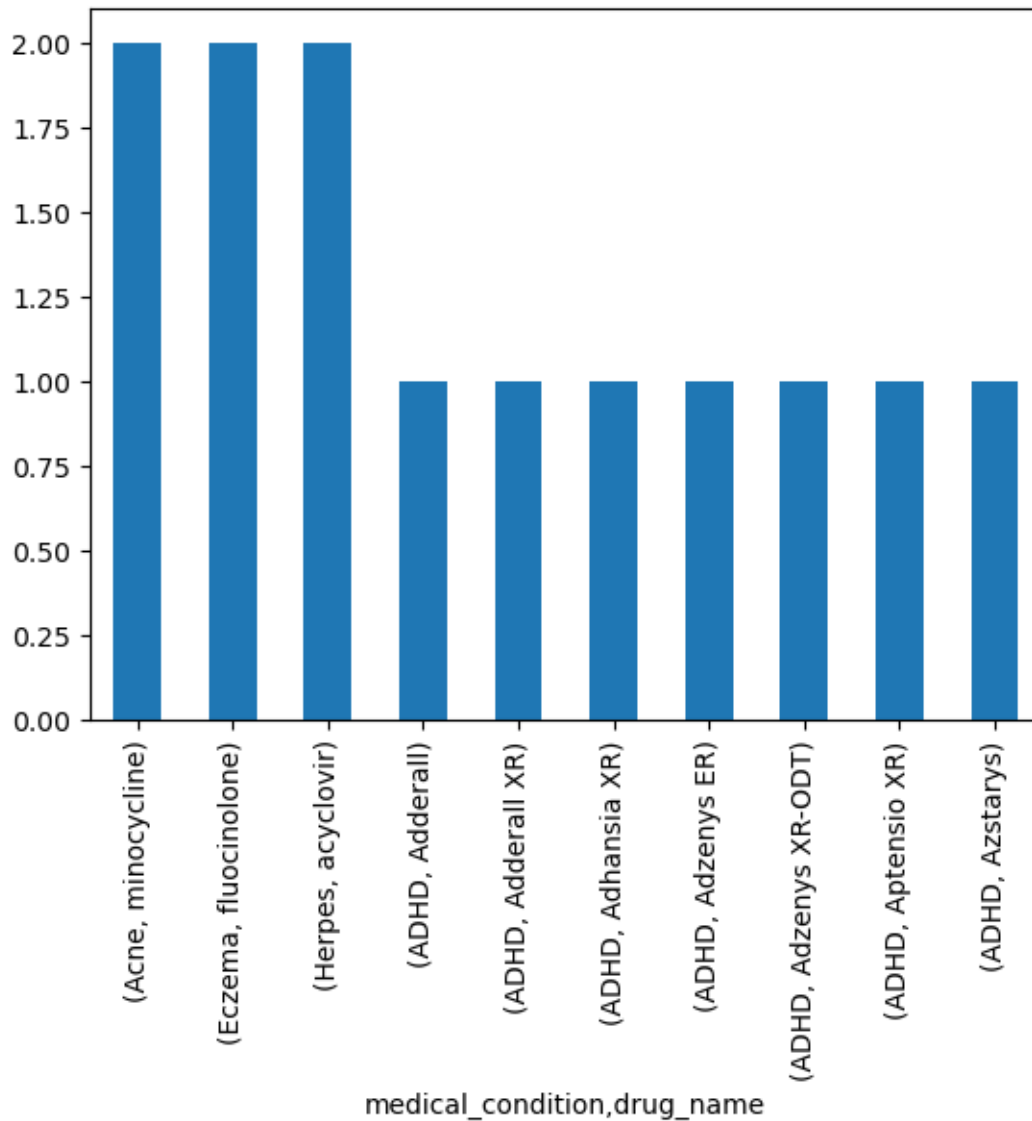
```

[19]: medical_condition  drug_name
Acne                    minocycline      2
Eczema                  fluocinolone      2
Herpes                  acyclovir         2
ADHD                    Adderall          1
                        Adderall XR       1
                        Adhansia XR       1
                        Adzenys ER        1
                        Adzenys XR-ODT    1
                        Aptensio XR       1
                        Azstarys          1
Name: count, dtype: int64

```

```
[20]: top_drugs.plot(kind = 'bar')
```

```
[20]: <Axes: xlabel='medical_condition,drug_name'>
```



### 2.1.2 Count of Drugs by Medical Condition by each medical condition

```
[21]: count_of_drugs = df_cleaned.groupby('medical_condition')['drug_name'].count().
      ↪sort_values(ascending=True )
count_of_drugs
```

```
[21]: medical_condition
Swine Flu
```

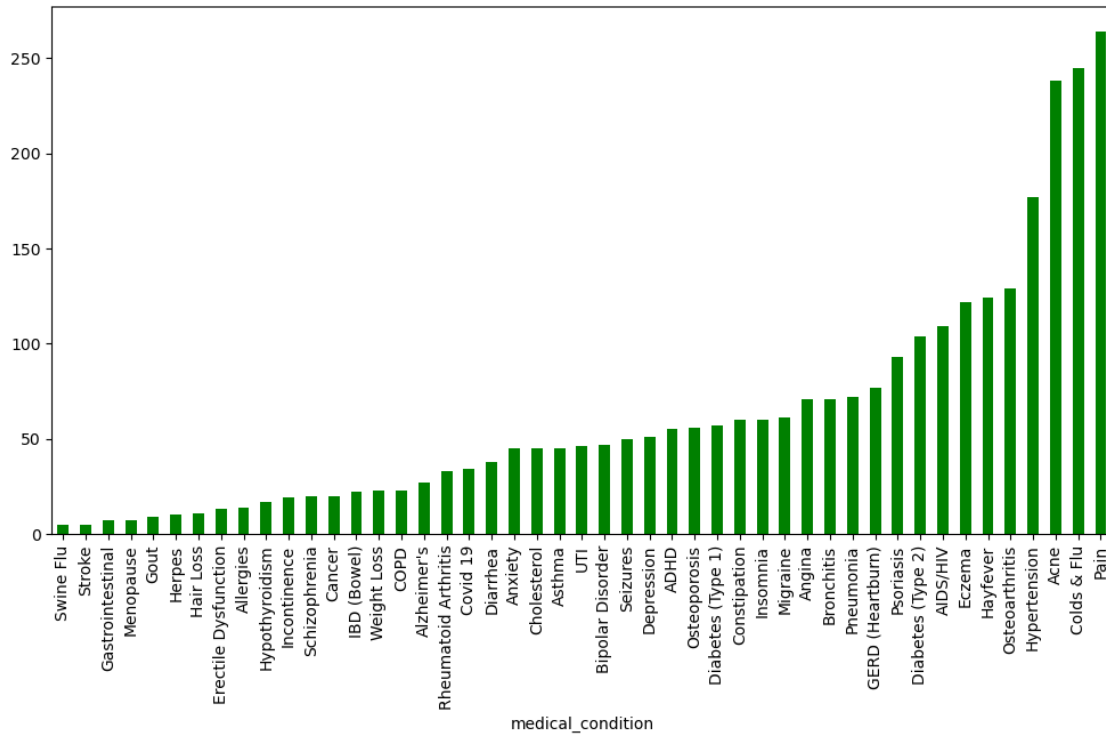
5



Stroke	5
Gastrointestinal	7
Menopause	7
Gout	9
Herpes	10
Hair Loss	11
Erectile Dysfunction	13
Allergies	14
Hypothyroidism	17
Incontinence	19
Schizophrenia	20
Cancer	20
IBD (Bowel)	22
Weight Loss	23
COPD	23
Alzheimer's	27
Rheumatoid Arthritis	33
Covid 19	34
Diarrhea	38
Anxiety	45
Cholesterol	45
Asthma	45
UTI	46
Bipolar Disorder	47
Seizures	50
Depression	51
ADHD	55
Osteoporosis	56
Diabetes (Type 1)	57
Constipation	60
Insomnia	60
Migraine	61
Angina	71
Bronchitis	71
Pneumonia	72
GERD (Heartburn)	77
Psoriasis	93
Diabetes (Type 2)	104
AIDS/HIV	109
Eczema	122
Hayfever	124
Osteoarthritis	129
Hypertension	177
Acne	238
Colds & Flu	245
Pain	264
Name: drug_name, dtype: int64	

```
[22]: count_of_drugs.plot(kind = 'bar',figsize=(12,6),color = 'green')
```

```
[22]: <Axes: xlabel='medical_condition'>
```



```
[23]: top_50_drug_class = df_cleaned['drug_classes'].value_counts().
      ↪sort_values(ascending = False).head(50)
top_50_drug_class
```

```
[23]: drug_classes
Upper respiratory combinations
245
Topical steroids
94
Topical acne agents
88
unknown
82
Nonsteroidal anti-inflammatory drugs
70
Topical rubefacient
56
Miscellaneous topical agents
56
```

Antihistamines  
55  
Topical anesthetics  
54  
Laxatives  
53  
Opioids (narcotic analgesics)  
51  
Antiviral combinations  
50  
Insulin  
50  
Analgesic combinations  
49  
Antidiabetic combinations  
41  
Urinary antispasmodics  
41  
Atypical antipsychotics  
41  
CNS stimulants  
39  
Miscellaneous analgesics  
38  
Antimigraine agents  
32  
Glucocorticoids  
30  
Miscellaneous anxiolytics, sedatives and hypnotics  
27  
Interleukin inhibitors  
25  
Topical acne agents, Topical keratolytics  
24  
Minerals and electrolytes  
24  
Topical antipsoriatics  
23  
Narcotic analgesic combinations  
22  
Contraceptives  
22  
Antacids  
21  
Proton pump inhibitors  
21  
Topical non-steroidal anti-inflammatories

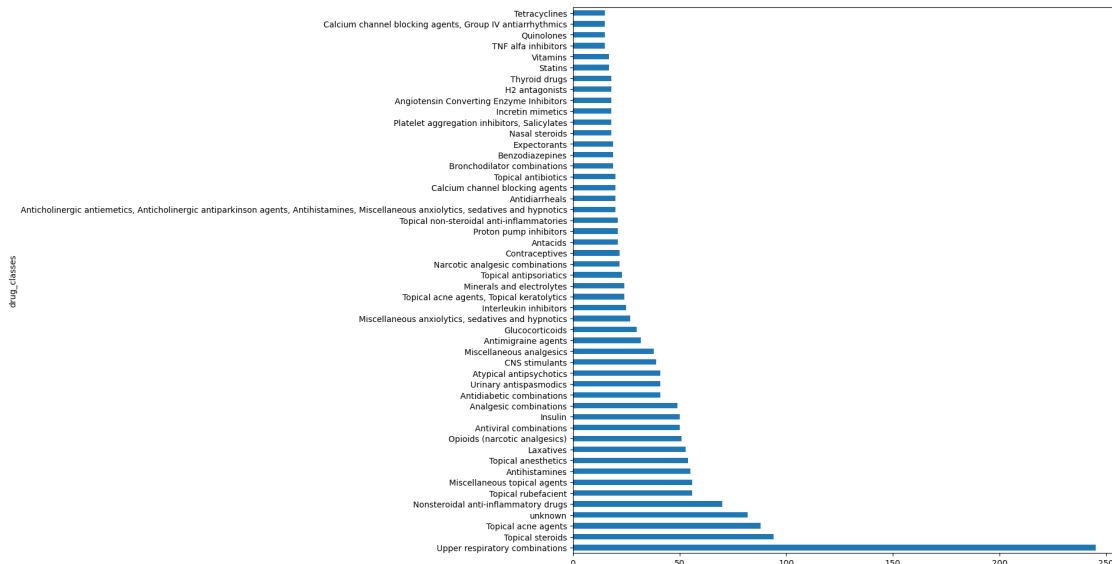
```

21
Anticholinergic antiemetics, Anticholinergic antiparkinson agents,
Antihistamines, Miscellaneous anxiolytics, sedatives and hypnotics      20
Antidiarrheals
20
Calcium channel blocking agents
20
Topical antibiotics
20
Bronchodilator combinations
19
Benzodiazepines
19
Expectorants
19
Nasal steroids
18
Platelet aggregation inhibitors, Salicylates
18
Incretin mimetics
18
Angiotensin Converting Enzyme Inhibitors
18
H2 antagonists
18
Thyroid drugs
18
Statins
17
Vitamins
17
TNF alfa inhibitors
15
Quinolones
15
Calcium channel blocking agents, Group IV antiarrhythmics
15
Tetracyclines
15
Name: count, dtype: int64

```

```
[24]: top_50_drug_class.plot(kind = 'barh',figsize = (12,12))
```

```
[24]: <Axes: ylabel='drug_classes'>
```



### 2.1.3 Extract Rows with Activity Greater Than 60%

```
[25]: df_cleaned[df_cleaned['activity']>60].index
```

```
[25]: Index([ 0, 1, 238, 239, 293, 294, 295, 296, 402, 403, 404, 416,
            417, 418, 419, 443, 444, 445, 514, 515, 516, 559, 560, 604,
            605, 651, 652, 653, 722, 742, 787, 1032, 1033, 1034, 1092, 1115,
            1149, 1200, 1201, 1202, 1257, 1361, 1362, 1399, 1521, 1522, 1523, 1534,
            1535, 1536, 1541, 1618, 1627, 1638, 1639, 1762, 1772, 1949, 1966, 1967,
            1988, 2007, 2067, 2074, 2135, 2136, 2264, 2265, 2656, 2749, 2802, 2852,
            2853, 2857, 2862, 2908],
            dtype='int64')
```

```
[26]: df_cleaned.iloc[df_cleaned[df_cleaned['activity']>60].index].reset_index().
      ↪head()
```

```
[26]:   index  drug_name  medical_condition \
0      0  doxycycline          Acne
1      1  spironolactone          Acne
2    238      Vyvanse          ADHD
3    239    Adderall          ADHD
4    293     Truvada        AIDS/HIV

      side_effects \
0  (hives, difficult breathing, swelling in your ...
1  hives ; difficulty breathing; swelling of your...
2  hives; difficult breathing; swelling of your f...
3  hives; difficult breathing; swelling of your f...
```

4 hives ; difficult breathing; swelling of your ...

	generic_name \
0	doxycycline
1	spironolactone
2	lisdexamfetamine
3	amphetamine and dextroamphetamine
4	emtricitabine and tenofovir disoproxil fumarate

	drug_classes \
0	Miscellaneous antimalarials, Tetracyclines
1	Aldosterone receptor antagonists, Potassium-sp...
2	CNS stimulants
3	CNS stimulants
4	Antiviral combinations

	brand_names	activity	rx_otc \
0	Acticlate, Adoxa CK, Adoxa Pak, Adoxa TT, Alod...	87	Rx
1	Aldactone, CaroSpir	82	Rx
2	unknown	100	Rx
3	Adderall XR, Mydayis	95	Rx
4	unknown	97	Rx

	pregnancy_category	csa	alcohol \
0	D	N	X
1	C	N	X
2	C	2	X
3	C	2	X
4	B	N	unknown

	related_drugs \
0	amoxicillin: <a href="https://www.drugs.com/amoxicillin...">https://www.drugs.com/amoxicillin...</a>
1	amlodipine: <a href="https://www.drugs.com/amlodipine.h...">https://www.drugs.com/amlodipine.h...</a>
2	unknown
3	Vyvanse: <a href="https://www.drugs.com/vyvanse.html">https://www.drugs.com/vyvanse.html</a>   ...
4	tenofovir: <a href="https://www.drugs.com/mtm/tenofovir...">https://www.drugs.com/mtm/tenofovir...</a>

	medical_condition_description	rating	no_of_reviews \
0	Acne Other names: Acne Vulgaris; Blackheads; B...	6.8	760.0
1	Acne Other names: Acne Vulgaris; Blackheads; B...	7.2	449.0
2	ADHD (Attention Deficit Hyperactivity Disorder...	7.3	768.0
3	ADHD (Attention Deficit Hyperactivity Disorder...	7.4	360.0
4	HIV Infection Other names: Acquired Immune Def...	7.1	14.0

	drug_link \
0	<a href="https://www.drugs.com/doxycycline.html">https://www.drugs.com/doxycycline.html</a>
1	<a href="https://www.drugs.com/spironolactone.html">https://www.drugs.com/spironolactone.html</a>

```

2      https://www.drugs.com/vyvanse.html
3      https://www.drugs.com/adderall.html
4      https://www.drugs.com/truvada.html

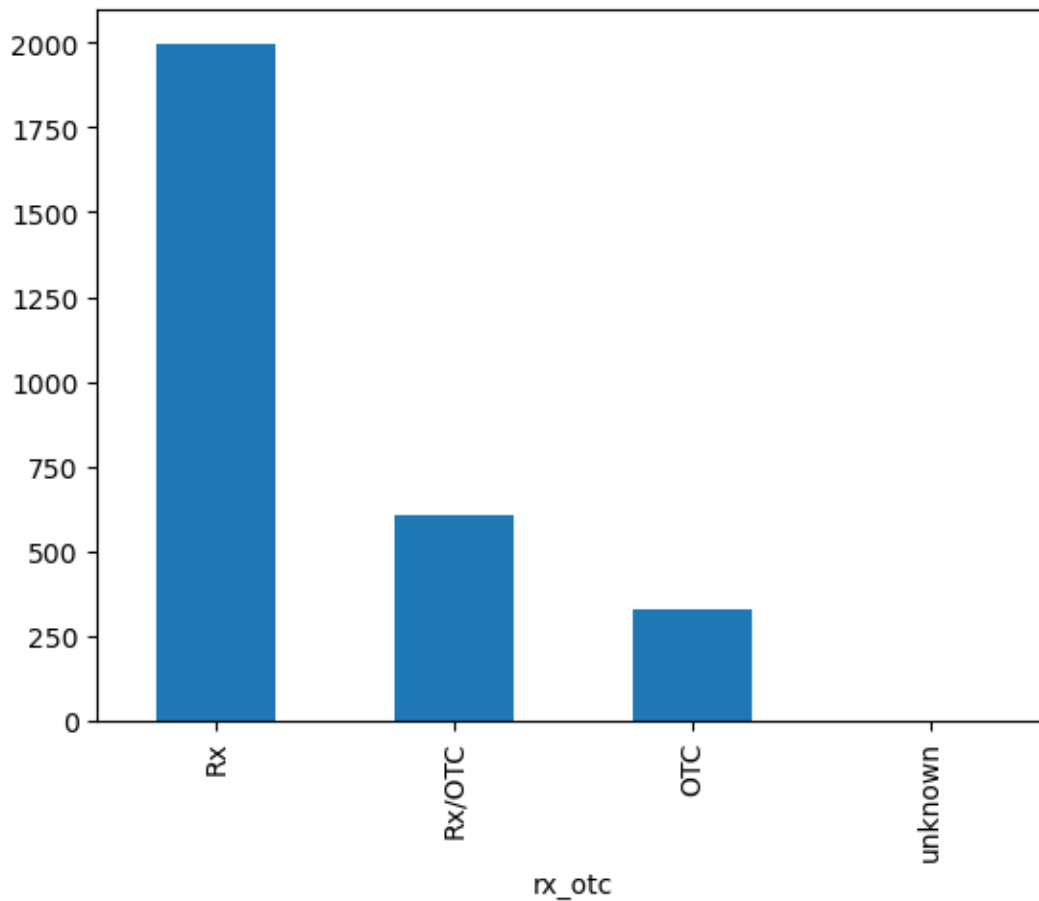
      medical_condition_url
0      https://www.drugs.com/condition/acne.html
1      https://www.drugs.com/condition/acne.html
2      https://www.drugs.com/condition/attention-defi...
3      https://www.drugs.com/condition/attention-defi...
4      https://www.drugs.com/condition/hiv-infection...

```

#### 2.1.4 Distribution of 'rx\_otc' with All Values Including 'Unknown'

```
[27]: df_cleaned['rx_otc'].value_counts().plot(kind = 'bar')
```

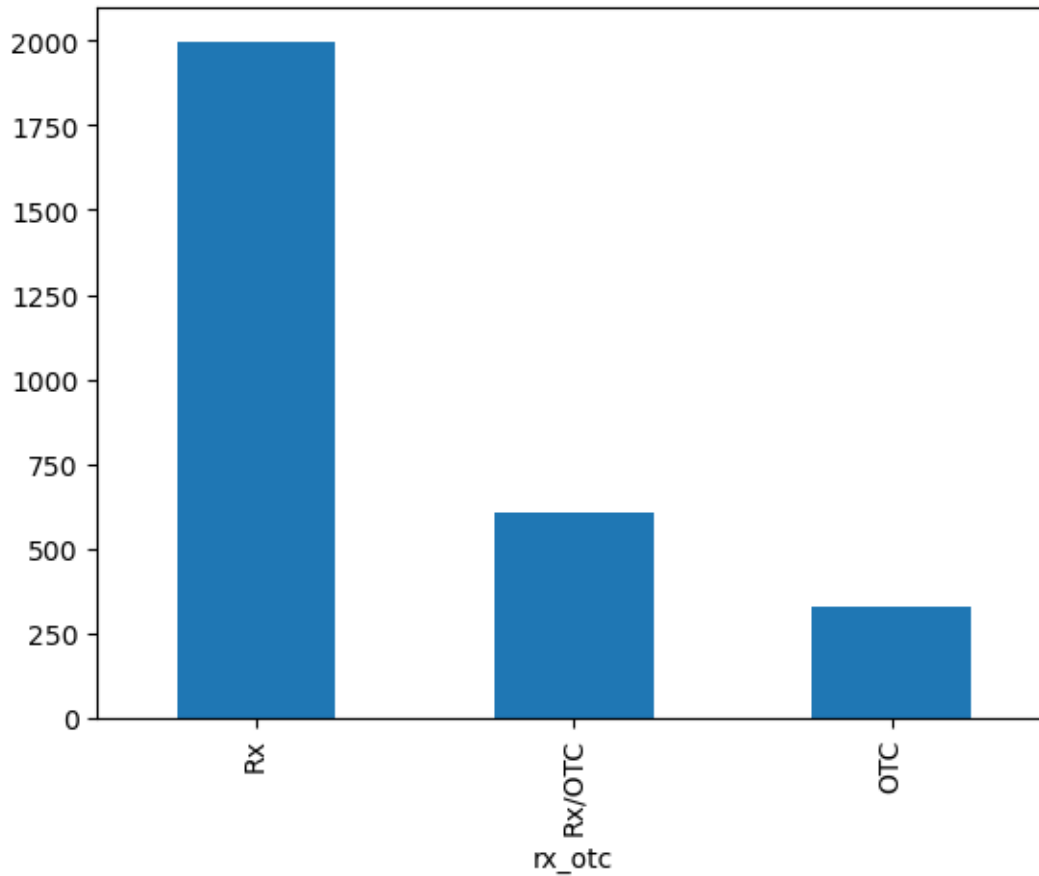
```
[27]: <Axes: xlabel='rx_otc'>
```



### 2.1.5 Distribution of 'rx\_otc' with All Values EXcluding 'Unknown'

```
[255]: df_cleaned['rx_otc'].value_counts()[df_cleaned['rx_otc'].value_counts().index !=  
      ↪ 'unknown'].plot(kind = 'bar')
```

```
[255]: <Axes: xlabel='rx_otc'>
```



### 2.1.6 pregnancy\_category Disbtribution

```
[45]: df_cleaned['pregnancy_category'].  
      ↪ value_counts()[df_cleaned['pregnancy_category'].value_counts().index !=  
      ↪ 'unknown']
```

```
[45]: pregnancy_category  
C      1382  
B       509  
N       436  
D       228  
X       129
```



```
A      18  
Name: count, dtype: int64
```

```
[44]: df_cleaned['pregnancy_category'].  
      ↪ value_counts()[df_cleaned['pregnancy_category'].value_counts().index !=  
      ↪ 'unknown'].plot(kind = 'bar')
```

```
[44]: <Axes: xlabel='pregnancy_category'>
```

