250k Medicines Usage, Side Effects and Substitutes

November 9, 2024

1 About Me

Mohammed Imam Uddin Riad
Data Analyst

Connect With Me - Portfolio - LinkedIn - Email Me

Email Me

[348]:
```
import pandas as pd
```

2 Importing Dataset and Checking Basic Info

[349]:
```
fname = "250k Medicines Usage, Side Effects and Substitutes.csv"
raw_data = pd.read_csv(fname)
df= raw_data.copy()
df.head()
```

```
C:\Users\Mohammed Riad\AppData\Local\Temp\ipykernel_14468\1691797094.py:2:
DtypeWarning: Columns (42,43,44,45,46,47,48) have mixed types. Specify dtype
option on import or set low_memory=False.
  raw_data = pd.read_csv(fname)
```

[349]:
```
   id                      name                     substitute0  \
0   1  augmentin 625 duo tablet  Penciclav 500 mg/125 mg Tablet
1   2       azithral 500 tablet         Zithrocare 500mg Tablet
2   3          ascoril ls syrup                 Solvin LS Syrup
3   4      allegra 120mg tablet                    Lcfex Tablet
4   5            avil 25 tablet               Eralet 25mg Tablet

             substitute1           substitute2           substitute3  \
0  Moxikind-CV 625 Tablet  Moxiforce-CV 625 Tablet     Fightox 625 Tablet
1         Azax 500 Tablet         Zady 500 Tablet  Cazithro 500mg Tablet
2       Ambrodil-LX Syrup       Zerotuss XP Syrup          Capex LS Syrup
3     Etofex 120mg Tablet     Nexofex 120mg Tablet     Fexise 120mg Tablet
4                     NaN                     NaN                    NaN

              substitute4 sideEffect0       sideEffect1     sideEffect2  …  \
0  Novamox CV 625mg Tablet    Vomiting            Nausea        Diarrhea  …
```

```
1    Trulimax 500mg Tablet    Vomiting           Nausea  Abdominal pain  …
2          Broxum LS Syrup      Nausea         Vomiting        Diarrhea  …
3     Histafree 120 Tablet    Headache       Drowsiness       Dizziness  …
4                      NaN  Sleepiness  Dryness in mouth             NaN  …

  sideEffect41                                               use0  \
0          NaN                 Treatment of Bacterial infections
1          NaN                 Treatment of Bacterial infections
2          NaN                     Treatment of Cough with mucus
3          NaN  Treatment of Sneezing and runny nose due to al…
4          NaN                  Treatment of Allergic conditions

                               use1 use2 use3 use4  \
0                               NaN  NaN  NaN  NaN
1                               NaN  NaN  NaN  NaN
2                               NaN  NaN  NaN  NaN
3  Treatment of Allergic conditions  NaN  NaN  NaN
4                               NaN  NaN  NaN  NaN

              Chemical Class Habit Forming Therapeutic Class  \
0                        NaN            No   ANTI INFECTIVES
1                 Macrolides            No   ANTI INFECTIVES
2                        NaN            No       RESPIRATORY
3  Diphenylmethane Derivative            No       RESPIRATORY
4      Pyridines Derivatives            No       RESPIRATORY

                             Action Class
0                                     NaN
1                              Macrolides
2                                     NaN
3  H1 Antihistaminics (second Generation)
4   H1 Antihistaminics (First Generation)

[5 rows x 58 columns]
```

[350]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 248218 entries, 0 to 248217
Data columns (total 58 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   id                  248218 non-null  int64
 1   name                248218 non-null  object
 2   substitute0         238621 non-null  object
 3   substitute1         233867 non-null  object
 4   substitute2         230233 non-null  object
 5   substitute3         226856 non-null  object
```

6	substitute4	223962 non-null	object
7	sideEffect0	248218 non-null	object
8	sideEffect1	238416 non-null	object
9	sideEffect2	229500 non-null	object
10	sideEffect3	207638 non-null	object
11	sideEffect4	163560 non-null	object
12	sideEffect5	131258 non-null	object
13	sideEffect6	91857 non-null	object
14	sideEffect7	67750 non-null	object
15	sideEffect8	48506 non-null	object
16	sideEffect9	37708 non-null	object
17	sideEffect10	27274 non-null	object
18	sideEffect11	20331 non-null	object
19	sideEffect12	16282 non-null	object
20	sideEffect13	14727 non-null	object
21	sideEffect14	10419 non-null	object
22	sideEffect15	7681 non-null	object
23	sideEffect16	6009 non-null	object
24	sideEffect17	5382 non-null	object
25	sideEffect18	4515 non-null	object
26	sideEffect19	3946 non-null	object
27	sideEffect20	3223 non-null	object
28	sideEffect21	3125 non-null	object
29	sideEffect22	3048 non-null	object
30	sideEffect23	2905 non-null	object
31	sideEffect24	2723 non-null	object
32	sideEffect25	1503 non-null	object
33	sideEffect26	1503 non-null	object
34	sideEffect27	1494 non-null	object
35	sideEffect28	1494 non-null	object
36	sideEffect29	1438 non-null	object
37	sideEffect30	1329 non-null	object
38	sideEffect31	1329 non-null	object
39	sideEffect32	1328 non-null	object
40	sideEffect33	1169 non-null	object
41	sideEffect34	1166 non-null	object
42	sideEffect35	2 non-null	object
43	sideEffect36	2 non-null	object
44	sideEffect37	2 non-null	object
45	sideEffect38	2 non-null	object
46	sideEffect39	2 non-null	object
47	sideEffect40	2 non-null	object
48	sideEffect41	2 non-null	object
49	use0	248218 non-null	object
50	use1	73365 non-null	object
51	use2	28307 non-null	object
52	use3	7379 non-null	object
53	use4	4971 non-null	object

```
 54  Chemical Class     137791 non-null  object
 55  Habit Forming      248218 non-null  object
 56  Therapeutic Class  248149 non-null  object
 57  Action Class       138036 non-null  object
dtypes: int64(1), object(57)
memory usage: 109.8+ MB
```

This Dataset Has Total 58 columns.It has

```
-5 substitutes columnns
-42 sideeffects columns
-5 uses columns
-and 4 class columns
```

3 Data Cleaning And Processing

3.0.1 Dropping Unneccessary column

[351]:
```
df = df.drop("id",axis =1)
df.head()
```

[351]:
```
                   name                       substitute0  \
0  augmentin 625 duo tablet  Penciclav 500 mg/125 mg Tablet
1      azithral 500 tablet         Zithrocare 500mg Tablet
2         ascoril ls syrup                 Solvin LS Syrup
3     allegra 120mg tablet                    Lcfex Tablet
4            avil 25 tablet               Eralet 25mg Tablet

            substitute1            substitute2           substitute3  \
0  Moxikind-CV 625 Tablet  Moxiforce-CV 625 Tablet     Fightox 625 Tablet
1         Azax 500 Tablet          Zady 500 Tablet  Cazithro 500mg Tablet
2       Ambrodil-LX Syrup        Zerotuss XP Syrup         Capex LS Syrup
3     Etofex 120mg Tablet     Nexofex 120mg Tablet    Fexise 120mg Tablet
4                     NaN                      NaN                    NaN

             substitute4 sideEffect0       sideEffect1     sideEffect2  \
0  Novamox CV 625mg Tablet    Vomiting            Nausea        Diarrhea
1    Trulimax 500mg Tablet    Vomiting            Nausea  Abdominal pain
2          Broxum LS Syrup      Nausea          Vomiting        Diarrhea
3     Histafree 120 Tablet    Headache        Drowsiness       Dizziness
4                      NaN  Sleepiness  Dryness in mouth             NaN

     sideEffect3  … sideEffect41  \
0            NaN  …          NaN
1       Diarrhea  …          NaN
2  Upset stomach  …          NaN
3         Nausea  …          NaN
4            NaN  …          NaN
```

```
                                                use0  \
0                Treatment of Bacterial infections
1                Treatment of Bacterial infections
2                    Treatment of Cough with mucus
3  Treatment of Sneezing and runny nose due to al…
4                 Treatment of Allergic conditions

                                use1 use2 use3 use4  \
0                                NaN  NaN  NaN  NaN
1                                NaN  NaN  NaN  NaN
2                                NaN  NaN  NaN  NaN
3  Treatment of Allergic conditions  NaN  NaN  NaN
4                                NaN  NaN  NaN  NaN

              Chemical Class Habit Forming Therapeutic Class  \
0                        NaN            No   ANTI INFECTIVES
1                 Macrolides            No   ANTI INFECTIVES
2                        NaN            No       RESPIRATORY
3  Diphenylmethane Derivative           No       RESPIRATORY
4       Pyridines Derivatives           No       RESPIRATORY

                            Action Class
0                                    NaN
1                             Macrolides
2                                    NaN
3  H1 Antihistaminics (second Generation)
4   H1 Antihistaminics (First Generation)

[5 rows x 57 columns]
```

3.0.2 Dropping Duplicated Rows across all the columns

[352]:
```
df.duplicated().any()
```

[352]: True

[353]:
```
# Drop duplicated rows, keeping only the first occurrence
df = df.drop_duplicates()
#checking again is there any duplicated rows after deletetion
df.duplicated().any()
```

[353]: False

3.0.3 Working With Missing Values

[354]:
```
df.isnull().sum().head()
```

[354]:
```
name              0
substitute0    8820
substitute1   13196
substitute2   16574
substitute3   19692
dtype: int64
```

[355]:
```
#making a dataframe with the missing values of the dataset

null_counts = df.isnull().sum()
missing_value =  pd.DataFrame(null_counts, columns=['null_count']).reset_index()
missing_value = missing_value.rename(columns={'index': 'Column'})

# Drop the first two row namem,as name has no missing value
missing_value = missing_value.drop([0]).reset_index(drop=True)



missing_value.head()
```

[355]:

	Column	null_count
0	substitute0	8820
1	substitute1	13196
2	substitute2	16574
3	substitute3	19692
4	substitute4	22357

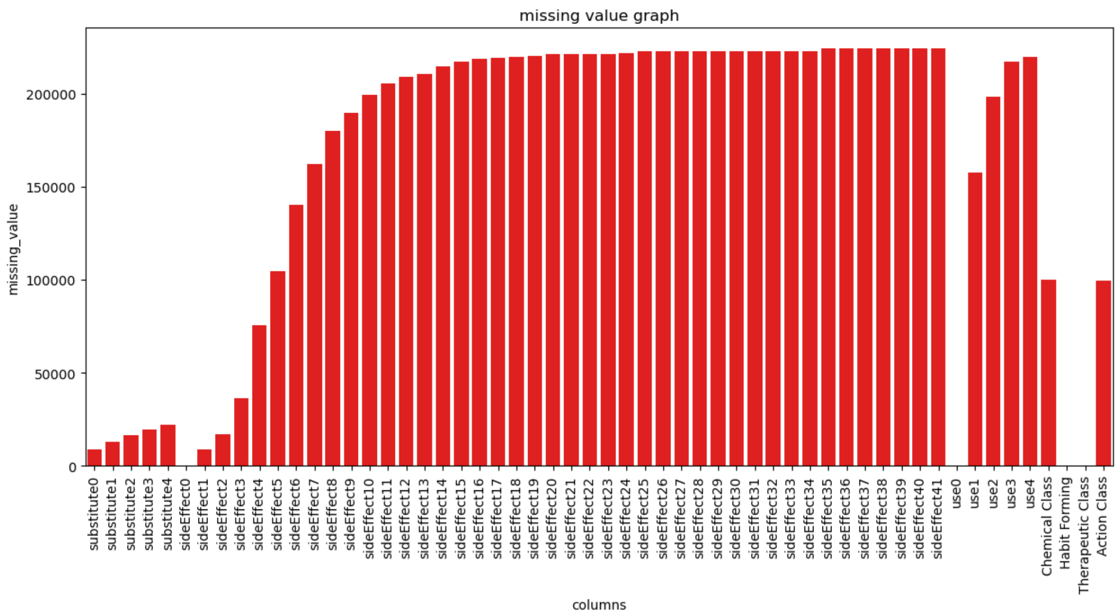
[356]:
```
#creating a barplot with the missing value

import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(14, 6))
sns.barplot(x='Column', y='null_count', data=missing_value, color ='red')

plt.xlabel('columns')
plt.ylabel('missing_value')
plt.title("missing value graph")
plt.xticks(rotation=90)  # Rotate x-axis labels if necessary
plt.show()

# Show the plot
plt.show()
```



[357]:
```
#filling the missing values with unknown
df.fillna("UnKnown",inplace = True)

#checking is there any missing value remaining
df.isnull().sum()
```

[357]:
```
name                0
substitute0         0
substitute1         0
substitute2         0
substitute3         0
substitute4         0
sideEffect0         0
sideEffect1         0
sideEffect2         0
sideEffect3         0
sideEffect4         0
sideEffect5         0
sideEffect6         0
sideEffect7         0
sideEffect8         0
sideEffect9         0
sideEffect10        0
sideEffect11        0
sideEffect12        0
sideEffect13        0
sideEffect14        0
```

```
sideEffect15        0
sideEffect16        0
sideEffect17        0
sideEffect18        0
sideEffect19        0
sideEffect20        0
sideEffect21        0
sideEffect22        0
sideEffect23        0
sideEffect24        0
sideEffect25        0
sideEffect26        0
sideEffect27        0
sideEffect28        0
sideEffect29        0
sideEffect30        0
sideEffect31        0
sideEffect32        0
sideEffect33        0
sideEffect34        0
sideEffect35        0
sideEffect36        0
sideEffect37        0
sideEffect38        0
sideEffect39        0
sideEffect40        0
sideEffect41        0
use0                0
use1                0
use2                0
use3                0
use4                0
Chemical Class      0
Habit Forming       0
Therapeutic Class   0
Action Class        0
dtype: int64
```

3.1 Creating Checkpoint 1

[358]:
```
df_cleaned = df.copy()
```

4 EDA(Exploratory Data Analysis.)

4.0.1 Explore how many sideeffects each medicine has and identify medicines with the most substitutes.

[359]:
```
#Create a list of side effect columns
side_effect_columns = [col for col in df_cleaned.columns if 'sideEffect' in col]

df_cleaned[side_effect_columns].head(3)
```

[359]:
```
  sideEffect0 sideEffect1     sideEffect2    sideEffect3   sideEffect4  \
0    Vomiting      Nausea        Diarrhea        UnKnown       UnKnown
1    Vomiting      Nausea  Abdominal pain       Diarrhea       UnKnown
2      Nausea    Vomiting        Diarrhea  Upset stomach  Stomach pain

         sideEffect5 sideEffect6 sideEffect7 sideEffect8 sideEffect9  …  \
0            UnKnown     UnKnown     UnKnown     UnKnown     UnKnown  …
1            UnKnown     UnKnown     UnKnown     UnKnown     UnKnown  …
2  Allergic reaction   Dizziness    Headache        Rash       Hives  …

  sideEffect32 sideEffect33 sideEffect34 sideEffect35 sideEffect36  \
0      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown
1      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown
2      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown

  sideEffect37 sideEffect38 sideEffect39 sideEffect40 sideEffect41
0      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown
1      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown
2      UnKnown      UnKnown      UnKnown      UnKnown      UnKnown

[3 rows x 42 columns]
```

[360]:
```
# Calculate the count of known side effects for each drug by counting entries␣
↪that are not "Unknown"
df_cleaned['total_side_effect_recorded'] = df_cleaned[side_effect_columns].
↪apply(lambda x: (x != "UnKnown").sum(), axis=1)
df_cleaned.head(3)
```

[360]:
```
                    name                      substitute0  \
0  augmentin 625 duo tablet  Penciclav 500 mg/125 mg Tablet
1       azithral 500 tablet         Zithrocare 500mg Tablet
2          ascoril ls syrup                 Solvin LS Syrup

            substitute1           substitute2           substitute3  \
0  Moxikind-CV 625 Tablet  Moxiforce-CV 625 Tablet     Fightox 625 Tablet
1          Azax 500 Tablet        Zady 500 Tablet  Cazithro 500mg Tablet
2       Ambrodil-LX Syrup      Zerotuss XP Syrup         Capex LS Syrup
```

```
               substitute4 sideEffect0 sideEffect1     sideEffect2  \
0  Novamox CV 625mg Tablet    Vomiting      Nausea        Diarrhea
1    Trulimax 500mg Tablet    Vomiting      Nausea  Abdominal pain
2          Broxum LS Syrup      Nausea    Vomiting        Diarrhea

      sideEffect3  …                            use0     use1     use2  \
0         UnKnown  …  Treatment of Bacterial infections  UnKnown  UnKnown
1        Diarrhea  …  Treatment of Bacterial infections  UnKnown  UnKnown
2   Upset stomach  …      Treatment of Cough with mucus  UnKnown  UnKnown

      use3     use4 Chemical Class Habit Forming Therapeutic Class  \
0  UnKnown  UnKnown        UnKnown            No   ANTI INFECTIVES
1  UnKnown  UnKnown     Macrolides            No   ANTI INFECTIVES
2  UnKnown  UnKnown        UnKnown            No       RESPIRATORY

  Action Class total_side_effect_recorded
0      UnKnown                          3
1   Macrolides                          4
2      UnKnown                         14

[3 rows x 58 columns]
```

[361]:
```
#Top 10 Medicine that has most Side effects
df_cleaned[['name','total_side_effect_recorded']].sort_values(by =␣
 ↪'total_side_effect_recorded',ascending=False).head(10)
```

[361]:
```
                                name  total_side_effect_recorded
30664                 balila capsule                          42
30832            balila 25mg capsule                          42
235196  waycef o 200mg/200mg tablet                           35
235198                wincef-o tablet                          35
37699   brefix o 200mg/200mg tablet                           35
46980                ceftrue-o tablet                          35
235822  winfex o 200mg/200mg tablet                           35
142110               milixim-o tablet                          35
169329               netfix-ox tablet                          35
48126               cefistar-o tablet                          35
```

4.0.2 Explore how many substitutes each medicine has

[362]:
```
#Create a list of substitute columns
substitute_columns = [col for col in df_cleaned.columns if 'substitute' in col]

df_cleaned[substitute_columns].head(3)
```

[362]:
```
                        substitute0             substitute1  \
0  Penciclav 500 mg/125 mg Tablet  Moxikind-CV 625 Tablet
```

```
1         Zithrocare 500mg Tablet         Azax 500 Tablet
2                 Solvin LS Syrup       Ambrodil-LX Syrup

               substitute2             substitute3              substitute4
0  Moxiforce-CV 625 Tablet     Fightox 625 Tablet  Novamox CV 625mg Tablet
1           Zady 500 Tablet  Cazithro 500mg Tablet    Trulimax 500mg Tablet
2         Zerotuss XP Syrup          Capex LS Syrup          Broxum LS Syrup
```

[363]:
```python
df_cleaned["total_substitute_counts"]=df_cleaned[substitute_columns].
 ↪apply(lambda x: (x!= 'UnKnown').sum(),axis= 1)
df_cleaned.head(3)
```

[363]:
```
                       name                    substitute0  \
0  augmentin 625 duo tablet  Penciclav 500 mg/125 mg Tablet
1       azithral 500 tablet         Zithrocare 500mg Tablet
2          ascoril ls syrup                 Solvin LS Syrup

              substitute1              substitute2            substitute3  \
0  Moxikind-CV 625 Tablet  Moxiforce-CV 625 Tablet     Fightox 625 Tablet
1         Azax 500 Tablet          Zady 500 Tablet  Cazithro 500mg Tablet
2       Ambrodil-LX Syrup        Zerotuss XP Syrup         Capex LS Syrup

               substitute4 sideEffect0 sideEffect1     sideEffect2  \
0  Novamox CV 625mg Tablet    Vomiting      Nausea        Diarrhea
1    Trulimax 500mg Tablet    Vomiting      Nausea  Abdominal pain
2          Broxum LS Syrup      Nausea    Vomiting        Diarrhea

     sideEffect3  …     use1     use2     use3     use4 Chemical Class  \
0        UnKnown  …  UnKnown  UnKnown  UnKnown  UnKnown        UnKnown
1       Diarrhea  …  UnKnown  UnKnown  UnKnown  UnKnown     Macrolides
2  Upset stomach  …  UnKnown  UnKnown  UnKnown  UnKnown        UnKnown

  Habit Forming Therapeutic Class Action Class total_side_effect_recorded  \
0            No   ANTI INFECTIVES      UnKnown                          3
1            No   ANTI INFECTIVES   Macrolides                          4
2            No       RESPIRATORY      UnKnown                         14

  total_substitute_counts
0                       5
1                       5
2                       5

[3 rows x 59 columns]
```

4.0.3 Identify which substitute is used in the highest number of medicines.

[364]:
```
# Count non-"Unknown" values for each substitute column
substitute_counts = pd.DataFrame(df_cleaned[substitute_columns].apply(lambda␣
↪col: (col != "UnKnown")).sum(axis = 0),columns= ['Counts']).reset_index()
substitute_counts=substitute_counts.rename(columns={'index':'substitutes'})
substitute_counts
```

[364]:

	substitutes	Counts
0	substitute0	215194
1	substitute1	210818
2	substitute2	207440
3	substitute3	204322
4	substitute4	201657

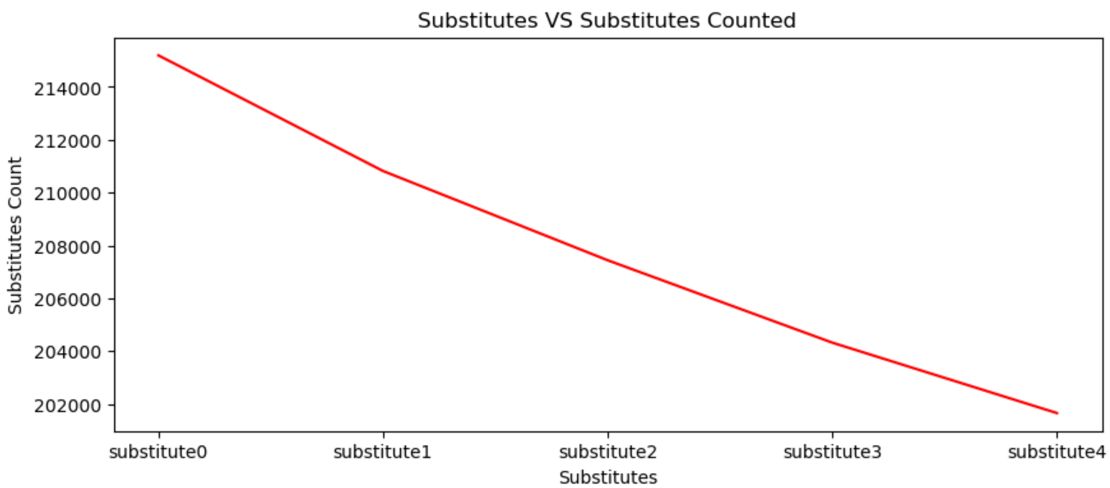
[365]:
```
#creating a line with the missing value

import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(10, 4))
sns.lineplot(x='substitutes', y='Counts', data=substitute_counts, color ='red')

plt.xlabel('Substitutes')
plt.ylabel('Substitutes Count')
plt.title("Substitutes VS Substitutes Counted")
plt.xticks(rotation=0)
plt.show()

# Show the plot
plt.show()
```



4.0.4 Explore the total uses of the medicines

[366]:
```python
use_columns = [f'use{i}' for i in range(5)]
use_columns
```

[366]:
```
['use0', 'use1', 'use2', 'use3', 'use4']
```

[367]:
```python
df_cleaned['total_uses'] = df_cleaned[use_columns].apply(lambda x: x !=␣
↪'UnKnown').sum(axis =1)
df_cleaned.sort_values(by= 'total_uses',ascending = False).head(3)
```

[367]:
```
                          name           substitute0         substitute1  \
152854   mincetam 800mg tablet     Cgtam 800mg Tablet    Leepan 800 Tablet
172310        navmit 4mg tablet      Predace 4 Tablet   Coelone 4mg Tablet
235271  walbeta la 20mg tablet  Prograin 20mg Tablet  Nigrain 20mg Tablet

                 substitute2          substitute3          substitute4  \
152854  Geocetam 800mg Tablet    Piramax 800 Tablet  Sumocetam 800 Tablet
172310     Melpred 4mg Tablet    Pilsone 4mg Tablet    Cortilog 4mg Tablet
235271    Betagem 20mg Tablet  Signolol 20mg Tablet   Arminol 20mg Tablet

              sideEffect0                 sideEffect1  \
152854        Weight gain                 Nervousness
172310  Thinning of skin  Increased risk of infection
235271          Tiredness                    Weakness

                             sideEffect2                         sideEffect3  \
152854  Abnormality of voluntary movements                              UnKnown
172310             Reduction in bone density                         Weight gain
235271                  Raynaud's phenomenon  Arrhythmia (irregular heartbeats)

        …                           use2                        use3  \
152854  …    Dementia in Parkinson's disease     Age related memory loss
172310  …  Treatment of Allergic conditions  Treatment of Eye disorders
235271  …               Prevention of migraine        Treatment of Anxiety

                             use4                Chemical Class  \
152854                  Head injury  Alpha Amino Acids Derivatives
172310  Treatment of Skin disorders               Glucocorticoids
235271      Treatment of Arrhythmia       Naphthalenes derivatives

       Habit Forming Therapeutic Class                  Action Class  \
152854             No         NEURO CNS               Nootropic agent
172310             No          HORMONES               Glucocorticoids
235271             No           CARDIAC  Beta blocker- Non selective
```

```
       total_side_effect_recorded total_substitute_counts total_uses
152854                          3                       5          5
172310                          7                       5          5
235271                          7                       5          5

[3 rows x 60 columns]
```

4.0.5 Explore the most common uses of the medicines.

[368]:
```python
# Combine all use columns into one
use_columns = [f'use{i}' for i in range(5)]

uses = df_cleaned[use_columns].apply(lambda x: x.value_counts()).sum(axis =1)
uses = pd.DataFrame(uses,columns = ['uses_counts']).reset_index()
uses = uses.rename(columns = {'index':'uses'})

# Filter out rows where 'uses' is "Unknown" or null
uses = uses[uses.apply(lambda x: (x!= 'UnKnown'))]
uses = uses.dropna()

#sorting the dataframe
uses = uses.sort_values("uses_counts",ascending = False)

#Top 10 using
uses.head(10)
```

[368]:
```
                                                 uses  uses_counts
549                 Treatment of Bacterial infections      37129.0
313                                       Pain relief      21152.0
736                 Treatment of Peptic ulcer disease      12581.0
627  Treatment of Gastroesophageal reflux disease (…      12556.0
799             Treatment of Type 2 diabetes mellitus      10250.0
41                                Bacterial infections       9719.0
649   Treatment of Hypertension (high blood pressure)       7178.0
525                  Treatment of Allergic conditions       5766.0
786  Treatment of Sneezing and runny nose due to al…       4904.0
637                            Treatment of Heartburn       4381.0
```

4.0.6 Explore 50 most used chemical (Excluding 'Unknown')

[369]:
```python
df_cleaned['Chemical Class'][df_cleaned['Chemical Class'].apply(lambda x:x!=
↪'UnKnown')].value_counts().head(50)
```

[369]:
```
Chemical Class
Fluoroquinolone                                             7007
Broad Spectrum (Third & fourth generation cephalosporins)    6407
```

Macrolides	4808
Sulfinylbenzimidazole Derivative	4395
Broad spectrum (Third & fourth generation cephalosporins}	3936
Gluco/mineralocorticoids, progestogins and derivatives	2604
Azoles {Triazoles}	2443
Intermediate spectrum {Second generation cephalosporins}	2375
Carbazole Derivative	2278
Glucocorticoids	2238
Azole derivatives {Imidazoles}	2065
Benzodiazepines Derivative	2025
Aminoglycosides	1951
Piperazine Derivatives	1818
P-Aminophenol Derivative	1563
Aminopenicillins {Penicillins}	1528
Pyrrole & heptanoic acid derivative	1505
Anabolic steroid	1358
Timoprazole Derivative	1327
Phenylacetic acid Derivative	1304
Progesterone Derivative	1263
Sulfone and Pyridine Derivative	1094
Diphenylmethane Derivative	1035
Carbapenem derivative	1011
Cobalamin Derivative	1008
Dihydropyridinecarboxylic acids derivatives	966
Phenylpyrimidine derivatives	947
Histamine Analog	944
Benzimidazole derivative	922
Azoles derivatives	916
Diphenylethers Derivative	845
2-Benzimidazolylcarbamic acid esters	790
Phenylbutylamine Derivative	752
Bile Acids and Salts	751
Thienobenzodiazepine Derivative	746
Dichlorobenzenes Derivative	742
Alcohol & phenols	718
Piperazine Derivative	700
Sesquiterpene lactones	680
Biguanides Derivative	680
Phenylmethyl Piperazinyl Derivative	635
Valiolamine derivative	584
Pyrrolidinone & Acetamide Derivative	559
Dibenzocycloheptenes Derivative	556
Benzenesulfonamide Derivative	552
Thiazole Derivative	540
Narrow-spectrum {First generation cephalosporins}	539
Oxazolidinone derivative	529
Nucleoside analog	520

```
Narrow- specttrum {First generation cephalosporins}          518
Name: count, dtype: int64
```

4.0.7 Explore the uses of Therapeutic class (Excluding 'Unknown')

[370]:
```python
df_cleaned['Therapeutic Class'][df_cleaned['Therapeutic Class'].apply(lambda x:
 ↪x!= 'UnKnown')].value_counts().sort_values(ascending = False)
```

[370]:
```
Therapeutic Class
ANTI INFECTIVES                 50459
GASTRO INTESTINAL               30442
PAIN ANALGESICS                 29187
NEURO CNS                       21524
RESPIRATORY                     21284
CARDIAC                         16774
ANTI DIABETIC                   10457
OPHTHAL                          9494
DERMA                            8949
HORMONES                         4927
GYNAECOLOGICAL                   4094
VITAMINS MINERALS NUTRIENTS      3885
ANTI NEOPLASTICS                 3202
BLOOD RELATED                    2464
UROLOGY                          1720
OPHTHAL OTOLOGICALS              1558
ANTI MALARIALS                   1466
SEX STIMULANTS REJUVENATORS       682
OTHERS                            464
OTOLOGICALS                       451
VACCINES                          306
STOMATOLOGICALS                   162
Name: count, dtype: int64
```

4.0.8 Explore the uses of Action class (Excluding 'Unknown')

[371]:
```python
df_cleaned['Action Class'][df_cleaned['Action Class'].apply(lambda x:x!=␣
 ↪'UnKnown')].value_counts().sort_values(ascending = False).head(50)
```

[371]:
```
Action Class
Cephalosporins: 3 generation                              10901
Quinolones/ Fluroquinolones                                7011
Fungal ergosterol synthesis inhibitor                      5990
Proton pump inhibitors                                     5734
Macrolides                                                 5109
Glucocorticoids                                            4779
H1 Antihistaminics (second Generation)                     3702
HMG CoA inhibitors (statins)                               2561
```

```
Cephalosporins: 2nd generation                               2448
Benzodiazepines                                              2447
NSAID's- Non-Selective COX 1&2 Inhibitors (acetic acid)      2426
Serotonin antagonists (5-HT3 antagonists)                    2369
Atypical Antipsychotics                                      2364
Aminoglycosides                                              1951
Selective Seretonin Reuptake inhibitors (SSRIs)              1829
Angiotensin receptor blockers(ARB)                           1747
Sodium channel modulators (AED)                              1600
Analgesic & Antipyretic-PCM                                  1563
Cell wall active agent -Extended spectrum Penicillin         1536
Vitamins                                                     1364
Anabolic steroid                                             1358
Natural Progesterone                                         1263
NSAID's -Selective COX-2 Inhibitors                          1213
Beta blocker- Cardioselective                                1206
Calcium channel blockers- Dihydropyridines (DHP)             1166
Cephalosporins: 1st generation                               1057
Cell wall active agent -Carbapenems                          1011
Antiprotozoal agents                                          966
Nootropic agent                                               944
Histamine analog- Meniere's Disease                           944
Sulfonylureas (Insulin Secretogogues)                         896
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)             872
H1 Antihistaminics (First Generation)                         844
Hepatoprotectives                                             837
Antimalarial- Artemisinin and derivatives                     831
Tricyclic antidepressants                                     779
Synaptic vescicle 2 A protein ligand (AED)                    685
Biguanides                                                    680
Alpha-glucosidase inhibitors                                  679
Xanthine oxidase Inhibitors-gout                              655
DPP-4 inhibitors                                              648
Theophylline & its derivatives                                610
Ectoparasiticides                                             609
Progestins (First generation)                                 607
Serotonin-norepinephrine reuptake inhibitors (SNRIs)          582
Alpha 2 delta ligands (AED)                                   579
Proteolytic Enzymes                                           578
Angiotensin-converting enzyme (ACE) inhibitors                568
Phosphodiesterase-V inhibitors                                564
Oxazolidinone                                                 529
Name: count, dtype: int64
```

5 Classification Analysis

5.0.1 Top 50 Chemical Classes by Frequency of Medicines (Excluding 'Unknown')

[372]:
```
med_frequency_by_chemical_class = df_cleaned.groupby('Chemical Class')['name'].
  ↪count().loc[lambda x: x.index != "UnKnown"].sort_values(ascending=False)
med_frequency_by_chemical_class.head(50)
```

[372]:
```
Chemical Class
Fluoroquinolone                                              7007
Broad Spectrum (Third & fourth generation cephalosporins)    6407
Macrolides                                                   4808
Sulfinylbenzimidazole Derivative                             4395
Broad spectrum (Third & fourth generation cephalosporins}    3936
Gluco/mineralocorticoids, progestogins and derivatives       2604
Azoles {Triazoles}                                           2443
Intermediate spectrum {Second generation cephalosporins}     2375
Carbazole Derivative                                         2278
Glucocorticoids                                              2238
Azole derivatives {Imidazoles}                               2065
Benzodiazepines Derivative                                   2025
Aminoglycosides                                              1951
Piperazine Derivatives                                       1818
P-Aminophenol Derivative                                     1563
Aminopenicillins {Penicillins}                               1528
Pyrrole & heptanoic acid derivative                          1505
Anabolic steroid                                             1358
Timoprazole Derivative                                       1327
Phenylacetic acid Derivative                                 1304
Progesterone Derivative                                      1263
Sulfone and Pyridine Derivative                              1094
Diphenylmethane Derivative                                   1035
Carbapenem derivative                                        1011
Cobalamin Derivative                                         1008
Dihydropyridinecarboxylic acids derivatives                   966
Phenylpyrimidine derivatives                                  947
Histamine Analog                                              944
Benzimidazole derivative                                      922
Azoles derivatives                                            916
Diphenylethers Derivative                                     845
2-Benzimidazolylcarbamic acid esters                          790
Phenylbutylamine Derivative                                   752
Bile Acids and Salts                                          751
Thienobenzodiazepine Derivative                               746
Dichlorobenzenes Derivative                                   742
Alcohol & phenols                                             718
Piperazine Derivative                                         700
```

```
Biguanides Derivative                                        680
Sesquiterpene lactones                                       680
Phenylmethyl Piperazinyl Derivative                          635
Valiolamine derivative                                       584
Pyrrolidinone & Acetamide Derivative                         559
Dibenzocycloheptenes Derivative                              556
Benzenesulfonamide Derivative                                552
Thiazole Derivative                                          540
Narrow-spectrum {First generation cephalosporins}            539
Oxazolidinone derivative                                     529
Nucleoside analog                                            520
Narrow- specttrum {First generation cephalosporins}          518
Name: name, dtype: int64
```

5.0.2 Top 50 Chemical Classes by Reported Side Effects Count (Excluding 'Unknown')

[373]:
```python
#loc[lambda x: x.index != "UnKnown"] excluding the index "unknow" from the␣
↪result

side_effect_frequency_by_chemical_class = df_cleaned.groupby(['Chemical␣
↪Class'])[side_effect_columns].count().sum(axis = 1).loc[lambda x: x.index !=␣
↪'UnKnown'].sort_values(ascending = False)
side_effect_frequency_by_chemical_class.head(50)
```

[373]:
```
Chemical Class
Fluoroquinolone                                                 294294
Broad Spectrum (Third & fourth generation cephalosporins)       269094
Macrolides                                                      201936
Sulfinylbenzimidazole Derivative                                184590
Broad spectrum (Third & fourth generation cephalosporins}       165312
Gluco/mineralocorticoids, progestogins and derivatives          109368
Azoles {Triazoles}                                              102606
Intermediate spectrum {Second generation cephalosporins}         99750
Carbazole Derivative                                             95676
Glucocorticoids                                                  93996
Azole derivatives {Imidazoles}                                   86730
Benzodiazepines Derivative                                       85050
Aminoglycosides                                                  81942
Piperazine Derivatives                                           76356
P-Aminophenol Derivative                                         65646
Aminopenicillins {Penicillins}                                   64176
Pyrrole & heptanoic acid derivative                              63210
Anabolic steroid                                                 57036
Timoprazole Derivative                                           55734
Phenylacetic acid Derivative                                     54768
Progesterone Derivative                                          53046
```

```
Sulfone and Pyridine Derivative                             45948
Diphenylmethane Derivative                                  43470
Carbapenem derivative                                       42462
Cobalamin Derivative                                        42336
Dihydropyridinecarboxylic acids derivatives                 40572
Phenylpyrimidine derivatives                                39774
Histamine Analog                                            39648
Benzimidazole derivative                                    38724
Azoles derivatives                                          38472
Diphenylethers Derivative                                   35490
2-Benzimidazolylcarbamic acid esters                        33180
Phenylbutylamine Derivative                                 31584
Bile Acids and Salts                                        31542
Thienobenzodiazepine Derivative                             31332
Dichlorobenzenes Derivative                                 31164
Alcohol & phenols                                           30156
Piperazine Derivative                                       29400
Biguanides Derivative                                       28560
Sesquiterpene lactones                                      28560
Phenylmethyl Piperazinyl Derivative                         26670
Valiolamine derivative                                      24528
Pyrrolidinone & Acetamide Derivative                        23478
Dibenzocycloheptenes Derivative                             23352
Benzenesulfonamide Derivative                               23184
Thiazole Derivative                                         22680
Narrow-spectrum {First generation cephalosporins}           22638
Oxazolidinone derivative                                    22218
Nucleoside analog                                           21840
Narrow- specttrum {First generation cephalosporins}         21756
dtype: int64
```

5.0.3 Top 50 Chemical Classes by Usage Count Across All Use Columns (Excluding 'Unknown')

[374]:

```
total_usage_frequency_by_chemical_class = df_cleaned.groupby(['Chemical Class'])[use_columns].count().sum(axis = 1).loc[lambda x: x.index != 'UnKnown'].sort_values(ascending = False)
total_usage_frequency_by_chemical_class.head(50)
```

[374]:

```
Chemical Class
Fluoroquinolone                                            35035
Broad Spectrum (Third & fourth generation cephalosporins)  32035
Macrolides                                                 24040
Sulfinylbenzimidazole Derivative                           21975
Broad spectrum (Third & fourth generation cephalosporins}  19680
Gluco/mineralocorticoids, progestogins and derivatives     13020
Azoles {Triazoles}                                         12215
```

```
Intermediate spectrum {Second generation cephalosporins}      11875
Carbazole Derivative                                          11390
Glucocorticoids                                               11190
Azole derivatives {Imidazoles}                                10325
Benzodiazepines Derivative                                    10125
Aminoglycosides                                                9755
Piperazine Derivatives                                         9090
P-Aminophenol Derivative                                       7815
Aminopenicillins {Penicillins}                                 7640
Pyrrole & heptanoic acid derivative                            7525
Anabolic steroid                                               6790
Timoprazole Derivative                                         6635
Phenylacetic acid Derivative                                   6520
Progesterone Derivative                                        6315
Sulfone and Pyridine Derivative                                5470
Diphenylmethane Derivative                                     5175
Carbapenem derivative                                          5055
Cobalamin Derivative                                           5040
Dihydropyridinecarboxylic acids derivatives                    4830
Phenylpyrimidine derivatives                                   4735
Histamine Analog                                               4720
Benzimidazole derivative                                       4610
Azoles derivatives                                             4580
Diphenylethers Derivative                                      4225
2-Benzimidazolylcarbamic acid esters                           3950
Phenylbutylamine Derivative                                    3760
Bile Acids and Salts                                           3755
Thienobenzodiazepine Derivative                                3730
Dichlorobenzenes Derivative                                    3710
Alcohol & phenols                                              3590
Piperazine Derivative                                          3500
Biguanides Derivative                                          3400
Sesquiterpene lactones                                         3400
Phenylmethyl Piperazinyl Derivative                            3175
Valiolamine derivative                                         2920
Pyrrolidinone & Acetamide Derivative                           2795
Dibenzocycloheptenes Derivative                                2780
Benzenesulfonamide Derivative                                  2760
Thiazole Derivative                                            2700
Narrow-spectrum {First generation cephalosporins}              2695
Oxazolidinone derivative                                       2645
Nucleoside analog                                              2600
Narrow- specttrum {First generation cephalosporins}            2590
dtype: int64
```

5.0.4 Chemical Data Summary: Usage, Side Effects, and Total Frequency

[375]:
```
all_abt_chemicals = pd.DataFrame({'Med_frequency':␣
 ↪med_frequency_by_chemical_class, 'Side_effect_frequency':␣
 ↪side_effect_frequency_by_chemical_class, 'Total_usage_frequency':␣
 ↪total_usage_frequency_by_chemical_class})
all_abt_chemicals.head(50)
```

[375]:
```
                                                    Med_frequency  \
Chemical Class
Fluoroquinolone                                              7007
Broad Spectrum (Third & fourth generation cepha…           6407
Macrolides                                                   4808
Sulfinylbenzimidazole Derivative                             4395
Broad spectrum (Third & fourth generation cepha…           3936
Gluco/mineralocorticoids, progestogins and deri…           2604
Azoles {Triazoles}                                           2443
Intermediate spectrum {Second generation cephal…           2375
Carbazole Derivative                                         2278
Glucocorticoids                                              2238
Azole derivatives {Imidazoles}                               2065
Benzodiazepines Derivative                                   2025
Aminoglycosides                                              1951
Piperazine Derivatives                                       1818
P-Aminophenol Derivative                                     1563
Aminopenicillins {Penicillins}                               1528
Pyrrole & heptanoic acid derivative                          1505
Anabolic steroid                                             1358
Timoprazole Derivative                                       1327
Phenylacetic acid Derivative                                 1304
Progesterone Derivative                                      1263
Sulfone and Pyridine Derivative                              1094
Diphenylmethane Derivative                                   1035
Carbapenem derivative                                        1011
Cobalamin Derivative                                         1008
Dihydropyridinecarboxylic acids derivatives                   966
Phenylpyrimidine derivatives                                  947
Histamine Analog                                              944
Benzimidazole derivative                                      922
Azoles derivatives                                            916
Diphenylethers Derivative                                     845
2-Benzimidazolylcarbamic acid esters                          790
Phenylbutylamine Derivative                                   752
Bile Acids and Salts                                          751
Thienobenzodiazepine Derivative                               746
Dichlorobenzenes Derivative                                   742
Alcohol & phenols                                             718
```

```
Piperazine Derivative                                        700
Biguanides Derivative                                        680
Sesquiterpene lactones                                       680
Phenylmethyl Piperazinyl Derivative                          635
Valiolamine derivative                                       584
Pyrrolidinone & Acetamide Derivative                         559
Dibenzocycloheptenes Derivative                              556
Benzenesulfonamide Derivative                                552
Thiazole Derivative                                          540
Narrow-spectrum {First generation cephalosporins}            539
Oxazolidinone derivative                                     529
Nucleoside analog                                            520
Narrow- specttrum {First generation cephalospor…           518

                                                   Side_effect_frequency  \
Chemical Class
Fluoroquinolone                                                   294294
Broad Spectrum (Third & fourth generation cepha…                269094
Macrolides                                                        201936
Sulfinylbenzimidazole Derivative                                  184590
Broad spectrum (Third & fourth generation cepha…                165312
Gluco/mineralocorticoids, progestogins and deri…                109368
Azoles {Triazoles}                                                102606
Intermediate spectrum {Second generation cephal…                 99750
Carbazole Derivative                                               95676
Glucocorticoids                                                    93996
Azole derivatives {Imidazoles}                                     86730
Benzodiazepines Derivative                                         85050
Aminoglycosides                                                    81942
Piperazine Derivatives                                             76356
P-Aminophenol Derivative                                           65646
Aminopenicillins {Penicillins}                                     64176
Pyrrole & heptanoic acid derivative                                63210
Anabolic steroid                                                   57036
Timoprazole Derivative                                             55734
Phenylacetic acid Derivative                                       54768
Progesterone Derivative                                            53046
Sulfone and Pyridine Derivative                                    45948
Diphenylmethane Derivative                                         43470
Carbapenem derivative                                              42462
Cobalamin Derivative                                               42336
Dihydropyridinecarboxylic acids derivatives                        40572
Phenylpyrimidine derivatives                                       39774
Histamine Analog                                                   39648
Benzimidazole derivative                                           38724
Azoles derivatives                                                 38472
Diphenylethers Derivative                                          35490
```

2-Benzimidazolylcarbamic acid esters	33180
Phenylbutylamine Derivative	31584
Bile Acids and Salts	31542
Thienobenzodiazepine Derivative	31332
Dichlorobenzenes Derivative	31164
Alcohol & phenols	30156
Piperazine Derivative	29400
Biguanides Derivative	28560
Sesquiterpene lactones	28560
Phenylmethyl Piperazinyl Derivative	26670
Valiolamine derivative	24528
Pyrrolidinone & Acetamide Derivative	23478
Dibenzocycloheptenes Derivative	23352
Benzenesulfonamide Derivative	23184
Thiazole Derivative	22680
Narrow-spectrum {First generation cephalosporins}	22638
Oxazolidinone derivative	22218
Nucleoside analog	21840
Narrow- specttrum {First generation cephalospor…	21756

	Total_usage_frequency
Chemical Class	
Fluoroquinolone	35035
Broad Spectrum (Third & fourth generation cepha…	32035
Macrolides	24040
Sulfinylbenzimidazole Derivative	21975
Broad spectrum (Third & fourth generation cepha…	19680
Gluco/mineralocorticoids, progestogins and deri…	13020
Azoles {Triazoles}	12215
Intermediate spectrum {Second generation cephal…	11875
Carbazole Derivative	11390
Glucocorticoids	11190
Azole derivatives {Imidazoles}	10325
Benzodiazepines Derivative	10125
Aminoglycosides	9755
Piperazine Derivatives	9090
P-Aminophenol Derivative	7815
Aminopenicillins {Penicillins}	7640
Pyrrole & heptanoic acid derivative	7525
Anabolic steroid	6790
Timoprazole Derivative	6635
Phenylacetic acid Derivative	6520
Progesterone Derivative	6315
Sulfone and Pyridine Derivative	5470
Diphenylmethane Derivative	5175
Carbapenem derivative	5055
Cobalamin Derivative	5040

```
Dihydropyridinecarboxylic acids derivatives                    4830
Phenylpyrimidine derivatives                                   4735
Histamine Analog                                               4720
Benzimidazole derivative                                       4610
Azoles derivatives                                             4580
Diphenylethers Derivative                                      4225
2-Benzimidazolylcarbamic acid esters                           3950
Phenylbutylamine Derivative                                    3760
Bile Acids and Salts                                           3755
Thienobenzodiazepine Derivative                                3730
Dichlorobenzenes Derivative                                    3710
Alcohol & phenols                                              3590
Piperazine Derivative                                          3500
Biguanides Derivative                                          3400
Sesquiterpene lactones                                         3400
Phenylmethyl Piperazinyl Derivative                            3175
Valiolamine derivative                                         2920
Pyrrolidinone & Acetamide Derivative                           2795
Dibenzocycloheptenes Derivative                                2780
Benzenesulfonamide Derivative                                  2760
Thiazole Derivative                                            2700
Narrow-spectrum {First generation cephalosporins}              2695
Oxazolidinone derivative                                       2645
Nucleoside analog                                              2600
Narrow- specttrum {First generation cephalospor…             2590
```

5.0.5 Top 50 Action Classes by Frequency of Medicines (Excluding 'Unknown')

[376]:
```python
med_frequency_by_action_class = df_cleaned.groupby(['Action Class'])['name'].
 ↪count().loc[lambda x: x.index != "UnKnown"].sort_values(ascending = False)
med_frequency_by_action_class.head(50)
```

[376]:
```
Action Class
Cephalosporins: 3 generation                              10901
Quinolones/ Fluroquinolones                                7011
Fungal ergosterol synthesis inhibitor                      5990
Proton pump inhibitors                                     5734
Macrolides                                                 5109
Glucocorticoids                                            4779
H1 Antihistaminics (second Generation)                     3702
HMG CoA inhibitors (statins)                               2561
Cephalosporins: 2nd generation                             2448
Benzodiazepines                                            2447
NSAID's- Non-Selective COX 1&2 Inhibitors (acetic acid)    2426
Serotonin antagonists (5-HT3 antagonists)                  2369
Atypical Antipsychotics                                    2364
Aminoglycosides                                            1951
```

```
Selective Seretonin Reuptake inhibitors (SSRIs)              1829
Angiotensin receptor blockers(ARB)                           1747
Sodium channel modulators (AED)                              1600
Analgesic & Antipyretic-PCM                                  1563
Cell wall active agent -Extended spectrum Penicillin         1536
Vitamins                                                     1364
Anabolic steroid                                             1358
Natural Progesterone                                         1263
NSAID's -Selective COX-2 Inhibitors                          1213
Beta blocker- Cardioselective                                1206
Calcium channel blockers- Dihydropyridines (DHP)             1166
Cephalosporins: 1st generation                               1057
Cell wall active agent -Carbapenems                          1011
Antiprotozoal agents                                          966
Histamine analog- Meniere's Disease                           944
Nootropic agent                                               944
Sulfonylureas (Insulin Secretogogues)                         896
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)             872
H1 Antihistaminics (First Generation)                         844
Hepatoprotectives                                             837
Antimalarial- Artemisinin and derivatives                     831
Tricyclic antidepressants                                     779
Synaptic vescicle 2 A protein ligand (AED)                    685
Biguanides                                                    680
Alpha-glucosidase inhibitors                                  679
Xanthine oxidase Inhibitors-gout                              655
DPP-4 inhibitors                                              648
Theophylline & its derivatives                                610
Ectoparasiticides                                             609
Progestins (First generation)                                 607
Serotonin-norepinephrine reuptake inhibitors (SNRIs)          582
Alpha 2 delta ligands (AED)                                   579
Proteolytic Enzymes                                           578
Angiotensin-converting enzyme (ACE) inhibitors                568
Phosphodiesterase-V inhibitors                                564
Oxazolidinone                                                 529
Name: name, dtype: int64
```

5.0.6 Top 50 ACtion Classes by Reported Side Effects Count (Excluding 'Unknown')¶

[377]:
```python
side_effect_frequency_by_action_class =df_cleaned.groupby(['Action Class'])[side_effect_columns].count().sum(axis= 1).loc[lambda x: x.index != "UnKnown"].sort_values(ascending = False)
side_effect_frequency_by_action_class.head(50)
```

[377]:
```
Action Class
Cephalosporins: 3 generation                                 457842
```

Quinolones/ Fluroquinolones	294462
Fungal ergosterol synthesis inhibitor	251580
Proton pump inhibitors	240828
Macrolides	214578
Glucocorticoids	200718
H1 Antihistaminics (second Generation)	155484
HMG CoA inhibitors (statins)	107562
Cephalosporins: 2nd generation	102816
Benzodiazepines	102774
NSAID's- Non-Selective COX 1&2 Inhibitors (acetic acid)	101892
Serotonin antagonists (5-HT3 antagonists)	99498
Atypical Antipsychotics	99288
Aminoglycosides	81942
Selective Seretonin Reuptake inhibitors (SSRIs)	76818
Angiotensin receptor blockers(ARB)	73374
Sodium channel modulators (AED)	67200
Analgesic & Antipyretic-PCM	65646
Cell wall active agent -Extended spectrum Penicillin	64512
Vitamins	57288
Anabolic steroid	57036
Natural Progesterone	53046
NSAID's -Selective COX-2 Inhibitors	50946
Beta blocker- Cardioselective	50652
Calcium channel blockers- Dihydropyridines (DHP)	48972
Cephalosporins: 1st generation	44394
Cell wall active agent -Carbapenems	42462
Antiprotozoal agents	40572
Histamine analog- Meniere's Disease	39648
Nootropic agent	39648
Sulfonylureas (Insulin Secretogogues)	37632
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)	36624
H1 Antihistaminics (First Generation)	35448
Hepatoprotectives	35154
Antimalarial- Artemisinin and derivatives	34902
Tricyclic antidepressants	32718
Synaptic vescicle 2 A protein ligand (AED)	28770
Biguanides	28560
Alpha-glucosidase inhibitors	28518
Xanthine oxidase Inhibitors-gout	27510
DPP-4 inhibitors	27216
Theophylline & its derivatives	25620
Ectoparasiticides	25578
Progestins (First generation)	25494
Serotonin-norepinephrine reuptake inhibitors (SNRIs)	24444
Alpha 2 delta ligands (AED)	24318
Proteolytic Enzymes	24276
Angiotensin-converting enzyme (ACE) inhibitors	23856

```
Phosphodiesterase-V inhibitors                                     23688
Oxazolidinone                                                      22218
dtype: int64
```

5.0.7 Top 50 Action Classes by Usage Count Across All Use Columns (Excluding 'Unknown')

[378]:
```python
total_usage_frequency_by_action_class = df_cleaned.groupby(['Action␣
↪Class'])[use_columns].count().sum(axis =1).loc[lambda x: x.index !=␣
↪'UnKnown'].sort_values(ascending = False)
total_usage_frequency_by_action_class.head(50)
```

[378]:
```
Action Class
Cephalosporins: 3 generation                                    54505
Quinolones/ Fluroquinolones                                     35055
Fungal ergosterol synthesis inhibitor                           29950
Proton pump inhibitors                                          28670
Macrolides                                                      25545
Glucocorticoids                                                 23895
H1 Antihistaminics (second Generation)                          18510
HMG CoA inhibitors (statins)                                    12805
Cephalosporins: 2nd generation                                  12240
Benzodiazepines                                                 12235
NSAID's- Non-Selective COX 1&2 Inhibitors (acetic acid)        12130
Serotonin antagonists (5-HT3 antagonists)                       11845
Atypical Antipsychotics                                         11820
Aminoglycosides                                                  9755
Selective Seretonin Reuptake inhibitors (SSRIs)                  9145
Angiotensin receptor blockers(ARB)                               8735
Sodium channel modulators (AED)                                  8000
Analgesic & Antipyretic-PCM                                      7815
Cell wall active agent -Extended spectrum Penicillin             7680
Vitamins                                                         6820
Anabolic steroid                                                 6790
Natural Progesterone                                             6315
NSAID's -Selective COX-2 Inhibitors                              6065
Beta blocker- Cardioselective                                    6030
Calcium channel blockers- Dihydropyridines (DHP)                 5830
Cephalosporins: 1st generation                                   5285
Cell wall active agent -Carbapenems                              5055
Antiprotozoal agents                                             4830
Histamine analog- Meniere's Disease                              4720
Nootropic agent                                                  4720
Sulfonylureas (Insulin Secretogogues)                            4480
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)                4360
H1 Antihistaminics (First Generation)                            4220
Hepatoprotectives                                                4185
```

```
Antimalarial- Artemisinin and derivatives                     4155
Tricyclic antidepressants                                     3895
Synaptic vescicle 2 A protein ligand (AED)                    3425
Biguanides                                                    3400
Alpha-glucosidase inhibitors                                  3395
Xanthine oxidase Inhibitors-gout                              3275
DPP-4 inhibitors                                              3240
Theophylline & its derivatives                                3050
Ectoparasiticides                                             3045
Progestins (First generation)                                 3035
Serotonin-norepinephrine reuptake inhibitors (SNRIs)          2910
Alpha 2 delta ligands (AED)                                   2895
Proteolytic Enzymes                                           2890
Angiotensin-converting enzyme (ACE) inhibitors                2840
Phosphodiesterase-V inhibitors                                2820
Oxazolidinone                                                 2645
dtype: int64
```

5.0.8 Action Data Summary: Usage, Side Effects, and Total Frequency

[379]:
```
all_abt_action = pd.DataFrame({'Med_frequency': med_frequency_by_action_class,
 ↪'Side_effect_frequency': side_effect_frequency_by_action_class,
 ↪'Total_usage_frequency':total_usage_frequency_by_action_class})
all_abt_action.head(50)
```

[379]:
```
                                                   Med_frequency  \
Action Class
Cephalosporins: 3 generation                               10901
Quinolones/ Fluroquinolones                                 7011
Fungal ergosterol synthesis inhibitor                       5990
Proton pump inhibitors                                      5734
Macrolides                                                  5109
Glucocorticoids                                             4779
H1 Antihistaminics (second Generation)                      3702
HMG CoA inhibitors (statins)                                2561
Cephalosporins: 2nd generation                              2448
Benzodiazepines                                             2447
NSAID's- Non-Selective COX 1&2 Inhibitors (acet…          2426
Serotonin antagonists (5-HT3 antagonists)                   2369
Atypical Antipsychotics                                     2364
Aminoglycosides                                             1951
Selective Seretonin Reuptake inhibitors (SSRIs)             1829
Angiotensin receptor blockers(ARB)                          1747
Sodium channel modulators (AED)                             1600
Analgesic & Antipyretic-PCM                                 1563
Cell wall active agent -Extended spectrum Penic…          1536
Vitamins                                                    1364
```

```
Anabolic steroid                                               1358
Natural Progesterone                                           1263
NSAID's -Selective COX-2 Inhibitors                            1213
Beta blocker- Cardioselective                                  1206
Calcium channel blockers- Dihydropyridines (DHP)               1166
Cephalosporins: 1st generation                                 1057
Cell wall active agent -Carbapenems                            1011
Antiprotozoal agents                                            966
Histamine analog- Meniere's Disease                             944
Nootropic agent                                                 944
Sulfonylureas (Insulin Secretogogues)                           896
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)               872
H1 Antihistaminics (First Generation)                           844
Hepatoprotectives                                               837
Antimalarial- Artemisinin and derivatives                       831
Tricyclic antidepressants                                       779
Synaptic vescicle 2 A protein ligand (AED)                      685
Biguanides                                                      680
Alpha-glucosidase inhibitors                                    679
Xanthine oxidase Inhibitors-gout                                655
DPP-4 inhibitors                                                648
Theophylline & its derivatives                                  610
Ectoparasiticides                                               609
Progestins (First generation)                                   607
Serotonin-norepinephrine reuptake inhibitors (S…              582
Alpha 2 delta ligands (AED)                                     579
Proteolytic Enzymes                                             578
Angiotensin-converting enzyme (ACE) inhibitors                  568
Phosphodiesterase-V inhibitors                                  564
Oxazolidinone                                                   529

                                                  Side_effect_frequency  \
Action Class
Cephalosporins: 3 generation                                     457842
Quinolones/ Fluroquinolones                                      294462
Fungal ergosterol synthesis inhibitor                            251580
Proton pump inhibitors                                           240828
Macrolides                                                       214578
Glucocorticoids                                                  200718
H1 Antihistaminics (second Generation)                           155484
HMG CoA inhibitors (statins)                                     107562
Cephalosporins: 2nd generation                                   102816
Benzodiazepines                                                  102774
NSAID's- Non-Selective COX 1&2 Inhibitors (acet…               101892
Serotonin antagonists (5-HT3 antagonists)                         99498
Atypical Antipsychotics                                           99288
Aminoglycosides                                                   81942
```

Selective Seretonin Reuptake inhibitors (SSRIs)	76818
Angiotensin receptor blockers(ARB)	73374
Sodium channel modulators (AED)	67200
Analgesic & Antipyretic-PCM	65646
Cell wall active agent -Extended spectrum Penic…	64512
Vitamins	57288
Anabolic steroid	57036
Natural Progesterone	53046
NSAID's -Selective COX-2 Inhibitors	50946
Beta blocker- Cardioselective	50652
Calcium channel blockers- Dihydropyridines (DHP)	48972
Cephalosporins: 1st generation	44394
Cell wall active agent -Carbapenems	42462
Antiprotozoal agents	40572
Histamine analog- Meniere's Disease	39648
Nootropic agent	39648
Sulfonylureas (Insulin Secretogogues)	37632
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)	36624
H1 Antihistaminics (First Generation)	35448
Hepatoprotectives	35154
Antimalarial- Artemisinin and derivatives	34902
Tricyclic antidepressants	32718
Synaptic vescicle 2 A protein ligand (AED)	28770
Biguanides	28560
Alpha-glucosidase inhibitors	28518
Xanthine oxidase Inhibitors-gout	27510
DPP-4 inhibitors	27216
Theophylline & its derivatives	25620
Ectoparasiticides	25578
Progestins (First generation)	25494
Serotonin-norepinephrine reuptake inhibitors (S…	24444
Alpha 2 delta ligands (AED)	24318
Proteolytic Enzymes	24276
Angiotensin-converting enzyme (ACE) inhibitors	23856
Phosphodiesterase-V inhibitors	23688
Oxazolidinone	22218

	Total_usage_frequency
Action Class	
Cephalosporins: 3 generation	54505
Quinolones/ Fluroquinolones	35055
Fungal ergosterol synthesis inhibitor	29950
Proton pump inhibitors	28670
Macrolides	25545
Glucocorticoids	23895
H1 Antihistaminics (second Generation)	18510
HMG CoA inhibitors (statins)	12805

Cephalosporins: 2nd generation	12240
Benzodiazepines	12235
NSAID's- Non-Selective COX 1&2 Inhibitors (acet…	12130
Serotonin antagonists (5-HT3 antagonists)	11845
Atypical Antipsychotics	11820
Aminoglycosides	9755
Selective Seretonin Reuptake inhibitors (SSRIs)	9145
Angiotensin receptor blockers(ARB)	8735
Sodium channel modulators (AED)	8000
Analgesic & Antipyretic-PCM	7815
Cell wall active agent -Extended spectrum Penic…	7680
Vitamins	6820
Anabolic steroid	6790
Natural Progesterone	6315
NSAID's -Selective COX-2 Inhibitors	6065
Beta blocker- Cardioselective	6030
Calcium channel blockers- Dihydropyridines (DHP)	5830
Cephalosporins: 1st generation	5285
Cell wall active agent -Carbapenems	5055
Antiprotozoal agents	4830
Histamine analog- Meniere's Disease	4720
Nootropic agent	4720
Sulfonylureas (Insulin Secretogogues)	4480
NSAID's-Non-Selective COX 1&2 Inhibitors (Others)	4360
H1 Antihistaminics (First Generation)	4220
Hepatoprotectives	4185
Antimalarial- Artemisinin and derivatives	4155
Tricyclic antidepressants	3895
Synaptic vescicle 2 A protein ligand (AED)	3425
Biguanides	3400
Alpha-glucosidase inhibitors	3395
Xanthine oxidase Inhibitors-gout	3275
DPP-4 inhibitors	3240
Theophylline & its derivatives	3050
Ectoparasiticides	3045
Progestins (First generation)	3035
Serotonin-norepinephrine reuptake inhibitors (S…	2910
Alpha 2 delta ligands (AED)	2895
Proteolytic Enzymes	2890
Angiotensin-converting enzyme (ACE) inhibitors	2840
Phosphodiesterase-V inhibitors	2820
Oxazolidinone	2645

5.0.9 Top 50 Therapeutic Classes by Frequency of Medicines (Excluding 'Unknown')

[380]:
```
med_frequency_by_Therapeutic_class =  df_cleaned.groupby(['Therapeutic␣
↪Class'])['name'].count().loc[lambda x: x.index != 'UnKnown'].
↪sort_values(ascending = False)
med_frequency_by_Therapeutic_class.head(50)
```

[380]:
```
Therapeutic Class
ANTI INFECTIVES               50459
GASTRO INTESTINAL             30442
PAIN ANALGESICS               29187
NEURO CNS                     21524
RESPIRATORY                   21284
CARDIAC                       16774
ANTI DIABETIC                 10457
OPHTHAL                        9494
DERMA                          8949
HORMONES                       4927
GYNAECOLOGICAL                 4094
VITAMINS MINERALS NUTRIENTS    3885
ANTI NEOPLASTICS               3202
BLOOD RELATED                  2464
UROLOGY                        1720
OPHTHAL OTOLOGICALS            1558
ANTI MALARIALS                 1466
SEX STIMULANTS REJUVENATORS     682
OTHERS                          464
OTOLOGICALS                     451
VACCINES                        306
STOMATOLOGICALS                 162
Name: name, dtype: int64
```

5.0.10 Top 50 Therapeutic Classes by Reported Side Effects Count (Excluding 'Unknown')¶

[381]:
```
side_effect_frequency_by_Therapeutic_class = df_cleaned.groupby(['Therapeutic␣
↪Class'])[side_effect_columns].count().sum(axis =1).loc[lambda x: x.index !=␣
↪"UnKnown"].sort_values(ascending = False)
side_effect_frequency_by_Therapeutic_class.head(50)
```

[381]:
```
Therapeutic Class
ANTI INFECTIVES               2119278
GASTRO INTESTINAL             1278564
PAIN ANALGESICS               1225854
NEURO CNS                      904008
RESPIRATORY                    893928
CARDIAC                        704508
```

```
ANTI DIABETIC                   439194
OPHTHAL                         398748
DERMA                           375858
HORMONES                        206934
GYNAECOLOGICAL                  171948
VITAMINS MINERALS NUTRIENTS     163170
ANTI NEOPLASTICS                134484
BLOOD RELATED                   103488
UROLOGY                          72240
OPHTHAL OTOLOGICALS              65436
ANTI MALARIALS                   61572
SEX STIMULANTS REJUVENATORS      28644
OTHERS                           19488
OTOLOGICALS                      18942
VACCINES                         12852
STOMATOLOGICALS                   6804
dtype: int64
```

5.0.11 Top 50 Therapeutic Classes by Usage Count Across All Use Columns (Excluding 'Unknown')

[382]:
```
total_usage_frequency_by_Therapeutic_class = df_cleaned.groupby(['Therapeutic␣
↪Class'])[use_columns].count().sum(axis =1).loc[lambda x:x.index !=␣
↪"UnKnown"].sort_values(ascending = False)
total_usage_frequency_by_Therapeutic_class.head(50)
```

[382]:
```
Therapeutic Class
ANTI INFECTIVES                252295
GASTRO INTESTINAL              152210
PAIN ANALGESICS                145935
NEURO CNS                      107620
RESPIRATORY                    106420
CARDIAC                         83870
ANTI DIABETIC                   52285
OPHTHAL                         47470
DERMA                           44745
HORMONES                        24635
GYNAECOLOGICAL                  20470
VITAMINS MINERALS NUTRIENTS     19425
ANTI NEOPLASTICS                16010
BLOOD RELATED                   12320
UROLOGY                          8600
OPHTHAL OTOLOGICALS              7790
ANTI MALARIALS                   7330
SEX STIMULANTS REJUVENATORS      3410
OTHERS                           2320
OTOLOGICALS                      2255
```

```
VACCINES                            1530
STOMATOLOGICALS                      810
dtype: int64
```

[383]:
```
all_abt_Therapeutic_class = pd.DataFrame({'Med_frequency':␣
 ↪med_frequency_by_Therapeutic_class, 'Side_effect_frequency':␣
 ↪side_effect_frequency_by_Therapeutic_class, 'Total_usage_frequency':
 ↪total_usage_frequency_by_Therapeutic_class})
all_abt_Therapeutic_class.head(50)
```

[383]:
```
                             Med_frequency  Side_effect_frequency  \
Therapeutic Class
ANTI INFECTIVES                      50459                2119278
GASTRO INTESTINAL                    30442                1278564
PAIN ANALGESICS                      29187                1225854
NEURO CNS                            21524                 904008
RESPIRATORY                          21284                 893928
CARDIAC                              16774                 704508
ANTI DIABETIC                        10457                 439194
OPHTHAL                               9494                 398748
DERMA                                 8949                 375858
HORMONES                              4927                 206934
GYNAECOLOGICAL                        4094                 171948
VITAMINS MINERALS NUTRIENTS           3885                 163170
ANTI NEOPLASTICS                      3202                 134484
BLOOD RELATED                         2464                 103488
UROLOGY                               1720                  72240
OPHTHAL OTOLOGICALS                   1558                  65436
ANTI MALARIALS                        1466                  61572
SEX STIMULANTS REJUVENATORS            682                  28644
OTHERS                                 464                  19488
OTOLOGICALS                            451                  18942
VACCINES                               306                  12852
STOMATOLOGICALS                        162                   6804

                             Total_usage_frequency
Therapeutic Class
ANTI INFECTIVES                             252295
GASTRO INTESTINAL                           152210
PAIN ANALGESICS                             145935
NEURO CNS                                   107620
RESPIRATORY                                 106420
CARDIAC                                      83870
ANTI DIABETIC                                52285
OPHTHAL                                      47470
DERMA                                        44745
HORMONES                                     24635
```

```
GYNAECOLOGICAL                                20470
VITAMINS MINERALS NUTRIENTS                   19425
ANTI NEOPLASTICS                              16010
BLOOD RELATED                                 12320
UROLOGY                                        8600
OPHTHAL OTOLOGICALS                            7790
ANTI MALARIALS                                 7330
SEX STIMULANTS REJUVENATORS                    3410
OTHERS                                         2320
OTOLOGICALS                                    2255
VACCINES                                       1530
STOMATOLOGICALS                                 810
```

5.0.12 Creating Separate DataFrames for Drugs Based on Substitute Count

df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==5][['name','total_substitute_counts']] df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==4][['name','total_substitute_counts']] df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==3][['name','total_substitute_counts']] df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==2][['name','total_substitute_counts']] df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==1][['name','total_substitute_counts']] df_cleaned[df_cleaned[substitute_columns].apply(lambda x: x != 'UnKnown').sum(axis =1)==0][['name','total_substitute_counts']]

[384]:
```python
for i in range( 6):
    globals()[f"substitute_{i}"] = df_cleaned[df_cleaned[substitute_columns].
 ↪apply(lambda x: x != 'UnKnown').sum(axis=1) == i][['name',␣
 ↪'total_substitute_counts']]
```

[385]:
```python
#name of the medicine that has only 0 substitute
substitute_0.head(5)
```

[385]:
```
                         name  total_substitute_counts
14            ambrodil-s syrup                        0
72           aquasol a capsule                        0
86           aerocort inhaler                         0
93               atarax drops                         0
95  antid 300mcg/ml injection                         0
```

[386]:
```python
#name of the medicine that has only 1 substitute
substitute_1.head(5)
```

[386]:
```
                                 name  total_substitute_counts
4                      avil 25 tablet                        1
```

```
35              alkasol oral solution                          1
108  alkasol oral solution sugar free                          1
124                 avanair 100 tablet                         1
141            ambrolite-s expectorant                         1
```

[387]:
```python
#name of the medicine that has only 2 substitute
substitute_2.head(5)
```

[387]:
```
                                  name  total_substitute_counts
53   alex cough lozenges lemon ginger                         2
116                     alerid-d tablet                        2
147        adrenaline tartrate injection                       2
161                 addnok 0.2mg tablet                        2
234         asthalin respirator solution                       2
```

[388]:
```python
#name of the medicine that has only 3 substitute
substitute_3.head(5)
```

[388]:
```
                            name  total_substitute_counts
12                 anovate cream                         3
63                 af kit tablet                         3
82   angispan - tr 2.5mg capsule                         3
123       ajaduo 25mg/5mg tablet                         3
130        aldosmin 500mg tablet                         3
```

[389]:
```python
#name of the medicine that has only 4 substitute
substitute_4.head(5)
```

[389]:
```
                                 name  total_substitute_counts
49   asthakind expectorant sugar free                        4
54                  asthalin respules                        4
55                     avil injection                        4
133                  af 150 tablet dt                        4
138                   amantrel tablet                        4
```

[390]:
```python
#name of the medicine that has only 5 substitute
substitute_5.head(5)
```

[390]:
```
                        name  total_substitute_counts
0  augmentin 625 duo tablet                        5
1       azithral 500 tablet                        5
2          ascoril ls syrup                        5
3      allegra 120mg tablet                        5
5          allegra-m tablet                        5
```

5.0.13 Count of Habit-Forming Drugs by Chemical Class

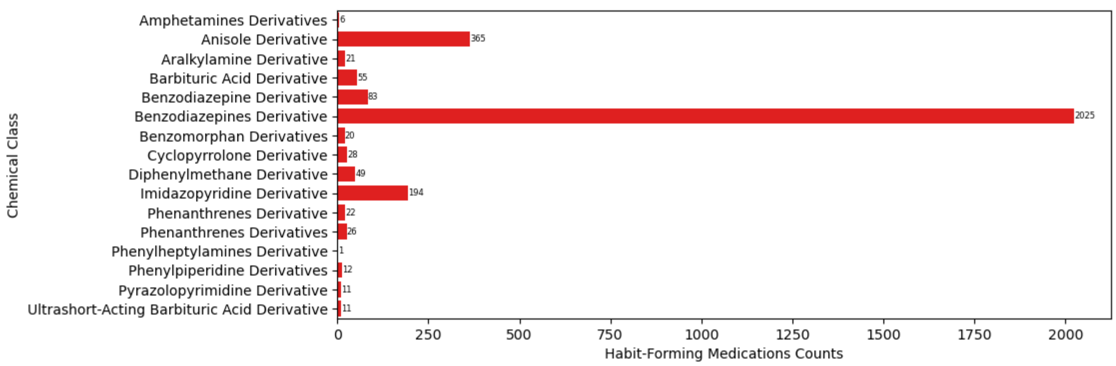
[391]:
```
habit_forming_counts_by_chemical_class = df_cleaned[df_cleaned['Habit Forming']␣
↪== 'Yes'].groupby(['Chemical Class'])['Habit Forming'].count().loc[lambda x:
↪x.index != 'UnKnown']
habit_forming_counts_by_chemical_class = pd.
↪DataFrame(habit_forming_counts_by_chemical_class).reset_index()
habit_forming_counts_by_chemical_class = habit_forming_counts_by_chemical_class.
↪rename(columns = {'Habit Forming' : 'Habit-Forming Medications per Chemical␣
↪Class'})
habit_forming_counts_by_chemical_class
```

[391]:
```
                              Chemical Class  \
0                    Amphetamines Derivatives
1                          Anisole Derivative
2                      Aralkylamine Derivative
3                  Barbituric Acid Derivative
4                   Benzodiazepine Derivative
5                  Benzodiazepines Derivative
6                    Benzomorphan Derivatives
7                   Cyclopyrrolone Derivative
8                  Diphenylmethane Derivative
9                  Imidazopyridine Derivative
10                   Phenanthrenes Derivative
11                  Phenanthrenes Derivatives
12               Phenylheptylamines Derivative
13               Phenylpiperidine Derivatives
14              Pyrazolopyrimidine Derivative
15  Ultrashort-Acting Barbituric Acid Derivative

    Habit-Forming Medications per Chemical Class
0                                              6
1                                            365
2                                             21
3                                             55
4                                             83
5                                           2025
6                                             20
7                                             28
8                                             49
9                                            194
10                                            22
11                                            26
12                                             1
13                                            12
14                                            11
15                                            11
```

[392]:
```
plt.figure(figsize=(10,4))
ax = sns.barplot(data=habit_forming_counts_by_chemical_class,y='Chemical␣
↪Class',x='Habit-Forming Medications per Chemical Class',color='red')
ax.bar_label(ax.containers[0], fontsize=6)
plt.xlabel('Habit-Forming Medications Counts')
```

[392]: Text(0.5, 0, 'Habit-Forming Medications Counts')



5.0.14 Count of Habit-Forming Drugs by Action Class

[393]:
```
habit_forming_counts_by_action_class = df_cleaned[df_cleaned['Habit Forming']␣
↪== 'Yes'].groupby(['Action Class'])['Habit Forming'].count().loc[lambda x:x.
↪index != 'UnKnown']
habit_forming_counts_by_action_class = pd.
↪DataFrame(habit_forming_counts_by_action_class).reset_index()
habit_forming_counts_by_action_class = habit_forming_counts_by_action_class.
↪rename(columns = {'Habit Forming' : 'Habit-Forming Medications per Action␣
↪Class'})
habit_forming_counts_by_action_class
```

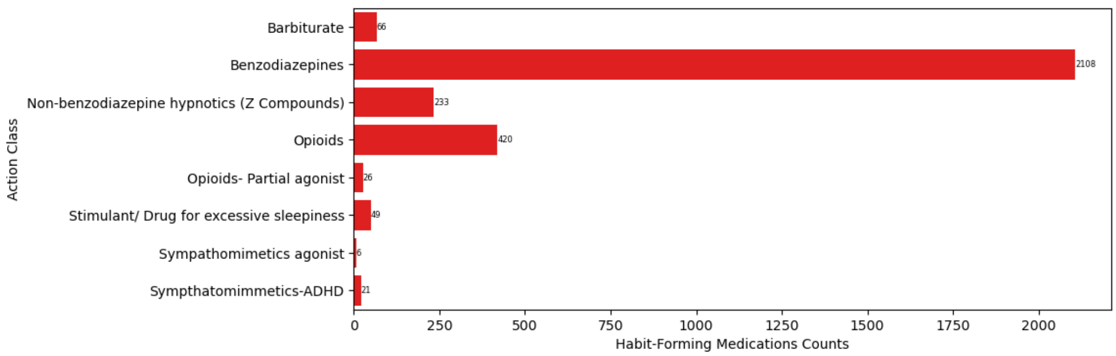
[393]:
```
                                    Action Class  \
0                                    Barbiturate
1                                Benzodiazepines
2   Non-benzodiazepine hypnotics (Z Compounds)
3                                        Opioids
4                        Opioids- Partial agonist
5     Stimulant/ Drug for excessive sleepiness
6                        Sympathomimetics agonist
7                          Sympthatomimmetics-ADHD

   Habit-Forming Medications per Action Class
0                                          66
1                                        2108
```

```
2                                        233
3                                        420
4                                         26
5                                         49
6                                          6
7                                         21
```

[394]:
```
plt.figure(figsize=(10,4))
ax = sns.barplot(data=habit_forming_counts_by_action_class,y='Action␣
↪Class',x='Habit-Forming Medications per Action Class',color='red')
ax.bar_label(ax.containers[0], fontsize=6)
plt.xlabel('Habit-Forming Medications Counts')
```

[394]:
```
Text(0.5, 0, 'Habit-Forming Medications Counts')
```



5.0.15 Count of Habit-Forming Drugs by Therapeutic Class

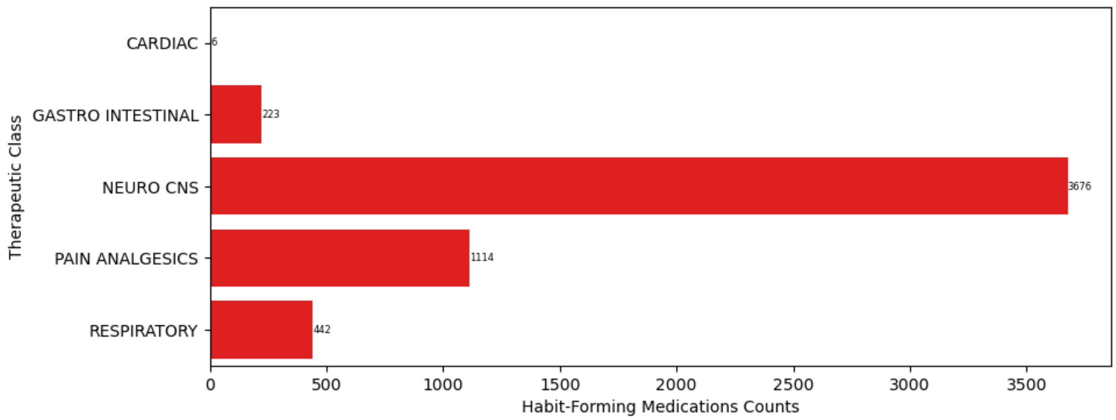
[395]:
```
habit_forming_counts_by_Therapeutic_class = df_cleaned[df_cleaned['Habit␣
↪Forming'] == 'Yes'].groupby(['Therapeutic Class'])['Habit Forming'].count().
↪loc[lambda x:x.index != 'UnKnown']
habit_forming_counts_by_Therapeutic_class = pd.
↪DataFrame(habit_forming_counts_by_Therapeutic_class).reset_index()
habit_forming_counts_by_Therapeutic_class =␣
↪habit_forming_counts_by_Therapeutic_class.rename(columns = {'Habit Forming' :
↪ 'Habit-Forming Medications per Therapeutic Class'})
habit_forming_counts_by_Therapeutic_class
```

[395]:

	Therapeutic Class	Habit-Forming Medications per Therapeutic Class
0	CARDIAC	6
1	GASTRO INTESTINAL	223
2	NEURO CNS	3676
3	PAIN ANALGESICS	1114
4	RESPIRATORY	442

[396]:
```python
plt.figure(figsize=(10,4))
ax = sns.barplot(data=habit_forming_counts_by_Therapeutic_class,y='Therapeutic␣
 ↪Class',x='Habit-Forming Medications per Therapeutic Class',color='red')
ax.bar_label(ax.containers[0], fontsize=6)
plt.xlabel('Habit-Forming Medications Counts')
```

[396]: Text(0.5, 0, 'Habit-Forming Medications Counts')



5.0.16 Identify the medicine that has less than 5 Side Effects

[397]:
```python
df_cleaned[df_cleaned['total_side_effect_recorded']<=5][['name','total_side_effect_recorded']]
```

[397]:

	name	total_side_effect_recorded
0	augmentin 625 duo tablet	3
1	azithral 500 tablet	4
3	allegra 120mg tablet	4
4	avil 25 tablet	2
6	amoxyclav 625 tablet	3
…	…	…
248209	zef cv 200mg/125mg tablet	5
248210	ziyapod 100mg oral suspension	3
248215	zivex 25mg tablet	5
248216	zi fast 500mg injection	5
248217	zyvocol 1% dusting powder	4

[104685 rows x 2 columns]

5.0.17 Discussion And Conclusion :"250k Medicines Usage, Side Effects, and Substitutes"

5.0.18 Project Overview:

The goal of this project was to analyze a dataset that includes detailed information about over 250,000 medicines. This dataset covers attributes such as the drug names, their substitutes, side effects, therapeutic uses, chemical classes, and more. By analyzing this data, I aimed to identify patterns in medicine usage, common side effects, and the availability of substitutes, with the ultimate goal of gaining insights that can be useful in pharmaceutical research, healthcare, and drug safety.

5.0.19 Step-by-Step Analysis:

1. **Data Loading and Initial Exploration**: I started by loading the dataset using Python (with Pandas) and conducted an initial inspection. This step involved:
   - Checking the basic structure of the dataset.
   - Identifying columns and their data types.
   - Reviewing the first few rows to get a sense of the data.
2. **Data Cleaning**: In this step, I addressed missing values and inconsistencies:
   - I filled missing data with "Unknown" or removed rows/columns that contained too many missing values.
   - This ensured that the analysis could proceed without errors or bias due to missing data.
   - I also removed duplicates to maintain data integrity and normalized text columns (e.g., lowercasing and stripping whitespace) for consistency.
3. **Substitute Analysis**: I analyzed the availability of substitutes for each medicine:
   - I counted how many substitutes each drug had by considering the columns for substitute names.
   - I explored the top medicines with the most substitutes to understand which drugs have readily available alternatives.
   - This is valuable in understanding the drug market and providing patients with more treatment options.
4. **Side Effect Analysis**: Side effects were another key focus:
   - I combined the various columns for side effects (up to 41 different columns) into one to assess the frequency of each side effect.
   - I identified the most common side effects across all medicines, such as nausea, vomiting, diarrhea, and abdominal pain, which could indicate general trends in medication safety.
   - This analysis helps in pharmacovigilance by identifying patterns in adverse reactions.
5. **Use Case Analysis**: I explored the therapeutic uses of the medicines by:
   - Combining the columns for different uses (up to 5 different uses per drug) and identifying the most common treatments provided by these medicines.
   - I found that conditions such as bacterial infections, pain relief, and type 2 diabetes treatment were among the most frequent.
   - This analysis provides insights into the most prevalent conditions treated by the pharmaceutical industry.
6. **Classification Analysis**: To better understand patterns in drug properties, I grouped medicines based on:
   - **Chemical Class**: Medicines with similar chemical structures were grouped, and I analyzed how side effects varied across these groups.

- **Therapeutic Class**: I classified medicines according to their therapeutic uses (e.g., respiratory, anti-infectives).
- **Action Class**: Medicines were also grouped by their actions (e.g., H2 receptor blockers), and I examined the side effects associated with these classes.
- This classification allowed I to identify trends and similarities across medicines with similar intended purposes or mechanisms.

7. **Visualization**: To enhance the understanding of Ir findings, I created visualizations:
   - I used tools like Matplotlib and Seaborn to visualize distributions, correlations, and patterns in side effects, substitutes, and usage.
   - This provided clear, interpretable results for stakeholders in drug research or healthcare.
8. **Advanced Analysis (Optional)**:
   - The project suggests further exploration using clustering or machine learning techniques, such as building models to predict side effects based on the chemical or therapeutic class of a medicine.
   - This could help in automating the process of identifying potential adverse reactions, thus aiding drug development and safety protocols.

5.0.20 Conclusion:

The project successfully explored the relationships between medicines, their substitutes, side effects, and therapeutic uses. Key findings included identifying the most common side effects (such as nausea, diarrhea, and vomiting), recognizing medicines with the most substitutes, and analyzing the variation of these factors across different chemical, therapeutic, and action classes. The insights derived from this project are valuable for drug safety, healthcare planning, and research in the pharmaceutical industry.

This analysis could be extended with more advanced techniques such as clustering, predictive modeling, and deeper exploration of correlations between drug classes and their effects, helping to refine drug recommendations or safety guidelines.

[ ]:

```
```