Healthcare Recommendation

April 24, 2024

[853]:
```
import pandas as pd
import numpy as np
from IPython.display import display
```

[2]:
```
df_drug_neg = pd.read_csv("Project Data/ChCh-Miner_durgbank-chem-chem.tsv",␣
 ↪sep="\t", header=None)
df_drug_neg
```

[2]:

	0	1
0	DB00862	DB00966
1	DB00575	DB00806
2	DB01242	DB08893
3	DB01151	DB08883
4	DB01235	DB01275
…	…	…
48509	DB00542	DB01354
48510	DB00476	DB01239
48511	DB00621	DB01120
48512	DB00808	DB01356
48513	DB00677	DB06287

[48514 rows x 2 columns]

[3]:
```
df_rev = pd.DataFrame(df_drug_neg[[1,0]])
df_rev
```

[3]:

	1	0
0	DB00966	DB00862
1	DB00806	DB00575
2	DB08893	DB01242
3	DB08883	DB01151
4	DB01275	DB01235
…	…	…
48509	DB01354	DB00542
48510	DB01239	DB00476
48511	DB01120	DB00621
48512	DB01356	DB00808
48513	DB06287	DB00677

```
[48514 rows x 2 columns]
```

[4]:
```
df_drug_neg_final = pd.concat((df_drug_neg, df_rev), ignore_index=True).
 ↪sort_values(by=0).reset_index(drop=True)
df_drug_neg_final
```

[4]:
```
              0        1
0       DB00005  DB08879
1       DB00005  DB08904
2       DB00005  DB00072
3       DB00005  DB00531
4       DB00005  DB06273
…         …       …
97023   DB09292  DB09477
97024   DB09330  DB09401
97025   DB09330  DB09570
97026   DB09330  DB09570
97027   DB09330  DB09401

[97028 rows x 2 columns]
```

[902]:
```
df_drug_vocab = pd.read_csv("Project Data/drugbank vocabulary.csv")
df_drug_vocab['Synonyms'] = df_drug_vocab['Common name'] + '|' +␣
 ↪df_drug_vocab['Synonyms']
df_drug_vocab.drop(columns='Common name', inplace=True)
df_drug_vocab
```

[902]:
```
       DrugBank ID     Accession Numbers            CAS        UNII  \
0          DB00001  BTD00024 | BIOD00024    138068-37-8  Y43GF64R34
1          DB00002  BTD00071 | BIOD00071    205923-56-4  PQX0D8J21J
2          DB00003  BTD00001 | BIOD00001    143831-71-4  953A26OA1Y
3          DB00004  BTD00084 | BIOD00084    173146-27-5  25E79B5CTM
4          DB00005  BTD00052 | BIOD00052    185243-69-0  OP401G7OJC
…            …                    …             …          …
16576      DB18713                  NaN           NaN         NaN
16577      DB18714                  NaN           NaN         NaN
16578      DB18715                  NaN  1971920-73-6  8CZ82ZYY9X
16579      DB18716                  NaN  1001404-83-6  80VUN7L00C
16580      DB18717                  NaN     7782-42-5  4QQN74LH4O

                                                 Synonyms  \
0       Lepirudin|[Leu1, Thr2]-63-desulfohirudin | Des…
1           Cetuximab|Cetuximab | Cétuximab | Cetuximabum
2       Dornase alfa|Deoxyribonuclease (human clone 18…
3       Denileukin diftitox|Denileukin | Denileukin di…
4       Etanercept|Etanercept | etanercept-szzs | etan…
```

```
…                                                     …
16576  Recombinant stabilized RSV A prefusion F antig…
16577  Recombinant stabilized RSV B prefusion F antig…
16578  Tolebrutinib|2h-imidazo(4,5-c)pyridin-2-one, 4…
16579  Enmetazobactam|(2s,3s,5r)-3-methyl-3-((3-methy…
16580  Graphite|Black lead | C.i. pigment black 10 | …

                Standard InChI Key
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN
…                            …
16576                        NaN
16577                        NaN
16578  KOEUOFPEZFUWRF-LJQANCHMSA-N
16579  HFZITXBUTWITPT-YWVKMMECSA-N
16580                        NaN

[16581 rows x 6 columns]
```

[903]:
```
df_uci_train = pd.read_csv("Project Data/drugsComTrain_raw.csv")
df_uci_train
```

[903]:
```
        uniqueID                 drugName                     condition  \
0         206461                Valsartan  Left Ventricular Dysfunction
1          95260                Guanfacine                          ADHD
2          92703                    Lybrel               Birth Control
3         138000                Ortho Evra               Birth Control
4          35696  Buprenorphine / naloxone           Opiate Dependence
…            …                        …                           …
161292    191035                   Campral          Alcohol Dependence
161293    127085            Metoclopramide             Nausea/Vomiting
161294    187382                   Orencia        Rheumatoid Arthritis
161295     47128        Thyroid desiccated         Underactive Thyroid
161296    215220              Lubiprostone       Constipation, Chronic

                                                   review  rating       date  \
0       "It has no side effect, I take it in combinati…       9  20-May-12
1       "My son is halfway through his fourth week of …       8  27-Apr-10
2       "I used to take another oral contraceptive, wh…       5  14-Dec-09
3       "This is my first time using any form of birth…       8   3-Nov-15
4       "Suboxone has completely turned my life around…       9  27-Nov-16
…                                                     …     …         …
161292  "I wrote my first report in Mid-October of 201…      10  31-May-15
161293  "I was given this in IV before surgey. I immed…       1   1-Nov-11
```

```
161294  "Limited improvement after 4 months, developed…       2  15-Mar-14
161295  "I&#039;ve been on thyroid medication 49 years…      10  19-Sep-15
161296  "I&#039;ve had chronic constipation all my adu…       9  13-Dec-14

        usefulCount
0                27
1               192
2                17
3                10
4                37
…               …
161292          125
161293           34
161294           35
161295           79
161296          116

[161297 rows x 7 columns]
```

[1168]:
```
df_uci_train[df_uci_train.drugName=='Erythromycin']
```

[1168]:
```
        uniqueID     drugName                     condition  \
6637      126857  Erythromycin  Upper Respiratory Tract In…
18659     126861  Erythromycin  Upper Respiratory Tract In…
30442     126846  Erythromycin                Dental Abscess
39198     126841  Erythromycin  Upper Respiratory Tract In…
40504     126849  Erythromycin                Dental Abscess
43835     126858  Erythromycin  Upper Respiratory Tract In…
45883     202824  Erythromycin     Conjunctivitis, Bacterial
55699     126856  Erythromycin                   Pharyngitis
72034     126855  Erythromycin                   Pharyngitis
74106     126860  Erythromycin                    Bronchitis
80220     126843  Erythromycin                Dental Abscess
100896    126848  Erythromycin                  Otitis Media
118230    126851  Erythromycin                  Strep Throat
131186    126859  Erythromycin  Skin or Soft Tissue Infection
131784    126838  Erythromycin                Dental Abscess
144832    126847  Erythromycin                   Pharyngitis
148737    126845  Erythromycin  Upper Respiratory Tract In…
156440    126850  Erythromycin                    Bronchitis
159521    119683  Erythromycin                          Acne

                             review  rating       date  usefulCount
6637    "I was prescribed this med…       9  15-Nov-09           31
18659   "I was prescribed to take …       8  20-Nov-08           32
30442   "I&#039;m currently taking…       2   4-Jan-14           25
39198   "I am taking LOFEPRAMINE a…       1  16-Jul-16            1
```

```
40504   "I was put on erythromycin…       1  30-Jun-12          26
43835   "I have been diagnosed wit…      10  16-Oct-09          17
45883   "I prefer the drops to the…       8   2-Jan-14          12
55699   "I have been poorly for 3 …       6  30-Mar-10           6
72034   "My experience at first wa…      10   5-Apr-10           6
74106   "I&#039;ve always had good…       7  26-Sep-09          36
80220   "This medicine gave me ext…       1  27-Aug-15           6
100896  "I&#039;m taking this due …       7   8-Jul-12          19
118230  "I am a skeptic on taking …      10   2-Dec-11          29
131186  "This drug worked miracles…       5   1-Oct-09          39
131784  "I&rsquo;ve had an abscess…      10  29-Oct-17           0
144832  "This is very effective fo…       7  26-Dec-13           6
148737  "Almost immediate relief w…      10  11-Dec-14          21
156440  "I&#039;ve been taking it …       6   4-May-12          17
159521  "This is a very good treat…      10  23-Jun-09          27
```

[904]:
```
df_drug_vocab['Synonyms'] = df_drug_vocab['Synonyms'].str.split('|')
df_drug_vocab_exploded = df_drug_vocab.explode('Synonyms')
df_drug_vocab_exploded['Synonyms'] = df_drug_vocab_exploded['Synonyms'].str.
 ↪strip()
df_drug_vocab_exploded = df_drug_vocab_exploded.
 ↪drop_duplicates(subset=['DrugBank ID','Synonyms'])
df_drug_vocab_exploded[df_drug_vocab_exploded.Synonyms=='Etanercept']
```

[904]:
```
  DrugBank ID     Accession Numbers          CAS        UNII    Synonyms  \
4     DB00005  BTD00052 | BIOD00052  185243-69-0  OP401G7OJC  Etanercept

  Standard InChI Key
4                NaN
```

[905]:
```
df_drug_vocab_exploded[df_drug_vocab_exploded['Synonyms']=='Valsartan']
```

[905]:
```
    DrugBank ID Accession Numbers          CAS        UNII   Synonyms  \
166     DB00177         APRD00133  137862-53-4  80M03YXJ7I  Valsartan

              Standard InChI Key
166  ACWBQPMHZXGDFX-QFIPXVFZSA-N
```

[915]:
```
df_uci_train_updated = pd.merge(df_uci_train,␣
 ↪df_drug_vocab_exploded[['Synonyms', 'DrugBank ID']], left_on="drugName",␣
 ↪right_on="Synonyms").drop(columns='Synonyms')
df_uci_train_updated.sort_values(by='DrugBank ID')
```

[915]:
```
       uniqueID    drugName                     condition  \
73782     18312   Cetuximab              Colorectal Cance
73781     18318   Cetuximab              Colorectal Cance
73780     18315   Cetuximab       Squamous Cell Carcinoma
```

```
73779     18316   Cetuximab              Colorectal Cance
73778     18314   Cetuximab           Head and Neck Cance
…           …         …                          …
71533    215349       Senna                  Constipation
71532    215342       Senna                  Constipation
71531    215406       Senna                  Constipation
71562    215394       Senna                  Constipation
74558    158936  Goldenseal  Skin and Structure Infection

                                                  review  rating       date  \
73782  "I have stage 4 colon cancer with liver mets. …       9  28-Dec-16
73781  "I have Stage 4 colorectal cancer with mets to…       8   1-Jun-11
73780  "Taking for inoperable squamous cell thyroid. …       7   6-Oct-13
73779  "Was also diagnosed with stage four colon canc…       8  19-Sep-13
73778  "I heard good things about this drug for metas…       1  16-Oct-15
…                                                    …     …         …
71533  "I have constipation regularly. Milk of magnes…      10   3-Oct-17
71532  "I use this for 2 days before  I was able to p…       6   2-Dec-17
71531  "Easy to use. No pain just gentle comfort. Exc…      10  26-Jan-10
71562  "Took only one tablet (directions said to take…       5  21-Sep-15
74558  "Can&#039;t breathe over a year. My allergies …      10  22-Jul-17

       usefulCount DrugBank ID
73782            2     DB00002
73781           47     DB00002
73780            8     DB00002
73779            7     DB00002
73778            1     DB00002
…              …           …
71533            3     DB15889
71532            0     DB15889
71531           65     DB15889
71562           34     DB15889
74558            1     DB16594

[74788 rows x 8 columns]
```

[1149]:
```
df_uci_train_updated[df_uci_train_updated.drugName=='Etanercept']
```

[1149]:
```
       uniqueID    drugName               condition  \
11140     41928  Etanercept    Rheumatoid Arthritis
11141     12780  Etanercept  Ankylosing Spondylitis
11142     12651  Etanercept  Ankylosing Spondylitis
11143     12764  Etanercept  Ankylosing Spondylitis
11144     12733  Etanercept    Rheumatoid Arthritis
…           …         …                      …
11287     41942  Etanercept    Rheumatoid Arthritis
```

```
11288     12684  Etanercept        Plaque Psoriasis
11289     41945  Etanercept        Behcet's Disease
11290     12843  Etanercept               Psoriasis
11291     12844  Etanercept               Psoriasis

                              review  rating       date  usefulCount     DBID
11140  "I live in Western Austral…      10  16-Sep-17           4  DB00005
11141  "I started using Enbrel tw…      10   5-Nov-09          22  DB00005
11142  "I am 35 years old and hav…       9   8-Oct-16           5  DB00005
11143  "I was diagnosed with anky…      10  18-Jul-12          34  DB00005
11144  "I&#039;ve had RA for 8 ye…      10  22-Nov-15          39  DB00005
…                                     …     …         …       …         …
11287  "Enbrel has done a fantast…      10  18-Mar-17          14  DB00005
11288  "Enbrel worked fantastic w…       8   6-Oct-13          38  DB00005
11289  "I&#039;ve been on this dr…       8  10-Feb-17           0  DB00005
11290  "I have been on Enbrel for…       9   5-Aug-10           9  DB00005
11291  "I personally was very ple…      10  26-Feb-08           3  DB00005

[152 rows x 8 columns]
```

[917]:
```
df_uci_train_updated.dropna(inplace=True)
df_uci_train_updated.rename(columns={'DrugBank ID': 'DBID'}, inplace=True)
df_uci_train_updated
```

[917]:
```
       uniqueID        drugName                    condition  \
0        206461       Valsartan  Left Ventricular Dysfunction
1        206546       Valsartan           High Blood Pressure
2        206572       Valsartan           High Blood Pressure
3        206538       Valsartan           High Blood Pressure
4        206462       Valsartan           High Blood Pressure
…           …              …                             …
74783       881      Clemastine             Allergic Rhinitis
74784    148324  Valganciclovir               CMV Prophylaxis
74785      4237        Acarbose              Diabetes, Type 2
74786     20575        Zileuton           Asthma, Maintenance
74787    173878        Oxytocin               Labor Induction

                                              review  rating       date  \
0      "It has no side effect, I take it in combinati…       9  20-May-12
1      "I took Diovan for two months to reduce blood …       6  22-Mar-09
2      "Been on Diovan for 9 years. Started with ring…       6   8-Nov-09
3      "I was prescribed Valsartan (generic of Diovan…       4  26-Apr-16
4                   "Experience tiredness, no energy."       8  13-Feb-12
…                                                     …     …         …
74783  "I have used this medicine for a number of yea…      10  27-Jan-10
74784  "Valgan takes care of cmv well I had a spk tra…       8  25-Oct-16
74785  "I took acarbose for 2 months. It did nothing …       1   4-Jun-16
```

```
74786  "I was given this for Allergic rhinitis . I go…       1   8-Jun-16
74787  "They used Pitocin to induce all of my three l…      10  15-Aug-14

       usefulCount     DBID
0               27  DB00177
1               40  DB00177
2               14  DB00177
3               46  DB00177
4              118  DB00177
…              …       …
74783           11  DB00283
74784            2  DB01610
74785            4  DB00284
74786            2  DB00744
74787           16  DB00107

[74354 rows x 8 columns]
```

[ ]:

[918]:
```
df_drug_neg_temp = pd.merge(df_drug_neg_final,␣
↪df_uci_train_updated[['drugName', 'DBID']], left_on=1, right_on='DBID').
↪drop(columns=['DBID',0])
df_drug_map = df_drug_neg_temp.groupby([1,'drugName']).count().reset_index()
df_drug_map = dict(zip(df_drug_map[1], df_drug_map['drugName']))

df_drug_map
```

[918]:
```
{'DB00014': 'Goserelin',
 'DB00026': 'Anakinra',
 'DB00030': 'Insulin regular',
 'DB00046': 'Insulin lispro',
 'DB00047': 'Insulin glargine',
 'DB00051': 'Adalimumab',
 'DB00065': 'Infliximab',
 'DB00072': 'Trastuzumab',
 'DB00073': 'Rituximab',
 'DB00083': 'OnabotulinumtoxinA',
 'DB00086': 'Streptokinase',
 'DB00087': 'Alemtuzumab',
 'DB00091': 'Cyclosporine',
 'DB00104': 'Octreotide',
 'DB00107': 'Oxytocin',
 'DB00108': 'Natalizumab',
 'DB00126': 'Vitamin C',
 'DB00159': 'Icosapent',
 'DB00175': 'Pravastatin',
```

```
'DB00176': 'Fluvoxamine',
'DB00177': 'Valsartan',
'DB00178': 'Ramipril',
'DB00181': 'Baclofen',
'DB00182': 'Amphetamine',
'DB00186': 'Lorazepam',
'DB00188': 'Bortezomib',
'DB00191': 'Phentermine',
'DB00193': 'Tramadol',
'DB00196': 'Fluconazole',
'DB00199': 'Erythromycin',
'DB00201': 'Caffeine',
'DB00202': 'Succinylcholine',
'DB00203': 'Sildenafil',
'DB00204': 'Dofetilide',
'DB00205': 'Pyrimethamine',
'DB00206': 'Reserpine',
'DB00207': 'Azithromycin',
'DB00211': 'Midodrine',
'DB00213': 'Pantoprazole',
'DB00214': 'Torsemide',
'DB00215': 'Citalopram',
'DB00216': 'Eletriptan',
'DB00218': 'Moxifloxacin',
'DB00222': 'Glimepiride',
'DB00223': 'Diflorasone',
'DB00227': 'Lovastatin',
'DB00230': 'Pregabalin',
'DB00231': 'Temazepam',
'DB00237': 'Butabarbital',
'DB00238': 'Nevirapine',
'DB00240': 'Alclometasone',
'DB00242': 'Cladribine',
'DB00243': 'Ranolazine',
'DB00244': 'Mesalamine',
'DB00245': 'Benztropine',
'DB00246': 'Ziprasidone',
'DB00248': 'Cabergoline',
'DB00250': 'Dapsone',
'DB00252': 'Phenytoin',
'DB00254': 'Doxycycline',
'DB00257': 'Clotrimazole',
'DB00258': 'Calcium acetate',
'DB00264': 'Metoprolol',
'DB00268': 'Ropinirole',
'DB00270': 'Isradipine',
'DB00273': 'Topiramate',
```

```
'DB00275': 'Olmesartan',
'DB00277': 'Theophylline',
'DB00279': 'Liothyronine',
'DB00280': 'Disopyramide',
'DB00281': 'Lidocaine',
'DB00283': 'Clemastine',
'DB00284': 'Acarbose',
'DB00285': 'Venlafaxine',
'DB00287': 'Travoprost',
'DB00289': 'Atomoxetine',
'DB00291': 'Chlorambucil',
'DB00294': 'Etonogestrel',
'DB00295': 'Morphine',
'DB00296': 'Ropivacaine',
'DB00297': 'Bupivacaine',
'DB00302': 'Tranexamic acid',
'DB00307': 'Bexarotene',
'DB00310': 'Chlorthalidone',
'DB00312': 'Pentobarbital',
'DB00313': 'Valproic acid',
'DB00315': 'Zolmitriptan',
'DB00316': 'Acetaminophen',
'DB00317': 'Gefitinib',
'DB00318': 'Codeine',
'DB00320': 'Dihydroergotamine',
'DB00321': 'Amitriptyline',
'DB00327': 'Hydromorphone',
'DB00328': 'Indomethacin',
'DB00331': 'Metformin',
'DB00332': 'Ipratropium',
'DB00333': 'Methadone',
'DB00334': 'Olanzapine',
'DB00335': 'Atenolol',
'DB00337': 'Pimecrolimus',
'DB00338': 'Omeprazole',
'DB00341': 'Cetirizine',
'DB00343': 'Diltiazem',
'DB00344': 'Protriptyline',
'DB00346': 'Alfuzosin',
'DB00349': 'Clobazam',
'DB00350': 'Minoxidil',
'DB00356': 'Chlorzoxazone',
'DB00358': 'Mefloquine',
'DB00363': 'Clozapine',
'DB00364': 'Sucralfate',
'DB00366': 'Doxylamine',
'DB00367': 'Levonorgestrel',
```

```
'DB00370': 'Mirtazapine',
'DB00371': 'Meprobamate',
'DB00373': 'Timolol',
'DB00374': 'Treprostinil',
'DB00376': 'Trihexyphenidyl',
'DB00377': 'Palonosetron',
'DB00379': 'Mexiletine',
'DB00381': 'Amlodipine',
'DB00384': 'Triamterene',
'DB00387': 'Procyclidine',
'DB00388': 'Phenylephrine',
'DB00390': 'Digoxin',
'DB00393': 'Nimodipine',
'DB00395': 'Carisoprodol',
'DB00396': 'Progesterone',
'DB00398': 'Sorafenib',
'DB00399': 'Zoledronic acid',
'DB00400': 'Griseofulvin',
'DB00401': 'Nisoldipine',
'DB00402': 'Eszopiclone',
'DB00404': 'Alprazolam',
'DB00408': 'Loxapine',
'DB00412': 'Rosiglitazone',
'DB00413': 'Pramipexole',
'DB00415': 'Ampicillin',
'DB00418': 'Secobarbital',
'DB00421': 'Spironolactone',
'DB00422': 'Methylphenidate',
'DB00423': 'Methocarbamol',
'DB00424': 'Hyoscyamine',
'DB00425': 'Zolpidem',
'DB00427': 'Triprolidine',
'DB00433': 'Prochlorperazine',
'DB00434': 'Cyproheptadine',
'DB00437': 'Allopurinol',
'DB00440': 'Trimethoprim',
'DB00441': 'Gemcitabine',
'DB00443': 'Betamethasone',
'DB00446': 'Chloramphenicol',
'DB00448': 'Lansoprazole',
'DB00450': 'Droperidol',
'DB00451': 'Levothyroxine',
'DB00454': 'Meperidine',
'DB00455': 'Loratadine',
'DB00457': 'Prazosin',
'DB00458': 'Imipramine',
'DB00459': 'Acitretin',
```

```
'DB00461': 'Nabumetone',
'DB00465': 'Ketorolac',
'DB00468': 'Quinine',
'DB00470': 'Dronabinol',
'DB00471': 'Montelukast',
'DB00472': 'Fluoxetine',
'DB00475': 'Chlordiazepoxide',
'DB00476': 'Duloxetine',
'DB00477': 'Chlorpromazine',
'DB00479': 'Amikacin',
'DB00480': 'Lenalidomide',
'DB00481': 'Raloxifene',
'DB00482': 'Celecoxib',
'DB00484': 'Brimonidine',
'DB00485': 'Dicloxacillin',
'DB00486': 'Nabilone',
'DB00488': 'Altretamine',
'DB00489': 'Sotalol',
'DB00490': 'Buspirone',
'DB00492': 'Fosinopril',
'DB00493': 'Cefotaxime',
'DB00494': 'Entacapone',
'DB00496': 'Darifenacin',
'DB00497': 'Oxycodone',
'DB00500': 'Tolmetin',
'DB00501': 'Cimetidine',
'DB00502': 'Haloperidol',
'DB00512': 'Vancomycin',
'DB00514': 'Dextromethorphan',
'DB00515': 'Cisplatin',
'DB00518': 'Albendazole',
'DB00519': 'Trandolapril',
'DB00524': 'Metolazone',
'DB00526': 'Oxaliplatin',
'DB00530': 'Erlotinib',
'DB00531': 'Cyclophosphamide',
'DB00535': 'Cefdinir',
'DB00537': 'Ciprofloxacin',
'DB00540': 'Nortriptyline',
'DB00541': 'Vincristine',
'DB00542': 'Benazepril',
'DB00543': 'Amoxapine',
'DB00544': 'Fluorouracil',
'DB00545': 'Pyridostigmine',
'DB00547': 'Desoximetasone',
'DB00549': 'Zafirlukast',
'DB00554': 'Piroxicam',
```

```
'DB00555': 'Lamotrigine',
'DB00557': 'Hydroxyzine',
'DB00559': 'Bosentan',
'DB00563': 'Methotrexate',
'DB00564': 'Carbamazepine',
'DB00567': 'Cephalexin',
'DB00569': 'Fondaparinux',
'DB00571': 'Propranolol',
'DB00572': 'Atropine',
'DB00573': 'Fenoprofen',
'DB00575': 'Clonidine',
'DB00577': 'Valacyclovir',
'DB00581': 'Lactulose',
'DB00583': 'L-Carnitine',
'DB00584': 'Enalapril',
'DB00585': 'Nizatidine',
'DB00586': 'Diclofenac',
'DB00590': 'Doxazosin',
'DB00593': 'Ethosuximide',
'DB00594': 'Amiloride',
'DB00596': 'Halobetasol',
'DB00598': 'Labetalol',
'DB00599': 'Thiopental',
'DB00601': 'Linezolid',
'DB00602': 'Ivermectin',
'DB00605': 'Sulindac',
'DB00611': 'Butorphanol',
'DB00612': 'Bisoprolol',
'DB00618': 'Demeclocycline',
'DB00619': 'Imatinib',
'DB00620': 'Triamcinolone',
'DB00623': 'Fluphenazine',
'DB00624': 'Testosterone',
'DB00625': 'Efavirenz',
'DB00627': 'Niacin',
'DB00628': 'Clorazepate',
'DB00630': 'Alendronate',
'DB00635': 'Prednisone',
'DB00639': 'Butoconazole',
'DB00640': 'Adenosine',
'DB00641': 'Simvastatin',
'DB00642': 'Pemetrexed',
'DB00643': 'Mebendazole',
'DB00652': 'Pentazocine',
'DB00654': 'Latanoprost',
'DB00656': 'Trazodone',
'DB00658': 'Sevelamer',
```

```
'DB00660': 'Metaxalone',
'DB00661': 'Verapamil',
'DB00662': 'Trimethobenzamide',
'DB00666': 'Nafarelin',
'DB00668': 'Epinephrine',
'DB00669': 'Sumatriptan',
'DB00671': 'Cefixime',
'DB00673': 'Aprepitant',
'DB00674': 'Galantamine',
'DB00675': 'Tamoxifen',
'DB00678': 'Losartan',
'DB00679': 'Thioridazine',
'DB00682': 'Warfarin',
'DB00683': 'Midazolam',
'DB00684': 'Tobramycin',
'DB00687': 'Fludrocortisone',
'DB00688': 'Mycophenolate mofetil',
'DB00690': 'Flurazepam',
'DB00691': 'Moexipril',
'DB00695': 'Furosemide',
'DB00696': 'Ergotamine',
'DB00697': 'Tizanidine',
'DB00698': 'Nitrofurantoin',
'DB00700': 'Eplerenone',
'DB00704': 'Naltrexone',
'DB00706': 'Tamsulosin',
'DB00709': 'Lamivudine',
'DB00710': 'Ibandronate',
'DB00712': 'Flurbiprofen',
'DB00714': 'Apomorphine',
'DB00715': 'Paroxetine',
'DB00717': 'Norethindrone',
'DB00722': 'Lisinopril',
'DB00724': 'Imiquimod',
'DB00727': 'Nitroglycerin',
'DB00734': 'Risperidone',
'DB00736': 'Esomeprazole',
'DB00737': 'Meclizine',
'DB00740': 'Riluzole',
'DB00741': 'Hydrocortisone',
'DB00744': 'Zileuton',
'DB00745': 'Modafinil',
'DB00747': 'Scopolamine',
'DB00748': 'Carbinoxamine',
'DB00749': 'Etodolac',
'DB00752': 'Tranylcypromine',
'DB00755': 'Tretinoin',
```

```
'DB00758': 'Clopidogrel',
'DB00759': 'Tetracycline',
'DB00760': 'Meropenem',
'DB00761': 'Potassium chloride',
'DB00762': 'Irinotecan',
'DB00763': 'Methimazole',
'DB00764': 'Mometasone',
'DB00768': 'Olopatadine',
'DB00770': 'Alprostadil',
'DB00776': 'Oxcarbazepine',
'DB00780': 'Phenelzine',
'DB00783': 'Estradiol',
'DB00784': 'Mefenamic acid',
'DB00787': 'Acyclovir',
'DB00788': 'Naproxen',
'DB00790': 'Perindopril',
'DB00794': 'Primidone',
'DB00795': 'Sulfasalazine',
'DB00804': 'Dicyclomine',
'DB00806': 'Pentoxifylline',
'DB00808': 'Indapamide',
'DB00811': 'Ribavirin',
'DB00813': 'Fentanyl',
'DB00814': 'Meloxicam',
'DB00816': 'Metaproterenol',
'DB00818': 'Propofol',
'DB00819': 'Acetazolamide',
'DB00820': 'Tadalafil',
'DB00822': 'Disulfiram',
'DB00829': 'Diazepam',
'DB00831': 'Trifluoperazine',
'DB00834': 'Mifepristone',
'DB00835': 'Brompheniramine',
'DB00836': 'Loperamide',
'DB00838': 'Clocortolone',
'DB00842': 'Oxazepam',
'DB00843': 'Donepezil',
'DB00844': 'Nalbuphine',
'DB00846': 'Flurandrenolide',
'DB00849': 'Mephobarbital',
'DB00850': 'Perphenazine',
'DB00852': 'Pseudoephedrine',
'DB00853': 'Temozolomide',
'DB00854': 'Levorphanol',
'DB00857': 'Terbinafine',
'DB00859': 'Penicillamine',
'DB00860': 'Prednisolone',
```

```
'DB00861': 'Diflunisal',
'DB00862': 'Vardenafil',
'DB00863': 'Ranitidine',
'DB00864': 'Tacrolimus',
'DB00865': 'Benzphetamine',
'DB00869': 'Dorzolamide',
'DB00871': 'Terbutaline',
'DB00872': 'Conivaptan',
'DB00880': 'Chlorothiazide',
'DB00881': 'Quinapril',
'DB00883': 'Isosorbide dinitrate',
'DB00884': 'Risedronate',
'DB00887': 'Bumetanide',
'DB00889': 'Granisetron',
'DB00897': 'Triazolam',
'DB00903': 'Ethacrynic acid',
'DB00904': 'Ondansetron',
'DB00905': 'Bimatoprost',
'DB00906': 'Tiagabine',
'DB00908': 'Quinidine',
'DB00909': 'Zonisamide',
'DB00911': 'Tinidazole',
'DB00912': 'Repaglinide',
'DB00915': 'Amantadine',
'DB00916': 'Metronidazole',
'DB00917': 'Dinoprostone',
'DB00918': 'Almotriptan',
'DB00921': 'Buprenorphine',
'DB00924': 'Cyclobenzaprine',
'DB00927': 'Famotidine',
'DB00928': 'Azacitidine',
'DB00929': 'Misoprostol',
'DB00930': 'Colesevelam',
'DB00934': 'Maprotiline',
'DB00935': 'Oxymetazoline',
'DB00937': 'Diethylpropion',
'DB00945': 'Aspirin',
'DB00949': 'Felbamate',
'DB00950': 'Fexofenadine',
'DB00951': 'Isoniazid',
'DB00952': 'Naratriptan',
'DB00953': 'Rizatriptan',
'DB00956': 'Hydrocodone',
'DB00958': 'Carboplatin',
'DB00959': 'Methylprednisolone',
'DB00960': 'Pindolol',
'DB00961': 'Mepivacaine',
```

```
'DB00962': 'Zaleplon',
'DB00963': 'Bromfenac',
'DB00966': 'Telmisartan',
'DB00967': 'Desloratadine',
'DB00968': 'Methyldopa',
'DB00969': 'Alosetron',
'DB00972': 'Azelastine',
'DB00973': 'Ezetimibe',
'DB00979': 'Cyclopentolate',
'DB00980': 'Ramelteon',
'DB00982': 'Isotretinoin',
'DB00983': 'Formoterol',
'DB00985': 'Dimenhydrinate',
'DB00989': 'Rivastigmine',
'DB00990': 'Exemestane',
'DB00991': 'Oxaprozin',
'DB00993': 'Azathioprine',
'DB00996': 'Gabapentin',
'DB00997': 'Doxorubicin',
'DB00998': 'Frovatriptan',
'DB00999': 'Hydrochlorothiazide',
'DB01001': 'Albuterol',
'DB01003': 'Cromolyn',
'DB01004': 'Ganciclovir',
'DB01005': 'Hydroxyurea',
'DB01006': 'Letrozole',
'DB01007': 'Tioconazole',
'DB01008': 'Busulfan',
'DB01009': 'Ketoprofen',
'DB01012': 'Cinacalcet',
'DB01014': 'Balsalazide',
'DB01016': 'Glyburide',
'DB01017': 'Minocycline',
'DB01018': 'Guanfacine',
'DB01019': 'Bethanechol',
'DB01020': 'Isosorbide mononitrate',
'DB01022': 'Phytonadione',
'DB01023': 'Felodipine',
'DB01024': 'Mycophenolic acid',
'DB01026': 'Ketoconazole',
'DB01029': 'Irbesartan',
'DB01032': 'Probenecid',
'DB01033': 'Mercaptopurine',
'DB01036': 'Tolterodine',
'DB01037': 'Selegiline',
'DB01039': 'Fenofibrate',
'DB01043': 'Memantine',
```

```
'DB01045': 'Rifampin',
'DB01046': 'Lubiprostone',
'DB01047': 'Fluocinonide',
'DB01050': 'Ibuprofen',
'DB01058': 'Praziquantel',
'DB01059': 'Norfloxacin',
'DB01060': 'Amoxicillin',
'DB01062': 'Oxybutynin',
'DB01065': 'Melatonin',
'DB01066': 'Cefditoren',
'DB01067': 'Glipizide',
'DB01068': 'Clonazepam',
'DB01069': 'Promethazine',
'DB01072': 'Atazanavir',
'DB01073': 'Fludarabine',
'DB01075': 'Diphenhydramine',
'DB01076': 'Atorvastatin',
'DB01083': 'Orlistat',
'DB01085': 'Pilocarpine',
'DB01086': 'Benzocaine',
'DB01093': 'Dimethyl sulfoxide',
'DB01097': 'Leflunomide',
'DB01098': 'Rosuvastatin',
'DB01100': 'Pimozide',
'DB01101': 'Capecitabine',
'DB01104': 'Sertraline',
'DB01110': 'Miconazole',
'DB01112': 'Cefuroxime',
'DB01114': 'Chlorpheniramine',
'DB01115': 'Nifedipine',
'DB01117': 'Atovaquone',
'DB01118': 'Amiodarone',
'DB01119': 'Diazoxide',
'DB01126': 'Dutasteride',
'DB01127': 'Econazole',
'DB01128': 'Bicalutamide',
'DB01129': 'Rabeprazole',
'DB01130': 'Prednicarbate',
'DB01132': 'Pioglitazone',
'DB01136': 'Carvedilol',
'DB01137': 'Levofloxacin',
'DB01138': 'Sulfinpyrazone',
'DB01142': 'Doxepin',
'DB01144': 'Dichlorphenamide',
'DB01148': 'Flavoxate',
'DB01149': 'Nefazodone',
'DB01150': 'Cefprozil',
```

```
'DB01151': 'Desipramine',
'DB01155': 'Gemifloxacin',
'DB01156': 'Bupropion',
'DB01162': 'Terazosin',
'DB01165': 'Ofloxacin',
'DB01166': 'Cilostazol',
'DB01167': 'Itraconazole',
'DB01169': 'Arsenic trioxide',
'DB01173': 'Orphenadrine',
'DB01174': 'Phenobarbital',
'DB01175': 'Escitalopram',
'DB01176': 'Cyclizine',
'DB01181': 'Ifosfamide',
'DB01182': 'Propafenone',
'DB01190': 'Clindamycin',
'DB01192': 'Oxymorphone',
'DB01193': 'Acebutolol',
'DB01194': 'Brinzolamide',
'DB01195': 'Flecainide',
'DB01197': 'Captopril',
'DB01200': 'Bromocriptine',
'DB01202': 'Levetiracetam',
'DB01203': 'Nadolol',
'DB01204': 'Mitoxantrone',
'DB01211': 'Clarithromycin',
'DB01212': 'Rocephin',
'DB01215': 'Estazolam',
'DB01217': 'Anastrozole',
'DB01219': 'Dantrolene',
'DB01220': 'Rifaximin',
'DB01221': 'Ketamine',
'DB01222': 'Budesonide',
'DB01224': 'Quetiapine',
'DB01229': 'Paclitaxel',
'DB01233': 'Metoclopramide',
'DB01234': 'Dexamethasone',
'DB01235': 'Levodopa',
'DB01238': 'Aripiprazole',
'DB01240': 'Epoprostenol',
'DB01241': 'Gemfibrozil',
'DB01242': 'Clomipramine',
'DB01247': 'Isocarboxazid',
'DB01248': 'Docetaxel',
'DB01254': 'Dasatinib',
'DB01255': 'Lisdexamfetamine',
'DB01256': 'Retapamulin',
'DB01261': 'Sitagliptin',
```

```
'DB01264': 'Darunavir',
'DB01267': 'Paliperidone',
'DB01268': 'Sunitinib',
'DB01269': 'Panitumumab',
'DB01273': 'Varenicline',
'DB01274': 'Arformoterol',
'DB01275': 'Hydralazine',
'DB01276': 'Exenatide',
'DB01278': 'Pramlintide',
'DB01281': 'Abatacept',
'DB01285': 'Corticotropin',
'DB01291': 'Pirbuterol',
'DB01294': 'Bismuth subsalicylate',
'DB01296': 'Glucosamine',
'DB01306': 'Insulin aspart',
'DB01307': 'Insulin detemir',
'DB01309': 'Insulin glulisine',
'DB01327': 'Cefazolin',
'DB01362': 'Iohexol',
'DB01364': 'Ephedrine',
'DB01367': 'Rasagiline',
'DB01377': 'Magnesium oxide',
'DB01381': 'Ginkgo',
'DB01390': 'Sodium bicarbonate',
'DB01392': 'Yohimbine',
'DB01394': 'Colchicine',
'DB01397': 'Magnesium salicylate',
'DB01399': 'Salsalate',
'DB01400': 'Neostigmine',
'DB01406': 'Danazol',
'DB01415': 'Ceftibuten',
'DB01416': 'Cefpodoxime',
'DB01432': 'Cholestyramine',
'DB01438': 'Phenazopyridine',
'DB01563': 'Chloral hydrate',
'DB01576': 'Dextroamphetamine',
'DB01577': 'Methamphetamine',
'DB01579': 'Phendimetrazine',
'DB01586': 'Ursodiol',
'DB01589': 'Quazepam',
'DB01590': 'Everolimus',
'DB01591': 'Solifenacin',
'DB01610': 'Valganciclovir',
'DB01611': 'Hydroxychloroquine',
'DB01618': 'Molindone',
'DB01623': 'Thiothixene',
'DB01627': 'Lincomycin',
```

```
'DB01656': 'Roflumilast',
'DB01708': 'Dehydroepiandrosterone',
'DB02300': 'Calcipotriene',
'DB02701': 'Niacinamide',
'DB04835': 'Maraviroc',
'DB04854': 'Febuxostat',
'DB04855': 'Dronedarone',
'DB04861': 'Nebivolol',
'DB04868': 'Nilotinib',
'DB04871': 'Lorcaserin',
'DB04896': 'Milnacipran',
'DB04908': 'Flibanserin',
'DB04938': 'Ospemifene',
'DB04946': 'Iloperidone',
'DB05245': 'Silver sulfadiazine',
'DB05259': 'Glatiramer',
'DB05271': 'Rotigotine',
'DB05351': 'Dexlansoprazole',
'DB05521': 'Telaprevir',
'DB05676': 'Apremilast',
'DB05679': 'Ustekinumab',
'DB05812': 'Abiraterone',
'DB05829': 'Parathyroid hormone',
'DB06176': 'Romidepsin',
'DB06196': 'Icatibant',
'DB06201': 'Rufinamide',
'DB06203': 'Alogliptin',
'DB06204': 'Tapentadol',
'DB06207': 'Silodosin',
'DB06209': 'Prasugrel',
'DB06210': 'Eltrombopag',
'DB06213': 'Regadenoson',
'DB06216': 'Asenapine',
'DB06218': 'Lacosamide',
'DB06228': 'Rivaroxaban',
'DB06237': 'Avanafil',
'DB06273': 'Tocilizumab',
'DB06282': 'Levocetirizine',
'DB06283': 'Ziconotide',
'DB06287': 'Temsirolimus',
'DB06290': 'Simeprevir',
'DB06292': 'Dapagliflozin',
'DB06335': 'Saxagliptin',
'DB06372': 'Rilonacept',
'DB06403': 'Ambrisentan',
'DB06412': 'Oxymetholone',
'DB06413': 'Armodafinil',
```

```
'DB06414': 'Etravirine',
'DB06589': 'Pazopanib',
'DB06605': 'Apixaban',
'DB06637': 'Dalfampridine',
'DB06643': 'Denosumab',
'DB06655': 'Liraglutide',
'DB06674': 'Golimumab',
'DB06684': 'Vilazodone',
'DB06688': 'Sipuleucel-T',
'DB06699': 'Degarelix',
'DB06700': 'Desvenlafaxine',
'DB06701': 'Dexmethylphenidate',
'DB06702': 'Fesoterodine',
'DB06709': 'Methacholine',
'DB06710': 'Methyltestosterone',
'DB06711': 'Naphazoline',
'DB06724': 'Calcium carbonate',
'DB06764': 'Tetrahydrozoline',
'DB06769': 'Bendamustine',
'DB06781': 'Difluprednate',
'DB06786': 'Halcinonide',
'DB06788': 'Histrelin',
'DB06792': 'Lanthanum carbonate',
'DB06799': 'Methenamine',
'DB06800': 'Methylnaltrexone',
'DB06802': 'Nepafenac',
'DB06813': 'Pralatrexate',
'DB06817': 'Raltegravir',
'DB06822': 'Tinzaparin',
'DB06825': 'Triptorelin',
'DB08815': 'Lurasidone',
'DB08816': 'Ticagrelor',
'DB08819': 'Tafluprost',
'DB08822': 'Azilsartan medoxomil',
'DB08860': 'Pitavastatin',
'DB08865': 'Crizotinib',
'DB08867': 'Ulipristal',
'DB08868': 'Fingolimod',
'DB08869': 'Tesamorelin',
'DB08871': 'Eribulin',
'DB08872': 'Gabapentin enacarbil',
'DB08873': 'Boceprevir',
'DB08874': 'Fidaxomicin',
'DB08875': 'Cabozantinib',
'DB08877': 'Ruxolitinib',
'DB08879': 'Belimumab',
'DB08880': 'Teriflunomide',
```

```
'DB08881': 'Vemurafenib',
'DB08882': 'Linagliptin',
'DB08883': 'Perampanel',
'DB08889': 'Carfilzomib',
'DB08893': 'Mirabegron',
'DB08895': 'Tofacitinib',
'DB08896': 'Regorafenib',
'DB08897': 'Aclidinium',
'DB08899': 'Enzalutamide',
'DB08901': 'Ponatinib',
'DB08907': 'Canagliflozin',
'DB08910': 'Pomalidomide',
'DB08911': 'Trametinib',
'DB08912': 'Dabrafenib',
'DB08916': 'Afatinib',
'DB08918': 'Levomilnacipran',
'DB08930': 'Dolutegravir',
'DB08931': 'Riociguat',
'DB08932': 'Macitentan',
'DB08934': 'Sofosbuvir',
'DB08941': 'Isoxsuprine',
'DB09020': 'Bisacodyl',
'DB09026': 'Aliskiren',
'DB09029': 'Secukinumab',
'DB09033': 'Vedolizumab',
'DB09034': 'Suvorexant',
'DB09035': 'Nivolumab',
'DB09038': 'Empagliflozin',
'DB09043': 'Albiglutide',
'DB09045': 'Dulaglutide',
'DB09049': 'Naloxegol',
'DB09053': 'Ibrutinib',
'DB09068': 'Vortioxetine',
'DB09072': 'Sodium oxybate',
'DB09073': 'Palbociclib',
'DB09076': 'Umeclidinium',
'DB09080': 'Olodaterol',
'DB09083': 'Ivabradine',
'DB09085': 'Tetracaine',
'DB09096': 'Benzoyl peroxide',
'DB09102': 'Daclatasvir',
'DB09104': 'Magnesium hydroxide',
'DB09122': 'Peginterferon beta-1a',
'DB09128': 'Brexpiprazole',
'DB09130': 'Copper',
'DB09146': 'Sucroferric oxyhydroxide',
'DB09153': 'Sodium chloride',
```

```
 'DB09208': 'Pegloticase',
 'DB09270': 'Ubiquinone',
 'DB09272': 'Eluxadoline',
 'DB09287': 'Polyethylene glycol 3350',
 'DB09291': 'Rolapitant',
 'DB09302': 'Alirocumab',
 'DB09303': 'Evolocumab',
 'DB09322': 'Zinc sulfate',
 'DB09564': 'Insulin degludec',
 'DB11110': 'Magnesium citrate',
 'DB11133': 'Omega-3 polyunsaturated fatty acids',
 'DB11156': 'Pyrantel',
 'DB11248': 'Zinc gluconate',
 'DB11256': 'L-methylfolate',
 'DB11315': 'Methscopolamine'}
```

[936]:
```
df_drug_neg_names = df_drug_neg_final.copy(deep=True)

df_drug_neg_names[0] = df_drug_neg_names[0].map(df_drug_map)
df_drug_neg_names[1] = df_drug_neg_names[1].map(df_drug_map)
df_drug_neg_names.dropna(inplace=True)
df_drug_neg_names
```

[936]:

	0	1
886	Goserelin	Alogliptin
888	Goserelin	Ofloxacin
891	Goserelin	Quetiapine
892	Goserelin	Crizotinib
893	Goserelin	Amoxapine
…	…	…
96896	Umeclidinium	Methscopolamine
96964	Brexpiprazole	Rolapitant
96965	Brexpiprazole	Rolapitant
96996	Eluxadoline	Methscopolamine
96997	Eluxadoline	Methscopolamine

[38980 rows x 2 columns]

[919]:
```
df_uci_train_final = pd.merge(df_uci_train_updated, df_drug_neg_final,␣
 ↪left_on='DBID', right_on=0).drop(columns=0)
df_uci_train_final['DBID'] = df_uci_train_final['DBID'].map(df_drug_map)
df_uci_train_final[1] = df_uci_train_final[1].map(df_drug_map)
df_uci_train_final
```

[919]:

	uniqueID	drugName	condition	\
0	206461	Valsartan	Left Ventricular Dysfunction	
1	206461	Valsartan	Left Ventricular Dysfunction	

```
2          206461  Valsartan  Left Ventricular Dysfunction
3          206461  Valsartan  Left Ventricular Dysfunction
4          206461  Valsartan  Left Ventricular Dysfunction
…             …         …                           …
9747449    173878   Oxytocin               Labor Induction
9747450    173878   Oxytocin               Labor Induction
9747451    173878   Oxytocin               Labor Induction
9747452    173878   Oxytocin               Labor Induction
9747453    173878   Oxytocin               Labor Induction

                                                    review  rating       date  \
0        "It has no side effect, I take it in combinati…       9  20-May-12
1        "It has no side effect, I take it in combinati…       9  20-May-12
2        "It has no side effect, I take it in combinati…       9  20-May-12
3        "It has no side effect, I take it in combinati…       9  20-May-12
4        "It has no side effect, I take it in combinati…       9  20-May-12
…                                                      …     …         …
9747449  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747450  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747451  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747452  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747453  "They used Pitocin to induce all of my three l…      10  15-Aug-14

         usefulCount       DBID                    1
0                 27  Valsartan                  NaN
1                 27  Valsartan                  NaN
2                 27  Valsartan  Azilsartan medoxomil
3                 27  Valsartan            Piroxicam
4                 27  Valsartan            Labetalol
…                 …         …                    …
9747449           16   Oxytocin           Misoprostol
9747450           16   Oxytocin          Dinoprostone
9747451           16   Oxytocin          Mifepristone
9747452           16   Oxytocin                   NaN
9747453           16   Oxytocin            Citalopram

[9747454 rows x 9 columns]
```

[920]:
```
df_uci_train_final.dropna(inplace=True)
df_uci_train_final
```

[920]:
```
         uniqueID   drugName                     condition  \
2          206461  Valsartan  Left Ventricular Dysfunction
3          206461  Valsartan  Left Ventricular Dysfunction
4          206461  Valsartan  Left Ventricular Dysfunction
5          206461  Valsartan  Left Ventricular Dysfunction
6          206461  Valsartan  Left Ventricular Dysfunction
```

```
…           …        …                              …
9747448    173878   Oxytocin                Labor Induction
9747449    173878   Oxytocin                Labor Induction
9747450    173878   Oxytocin                Labor Induction
9747451    173878   Oxytocin                Labor Induction
9747453    173878   Oxytocin                Labor Induction

                                                    review  rating       date  \
2        "It has no side effect, I take it in combinati…       9  20-May-12
3        "It has no side effect, I take it in combinati…       9  20-May-12
4        "It has no side effect, I take it in combinati…       9  20-May-12
5        "It has no side effect, I take it in combinati…       9  20-May-12
6        "It has no side effect, I take it in combinati…       9  20-May-12
…                                                      …      …         …
9747448  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747449  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747450  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747451  "They used Pitocin to induce all of my three l…      10  15-Aug-14
9747453  "They used Pitocin to induce all of my three l…      10  15-Aug-14

         usefulCount       DBID                     1
2                 27  Valsartan  Azilsartan medoxomil
3                 27  Valsartan             Piroxicam
4                 27  Valsartan             Labetalol
5                 27  Valsartan              Pindolol
6                 27  Valsartan            Metolazone
…               …         …                      …
9747448           16   Oxytocin          Dinoprostone
9747449           16   Oxytocin           Misoprostol
9747450           16   Oxytocin          Dinoprostone
9747451           16   Oxytocin          Mifepristone
9747453           16   Oxytocin            Citalopram

[5512052 rows x 9 columns]
```

[921]:
```
df_uci_train_final.columns
```

[921]:
```
Index([   'uniqueID',    'drugName',   'condition',      'review',
         'rating',        'date', 'usefulCount',        'DBID',
                1],
      dtype='object')
```

[923]:
```
# Aggregating related DrugIDs into a list
aggregated_df = df_uci_train_final.
 ↪groupby(['uniqueID','drugName','condition','review','rating','date',
                                          'usefulCount']).agg({
    1: lambda x: -1*len(list(x))
```

```
})

# Output the final DataFrame
aggregated_df = aggregated_df.reset_index()
aggregated_df.rename(columns={1:'negDrugNo'}, inplace=True)
aggregated_df
```

[923]:

```
       uniqueID       drugName                     condition  \
0           766  Phenylephrine                   Eye Redness
1           767      Silodosin  Benign Prostatic Hyperplasia
2           768      Silodosin  Benign Prostatic Hyperplasia
3           773      Silodosin  Benign Prostatic Hyperplasia
4           777      Silodosin  Benign Prostatic Hyperplasia
…           …            …                            …
64757    232220     Paroxetine            Anxiety and Stress
64758    232221     Paroxetine                   Hot Flashes
64759    232223     Paroxetine   Generalized Anxiety Disorde
64760    232224     Paroxetine                       Anxiety
64761    232225     Paroxetine                       Anxiety

                                                  review  rating       date  \
0      "Best eye drop I ever tried. Provided soothing…      10  17-Apr-16
1      "Really not sure why my urologist gave me samp…       1   2-Apr-16
2      "It reduces urination frequency but has a very…       2   2-Mar-16
3      "Having trouble emptying bladder so doc gave R…       4  19-Sep-15
4      "I&#039;m 65 and have had weak flow for years…       5  13-May-15
…                                                    …     …         …
64757  "I took the lowest dose 10mg. The first two da…       1  22-Nov-15
64758  "I had been taking Sertraline for anxiety. Sta…       8  20-Nov-15
64759  "I&#039;ve been on Paxil for about two months…       5  18-Nov-15
64760  "I took this about 3 years ago for a massive m…       1  18-Nov-15
64761  "My physician prescribed me 20mg of paxil dail…       3  16-Nov-15

       usefulCount  negDrugNo
0                0        -68
1               28        -22
2               18        -22
3               21        -22
4               44        -22
…              …          …
64757           27       -244
64758           11       -244
64759           11       -244
64760           14       -244
64761           17       -244

[64762 rows x 8 columns]
```

[924]:
```
aggregated_df[aggregated_df.drugName=='Abiraterone']
```

[924]:

	uniqueID	drugName	condition	review	rating	date	usefulCount	negDrugNo
50785	188800	Abiraterone	Prostate Cance	"effectiveness lasted 8 months side effects un…	4	15-Nov-17	0	-24
50786	188802	Abiraterone	Prostate Cance	"Very satisfied with this drug. PSA dropped fr…	9	23-Dec-16	7	-24
50787	188805	Abiraterone	Prostate Cance	"Began Zytiga with Prednisone 4/23/14 when con…	10	31-May-15	17	-24
50788	188808	Abiraterone	Prostate Cance	"Have taken Zytiga for a year now, one pill ea…	10	7-Apr-14	36	-24
50789	188809	Abiraterone	Prostate Cance	"PSA went from 19 to 9 in 3 weeks. No signifi…	10	27-Oct-13	33	-24
50790	188810	Abiraterone	Prostate Cance	"On Zytiga for 6 months no side effects so far…	10	9-Jun-13	57	-24
50791	188811	Abiraterone	Prostate Cance	"One month later- PSA is 0.16! Sure, it'…	10	26-Apr-13	47	-24
50792	188812	Abiraterone	Prostate Cance	"PSA dropped from 7.46 to 0.34 in 30 days. L…	10	5-Apr-13	53	-24
50793	188813	Abiraterone	Prostate Cance	"PSA 28 to 4 in one week. Fatigue only side ef…	9	9-Aug-11	110	-24

[ ]:
```

```

[925]:
```
aggregated_df['usefulCount'] = aggregated_df['usefulCount'].replace(0,1)
aggregated_df
```

[925]:

	uniqueID	drugName	condition
0	766	Phenylephrine	Eye Redness
1	767	Silodosin	Benign Prostatic Hyperplasia
2	768	Silodosin	Benign Prostatic Hyperplasia
3	773	Silodosin	Benign Prostatic Hyperplasia
4	777	Silodosin	Benign Prostatic Hyperplasia
…	…	…	…

```
64757    232220     Paroxetine            Anxiety and Stress
64758    232221     Paroxetine                  Hot Flashes
64759    232223     Paroxetine   Generalized Anxiety Disorde
64760    232224     Paroxetine                       Anxiety
64761    232225     Paroxetine                       Anxiety

                                                    review  rating       date  \
0      "Best eye drop I ever tried. Provided soothing…      10  17-Apr-16
1      "Really not sure why my urologist gave me samp…       1   2-Apr-16
2      "It reduces urination frequency but has a very…       2   2-Mar-16
3      "Having trouble emptying bladder so doc gave R…       4  19-Sep-15
4      "I&#039;m 65 and have had weak flow for years…       5  13-May-15
…                                                    …     …         …
64757  "I took the lowest dose 10mg. The first two da…       1  22-Nov-15
64758  "I had been taking Sertraline for anxiety. Sta…       8  20-Nov-15
64759  "I&#039;ve been on Paxil for about two months…       5  18-Nov-15
64760  "I took this about 3 years ago for a massive m…       1  18-Nov-15
64761  "My physician prescribed me 20mg of paxil dail…       3  16-Nov-15

       usefulCount  negDrugNo
0                1        -68
1               28        -22
2               18        -22
3               21        -22
4               44        -22
…              …         …
64757           27       -244
64758           11       -244
64759           11       -244
64760           14       -244
64761           17       -244

[64762 rows x 8 columns]
```

[1165]:
```
aggregated_df[aggregated_df_update.drugName=='Erythromycin']
```

[1165]:
```
       uniqueID            drugName          condition  \
30954    132451  AbobotulinumtoxinA  Cervical Dystonia
30955    132452  AbobotulinumtoxinA     acial Wrinkles
30956    132454             Doxepin            Anxiety
30957    132455             Doxepin         Depression
30958    132458             Doxepin            Anxiety
30959    132460             Doxepin            Anxiety
30960    132461             Doxepin            Anxiety
30961    132462             Doxepin           Insomnia
30962    132465             Doxepin           Insomnia
30963    132466             Doxepin           Insomnia
```

30964	132467	Doxepin	Insomnia
30965	132468	Doxepin	Insomnia
30966	132473	Doxepin	Depression
30967	132474	Doxepin	Insomnia
30968	132476	Doxepin	Insomnia
30969	132477	Doxepin	Insomnia
30970	132481	Doxepin	Insomnia
30971	132483	Doxepin	Anxiety
30972	132484	Doxepin	Insomnia

	review	rating	date	usefulCount	\
30954	"I receive Dysport inj&#03…	8	14-Aug-14	9	
30955	"Experienced dizziness and…	5	13-Feb-14	11	
30956	"Started on Doxepin 100mg …	7	2-Dec-17	1	
30957	"I took this as prescribed…	10	1-Dec-17	1	
30958	"Had to stop it because of…	3	21-Oct-17	2	
30959	"Honestly, a very good dru…	9	28-Sep-17	2	
30960	"Hi milemile I have been o…	9	24-Sep-17	4	
30961	"Use to take Ambien but si…	10	17-Sep-17	23	
30962	"changed my life - I now s…	10	1-Aug-17	23	
30963	"Doxepin does not help me …	4	28-Jul-17	15	
30964	"Caused restless leg syndr…	1	28-Jul-17	11	
30965	"I have had insomnia since…	10	19-Jul-17	18	
30966	"I was diagnosed with majo…	10	23-May-17	21	
30967	"For the PTSD, Insomnia, a…	3	7-May-17	9	
30968	"Started off with 25mgs of…	4	28-Mar-17	15	
30969	"I take oxazepam 30mg as w…	9	24-Mar-17	25	
30970	"I had suffered an unspeak…	10	6-Mar-17	42	
30971	"I started doxepin last ni…	10	25-Jan-17	26	
30972	"I have been on Silenor si…	8	13-Jan-17	40	

	negDrugNo
30954	-144
30955	-144
30956	-50
30957	-50
30958	-50
30959	-50
30960	-50
30961	-50
30962	-50
30963	-50
30964	-50
30965	-50
30966	-50
30967	-50
30968	-50

```
30969        -50
30970        -50
30971        -50
30972        -50
```

[927]:
```
df_drug_eff = pd.read_csv("Project Data/Drug_clean.csv")
df_drug_eff
```

[927]:
```
                    Condition         Drug  EaseOfUse  Effective  \
0    Acute Bacterial Sinusitis  Amoxicillin   3.852353   3.655882
1    Acute Bacterial Sinusitis  Amoxicillin   3.470000   3.290000
2    Acute Bacterial Sinusitis  Amoxicillin   3.121429   2.962857
3    Acute Bacterial Sinusitis   Ampicillin   2.000000   3.000000
4    Acute Bacterial Sinusitis   Ampicillin   3.250000   3.000000
..                         …          …         …         …
680   vulvovaginal candidiasis   Miconazole   3.465000   2.770000
681   vulvovaginal candidiasis   Miconazole   4.750000   3.000000
682   vulvovaginal candidiasis   Miconazole   4.000000   1.000000
683   vulvovaginal candidiasis  Terconazole   3.525000   3.047500
684   vulvovaginal candidiasis  Tioconazole   3.852500   2.022500

                Form Indication       Price     Reviews  Satisfaction    Type
0            Capsule   On Label   12.590000   86.294118      3.197647      RX
1    Liquid (Drink)  Off Label  287.370000   43.000000      2.590000      RX
2             Tablet   On Label   70.608571  267.285714      2.248571      RX
3            Capsule   On Label   12.590000    1.000000      1.000000      RX
4             Tablet   On Label  125.240000   15.000000      3.000000      RX
..               …        …          …          …             …     …
680           Cream   On Label   13.990000   19.500000      2.345000  RX/OTC
681           Cream   On Label   13.990000    4.000000      3.000000     OTC
682           Other   On Label  125.990000    1.000000      1.000000     OTC
683           Cream   On Label   68.990000   20.000000      2.717500      RX
684           Other   On Label   22.990000  145.000000      1.827500     OTC

[685 rows x 10 columns]
```

[928]:
```
numerical_data = df_drug_eff.select_dtypes(include='number')
numerical_data['Drug'] = df_drug_eff['Drug']

drug_group = numerical_data.groupby(by='Drug').mean().round(2).reset_index()

drug_group
```

[928]:
```
                Drug  EaseOfUse  Effective   Price  Reviews  Satisfaction
0                ASA       3.00       2.60   17.99     6.00          2.80
1         Acebutolol       4.32       3.75   24.49    14.50          4.20
2      Acetaminophen       3.75       3.47   14.76     5.45          3.31
```

```
3      Acetazolamide       3.34       3.60   70.49   104.00          2.89
4    Acetohydroxamic       4.00       2.50  638.99     3.00          2.00
..               …          …          …      …       …             …
272        Valsartan       4.08       3.58  168.99   444.25          3.06
273        Verapamil       4.71       4.13  124.36    42.33          4.03
274              Vit       4.18       3.23  110.66     8.67          3.76
275        Zanamivir       2.35       2.12   81.99    10.00          2.00
276             Zinc       1.00       1.00   19.44     1.00          1.00

[277 rows x 6 columns]
```

[929]:
```
final_df = pd.merge(aggregated_df, drug_group, left_on='drugName',␣
 ↪right_on='Drug').drop(columns=['Drug'])
final_df['Price'] = -1*final_df['Price']
final_df.drop(columns='uniqueID', inplace=True)
final_df
```

[929]:
```
            drugName                  condition  \
0       Phenylephrine                Eye Redness
1       Phenylephrine  mulation) (phenylephrine)
2       Phenylephrine  mulation) (phenylephrine)
3       Phenylephrine  mulation) (phenylephrine)
4       Phenylephrine           Nasal Congestion
…                 …                          …
23720     Milnacipran                ibromyalgia
23721     Milnacipran                ibromyalgia
23722     Milnacipran                ibromyalgia
23723     Milnacipran                ibromyalgia
23724     Milnacipran                ibromyalgia

                                                  review  rating       date  \
0       "Best eye drop I ever tried. Provided soothing…      10  17-Apr-16
1       "Coming down with a cold and couldn&#039;t bre…      10  10-Oct-17
2       "I caught a bad sinus infection/ congestion. T…       9   5-May-17
3       "This product didn&#039;t make the slightest d…       1  10-Feb-17
4                                             "love it"      10   7-Feb-17
…                                                   …      …         …
23720   "I was taking Cymbalta and decided to go off d…       6  26-Jun-09
23721   "I&#039;m on day 4, my pain level in limbs see…       8  19-Jun-09
23722   "I am more than half way through the starter p…       4  16-Jun-09
23723   "At first I took Cymbalta and it really helped…       9  15-Jun-09
23724   "Having been diagnosed with fibrositis in 1987…       6  25-May-09

        usefulCount  negDrugNo  EaseOfUse  Effective   Price  Reviews  \
0                 1        -68       4.19       4.04  -65.05     2.22
1                 1        -68       4.19       4.04  -65.05     2.22
2                 2        -68       4.19       4.04  -65.05     2.22
```

```
3                3        -68       4.19       4.04  -65.05     2.22
4                1        -68       4.19       4.04  -65.05     2.22
…              …        …         …          …      …          …
23720           22         -8       4.29       3.57 -477.99     9.00
23721           37         -8       4.29       3.57 -477.99     9.00
23722           23         -8       4.29       3.57 -477.99     9.00
23723           72         -8       4.29       3.57 -477.99     9.00
23724            1         -8       4.29       3.57 -477.99     9.00

       Satisfaction
0              3.50
1              3.50
2              3.50
3              3.50
4              3.50
…              …
23720          3.57
23721          3.57
23722          3.57
23723          3.57
23724          3.57

[23725 rows x 12 columns]
```

[965]:
```
pd.set_option('display.max_colwidth', 30)
```

[1163]:
```
final_df[final_df.drugName=='Cedax']
```

[1163]:
```
Empty DataFrame
Columns: [drugName, condition, review, rating, date, usefulCount, negDrugNo,
EaseOfUse, Effective, Price, Reviews, Satisfaction]
Index: []
```

[981]:
```
final_df.condition.unique()
```

[981]:
```
array(['Eye Redness', 'mulation) (phenylephrine)', 'Nasal Congestion',
       'High Blood Pressure', 'Edema', 'Gout, Acute',
       'Bacterial Vaginitis', 'Rosacea', 'Perioral Dermatitis',
       'Bacterial Infection', 'Trichomoniasis', 'Dental Abscess',
       'Intraabdominal Infection', 'Diverticulitis',
       'Clostridial Infection', 'Giardiasis',
       'Pelvic Inflammatory Disease', 'Skin or Soft Tissue Infection',
       'Amebiasis', 'Helicobacter Pylori Infection', 'STD Prophylaxis',
       'Pseudomembranous Colitis', "Crohn's Disease, Acute",
       "Crohn's Disease, Maintenance", 'Surgical Prophylaxis',
       'Not Listed / Othe', 'Bone infection', 'Deep Neck Infection',
       'Supraventricular Tachycardia', 'Angina Pectoris Prophylaxis',
```

'Atrial Flutte', 'Atrial Fibrillation', "Raynaud's Syndrome",
'Bladder Infection', 'Sinusitis', 'Bronchitis',
'Urinary Tract Infection', 'Otitis Media', 'Pneumonia',
'Prevention of Bladder infection', 'Brain Tum',
'Conjunctivitis, Bacterial', 'Kidney Infections', 'Prostatitis',
'Infectious Diarrhea', "Crohn's Disease",
'Epididymitis, Sexually Transmitted', 'Typhoid Feve',
'Gonococcal Infection, Uncomplicated', 'Cholera', 'Anthrax',
"Traveler's Diarrhea", 'Lyme Disease',
'Upper Respiratory Tract Infection', 'Strep Throat',
'Skin and Structure Infection', 'Tonsillitis/Pharyngitis', 'Acne',
'Hirsutism', 'Alopecia', 'Heart Failure',
'Primary Hyperaldosteronism Diagnosis',
'Primary Hyperaldosteronism', 'Gender Dysphoria', 'Hypokalemia',
'Prevention of Cardiovascular Disease', 'Restless Legs Syndrome',
'Periodic Limb Movement Disorde', "Parkinson's Disease",
'Asthma, acute', 'Inflammatory Conditions', 'Immunosuppression',
'Nephrotic Syndrome', 'Dermatitis',
'Postoperative Ocular Inflammation', 'SIADH', 'Gouty Arthritis',
'Sarcoidosis', 'Rheumatoid Arthritis', 'Asthma',
'Allergic Rhinitis', 'Allergic Reactions', 'Psoriasis',
'Pityriasis rubra pilaris', 'Osteoarthritis', 'Iritis',
'Systemic Lupus Erythematosus', 'Lichen Planus', 'Neuritis',
'Atopic Dermatitis', 'Eczema', 'Aphthous Ulce', 'Pancreatic Cance',
'Breast Cancer, Metastatic', 'Neuropathic Pain',
'Migraine Prevention', 'Generalized Anxiety Disorde', 'Neuralgia',
'Pain', 'ibromyalgia', 'Postherpetic Neuralgia',
'Diabetic Peripheral Neuropathy', 'Peripheral Neuropathy',
'Reflex Sympathetic Dystrophy Syndrome', "Dercum's Disease",
'Epilepsy', 'Herpes Simplex', 'Herpes Simplex, Suppression',
'Cold Sores', 'Mononucleosis',
'Herpes Simplex, Mucocutaneous/Immunocompromised Host',
'Herpes Zoste',
'Herpes Simplex, Mucocutaneous/Immunocompetent Host',
'Schizophrenia', 'Schizoaffective Disorde', 'Bipolar Disorde',
'Depression', 'Autism', 'High Cholesterol', 'Anesthesia',
'Manscaping Pain', 'Hemorrhoids', 'Sunburn', 'Burns, External',
'Costochondritis', 'Hyperlipoproteinemia', 'Hypertriglyceridemia',
'Hyperlipoproteinemia Type IIa, Elevated LDL', 'Panic Disorde',
'Major Depressive Disorde', 'Insomnia', 'Nausea/Vomiting',
'Motion Sickness', 'Urticaria', 'Cold Symptoms',
'Diabetic Kidney Disease', 'Coronary Artery Disease',
'Heart Attack', 'Chronic Pain', 'Anxiety', 'Back Pain',
'Vulvodynia', 'Opiate Withdrawal', 'Obsessive Compulsive Disorde',
'Periodontitis', 'Streptococcal Infection', 'Stomach Ulce', 'GERD',
"Addison's Disease", 'Adrenocortical Insufficiency',
'Ulcerative Proctitis', 'Skin Rash', 'Pruritus',

"Barrett's Esophagus", 'Erosive Esophagitis',
'NSAID-Induced Gastric Ulce', 'Duodenal Ulcer Prophylaxis',
'Perimenopausal Symptoms', 'ADHD', "Tourette's Syndrome",
'Benzodiazepine Withdrawal', 'Insomnia, Stimulant-Associated',
'Hyperhidrosis', 'Hypertensive Emergency', 'Alcohol Withdrawal',
'Postural Orthostatic Tachycardia Syndrome',
'Mitral Valve Prolapse', 'Ventricular Tachycardia', 'Angina',
'Corneal Ulce', 'Campylobacter Gastroenteritis',
'Left Ventricular Dysfunction', 'Cardiovascular Risk Reduction',
'Duodenal Ulce', 'Obstructive Sleep Apnea/Hypopnea Syndrome',
'Shift Work Sleep Disorde', 'Hypersomnia', 'Narcolepsy', 'atigue',
'Chronic Fatigue Syndrome', 'Multiple Sclerosis',
'Chlamydia Infection', 'Bacterial Endocarditis Prevention',
'Pharyngitis', 'Cystic Fibrosis', 'COPD, Acute',
'Mycobacterium avium-intracellulare, Treatment',
'Gastritis/Duodenitis', 'Peptic Ulce', 'Stress Ulcer Prophylaxis',
'High Cholesterol, Familial Heterozygous', 'Women (minoxidil)',
'Vertig', 'Indigestion', 'Gastric Ulcer Maintenance Treatment',
'Pathological Hypersecretory Conditions', 'Niacin Deficiency',
'Hyperlipoproteinemia Type IV, Elevated VLDL', 'Angioedema',
'Endometriosis', 'ibrocystic Breast Disease',
'Aspiration Pneumonia', 'Benign Prostatic Hyperplasia',
'Colorectal Cance', 'Nephrocalcinosis', 'mance Anxiety',
'Arrhythmia', 'Benign Essential Trem', 'Portal Hypertension', 'Pe',
'Thyrotoxicosis', 'Hemangioma', 'Varicella-Zoste',
'Ramsay Hunt Syndrome', 'CMV Prophylaxis',
'Multiple Endocrine Adenomas', 'Zollinger-Ellison Syndrome',
'Impetig', 'Vaginal Yeast Infection', 'Tinea Cruris',
'Tinea Versicol', 'Oral Thrush', 'Ophthalmic Surgery',
'Period Pain', 'Headache', 'Tendonitis', 'Sciatica', 'Muscle Pain',
'Ankylosing Spondylitis', 'Neck Pain',
'Juvenile Rheumatoid Arthritis', 'Chronic Myofascial Pain',
'Bursitis', 'zen Shoulde', 'Transient Ischemic Attack',
'Thromboembolic Stroke Prophylaxis', 'eve', 'Allergic Urticaria',
'Borderline Personality Disorde', 'Agitated State',
'Post Traumatic Stress Disorde', 'Dermatologic Lesion',
'Asthma, Maintenance', 'Meningitis', 'Cluster Headaches',
'Organ Transplant, Rejection Prophylaxis', 'Glaucoma, Open Angle',
'Hyperlipoproteinemia Type III, Elevated beta-VLDL   IDL',
'Esophageal Variceal Hemorrhage Prophylaxis',
'Irritable Bowel Syndrome', 'Cyclic Vomiting Syndrome',
'Smoking Cessation', 'Neurosis', 'Primary Nocturnal Enuresis',
'Migraine', 'Gastroparesis', 'Lactation Augmentation',
'Nausea/Vomiting, Chemotherapy Induced',
'Nausea/Vomiting, Postoperative', 'Dermatological Disorders',
'Lichen Sclerosus', 'Birth Control', 'Emergency Contraception',
'Abnormal Uterine Bleeding', 'Ulcerative Colitis', 'COPD',

```
       'Pemphigus', 'Autoimmune Hemolytic Anemia',
       'Mixed Connective Tissue Disease', 'Inflammatory Bowel Disease',
       'Epicondylitis, Tennis Elbow', 'Psoriatic Arthritis', 'Lymphoma',
       'Conjunctivitis, Allergic', 'Leukemia', 'Systemic Sclerosis',
       'Bullous Pemphigoid', 'Prostate Cance', 'Anxiety and Stress',
       'Nightmares', 'Breast Cance', 'Cance', 'Amenorrhea',
       'Hypercalcemia', 'Ehrlichiosis', 'Malaria Prevention',
       'Lyme Disease, Arthritis', 'Ocular Rosacea', 'Malaria',
       'Bronchiectasis', 'Q Feve', 'Gastroenteritis',
       'Lyme Disease, Neurologic',
       'Premature Ventricular Depolarizations', 'Agitation',
       'Paranoid Disorde', 'Tic Disorde', 'Asperger Syndrome',
       'Severe Mood Dysregulation', 'Head Injury', 'Mania',
       'Social Anxiety Disorde', 'Cough', 'Hot Flashes', 'Toothache',
       'Temporomandibular Joint Disorde', 'Spondylolisthesis',
       'Patent Ductus Arteriosus', 'Keratosis', 'Warts', 'Skin Cance',
       'Basal Cell Carcinoma', 'Actinic Keratosis', 'Night Terrors',
       'Interstitial Cystitis', 'Pseudotumor Cerebri', 'Glaucoma',
       'Mountain Sickness / Altitude Sickness', 'Seizure Prevention',
       'Hydrocephalus', 'Urinary Incontinence', 'Burning Mouth Syndrome',
       'Neurotic Depression', 'Pudendal Neuralgia', 'Dysautonomia',
       'Persistent Depressive Disorde', 'Small Fiber Neuropathy',
       'Human Papilloma Virus', 'Trichotillomania', 'Bleeding Disorde',
       'Multiple Myeloma', 'Croup', 'Cerebral Edema',
       'Dermatitis Herpeti', 'Macular Edema', 'Stomach Cance',
       'Post-Cholecystectomy Diarrhea', 'Dumping Syndrome',
       'Pruritus of Partial Biliary Obstruction', 'Muscle Spasm',
       't Pac with Cyclobenzaprine (cyclobenzaprine)',
       'High Cholesterol, Familial Homozygous',
       'Anorexia/Feeding Problems', 'Postmenopausal Symptoms',
       'Postpartum Depression', 'Premenstrual Dysphoric Disorde',
       'Premature Lab', 'Ovarian Cance', 'Atherosclerosis',
       'Mycoplasma Pneumonia', 'Pertussis', 'Legionella Pneumonia',
       'llicular Lymphoma', 'Onychomycosis, Toenail', 'ungal Pneumonia',
       'Candida Urinary Tract Infection', 'Systemic Candidiasis',
       'Tinea Corporis', 'Candidemia', 'Esophageal Candidiasis',
       'm Pain Disorde'], dtype=object)
```

[1171]:
```python
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import TruncatedSVD
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

def recommend_drugs_for_condition(condition, final_df, top_n=5):
    # Filter the dataset for the current condition
    condition_data = final_df[final_df['condition'] == condition]
```

```
    # Select relevant features
    features = condition_data[['rating', 'usefulCount', 'negDrugNo',␣
↪'EaseOfUse', 'Effective', 'Price', 'Reviews', 'Satisfaction']]

    # Group by drugName and aggregate the numerical features
    aggregated_data = condition_data.groupby('drugName').agg({
        'rating': 'mean',
        'usefulCount': 'mean',
        'negDrugNo': 'mean',
        'EaseOfUse': 'mean',
        'Effective': 'mean',
        'Price': 'mean',
        'Reviews': 'mean',
        'Satisfaction': 'mean'
    }).reset_index()

    # Scale the aggregated features
    scaler = StandardScaler()
    features_scaled = scaler.fit_transform(aggregated_data[['rating',␣
↪'usefulCount', 'negDrugNo', 'EaseOfUse', 'Effective', 'Price', 'Reviews',␣
↪'Satisfaction']])
    print(features_scaled)

    # Apply SVD to reduce dimensionality
    svd = TruncatedSVD(n_components=5) # Choose the number of components
    features_reduced = svd.fit_transform(features_scaled)
    print(features_reduced)

    #similarity
    similarity_matrix = np.dot(features_reduced, features_reduced.T)

    print(similarity_matrix)

    # Find the mean similarity score with all other drugs for each drug
    mean_similarity = np.mean(similarity_matrix, axis=1)

    print(mean_similarity)

    # Get indices of top N similar drugs
    top_indices = np.argsort(mean_similarity)[-top_n:][::-1]

    # Retrieve the names of the top recommended drugs
    recommended_drugs = aggregated_data.iloc[top_indices]

    df_temp = pd.merge(recommended_drugs, df_drug_neg_names,␣
↪left_on='drugName', right_on=0).drop(columns=0)
```

```
    df_merged = df_temp.groupby(['drugName', 'rating', 'usefulCount',␣
↪'negDrugNo', 'EaseOfUse', 'Effective', 'Price', 'Reviews', 'Satisfaction']).
↪agg({
        1: lambda x: ', '.join(x)
    }).reset_index()
    df_merged.rename(columns={1: 'negDrugNames'}, inplace=True)


    # Visualize explained variance ratio of each component
    plt.figure(figsize=(6, 4))
    plt.bar(range(len(svd.explained_variance_ratio_)), svd.
↪explained_variance_ratio_, color='skyblue', alpha=0.7)
    plt.xlabel('Component')
    plt.ylabel('Explained Variance Ratio')
    plt.title('Explained Variance Ratio of SVD Components')
    plt.xticks(range(len(svd.explained_variance_ratio_)))
    plt.grid(True)
    plt.show()

        # Cumulative explained variance
    plt.figure(figsize=(6, 4))
    plt.plot(np.cumsum(svd.explained_variance_ratio_), marker='o',␣
↪linestyle='--', color='b')
    plt.title('Cumulative Explained Variance')
    plt.xlabel('Number of Components')
    plt.ylabel('Cumulative Explained Variance')
    plt.show()

    # Visualize feature influence with radar chart
    features = features.columns.tolist()  # Extract feature names
    num_features = len(features)
    angles = np.linspace(0, 2*np.pi, num_features, endpoint=False).tolist()  #␣
↪Angles for radar chart

    # Create radar chart for each component
    colors = plt.cm.viridis(np.linspace(0, 1, len(svd.components_)))
    for i in range(len(svd.components_)):
        plt.figure(figsize=(6, 4))
        component_weights = svd.components_[i].tolist()
        plt.subplot(111, polar=True)
        plt.xticks(angles, features, fontsize=6)  # Adjust fontsize for␣
↪readability
        plt.yticks([-1, 0, 1], color='grey', alpha=0.7)  # Y-axis ticks
        color = colors[i]
        plt.plot(angles, component_weights, '-o', label=f'Component {i+1}',␣
↪linewidth=2, color = color)
```

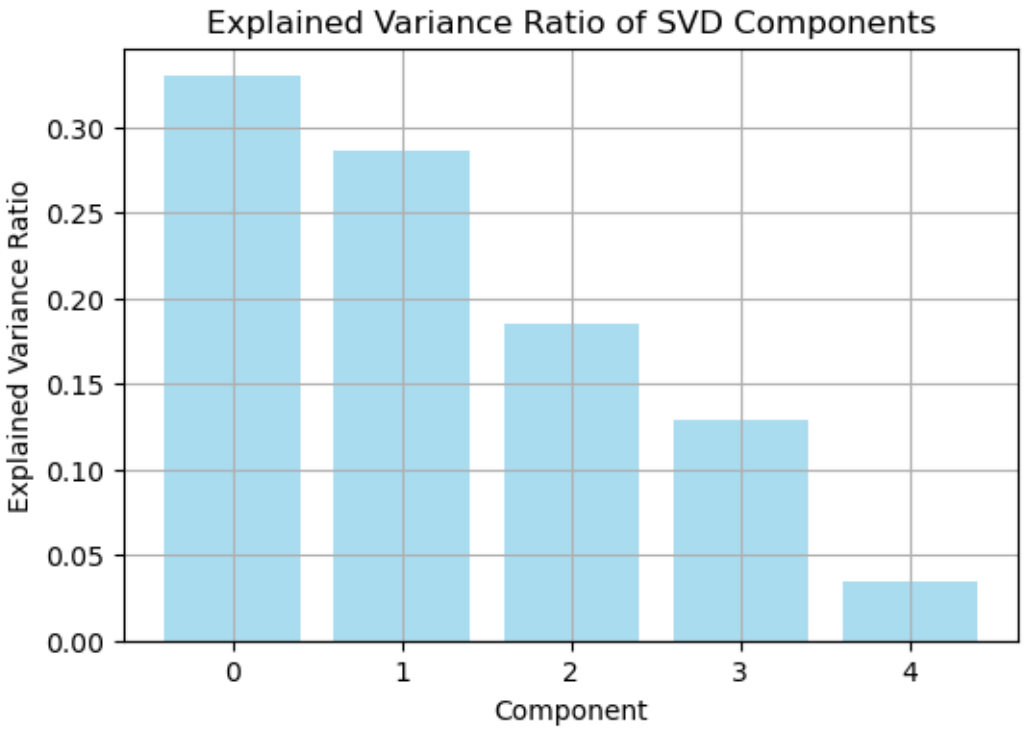
```
        # Add title and legend
        plt.title(f'Feature Influence on SVD Component {i+1} (Higher = More␣
↪Influence)')
        plt.legend()
        plt.grid(True)

        # Mark values
        for j, (angle, weight) in enumerate(zip(angles, component_weights)):
            if weight < 0:
                plt.text(angle, weight, round(weight,2), ha='right',␣
↪va='center')
            else:
                plt.text(angle, weight, round(weight,2), ha='left', va='center')
        plt.ylim(-1, 1)  # Fixing the y-axis limits
        plt.savefig("Project Data/SVD-Component"+condition+str(i)+".png")
        plt.show()

    # Merge with negative drug names
    df_temp = pd.merge(recommended_drugs, df_drug_neg_names,␣
↪left_on='drugName', right_on=0).drop(columns=0)

    df_merged = df_temp.groupby(['drugName', 'rating', 'usefulCount',␣
↪'negDrugNo', 'EaseOfUse', 'Effective', 'Price', 'Reviews', 'Satisfaction']).
↪agg({
        1: lambda x: ', '.join(x)
    }).reset_index()
    df_merged.rename(columns={1: 'negDrugNames'}, inplace=True)

    return df_merged
```

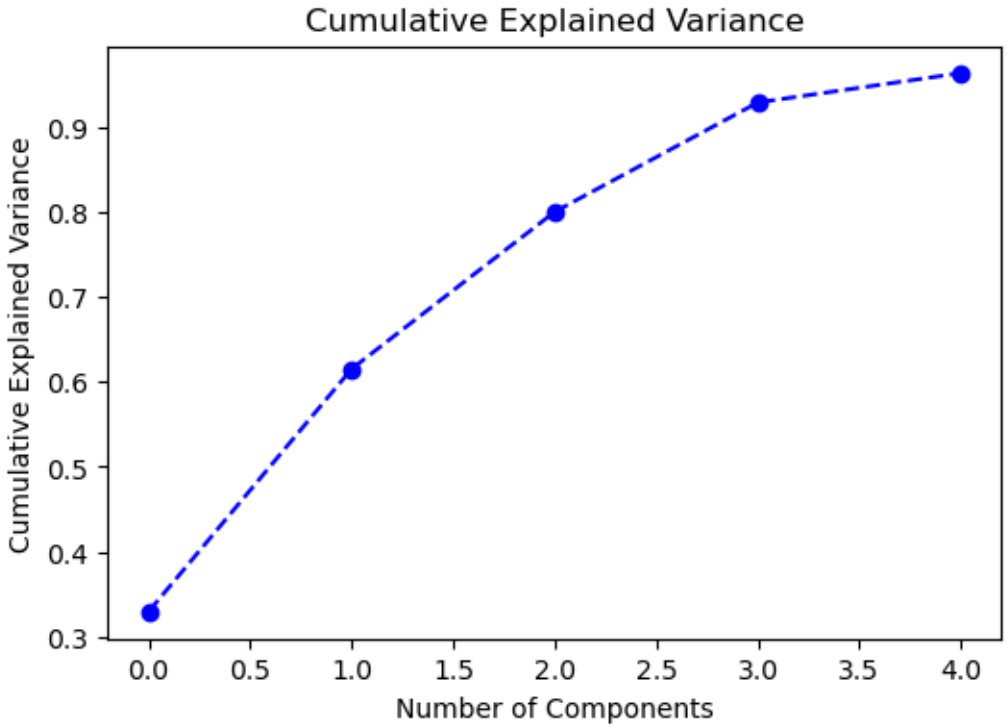
[1131]:
```
train_df, test_df = train_test_split(final_df, test_size=0.5, random_state=10)
```

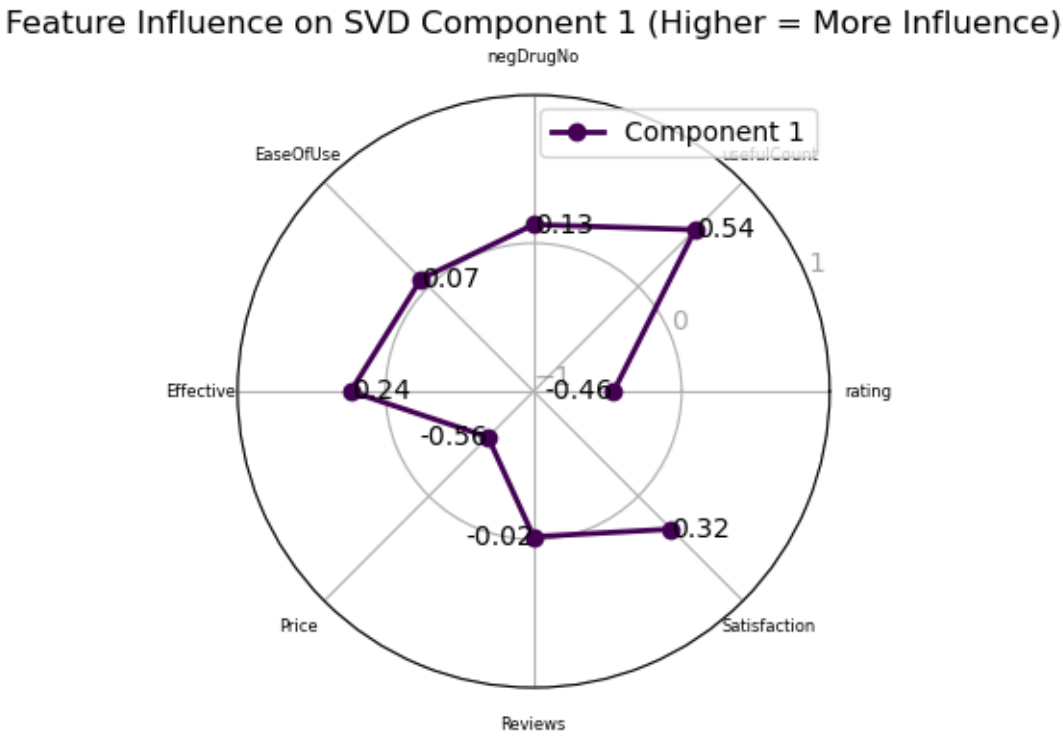
[1172]:
```
recommended_drugs = recommend_drugs_for_condition("Pneumonia", final_df,␣
↪top_n=3)
print("Recommended Drugs for Pneumonia:")
display(recommended_drugs)
```

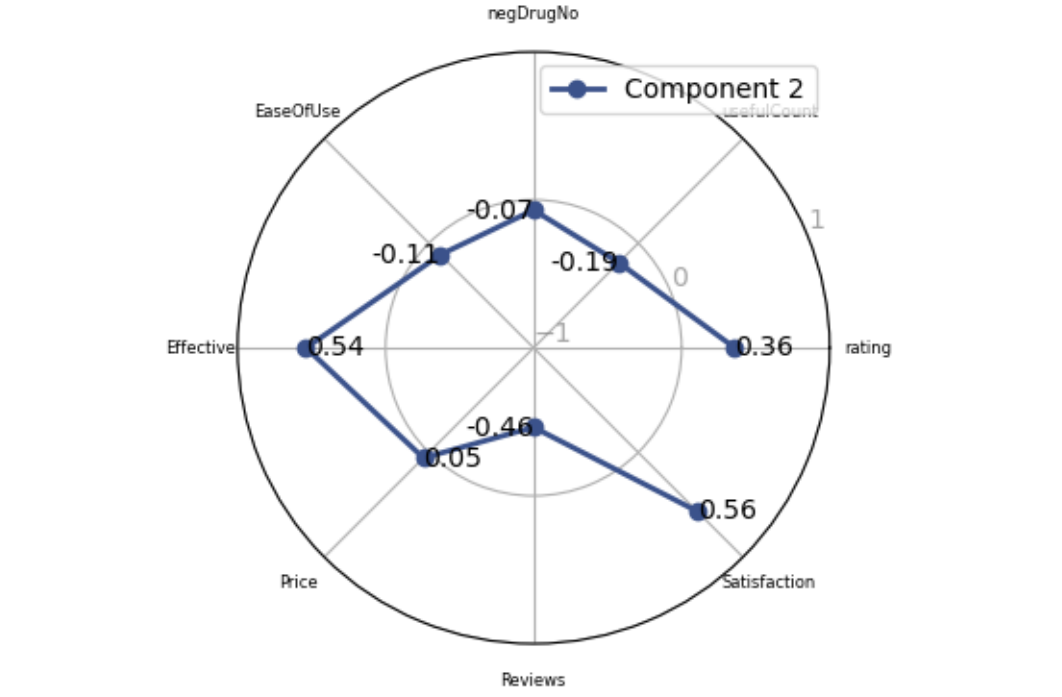
```
[[-0.37252805  0.12609735 -0.62476641  1.00835441  0.40782893  0.18357358
   1.68808786 -0.30702446]
 [-0.23904573 -0.83593832  0.64492016 -0.06071244 -1.5180299   0.63043542
   0.91063956 -1.68863452]
 [ 1.77987433 -1.60277835  0.64492016 -0.10822652  1.3367726   0.80965612
  -1.30070266  1.19368992]
 [ 0.67864521 -0.68257032  0.70538143  1.76857973 -0.81565786 -0.00663102
  -1.44978191 -0.35466618]
 [-1.52381304  1.87356312  0.5844589   0.41442838  1.38208693 -2.55361533
  -0.45442945  1.69392804]
```

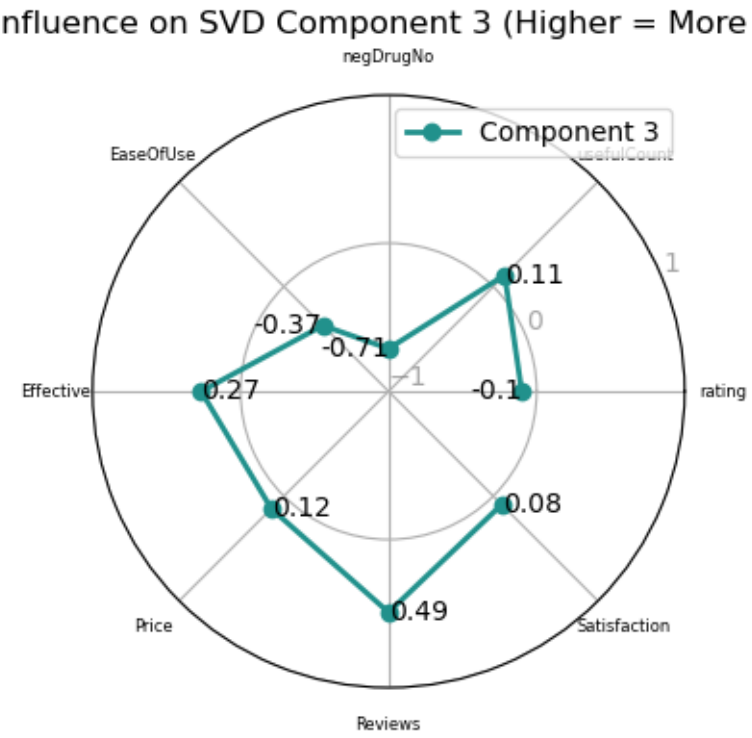
```
 [ 0.65111448 -0.70557552 -2.55952689 -0.27452581  0.58908623  0.38329757
   0.63956379  0.14557194]
 [-0.920759    0.37639925 -0.14107629 -1.15353633  0.22657163  0.8275965
  -0.50809798  0.55052661]
 [-0.91567159  0.88086772  0.5844589   0.20061501 -1.29145828 -0.64845716
   0.75746063 -1.14075467]
 [ 0.86218339  0.56993507  0.16123004 -1.79497644 -0.31720028  0.37414432
  -0.28273984 -0.09263669]]
[[ 0.09981872 -0.95051835  1.05626333 -1.38674621 -0.73604066]
 [-1.52973412 -2.11898158 -0.5202006   0.27012047 -0.43472128]
 [-1.3561333   2.93555473 -0.86373183 -0.09081867 -0.63782775]
 [-0.7360457   0.16148184 -2.26128819 -1.04187351  0.71667635]
 [ 4.12879921  0.78859591 -0.23776536 -0.23592129  0.01026134]
 [-1.06669781  0.69253908  2.30972974 -0.95187357  0.74559869]
 [ 0.29410686  0.44550907  0.61748641  1.29358146 -0.2687326 ]
 [ 0.67475985 -2.27217289 -0.4428962   0.27935396  0.17805087]
 [-0.50887371  0.3179922   0.3424027   1.86417736  0.42673504]]
[[ 4.49396202e+00  1.25734996e+00 -3.24258477e+00 -1.69816717e+00
  -2.68976085e-01  2.44615143e+00 -1.33794998e+00  1.24083495e+00
  -2.89062036e+00]
 [ 1.25734996e+00  7.36272575e+00 -3.44380388e+00  1.36711482e+00
  -7.93148757e+00 -1.61848447e+00 -1.24890101e+00  4.01094093e+00
   2.44545442e-01]
 [-3.24258477e+00 -3.44380388e+00  1.16176840e+01  3.06286697e+00
  -3.06398906e+00  1.09546805e+00  4.29549230e-01 -7.34154498e+00
   8.86354407e-01]
 [-1.69816717e+00  1.36711482e+00  3.06286697e+00  7.28038933e+00
  -2.12083078e+00 -2.79990896e+00 -3.08119175e+00 -2.54994540e-02
  -1.98477302e+00]
 [-2.68976085e-01 -7.93148757e+00 -3.06398906e+00 -2.12083078e+00
   1.77811630e+01 -4.17500323e+00  1.11087698e+00  1.03534855e+00
  -2.36710180e+00]
 [ 2.44615143e+00 -1.61848447e+00  1.09546805e+00 -2.79990896e+00
  -4.17500323e+00  8.41428676e+00 -1.06566546e-02 -3.44945904e+00
   9.76061257e-02]
 [-1.33794998e+00 -1.24890101e+00  4.29549230e-01 -3.08119175e+00
   1.11087698e+00 -1.06566546e-02  2.41183684e+00 -7.73785479e-01
   2.50022182e+00]
 [ 1.24083495e+00  4.01094093e+00 -7.34154498e+00 -2.54994540e-02
   1.03534855e+00 -3.44945904e+00 -7.73785479e-01  5.92396830e+00
  -6.20803778e-01]
 [-2.89062036e+00  2.44545442e-01  8.86354407e-01 -1.98477302e+00
  -2.36710180e+00  9.76061257e-02  2.50022182e+00 -6.20803778e-01
   4.13457115e+00]]
[-5.42775701e-16 -5.36607795e-16 -7.03141249e-16 -2.22044605e-15
  1.28292438e-15  2.96059473e-16  1.48029737e-15 -3.70074342e-16
  1.48029737e-15]
```

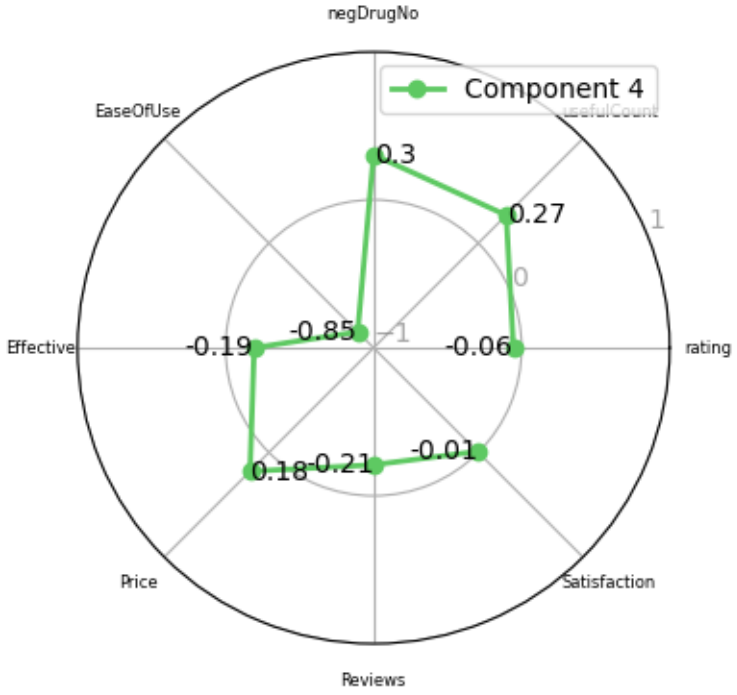

Explained Variance Ratio of SVD Components
Explained Variance Ratio
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
Component

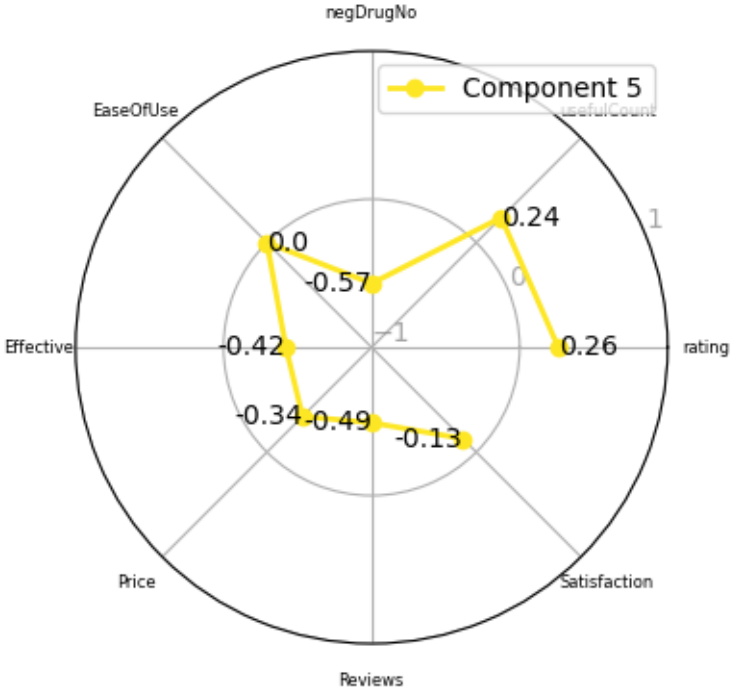

Cumulative Explained Variance
Cumulative Explained Variance
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of Components


Feature Influence on SVD Component 1 (Higher = More Influence)
negDrugNo
Component 1
EaseOfUse
usefulCount
0.13
0.54
1
0.07
0
Effective
0.24
−1
-0.46
rating
-0.56
-0.02
0.32
Price
Satisfaction
Reviews


Feature Influence on SVD Component 2 (Higher = More Influence)
negDrugNo
Component 2
EaseOfUse
usefulCount
-0.07
1
-0.11
-0.19
0
Effective
0.54
−1
0.36
rating
-0.46
0.05
0.56
Price
Satisfaction
Reviews


Feature Influence on SVD Component 3 (Higher = More Influence)
Component 3
negDrugNo
usefulCount
rating
Satisfaction
Reviews
Price
Effective
EaseOfUse
1
0
−1
0.11
-0.1
0.08
0.49
0.12
0.27
-0.37
-0.71


Feature Influence on SVD Component 4 (Higher = More Influence)
Component 4
negDrugNo
usefulCount
rating
Satisfaction
Reviews
Price
Effective
EaseOfUse
1
0
−1
0.3
0.27
-0.06
-0.01
-0.21
0.18
-0.19
-0.85



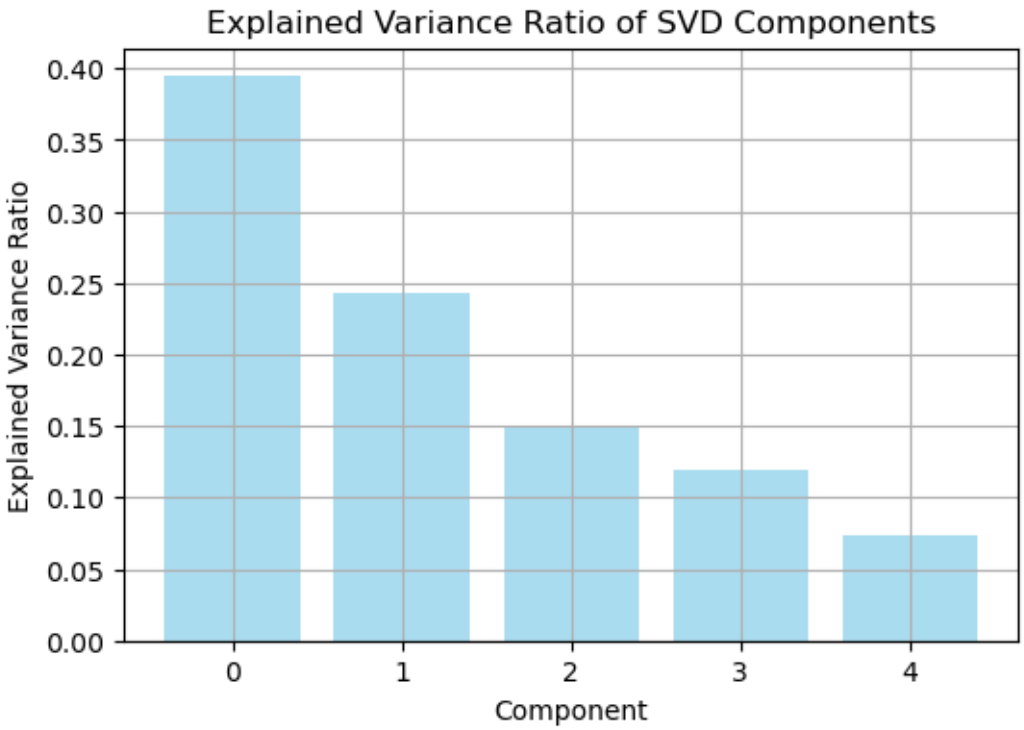
Recommended Drugs for Pneumonia:

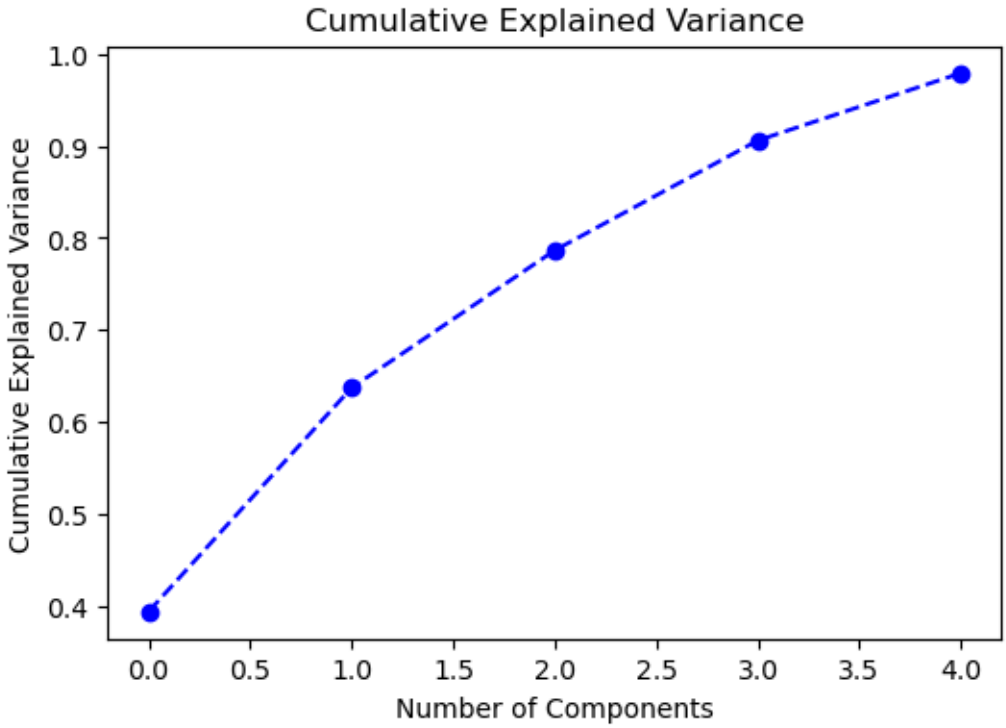
	drugName	rating	usefulCount	negDrugNo	EaseOfUse	Effective	\
0	Cefuroxime	4.000000	28.666667	-8.0	3.76	3.64	
1	Doxycycline	5.095238	18.904762	-32.0	3.10	3.13	
2	Moxifloxacin	8.333333	20.166667	-22.0	2.83	2.89	

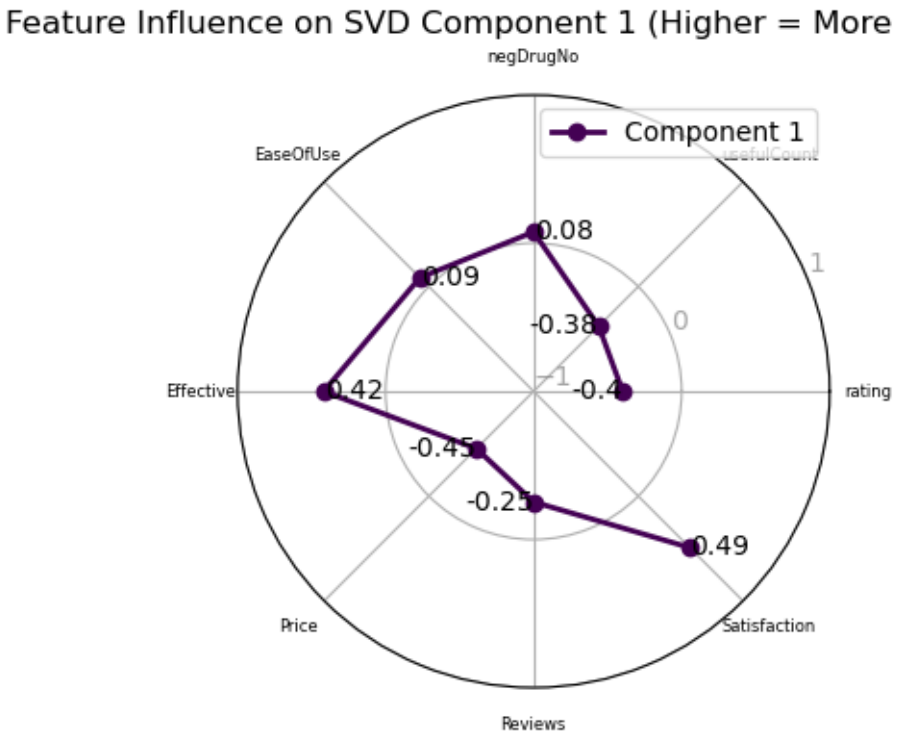
	Price	Reviews	Satisfaction	negDrugNames
0	-259.92	95.12	3.19	Magnesium hydroxide, Calci…
1	-75.22	90.80	2.71	Magnesium oxide, Ampicilli…
2	-99.99	108.94	2.44	Mifepristone, Magnesium ox…

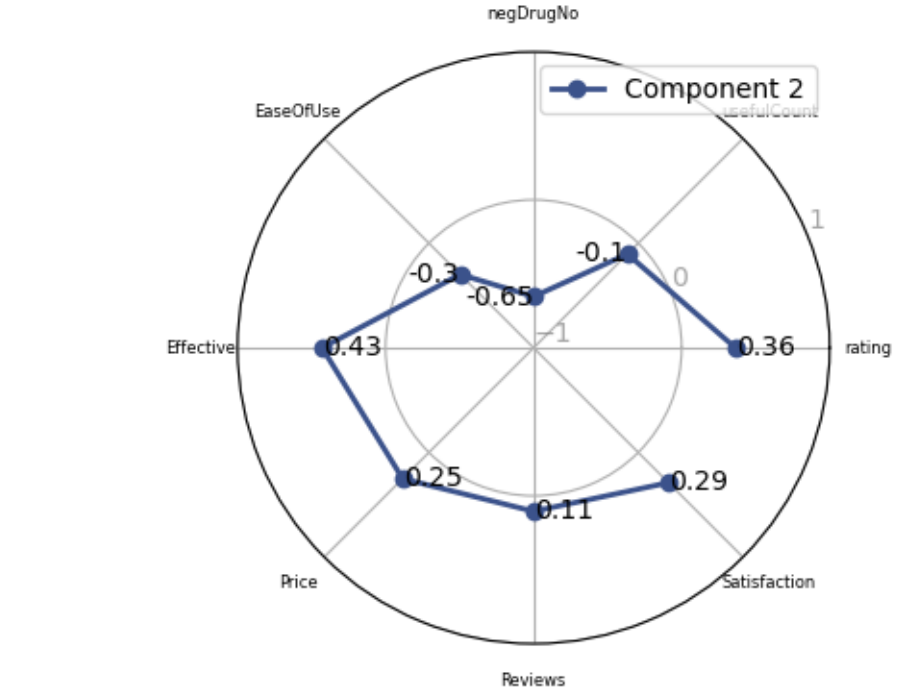
[1133]:
```
recommended_drugs.to_csv("Project Data/results.csv", index=False)
```

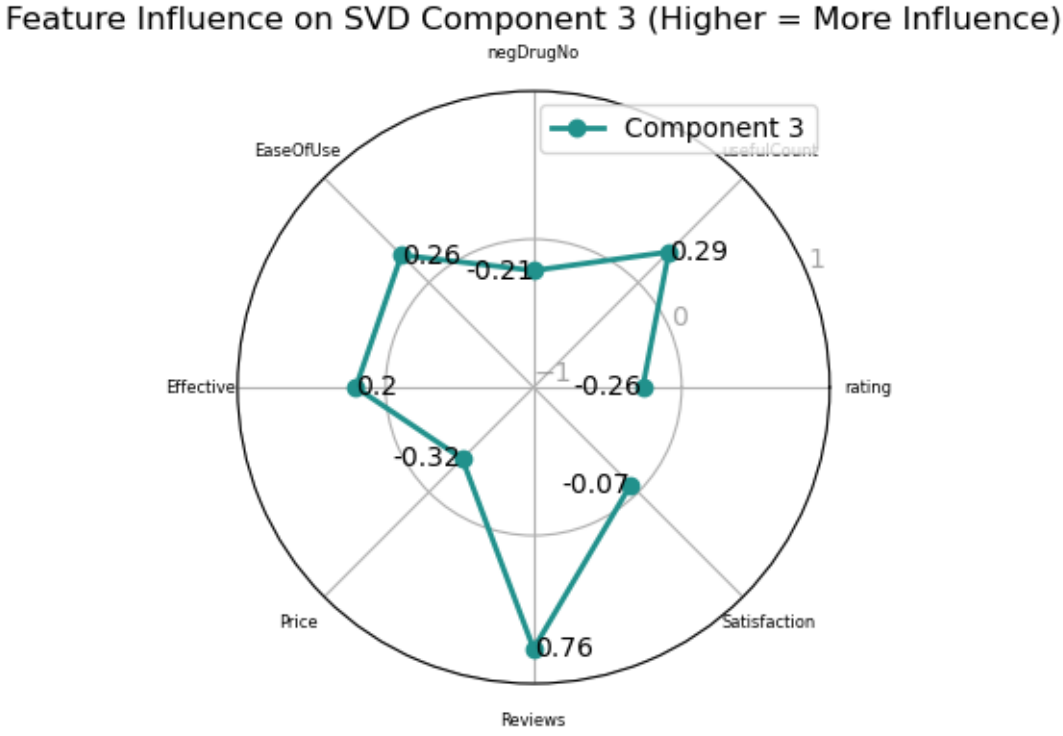
[1153]:
```
recommended_drugs = recommend_drugs_for_condition("Pneumonia", test_df, top_n=3)
print("Recommended Drugs for Pneumonia:")
display(recommended_drugs)
```

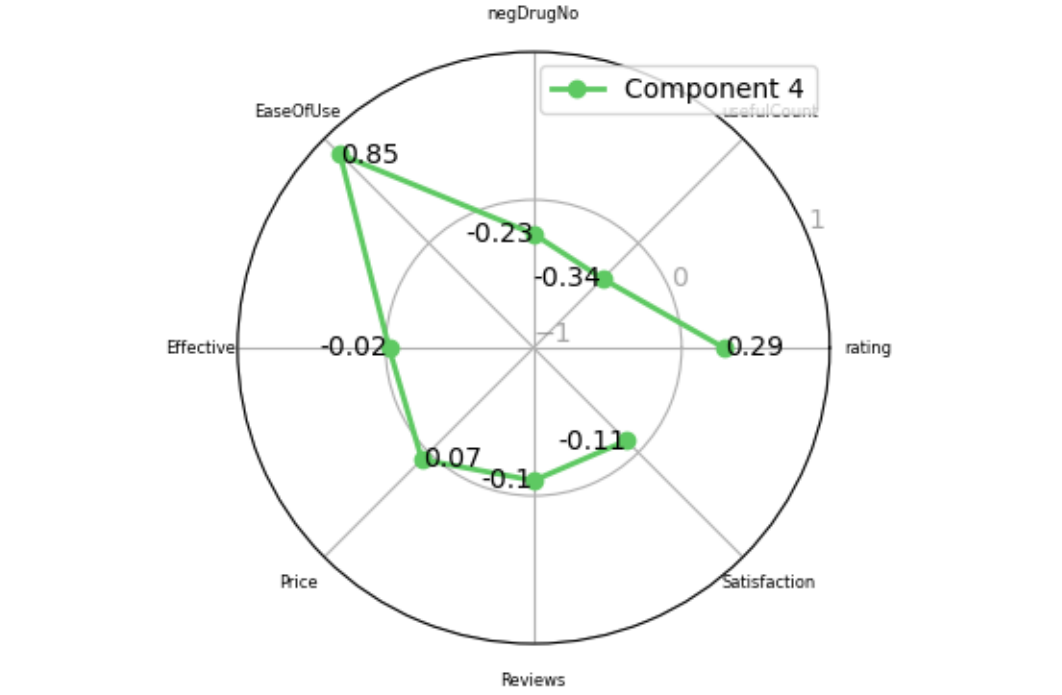

Explained Variance Ratio of SVD Components
Explained Variance Ratio
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
Component

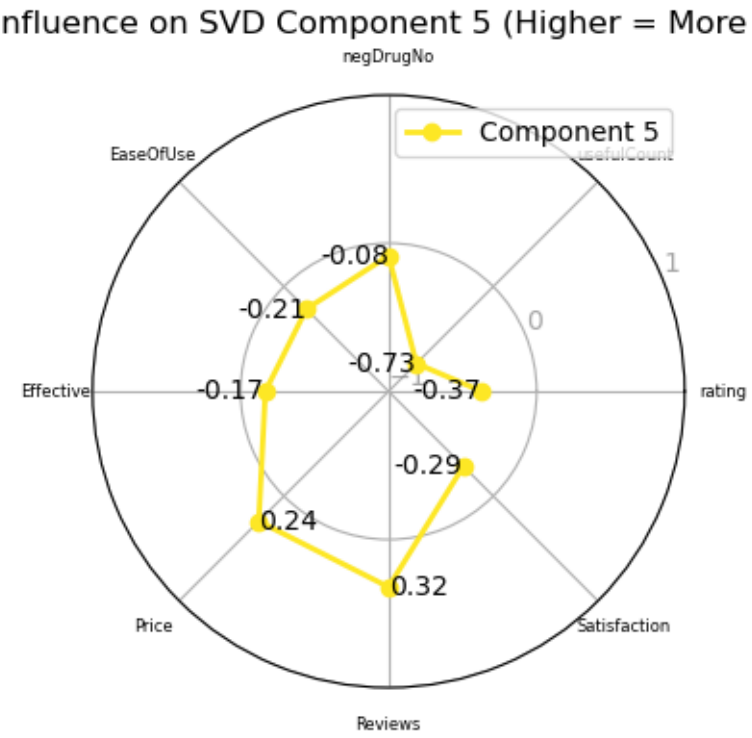

Cumulative Explained Variance
Cumulative Explained Variance
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of Components


Feature Influence on SVD Component 1 (Higher = More Influence)
negDrugNo
Component 1
EaseOfUse
usefulCount
0.08
0.09
-0.38
1
0
−1
Effective
0.42
-0.4
rating
-0.45
-0.25
0.49
Price
Satisfaction
Reviews


Feature Influence on SVD Component 2 (Higher = More Influence)
negDrugNo
Component 2
EaseOfUse
usefulCount
-0.1
1
-0.3
0
-0.65
−1
Effective
0.43
0.36
rating
0.25
0.29
0.11
Price
Satisfaction
Reviews


Feature Influence on SVD Component 3 (Higher = More Influence)
negDrugNo
EaseOfUse
usefulCount
Effective
rating
Price
Satisfaction
Reviews
Component 3
0.26
-0.21
0.29
-0.26
0.2
-0.32
-0.07
0.76
1
0
−1


Feature Influence on SVD Component 4 (Higher = More Influence)
negDrugNo
EaseOfUse
usefulCount
Effective
rating
Price
Satisfaction
Reviews
Component 4
0.85
-0.23
-0.34
-0.02
0.29
-0.11
0.07
-0.1
1
0
−1



Recommended Drugs for Pneumonia:

	drugName	rating	usefulCount	negDrugNo	EaseOfUse	Effective	\
0	Cefuroxime	1.000000	6.000000	-8.0	3.76	3.64	
1	Clarithromycin	9.111111	11.111111	-112.0	3.47	3.29	
2	Doxycycline	7.000000	12.666667	-32.0	3.10	3.13	

	Price	Reviews	Satisfaction	negDrugNames
0	-259.92	95.12	3.19	Magnesium hydroxide, Calci…
1	-99.49	183.18	2.54	Flibanserin, Retapamulin, …
2	-75.22	90.80	2.71	Magnesium oxide, Ampicilli…

[933]:
```
final_df.to_csv("Project Data/final_df.csv", index=False)
```

1 IGNORE BELOW

[1175]:
```
grouped_data.to_csv("Project Data/grouped_data.csv", index=False)
```

[1141]:
```
import pandas as pd

# Group data by 'drugName' and 'condition' and aggregate columns
grouped_data = final_df.groupby(by=['drugName', 'condition']).agg({
    'rating': 'mean',
```

```
    'usefulCount': 'mean',
    'negDrugNo': 'mean',
    'EaseOfUse': 'mean',
    'Effective': 'mean',
    'Reviews': 'mean',
    'Price': 'mean',
    'Satisfaction': 'mean'
}).reset_index()

grouped_data
```

[1141]:
```
           drugName                     condition    rating  usefulCount  \
0        Acebutolol           High Blood Pressure  9.500000    13.000000
1     Acetaminophen                   Muscle Pain  9.500000    29.000000
2     Acetaminophen                          Pain  7.761905    11.666667
3     Acetaminophen                      Sciatica  5.500000    10.000000
4     Acetaminophen                           eve  9.600000    11.600000
…               …                             …         …           …
1041      Verapamil                    Arrhythmia  4.600000    10.800000
1042      Verapamil             Cluster Headaches  9.025641    21.743590
1043      Verapamil           High Blood Pressure  6.250000    20.500000
1044      Verapamil           Migraine Prevention  8.233333    42.000000
1045      Verapamil  Supraventricular Tachycardia  8.000000    14.750000

      negDrugNo  EaseOfUse  Effective  Reviews   Price  Satisfaction
0         -44.0       4.32       3.75    14.50  -24.49          4.20
1         -76.0       3.75       3.47     5.45  -14.76          3.31
2         -76.0       3.75       3.47     5.45  -14.76          3.31
3         -76.0       3.75       3.47     5.45  -14.76          3.31
4         -76.0       3.75       3.47     5.45  -14.76          3.31
…           …          …          …        …       …             …
1041     -210.0       4.71       4.13    42.33 -124.36          4.03
1042     -210.0       4.71       4.13    42.33 -124.36          4.03
1043     -210.0       4.71       4.13    42.33 -124.36          4.03
1044     -210.0       4.71       4.13    42.33 -124.36          4.03
1045     -210.0       4.71       4.13    42.33 -124.36          4.03

[1046 rows x 10 columns]
```

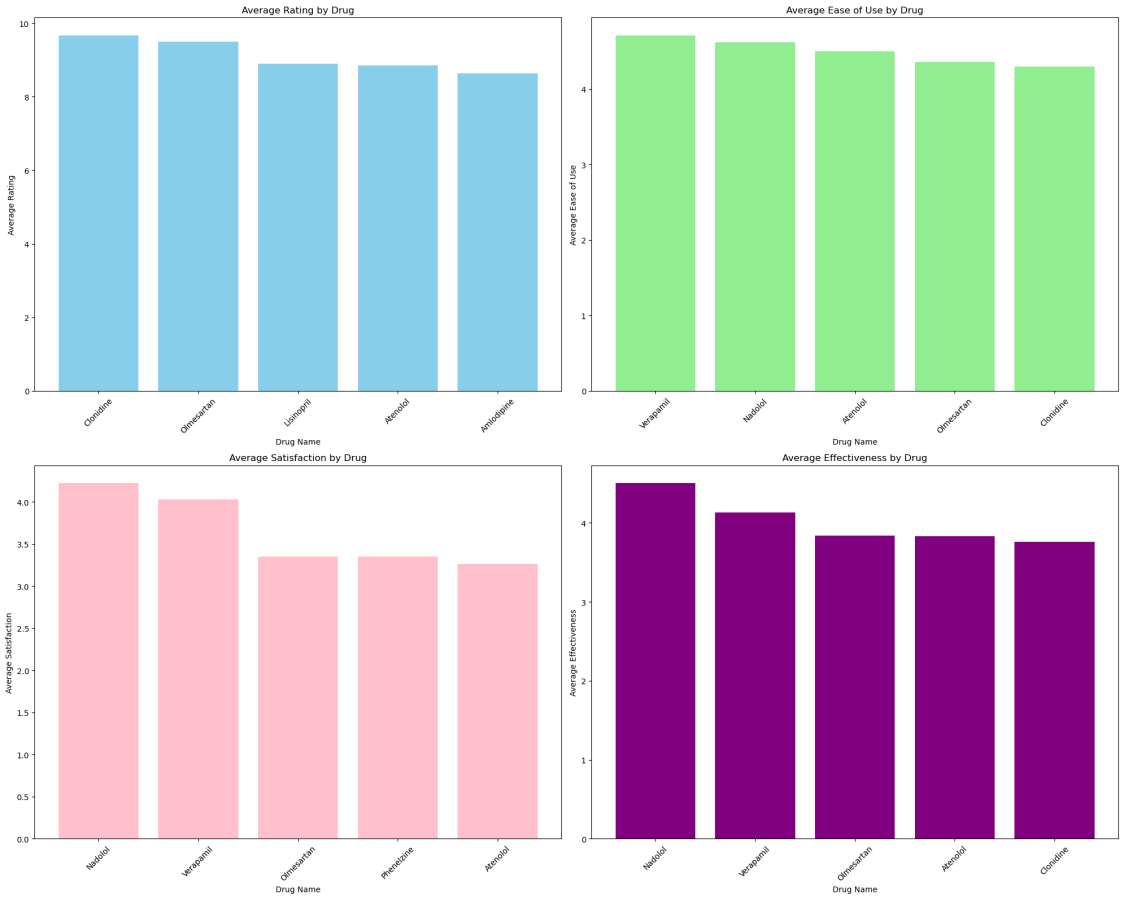
[1154]:
```
grouped_data[grouped_data.condition=='Pneumonia'].sort_values('rating',␣
↪ascending=False)
```

[1154]:
```
           drugName  condition     rating  usefulCount  negDrugNo  EaseOfUse  \
152        Cefixime  Pneumonia  10.000000     6.000000       -6.0       3.54
694    Moxifloxacin  Pneumonia   8.333333    20.166667      -22.0       2.83
157     Cefpodoxime  Pneumonia   8.000000    12.000000       -4.0       4.33
239  Clarithromycin  Pneumonia   7.950000    11.850000     -112.0       3.47
```

```
141        Cefdinir  Pneumonia   6.333333    11.000000       -6.0       3.56
110    Azithromycin  Pneumonia   6.090909    17.272727      -48.0       4.01
552    Levofloxacin  Pneumonia   5.104478    22.194030       -8.0       3.67
368     Doxycycline  Pneumonia   5.095238    18.904762      -32.0       3.10
164      Cefuroxime  Pneumonia   4.000000    28.666667       -8.0       3.76

     Effective  Reviews   Price  Satisfaction
152       3.62    27.00  -76.20          2.98
694       2.89   108.94  -99.99          2.44
157       2.67    15.00 -120.79          2.33
239       3.29   183.18  -99.49          2.54
141       2.36   205.00  -85.99          1.77
110       3.21   267.58 -110.40          2.35
552       2.46   192.67 -155.85          2.00
368       3.13    90.80  -75.22          2.71
164       3.64    95.12 -259.92          3.19
```

[1108]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Create a subplot object with 1 row and 2 columns
fig, axes = plt.subplots(2, 2, figsize=(20, 16))

# Average Rating subplot
axes[0,0].bar(grouped_data.sort_values(by='rating',␣
 ↪ascending=False)['drugName'][:5], grouped_data.sort_values(by='rating',␣
 ↪ascending=False)['rating'][:5], color='skyblue')
axes[0,0].set_title('Average Rating by Drug')
axes[0,0].set_xlabel('Drug Name')
axes[0,0].set_ylabel('Average Rating')
axes[0,0].tick_params(axis='x', rotation=45)

# Useful Count subplot
axes[0,1].bar(grouped_data.sort_values(by='EaseOfUse',␣
 ↪ascending=False)['drugName'][:5], grouped_data.sort_values(by='EaseOfUse',␣
 ↪ascending=False)['EaseOfUse'][:5], color='lightgreen')
axes[0,1].set_title('Average Ease of Use by Drug')
axes[0,1].set_xlabel('Drug Name')
axes[0,1].set_ylabel('Average Ease of Use')
axes[0,1].tick_params(axis='x', rotation=45)

# Average Rating subplot
axes[1,0].bar(grouped_data.sort_values(by='Satisfaction',␣
 ↪ascending=False)['drugName'][:5], grouped_data.
 ↪sort_values(by='Satisfaction', ascending=False)['Satisfaction'][:5],␣
 ↪color='pink')
axes[1,0].set_title('Average Satisfaction by Drug')
```

```
axes[1,0].set_xlabel('Drug Name')
axes[1,0].set_ylabel('Average Satisfaction')
axes[1,0].tick_params(axis='x', rotation=45)

# Useful Count subplot
axes[1,1].bar(grouped_data.sort_values(by='Effective',␣
 ↪ascending=False)['drugName'][:5], grouped_data.sort_values(by='Effective',␣
 ↪ascending=False)['Effective'][:5], color='purple')
axes[1,1].set_title('Average Effectiveness by Drug')
axes[1,1].set_xlabel('Drug Name')
axes[1,1].set_ylabel('Average Effectiveness')
axes[1,1].tick_params(axis='x', rotation=45)

plt.tight_layout()

#
plt.show()
```

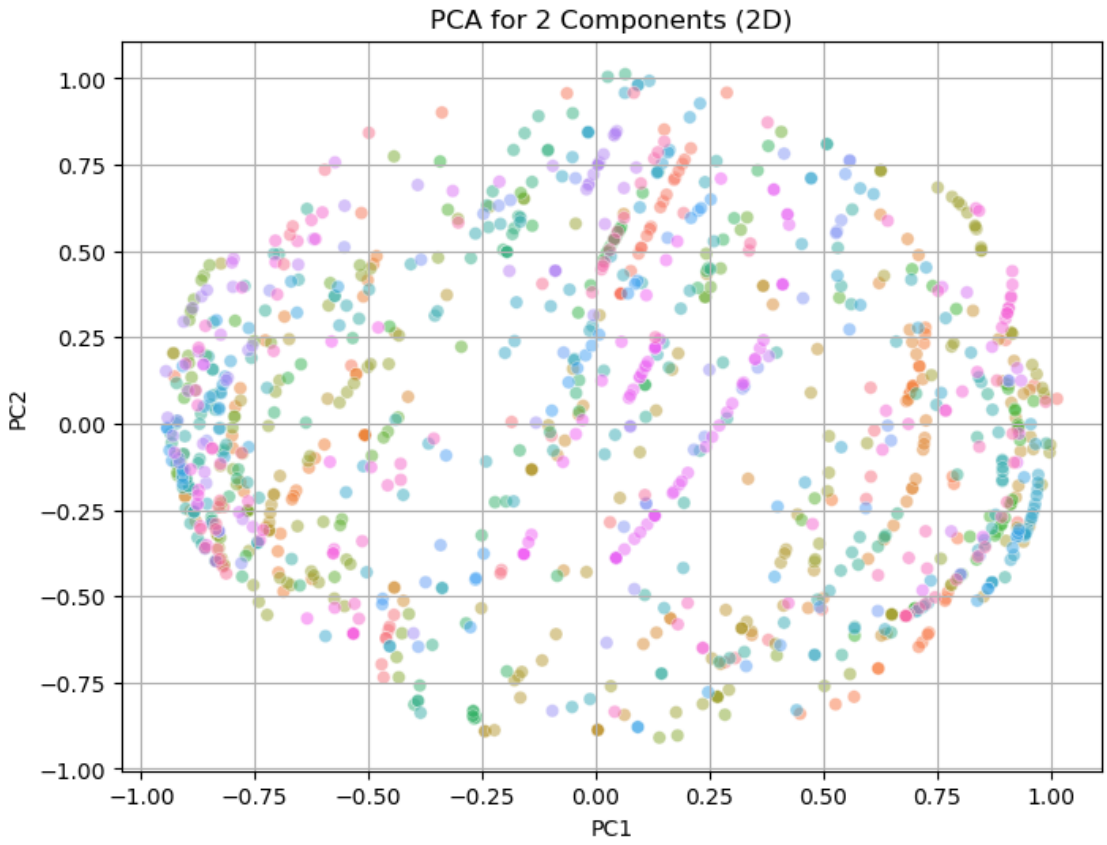


[614]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler, normalize
from sklearn.decomposition import PCA

# Select numerical columns for PCA
numerical_cols = grouped_data.select_dtypes(include=np.number).columns

# Create a StandardScaler object
scaler = StandardScaler()

# Fit the scaler to the data
scaled_data = scaler.fit_transform(grouped_data[numerical_cols])
Migraine
# Normalize the data
scaled_data = normalize(scaled_data)

# Perform PCA with the number of components set to 2
pca_2d = PCA(n_components=2)
pca_data_2d = pca_2d.fit_transform(scaled_data)

# Perform PCA with the number of components set to 3
pca_3d = PCA(n_components=3)
pca_data_3d = pca_3d.fit_transform(scaled_data)

# Create a new DataFrame with the PCA components
pca_df_2d = pd.DataFrame(pca_data_2d, columns=['PC1', 'PC2'])
pca_df_3d = pd.DataFrame(pca_data_3d, columns=['PC1', 'PC2', 'PC3'])

# Concatenate the PCA DataFrame with the original DataFrame for user condition
df_with_pca_2d = pd.concat([grouped_data[['condition','drugName']], pca_df_2d],␣
 ↪axis=1)
df_with_pca_3d = pd.concat([grouped_data[['condition','drugName']], pca_df_3d],␣
 ↪axis=1)

# Getting the most contributing features for each principal component
most_contributing_features_2d = pd.DataFrame(pca_2d.components_,␣
 ↪columns=numerical_cols, index=['PC1', 'PC2'])
most_contributing_features_3d = pd.DataFrame(pca_3d.components_,␣
 ↪columns=numerical_cols, index=['PC1', 'PC2', 'PC3'])

print("Most Contributing Features for 2D PCA:")
print(most_contributing_features_2d)
```

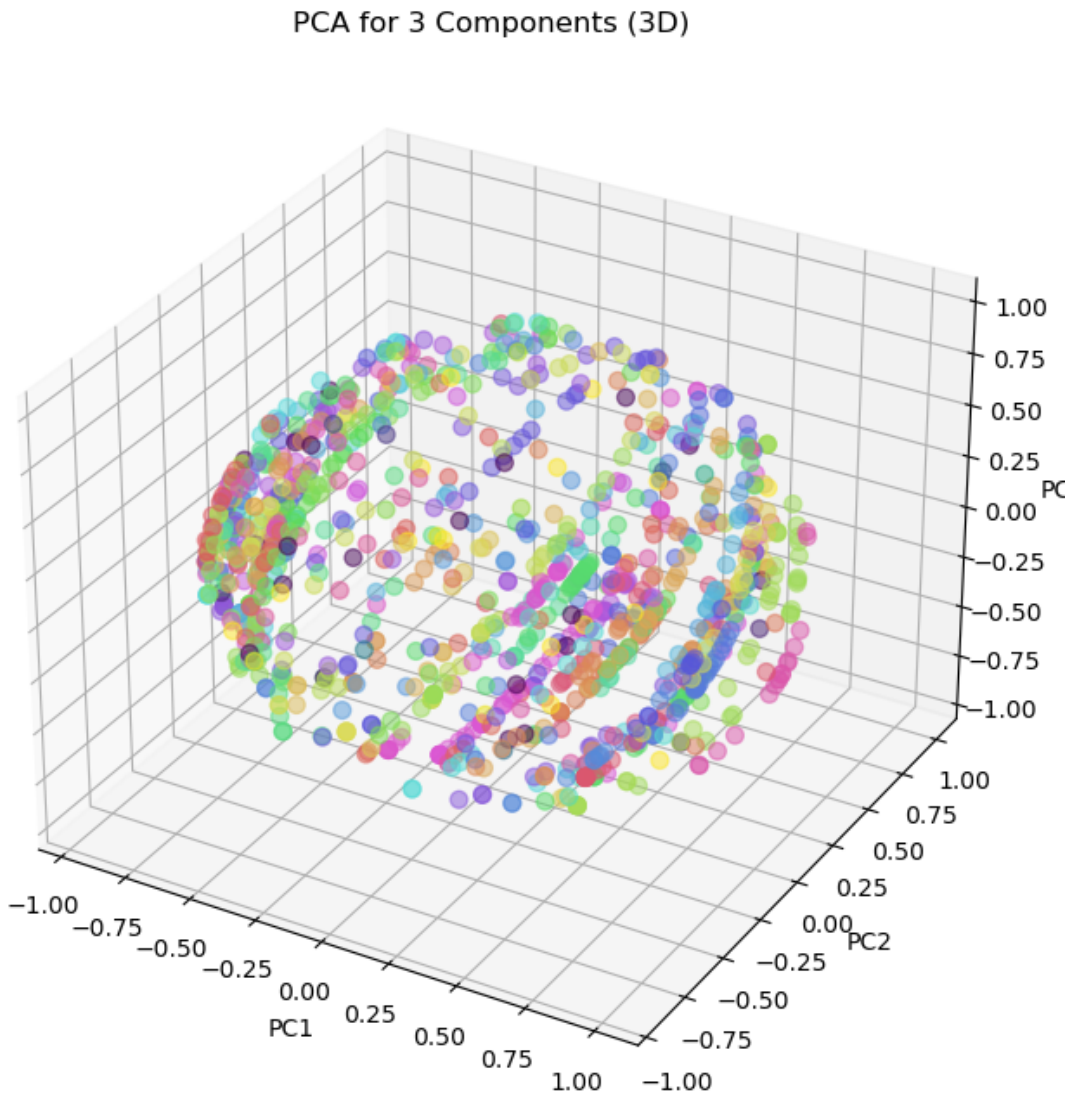
```
print("\nMost Contributing Features for 3D PCA:")
print(most_contributing_features_3d)
```

```
Most Contributing Features for 2D PCA:
       rating  negDrugNo  EaseOfUse  Effective   Reviews     Price  \
PC1 -0.260365  -0.034787  -0.518151  -0.554973  0.060713 -0.008457
PC2 -0.656687  -0.695635   0.186742   0.118005  0.175061  0.002630

     Satisfaction
PC1     -0.592256
PC2      0.073503

Most Contributing Features for 3D PCA:
       rating  negDrugNo  EaseOfUse  Effective   Reviews     Price  \
PC1 -0.260365  -0.034787  -0.518151  -0.554973  0.060713 -0.008457
PC2 -0.656687  -0.695635   0.186742   0.118005  0.175061  0.002630
PC3 -0.704069   0.675059   0.077811   0.113406 -0.151800  0.016340

     Satisfaction
PC1     -0.592256
PC2      0.073503
PC3      0.079733
```

[615]:
```
# Visualization for 2D
plt.figure(figsize=(8, 6))
sns.scatterplot(x='PC1', y='PC2', data=df_with_pca_2d,␣
 ↪hue=df_with_pca_2d['drugName'], alpha=0.5, legend=False)
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.title('PCA for 2 Components (2D)')
plt.grid(True)
plt.show()
```



[616]:
```
# Visualization for 3D
plt.figure(figsize=(16, 8))
ax = plt.axes(projection='3d')
# Define a colormap
palette = sns.color_palette("hls", len(df_with_pca_3d['drugName'].unique()))
colors = {k: v for k, v in zip(df_with_pca_3d['drugName'].unique(), palette)}

for name, group in df_with_pca_3d.groupby('drugName'):
    ax.scatter(group['PC1'], group['PC2'], group['PC3'], c=colors[name],␣
 ↪label=name, s=50, alpha=0.5)

ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
ax.set_zlabel('PC3')
ax.set_title('PCA for 3 Components (3D)')

plt.show()
```



[1157]:
```
import pandas as pd
import numpy as np
from sklearn.decomposition import TruncatedSVD
from sklearn.preprocessing import LabelEncoder

data = grouped_data.copy(deep=True)

features_scaled = scaler.fit_transform(data[['rating', 'usefulCount',␣
 ↪'negDrugNo', 'EaseOfUse', 'Effective', 'Price', 'Reviews', 'Satisfaction']])

# Apply SVD to reduce dimensionality
svd = TruncatedSVD(n_components=8) # Choose the number of components
```

```
features_reduced = svd.fit_transform(features_scaled)


# Create a user-item matrix
user_item_matrix = data.pivot_table(index='condition', columns='drugName',␣
 ↪values=['rating','usefulCount','negDrugNo','EaseOfUse','Effective','Price','Reviews','Satis
 ↪fill_value=0)

# Transpose the user-item matrix to have items as rows and users as columns
item_user_matrix = user_item_matrix.T

# Perform Singular Value Decomposition (SVD)
U, sigma, Vt = np.linalg.svd(item_user_matrix)

# Make predictions
predicted_ratings = np.dot(U[:, :len(sigma)], np.dot(np.diag(sigma), Vt))

# Convert the predicted ratings back to a DataFrame
predicted_ratings_df = pd.DataFrame(predicted_ratings, index=item_user_matrix.
 ↪index, columns=item_user_matrix.columns)

# Generate top N recommendations for the specified condition
def get_top_n_recommendations(predictions, condition, n=3):
    if condition not in predictions.columns:
        return None
    else:
        return predictions.loc[:, condition].nlargest(n).index.tolist()



predicted_ratings_df
```

[1157]:
```
condition                          ADHD  Abnormal Uterine Bleeding  \
            drugName
EaseOfUse   Acebutolol     1.320016e-15              -4.228388e-18
            Acetaminophen  2.396087e-16              -7.742288e-16
            Acetazolamide -3.942024e-16               1.072047e-17
            Acyclovir      6.922834e-16               6.740078e-16
            Amikacin      -5.563854e-16              -1.141990e-15
…                                   …                         …
usefulCount Trimethoprim  -2.154527e-15               2.556982e-15
            Triptorelin    7.905749e-16               2.934393e-16
            Valacyclovir   8.474124e-15               4.143821e-15
            Valsartan      5.584075e-15              -5.243202e-16
            Verapamil      2.813028e-14              -7.972789e-15

condition                          Acne  Actinic Keratosis  Addison's Disease  \
```

```
            drugName
EaseOfUse   Acebutolol    -6.262487e-16       7.612644e-15      -3.139410e-14
            Acetaminophen  8.738670e-16       5.512003e-14      -1.418719e-15
            Acetazolamide  1.195568e-15       1.175134e-13       1.497985e-13
            Acyclovir     -2.730021e-16      -9.139173e-14      -2.795262e-14
            Amikacin       1.112391e-16      -4.837224e-14      -2.929921e-14
…                                 …                 …                 …
usefulCount Trimethoprim   1.530026e-15       1.486912e-13      -1.092876e-15
            Triptorelin    1.527696e-15       1.331964e-17      -3.489656e-16
            Valacyclovir  -5.722419e-15      -1.161913e-13       3.233525e-15
            Valsartan     -6.736799e-15       1.000252e-14      -5.858486e-15
            Verapamil     -2.657596e-15      -1.964297e-13      -6.120104e-15

condition                 Adrenocortical Insufficiency  Agitated State  \
            drugName
EaseOfUse   Acebutolol                   -3.485016e-14    1.080299e-15
            Acetaminophen                -1.900628e-14    7.242471e-16
            Acetazolamide                 1.130238e-13    6.923986e-17
            Acyclovir                    -7.146284e-14    8.716037e-16
            Amikacin                     -1.550829e-13   -4.388850e-16
…                                                 …               …
usefulCount Trimethoprim                 -5.224120e-14    1.589007e-15
            Triptorelin                  -9.235086e-17    2.158702e-16
            Valacyclovir                  6.008041e-14    2.556115e-15
            Valsartan                     2.480221e-15    3.851086e-15
            Verapamil                     6.700196e-14   -2.081668e-16

condition                    Agitation  Alcohol Withdrawal  \
            drugName
EaseOfUse   Acebutolol    -6.352904e-14       -2.107689e-15
            Acetaminophen  2.074057e-14        3.747978e-15
            Acetazolamide -5.705983e-14       -3.100713e-15
            Acyclovir     -5.552403e-15       -9.873897e-16
            Amikacin      -5.655372e-14        4.665322e-16
…                                 …                 …
usefulCount Trimethoprim   4.780898e-14        7.095886e-15
            Triptorelin   -1.701034e-16        4.053135e-16
            Valacyclovir  -4.139484e-14       -7.454107e-15
            Valsartan      1.263399e-14        1.591175e-14
            Verapamil     -6.040134e-14        8.326673e-16

condition                 Allergic Reactions  …           eve  \
            drugName                          …
EaseOfUse   Acebutolol          -1.452831e-17  … -4.571647e-15
            Acetaminophen        4.100453e-15  …  3.750000e+00
            Acetazolamide       -9.700785e-16  … -3.163898e-15
            Acyclovir            1.696302e-16  … -5.974632e-17
```

```
            Amikacin            -6.714464e-16  … -9.183192e-16
…                                         …  …           …
usefulCount Trimethoprim         9.749146e-15  …  1.061998e-14
            Triptorelin          5.367341e-16  … -5.288260e-17
            Valacyclovir        -1.432448e-14  … -1.104151e-14
            Valsartan            7.054687e-15  … -2.714842e-16
            Verapamil           -1.273634e-14  … -5.551115e-15

condition                  ibrocystic Breast Disease   ibromyalgia  \
            drugName
EaseOfUse   Acebutolol                -1.654357e-16  1.332701e-15
            Acetaminophen             -1.962419e-15 -1.987126e-15
            Acetazolamide             -1.753703e-15  3.464433e-16
            Acyclovir                 -1.164867e-16 -3.665837e-15
            Amikacin                  -1.299769e-15 -2.964995e-15
…                                               …           …
usefulCount Trimethoprim              -1.204803e-14 -1.117162e-15
            Triptorelin               -1.818207e-16 -3.971318e-15
            Valacyclovir               8.286015e-15  5.559789e-16
            Valsartan                  2.856331e-16  9.487203e-15
            Verapamil                  1.809013e-14  1.574262e-15

condition                  llicular Lymphoma  m Pain Disorde  mance Anxiety  \
            drugName
EaseOfUse   Acebutolol        -5.799873e-15   -1.713039e-16  -7.741204e-16
            Acetaminophen     -2.484341e-15    9.671734e-15  -1.646253e-15
            Acetazolamide     -2.655206e-14    2.321873e-15  -5.212343e-15
            Acyclovir         -4.894551e-15   -9.606099e-16  -3.370100e-15
            Amikacin          -5.764611e-14    4.305909e-16   6.553406e-15
…                                       …              …              …
usefulCount Trimethoprim      -6.793177e-15    2.279774e-14  -3.747003e-15
            Triptorelin        1.167923e-15   -6.739585e-16  -4.903903e-17
            Valacyclovir       3.549678e-16   -2.862267e-14   1.867430e-15
            Valsartan          1.402009e-16    5.341214e-15   2.472848e-15
            Verapamil          1.559516e-15   -4.581058e-14   9.658940e-15

condition                  mulation) (phenylephrine)  \
            drugName
EaseOfUse   Acebutolol               -3.946496e-17
            Acetaminophen            -7.836071e-16
            Acetazolamide             3.570672e-16
            Acyclovir                -3.123858e-17
            Amikacin                 -9.898766e-17
…                                            …
usefulCount Trimethoprim              1.609823e-15
            Triptorelin              -2.493559e-18
            Valacyclovir              8.213916e-16
```

```
             Valsartan                  -3.122502e-17
             Verapamil                   1.190714e-14

condition                 t Pac with Cyclobenzaprine (cyclobenzaprine)  \
             drugName
EaseOfUse    Acebutolol                     1.216475e-16
             Acetaminophen                 -9.627715e-16
             Acetazolamide                  9.333354e-16
             Acyclovir                     -1.478493e-15
             Amikacin                      -6.500199e-16
…                                                  …
usefulCount  Trimethoprim                  -6.848688e-15
             Triptorelin                   -2.684462e-16
             Valacyclovir                  -2.889074e-15
             Valsartan                      4.683753e-15
             Verapamil                      2.942091e-15

condition                 ungal Pneumonia   zen Shoulde
             drugName
EaseOfUse    Acebutolol        4.526435e-14 -7.318744e-15
             Acetaminophen    -2.163666e-17  3.219647e-15
             Acetazolamide     8.949969e-14  2.675269e-16
             Acyclovir        -1.754488e-14 -4.449057e-16
             Amikacin          7.299609e-14  2.021197e-15
…                                     …           …
usefulCount  Trimethoprim     -4.277016e-15  2.491063e-15
             Triptorelin      -1.238599e-17 -1.293386e-16
             Valacyclovir      4.692437e-15 -5.993470e-16
             Valsartan         1.707179e-16 -7.612183e-15
             Verapamil         6.032452e-15 -7.771561e-15

[1392 rows x 361 columns]
```

[1158]:
```python
# Prompt the user to input their condition
user_condition = input("Enter your condition: ")

# Check if the condition exists in the predictions
if user_condition not in predicted_ratings_df.columns:
    print("No recommendations available for the specified condition.")
else:
    # Get top N recommendations for the specified condition
    top_n = get_top_n_recommendations(predicted_ratings_df, user_condition)
    print(top_n)
    # Print the top 3 recommendations for the specified condition
    if top_n is None:
        print("No recommendations available for the specified condition.")
    else:
```

```
        print(f"Top 3 drug recommendations for condition '{user_condition}':")
        for drug in top_n:
            print("Drug Name:", drug[1])
```

```
Enter your condition: Pneumonia
[('Reviews', 'Azithromycin'), ('Reviews', 'Cefdinir'), ('Reviews',
'Levofloxacin')]
Top 3 drug recommendations for condition 'Pneumonia':
Drug Name: Azithromycin
Drug Name: Cefdinir
Drug Name: Levofloxacin
```

[ ]:

[1162]:
```
grouped_data[(grouped_data['drugName']=='Cedax')]
```

[1162]:
```
Empty DataFrame
Columns: [drugName, condition, rating, usefulCount, negDrugNo, EaseOfUse,
Effective, Reviews, Price, Satisfaction]
Index: []
```

[678]:
```
# Selecting relevant features for matrix factorization
X = df[['PC1', 'PC2', 'PC3']].values

# Fit the Truncated SVD
svd = TruncatedSVD(n_components=2)
X_transformed = svd.fit_transform(X)

X_transformed
```

[678]:
```
array([[-0.80402973, -0.28574754],
       [-0.21740457,  0.28439644],
       [-0.04487198, -0.42982595],
       …,
       [-0.61873541,  0.39422298],
       [-0.66446759,  0.62678767],
       [-0.68135606,  0.2644798 ]])
```

[680]:
```
# Calculate the item features matrix
item_features = np.dot(np.diag(svd.singular_values_), svd.components_)

item_features
```

[680]:
```
array([[ 1.95523396e+01,  0.00000000e+00,  0.00000000e+00],
       [ 0.00000000e+00,  1.49635811e+01, -8.28572373e-14]])
```

[681]:
```
# Function to get top N recommendations based on a given condition
def get_top_n_recommendations(df, condition, n=3):
```

```
    condition_code = le_condition.transform([condition])[0]

    drug_ratings = {}
    for i, row in df.iterrows():
        if row['condition'] == condition:
            drug_ratings[row['drugName']] = np.
↪dot(X_transformed[row['drugCode']], X_transformed[condition_code])

    # Sort the drugs by their ratings
    top_n_drugs = sorted(drug_ratings.items(), key=lambda x: x[1],␣
↪reverse=True)[:n]
    return top_n_drugs
```

[682]:
```
grouped_data['condition'].unique()
```

[682]:
```
array(['High Blood Pressure', 'Muscle Pain', 'Pain', 'Sciatica', 'eve',
       'Edema', 'Epilepsy', 'Glaucoma', 'Hydrocephalus',
       'Mountain Sickness / Altitude Sickness', 'Pseudotumor Cerebri',
       'Seizure Prevention', 'Cold Sores', 'Herpes Simplex',
       'Herpes Simplex, Mucocutaneous/Immunocompetent Host',
       'Herpes Simplex, Mucocutaneous/Immunocompromised Host',
       'Herpes Simplex, Suppression', 'Herpes Zoste', 'Mononucleosis',
       'Not Listed / Othe', 'Psoriasis', 'Urinary Tract Infection',
       'Anxiety and Stress', 'Burning Mouth Syndrome',
       'Chronic Myofascial Pain', 'Cyclic Vomiting Syndrome',
       'Depression', 'Dysautonomia', 'Hyperhidrosis', 'Insomnia',
       'Interstitial Cystitis', 'Irritable Bowel Syndrome',
       'Migraine Prevention', 'Neurotic Depression',
       'Persistent Depressive Disorde', 'Post Traumatic Stress Disorde',
       'Pudendal Neuralgia', 'Reflex Sympathetic Dystrophy Syndrome',
       'Urinary Incontinence', 'Vulvodynia', 'ibromyalgia',
       'Coronary Artery Disease', 'Heart Failure', "Raynaud's Syndrome",
       'Bacterial Infection', 'Bladder Infection', 'Bronchitis',
       'Dental Abscess', 'Helicobacter Pylori Infection', 'Otitis Media',
       'Pneumonia', 'Sinusitis', 'Skin or Soft Tissue Infection',
       'Tonsillitis/Pharyngitis', 'Upper Respiratory Tract Infection',
       'Meningitis', 'Agitated State', 'Autism', 'Bipolar Disorde',
       'Borderline Personality Disorde', 'Major Depressive Disorde',
       'Obsessive Compulsive Disorde', 'Schizoaffective Disorde',
       'Schizophrenia', "Tourette's Syndrome", 'ADHD',
       'Chronic Fatigue Syndrome', 'Hypersomnia', 'Narcolepsy',
       'Obstructive Sleep Apnea/Hypopnea Syndrome',
       'Shift Work Sleep Disorde', 'Angina', 'Back Pain', 'Heart Attack',
       'Osteoarthritis', 'Rheumatoid Arthritis',
       'Thromboembolic Stroke Prophylaxis', 'Transient Ischemic Attack',
       'Alcohol Withdrawal', 'Anxiety', 'Mitral Valve Prolapse',
       'Supraventricular Tachycardia', 'Ventricular Tachycardia',
```

'High Cholesterol', 'High Cholesterol, Familial Heterozygous',
'High Cholesterol, Familial Homozygous',
'Prevention of Cardiovascular Disease',
'Bacterial Endocarditis Prevention', 'COPD, Acute',
'Chlamydia Infection', 'Conjunctivitis, Bacterial',
'Cystic Fibrosis', 'Gonococcal Infection, Uncomplicated',
'Mycobacterium avium-intracellulare, Treatment', 'Pharyngitis',
'STD Prophylaxis', 'Bacterial Skin Infection',
'Left Ventricular Dysfunction', 'Oral and Dental Conditions',
'Pruritus', 'Sore Throat', 'Atopic Dermatitis', 'Dermatitis',
'Dermatological Disorders', 'Inflammatory Conditions',
'Lichen Sclerosus', 'Premature Ventricular Depolarizations',
'Breast Cance', 'Breast Cancer, Metastatic', 'Colorectal Cance',
'Stomach Cance', 'Atrial Fibrillation',
'Skin and Structure Infection', 'Strep Throat',
'Gonococcal Infection, Disseminated', 'Intraabdominal Infection',
'Kidney Infections', 'Lyme Disease', 'Lyme Disease, Neurologic',
'Pelvic Inflammatory Disease', 'Ankylosing Spondylitis',
'Juvenile Rheumatoid Arthritis', 'Period Pain', 'Acne',
'Bone infection', 'Allergic Rhinitis', 'Urticaria',
'Allergic Reactions', 'Cold Symptoms', "Crohn's Disease",
'Dumping Syndrome', 'Hyperlipoproteinemia',
'Hyperlipoproteinemia Type IIa, Elevated LDL',
'Post-Cholecystectomy Diarrhea',
'Pruritus of Partial Biliary Obstruction', 'GERD',
'Human Papilloma Virus', 'Stomach Ulce', 'Anthrax', 'Cholera',
'Diverticulitis', 'Epididymitis, Sexually Transmitted',
'Infectious Diarrhea', 'Prostatitis', "Traveler's Diarrhea",
'Typhoid Feve', 'Brain Tum', 'Legionella Pneumonia',
'Mycoplasma Pneumonia', 'Pertussis', 'llicular Lymphoma',
'Aspiration Pneumonia', 'Bacterial Vaginitis',
'Deep Neck Infection', 'Peritonitis',
'Prevention of Perinatal Group B Streptococcal Disease',
'Surgical Prophylaxis', 'Panic Disorde', 'Trichotillomania',
'Benzodiazepine Withdrawal', 'Hypertensive Emergency',
'Insomnia, Stimulant-Associated', 'Opiate Withdrawal',
'Perimenopausal Symptoms',
'Postural Orthostatic Tachycardia Syndrome',
'Restless Legs Syndrome', 'Oral Thrush', 'Tinea Cruris',
'Tinea Versicol', 'Vaginal Yeast Infection', 'Migraine',
'Muscle Spasm', 'Temporomandibular Joint Disorde',
't Pac with Cyclobenzaprine (cyclobenzaprine)', 'Angioedema',
'Endometriosis', 'ibrocystic Breast Disease', 'SIADH', 'Eczema',
"Addison's Disease", 'Asthma', 'Cerebral Edema', 'Croup',
'Dermatitis Herpeti', 'Lymphoma', 'Macular Edema',
'Multiple Myeloma', 'Multiple Sclerosis',
'Nausea/Vomiting, Chemotherapy Induced', 'Cough', 'Keratosis',

'zen Shoulde', 'Skin Rash', 'Angina Pectoris Prophylaxis',
'Atrial Flutte', 'Motion Sickness', 'Nausea/Vomiting',
'Benign Prostatic Hyperplasia', 'Cance', 'Bronchiectasis',
'Bullous Pemphigoid', 'Ehrlichiosis', 'Gastroenteritis',
'Lyme Disease, Arthritis', 'Malaria', 'Malaria Prevention',
'Ocular Rosacea', 'Q Feve', 'Rosacea', 'Chronic Pain',
'Diabetic Peripheral Neuropathy', 'Generalized Anxiety Disorde',
'Hot Flashes', 'Peripheral Neuropathy', 'Small Fiber Neuropathy',
'Diabetic Kidney Disease', "Barrett's Esophagus",
'Duodenal Ulcer Prophylaxis', 'Erosive Esophagitis',
'NSAID-Induced Gastric Ulce', 'Allergic Urticaria',
'Duodenal Ulce', 'Indigestion', 'Hypertriglyceridemia',
'Gout, Acute', 'Candida Urinary Tract Infection', 'Candidemia',
'Esophageal Candidiasis', 'Onychomycosis, Toenail',
'Systemic Candidiasis', 'Tinea Corporis', 'ungal Pneumonia',
'Granuloma Annulare', 'Lichen Planus', 'Actinic Keratosis',
'Basal Cell Carcinoma', 'Skin Cance', 'Warts',
'Anorexia/Feeding Problems', 'Social Anxiety Disorde',
'Hypercalcemia', 'Pancreatic Cance', 'Bleeding Disorde',
'Prostate Cance', 'Nephrocalcinosis',
'Adrenocortical Insufficiency', 'Hemorrhoids',
'Ulcerative Proctitis', 'Headache', 'Neck Pain',
'Patent Ductus Arteriosus', 'Spondylolisthesis', 'Toothache',
'Night Terrors', 'Primary Nocturnal Enuresis', 'Bursitis',
'Cluster Headaches', 'Gastritis/Duodenitis',
'Streptococcal Infection', 'Abnormal Uterine Bleeding',
'Birth Control', 'Emergency Contraception', 'Anesthesia',
'Burns, External', 'Costochondritis', 'Manscaping Pain',
'Neuropathic Pain', 'Postherpetic Neuralgia', 'Sunburn', 'Vertig',
'Tendonitis', 'Asthma, Maintenance', 'Asthma, acute',
'Dermatologic Lesion', 'Immunosuppression', 'Neuralgia',
'Gastroparesis', 'Lactation Augmentation',
'Nausea/Vomiting, Postoperative', 'Benign Essential Trem',
'Amebiasis', 'Clostridial Infection', "Crohn's Disease, Acute",
"Crohn's Disease, Maintenance", 'Giardiasis',
'Perioral Dermatitis', 'Pseudomembranous Colitis',
'Trichomoniasis', 'm Pain Disorde', 'Periodontitis', 'Alopecia',
'Women (minoxidil)', 'atigue', 'Nasal Polyps', 'Sinus Symptoms',
'Ophthalmic Surgery', 'Impetig',
'Nasal Carriage of Staphylococcus aureus',
'Secondary Cutaneous Bacterial Infections',
'Esophageal Variceal Hemorrhage Prophylaxis',
'Hyperlipoproteinemia Type IV, Elevated VLDL', 'Niacin Deficiency',
'Premature Lab', 'Prevention of Bladder infection', 'Amenorrhea',
'Neurosis', 'Smoking Cessation', 'Campylobacter Gastroenteritis',
'Corneal Ulce', 'Multiple Endocrine Adenomas',
'Zollinger-Ellison Syndrome', 'Avian Influenza', 'Influenza',

```
       'Influenza Prophylaxis', 'Swine Flu', 'Ovarian Cance',
       'Peptic Ulce', 'Stress Ulcer Prophylaxis',
       'Postmenopausal Symptoms', 'Postpartum Depression',
       'Premenstrual Dysphoric Disorde', 'Eye Redness',
       'Nasal Congestion', 'mulation) (phenylephrine)',
       "Parkinson's Disease", 'Periodic Limb Movement Disorde',
       'Hyperlipoproteinemia Type III, Elevated beta-VLDL   IDL',
       'Nightmares', 'Nephrotic Syndrome',
       'Postoperative Ocular Inflammation', 'Autoimmune Hemolytic Anemia',
       'COPD', 'Conjunctivitis, Allergic', 'Epicondylitis, Tennis Elbow',
       'Gouty Arthritis', 'Inflammatory Bowel Disease', 'Leukemia',
       'Mixed Connective Tissue Disease', 'Pemphigus',
       'Psoriatic Arthritis', 'Sarcoidosis',
       'Systemic Lupus Erythematosus', 'Systemic Sclerosis',
       'Ulcerative Colitis', "Dercum's Disease", 'Arrhythmia',
       'Hemangioma', 'Pe', 'Portal Hypertension', 'Thyrotoxicosis',
       'mance Anxiety', 'Cardiovascular Risk Reduction',
       'Gastric Ulcer Maintenance Treatment',
       'Pathological Hypersecretory Conditions', 'Agitation',
       'Asperger Syndrome', 'Head Injury', 'Mania', 'Paranoid Disorde',
       'Severe Mood Dysregulation', 'Tic Disorde', 'Atherosclerosis',
       'Gender Dysphoria', 'Hirsutism', 'Hypokalemia',
       'Primary Hyperaldosteronism',
       'Primary Hyperaldosteronism Diagnosis',
       'Orgran Transplant, Rejection Prophylaxis', 'Glaucoma, Open Angle',
       'Aphthous Ulce', 'Iritis', 'Neuritis', 'Pityriasis rubra pilaris',
       'Biliary Cirrhosis', 'Gallbladder Disease', 'CMV Prophylaxis',
       'Ramsay Hunt Syndrome', 'Varicella-Zoste'], dtype=object)
```

[683]:
```
# Example usage
user_condition = 'Back Pain'
top_recommendations = get_top_n_recommendations(df, user_condition)
print(f"Top 3 drugs for {user_condition}:")
for drug, rating in top_recommendations:
    print(f"{drug}: {rating:.2f}")
```

```
Top 3 drugs for Back Pain:
Duloxetine: 0.08
Diclofenac: 0.01
Ketoprofen: 0.01
```

[684]:
```
grouped_data[grouped_data['condition']=='Back Pain'].sort_values(by='rating',␣
↪ascending=False)
```

[684]:

	drugName	condition	rating	negDrugNo	EaseOfUse	Effective	\
525	Flurbiprofen	Back Pain	9.666667	-54.0	4.66	4.75	
610	Ketoprofen	Back Pain	9.000000	-42.0	4.36	4.38	
91	Aspirin	Back Pain	7.600000	-116.0	3.92	3.72	

```
366     Diclofenac  Back Pain  7.355556      -85.0       4.25       3.90
569      Ibuprofen  Back Pain  7.285714      -38.0       4.30       3.98
444     Duloxetine  Back Pain  7.263158     -175.0       4.01       3.71
597   Indomethacin  Back Pain  7.000000      -75.0       4.40       3.66
1094      Tramadol  Back Pain  6.703704     -214.0       4.24       3.20
798       Naproxen  Back Pain  5.800000      -53.0       4.21       3.33

      Reviews   Price  Satisfaction
525      7.00  -99.49          4.42
610      9.88  -25.46          4.38
91      12.10  -12.11          3.61
366    152.33 -157.19          3.58
569     48.45 -154.05          3.73
444   2367.67 -100.74          3.28
597    133.67  -15.99          3.60
1094   154.50  -26.89          3.02
798    132.33  -43.33          3.12
```

[65]:
```
aggregated_df.to_csv("Project Data/aggregated_df.csv", index=False)
```

[37]:
```
pivot_df = aggregated_df.pivot_table(index=['condition','drugName'],␣
 ↪values='review', aggfunc=lambda x: ', '.join(x))
pivot_df
```

[37]:

condition	drugName	review
45 users found this comment helpful.	Vitamin B12	"Vitam B12 really helps with depression and ha…
ADHD	Amantadine	"Primary related symptom was mild cognitive im…
	Amphetamine	"I have taken Adderall 30 mg XR for a little o…
	Armodafinil	"This drug nuvigil I took for 3 days only half…
	Atomoxetine	"Strattera was my first treatment after I was …
…		
…		
zen Shoulde	Diclofenac	"This medication has been a God send for me. …
	Ibuprofen	"I've found that taking ibuprofen (200 mg…
	Indomethacin	"It works."
	Nabumetone	"The only side effect I have experienced with …
	Naproxen	"Very little relief.

```
I finished PT and after …

[2761 rows x 1 columns]
```

[80]:
```
import nltk
import string
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
import warnings
warnings.filterwarnings("ignore")
from collections import Counter
import re

def clean_doc(row):
    #Converting doc to lowercase
    doc = row['review'][0]

    doc = doc.lower()

    doc = re.sub(r"\W+|_|\\|/"," ",doc)

    #Tokenizing the docs into words
    word_tokens = word_tokenize(doc)

    #Removing stopwords
    filtered_doc = [w for w in word_tokens if not w in stop_words]

    #Joining all words into a string with spaces
    cleaned_doc = ' '.join(filtered_doc)

    #Return cleaned document
    return cleaned_doc
```

[83]:
```
aggregated_df['review']
# = aggregated_df.apply(clean_doc, axis=1)
# aggregated_df
```

[83]:
```
0
1
2
3
4
       ..
69012
69013
69014
```

```
69015
69016
Name: review, Length: 69017, dtype: object
```

[76]:
```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import PCA
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.neighbors import NearestNeighbors
from sklearn.metrics.pairwise import linear_kernel


# Text data preprocessing
tfidf = TfidfVectorizer(stop_words='english')


tfidf_matrix = tfidf.fit_transform(aggregated_df['review'])

tfidf_matrix.shape
```

[76]:
```
(69017, 40104)
```

[77]:
```
cosine_sim_review = linear_kernel(tfidf_matrix, tfidf_matrix)
cosine_sim_review
```

```
---------------------------------------------------------------------------
MemoryError                               Traceback (most recent call last)
Cell In[77], line 1
----> 1 cosine_sim_review = linear_kernel(tfidf_matrix, tfidf_matrix)
      2 cosine_sim_review

File ~\anaconda3\Lib\site-packages\sklearn\utils\_param_validation.py:211, in␣
 ↪validate_params.<locals>.decorator.<locals>.wrapper(*args, **kwargs)
    205 try:
    206     with config_context(
    207         skip_parameter_validation=(
    208             prefer_skip_nested_validation or global_skip_validation
    209         )
    210     ):
--> 211         return func(*args, **kwargs)
    212 except InvalidParameterError as e:
    213     # When the function is just a wrapper around an estimator, we allow
    214     # the function to delegate validation to the estimator, but we␣
 ↪replace
```

```
    215     # the name of the estimator by the name of the function in the erro
    216     # message to avoid confusion.
    217     msg = re.sub(
    218         r"parameter of \w+ must be",
    219         f"parameter of {func.__qualname__} must be",
    220         str(e),
    221     )

File ~\anaconda3\Lib\site-packages\sklearn\metrics\pairwise.py:1329, in␣
 ↪linear_kernel(X, Y, dense_output)
   1304 """
   1305 Compute the linear kernel between X and Y.
   1306
   (…)
   1326     The Gram matrix of the linear kernel, i.e. `X @ Y.T`.
   1327 """
   1328 X, Y = check_pairwise_arrays(X, Y)
-> 1329 return safe_sparse_dot(X, Y.T, dense_output=dense_output)

File ~\anaconda3\Lib\site-packages\sklearn\utils\extmath.py:201, in␣
 ↪safe_sparse_dot(a, b, dense_output)
    193     ret = a @ b
    195 if (
    196     sparse.issparse(a)
    197     and sparse.issparse(b)
    198     and dense_output
    199     and hasattr(ret, "toarray")
    200 ):
--> 201     return ret.toarray()
    202 return ret

File ~\anaconda3\Lib\site-packages\scipy\sparse\_compressed.py:1050, in␣
 ↪_cs_matrix.toarray(self, order, out)
   1048 if out is None and order is None:
   1049     order = self._swap('cf')[0]
-> 1050 out = self._process_toarray_args(order, out)
   1051 if not (out.flags.c_contiguous or out.flags.f_contiguous):
   1052     raise ValueError('Output array must be C or F contiguous')

File ~\anaconda3\Lib\site-packages\scipy\sparse\_base.py:1267, in _spbase.
 ↪_process_toarray_args(self, order, out)
   1265     return out
   1266 else:
-> 1267     return np.zeros(self.shape, dtype=self.dtype, order=order)

MemoryError: Unable to allocate 35.5 GiB for an array with shape (69017, 69017),
 ↪and data type float64
```

[ ]:
```
# Categorical data preprocessing
preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), ['rating', 'usefulCount', 'negDrugNo',␣
 ↪'noConditionTreated']),
        ('cat', OneHotEncoder(), ['drugName', 'condition'])
    ],
    remainder='passthrough'
)



# Combine the transformers into a single pipeline
pipeline = Pipeline([
    ('preprocessor', preprocessor),
    ('pca', PCA(n_components=0.95))  # 95% of the variance
])

# Fit and transform the data
X = pipeline.fit_transform(df.drop(columns=["review", "date", "DBID"]))

# Find similar items
knn = NearestNeighbors(metric='cosine')
knn.fit(X)

# Function to get the most similar items
def get_recommendations(condition, k=5):
    condition_df = df[df['condition'] == condition]
    condition_X = pipeline.transform(condition_df.drop(columns=["review",␣
 ↪"date", "DBID"]))
    distances, indices = knn.kneighbors(condition_X)
    return indices[0][:k], distances[0][:k]

# Get recommendations for a specific condition
condition = "Eye Redness"
rec_indices, distances = get_recommendations(condition)

# Output recommendations
print("Recommendations:")
recommended_drugs = []
for i, distance in zip(rec_indices, distances):
    recommended_drugs.append(df.iloc[i]['drugName'])
    print(f"Index: {i}, Distance: {distance}, Data: {df.iloc[i]}")

# Possible list of conditions based on the recommended drugs
possible_conditions = df[df['drugName'].isin(recommended_drugs)]['condition'].
 ↪unique()
```

```
print("\nPossible list of conditions based on the recommended drugs:")
print(possible_conditions)
```