# Building a Model that Classifies the Side - Effects of a Drug

**Objective Of This Project**

Is to build a classification model that classifies the side effects of a particular drug by age, gender and race

## Loading Important Libraries

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pandas as pd
from nltk.corpus import stopwords
pd.options.mode.chained_assignment = None  # default='warn'
from IPython.html import widgets
from IPython.display import display
from collections import OrderedDict
from textblob import TextBlob
from sklearn.preprocessing import LabelEncoder #One hot encoding for Categorical variab
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split # splitting the data
from sklearn.metrics import accuracy_score # model precision
import matplotlib.pyplot as plt # visualization
from matplotlib import rcParams # figure size
from termcolor import colored as cl # text customization
from sklearn.metrics import accuracy_score, classification_report, precision_score, rec
from sklearn.metrics import confusion_matrix, precision_recall_curve, roc_curve, auc, l
from collections import Counter
import re
import string
```

```
D:\Python\lib\site-packages\IPython\html.py:12: ShimWarning: The `IPython.html` package
has been deprecated since IPython 4.0. You should import from `notebook` instead. `IPyth
on.html.widgets` has moved to `ipywidgets`.
  warn("The `IPython.html` package has been deprecated since IPython 4.0. "
```

## Importing DataSet - "webmd"

In [2]:
```
df = pd.read_csv("D:\TCS project\Data sets\webmd.csv")
```

## Checking Columns name present in the DataSet

In [3]:
```
df.columns
```

Out[3]:
```
Index(['Age', 'Condition', 'Date', 'Drug', 'DrugId', 'EaseofUse',
       'Effectiveness', 'Reviews', 'Satisfaction', 'Sex', 'Sides',
       'UsefulCount'],
      dtype='object')
```

# About this DataSet

1. **Drug (categorical):** name of drug
2. **DrugId (numerical):** drug id
3. **Condition (categorical):** name of condition
4. **Review (text):** patient review
5. **Side (text):** side effects associated with drug (if any)
6. **EaseOfUse (numerical):** 5 star rating
7. **Effectiveness (numerical):** 5 star rating
8. **Satisfaction (numerical):** 5 star rating
9. **Date (date):** date of review entry
10. **UsefulCount (numerical):** number of users who found review useful.
11. **Age (numerical):** age group range of user
12. **Sex (categorical):** gender of user

# Cleaning the Data

By going through the DataSet we can observe that there are empty values/empty string present in the Review and Sides Columns. So we are gonna convert the empty value/empty string into np.nan object and then remove them

In [4]:
```
df.head(10)
```

Out[4]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 75 or over | Stuffy Nose | 9/21/2014 | 25dph-7.5peh | 146724 | 5 | 5 | I'm a retired physician and of all the meds I ... | 5 | N |
| **1** | 25-34 | Cold Symptoms | 1/13/2011 | 25dph-7.5peh | 146724 | 5 | 5 | cleared me right up even with my throat hurtin... | 5 | Fer |
| **2** | 65-74 | Other | 7/16/2012 | warfarin (bulk) 100 % powder | 144731 | 2 | 3 | why did my PTINR go from a normal of 2.5 to ov... | 3 | Fer |
| **3** | 75 or over | Other | 9/23/2010 | warfarin (bulk) 100 % powder | 144731 | 2 | 2 | FALLING AND DON'T REALISE IT | 1 | Fer |

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **4** | 35-44 | Other | 1/6/2009 | warfarin (bulk) 100 % powder | 144731 | 1 | 1 | My grandfather was prescribed this medication ... | 1 | N |
| **5** | 55-64 | Other | 7/19/2008 | warfarin (bulk) 100 % powder | 144731 | 4 | 4 | help heart condition operation well | 4 | N |
| **6** | 25-34 | Birth Control | 6/15/2017 | wymzya fe | 163180 | 5 | 5 | Haven't gotten pregnant so it does it's job. I... | 2 | Fer |
| **7** | 45-54 | Disease of Ovaries with Cysts | 1/30/2017 | wymzya fe | 163180 | 5 | 5 | I have take this for 5 years age 45-50 to prev... | 5 | Fer |
| **8** | 25-34 | Acne | 4/27/2016 | wymzya fe | 163180 | 4 | 2 | | 2 | Fer |
| **9** | 55-64 | Stuffy Nose | 10/29/2012 | 12 hour nasal relief spray, non-aerosol | 9800 | 4 | 2 | The 12 hour spray only works for me for 6 hours. | 2 | N |

Cleaning the 'Review' and 'Sides' part from the columns is carried out to make data more readable and less redundant.

Cleaning the 'Review' and 'Sides' part from the columns is carried out to make data more readable and less redundant.

In [5]:
```
df['Reviews'].replace(' ', np.nan, inplace=True)
```

In [6]:
```
df['Reviews'].isnull().sum()
```

Out[6]: 41861

Next we are gonna replace all the empty strings in the "Reviews" column with np.nan objects in the following manner

In [7]:
```
df['Sides'].replace(' ', np.nan, inplace=True)
```

In [8]:
```
df['Sides'].isnull().sum()
```

Out[8]: 17460

Next we are gonna replace all the empty strings in the "Sides" column with np.nan objects in the following manner

Now we are gonna remove all the records which had no value in their "Reviews" Column

In [9]:
```
df.dropna(subset=['Reviews'], inplace=True)
```

Now we are gonna remove all the records which had no value in their "Reviews" Column

In [10]:
```
df.dropna(subset=['Sides'], inplace=True)
```

In [11]:
```
df['Reviews'].isnull().sum()
```

Out[11]: 0

In [12]:
```
df['Sides'].isnull().sum()
```

Out[12]: 0

# Determining if the DataFrame still contains any Null-Values or Not

In [13]:
```
df.isnull().values.any()
```

Out[13]: False

Since there are no more null-values present in the dataset therefore we can proceed with Exploratory Data Analysis.

---

# Removing Empty Strings from the column "Sex"

We are gonna check if the "Sex" column has any empty strings/ Empty values

In [14]:
```
print(df["Sex"].value_counts())
```

```
Female    205842
Male       78138
           21367
Name: Sex, dtype: int64
```

Here we can see in below that "Sex" column have values (21367) which have not been categorised.

Pandas will recognise a value as null if it is a np.nan object, which will print as NaN in the DataFrame. But the missing values are probably empty strings, which Pandas doesn't recognise as null. To fix this, We can convert the empty stings (or whatever is in the empty cells) to np.nan objects using replace(), and then call dropna()on your DataFrame to delete rows with null values present in the "Sex" column.

Therefore we will replace all the empty strings in the Sex column with np.nan objects.

In [15]:
```
df['Sex'].replace(' ', np.nan, inplace=True)
```

In [16]:
```
df['Sex'].isnull().sum()
```

Out[16]:
```
21367
```

Now we will drop all the null values present in the "Sex" column.

In [17]:
```
df.dropna(subset=['Sex'], inplace=True)
```

All the null values have been removed that were present in the form of empty string/ empty cells in "Sex" column.

In [18]:
```
print(df["Sex"].value_counts())
```

```
Female    205842
Male       78138
Name: Sex, dtype: int64
```

So now we can see all the values which had no category have been removed from the column "Sex"

---

# Removing Empty Strings from the column "Age"

We are now gonna check if the column "Age" has empty strings or not

In [19]:
```
df.Age.value_counts()
```

Out[19]:
```
45-54         64129
55-64         57473
35-44         45476
25-34         42800
65-74         29986
19-24         21236
75 or over    10730
13-18          5975
               3825
7-12           1334
3-6             623
0-2             393
Name: Age, dtype: int64
```

Here we can observe the same issue that the column "Age" has empty values/empty string (5975, 3825). Therefore we are gonna convert these empty values into 'np.nan' objects

In [20]:
```
df['Age'].replace(' ', np.nan, inplace=True)
```

Now we can drop the null values present in the "Age" column

In [21]:
```
df.dropna(subset=['Age'], inplace=True)
```

In [22]:
```
df.Age.value_counts()
```

Out[22]:
```
45-54         64129
55-64         57473
35-44         45476
25-34         42800
65-74         29986
19-24         21236
75 or over    10730
```

```
13-18          5975
7-12           1334
3-6             623
0-2             393
Name: Age, dtype: int64
```

So now we can see all the values which had no category have been removed from the column "Age"

**For easy convetion we gonna categorise each age group as numbers and store them in new column "Age_Group_Number" in the following manner**

**Age**

**0-2 > 1**
**3-6 > 2**
**7-12 > 3**
**13-18 > 4**
**19-24 > 5**
**25-34 > 6**
**35-44 > 7**
**45-54 > 8**
**55-64 > 9**
**65-74 > 10**
**75 or over > 11**

In [23]:
```
def age_fun(df):
    if df['Age'] == '0-2':
        return 1
    elif df['Age'] == '13-18':
        return 4
    elif df['Age'] == '19-24':
        return 5
    elif df['Age'] == '25-34':
        return 6
    elif df['Age'] == '3-6':
        return 2
    elif df['Age'] == '35-44':
        return 7
    elif df['Age'] == '45-54':
        return 8
    elif df['Age'] == '55-64':
        return 9
    elif df['Age'] == '65-74':
        return 10
    elif df['Age'] == '7-12':
        return 3
    elif df['Age'] == '75 or over':
        return 11
```

In [24]:
```
df['Age_Group_number'] = df.apply(age_fun, axis = 1)
```

In [25]:
```
df.head(1)
```

Out[25]:

| Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | Sex |
|---|---|---|---|---|---|---|---|---|---|

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | Sex |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 75 or over | Stuffy Nose | 9/21/2014 | 25dph-7.5peh | 146724 | 5 | 5 | I'm a retired physician and of all the meds I ... | 5 | Male |

# "Effectiveness" column contains rating value from 1 to 5, Therefore values that indicates rating of 6 will be removed in order to get a clean Data Set

In [26]:
```
df.Effectiveness.value_counts()
```

Out[26]:
```
5    101860
4     62305
1     47150
3     45180
2     23658
6         2
Name: Effectiveness, dtype: int64
```

In [27]:
```
index_names = df[ df['Effectiveness'] == 6 ].index
```

In [28]:
```
df.drop(index_names, inplace = True)
```

In [29]:
```
df.Effectiveness.value_counts()
```

Out[29]:
```
5    101860
4     62305
1     47150
3     45180
2     23658
Name: Effectiveness, dtype: int64
```

# Cleaning Text Data from column "Sides" and "Reviews"

Applying a first round of text cleaning techniques

In [30]:
```
def clean_text_round1(text):
    '''Make text lowercase, remove text in square brackets, remove punctuation and remo
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\w*\d\w*', '', text)
    return text

round1 = lambda x: clean_text_round1(x)
```

In [31]:
```
df['Sides'] = pd.DataFrame(df['Sides'].apply(round1))
df['Reviews'] = pd.DataFrame(df['Reviews'].apply(round1))
```

Applying a second round of text cleaning techniques

In [32]:
```
def clean_text_round2(text):
    '''Get rid of some additional punctuation and non-sensical text that was missed the
    text = re.sub('[‘’“”…]', '', text)
    text = re.sub('\n', '', text)
    return text

round2 = lambda x: clean_text_round2(x)
```

In [33]:
```
df['Sides'] = pd.DataFrame(df['Sides'].apply(round2))
df['Reviews'] = pd.DataFrame(df['Reviews'].apply(round2))
```

In [34]:
```
stop = stopwords.words('english')
df['Sides'] = df['Sides'].apply(lambda x: ' '.join([word for word in x.split() if word
```

In [35]:
```
result = Counter(" ".join(df['Sides'].values.tolist()).split(" ")).items()
print(result)
```

```
dict_items([('drowsiness', 123616), ('dizziness', 176464), ('dry', 78863), ('mouth', 820
90), ('nosethroat', 1588), ('headache', 102775), ('upset', 48827), ('stomach', 65314),
('constipation', 88824), ('trouble', 54410), ('sleeping', 56767), ('may', 287863), ('occ
ur', 286809), ('nausea', 167326), ('vomiting', 109986), ('bloating', 18662), ('breast',
17630), ('tenderness', 17196), ('swelling', 20883), ('ankles', 14611), ('feet', 14519),
('fluid', 12603), ('retention', 11325), ('weight', 51203), ('change', 22386), ('temporar
y', 2159), ('burning', 7019), ('stinging', 3322), ('dryness', 3645), ('nose', 8153), ('r
unny', 2555), ('sneezing', 327), ('gain', 17574), ('diarrhea', 65416), ('muscle', 7152),
('pain', 30445), ('loss', 86044), ('appetite', 56184), ('way', 4), ('food', 5), ('taste
s', 5), ('tiredness', 62680), ('bad', 6129), ('taste', 18057), ('joint', 4128), ('back',
1562), ('difficulty', 8492), ('concentrating', 2917), ('weakness', 20642), ('blurred', 3
6713), ('vision', 37329), ('pyridoxine', 2), ('usually', 2057), ('side', 1811), ('effect
s', 7165), ('used', 542), ('recommended', 7), ('doses', 298), ('abdominal', 15327), ('cr
amps', 4414), ('increased', 34811), ('saliva', 4980), ('sweating', 47953), ('decreased',
5457), ('pupil', 105), ('size', 531), ('urination', 3038), ('changes', 13470), ('itchin
g', 9360), ('apply', 419), ('medication', 22997), ('last', 1908), ('short', 108), ('tim
e', 260), ('redness', 9801), ('little', 31), ('bit', 31), ('application', 3959), ('sit
e', 12112), ('lightheadedness', 61669), ('heartburn', 14761), ('mild', 4330), ('irritati
on', 7537), ('bruising', 1675), ('injection', 8026), ('burping', 482), ('strange', 947),
('pimple', 126), ('like', 1111), ('bumps', 240), ('flaking', 157), ('treated', 420), ('s
kin', 10403), ('tingling', 5494), ('body', 19082), ('adjusts', 18152), ('cough', 11909),
('hands', 1175), ('also', 9694), ('effect', 97), ('persists', 80), ('worsens', 80), ('te
ll', 4787), ('doctor', 5224), ('pharmacist', 5008), ('promptly', 5006), ('eye', 9518),
('discomfort', 1340), ('stingingburning', 276), ('eyes', 2367), ('minutes', 540), ('pers
ist', 3871), ('worsen', 3437), ('feeling', 2844), ('something', 923), ('experience', 23
5), ('fatigue', 3010), ('notify', 362), ('painburningswellingredness', 25), ('unsteadine
ss', 2364), ('coordination', 18385), ('sexual', 6673), ('interestability', 3623), ('nerv
ousness', 14466), ('thirst', 927), ('drug', 1077), ('slow', 3271), ('heartbeat', 3869),
('unusual', 15542), ('hair', 8459), ('lip', 11), ('sores', 974), ('runnystuffy', 808),
('carbidopa', 15), ('rare', 58), ('due', 128), ('levodopacontaining', 15), ('product', 5
86), ('taken', 31), ('together', 15), ('first', 6405), ('applied', 1856), ('disappear',
1230), ('days', 1528), ('retaining', 832), ('simethicone', 27), ('frequency', 768), ('mi
ldly', 768), ('shaking', 12540), ('fine', 768), ('tremor', 12973), ('irritability', 488
8), ('restlessness', 1616), ('peeling', 774), ('temporarily', 440), ('slight', 193), ('c
ause', 192), ('symptoms', 1447), ('become', 29), ('severe', 29), ('achespains', 26), ('a
nal', 5), ('gas', 8043), ('months', 3411), ('starting', 187), ('especially', 518), ('chi
ldren', 385), ('metallic', 5819), ('bleeding', 3733), ('placement', 1718), ('device', 17
18), ('blisters', 335), ('color', 1163), ('tinglingburning', 2), ('lightening', 9), ('sc
alp', 470), ('including', 124), ('head', 56), ('lice', 56), ('treatment', 4224), ('perme
thrin', 115), ('numbness', 113), ('masklike', 282), ('facial', 501), ('expression', 28
2), ('inability', 444), ('keep', 311), ('still', 311), ('agitation', 693), ('painirritat
ion', 54), ('hot', 4048), ('flashes', 4022), ('flushing', 9738), ('night', 954), ('sweat
s', 630), ('acne', 10251), ('seborrhea', 11), ('aches', 2311), ('vaginal', 2236), ('disc
```

omfortdryness', 285), ('discharge', 1429), ('reduced', 305), ('interest', 1533), ('burni
ngpainbruising', 593), ('abnormal', 200), ('unpleasant', 793), ('directed', 109), ('us
e', 625), ('remember', 76), ('judged', 76), ('benefit', 76), ('greater', 76), ('risk', 7
6), ('burningstingingirritationredness', 85), ('watering', 150), ('sensitivity', 316),
('light', 337), ('ache', 300), ('fever', 4128), ('tears', 48), ('affected', 60), ('spinn
ing', 1009), ('emollients', 131), ('safely', 131), ('effectively', 131), ('cramping', 21
47), ('rednessdiscomfort', 97), ('eyelid', 545), ('burningstinging', 459), ('eyelash', 2
37), ('numbercolorlengththickness', 107), ('darkening', 200), ('lid', 107), ('crustingdi
scomfort', 107), ('rash', 456), ('reactions', 1544), ('infections', 2), ('earnosethroa
t', 2), ('rumblingpain', 204), ('increase', 367), ('goes', 199), ('away', 392), ('contin
ued', 149), ('problems', 2819), ('discomfortirritationredness', 25), ('tearing', 162),
('sticky', 27), ('eyelashes', 73), ('double', 7867), ('balance', 115), ('walking', 130),
('shakiness', 1746), ('memory', 252), ('painrednessswelling', 1051), ('folic', 14), ('ac
id', 15), ('handslower', 16), ('legsfeet', 48), ('odor', 27), ('fishy', 39), ('smell', 2
6), ('stuffy', 4534), ('handsfeet', 4982), ('coughing', 924), ('sore', 4649), ('throat',
6950), ('aching', 89), ('bones', 89), ('muscles', 89), ('infrequently', 6), ('bone', 67
8), ('hoarseness', 1344), ('bleeds', 244), ('chills', 1300), ('shortness', 662), ('breat
h', 673), ('infusion', 22), ('fast', 436), ('rednessdiscomfortpainswelling', 6), ('bitte
rsourunusual', 60), ('loose', 2400), ('stools', 2467), ('painredness', 163), ('irritatio
nstingingburning', 59), ('beginning', 38), ('weeks', 976), ('subside', 68), ('bitter', 2
21), ('inside', 118), ('fits', 118), ('red', 119), ('sting', 194), ('burn', 194), ('minu
te', 146), ('two', 146), ('rednesspainbruising', 33), ('tinglingnumbness', 468), ('are
a', 764), ('urinating', 1948), ('frequent', 814), ('urge', 177), ('urinate', 177), ('moo
d', 554), ('swings', 540), ('gainloss', 379), ('swellingtenderness', 127), ('unwanted',
127), ('growth', 461), ('bleed', 13), ('discomfortirritation', 559), ('feethands', 21),
('confusion', 120), ('feelings', 1800), ('fullness', 622), ('abdomen', 503), ('common',
520), ('movements', 7221), ('excitation', 133), ('dryflushed', 190), ('methoxsalen', 1
8), ('dryingreddeningdarkening', 9), ('along', 297), ('uva', 9), ('staining', 117), ('pa
tients', 18), ('blonde', 18), ('bleached', 18), ('dyed', 18), ('gray', 18), ('watery', 9
96), ('increaseddecreased', 254), ('oily', 449), ('nosebleeds', 832), ('swellingrednes
s', 144), ('nosebleed', 1001), ('nasal', 253), ('drynessirritationscabbing', 6), ('tickl
ing', 3), ('irritationredness', 151), ('rednessirritation', 988), ('menstrual', 2037),
('cycle', 130), ('blood', 128), ('tinged', 128), ('mucus', 133), ('phlegm', 128), ('brie
f', 263), ('right', 290), ('tastesmell', 92), ('sensations', 121), ('tinglingnumbnesspri
cklingheat', 1143), ('stingingredness', 43), ('widened', 205), ('pupils', 205), ('bladde
r', 41), ('control', 235), ('anxiety', 2420), ('injected', 145), ('high', 141), ('exagge
rated', 128), ('sense', 273), ('wellbeing', 130), ('euphoria', 22), ('voice', 376), ('pe
elingburningdryreddened', 14), ('start', 338), ('substitutes', 13), ('sorespain', 1),
('nail', 63), ('spasms', 174), ('get', 1652), ('worse', 1604), ('texture', 67), ('discol
oration', 263), ('burningtingling', 21), ('tongue', 141), ('well', 17), ('tolerated', 1
7), ('ear', 256), ('nilutamide', 4), ('medications', 23), ('one', 272), ('medicine', 6),
('combination', 10), ('tastedryness', 1), ('leg', 805), ('tooth', 687), ('hiccups', 85),
('irritationblisteringtingling', 4), ('bloody', 82), ('teeth', 51), ('jaw', 56), ('fac
e', 482), ('neck', 336), ('warmth', 949), ('within', 798), ('hours', 1176), ('taking', 7
17), ('niacinamide', 9), ('generally', 9), ('usual', 9), ('dandruff', 18), ('although',
125), ('uncommon', 135), ('vitamin', 153), ('normal', 126), ('rednesswarmthbruising', 92
1), ('period', 214), ('bowel', 1430), ('function', 187), ('often', 191), ('unabsorbed',
187), ('fat', 187), ('fattyoily', 187), ('stool', 192), ('spotting', 229), ('intestina
l', 187), ('needing', 187), ('movement', 188), ('number', 191), ('poor', 198), ('cold',
202), ('headnose', 173), ('irritationpain', 457), ('irritationnumbnesstingling', 38),
('using', 709), ('drooling', 1123), ('rednesspainburning', 21), ('given', 109), ('condit
ion', 109), ('close', 109), ('folliculitis', 52), ('rarely', 283), ('aftertaste', 23),
('made', 23), ('fish', 23), ('oil', 36), ('people', 84), ('serious', 4), ('burningstingi
ngrednessdryness', 131), ('irritationsoreness', 50), ('rednesspainswelling', 26), ('eart
hroat', 8), ('drynessirritation', 241), ('looseoily', 39), ('fastpounding', 20), ('irrit
ationdiscomfort', 65), ('earache', 10), ('penis', 84), ('testiclegroin', 42), ('minor',
1034), ('veins', 43), ('rapid', 53), ('drynessrednessitchiness', 103), ('worsening', 50
2), ('allergy', 13), ('expected', 561), ('commonly', 190), ('swellingpain', 48), ('vagin
alurethral', 503), ('lower', 363), ('lasts', 76), ('irritated', 125), ('dryirritated', 1
89), ('lips', 437), ('eyelids', 231), ('crusty', 190), ('thinning', 568), ('reports', 10
9), ('however', 109), ('upsetpain', 3246), ('swellingrednessirritation', 5), ('mineral',
13), ('leak', 13), ('rectum', 43), ('difficulties', 21), ('sensitive', 1), ('flu', 101
0), ('rectal', 202), ('inserting', 80), ('suppository', 57), ('bottle', 23), ('tip', 2
3), ('excitement', 23), ('k', 3), ('soreness', 311), ('rednesspain', 27), ('spottingblee
ding', 7), ('abilitydesire', 956), ('decrease', 975), ('rednessburning', 128), ('headach

es', 8), ('falling', 96), ('asleep', 96), ('shakynervous', 96), ('discomfortpain', 35), ('goose', 61), ('urinary', 61), ('strongfrequent', 61), ('see', 26), ('section', 26), ('crusting', 125), ('yawning', 11560), ('dreams', 2001), ('metallicsalty', 121), ('stiffness', 14), ('irritationburningstinging', 15), ('numbtingling', 392), ('numbnesstingling', 295), ('around', 231), ('sex', 3172), ('painswelling', 237), ('bodyfacial', 4), ('swallowing', 1), ('excessive', 1311), ('since', 3), ('darunavir', 6), ('always', 3), ('hiv', 3), ('difficult', 3), ('whether', 3), ('causing', 3), ('certain', 3), ('gum', 72), ('eg', 833), ('hard', 117), ('lump', 120), ('drink', 149), ('bothersome', 42), ('include', 494), ('easy', 1661), ('cuts', 802), ('soredry', 8), ('irritationburning', 8), ('scabies', 59), ('reddened', 55), ('tartar', 59), ('fooddrinks', 59), ('belching', 30), ('shapecolor', 32), ('better', 48), ('lack', 302), ('energy', 16), ('flakingpeeling', 51), ('dim', 9), ('brow', 9), ('twitching', 2), ('top', 28), ('layer', 28), ('scaling', 462), ('bleedingbruising', 857), ('backbone', 7), ('stingingburningrednessitchiness', 39), ('small', 281), ('altered', 33), ('numbnesstinglingredness', 24), ('tiredweak', 24), ('armslegs', 396), ('numbnesstinglingburning', 27), ('daytime', 212), ('calcitriol', 30), ('several', 67), ('ejaculation', 834), ('tremors', 172), ('rednessswellingpain', 54), ('urine', 12), ('turn', 12), ('bright', 15), ('yellow', 12), ('low', 171), ('tastehearing', 142), ('chest', 16), ('swollen', 36), ('edema', 38), ('sensation', 288), ('immediately', 166), ('applying', 240), ('discomfortredness', 301), ('corneal', 7), ('dermatitis', 6), ('rosacea', 6), ('pimples', 6), ('painsores', 30), ('dose', 67), ('ringing', 3637), ('ears', 3637), ('hearing', 83), ('foods', 442), ('drycrackedthickened', 56), ('legs', 135), ('move', 133), ('talk', 133), ('frequently', 235), ('amounts', 194), ('implant', 22), ('palms', 90), ('soles', 67), ('gums', 9), ('bitterstrangechange', 9), ('spray', 35), ('swellingflaking', 17), ('lessen', 57), ('stop', 3), ('stingingburningredness', 130), ('hardening', 48), ('increasing', 29), ('slowly', 36), ('itchy', 42), ('lumps', 4), ('shortterm', 117), ('take', 117), ('delayedirregularabsent', 82), ('periods', 1307), ('flushingwarmth', 192), ('rednessdischarge', 22), ('flushes', 26), ('coldlike', 26), ('irritationpaindrynessredness', 8), ('paintenderness', 446), ('thickeninghardening', 446), ('peelingflakingscabbingcrusting', 446), ('leaking', 446), ('clear', 446), ('abdominalpelvic', 63), ('gets', 176), ('near', 30), ('sleepiness', 5), ('waking', 33), ('drycracked', 55), ('flaky', 39), ('mentalmood', 22), ('burningstingingirritation', 24), ('jawmusclebone', 7), ('shrinking', 12), ('testicles', 12), ('tenderswelling', 12), ('result', 12), ('lowered', 12), ('testosterone', 12), ('levels', 12), ('ability', 3047), ('numberlengththickness', 41), ('achepain', 1655), ('occurs', 134), ('hydrocodone', 127), ('sit', 4), ('lie', 4), ('painrednesssoreness', 4), ('large', 16), ('unwanteduncontrolled', 22), ('speaking', 2), ('black', 42), ('hairy', 42), ('desire', 325), ('rednessswelling', 325), ('tinglingburningnumbnessprickling', 24), ('painrednesswarmthbruisingswelling', 1), ('upper', 1), ('man', 1), ('syndrome', 1), ('drunk', 42), ('drynessrednessirritation', 89), ('baloxavir', 14), ('marboxil', 14), ('sometimes', 7), ('lessened', 7), ('giving', 7), ('production', 4709), ('drive', 2195), ('dryflaky', 16), ('arms', 2), ('prescribed', 9), ('child', 5), ('excess', 136), ('fewer', 189), ('corticosteroids', 189), ('budesonide', 189), ('works', 189), ('gut', 189), ('absorbed', 189), ('spasm', 7), ('drowsinesstiredness', 26), ('e', 3), ('burningpainswellingreddening', 30), ('ask', 2), ('ways', 2), ('anti', 2), ('abilityinterest', 1), ('tinglingnumbnessprickling', 58), ('amount', 225), ('semen', 225), ('sperm', 225), ('testicle', 225), ('sunscreen', 1), ('products', 1), ('containing', 1), ('aminobenzoic', 1), ('paraaminobenzoic', 1), ('acidpaba', 1), ('stain', 1), ('clothing', 1), ('burningstingingdiscomfort', 5), ('corner', 5), ('insomnia', 120), ('breathing', 29), ('started', 10), ('follicle', 65), ('painswellingredness', 217), ('limited', 1), ('arm', 2), ('metallicgarliclike', 128), ('sleep', 47), ('patterns', 43), ('patch', 53), ('week', 268), ('burningstingingredness', 13), ('discomfortpainsoreness', 20), ('toothache', 20), ('irritationrednesspain', 6), ('fainting', 37), ('indigestion', 5), ('canal', 6), ('dropped', 12), ('drypeelingoily', 2), ('changesloss', 3), ('stomachache', 26), ('backache', 26), ('feverchills', 26), ('pale', 19), ('soft', 28), ('sorenessrednessswellingbruising', 4), ('painful', 2), ('male', 1), ('tendernessenlargement', 1), ('mildmoderate', 96), ('tendernesspain', 9), ('unlikely', 2), ('persistent', 2), ('bruisingbleeding', 2), ('painrednessirritation', 17), ('hardened', 1), ('flutamide', 6), ('caused', 6), ('lethargy', 28), ('tape', 3), ('redsore', 4), ('lipsthroat', 1), ('focusing', 180), ('seeing', 180), ('spots', 180), ('hypotension', 32), ('poundingfast', 100), ('nightmares', 2), ('painsorenessrednessswelling', 1), ('disturbances', 4), ('musclebone', 1), ('painirritationbruising', 16), ('shivering', 19), ('ingrown', 152), ('toenail', 152), ('numb', 1), ('irritationcongestion', 31), ('fingernails', 7), ('toenails', 7), ('stingingdiscomfort', 4), ('waterydryitchyred', 4), ('swellingtendernesspain', 57), ('difficultpainfulfrequentbloody', 1), ('nails', 14), ('painburningredness', 5), ('impotence', 558), ('spotsfreckles', 23), ('handssoles', 23), ('burningpainful', 157), ('vein', 12), ('walkingclumsiness', 2), ('oiliness', 1), ('pitch', 20), ('women', 20), ('pain

```
bruisingrednessswelling', 1), ('achespain', 21), ('diaper', 237), ('young', 237), ('pain discomfort', 1), ('calcifediol', 1), ('feeding', 2)])
```

In [36]:
```
sorted(result, key=lambda x: (-x[1], x[0]))
```

Out[36]:
```
[('may', 287863),
 ('occur', 286809),
 ('dizziness', 176464),
 ('nausea', 167326),
 ('drowsiness', 123616),
 ('vomiting', 109986),
 ('headache', 102775),
 ('constipation', 88824),
 ('loss', 86044),
 ('mouth', 82090),
 ('dry', 78863),
 ('diarrhea', 65416),
 ('stomach', 65314),
 ('tiredness', 62680),
 ('lightheadedness', 61669),
 ('sleeping', 56767),
 ('appetite', 56184),
 ('trouble', 54410),
 ('weight', 51203),
 ('upset', 48827),
 ('sweating', 47953),
 ('vision', 37329),
 ('blurred', 36713),
 ('increased', 34811),
 ('pain', 30445),
 ('medication', 22997),
 ('change', 22386),
 ('swelling', 20883),
 ('weakness', 20642),
 ('body', 19082),
 ('bloating', 18662),
 ('coordination', 18385),
 ('adjusts', 18152),
 ('taste', 18057),
 ('breast', 17630),
 ('gain', 17574),
 ('tenderness', 17196),
 ('unusual', 15542),
 ('abdominal', 15327),
 ('heartburn', 14761),
 ('ankles', 14611),
 ('feet', 14519),
 ('nervousness', 14466),
 ('changes', 13470),
 ('tremor', 12973),
 ('fluid', 12603),
 ('shaking', 12540),
 ('site', 12112),
 ('cough', 11909),
 ('yawning', 11560),
 ('retention', 11325),
 ('skin', 10403),
 ('acne', 10251),
 ('redness', 9801),
 ('flushing', 9738),
 ('also', 9694),
 ('eye', 9518),
 ('itching', 9360),
 ('difficulty', 8492),
 ('hair', 8459),
```

```
('nose', 8153),
('gas', 8043),
('injection', 8026),
('double', 7867),
('irritation', 7537),
('movements', 7221),
('effects', 7165),
('muscle', 7152),
('burning', 7019),
('throat', 6950),
('sexual', 6673),
('first', 6405),
('bad', 6129),
('metallic', 5819),
('tingling', 5494),
('decreased', 5457),
('doctor', 5224),
('pharmacist', 5008),
('promptly', 5006),
('handsfeet', 4982),
('saliva', 4980),
('irritability', 4888),
('tell', 4787),
('production', 4709),
('sore', 4649),
('stuffy', 4534),
('cramps', 4414),
('mild', 4330),
('treatment', 4224),
('fever', 4128),
('joint', 4128),
('hot', 4048),
('flashes', 4022),
('application', 3959),
('persist', 3871),
('heartbeat', 3869),
('bleeding', 3733),
('dryness', 3645),
('ears', 3637),
('ringing', 3637),
('interestability', 3623),
('worsen', 3437),
('months', 3411),
('stinging', 3322),
('slow', 3271),
('upsetpain', 3246),
('sex', 3172),
('ability', 3047),
('urination', 3038),
('fatigue', 3010),
('concentrating', 2917),
('feeling', 2844),
('problems', 2819),
('runny', 2555),
('stools', 2467),
('anxiety', 2420),
('loose', 2400),
('eyes', 2367),
('unsteadiness', 2364),
('aches', 2311),
('vaginal', 2236),
('drive', 2195),
('temporary', 2159),
('cramping', 2147),
('usually', 2057),
```

```
('menstrual', 2037),
('dreams', 2001),
('urinating', 1948),
('last', 1908),
('applied', 1856),
('side', 1811),
('feelings', 1800),
('shakiness', 1746),
('device', 1718),
('placement', 1718),
('bruising', 1675),
('easy', 1661),
('achepain', 1655),
('get', 1652),
('restlessness', 1616),
('worse', 1604),
('nosethroat', 1588),
('back', 1562),
('reactions', 1544),
('interest', 1533),
('days', 1528),
('symptoms', 1447),
('bowel', 1430),
('discharge', 1429),
('hoarseness', 1344),
('discomfort', 1340),
('excessive', 1311),
('periods', 1307),
('chills', 1300),
('disappear', 1230),
('hours', 1176),
('hands', 1175),
('color', 1163),
('tinglingnumbnesspricklingheat', 1143),
('drooling', 1123),
('like', 1111),
('drug', 1077),
('painrednessswelling', 1051),
('minor', 1034),
('flu', 1010),
('spinning', 1009),
('nosebleed', 1001),
('watery', 996),
('rednessirritation', 988),
('weeks', 976),
('decrease', 975),
('sores', 974),
('abilitydesire', 956),
('night', 954),
('warmth', 949),
('strange', 947),
('thirst', 927),
('coughing', 924),
('something', 923),
('rednesswarmthbruising', 921),
('bleedingbruising', 857),
('ejaculation', 834),
('eg', 833),
('nosebleeds', 832),
('retaining', 832),
('frequent', 814),
('runnystuffy', 808),
('leg', 805),
('cuts', 802),
('within', 798),
```

```
('unpleasant', 793),
('peeling', 774),
('fine', 768),
('frequency', 768),
('mildly', 768),
('area', 764),
('taking', 717),
('using', 709),
('agitation', 693),
('tooth', 687),
('bone', 678),
('breath', 673),
('shortness', 662),
('sweats', 630),
('use', 625),
('fullness', 622),
('burningpainbruising', 593),
('product', 586),
('thinning', 568),
('expected', 561),
('discomfortirritation', 559),
('impotence', 558),
('mood', 554),
('eyelid', 545),
('used', 542),
('minutes', 540),
('swings', 540),
('size', 531),
('common', 520),
('especially', 518),
('abdomen', 503),
('vaginalurethral', 503),
('worsening', 502),
('facial', 501),
('include', 494),
('burping', 482),
('face', 482),
('scalp', 470),
('tinglingnumbness', 468),
('scaling', 462),
('growth', 461),
('burningstinging', 459),
('irritationpain', 457),
('rash', 456),
('oily', 449),
('clear', 446),
('leaking', 446),
('paintenderness', 446),
('peelingflakingscabbingcrusting', 446),
('thickeninghardening', 446),
('inability', 444),
('foods', 442),
('temporarily', 440),
('lips', 437),
('fast', 436),
('treated', 420),
('apply', 419),
('armslegs', 396),
('away', 392),
('numbtingling', 392),
('children', 385),
('gainloss', 379),
('voice', 376),
('increase', 367),
('lower', 363),
```

```
('notify', 362),
('start', 338),
('light', 337),
('neck', 336),
('blisters', 335),
('sneezing', 327),
('desire', 325),
('rednessswelling', 325),
('sensitivity', 316),
('keep', 311),
('soreness', 311),
('still', 311),
('reduced', 305),
('lack', 302),
('discomfortredness', 301),
('ache', 300),
('doses', 298),
('along', 297),
('numbnesstingling', 295),
('right', 290),
('sensation', 288),
('discomfortdryness', 285),
('rarely', 283),
('expression', 282),
('masklike', 282),
('small', 281),
('stingingburning', 276),
('sense', 273),
('one', 272),
('week', 268),
('brief', 263),
('discoloration', 263),
('time', 260),
('ear', 256),
('increaseddecreased', 254),
('nasal', 253),
('memory', 252),
('bleeds', 244),
('drynessirritation', 241),
('applying', 240),
('bumps', 240),
('diaper', 237),
('eyelash', 237),
('painswelling', 237),
('young', 237),
('control', 235),
('experience', 235),
('frequently', 235),
('around', 231),
('eyelids', 231),
('spotting', 229),
('amount', 225),
('semen', 225),
('sperm', 225),
('testicle', 225),
('bitter', 221),
('painswellingredness', 217),
('period', 214),
('daytime', 212),
('pupils', 205),
('widened', 205),
('rumblingpain', 204),
('cold', 202),
('rectal', 202),
('abnormal', 200),
```

```
('darkening', 200),
('goes', 199),
('poor', 198),
('amounts', 194),
('burn', 194),
('sting', 194),
('slight', 193),
('cause', 192),
('flushingwarmth', 192),
('stool', 192),
('number', 191),
('often', 191),
('commonly', 190),
('crusty', 190),
('dryflushed', 190),
('absorbed', 189),
('budesonide', 189),
('corticosteroids', 189),
('dryirritated', 189),
('fewer', 189),
('gut', 189),
('works', 189),
('movement', 188),
('fat', 187),
('fattyoily', 187),
('function', 187),
('intestinal', 187),
('needing', 187),
('starting', 187),
('unabsorbed', 187),
('focusing', 180),
('seeing', 180),
('spots', 180),
('urge', 177),
('urinate', 177),
('gets', 176),
('spasms', 174),
('headnose', 173),
('tremors', 172),
('low', 171),
('immediately', 166),
('painredness', 163),
('tearing', 162),
('burningpainful', 157),
('flaking', 157),
('vitamin', 153),
('ingrown', 152),
('toenail', 152),
('irritationredness', 151),
('watering', 150),
('continued', 149),
('drink', 149),
('minute', 146),
('two', 146),
('injected', 145),
('swellingredness', 144),
('tastehearing', 142),
('high', 141),
('tongue', 141),
('excess', 136),
('legs', 135),
('uncommon', 135),
('occurs', 134),
('excitation', 133),
('move', 133),
```

```
('mucus', 133),
('talk', 133),
('burningstingingrednessdryness', 131),
('effectively', 131),
('emollients', 131),
('safely', 131),
('cycle', 130),
('stingingburningredness', 130),
('walking', 130),
('wellbeing', 130),
('blood', 128),
('due', 128),
('exaggerated', 128),
('metallicgarliclike', 128),
('phlegm', 128),
('rednessburning', 128),
('tinged', 128),
('hydrocodone', 127),
('swellingtenderness', 127),
('unwanted', 127),
('normal', 126),
('pimple', 126),
('although', 125),
('crusting', 125),
('irritated', 125),
('including', 124),
('metallicsalty', 121),
('sensations', 121),
('confusion', 120),
('insomnia', 120),
('lump', 120),
('red', 119),
('fits', 118),
('inside', 118),
('hard', 117),
('shortterm', 117),
('staining', 117),
('take', 117),
('balance', 115),
('permethrin', 115),
('numbness', 113),
('close', 109),
('condition', 109),
('directed', 109),
('given', 109),
('however', 109),
('reports', 109),
('short', 108),
('crustingdiscomfort', 107),
('lid', 107),
('numbercolorlengththickness', 107),
('pupil', 105),
('drynessrednessitchiness', 103),
('poundingfast', 100),
('effect', 97),
('rednessdiscomfort', 97),
('asleep', 96),
('falling', 96),
('mildmoderate', 96),
('shakynervous', 96),
('tastesmell', 92),
('palms', 90),
('aching', 89),
('bones', 89),
('drynessrednessirritation', 89),
```

```
('muscles', 89),
('burningstingingirritationredness', 85),
('hiccups', 85),
('penis', 84),
('people', 84),
('hearing', 83),
('bloody', 82),
('delayedirregularabsent', 82),
('inserting', 80),
('persists', 80),
('worsens', 80),
('benefit', 76),
('greater', 76),
('judged', 76),
('lasts', 76),
('remember', 76),
('risk', 76),
('eyelashes', 73),
('gum', 72),
('subside', 68),
('dose', 67),
('several', 67),
('soles', 67),
('texture', 67),
('follicle', 65),
('irritationdiscomfort', 65),
('abdominalpelvic', 63),
('nail', 63),
('goose', 61),
('strongfrequent', 61),
('urinary', 61),
('affected', 60),
('bittersourunusual', 60),
('fooddrinks', 59),
('irritationstingingburning', 59),
('scabies', 59),
('tartar', 59),
('rare', 58),
('tinglingnumbnessprickling', 58),
('lessen', 57),
('suppository', 57),
('swellingtendernesspain', 57),
('drycrackedthickened', 56),
('head', 56),
('jaw', 56),
('lice', 56),
('drycracked', 55),
('reddened', 55),
('painirritation', 54),
('rednessswellingpain', 54),
('patch', 53),
('rapid', 53),
('folliculitis', 52),
('flakingpeeling', 51),
('teeth', 51),
('irritationsoreness', 50),
('better', 48),
('hardening', 48),
('legsfeet', 48),
('swellingpain', 48),
('tears', 48),
('sleep', 47),
('patterns', 43),
('rectum', 43),
('stingingredness', 43),
```

```
('veins', 43),
('black', 42),
('bothersome', 42),
('drunk', 42),
('hairy', 42),
('itchy', 42),
('testiclegroin', 42),
('bladder', 41),
('numberlengththickness', 41),
('fishy', 39),
('flaky', 39),
('looseoily', 39),
('stingingburningrednessitchiness', 39),
('beginning', 38),
('edema', 38),
('irritationnumbnesstingling', 38),
('fainting', 37),
('oil', 36),
('slowly', 36),
('swollen', 36),
('discomfortpain', 35),
('spray', 35),
('altered', 33),
('rednesspainbruising', 33),
('waking', 33),
('hypotension', 32),
('shapecolor', 32),
('bit', 31),
('irritationcongestion', 31),
('little', 31),
('taken', 31),
('belching', 30),
('burningpainswellingreddening', 30),
('calcitriol', 30),
('near', 30),
('painsores', 30),
('become', 29),
('breathing', 29),
('increasing', 29),
('severe', 29),
('layer', 28),
('lethargy', 28),
('soft', 28),
('top', 28),
('numbnesstinglingburning', 27),
('odor', 27),
('rednesspain', 27),
('simethicone', 27),
('sticky', 27),
('achespains', 26),
('backache', 26),
('coldlike', 26),
('drowsinesstiredness', 26),
('feverchills', 26),
('flushes', 26),
('rednesspainswelling', 26),
('section', 26),
('see', 26),
('smell', 26),
('stomachache', 26),
('discomfortirritationredness', 25),
('painburningswellingredness', 25),
('burningstingingirritation', 24),
('numbnesstinglingredness', 24),
('tinglingburningnumbnessprickling', 24),
```

```
('tiredweak', 24),
('aftertaste', 23),
('bottle', 23),
('excitement', 23),
('fish', 23),
('handssoles', 23),
('made', 23),
('medications', 23),
('spotsfreckles', 23),
('tip', 23),
('euphoria', 22),
('implant', 22),
('infusion', 22),
('mentalmood', 22),
('rednessdischarge', 22),
('unwanteduncontrolled', 22),
('achespain', 21),
('burningtingling', 21),
('difficulties', 21),
('feethands', 21),
('rednesspainburning', 21),
('discomfortpainsoreness', 20),
('fastpounding', 20),
('pitch', 20),
('toothache', 20),
('women', 20),
('pale', 19),
('shivering', 19),
('bleached', 18),
('blonde', 18),
('dandruff', 18),
('dyed', 18),
('gray', 18),
('methoxsalen', 18),
('patients', 18),
('painrednessirritation', 17),
('swellingflaking', 17),
('tolerated', 17),
('well', 17),
('chest', 16),
('dryflaky', 16),
('energy', 16),
('handslower', 16),
('large', 16),
('painirritationbruising', 16),
('acid', 15),
('bright', 15),
('carbidopa', 15),
('irritationburningstinging', 15),
('levodopacontaining', 15),
('together', 15),
('baloxavir', 14),
('folic', 14),
('marboxil', 14),
('nails', 14),
('peelingburningdryreddened', 14),
('stiffness', 14),
('allergy', 13),
('bleed', 13),
('burningstingingredness', 13),
('leak', 13),
('mineral', 13),
('substitutes', 13),
('dropped', 12),
('levels', 12),
```

```
('lowered', 12),
('result', 12),
('shrinking', 12),
('tendernessswelling', 12),
('testicles', 12),
('testosterone', 12),
('turn', 12),
('urine', 12),
('vein', 12),
('yellow', 12),
('lip', 11),
('seborrhea', 11),
('combination', 10),
('earache', 10),
('started', 10),
('bitterstrangechange', 9),
('brow', 9),
('dim', 9),
('dryingreddeningdarkening', 9),
('generally', 9),
('gums', 9),
('lightening', 9),
('niacinamide', 9),
('prescribed', 9),
('tendernesspain', 9),
('usual', 9),
('uva', 9),
('earthroat', 8),
('headaches', 8),
('irritationburning', 8),
('irritationpaindrynessredness', 8),
('soredry', 8),
('backbone', 7),
('corneal', 7),
('fingernails', 7),
('giving', 7),
('jawmusclebone', 7),
('lessened', 7),
('recommended', 7),
('sometimes', 7),
('spasm', 7),
('spottingbleeding', 7),
('toenails', 7),
('canal', 6),
('caused', 6),
('darunavir', 6),
('dermatitis', 6),
('drynessirritationscabbing', 6),
('flutamide', 6),
('infrequently', 6),
('irritationrednesspain', 6),
('medicine', 6),
('pimples', 6),
('rednessdiscomfortpainswelling', 6),
('rosacea', 6),
('anal', 5),
('burningstingingdiscomfort', 5),
('child', 5),
('corner', 5),
('food', 5),
('indigestion', 5),
('painburningredness', 5),
('sleepiness', 5),
('swellingrednessirritation', 5),
('tastes', 5),
```

```
('bodyfacial', 4),
('disturbances', 4),
('irritationblisteringtingling', 4),
('lie', 4),
('lumps', 4),
('nilutamide', 4),
('painrednesssoreness', 4),
('redsore', 4),
('serious', 4),
('sit', 4),
('sorenessrednessswellingbruising', 4),
('stingingdiscomfort', 4),
('waterydryitchyred', 4),
('way', 4),
('always', 3),
('causing', 3),
('certain', 3),
('changesloss', 3),
('difficult', 3),
('e', 3),
('hiv', 3),
('k', 3),
('since', 3),
('stop', 3),
('tape', 3),
('tickling', 3),
('whether', 3),
('anti', 2),
('arm', 2),
('arms', 2),
('ask', 2),
('bruisingbleeding', 2),
('drypeelingoily', 2),
('earnosethroat', 2),
('feeding', 2),
('infections', 2),
('nightmares', 2),
('painful', 2),
('persistent', 2),
('pyridoxine', 2),
('speaking', 2),
('tinglingburning', 2),
('twitching', 2),
('unlikely', 2),
('walkingclumsiness', 2),
('ways', 2),
('abilityinterest', 1),
('acidpaba', 1),
('aminobenzoic', 1),
('calcifediol', 1),
('clothing', 1),
('containing', 1),
('difficultpainfulfrequentbloody', 1),
('hardened', 1),
('limited', 1),
('lipsthroat', 1),
('male', 1),
('man', 1),
('musclebone', 1),
('numb', 1),
('oiliness', 1),
('painbruisingrednessswelling', 1),
('paindiscomfort', 1),
('painrednesswarmthbruisingswelling', 1),
('painsorenessrednessswelling', 1),
```

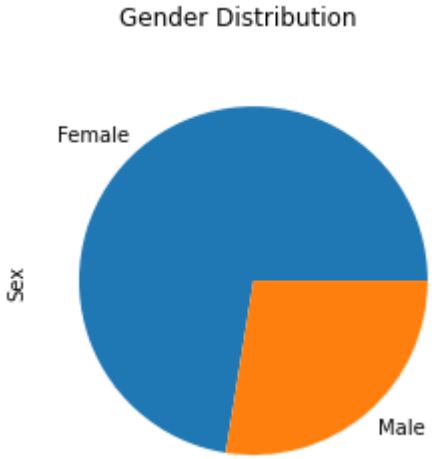
```
('paraaminobenzoic', 1),
('products', 1),
('sensitive', 1),
('sorespain', 1),
('stain', 1),
('sunscreen', 1),
('swallowing', 1),
('syndrome', 1),
('tastedryness', 1),
('tendernessenlargement', 1),
('upper', 1)]
```

---

# Exploratory Data Analysis

## Visual Representation of Age Distribution

In [37]:
```
data1 = df['Sex'].tolist()
male=data1.count('Male')
female=data1.count('Female')
ax = df.Sex.value_counts().plot.pie(subplots = True,figsize = (8,4), title = "Gender Di
```

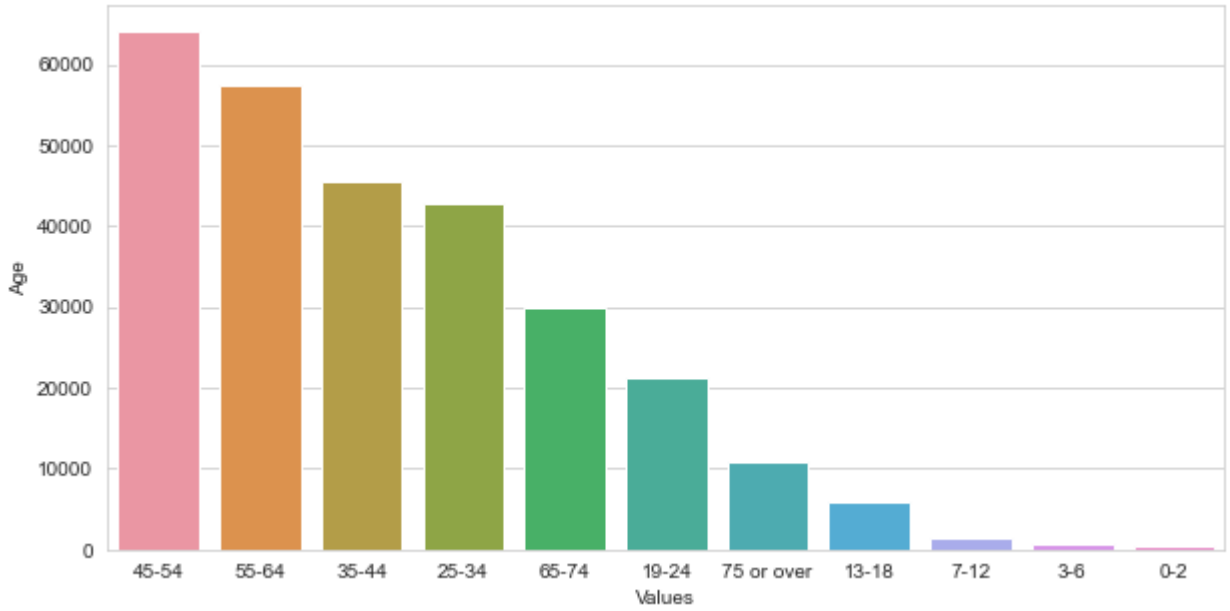



## Visual Representation of Gender Distribution

In [38]:
```
ageColourful = df['Age'].value_counts().reset_index()
ageColourful.columns = ["Values", "Age"]
ageColourful

# set style
sns.set_style("whitegrid");

plt.figure(figsize = (10,5));
sns.barplot(x = 'Values',y ='Age', data = ageColourful);
plt.show();
```

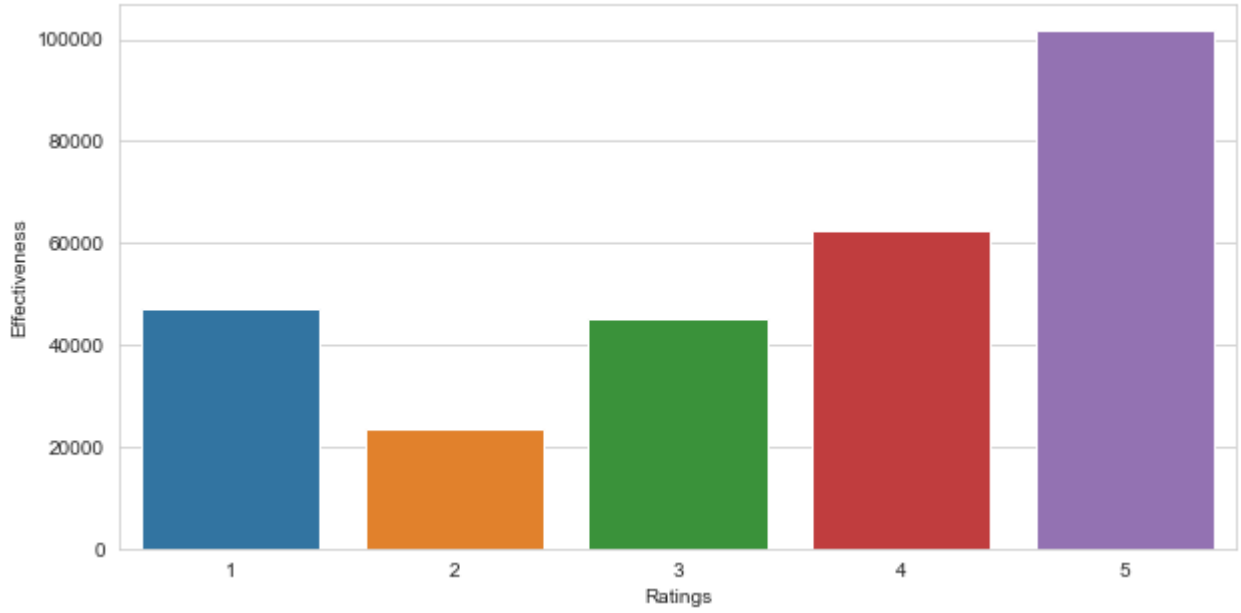



# Visual Representation of Effectiveness Rating

In [39]:
```
effectivenessRatingColourful = df['Effectiveness'].value_counts().reset_index()
effectivenessRatingColourful.columns = ["Ratings", "Effectiveness"]
effectivenessRatingColourful

# set style
sns.set_style("whitegrid");

plt.figure(figsize=(10,5));
sns.barplot(x = 'Ratings',y = 'Effectiveness', data = effectivenessRatingColourful);
plt.show();
```

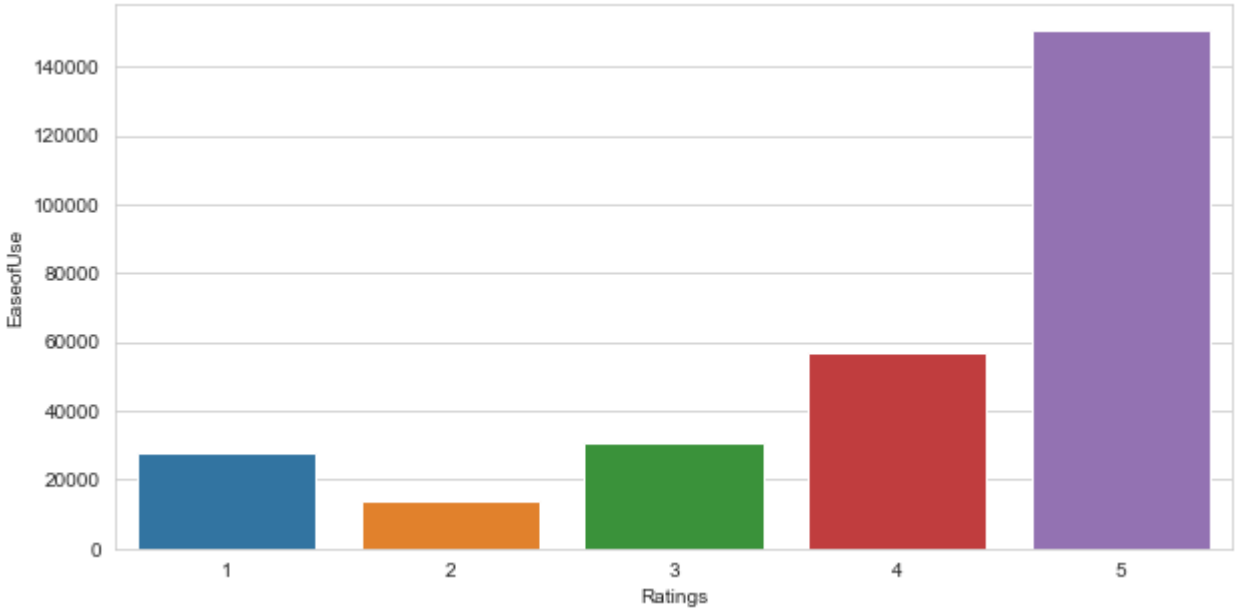



# Visual Representation of EaseofUse Rating

In [40]:
```
easeOfUseRatingColourful = df['EaseofUse'].value_counts().reset_index()
easeOfUseRatingColourful.columns = ["Ratings", "EaseofUse"]
```

```
easeOfUseRatingColourful

# set style
sns.set_style("whitegrid");
plt.figure(figsize = (10,5));
sns.barplot(x = 'Ratings',y = "EaseofUse", data = easeOfUseRatingColourful);
plt.show();
```




---

# Building Classification Model

# One Hot Encoding

In [41]:
```
le_age=LabelEncoder()
le_DrugId=LabelEncoder()
le_Sex=LabelEncoder()
le_Condition=LabelEncoder()
```

In [42]:
```
df['age_n']=le_age.fit_transform(df['Age_Group_number'])
df['DrugId_n']=le_DrugId.fit_transform(df['DrugId'])
df['Sex_n']=le_Sex.fit_transform(df['Sex'])
df['condition_n']=le_Condition.fit_transform(df['Condition'])
```

# Filter "Side Effect"

We are Gonna take 5 Random Side Effects for this model

1.Dizziness
2.Nausea
3.Drowsiness
4.Vomitting
5.Headache

In [43]:
```
SideEffect_Name=widgets.RadioButtons(
            options=['dizziness', 'nausea', 'drowsiness','vomiting','headache'],
            layout={'width': 'max-content'})
```

Now we are gonna create a list of Side Effects from the Data set

In [44]:
```
display(SideEffect_Name)
```

In [45]:
```
SideEffect_Name.value
```

Out[45]:
```
'dizziness'
```

Filter Only that Side Effect

In [46]:
```
df=df[df['Sides'].str.contains(SideEffect_Name.value)]
```

**Consider only useful reviews**

**It will filter and display only those values which have drug name as "Selected Drug name as per the user"**

In [47]:
```
df=df[df['UsefulCount']>1]
```

In [48]:
```
df.head()
```

Out[48]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **122** | 65-74 | Cancer of Ovary in Patient with Germline BRCA ... | 12/8/2016 | lynparza tablet | 167493 | 5 | 5 | low hgb for the first months but not as tired... | 5 | Fe |
| **123** | 45-54 | Cancer of Ovary in Patient with Germline BRCA ... | 7/13/2016 | lynparza tablet | 167493 | 5 | 5 | prescribed lynparza after my firstline ivip ch... | 5 | Fe |
| **124** | 55-64 | Cancer of Ovary in Patient with Germline BRCA ... | 6/21/2016 | lynparza tablet | 167493 | 5 | 5 | i take six pills twice a day and no food one h... | 5 | Fe |
| **125** | 45-54 | Cancer of Ovary in Patient with Germline BRCA ... | 5/23/2016 | lynparza tablet | 167493 | 5 | 5 | only issue i am having is sun exposure rash on... | 4 | Fe |

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **178** | 65-74 | Cancer of Ovary in Patient with Germline BRCA ... | 12/8/2016 | lynparza | 167493 | 5 | 5 | low hgb for the first months but not as tired... | 5 | Fe |

# Filter "Drug Name"

In [49]:
```
DrugList=df.Drug.value_counts().head(15)
```

In [50]:
```
DrugList=DrugList.to_dict()
print(DrugList)
```

```
{'lisinopril solution': 2888, 'lisinopril': 2888, 'lexapro': 2710, 'effexor xr': 2326,
'lyrica': 2095, 'tramadol hcl': 1750, 'tramadol hcl er': 1750, 'zoloft': 1743, 'topamax
capsule, sprinkle': 1541, 'topamax': 1541, 'hydrocodone-acetaminophen': 1525, 'phentermi
ne hcl': 1509, 'neurontin': 1460, 'neurontin capsule': 1460, 'trazodone hcl': 1421}
```

In [51]:
```
Drug_Name=widgets.Dropdown( options=DrugList.keys(),
labels='Enter the name of the Drug: \t',
disabled=False,
)
```

In [52]:
```
display(Drug_Name)
```

In [53]:
```
Drug_Name.value
```

Out[53]:
```
'lisinopril solution'
```

In [54]:
```
df = df[df['Drug'].isin([Drug_Name.value])]
```

It will filter and display only those values which have drug name as "Selected Drug name as per the user"

In [55]:
```
df.head()
```

Out[55]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction |
|---|---|---|---|---|---|---|---|---|---|
| **9203** | 75 or over | High Blood Pressure | 2/8/2020 | lisinopril solution | 6873 | 5 | 4 | i have developed a very itch rash on my body f... | 4 |
| **9209** | 55-64 | High Blood Pressure | 1/14/2020 | lisinopril solution | 6873 | 4 | 5 | tried alternative medicine beet root hibiscus ... | 4 |

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction |
|---|---|---|---|---|---|---|---|---|---|
| **9215** | 65-74 | High Blood Pressure | 10/17/2019 | lisinopril solution | 6873 | 5 | 5 | have been taking mg doses per day for over y... | 5 |
| **9221** | 55-64 | High Blood Pressure | 10/13/2019 | lisinopril solution | 6873 | 2 | 4 | i dont like that im taking this and not knowin... | 2 |
| **9227** | 55-64 | High Blood Pressure | 8/16/2019 | lisinopril solution | 6873 | 3 | 1 | was on this medicine for about months and hav... | 1 |

# Filter "Gender"

Select either in order to filter the gender

1. Male

   or
2. Female

In [56]:
```
gender_opt=widgets.RadioButtons(
            options=['Male', 'Female'],
            layout={'width': 'max-content'})
```

In [58]:
```
display(gender_opt)
```

In [59]:
```
gender_opt.value
```

Out[59]:
```
'Female'
```

In [60]:
```
df=df[df['Sex'].str.contains(gender_opt.value)]
```

In [61]:
```
df.head()
```

Out[61]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **9203** | 75 or over | High Blood Pressure | 2/8/2020 | lisinopril solution | 6873 | 5 | 4 | i have developed a very itch rash on my body f... | 4 | F |

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **9227** | 55-64 | High Blood Pressure | 8/16/2019 | lisinopril solution | 6873 | 3 | 1 | was on this medicine for about months and hav... | 1 | F |
| **9233** | 25-34 | High Blood Pressure | 6/3/2019 | lisinopril solution | 6873 | 5 | 2 | i was placed on a low dose of lisinopril after... | 1 | F |
| **9239** | 35-44 | Other | 6/2/2019 | lisinopril solution | 6873 | 5 | 5 | no script or health insurance needed to place ... | 5 | F |
| **9251** | 65-74 | High Blood Pressure | 5/6/2019 | lisinopril solution | 6873 | 3 | 2 | i just started taking it | 2 | F |

# Converting data type of "Reviews" column

In [62]:
```
df['Reviews'] = df['Reviews'].astype(str)
```

In [63]:
```
df.head(1)
```

Out[63]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **9203** | 75 or over | High Blood Pressure | 2/8/2020 | lisinopril solution | 6873 | 5 | 4 | i have developed a very itch rash on my body f... | 4 | Fe |

In [64]:
```
def sentiment_calc(text):
    try:
        return TextBlob(text).sentiment.polarity
    except:
        return None

df['sentiment'] = df['Reviews'].apply(sentiment_calc)
```

In [65]:
```
df.sentiment.head()
```

Out[65]:
```
9203    0.04375
9227   -0.17000
9233   -0.04376
```

```
9239    0.30000
9251    0.00000
Name: sentiment, dtype: float64
```

In [66]:
```
def getAnalysis(score):
  if score < 0:
    return 'Negative'
  elif score == 0:
    return 'Neutral'
  else:
    return 'Positive'

df ['Review_Sentiment'] = df['sentiment'].apply(getAnalysis )
```

Here we are categorising Side effects in terms of following two score :-

1. Score < 0 - Negative (Which indicates that the patient did suffer side effects using the Drug)
2. Score > 0 - Positive and Neutral (Which indicates that the patient did not suffer any side effects using the Drug)

Side-effect prediction based on sentiment score and store it in a new column

In [67]:
```
def getAnalysis(score):
  if score < 0:
    return 1
  else:
    return 0

df ['Review_Score'] = df['sentiment'].apply(getAnalysis )
```

Here we are storing the values (0 and 1) which will indicate if the patient suffer from side effect or not
\ 0 (Positive) - Patient did not suffer any side effects
1 (Negative) - Patient did suffer side effects

In [68]:
```
df.head(5)
```

Out[68]:

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **9203** | 75 or over | High Blood Pressure | 2/8/2020 | lisinopril solution | 6873 | 5 | 4 | i have developed a very itch rash on my body f... | 4 | F |
| **9227** | 55-64 | High Blood Pressure | 8/16/2019 | lisinopril solution | 6873 | 3 | 1 | was on this medicine for about months and hav... | 1 | F |
| **9233** | 25-34 | High Blood Pressure | 6/3/2019 | lisinopril solution | 6873 | 5 | 2 | i was placed on a low dose of lisinopril after... | 1 | F |

| | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfaction | |
|---|---|---|---|---|---|---|---|---|---|---|
| **9239** | 35-44 | Other | 6/2/2019 | lisinopril solution | 6873 | 5 | 5 | no script or health insurance needed to place ... | 5 | F |
| **9251** | 65-74 | High Blood Pressure | 5/6/2019 | lisinopril solution | 6873 | 3 | 2 | i just started taking it | 2 | F |

# Negative Reviews age wise

In [69]:
```
df_3=df[df.Review_Sentiment=="Negative"]
```

In [70]:
```
print(df_3["Age"].value_counts())
```

```
55-64         242
45-54         242
65-74         132
35-44         116
75 or over     47
25-34          39
19-24           2
13-18           1
Name: Age, dtype: int64
```

**Total Reviews**

In [71]:
```
print(df["Age"].value_counts())
```

```
55-64         539
45-54         530
65-74         300
35-44         261
75 or over    103
25-34          97
19-24          11
13-18           2
Name: Age, dtype: int64
```

# Building a Logistic Regression Model

In [72]:
```
logModel = LogisticRegression()
```

Testing

In [73]:
```
X_var = df[['EaseofUse','Effectiveness','Satisfaction','Sex_n','Age_Group_number']].val
y_var = df['Review_Score'].values # dependent variable
```

Train - Test Split

Dropping the whole index column

In [74]:
```
df.reset_index(drop=True, inplace=True)
```

Adding and reseting the index column

In [75]:
```
df.reset_index()
```

Out[75]:

| | index | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfac |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 75 or over | High Blood Pressure | 2/8/2020 | lisinopril solution | 6873 | 5 | 4 | i have developed a very itch rash on my body f... | |
| **1** | 1 | 55-64 | High Blood Pressure | 8/16/2019 | lisinopril solution | 6873 | 3 | 1 | was on this medicine for about months and hav... | |
| **2** | 2 | 25-34 | High Blood Pressure | 6/3/2019 | lisinopril solution | 6873 | 5 | 2 | i was placed on a low dose of lisinopril after... | |
| **3** | 3 | 35-44 | Other | 6/2/2019 | lisinopril solution | 6873 | 5 | 5 | no script or health insurance needed to place ... | |
| **4** | 4 | 65-74 | High Blood Pressure | 5/6/2019 | lisinopril solution | 6873 | 3 | 2 | i just started taking it | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **1838** | 1838 | 35-44 | High Blood Pressure | 5/2/2011 | lisinopril solution | 6873 | 5 | 4 | today is my day on lisinopril so far so good... | |
| **1839** | 1839 | 25-34 | High Blood Pressure | 5/1/2011 | lisinopril solution | 6873 | 5 | 3 | i have been taking this med for about a week a... | |
| **1840** | 1840 | 55-64 | High Blood Pressure | 4/27/2011 | lisinopril solution | 6873 | 1 | 1 | i used to take hctz for years with no problem ... | |

| | index | Age | Condition | Date | Drug | DrugId | EaseofUse | Effectiveness | Reviews | Satisfac |
|---|---|---|---|---|---|---|---|---|---|---|
| **1841** | 1841 | 55-64 | High Blood Pressure | 4/24/2011 | lisinopril solution | 6873 | 5 | 3 | dry cough causing difficulty sleeping and havi... | |
| **1842** | 1842 | 45-54 | High Blood Pressure | 4/24/2011 | lisinopril solution | 6873 | 2 | 3 | i started this drug six days ago for high bloo... | |

1843 rows × 21 columns

In [76]:
```
X_train, X_test, y_train, y_test = train_test_split(X_var, y_var, test_size = 0.3, rand
```

Samples

In [77]:
```
print(cl('X variable samples : {}'.format(X_var[:5]), attrs = ['bold']))
print(cl('Y variable samples : {}'.format(y_var[:5]), attrs = ['bold']))
```

```
X variable samples : [[ 5  4  4  0 11]
 [ 3  1  1  0  9]
 [ 5  2  1  0  6]
 [ 5  5  5  0  7]
 [ 3  2  2  0 10]]
Y variable samples : [0 1 1 0 0]
```

In [78]:
```
logModel.fit(X_train, y_train)
```

Out[78]:
```
LogisticRegression()
```

In [79]:
```
logModel.score(X_train, y_train)
```

Out[79]:
```
0.5782945736434109
```

In [80]:
```
logModel.score(X_test,y_test)
```

Out[80]:
```
0.5768535262206148
```

In [81]:
```
logModel.coef_
```

Out[81]:
```
array([[ 0.00639378,  0.02338568, -0.32268264,  0.        , -0.00863536]])
```

In [82]:
```
logModel.intercept_
```

Out[82]:
```
array([0.47388002])
```

# Making Prediction

In [83]:
```
predictions = logModel.predict(X_test)
```

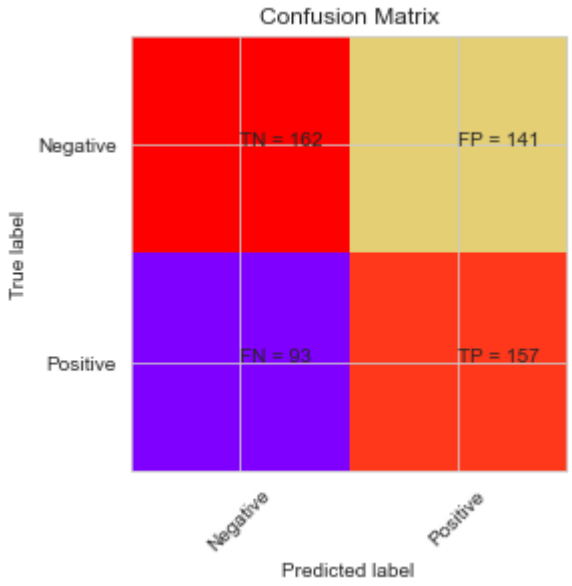
In [84]:
```
cm=confusion_matrix(y_test,predictions)
```

In [85]:
```
confusion_matrix(y_test,predictions)
```

Out[85]:
```
array([[162, 141],
       [ 93, 157]], dtype=int64)
```

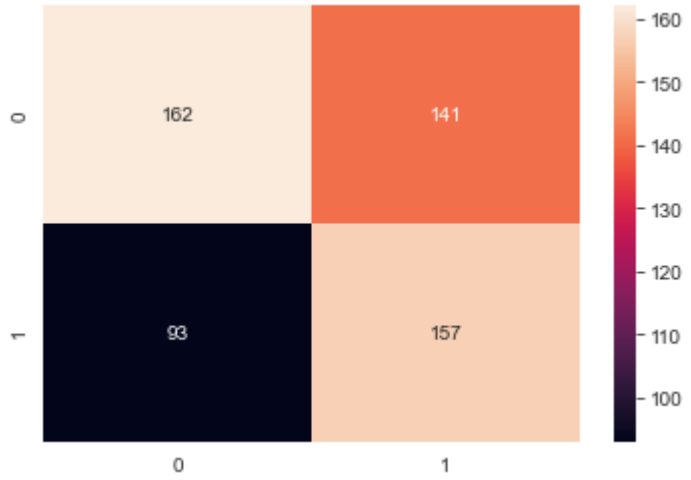
### Displaying the Confusion Matrix

In [86]:
```
plt.clf()
plt.imshow(cm, interpolation='nearest', cmap=plt.cm.rainbow)
classNames = ['Negative','Positive']
plt.title('Confusion Matrix')
plt.ylabel('True label')
plt.xlabel('Predicted label')
tick_marks = np.arange(len(classNames))
plt.xticks(tick_marks, classNames, rotation=45)
plt.yticks(tick_marks, classNames)
s = [['TN','FP'], ['FN', 'TP']]
for i in range(2):
    for j in range(2):
        plt.text(j,i, str(s[i][j])+" = "+str(cm[i][j]))
plt.show()
```




In [87]:
```
sns.heatmap(cm, annot = True, fmt = 'd')
```

Out[87]:
```
<AxesSubplot:>
```




In [88]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test, predictions))
```

```
              precision    recall  f1-score   support

           0       0.64      0.53      0.58       303
           1       0.53      0.63      0.57       250

    accuracy                           0.58       553
   macro avg       0.58      0.58      0.58       553
weighted avg       0.59      0.58      0.58       553
```

# Conclusion

Lets understand the above confusion matrix as per the logistic regression algorithm

As we can see the values :-

**TN (True Negative)** = [162] - It indicates the number of patient who had no side effects and were correctly identified by the model also.
**FP (False Positive)** = [141] - It indicates the number of patient that did not had any side effect but our algorithm detected them as patients who had side effects.
**FN (False Negative)** = [93] - It indicates the number of patient that had side effect but the model detected them as patients who did not had any side effect.
**TP (True positive)** = [157] - It indicates the number of patients who had side effect and were correctly identified by the model also.

**Precision** - (What proportion of patient had side effect was actually correct?) = 0.53
**Recall** - (What proportion of actual patient who had side effect was identified correctly?) = 0.63