In [1]:
```
import os
import pandas as pd
import numpy as np
from sqlalchemy import create_engine
from sqlalchemy.types import Integer, Text, String, Float, DateTime
from datetime import datetime
```

In [2]:
```
def build_DB_URI(db_type, db_lib, user_id, password, db_name,  db_location='localhost', port='5432' ):
    '''
        A method which generates a DB_URI for SQL-Alchemey. Assumption that this will be
        used with Postgresql, however written to be generic.

        arg:

        db_type     --> the type of database, e.g 'postgres', 'mysql'

        db_lib      --> the appropriate sql-alchemy plughin for
                        db_type, e.g 'psycopg2' or 'pymysql'

        user_id     --> the user name for the database, who has
                        appropriate permissions

        password    --> the password for the db-user-id.
        db_name     --> the name of the db, e.g. 'esomeprazole'
        db_location --> the address / URL for the database. DEFAULT = localhost
        port        --> the port for the database. DEFAULT = 5432

        returns:
        db_URI     --> The URI for SQL-Alchemy of the form:
                       postgres+psycop2://user_id:password@db_location:5432/db_name

    '''

    db_URI = db_type+'+'+db_lib+'://'+user_id+':'+password+'@'+db_location+':'+port+'/'+db_name

    return db_URI
```

In [3]:
```
db_type = 'postgres'
db_lib = 'psycopg2'
user_id = 'bhima'
password= ''
db_name = 'openfda'

db_URI = build_DB_URI(db_type, db_lib, user_id, password, db_name)
db_engine = create_engine(db_URI, echo=False)
db_engine.connect()
connection= db_engine.connect()
```

In [4]:
```
sql_query = 'SELECT * FROM drugs_reactions_countries;'
```

In [5]:
```
df = pd.read_sql(sql_query,
                      con=db_engine)
```

In [ ]:
```

```

In [6]:
```
df.describe()
```

Out[6]:

	safetyreportid	activesubstancename	openfda_generic_name	reactionmeddrapt	primarysource_repor
count	87828606	87828606	87828606	87828606	
unique	237943	351	2250	9613	
top	9350740			Pain	
freq	3360000	87764046	26877200	725617	

In [ ]:
```

```

In [7]:
```
drug_and_raction_grouped = df.groupby(['reactionmeddrapt', 'activesubstancename
'])
print(drug_and_raction_grouped.agg(np.size))
```

```
                                                    safetyreportid  \
reactionmeddrapt               activesubstancename
ACTH-producing pituitary tumour                                 70
AIDS dementia complex                                           34
Abasia                                                       56098
Abdomen crushing                                                20
Abdominal abscess                                             3700
...                                                            ...
Zinc deficiency                                                81
Zygomycosis                                                   490
pH body fluid abnormal                                        480
pH body fluid decreased                                      1014
pH urine increased                                             52

                                                    openfda_generic_name  \
reactionmeddrapt               activesubstancename
ACTH-producing pituitary tumour                                   70
AIDS dementia complex                                             34
Abasia                                                         56098
Abdomen crushing                                                  20
Abdominal abscess                                               3700
...                                                              ...
Zinc deficiency                                                  81
Zygomycosis                                                     490
pH body fluid abnormal                                          480
pH body fluid decreased                                        1014
pH urine increased                                               52

                                                    primarysource_reportercou
ntry
reactionmeddrapt               activesubstancename
ACTH-producing pituitary tumour
70
AIDS dementia complex
34
Abasia                                                                      5
6098
Abdomen crushing
20
Abdominal abscess
3700
...
...
Zinc deficiency
81
Zygomycosis
490
pH body fluid abnormal
480
pH body fluid decreased
1014
pH urine increased
52

[12692 rows x 3 columns]
```

In [8]:
```
drug_and_raction_df = df[['activesubstancename', 'reactionmeddrapt']].drop_dupl
icates()
```

In [9]:
```
drug_and_raction_df.describe()
```

Out[9]:

	activesubstancename	reactionmeddrapt
count	12692	12692
unique	351	9613
top		Pyrexia
freq	9612	72

In [10]:
```
drug_and_raction_df.groupby(['reactionmeddrapt', 'activesubstancename']).groups
```

Out[10]:
```
{('ACTH-producing pituitary tumour', ''): Int64Index([7178205], dtype='int64
'),
 ('AIDS dementia complex', ''): Int64Index([570732], dtype='int64'),
 ('Abasia', ''): Int64Index([4266], dtype='int64'),
 ('Abdomen crushing', ''): Int64Index([85201058], dtype='int64'),
 ('Abdominal abscess', ''): Int64Index([1492694], dtype='int64'),
 ('Abdominal adhesions', ''): Int64Index([1599829], dtype='int64'),
 ('Abdominal cavity drainage', ''): Int64Index([7034116], dtype='int64'),
 ('Abdominal compartment syndrome', ''): Int64Index([1094352], dtype='int64'),
 ('Abdominal compartment syndrome',
  'BEVACIZUMAB'): Int64Index([1999486], dtype='int64'),
 ('Abdominal compartment syndrome',
  'BLEOMYCIN SULFATE'): Int64Index([1999232], dtype='int64'),
 ('Abdominal compartment syndrome',
  'CARBOPLATIN'): Int64Index([1999184], dtype='int64'),
 ('Abdominal compartment syndrome',
  'CISPLATIN'): Int64Index([1999209], dtype='int64'),
 ('Abdominal compartment syndrome',
  'CYCLOPHOSPHAMIDE'): Int64Index([1999252], dtype='int64'),
 ('Abdominal compartment syndrome',
  'DOXORUBICIN'): Int64Index([1999269], dtype='int64'),
 ('Abdominal compartment syndrome',
  'ETOPOSIDE'): Int64Index([1999296], dtype='int64'),
 ('Abdominal compartment syndrome',
  'MELPHALAN'): Int64Index([1999370], dtype='int64'),
 ('Abdominal compartment syndrome',
  'VINBLASTINE'): Int64Index([1999340], dtype='int64'),
 ('Abdominal discomfort', ''): Int64Index([10358], dtype='int64'),
 ('Abdominal discomfort',
  'AMITRIPTYLINE'): Int64Index([13934044], dtype='int64'),
 ('Abdominal discomfort',
  'CYANOCOBALAMIN\\PYRIDOXAMINE HYDROCHLORIDE\\THIAMINE HYDROCHLORIDE'): Int64
Index([13934104], dtype='int64'),
 ('Abdominal discomfort', 'FLUPIRTINE'): Int64Index([13934098], dtype='int64
'),
 ('Abdominal discomfort',
  'PANTOPRAZOLE SODIUM'): Int64Index([13934049], dtype='int64'),
 ('Abdominal discomfort', 'QUETIAPINE'): Int64Index([13934113], dtype='int64
'),
 ('Abdominal discomfort',
  'VENLAFAXINE HYDROCHLORIDE'): Int64Index([13934062], dtype='int64'),
 ('Abdominal distension', ''): Int64Index([7594], dtype='int64'),
 ('Abdominal hernia', ''): Int64Index([215105], dtype='int64'),
 ('Abdominal hernia obstructive', ''): Int64Index([13024631], dtype='int64'),
 ('Abdominal hernia repair', ''): Int64Index([188249], dtype='int64'),
 ('Abdominal infection', ''): Int64Index([1490183], dtype='int64'),
 ('Abdominal injury', ''): Int64Index([263472], dtype='int64'),
 ('Abdominal lymphadenopathy', ''): Int64Index([878793], dtype='int64'),
 ('Abdominal mass', ''): Int64Index([712806], dtype='int64'),
 ('Abdominal neoplasm', ''): Int64Index([5565949], dtype='int64'),
 ('Abdominal operation', ''): Int64Index([6035640], dtype='int64'),
 ('Abdominal pain', ''): Int64Index([1475], dtype='int64'),
 ('Abdominal pain', 'ATAZANAVIR'): Int64Index([3221872], dtype='int64'),
 ('Abdominal pain',
  'AZITHROMYCIN ANHYDROUS'): Int64Index([3221843], dtype='int64'),
 ('Abdominal pain', 'CEFTRIAXONE'): Int64Index([3221752], dtype='int64'),
 ('Abdominal pain', 'DIDANOSINE'): Int64Index([3221740], dtype='int64'),
 ('Abdominal pain', 'RITONAVIR'): Int64Index([3221814], dtype='int64'),
 ('Abdominal pain', 'TENOFOVIR'): Int64Index([3221954], dtype='int64'),
 ('Abdominal pain lower', ''): Int64Index([3534], dtype='int64'),
 ('Abdominal pain lower',
  'ACETAMINOPHEN'): Int64Index([2379022], dtype='int64'),
 ('Abdominal pain lower', 'ASPIRIN'): Int64Index([2379051], dtype='int64'),
 ('Abdominal pain lower',
  'ATORVASTATIN'): Int64Index([2379028], dtype='int64'),
 ('Abdominal pain lower', 'BISOPROLOL'): Int64Index([2379036], dtype='int64'),
 ('Abdominal pain lower', 'CANAKINUMAB'): Int64Index([2379080], dtype='int64
'),
 ('Abdominal pain lower', 'FUROSEMIDE'): Int64Index([2379054], dtype='int64'),
 ('Abdominal pain lower', 'PERINDOPRIL'): Int64Index([2379096], dtype='int64
```

In [ ]:

In [ ]:

In [20]:
```
genericname_and_raction_grouped = df.groupby(['activesubstancename', 'openfda_generic_name'])
counted_genericname_and_raction_grouped = genericname_and_raction_grouped.agg(np.size)
```

In [12]:
```
genericname_and_raction_df = df[['activesubstancename', 'openfda_generic_name']].drop_duplicates()
```

In [13]:
```
genericname_and_raction_df.describe()
```

Out[13]:

	activesubstancename	openfda_generic_name
count	2630	2630
unique	351	2250
top		
freq	2250	276

In [15]:
```
genericname_and_raction_df.groupby(['activesubstancename', 'openfda_generic_nam
e']).groups
```

Out[15]:
```
{('', ''): Int64Index([22], dtype='int64'),
 ('', '(ESTRADIOL TRANSDERMAL SYSTEM)'): Int64Index([2121110], dtype='int64'),
 ('',
  '(SUMATRIPTAN INJECTION) 6MG/0.5ML, AUTO-INJECTOR;SUMATRIPTAN'): Int64Index
([10482796], dtype='int64'),
 ('',
  '.ALPHA.-TOCOPHEROL ACETATE, DL-, ASCORBIC ACID, CYANOCOBALAMIN, SODIUM FLUO
RIDE, FOLIC ACID, NIACIN, PYRIDOXINE, RIBOFLAVIN, THIAMINE, VITAMIN A, AND VIT
AMIN D;SODIUM FLUORIDE'): Int64Index([11381808], dtype='int64'),
 ('',
  '.BETA.-CAROTENE, ASCORBIC ACID, CHOLECALCIFEROL, .ALPHA.-TOCOPHEROL ACETAT
E, DL-, THIAMINE MONONITRATE, RIBOFLAVIN, NIACINAMIDE, PYRIDOXINE HYDROCHLORID
E, FOLIC ACID, 5-METHYLTETRAHYDROFOLIC ACID, CALCIUM FORMATE, FERROUS ASPARTO
GLYCINATE, CYANOCOBALAMIN, BIOTIN, POTASSIUM IODIDE, MAGNESIUM OXIDE, ZINC OXI
DE AND CUPRIC OXIDE;.BETA.-CAROTENE, ASCORBIC ACID, CHOLECALCIFEROL, .ALPHA.-T
OCOPHEROL ACETATE, DL-, THIAMINE MONONITRATE, RIBOFLAVIN, NIACINAMIDE, PYRIDOX
INE HYDROCHLORIDE, FOLIC ACID, CYANOCOBALAMIN, BIOTIN, CALCIUM PANTOTHENATE, C
ALCIUM CARBONATE, FERROUS FUMARATE, POTASSIUM IODIDE, MAGNESIUM OXIDE, ZINC OX
IDE AND CUPRIC OXIDE'): Int64Index([52550644], dtype='int64'),
 ('',
  '.BETA.-CAROTENE, SODIUM ACETATE, ASCORBIC ACID, CHOLECALCIFEROL, .ALPHA.-TO
COPHEROL ACETATE, DL-, THIAMINE MONONITRATE, RIBOFLAVIN, NIACINAMIDE, PYRIDOXI
NE HYDROCHLORIDE, FOLIC ACID, CYANOCOBALAMIN, CALCIUM CARBONATE, FERROUS FUMAR
ATE, ZINC OXIDE AND CUPRIC OXIDE;VITAMIN A ACETATE, .BETA.-CAROTENE, ASCORBIC
ACID, CHOLECALCIFEROL, .ALPHA.-TOCOPHEROL ACETATE, DL-, THIAMINE MONONITRATE,
RIBOFLAVIN, NIACINAMIDE, PYRIDOXINE HYDROCHLORIDE, FOLIC ACID, CYANOCOBALAMIN,
CALCIUM CARBONATE, FERROUS FUMARATE, ZINC OXIDE, CUPRIC OXIDE'): Int64Index([6
23634], dtype='int64'),
 ('',
  '0.63% STANNOUS FLUORIDE;STANNOUS FLUORIDE'): Int64Index([14301153], dtype='
int64'),
 ('', '0XYGEN;OXYGEN;OXYGEN COMPRESSED'): Int64Index([4315], dtype='int64'),
 ('',
  '1.1% SODIUM FLUORIDE PRESCRIPTION DENTAL CREAM'): Int64Index([8633137], dty
pe='int64'),
 ('', 'ABACAVIR SULFATE'): Int64Index([976634], dtype='int64'),
 ('', 'ABACAVIR SULFATE AND LAMIVUDINE'): Int64Index([217342], dtype='int64'),
 ('',
  'ABACAVIR SULFATE, LAMIVUDINE, AND ZIDOVUDINE'): Int64Index([7314480], dtype
='int64'),
 ('', 'ABACAVIR;ABACAVIR SULFATE'): Int64Index([512646], dtype='int64'),
 ('', 'ABATACEPT'): Int64Index([56646], dtype='int64'),
 ('', 'ABCIXIMAB'): Int64Index([183700], dtype='int64'),
 ('', 'ABIRATERONE ACETATE'): Int64Index([43112], dtype='int64'),
 ('', 'ACAMPROSATE CALCIUM'): Int64Index([1155823], dtype='int64'),
 ('', 'ACARBOSE'): Int64Index([996030], dtype='int64'),
 ('', 'ACEBUTOLOL HYDROCHLORIDE'): Int64Index([378877], dtype='int64'),
 ('', 'ACETAMINOPHEN'): Int64Index([17789], dtype='int64'),
 ('',
  'ACETAMINOPHEN AND CHLORPHENIRAMINE MALEATE'): Int64Index([14850], dtype='in
t64'),
 ('',
  'ACETAMINOPHEN AND CHLORPHENIRAMINE MALEATE;DEXTROMETHORPHAN HYDROBROMIDE, G
UAIFENESIN, ACETAMINOPHEN, CHLORPHENIRAMINE MALEATE'): Int64Index([11654762],
dtype='int64'),
 ('',
  'ACETAMINOPHEN AND CODEINE PHOSPHATE'): Int64Index([734179], dtype='int64'),
 ('',
  'ACETAMINOPHEN AND CODEINE;ACETAMINOPHEN AND CODEINE PHOSPHATE'): Int64Index
([6300446], dtype='int64'),
 ('',
  'ACETAMINOPHEN AND DIPHENHYDRAMINE CITRATE'): Int64Index([85985019], dtype='
int64'),
 ('',
  'ACETAMINOPHEN AND DIPHENHYDRAMINE HCL;ACETAMINOPHEN AND DIPHENHYDRAMINE HYD
ROCHLORIDE;ACETAMINOPHEN PM;ACETAMINOPHEN, DIPHENHYDRAMINE HCL'): Int64Index
([7324312], dtype='int64'),
 ('',
  'ACETAMINOPHEN AND DIPHENHYDRAMINE HCL;ACETAMINOPHEN, CHLORPHENIRAMINE MALEA
TE, PHENYLEPHRINE HCL;ALLIUM CEPA, APIS MELLIFICA, ARALIA RACEMOSA, ARUNDO MAU
```

In [ ]:

```
```

In [16]:

```
genericname_and_raction_grouped = df.groupby(['activesubstancename', 'openfda_generic_name'])
counted_genericname_and_raction = genericname_and_raction_grouped.agg(np.size)
```

Out[16]:

		safetyreportid	activesubstancename	openfda_generic_nam
reactionmeddrapt	**primarysource_reportercountry**			
ACTH-producing pituitary tumour	**US**	70	70	7
AIDS dementia complex	**COUNTRY NOT SPECIFIED**	12	12	1
	JP	12	12	1
	US	10	10	1
Abasia	**AE**	44	44	4
...	**...**	...	...	
Zygomycosis	**JP**	77	77	7
	US	200	200	20
pH body fluid abnormal	**US**	480	480	48
pH body fluid decreased	**DE**	1014	1014	101
pH urine increased	**IR**	52	52	5

54110 rows × 3 columns

In [17]:

```
country_and_reaction_df = df[['reactionmeddrapt', 'primarysource_reportercountry']].drop_duplicates()
```

In [18]:

```
country_and_reaction_df.describe()
```

Out[18]:

	reactionmeddrapt	primarysource_reportercountry
count	54110	54110
unique	9613	141
top	Pyrexia	US
freq	82	7771

In [19]:
```
country_and_reaction_df.groupby(['reactionmeddrapt', 'primarysource_reportercou
ntry']).groups
```

Out[19]:
```
{('ACTH-producing pituitary tumour',
  'US'): Int64Index([7178205], dtype='int64'),
 ('AIDS dementia complex',
  'COUNTRY NOT SPECIFIED'): Int64Index([570732], dtype='int64'),
 ('AIDS dementia complex', 'JP'): Int64Index([76050869], dtype='int64'),
 ('AIDS dementia complex', 'US'): Int64Index([6656662], dtype='int64'),
 ('Abasia', 'AE'): Int64Index([78975707], dtype='int64'),
 ('Abasia', 'AR'): Int64Index([830248], dtype='int64'),
 ('Abasia', 'AU'): Int64Index([1721108], dtype='int64'),
 ('Abasia', 'BE'): Int64Index([72820009], dtype='int64'),
 ('Abasia', 'BR'): Int64Index([965606], dtype='int64'),
 ('Abasia', 'CA'): Int64Index([1307971], dtype='int64'),
 ('Abasia', 'CM'): Int64Index([62839480], dtype='int64'),
 ('Abasia', 'CN'): Int64Index([77114388], dtype='int64'),
 ('Abasia', 'CO'): Int64Index([6809291], dtype='int64'),
 ('Abasia', 'COUNTRY NOT SPECIFIED'): Int64Index([4266], dtype='int64'),
 ('Abasia', 'DE'): Int64Index([2330719], dtype='int64'),
 ('Abasia', 'DK'): Int64Index([68844902], dtype='int64'),
 ('Abasia', 'ES'): Int64Index([11782057], dtype='int64'),
 ('Abasia', 'FR'): Int64Index([2274904], dtype='int64'),
 ('Abasia', 'GB'): Int64Index([94651], dtype='int64'),
 ('Abasia', 'IE'): Int64Index([13442964], dtype='int64'),
 ('Abasia', 'IN'): Int64Index([7363114], dtype='int64'),
 ('Abasia', 'JP'): Int64Index([227052], dtype='int64'),
 ('Abasia', 'KW'): Int64Index([10698735], dtype='int64'),
 ('Abasia', 'MX'): Int64Index([97369], dtype='int64'),
 ('Abasia', 'NL'): Int64Index([5464325], dtype='int64'),
 ('Abasia', 'PT'): Int64Index([11131277], dtype='int64'),
 ('Abasia', 'SE'): Int64Index([7520466], dtype='int64'),
 ('Abasia', 'TN'): Int64Index([79129512], dtype='int64'),
 ('Abasia', 'TR'): Int64Index([7651043], dtype='int64'),
 ('Abasia', 'US'): Int64Index([41854], dtype='int64'),
 ('Abasia', 'UY'): Int64Index([6734867], dtype='int64'),
 ('Abdomen crushing', 'US'): Int64Index([85201058], dtype='int64'),
 ('Abdominal abscess', 'BR'): Int64Index([6137040], dtype='int64'),
 ('Abdominal abscess', 'CA'): Int64Index([1492694], dtype='int64'),
 ('Abdominal abscess',
  'COUNTRY NOT SPECIFIED'): Int64Index([6300196], dtype='int64'),
 ('Abdominal abscess', 'JP'): Int64Index([2948055], dtype='int64'),
 ('Abdominal abscess', 'NL'): Int64Index([3577265], dtype='int64'),
 ('Abdominal abscess', 'TW'): Int64Index([85594061], dtype='int64'),
 ('Abdominal abscess', 'US'): Int64Index([7726961], dtype='int64'),
 ('Abdominal adhesions', 'BR'): Int64Index([1680988], dtype='int64'),
 ('Abdominal adhesions', 'CA'): Int64Index([3961794], dtype='int64'),
 ('Abdominal adhesions', 'CN'): Int64Index([77369612], dtype='int64'),
 ('Abdominal adhesions',
  'COUNTRY NOT SPECIFIED'): Int64Index([1888664], dtype='int64'),
 ('Abdominal adhesions', 'DE'): Int64Index([7615145], dtype='int64'),
 ('Abdominal adhesions', 'GB'): Int64Index([5598186], dtype='int64'),
 ('Abdominal adhesions', 'US'): Int64Index([1599829], dtype='int64'),
 ('Abdominal adhesions', 'VE'): Int64Index([6550516], dtype='int64'),
 ('Abdominal cavity drainage', 'US'): Int64Index([7034116], dtype='int64'),
 ('Abdominal compartment syndrome',
  'DK'): Int64Index([2925004], dtype='int64'),
 ('Abdominal compartment syndrome',
  'US'): Int64Index([1094352], dtype='int64'),
 ('Abdominal discomfort', 'AR'): Int64Index([54874331], dtype='int64'),
 ('Abdominal discomfort', 'AT'): Int64Index([8631029], dtype='int64'),
 ('Abdominal discomfort', 'AU'): Int64Index([610892], dtype='int64'),
 ('Abdominal discomfort', 'BR'): Int64Index([1647474], dtype='int64'),
 ('Abdominal discomfort', 'CA'): Int64Index([1147706], dtype='int64'),
 ('Abdominal discomfort', 'CH'): Int64Index([866899], dtype='int64'),
 ('Abdominal discomfort', 'CL'): Int64Index([7095151], dtype='int64'),
 ('Abdominal discomfort', 'CN'): Int64Index([946438], dtype='int64'),
 ('Abdominal discomfort',
  'COUNTRY NOT SPECIFIED'): Int64Index([10358], dtype='int64'),
 ('Abdominal discomfort', 'CZ'): Int64Index([11505340], dtype='int64'),
 ('Abdominal discomfort', 'DE'): Int64Index([2330727], dtype='int64'),
 ('Abdominal discomfort', 'DK'): Int64Index([690695], dtype='int64'),
 ('Abdominal discomfort', 'EG'): Int64Index([85409219], dtype='int64'),
```

In [21]:
```
country_and_reaction__grouped = df.groupby(['reactionmeddrapt', 'primarysource_
reportercountry'])
counted_genericname_and_raction = country_and_reaction__grouped.agg(np.size)
```

In [ ]:
```

```