Project Title	250k Medicines Usage, Side Effects and Substitutes Comprehensive collection of medical drugs info worldwide from various sources.
Tools	Python, ML, SQL, Excel
Technologies	Data Analyst & Data scientist
Project Difficulties level	intermediate

Dataset : Dataset is available in the given link. You can download it at your convenience.

Click here to download data set

About Dataset

This dataset contains comprehensive information on over **248,000** medical drugs from all manufacturers available worldwide. The data includes details such as drug names, active ingredients, therapeutic uses, dosage, side effects, and substitutes. The dataset aims to provide a useful resource for *medical researchers, healthcare professionals, and drug manufacturers*.

The dataset contains the following information for each drug:

1. **Drug name**
2. **Adverse reactions and side effects**
3. **Drug interactions**
4. **Drug class**
5. **Substitute drugs**
6. **Active ingredients** [Available in previous dataset published by us. Link added Below]

Dataset Related Info.:

1. There are total of 5 substitutes columns in the dataset.
2. There are total of 41 Side Effects columns in the dataset. Drugs with less side effects have those columns as blank.
3. There are total of 5 usage columns in the dataset.
4. Rest Columns have their own individual identity.

Substitute : It indicates that whatever is the name of the medicine, if the consumer want's to know alternative medicine that have same compositions but manufactured by different manufacturers. So we have provided 5 substitutes for each drug/medicine.

Therapeutic Class : It is a way of classifying medical drugs according to their functions. Each therapeutic class is a group of similar medications classified together because they are intended to treat the same medical conditions. for eg.: Respiratory is one of the class, so all medication treating this illness comes under the same class.

Action Class : It is the way of classifying medications based on actions they perform such as "H2 Receptor Blocker" It Block H2 receptors in parietal cells of the stomach → decrease gastric acid secretion. So drugs with similar action are grouped under "H2 Receptor Blocker".

Chemical Class : As name suggests, it is grouping based on chemical compound used.

Habit Forming : It is classified as YES or NO. It is defined as the process of forming a habit, referring generally to psychological dependence on the continued use of a drug to maintain a sense of well-being, which can result in drug addiction. If there are chances of drug addiction, then it is set to YES otherwise NO.

Similar Dataset that has market based information & Compositions of each drug :
https://www.kaggle.com/datasets/shudhanshusingh/az-medicine-dataset-of-india

Important Info.:

The dataset can be used for various purposes such as drug research and development, pharmacovigilance, clinical trials, and drug safety analysis.

Announcement :
I have released a new dataset on **Real Estate Properties** , if you are interested must checkout here: https://www.kaggle.com/datasets/shudhanshusingh/real-estate-properties-dataset ,If you liked it, do give an upvote :

NOTE :

1. this project is only for your guidance, not exactly the same you have to create. Here I am trying to show the way or idea of what steps you can follow and how your projects look. Some projects are very advanced (because it will be made with the help of flask, nlp, advance ai, advance DL and some advanced things) which you can not understand .

2. You can make or analyze your project with yourself, with your idea, make it more creative from where we can get some information and understand about our business. make sure what overall things you have created all things you understand very well.

Example

what steps you should have to follow

Here's a step-by-step guide to start your data analytics project titled "250k Medicines Usage, Side Effects, and Substitutes" using the provided columns.

1. Objective

The project aims to explore the relationships between medicines, their substitutes, side effects, and uses. You'll also analyze how these factors vary across different chemical, therapeutic, and action classes.

2. Data Overview

You have the following columns:

- **Basic Info**: `id`, `name`
- **Substitutes**: `substitute0`, `substitute1`, `substitute2`, `substitute3`, `substitute4`
- **Side Effects**: `sideEffect0` to `sideEffect41`
- **Uses**: `use0` to `use4`
- **Classification**: `Chemical Class`, `Habit Forming`, `Therapeutic Class`, `Action Class`

3. Steps to Follow

Step 1: Data Loading and Initial Exploration

Load the dataset and get a basic overview.

```
import pandas as pd


# Load the dataset

df = pd.read_csv('medicines_data.csv')


# Display the first few rows

print(df.head())


# Display basic statistics

print(df.describe())


# Display information about the dataset

print(df.info())
```

Output: You will see the first few rows of your dataset, summary statistics for numerical columns, and information about the data types and non-null values in each column.

Step 2: Data Cleaning

Check for missing values and handle them.

```
# Check for missing values

print(df.isnull().sum())


# Handle missing values, for example, by filling them with
'Unknown' or dropping rows/columns
```

```
df.fillna('Unknown', inplace=True)
```

Output: A summary of missing values per column before and after handling them.

Step 3: Analyzing Substitutes

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

```
# Count the number of substitutes for each medicine

df['num_substitutes'] = df[['substitute0', 'substitute1',
'substitute2', 'substitute3',
'substitute4']].notnull().sum(axis=1)


# Display the top 10 medicines with the most substitutes

print(df[['name',
'num_substitutes']].sort_values(by='num_substitutes',
ascending=False).head(10))
```

Output: A list of medicines with the most substitutes and the number of substitutes they have.

Step 4: Analyzing Side Effects

Identify the most common side effects across all medicines.

```
# Combine all side effects columns into one

side_effects_columns = [f'sideEffect{i}' for i in range(42)]

side_effects = df[side_effects_columns].apply(lambda x:
x.value_counts()).sum(axis=1).sort_values(ascending=False)
```

```
# Display the top 10 most common side effects

print(side_effects.head(10))
```

Output: A list of the most common side effects and their frequency.

Step 5: Analyzing Uses

Explore the most common uses of the medicines.

```
# Combine all use columns into one

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

uses = df[use_columns].apply(lambda x:
x.value_counts()).sum(axis=1).sort_values(ascending=False)


# Display the top 10 most common uses

print(uses.head(10))
```

Output: A list of the most common uses and their frequency.

Step 6: Classification Analysis

Analyze how side effects and uses vary across different classes.

```
# Group by Chemical Class and analyze side effects

chemical_class_effects = df.groupby('Chemical
Class')[side_effects_columns].apply(lambda x:
x.notnull().sum()).sum(axis=1)


# Display the top 10 Chemical Classes with the most side
effects

print(chemical_class_effects.sort_values(ascending=False).head(
10))
```

```
# Similarly, group by Therapeutic Class and Action Class

therapeutic_class_effects = df.groupby('Therapeutic
Class')[side_effects_columns].apply(lambda x:
x.notnull().sum()).sum(axis=1)

action_class_effects = df.groupby('Action
Class')[side_effects_columns].apply(lambda x:
x.notnull().sum()).sum(axis=1)



# Display the results

print(therapeutic_class_effects.sort_values(ascending=False).he
ad(10))

print(action_class_effects.sort_values(ascending=False).head(10
))
```

Output: Lists of the top Chemical, Therapeutic, and Action Classes with the most side effects.

4. Conclusion

- **Summary**: Summarize the findings, such as the most common side effects, which medicines have the most substitutes, and how these factors vary by classification.
- **Visualization (Optional)**: You can create plots using libraries like Matplotlib or Seaborn to visualize your findings.

Next Steps

- **Further Analysis**: Explore correlations between side effects, substitutes, and classes.
- **Advanced Techniques**: For more advanced analysis, consider clustering or classification models to predict side effects based on the medicine's chemical or therapeutic class.

This guide will help a beginner get started with exploring the dataset and deriving

meaningful insights.

Sample code

```
import pandas as pd

df                                                                    =

pd.read_csv('/kaggle/input/250k-medicines-usage-side-effects-and-substitutes/medicin

e_dataset.csv')

df.head()
```

```
/tmp/ipykernel_28/3317851521.py:2: DtypeWarning: Columns (42,43,44,45,46,47,48) have

mixed types. Specify dtype option on import or set low_memory=False.

                                                    df                                =

pd.read_csv('/kaggle/input/250k-medicines-usage-side-effects-and-substitutes/medicin

e_dataset.csv')
```

Out[2]:

	id	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	...	sideEffect41	use0	use1	use2	use3	use4	Chemical Class	Habit Forming	Therapeutic Class	Action Class

																		g		
0	1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	...	NaN	Treatment of Bacterial infections	NaN	NaN	NaN	NaN	NaN	No	ANTI INFECTIVES	NaN
1	2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro 500mg Tablet	Trulimax 500mg Tablet	Vomiting	Nausea	Abdominal pain	...	NaN	Treatment of Bacterial infections	NaN	NaN	NaN	NaN	Macrolides	No	ANTI INFECTIVES	Macrolides
2	3	ascoril ls syrup	Solvin LS Syrup	Ambrodil-LX Syrup	Zerotuss XP Syrup	Capex LS Syrup	Broxum LS Syrup	Nausea	Vomiting	Diarrhea	...	NaN	Treatment of Cough with mucus	NaN	NaN	NaN	NaN	NaN	No	RESPIRATORY	NaN
3	4	allegra 12	Lcfex Tab	Etofex 12	Nexofex	Fexise	Histafree	Headach	Drowsines	Dizzin	.	NaN	Treatme	Treatme	Na	Na	Na	Diphenylmethane	No	RESPIRAT	H1 Antihistan

		0mg tablet	let	0mg Tablet	120mg Tablet	120mg Tablet	120 Tablet	e	s	ess	.		nt of Sneezing and runny nose due to al...	nt of Allergic conditions	N	N	N	Derivative		ORY	inics (sec nd Gen ratio n)
4	5	avil 25 tablet	Eralet 25 mg Tablet	NaN	NaN	NaN	NaN	Sleepiness	Dryness in mouth	NaN	...	NaN	Treatment of Allergic conditions	NaN	NaN	NaN	NaN	Pyridines Derivatives	No	RESPIRATORY	H1 Antistanics (First Generation)

5 rows × 58 columns

In [3]:

```
df.columns
```

Out[3]:

```
Index(['id', 'name', 'substitute0', 'substitute1', 'substitute2',
       'substitute3', 'substitute4', 'sideEffect0', 'sideEffect1',
       'sideEffect2', 'sideEffect3', 'sideEffect4', 'sideEffect5',
```

```
       'sideEffect6', 'sideEffect7', 'sideEffect8', 'sideEffect9',
       'sideEffect10', 'sideEffect11', 'sideEffect12', 'sideEffect13',
       'sideEffect14', 'sideEffect15', 'sideEffect16', 'sideEffect17',
       'sideEffect18', 'sideEffect19', 'sideEffect20', 'sideEffect21',
       'sideEffect22', 'sideEffect23', 'sideEffect24', 'sideEffect25',
       'sideEffect26', 'sideEffect27', 'sideEffect28', 'sideEffect29',
       'sideEffect30', 'sideEffect31', 'sideEffect32', 'sideEffect33',
       'sideEffect34', 'sideEffect35', 'sideEffect36', 'sideEffect37',
       'sideEffect38', 'sideEffect39', 'sideEffect40', 'sideEffect41', 'use0',
       'use1', 'use2', 'use3', 'use4', 'Chemical Class', 'Habit Forming',
       'Therapeutic Class', 'Action Class'],

      dtype='object')
```

In [4]:

```
df.shape
```

Out[4]:

```
(248218, 58)
```

In [5]:

```
missing_vals = df.isnull().sum() / len(df)
missing_vals
```

Out[5]:

```
id                    0.000000
```

```
name                0.000000
sideEffect25        0.993945
sideEffect26        0.993945
sideEffect27        0.993981
sideEffect28        0.993981
sideEffect39        0.999992
sideEffect40        0.999992
sideEffect41        0.999992
use0                0.000000
use1                0.704433
use2                0.885959
use3                0.970272
use4                0.979973
Chemical Class      0.444879
Habit Forming       0.000000
Therapeutic Class   0.000278
Action Class        0.443892

dtype: float64
```

In [6]:

```
missing_more_15 = missing_vals[missing_vals > 0.15]
missing_more_15
```

Out[6]:

```
sideEffect3       0.163485
```

```
sideEffect4       0.341063

sideEffect5       0.471199

sideEffect6       0.629934

sideEffect7       0.727054

sideEffect8       0.804583

sideEffect9       0.848085

sideEffect10      0.890121

sideEffect11      0.918092

sideEffect12      0.934404

sideEffect13      0.940669

sideEffect14      0.958025

sideEffect15      0.969055

sideEffect16      0.975791

sideEffect17      0.978317

sideEffect18      0.981810

sideEffect19      0.984103


sideEffect36      0.999992

sideEffect37      0.999992

sideEffect38      0.999992

sideEffect39      0.999992

sideEffect40      0.999992

sideEffect41      0.999992

use1              0.704433

use2              0.885959
```

```
use3              0.970272

use4              0.979973

Chemical Class    0.444879

Action Class      0.443892


dtype: float64
```

In [7]:

```
list(missing_more_15.index)
```

Out[7]:

```
['sideEffect3',

 'sideEffect4',

 'sideEffect5',

 'sideEffect6',

 'sideEffect7',

 'sideEffect8',

 'sideEffect9',

 'sideEffect10',

 'sideEffect11',

 'sideEffect12',

 'sideEffect13',

 'sideEffect14',

 'sideEffect15',

 'sideEffect16',

 'sideEffect17',
```

```
'sideEffect18',

'sideEffect19',

'sideEffect20',

'sideEffect21',

'sideEffect22',

'sideEffect23',

'sideEffect24',

'sideEffect25',

'sideEffect26',

'sideEffect27',

'sideEffect28',

'sideEffect29',

'sideEffect30',

'sideEffect31',

'sideEffect32',

'sideEffect33',

'sideEffect34',

'sideEffect35',

'sideEffect36',

'sideEffect37',

'sideEffect38',

'sideEffect39',

'sideEffect40',

'sideEffect41',

'use1',
```

```
 'use2',

 'use3',

 'use4',

 'Chemical Class',


 'Action Class']
```

In [8]:

```
missing_vals[missing_vals <= 0.15]
```

Out[8]:

```
id                  0.000000

name                0.000000

substitute0         0.038664

substitute1         0.057816

substitute2         0.072456

substitute3         0.086061

substitute4         0.097721

sideEffect0         0.000000

sideEffect1         0.039489

sideEffect2         0.075410

use0                0.000000

Habit Forming       0.000000

Therapeutic Class   0.000278


dtype: float64
```

In [9]:

```
df1 = df.drop(list(missing_more_15.index) , axis = 'columns')

df1.head()
```

Out[9]:

	id	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	Treatment of Bacterial infections	No	ANTI INFECTIVES
1	2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro 500mg Tablet	Trulimax 500mg Tablet	Vomiting	Nausea	Abdominal pain	Treatment of Bacterial infections	No	ANTI INFECTIVES
2	3	ascoril ls syrup	Solvin LS Syrup	Ambrodil-LX Syrup	Zerotuss XP Syrup	Capex LS Syrup	Broxum LS Syrup	Nausea	Vomiting	Diarrhea	Treatment of Cough with mucus	No	RESPIRATORY
3	4	allegra 120mg tablet	Lcfex Tablet	Etofex 120mg Tablet	Nexofex 120mg Tablet	Fexise 120mg Tablet	Histafree 120 Tablet	Headache	Drowsiness	Dizziness	Treatment of Sneezing and runny nose due to al...	No	RESPIRATORY

4	5	avil 25 tablet	Eralet 25mg Tablet	NaN	NaN	NaN	NaN	Sleepiness	Dryness in mouth	NaN	Treatment of Allergic conditions	No	RESPIRATORY

In [10]:

```
df1 = df1.drop_duplicates()
```

In [11]:

```
df1['name'] = df1['name'].str.lower()

df1['name'] = df1['name'].str.strip()
```

In [12]:

```
df1['name'].value_counts()
```

Out[12]:

```
name

ns 0.9% infusion                      13

meditrax s 1000mg/500mg injection      8

modace 100mg/500mg tablet              6

mucohelp 1gm injection                 6

matchfix-cv 200mg/125mg tablet         6

                                      ..

etoxiflam 120mg tablet                 1
```

```
etocrax 90mg tablet                    1
edikacin 500mg injection               1
eurotam 0.40mg tablet                  1
zyvocol 1% dusting powder              1

Name: count, Length: 222825, dtype: int64
```

In [13]:

```
df1.isnull().sum()
```

Out[13]:

```
id                        0
name                      0
sideEffect0               0
sideEffect1            9802
sideEffect2           18718
use0                      0
Habit Forming             0
Therapeutic Class        69

dtype: int64
```

In [14]:

```
df1['use0'] = df1['use0'].str.lower()
df1['use0'] = df1['use0'].str.strip()
```

In [15]:

```
df1['use0'].value_counts()
```

Out[15]:

```
use0
treatment of bacterial infections                                39573
pain relief                                                      22783
treatment of type 2 diabetes mellitus                            11350
bacterial infections                                             11288
treatment of gastroesophageal reflux disease (acid reflux)        9749
                                                                 ...
urticaria                                                           1
treatment of arthralgia (joint pain)                                1
benzodiazepine overdose                                             1
skin disorders                                                      1
severe acute pain                                                   1

Name: count, Length: 655, dtype: int64
```

In [16]:

```
top_treatments = list(df1['use0'].value_counts().head(15).index)
top_treatments
```

Out[16]:

```
['treatment of bacterial infections',
 'pain relief',
```

```
 'treatment of type 2 diabetes mellitus',
 'bacterial infections',
 'treatment of gastroesophageal reflux disease (acid reflux)',
 'treatment of hypertension (high blood pressure)',
 'treatment of sneezing and runny nose due to allergies',
 'treatment of heartburn',
 'treatment of neuropathic pain',
 'treatment of allergic conditions',
 'treatment of fungal infections',
 'hypertension (high blood pressure)',
 'treatment of bacterial & parasitic infections',
 'treatment of depression',

 'treatment of common cold']
```

In [17]:

```
df1['use0'].value_counts().head(15).sum()
```

Out[17]:

```
135930
```

In [18]:

```
df1.shape[0] - df1['use0'].value_counts().head(15).sum()
```

Out[18]:

```
112288
```

In [19]:

```
def filter_high_uses(x):
    return x in top_treatments
```

In [20]:

```
df2 = df1[df1['use0'].apply(filter_high_uses)]
df2.head()
```

Out[20]:

	id	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	treatment of bacterial infections	No	ANTI INFECTIVES
1	2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro 500mg Tablet	Trulimax 500mg Tablet	Vomiting	Nausea	Abdominal pain	treatment of bacterial infections	No	ANTI INFECTIVES
3	4	allegra 120mg tablet	Lcfex Tablet	Etofex 120mg Tablet	Nexofex 120mg Tablet	Fexise 120mg Tablet	Histafree 120 Tablet	Headache	Drowsiness	Dizziness	treatment of sneezing and	No	RESPIRATORY

											runny nose due to al...		
4	5	avil 25 tablet	Eralet 25mg Tablet	NaN	NaN	NaN	NaN	Sleepiness	Dryness in mouth	NaN	treatment of allergic conditions	No	RESPIRATORY
5	6	allegra-m tablet	Emlukast-FX Tablet	LCFEX-Mont Tablet	Fixar 10mg/120mg Tablet	Histakind-M Tablet	Histafree-M Tablet	Nausea	Diarrhea	Vomiting	treatment of sneezing and runny nose due to al...	No	RESPIRATORY

In [21]:

```
df2.isnull().sum()
```

Out[21]:

```
id                    0
name                  0
substitute0        2545
substitute1        3878
substitute2        4908
substitute3        5916
substitute4        6772
sideEffect0           0
```

```
sideEffect1          1845

sideEffect2          5704

use0                    0

Habit Forming           0

Therapeutic Class       9


dtype: int64
```

In [22]:

```
df3 = df2.dropna()

df3.head()
```

Out[22]:

	id	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	treatment of bacterial infections	No	ANTI INFECTIVES
1	2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro 500mg Tablet	Trulimax 500mg Tablet	Vomiting	Nausea	Abdominal pain	treatment of bacterial infections	No	ANTI INFECTIVES

3	4	allegra 120mg tablet	Lcfex Tablet	Etofex 120mg Tablet	Nexofex 120mg Tablet	Fexise 120mg Tablet	Histafree 120 Tablet	Headache	Drowsiness	Dizziness	treatment of sneezing and runny nose due to al...	No	RESPIRATORY
5	6	allegra-m tablet	Emlukast-FX Tablet	LCFEX-Mont Tablet	Fixar 10mg/120mg Tablet	Histakind-M Tablet	Histafree-M Tablet	Nausea	Diarrhea	Vomiting	treatment of sneezing and runny nose due to al...	No	RESPIRATORY
6	7	amoxyclav 625 tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fighto x 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	treatment of bacterial infections	No	ANTI INFECTIVES

In [23]:

```
df2.shape[0] - df3.shape[0]
```

Out[23]:

```
12050
```

In [24]:

```
df3.shape
```

Out[24]:

```
(123880, 13)
```

In [25]:

```
df3.isnull().sum()
```

Out[25]:

```
id                  0
name                0
substitute0         0
substitute1         0
substitute2         0
substitute3         0
substitute4         0
sideEffect0         0
sideEffect1         0
sideEffect2         0
use0                0
Habit Forming       0
Therapeutic Class   0

dtype: int64
```

In [26]:

```
df3.head()
```

Out[26]:

	id	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	1	augmentin 625 duo tablet	Penciclav 500 mg/125 mg Tablet	Moxikind-CV 625 Tablet	Moxiforce-CV 625 Tablet	Fightox 625 Tablet	Novamox CV 625mg Tablet	Vomiting	Nausea	Diarrhea	treatment of bacterial infections	No	ANTI INFECTIVES
1	2	azithral 500 tablet	Zithrocare 500mg Tablet	Azax 500 Tablet	Zady 500 Tablet	Cazithro 500mg Tablet	Trulimax 500mg Tablet	Vomiting	Nausea	Abdominal pain	treatment of bacterial infections	No	ANTI INFECTIVES
3	4	allegra 120mg tablet	Lcfex Tablet	Etofex 120mg Tablet	Nexofex 120mg Tablet	Fexise 120mg Tablet	Histafree 120 Tablet	Headache	Drowsiness	Dizziness	treatment of sneezing and runny nose due to al...	No	RESPIRATORY
5	6	allegra-m tablet	Emlukast-FX Tablet	LCFEX-Mont Tablet	Fixar 10mg/120mg Tablet	Histakind-M Tablet	Histafree-M Tablet	Nausea	Diarrhea	Vomiting	treatment of sneezing and runny nose due to al...	No	RESPIRATORY
6	7	amoxyclav 625	Penciclav 500 mg/125 mg	Moxikind-CV 625	Moxiforce-CV 625	Fightox 625 Tablet	Novamox CV 625mg	Vomiting	Nausea	Diarrhea	treatment of bacterial infection	No	ANTI INFECTIVES

		tablet	Tablet	Tablet	Tablet		Tablet				s		

In [27]:

```
df4 = df3.drop('id' , axis = 'columns')
```

In [28]:

```
df4 = df4.applymap(lambda x : x.lower())

df4 = df4.applymap(lambda x : x.strip())
```

In [29]:

```
df4.head()
```

Out[29]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	treatment of bacterial infections	no	anti infectives

3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	headache	drowsiness	dizziness	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	treatment of sneezing and runny nose due to al...	no	respiratory
6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	antiinfectives

In [30]:

```
list(df4.columns)
```

Out[30]:

```
['name',
 'substitute0',
 'substitute1',
 'substitute2',
 'substitute3',
 'substitute4',
 'sideEffect0',
```

```
 'sideEffect1',
 'sideEffect2',
 'use0',
 'Habit Forming',

 'Therapeutic Class']
```

In [31]:

```
for col in list(df4.columns):
    print(col , len(df4[col].unique()))
```

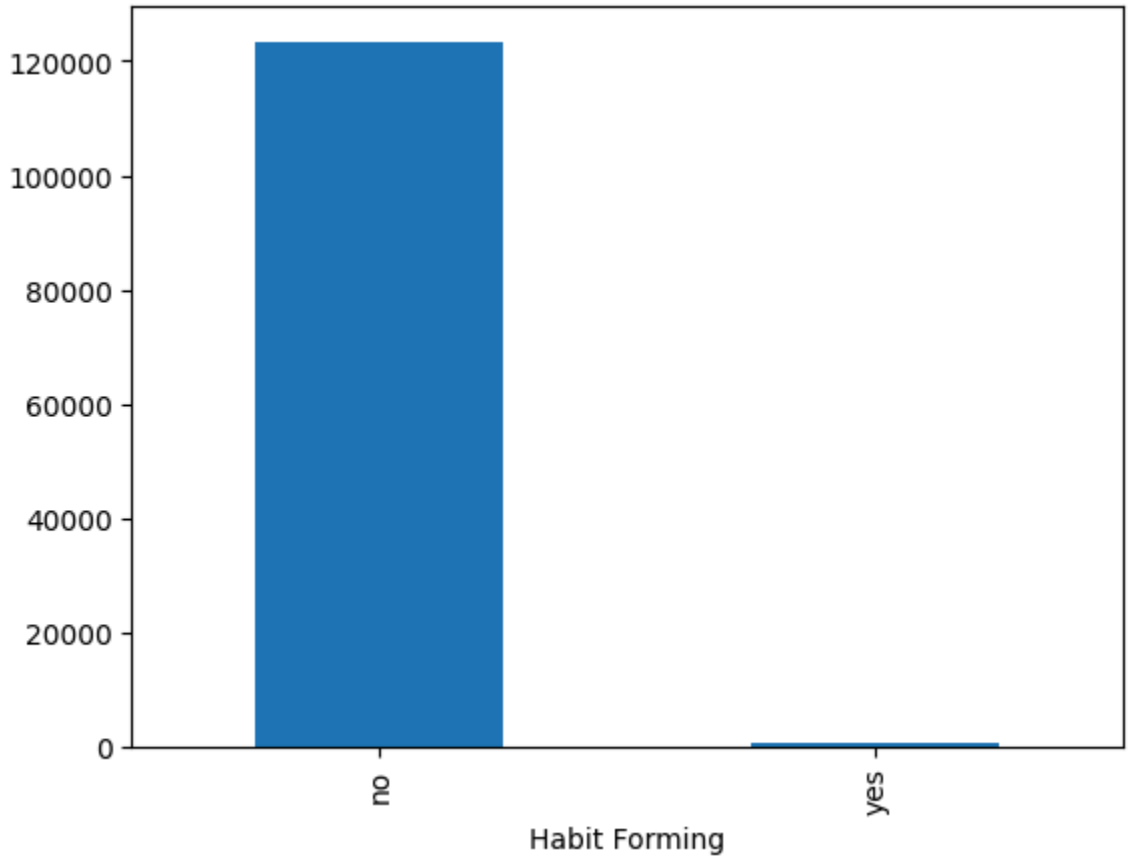
```
name 110212
substitute0 3551
substitute1 3991
substitute2 4201
substitute3 4351
substitute4 4412
sideEffect0 67
sideEffect1 71
sideEffect2 80
use0 15
Habit Forming 2
Therapeutic Class 12
```

In [32]:

```
df4['Habit Forming'].value_counts().plot.bar()
```

Out[32]:

```
<Axes: xlabel='Habit Forming'>
```



In [33]:

```
df4['sideEffect0'].value_counts()
```

Out[33]:

```
sideEffect0

nausea                      38072

diarrhea                    19940
```

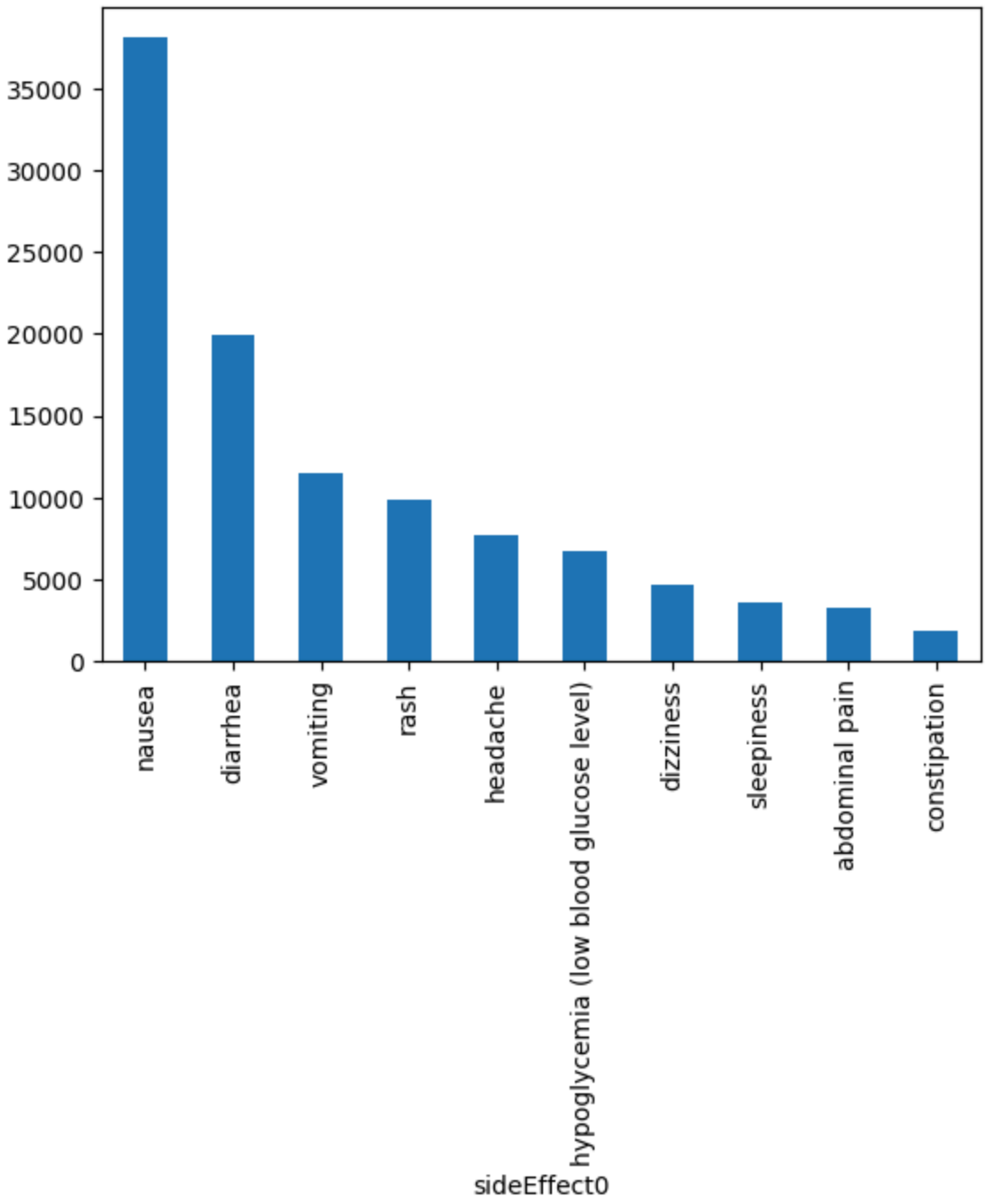
```
vomiting                     11478
rash                          9850
headache                      7727
                             ...
gastrointesinal symptoms         7
weight gain                      7
muscle rigidity                  7
loss of appetite                 6
urinary tract infection          5

Name: count, Length: 67, dtype: int64
```

In [34]:

```
df4['sideEffect0'].value_counts().head(10).plot.bar()
```

Out[34]:

```
<Axes: xlabel='sideEffect0'>
```



In [35]:

```
df4['sideEffect1'].value_counts().head(10).plot.bar()
```

Out[35]:

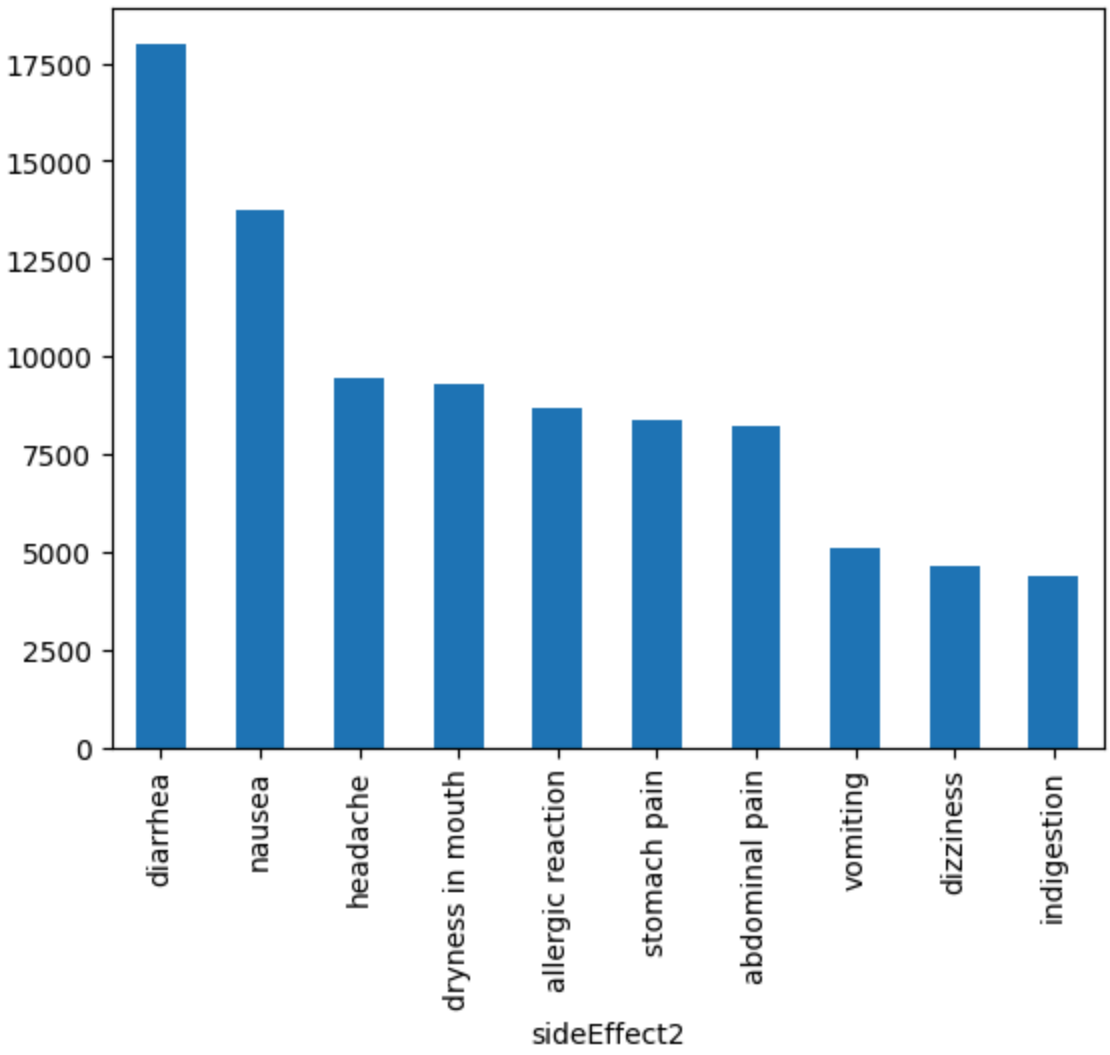
<Axes: xlabel='sideEffect1'>

In [36]:

```
df4['sideEffect2'].value_counts().head(10).plot.bar()
```

Out[36]:

<Axes: xlabel='sideEffect2'>



In [37]:

```
df4.head()
```

Out[37]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class

0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	headache	drowsiness	dizziness	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	treatment of sneezing and runny nose due to al...	no	respiratory
6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives

In [38]:

```
substitute0_sideeffects = {}

for sub,sideffect in df4.groupby('substitute0')['sideEffect0']:

    #print('Substtitute' , sub)
```

```
    #print('Sideeffect' , sideffect)

    substitute0_sideeffects[sub] = sideffect


#print(substitute0_sideeffects)
```

In [39]:

```
st0_sideeffects = pd.DataFrame(substitute0_sideeffects).fillna(0)

st0_sideeffects
```

Out[39]:

	1-alm syrup	3cef250mg tablet	3a250mg tablet dt	3d injection	5-zine-pt ablet	aflox200mg tablet	acold tablet	abdome 1mg oral suspension	abetin-nt200 tablet	abmetop 50 xl tablet	...	zycin 500mg injection	zydemf 80mg/500mg tablet	zydero 10mg/20mg tablet	zylocef t250mg/31.25mg injection	zylomefp ds oral suspension	zymostd tablet	zyn500mg tablet	zynaclox 250mg/250mg injection	zynol300mg tablet	zythro 200mg oral suspension
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
248210	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
248211	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
248212	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
248214	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

248216	0	0	0	0	0	0	0	0	0	0	...	injection site reactions (pain, swelling, redn...	0	0	0	0	0	0	0	0	0

123880 rows × 3551 columns

In [40]:

```
st0_sideeffects.columns
```

Out[40]:

```
Index(['1-al m syrup', '3 cef 250mg tablet', '3a 250mg tablet dt',
       '3d injection', '5-zine-p tablet', 'a flox 200mg tablet',
       'ab cold tablet', 'abdome 1mg oral suspension', 'abetin-nt 200 tablet',
       'abmetop 50 xl tablet',
       ...
       'zycin 500mg injection', 'zyde mf 80mg/500mg tablet',
       'zydero 10mg/20mg tablet', 'zylocef t 250 mg/31.25 mg injection',
       'zylomef p ds oral suspension', 'zymost d tablet', 'zyn 500mg tablet',
       'zynaclox 250mg/250mg injection', 'zynol 300mg tablet',
       'zythro 200mg oral suspension'],
```

```
      dtype='object', length=3551)
```

In [41]:

```
st0_sideeffects = st0_sideeffects.drop_duplicates()
```

In [42]:

```
st0_sideeffects.shape
```

Out[42]:

```
(3554, 3551)
```

In [43]:

```
st0_sideeffects['zycin 500mg injection'].unique()
```

Out[43]:

```
array([0, 'injection site reactions (pain, swelling, redness)'],
      dtype=object)
```

In [44]:

```
st0_sideeffects['lcfex tablet'].unique()
```

Out[44]:

```
array([0, 'headache'], dtype=object)
```

In [45]:

```
df4.head()
```

Out[45]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	headache	drowsiness	dizziness	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	treatment of sneezing and runny nose due to al...	no	respiratory

6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives

In [46]:

```
df4['Therapeutic Class'].unique()
```

Out[46]:

```
array(['anti infectives', 'respiratory', 'gastro intestinal', 'cardiac',
       'pain analgesics', 'neuro cns', 'anti diabetic', 'ophthal',
       'ophthal otologicals', 'otologicals',
       'vitamins minerals nutrients', 'derma'], dtype=object)
```

In [47]:

```
df4['Therapeutic Class'].value_counts()
```

Out[47]:

```
Therapeutic Class
anti infectives                43142
pain analgesics                19291
gastro intestinal              18747
anti diabetic                  10781
cardiac                        10240
respiratory                     9580
```

```
neuro cns                        6771
ophthal                          4933
ophthal otologicals               230
otologicals                       135
vitamins minerals nutrients        23
derma                               7

Name: count, dtype: int64
```

In [48]:

```
df4['Habit Forming'].value_counts()
```

Out[48]:

```
Habit Forming
no     123366
yes       514

Name: count, dtype: int64
```

In [49]:

```
df4.groupby('Therapeutic Class')['Habit Forming'].value_counts()
```

Out[49]:

```
Therapeutic Class            Habit Forming
anti diabetic                no               10781
anti infectives              no               43142
cardiac                      no               10240
```

```
derma                          no                7
gastro intestinal              no            18747
neuro cns                      no             6535
                               yes             236
ophthal                        no             4933
ophthal otologicals            no              230
otologicals                    no              135
pain analgesics                no            19018
                               yes             273
respiratory                    no             9575
                               yes               5
vitamins minerals nutrients    no               23

Name: count, dtype: int64
```

In [50]:

```
df4.groupby('Therapeutic Class')['Habit Forming'].value_counts().unstack().fillna(0)
```

Out[50]:

Habit Forming	no	yes
Therapeutic Class		
anti diabetic	10781.0	0.0

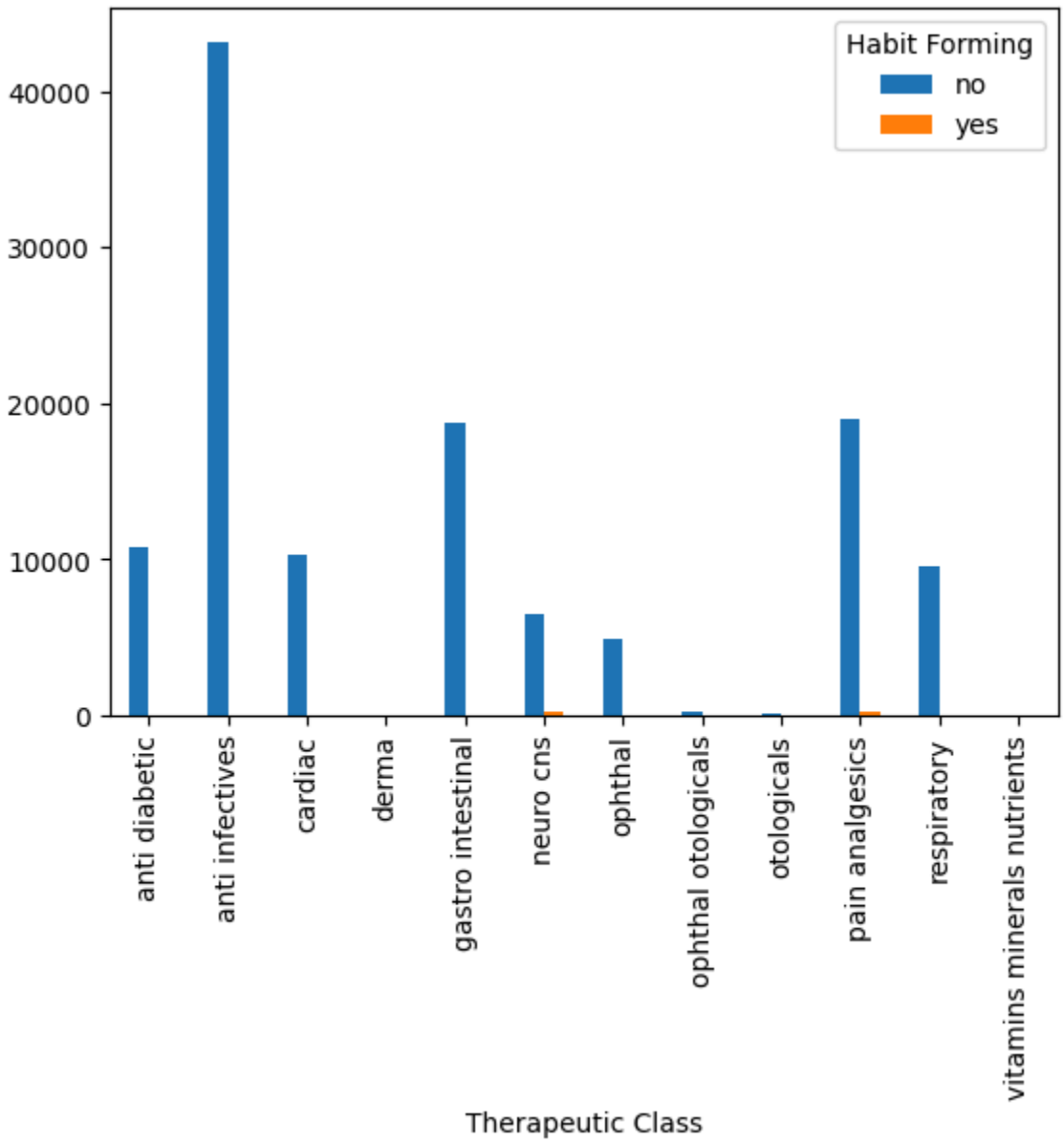
anti infectives	43142.0	0.0
cardiac	10240.0	0.0
derma	7.0	0.0
gastro intestinal	18747.0	0.0
neuro cns	6535.0	236.0
ophthal	4933.0	0.0
ophthal otologicals	230.0	0.0
otologicals	135.0	0.0
pain analgesics	19018.0	273.0
respiratory	9575.0	5.0
vitamins minerals nutrients	23.0	0.0

In [51]:

```
df4.groupby('Therapeutic Class')['Habit Forming'].value_counts().unstack().fillna(0).plot.bar()
```

Out[51]:

```
<Axes: xlabel='Therapeutic Class'>
```



In [52]:

```
df4.head()
```

Out[52]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	headache	drowsiness	dizziness	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	treatment of sneezing and runny nose due to al...	no	respiratory

6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives

In [53]:

```
X = df4.drop(['Therapeutic Class','Habit Forming'] , axis = 1)

Y = df4['Therapeutic Class']
```

In [54]:

```
df4.head()
```

Out[54]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	treatment of bacterial infections	no	anti infectives

3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	head ache	drowsiness	dizziness	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	treatment of sneezing and runny nose due to al...	no	respiratory
6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	treatment of bacterial infections	no	anti infectives

In [57]:

```
all_sideeffects = []

for i in range(0,len(df4)):

    sideeffects = list(df4.iloc[i , 6:9])

    sideeffects = sorted(sideeffects)

    all_sideeffects.append(sideeffects)
```

In [58]:

```
df4.insert(9,'All Sideeffects' , all_sideeffects)
```

In [59]:

```
df4.head()
```

Out[59]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	sideEffect0	sideEffect1	sideEffect2	All Sideeffects	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	[diarrhea, nausea, vomiting]	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	vomiting	nausea	abdominal pain	[abdominal pain, nausea, vomiting]	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	headache	drowsiness	dizziness	[dizziness, drowsiness, headache]	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	nausea	diarrhea	vomiting	[diarrhea, nausea, vomitin	treatment of sneezing and runny	no	respiratory

										g]	nose due to al...		
6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	vomiting	nausea	diarrhea	[diarrhea, nausea, vomiting]	treatment of bacterial infections	no	anti infectives

In [61]:

```
df5 = df4.drop(['sideEffect0','sideEffect1','sideEffect2'] , axis = 1)

df5.head()
```

Out[61]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	All Sideeffects	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	[diarrhea, nausea, vomiting]	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	[abdominal pain, nausea, vomiting]	treatment of bacterial infections	no	anti infectives
3	allegra 120mg	lcfex tablet	etofex 120mg	nexofex 120mg	fexise 120mg	histafree 120	[dizziness, drowsiness,	treatment of sneezing and runny nose	no	respiratory

	tablet		tablet	tablet	tablet	tablet	headache]	due to al...		
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120 mg tablet	histakind-m tablet	histafree-m tablet	[diarrhea, nausea, vomiting]	treatment of sneezing and runny nose due to al...	no	respiratory
6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamoxcv 625mg tablet	[diarrhea, nausea, vomiting]	treatment of bacterial infections	no	anti infectives

In [65]:

```
df5['All Sideeffects'] = df5['All Sideeffects'].astype(str)

df5['All Sideeffects'].nunique()
```

Out[65]:

```
232
```

In [78]:

```
count_all_sideeffects = df5['All Sideeffects'].value_counts()

count_all_sideeffects
```

Out[78]:

```
All Sideeffects
['diarrhea', 'nausea', 'vomiting']                    12052
['abdominal pain', 'nausea', 'vomiting']               6389
['nausea', 'stomach pain', 'vomiting']                 5211
```

```
['allergic reaction', 'rash', 'vomiting']              4502
['diarrhea', 'nausea', 'rash']                         4402
                                                       ...
['increased appetite', 'upset stomach', 'vomiting']       5
['nausea', 'urinary tract infection', 'vomiting']         5
['headache', 'nausea', 'tiredness']                       4
['diarrhea', 'dizziness', 'flatulence']                   2
['abdominal pain', 'headache', 'nausea']                  1

Name: count, Length: 232, dtype: int64
```

In [79]:

```
count_all_sideeffects[count_all_sideeffects > 100]
```

Out[79]:

```
All Sideeffects
['diarrhea',                          'nausea',                          'vomiting']
12052
['abdominal               pain',              'nausea',               'vomiting']
6389
['nausea',                'stomach                pain',               'vomiting']
5211
['allergic              reaction',              'rash',               'vomiting']
4502
['diarrhea',                           'nausea',                             'rash']
4402
```

```
...
['constipation',                    'dizziness',                    'headache']
110
['allergic          reaction',          'nausea',          'stomach          pain']
109
['injection site reactions (pain, swelling, redness)', 'stomach pain', 'vomiting']
107
['application site reactions (burning, irritation, itching and redness)', 'skin
rash', 'stomach pain']      106
['diarrhea',               'dizziness',               'muscle               damage']
106

Name: count, Length: 123, dtype: int64
```

In [67]:

```
all_subs = []
for i in range(0,len(df4)):
    subs = list(df4.iloc[i ,1: 6])
    subs = sorted(subs)
    all_subs.append(subs)
```

In [70]:

```
df5.insert(6 , 'All Substitutes' , all_subs)
df5.head()
```

Out[70]:

	name	substitute0	substitute1	substitute2	substitute3	substitute4	All Substitutes	All Substitute	All Sideeffects	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	zithrocare 500mg tablet	azax 500 tablet	zady 500 tablet	cazithro 500mg tablet	trulimax 500mg tablet	[azax 500 tablet, cazithro 500mg tablet, truli...	[azax 500 tablet, cazithro 500mg tablet, truli...	['abdominal pain', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	lcfex tablet	etofex 120mg tablet	nexofex 120mg tablet	fexise 120mg tablet	histafree 120 tablet	[etofex 120mg tablet, fexise 120mg tablet, his...	[etofex 120mg tablet, fexise 120mg tablet, his...	['dizziness', 'drowsiness', 'headache']	treatment of sneezing and runny nose due to al...	no	respiratory
5	allegra-m tablet	emlukast-fx tablet	lcfex-mont tablet	fixar 10mg/120mg tablet	histakind-m tablet	histafree-m tablet	[emlukast-fx tablet, fixar 10mg/120mg tablet, ...	[emlukast-fx tablet, fixar 10mg/120mg tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of sneezing and runny nose due to al...	no	respiratory

6	amoxyclav 625 tablet	penciclav 500 mg/125 mg tablet	moxikind-cv 625 tablet	moxiforce-cv 625 tablet	fightox 625 tablet	novamox cv 625mg tablet	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives

In [71]:

```
df6 = df5.drop(['substitute0','substitute1','substitute2',
                'substitute3','substitute4','All Substitute'] , axis=1)


df6.head()
```

Out[71]:

	name	All Substitutes	All Sideeffects	use0	Habit Forming	Therapeutic Class
0	augmentin 625 duo tablet	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives
1	azithral 500 tablet	[azax 500 tablet, cazithro 500mg tablet, truli...	['abdominal pain', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives
3	allegra 120mg tablet	[etofex 120mg tablet, fexise 120mg tablet, his...	['dizziness', 'drowsiness', 'headache']	treatment of sneezing and runny nose due to al...	no	respiratory

5	allegra-m tablet	[emlukast-fx tablet, fixar 10mg/120mg tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of sneezing and runny nose due to al...	no	respiratory
6	amoxyclav 625 tablet	[fightox 625 tablet, moxiforce-cv 625 tablet, ...	['diarrhea', 'nausea', 'vomiting']	treatment of bacterial infections	no	anti infectives

In [72]:

```
df6['All Substitutes'] = df6['All Substitutes'].astype(str)
```

In [73]:

```
df6['All Substitutes'].nunique()
```

Out[73]:

```
10974
```

In [74]:

```
df6.shape
```

Out[74]:

```
(123880, 6)
```

In [76]:

```
count_all_subs = df6['All Substitutes'].value_counts()

count_all_subs
```

Out[76]:

```
All Substitutes
['lecope-m  tablet', 'levocet  m  tablet', 'monticope  tablet', 'montina-l  tablet',
'solitair                                                                tablet']
1594
['oflamed  200  tablet', 'oflotas  200mg  tablet', 'olox  200mg  tablet', 'oxa  200mg
tablet',                  'zenflox                  200                  tablet']
1422
['bigcef  500  tablet', 'cefakind  500  tablet', 'cefoxim  500  tablet', 'pulmocef  500
tablet',                   'zefu                   500                   tablet']
1167
['acimol 100 mg/325 mg tablet', 'arflur-p tablet', 'ark-ap tablet', 'dolostat pc 100
mg/325          mg          tablet',          'topnac          p          tablet']
1145
['cefaxone  1gm  injection', 'ceftrax  1000mg  injection', 'ritecef  1000mg  injection',
'trixon     1000mg     injection',     'xone     hospital     1000mg     injection']
1096


...
['axetin tz 125mg/15.6mg injection', 'c tri xp  kid 125mg/15.6mg injection', 'montaz
125mg injection', 'taxone xp 125 mg/15.6 mg injection', 'traxol t 125 mg/15.6 mg
injection']       1
['expirtin  100mg  capsule',  'flit  100mg  capsule',  'flupidol  100mg  capsule',
```

'flupinock capsule', 'snepdol capsule']

1

['af 400 tablet', 'faze 400mg tablet', 'fazol 400mg tablet', 'flucus 400mg tablet', 'nuforce 400mg tablet']

1

['laxmo 4mg tablet', 'lezdes 4mg tablet', 'lornicam 4mg tablet', 'lorniz 4mg tablet', 'neucam 4mg tablet']

1

['biocef-s 1.5gm injection', 'c tum 1000mg/500mg injection', 'tazoquit sb 1000mg/500mg injection', 'tocef s 1000mg/500mg injection', 'ulfaz s 1000mg/500mg injection']

1

Name: count, Length: 10974, dtype: int64

In [77]:

linkcode

```
count_all_subs[count_all_subs > 100]
```

Out[77]:

All Substitutes

['lecope-m tablet', 'levocet m tablet', 'monticope tablet', 'montina-l tablet', 'solitair tablet']

1594

['oflamed 200 tablet', 'oflotas 200mg tablet', 'olox 200mg tablet', 'oxa 200mg tablet', 'zenflox 200 tablet']

```
1422
['bigcef 500 tablet', 'cefakind 500 tablet', 'cefoxim 500 tablet', 'pulmocef 500
tablet',                    'zefu                    500                    tablet']
1167
['acimol 100 mg/325 mg tablet', 'arflur-p tablet', 'ark-ap tablet', 'dolostat pc 100
mg/325          mg          tablet',          'topnac          p          tablet']
1145
['cefaxone 1gm injection', 'ceftrax 1000mg injection', 'ritecef 1000mg injection',
'trixon     1000mg     injection',     'xone     hospital     1000mg     injection']
1096


...
['glimpeg-vm 2 tablet', 'mg care v 2mg/500mg/0.2mg tablet', 'vogliboz-gm2 tablet',
'voglivance  gm  2mg/500mg/0.2mg  tablet',  'voglolysis  gm  2mg/500mg/0.2mg  tablet']
102
['rabalkem ls 75mg/20mg capsule', 'rabera-l capsule', 'rabesec ls capsule', 'rabipen
plus                capsule',                'rubizol-l                capsule']
102
['nervilax nt 75mg/10mg tablet', 'nervite tablet', 'neurica nt tablet', 'pregalift
nt         tablet',         'pregalin         nt         75mg/10mg         tablet']
101
['olmesafe h 12.5mg/40mg tablet', 'olmetime-h 40 tablet', 'olmetrack h 40 tablet',
'olmighty        40h        tablet',        'olsertain-h        40        tablet']
101
['caz fx 100mg/5ml dry syrup', 'favonix 100mg/5ml dry syrup', 'nelpod 100mg/5ml dry
```

```
syrup',   'prodox   100mg/5ml   dry   syrup',   'safenix   100mg/5ml   dry   syrup']

101


Name: count, Length: 197, dtype: int64
```

1 Reference link
2 Reference link for ML project