


```
In [1]: import pandas as pd
```

```
In [2]: df = pd.read_csv('processed.csv.gz')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	Unnamed: 0	Id	review	cleanReview	rating	ratingSentiment	ratingSentimentLabel
0	0	206461	"It has no side effect, I take it in combinati...	"It side effect, I take combination Bystolic 5...	9.0	2	positive
1	1	95260	"My son is halfway through his fourth week of ...	"My son halfway fourth week Intuniv. We became...	8.0	2	positive
2	2	92703	"I used to take another oral contraceptive, wh...	"I used take another oral contraceptive, 21 pi...	5.0	0	neutral
3	3	138000	"This is my first time using any form of birth...	"This first time using form birth control. I&#...	8.0	2	positive
4	4	35696	"Suboxone has completely turned my life around...	"Suboxone completely turned life around. I fee...	9.0	2	positive



```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215063 entries, 0 to 215062
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Unnamed: 0                            215063 non-null int64
1   Id                                     215063 non-null int64
2   review                               215063 non-null object
3   cleanReview                           215063 non-null object
4   rating                                215063 non-null float64
5   ratingSentiment                       215063 non-null int64
6   ratingSentimentLabel                  215063 non-null object
7   vaderReviewScore                      215063 non-null float64
8   vaderSentiment                        215063 non-null int64
9   vaderSentimentLabel                   215063 non-null object
dtypes: float64(2), int64(4), object(4)
memory usage: 16.4+ MB
```

```
In [5]: df = df.drop(columns=df.columns[0])
```

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

```
Out[6]:
```

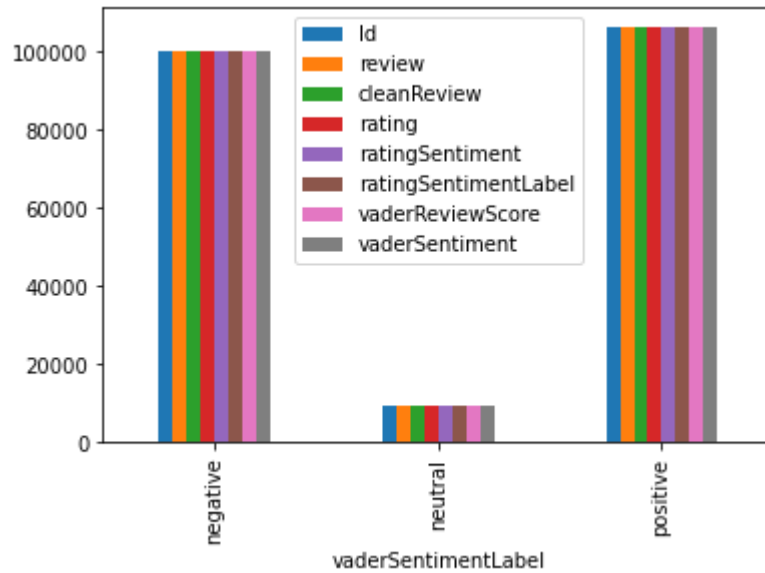
	Id	review	cleanReview	rating	ratingSentiment	ratingSentimentLabel	vaderReview
0	206461	"It has no side effect, I take it in combinati...	"It side effect, I take combination Bystolic 5...	9.0	2	positive	
1	95260	"My son is halfway through his fourth week of ...	"My son halfway fourth week Intuniv. We became...	8.0	2	positive	
2	92703	"I used to take another oral contraceptive, wh...	"I used take another oral contraceptive, 21 pi...	5.0	0	neutral	
3	138000	"This is my first time using any form of birth...	"This first time using form birth control. I&#...	8.0	2	positive	
4	35696	"Suboxone has completely turned my life around...	"Suboxone completely turned life around. I fee...	9.0	2	positive	

```
In [7]: df.groupby('vaderSentimentLabel').size()
```

```
Out[7]: vaderSentimentLabel
negative    99830
neutral     9035
positive   106198
dtype: int64
```

```
In [8]: import matplotlib.pyplot as plt
```

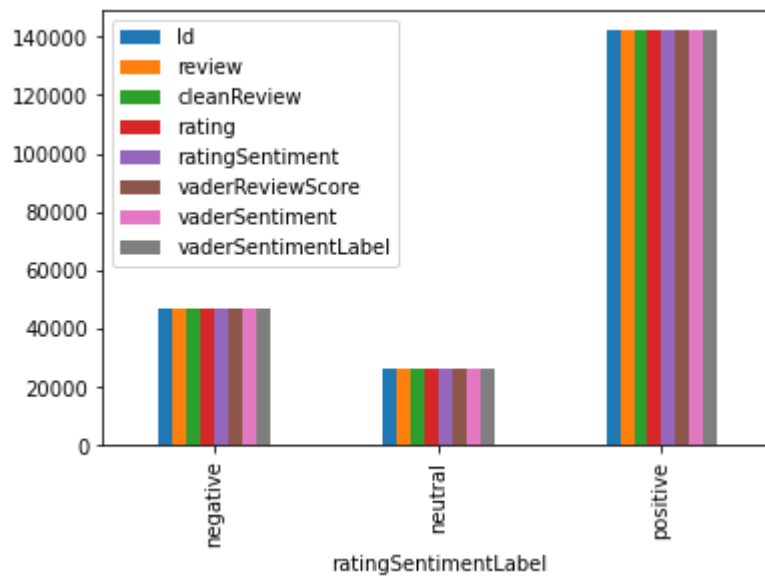
```
In [9]: df.groupby('vaderSentimentLabel').count().plot.bar()
plt.show()
```



```
In [10]: df.groupby('ratingSentimentLabel').size()
```

```
Out[10]: ratingSentimentLabel
negative      46901
neutral       25856
positive      142306
dtype: int64
```

```
In [11]: df.groupby('ratingSentimentLabel').count().plot.bar()
plt.show()
```



```
In [12]: df.groupby('ratingSentiment').size()
```

```
Out[12]: ratingSentiment
0      25856
1      46901
2      142306
dtype: int64
```

```
In [13]: positive_vader_sentiments = df[df.ratingSentiment == 2]
positive_string = []
for s in positive_vader_sentiments.cleanReview:
    positive_string.append(s)
positive_string = pd.Series(positive_string).str.cat(sep=' ')
```

```
In [14]: from wordcloud import WordCloud
wordcloud = WordCloud(width=2000,height=1000,max_font_size=200).generate(positive_string)
plt.imshow(wordcloud,interpolation='bilinear')
plt.show()
```



```
In [15]: for s in positive_vader_sentiments.cleanReview[:20]:
        if 'side effect' in s:
            print(s)
```

"It side effect, I take combination Bystolic 5 Mg Fish Oil"

"Abilify changed life. There hope. I Zoloft Clonidine I first started Abilify age 15.. Zoloft depression Clondine manage complete rage. My moods control. I depressed hopeless one second mean, irrational, full rage next. My Dr. prescribed 2mg Abilify point I feel like I cured though I know I'm not.. Bi-polar disorder constant battle. I know Abilify works I tried get lost complete control emotions. Went back I golden again. I 5mg 2x daily. I 21 better I ever past. Only side effect I like eat lot."

"I medication almost two weeks, started 25mg working way 100mg, currently 50mg. No headaches far I 2-3 crippling migraines week. I lost 5.2lbs far note I really paying close attention I eating, I lot weight lose weight loss side effect I want help along much I can. Now, side effects, word recall issues exist, memory issues, worst seems vision disturbances, times I driven I'm sure would safe. The good news seems wearing off...I tons energy I great mood."

"I taken anti-depressants years, improvement mostly moderate severe side affects, makes go them. I take Cymbalta mostly pain. When I began Deplin, I noticed major improvement overnight. More energy, better disposition, sinking low lows major depression. I taking 3 months feel like normal person first time ever. Best thing, side effects."

"I taking Saxenda since July 2016. I severe nausea month I got 2.6 dosage. It since subsided side effect I notice dry mouth. I make sure drink 2.5 litres water day (about 10 glasses). This helps weight loss well constipation. I reducing dose find comfortable spot I still losing weight don't feel like I medicating. For me, 1.8 working well. I also feel wearing Fitbit really helped. I track food, water, exercise steps - keeps moving more. When started I could barely walk length without getting winded - I lost 58 lbs far."

"This drug worked well cleared UTI matter 48hrs, although I 7 day course 2x200mg/daily. Unfortunately tablets finished infection returned needed 2nd course. I'm currently taking low dose evening keep infection bay specialists tests. I normally get every side effect going dread taking new medication for m especially horrific experience Ciprofloxacin side effect I experienced itchy ankles feet day two didn't last third day I would class minor side effect insignificant. Good luck."

"I Tassigna 3 years (300mg x 2 times day) Tassigna worked within weeks I remission basically entire 3 years. As side effects, I feel sick stomach, tired, severe leg arm pain, I can't walk long without needing break, I went working full time working I tired time.... hey I remission alive really problem? Nope"

"Nexplanon job. I worry free sex. The thing periods sometimes light sometimes heavy. Sometimes go away sometimes show unexpected. I also feel somewhat depressed. Not sure Nexplanon not. I've Nexplanon 2 months now, despite side effects effective birth control I've ever used I plan taking out."

```
In [16]: negative_vader_sentiments = df[df.ratingSentiment == 1]
        negative_string = []
        for s in negative_vader_sentiments.cleanReview:
            negative_string.append(s)
        negative_string = pd.Series(negative_string).str.cat(sep='')
```



```
In [20]: for s in neutral_vader_sentiments.cleanReview[:20]:  
        if 'side effect' in s:  
            print(s)
```

"I used take another oral contraceptive, 21 pill cycle, happy- light periods, max 5 days, side effects. But contained hormone gestodene, available US, I sw itched Lybrel, ingredients similar. When pills ended, I started Lybrel immedi ately, first day period, instructions said. And period lasted two weeks. When taking second pack- two weeks. And now, third pack things got even worse- thi rd period lasted two weeks it's end third week- I still daily brown disc harge. The positive side I didn't side effects. The idea period free tem pting... Alas."

"Have little bit lingering cough cold. Not giving much trouble except keeps n ight. I heard good I took I could get sleep. Helped tremendously cough I bad stomach cramps diarrhea. I hadn't eaten anything upset stomach didn't really feel like 'bug' I looked side effects Delsym. Now I wish I done first I probably wouldn't taken it. So, worked cough I still didn't get sleep due stomach issues."

"my gp started Venlafaxine yesterday help depression change, a hour taking fee ling sick couldn't stomach food fluids, thought keep told come side effects w ould get better, took another one last night ill couldn't stand ,being sick sw eating shaking thought going pass out. Did get sleep hoping feel better morn ing, took another one felt spaced dry mouth shaking ,sick, booked see gp make sure feeling like this, find put wrong dose 37.5mg put 150mg, right dose hope better"

"24 Year Old, Male, UK ,Normally I would go every hour, even drinking small a mounts water drinks. I took medication two weeks, I noticed right away I woul d go less often toilet, much easier dispense urine. Unfortunately I stop taki ng medications side effects outweighed benefits. I found I would incredibly i ntense mood swings, aggression, feel clumsy dizzy time. Blurry vision. Memory problems mind blanks. runny nose, weakness, tiredness. I trying Betmiga.. let s see"

"Was med 5 years. Worked fine great. Stopped panic attacks gave relief every day nervousness I experiencing. Took months side effects diminish I first sta rted taking it. What leave - insomnia, night sweats, lack sex drive poor orga sms (or orgasms all)."

"I birth control considering getting pregnant. I used types birth control bef ore, pill, patch, etc. I decided get back birth control didn't want worr y taking daily pill I requested Nuvaring. When I started Nuvaring I felt insi de body lot cramps. A days later I started experiencing cramps strong upper s tomach pain. They bad I missed work week. Finally I ended going ER couldn't figure wrong. I finally asked doctor Nuvaring side effects."

"I torn Nuvaring. The convenience great. There daily time take pill, month ea sily marked calender put phone. It never falls out, I never feel changing sim ple, too. I would give Nuvaring 10 10 based upon simplicity price really reas onable, too. But, downsides (for personally). For first months, I notice side effects, although there. I VERY low tolerance alcohol Nuvaring. Very emotiona l, tired, strung time. I pill none happened. I also longer periods. It's worth try, every girl seems different experience!"

"I really glad I experienced none possible side effects. After four weeks I n otice mood getting better, things happened life (my dog surgery, break-up, ha ir experiment gone wrong) I able deal I got worse. I felt spiraling downward anxiety shot way up. It's month number two I switched lexapro. Hope help s."

```
In [21]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [22]: tfidf = TfidfVectorizer(stop_words='english',ngram_range=(1,2))
features = tfidf.fit_transform(df.cleanReview)
labels   = df.vaderSentiment
```

```
In [23]: features.shape
```

```
Out[23]: (215063, 1619394)
```

```
In [24]: from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.naive_bayes import MultinomialNB
```

```
In [25]: x_train,x_test,y_train,y_test = train_test_split(df['cleanReview'],df['ratingSentimentLabel'],random_state=0)
```

```
In [26]: from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import LinearSVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import StandardScaler
```

```
In [27]: models = [RandomForestClassifier(n_estimators=200,max_depth=3,random_state=0),
LinearSVC(),MultinomialNB(),LogisticRegression(random_state=0,solver='lbfgs',max_iter=2000,multi_class='auto')]
CV = 5
cv_df = pd.DataFrame(index=range(CV * len(models)))
entries = []
for model in models:
    model_name = model.__class__.__name__
    accuracies = cross_val_score(model,features,labels,scoring='accuracy',cv=CV)
    for fold_idx,accuracy in enumerate(accuracies):
        entries.append((model_name,fold_idx,accuracy))
cv_df = pd.DataFrame(entries,columns=['model_name','fold_idx','accuracy'])
```



```
In [28]: cv_df
```

```
Out[28]:
```

	model_name	fold_idx	accuracy
0	RandomForestClassifier	0	0.495478
1	RandomForestClassifier	1	0.493851
2	RandomForestClassifier	2	0.494571
3	RandomForestClassifier	3	0.494071
4	RandomForestClassifier	4	0.494118
5	LinearSVC	0	0.944645
6	LinearSVC	1	0.944900
7	LinearSVC	2	0.945575
8	LinearSVC	3	0.945225
9	LinearSVC	4	0.943225
10	MultinomialNB	0	0.829586
11	MultinomialNB	1	0.834236
12	MultinomialNB	2	0.830888
13	MultinomialNB	3	0.831512
14	MultinomialNB	4	0.833302
15	LogisticRegression	0	0.901727
16	LogisticRegression	1	0.902704
17	LogisticRegression	2	0.902332
18	LogisticRegression	3	0.903353
19	LogisticRegression	4	0.902469

```
In [29]: cv_df.groupby('model_name').accuracy.mean()
```

```
Out[29]: model_name
LinearSVC          0.944714
LogisticRegression 0.902517
MultinomialNB      0.831905
RandomForestClassifier 0.494418
Name: accuracy, dtype: float64
```

```
In [30]: from sklearn.preprocessing import Normalizer
```

```
In [31]: model = LinearSVC('l2')
x_train,x_test,y_train,y_test = train_test_split(features,labels,test_size=0.2
5,random_state=0)
normalize = Normalizer()
x_train = normalize.fit_transform(x_train)
x_test = normalize.transform(x_test)
model.fit(x_train,y_train)
y_pred = model.predict(x_test)
```

```
In [32]: from sklearn.metrics import accuracy_score
print(accuracy_score(y_test, y_pred))
```

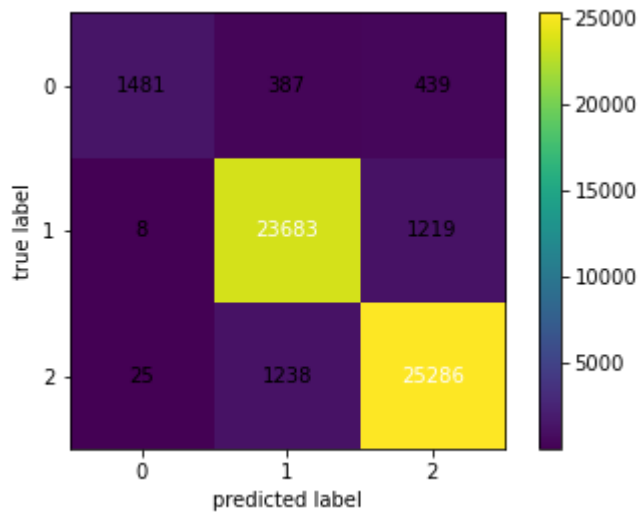
0.93832533571402

```
In [33]: from sklearn.metrics import confusion_matrix
conf_mat = confusion_matrix(y_test,y_pred)
conf_mat
```

```
Out[33]: array([[ 1481,   387,   439],
               [    8, 23683,  1219],
               [   25,  1238, 25286]], dtype=int64)
```

```
In [34]: from mlxtend.plotting import plot_confusion_matrix
```

```
In [35]: fig,ax = plot_confusion_matrix(conf_mat=conf_mat,colorbar=True,show_absolute=True,
cmap='viridis')
```



```
In [36]: from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred,target_names= df['ratingSentimentLabel'].unique()))
```

	precision	recall	f1-score	support
positive	0.98	0.64	0.78	2307
neutral	0.94	0.95	0.94	24910
negative	0.94	0.95	0.95	26549
accuracy			0.94	53766
macro avg	0.95	0.85	0.89	53766
weighted avg	0.94	0.94	0.94	53766

```
In [37]: df.head()
```

```
Out[37]:
```

	Id	review	cleanReview	rating	ratingSentiment	ratingSentimentLabel	vaderReview
0	206461	"It has no side effect, I take it in combinati...	"It side effect, I take combination Bystolic 5...	9.0	2	positive	
1	95260	"My son is halfway through his fourth week of ...	"My son halfway fourth week Intuniv. We became...	8.0	2	positive	
2	92703	"I used to take another oral contraceptive, wh...	"I used take another oral contraceptive, 21 pi...	5.0	0	neutral	
3	138000	"This is my first time using any form of birth...	"This first time using form birth control. l&#...	8.0	2	positive	
4	35696	"Suboxone has completely turned my life around...	"Suboxone completely turned life around. I fee...	9.0	2	positive	



```
In [38]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 215063 entries, 0 to 215062
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Id                                    215063 non-null int64
1   review                               215063 non-null object
2   cleanReview                           215063 non-null object
3   rating                                215063 non-null float64
4   ratingSentiment                       215063 non-null int64
5   ratingSentimentLabel                  215063 non-null object
6   vaderReviewScore                      215063 non-null float64
7   vaderSentiment                        215063 non-null int64
8   vaderSentimentLabel                   215063 non-null object
dtypes: float64(2), int64(3), object(4)
memory usage: 14.8+ MB
```

```
In [39]: y0 = df['vaderSentimentLabel']  
y0
```

```
Out[39]: 0          neutral  
1         positive  
2         positive  
3         positive  
4         positive  
        ...  
215058    negative  
215059    positive  
215060    negative  
215061    negative  
215062     neutral  
Name: vaderSentimentLabel, Length: 215063, dtype: object
```