



# PATIENT'S CONDITION CLASSIFICATION USING DRUG REVIEWS



# EDA

# BUSINESS OBJECTIVE:

This is a sample dataset which consists of 161297 drug name, condition reviews and ratings from different patients and our goal is to examine how patients are feeling using the drugs their positive and negative experiences so that we can recommend him a suitable drug. By analyzing the reviews, we can understand the drug effectiveness and its side effects.

# DATASET

The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. So in this dataset, we can see many patients conditions but we will focus only on the below, classify the below conditions from the patients reviews

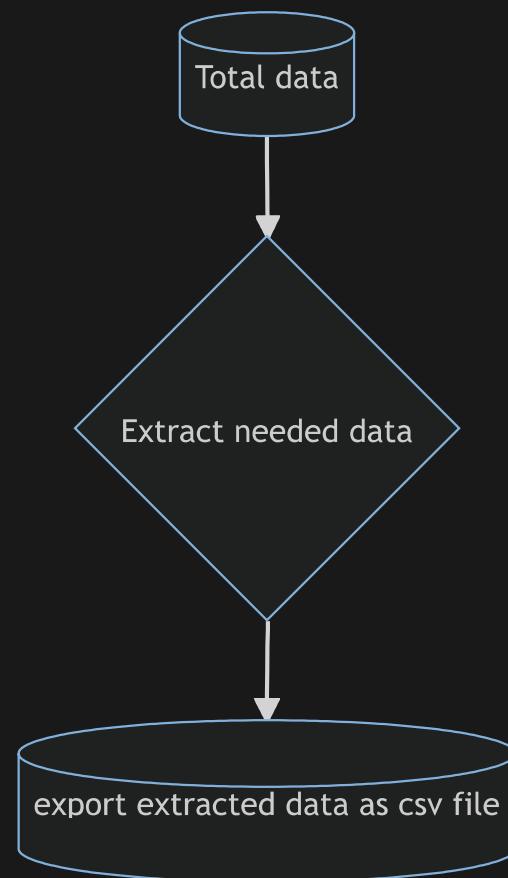
- a. Depression
- b. High Blood Pressure
- c. Diabetes, Type 2

# DATA SET EXPLANATION

## Attribute Information:

1. DrugName (categorical): name of drug
2. condition (categorical): name of condition
3. review (text): patient review
4. rating (numerical): 10 star patient rating
5. date (date): date of review entry
6. usefulCount (numerical): number of users who found review useful

# Project data extaction flow



## Total data set dimensions

```
df.shape
```

We have total 161297 rows and 7 columns

```
(161297, 7)
```

# Extracting the following diseases data:

```
df_proj = df_data.loc[df['condition']
                      .isin(["Diabetes, Type 2",
                             "Diabetes, Type 1",
                             "Diabetic Peripheral Neuropathy",
                             "Diabetic Kidney Disease",
                             "Diabetes Insipidus",
                             "Diabetic Macular Edema",
                             "Gestational Diabetes",
                             "Depression",
                             "Postpartum Depression",
                             "Neurotic Depression",
                             "High Blood Pressure",
                             "Postoperative Increased Intraocular Pressure"
                            ]) ]
```

## Extracted data set dimensions

```
df_proj.shape
```

We have total 161297 rows and 7 columns

```
(14386, 7)
```

# FORTUNATELY WE DON'T HAVE NULL VALUES

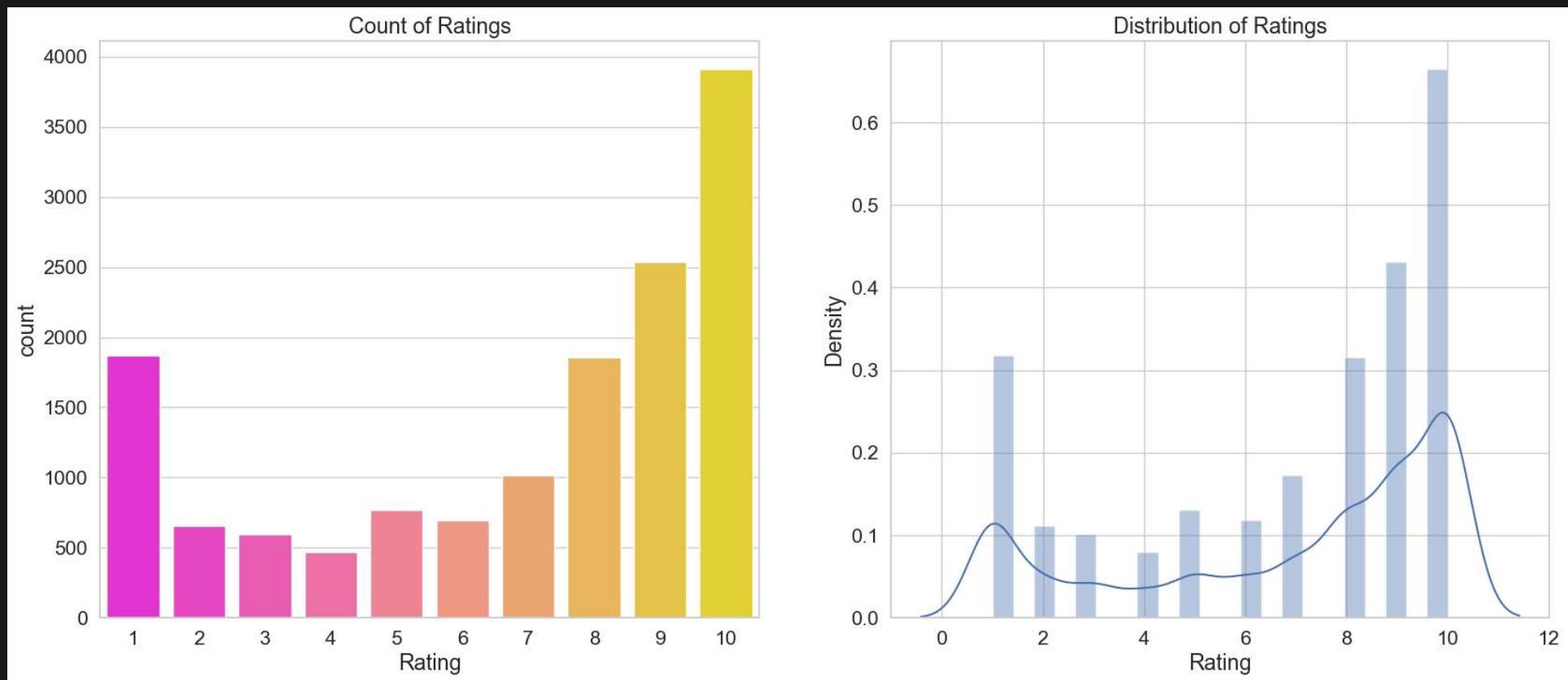
Null values finding

```
df_data['condition'].isnull().sum(axis = 0)
```

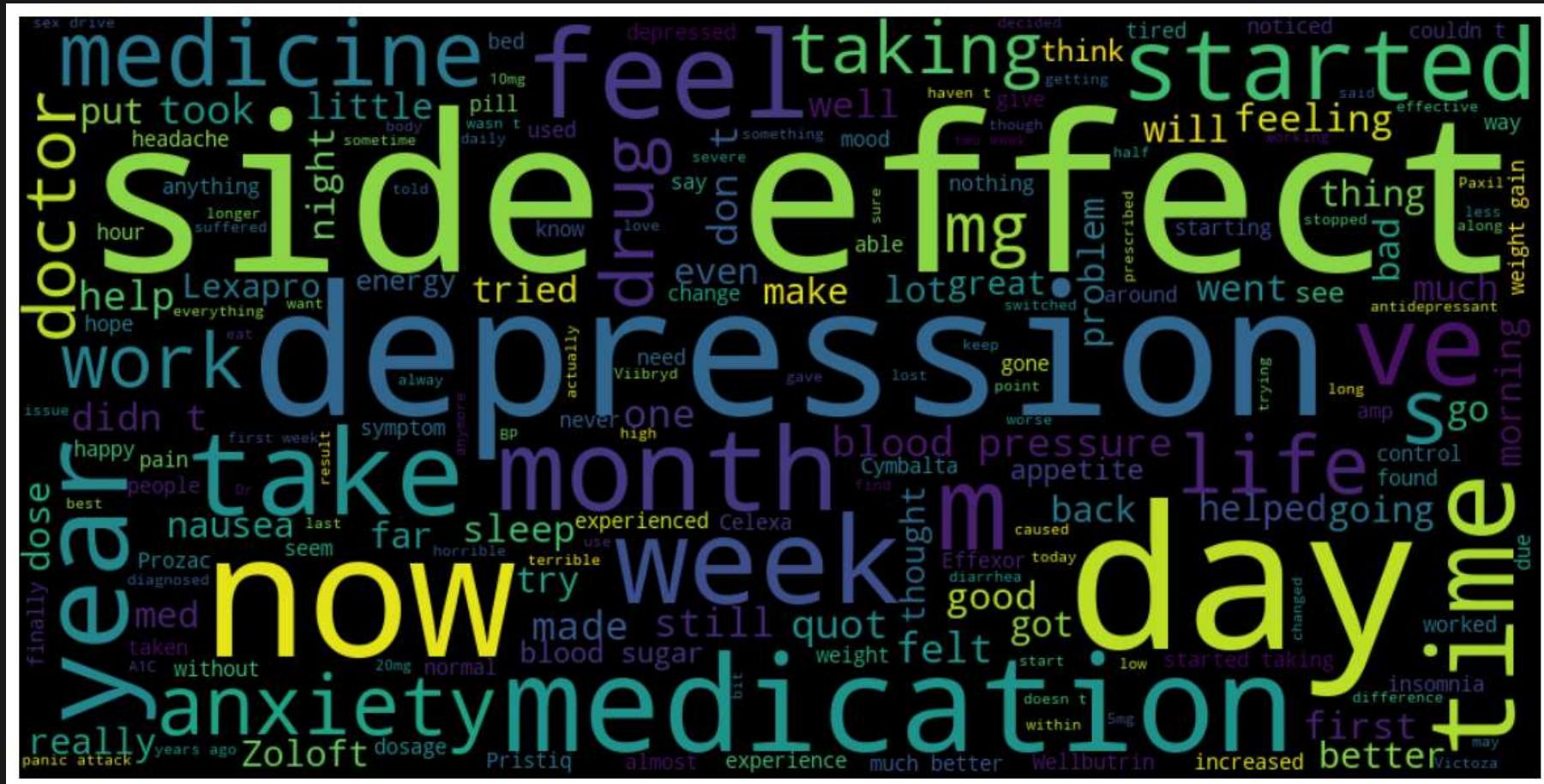
0 null values

0

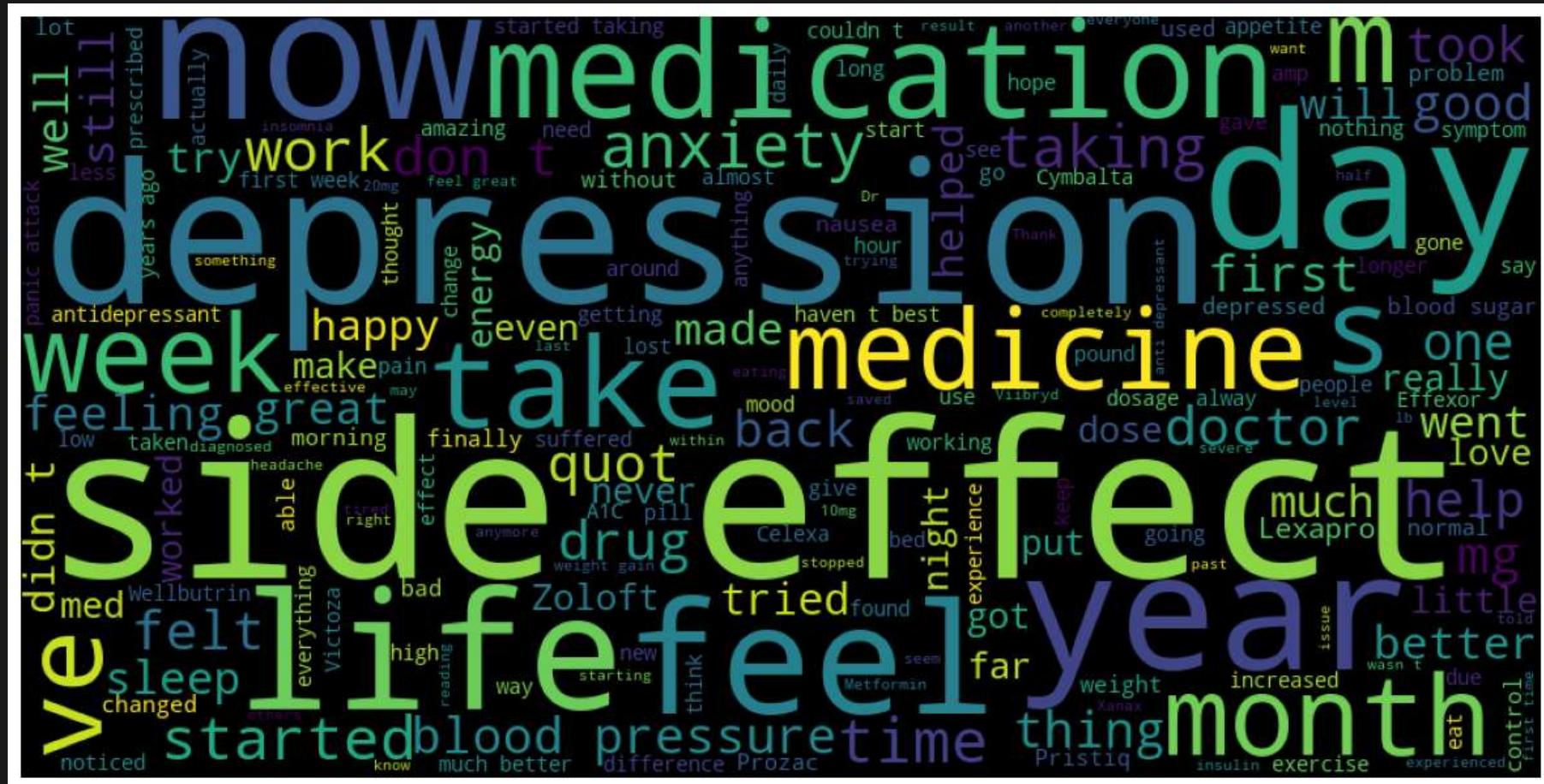
# Bar graph of count of ratings



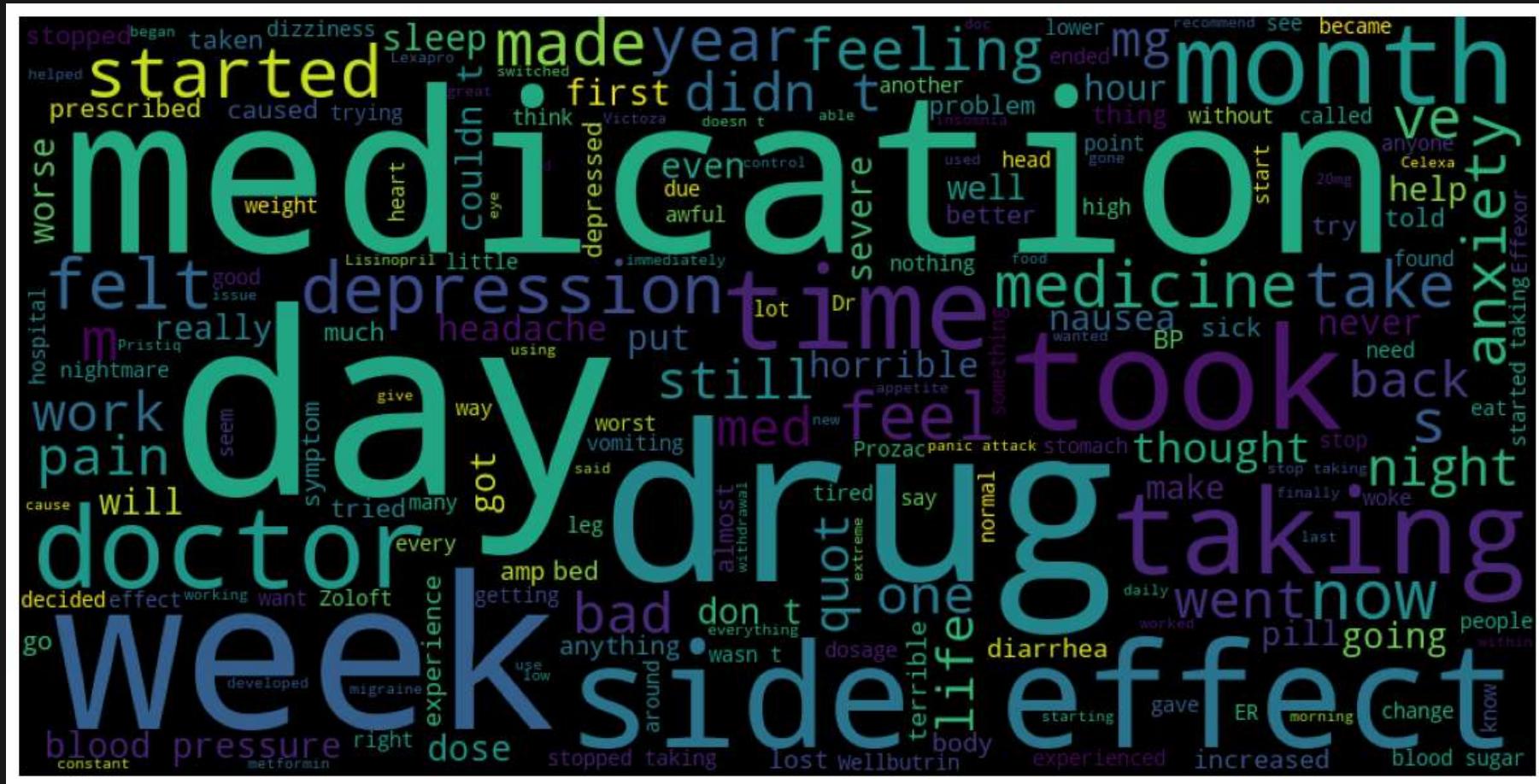
# Word Cloud for all the reviews in the data set.



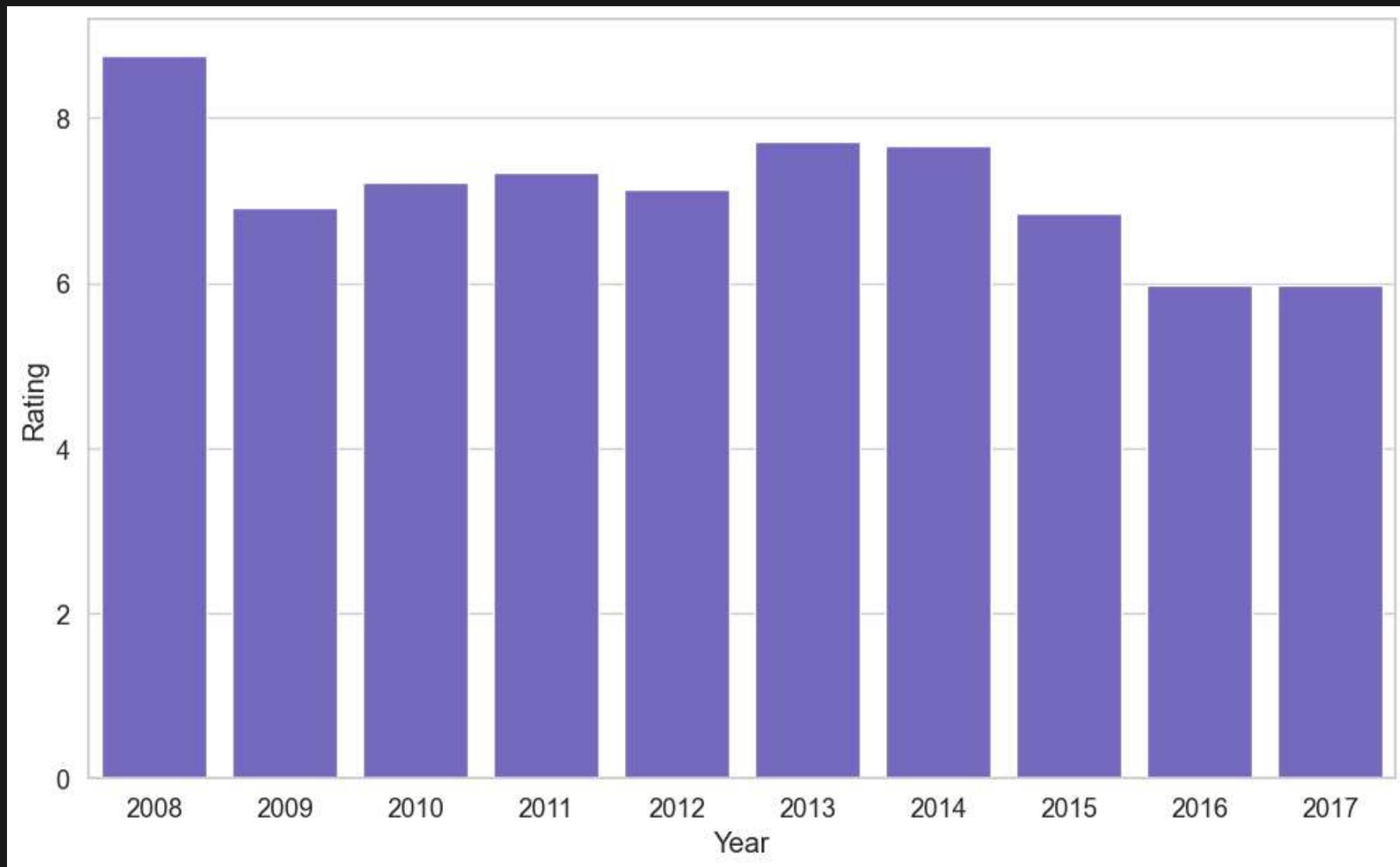
# Word cloud of the reviews with rating equal to 10



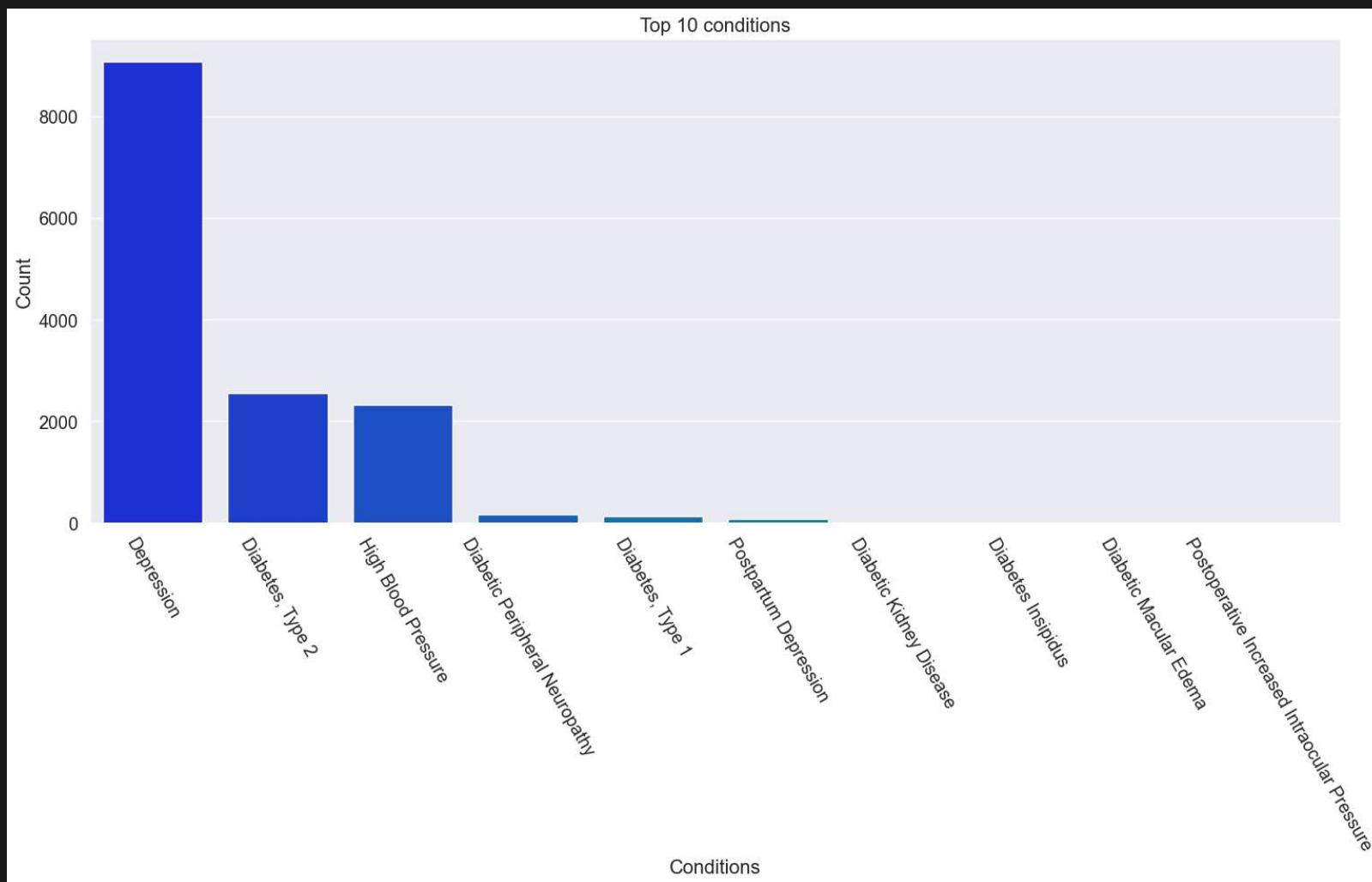
# Word cloud of the reviews with rating equal to 1



This barplot shows the mean rating per year

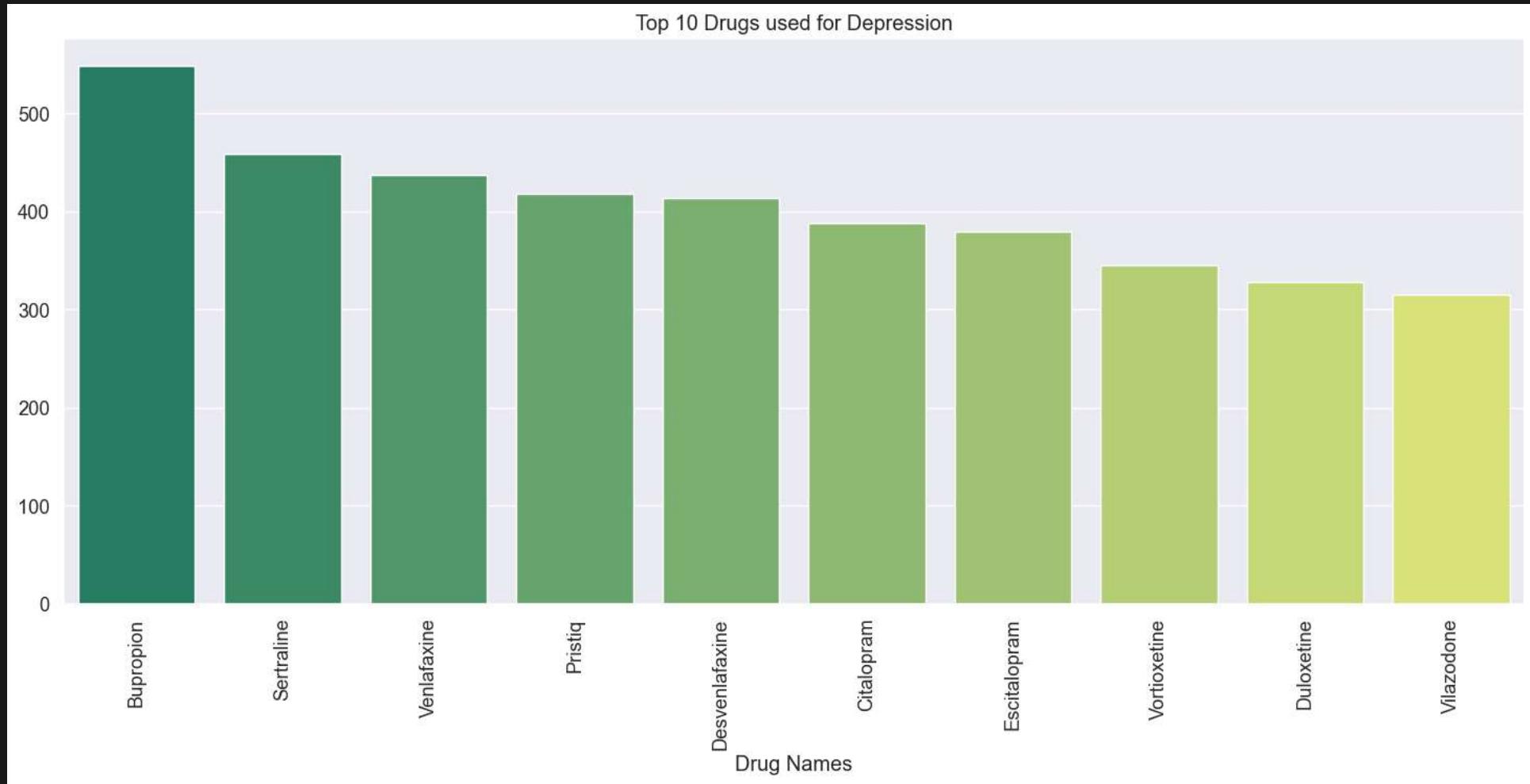


# Top 10 conditions the people are suffering.

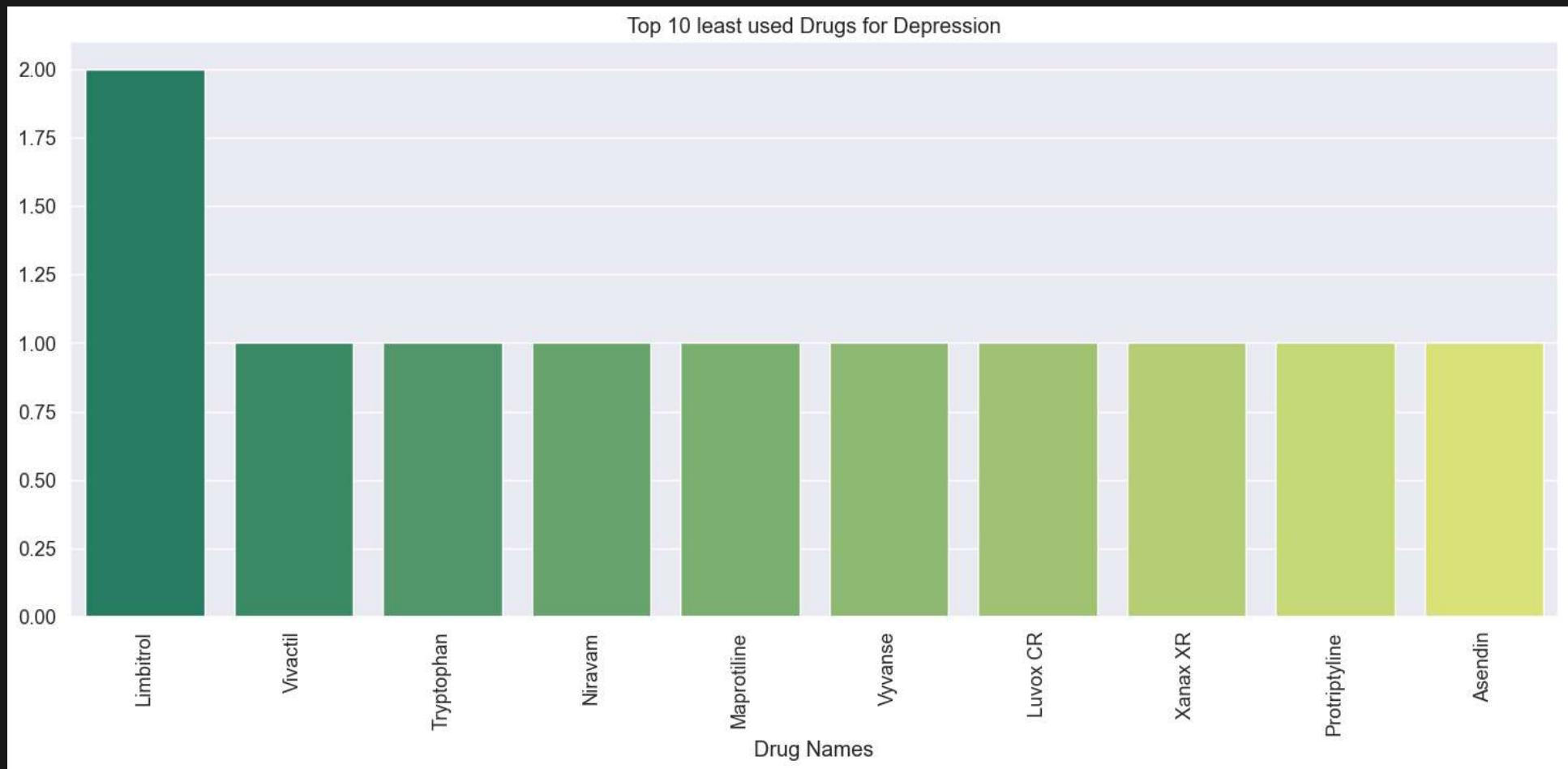


# DEPRESSION ANALYSIS

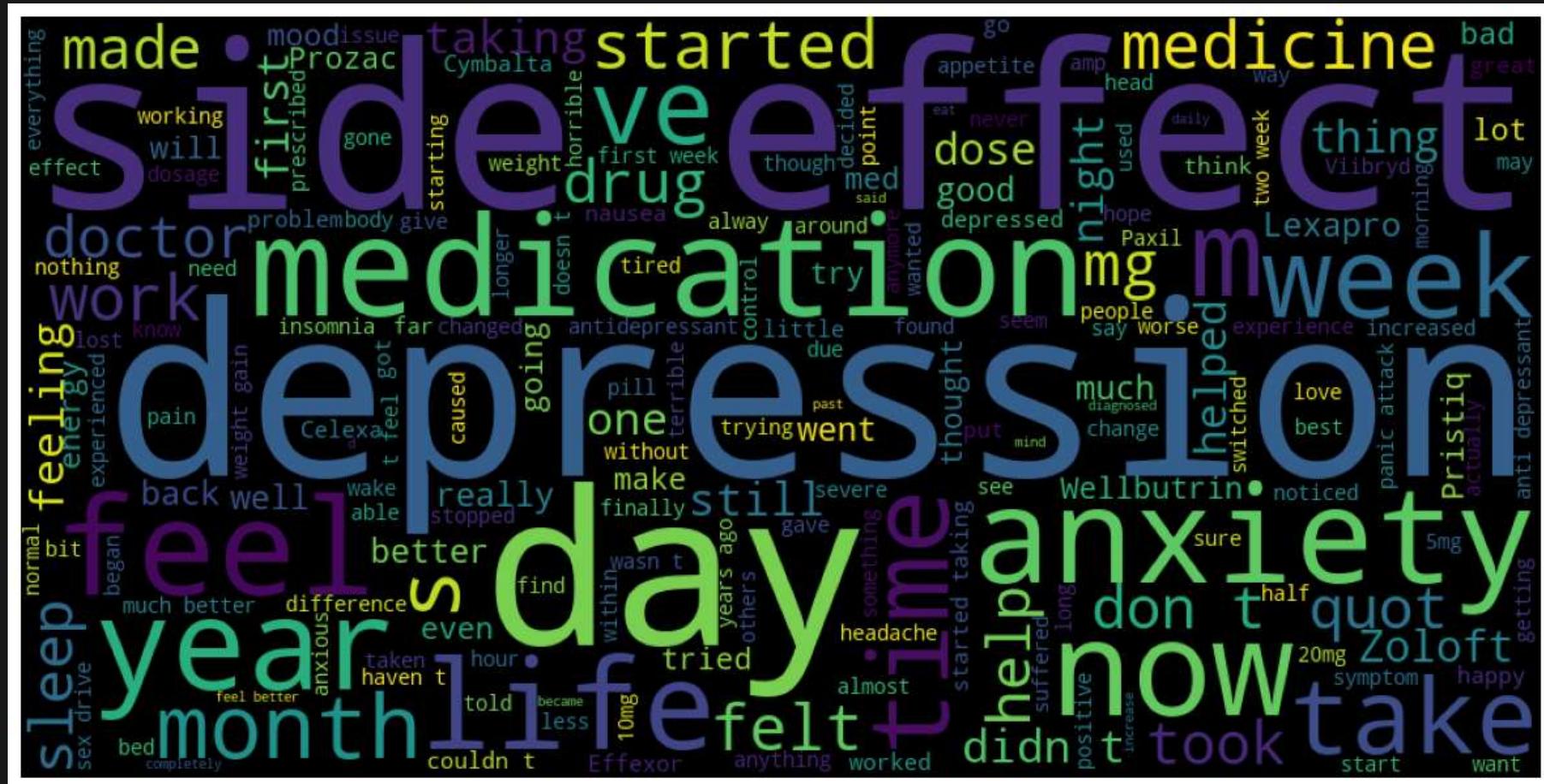
# Top 10 Drugs for Depression



# Top 10 Least useful Drugs for Depression based on ratings

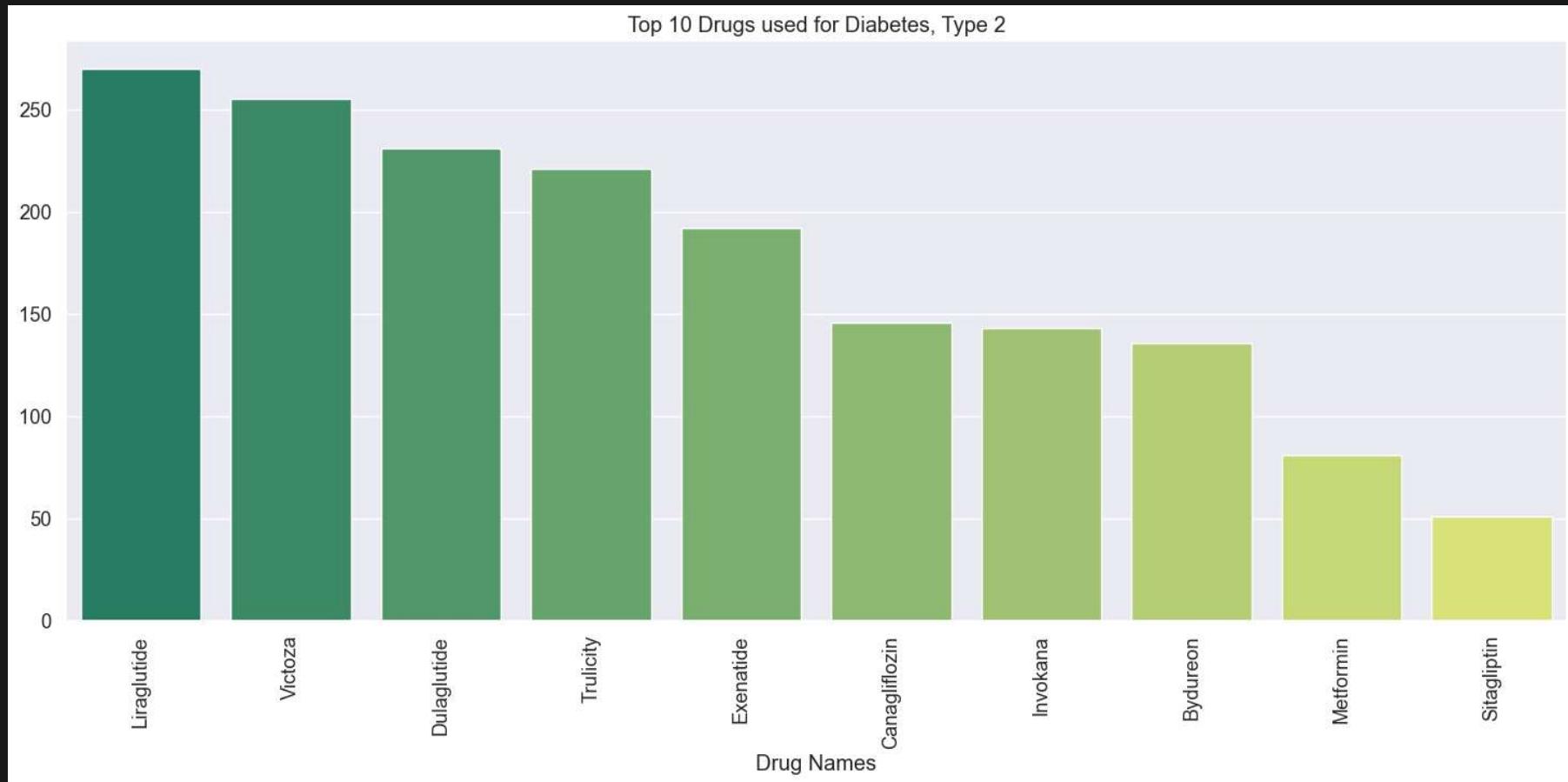


# Word cloud of reviews of Depression

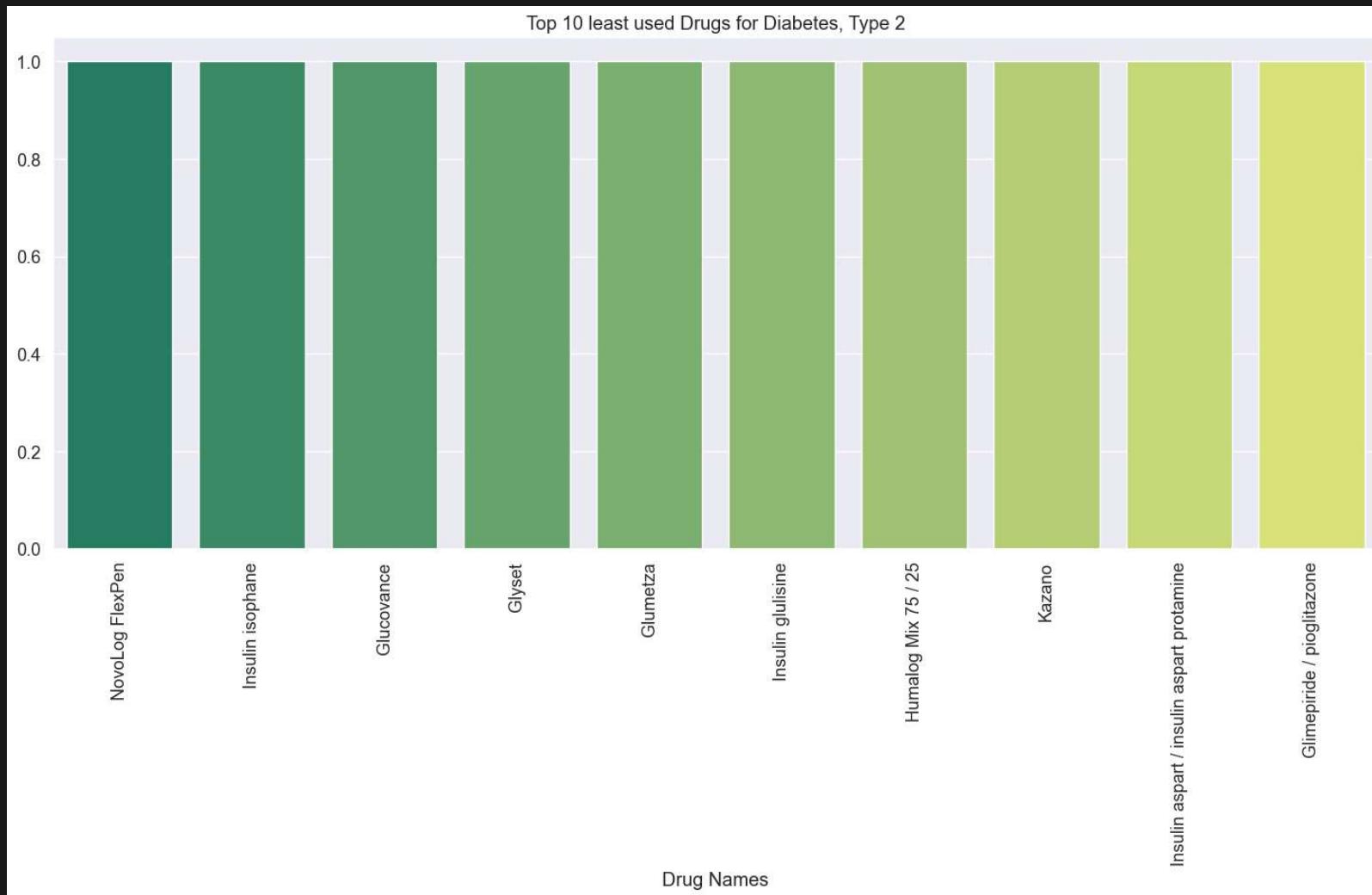


# DIABETES TYPE 2 ANALYSIS

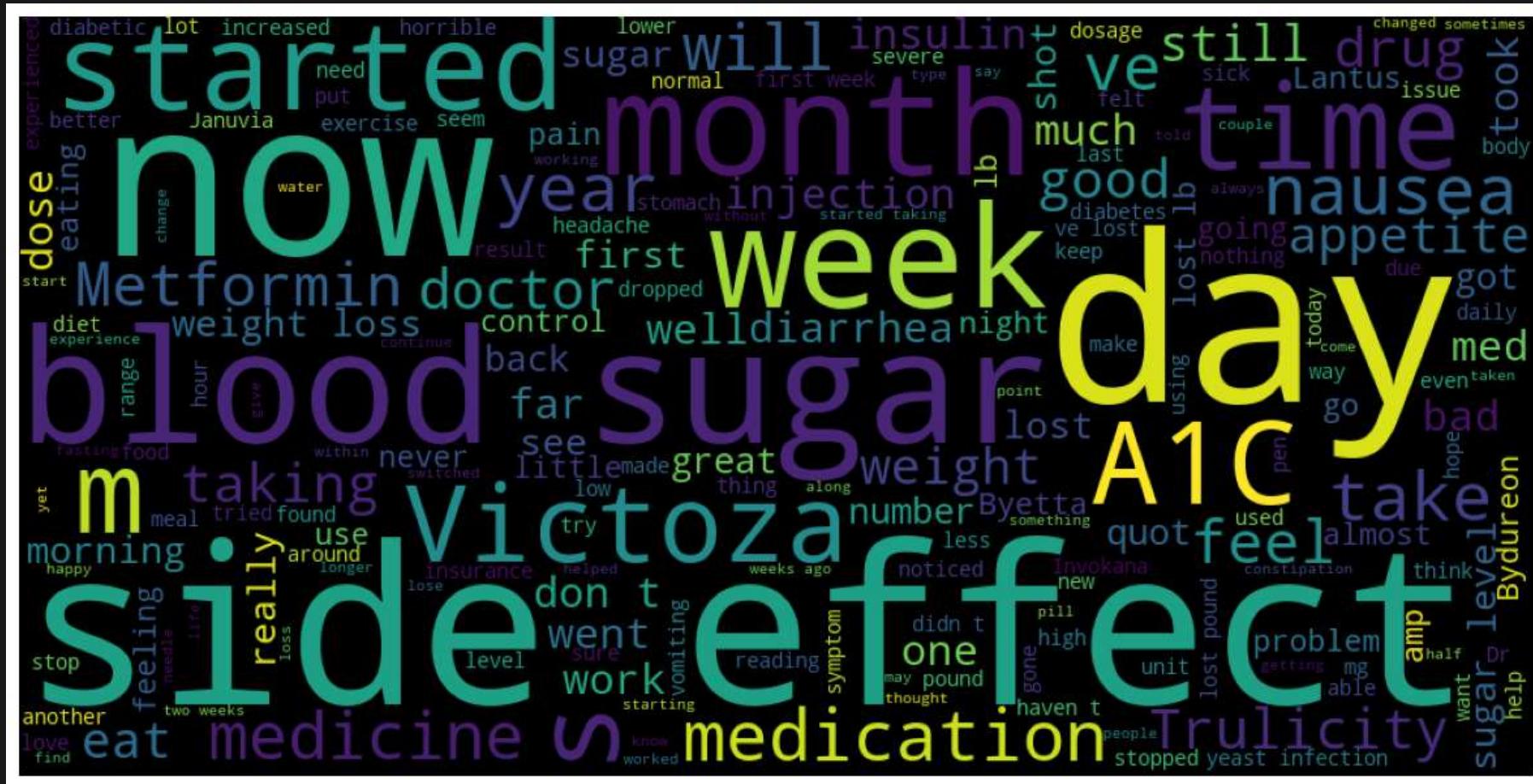
# Top 10 Drugs for Diabetes Type 2



# Top 10 Least useful Drugs for Diabetes type 2 based on ratings

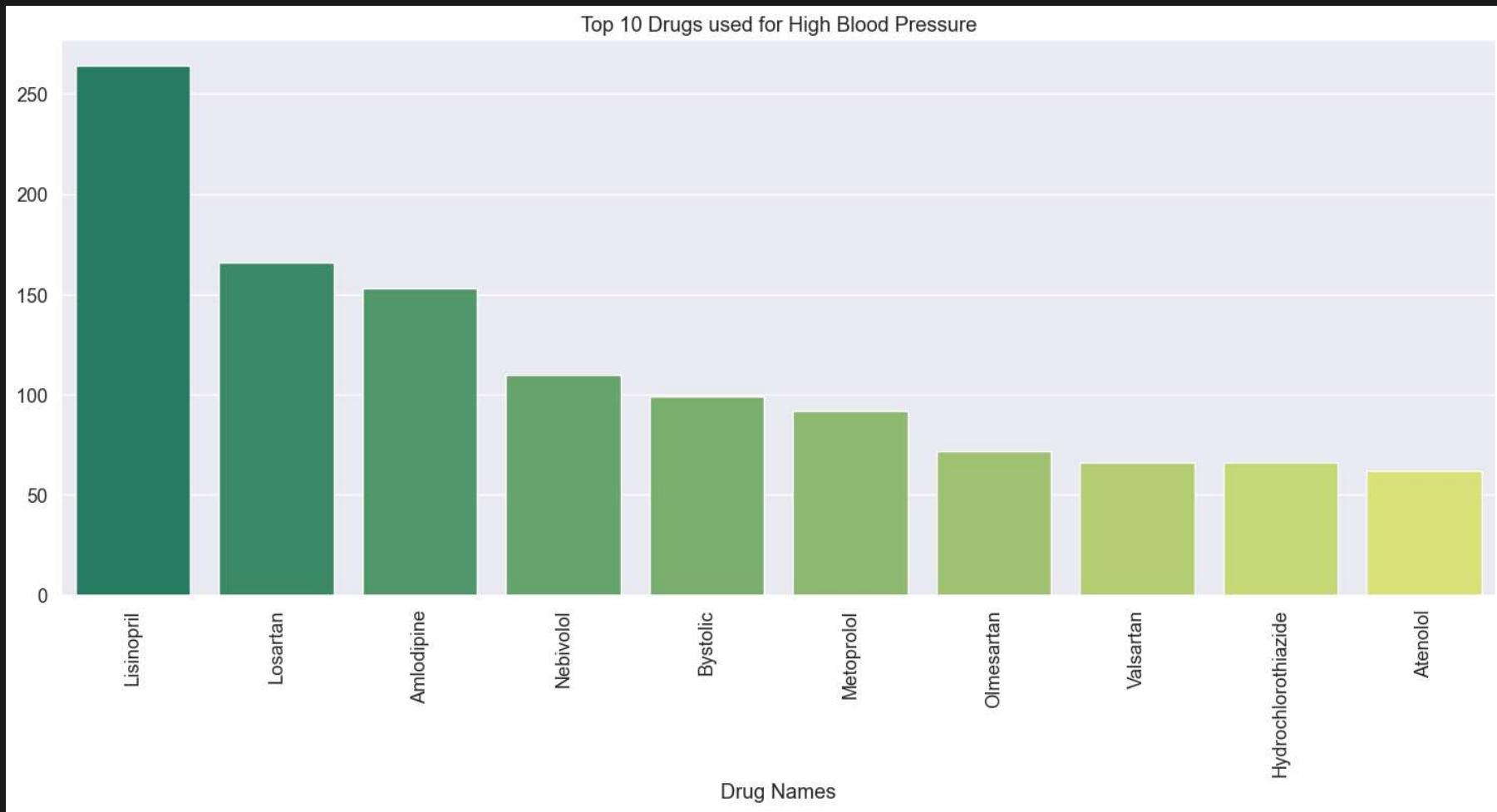


# Word cloud of reviews of Diabetes Type 2

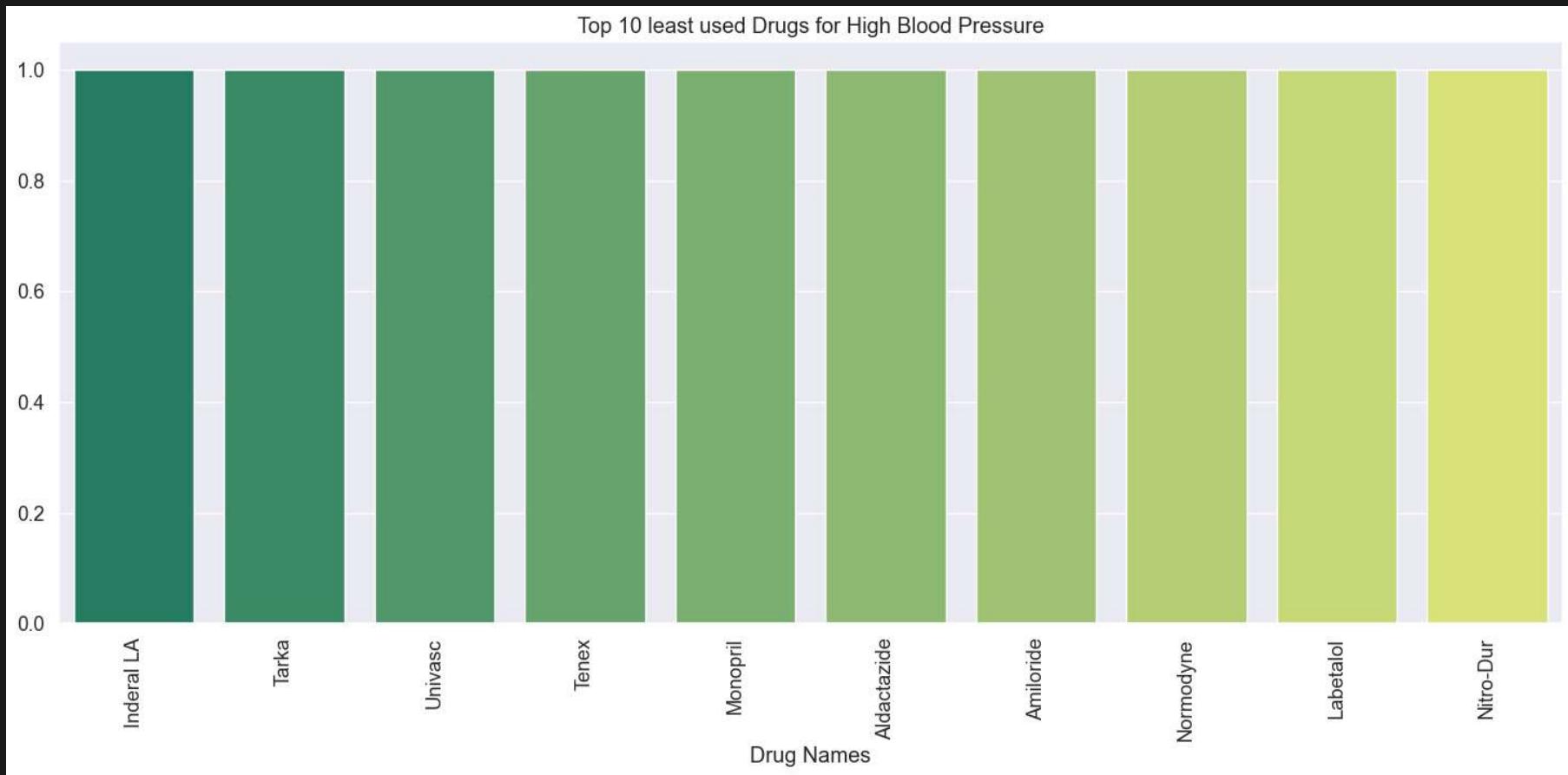


# HIGH BLOOD PRESSURE ANALYSIS

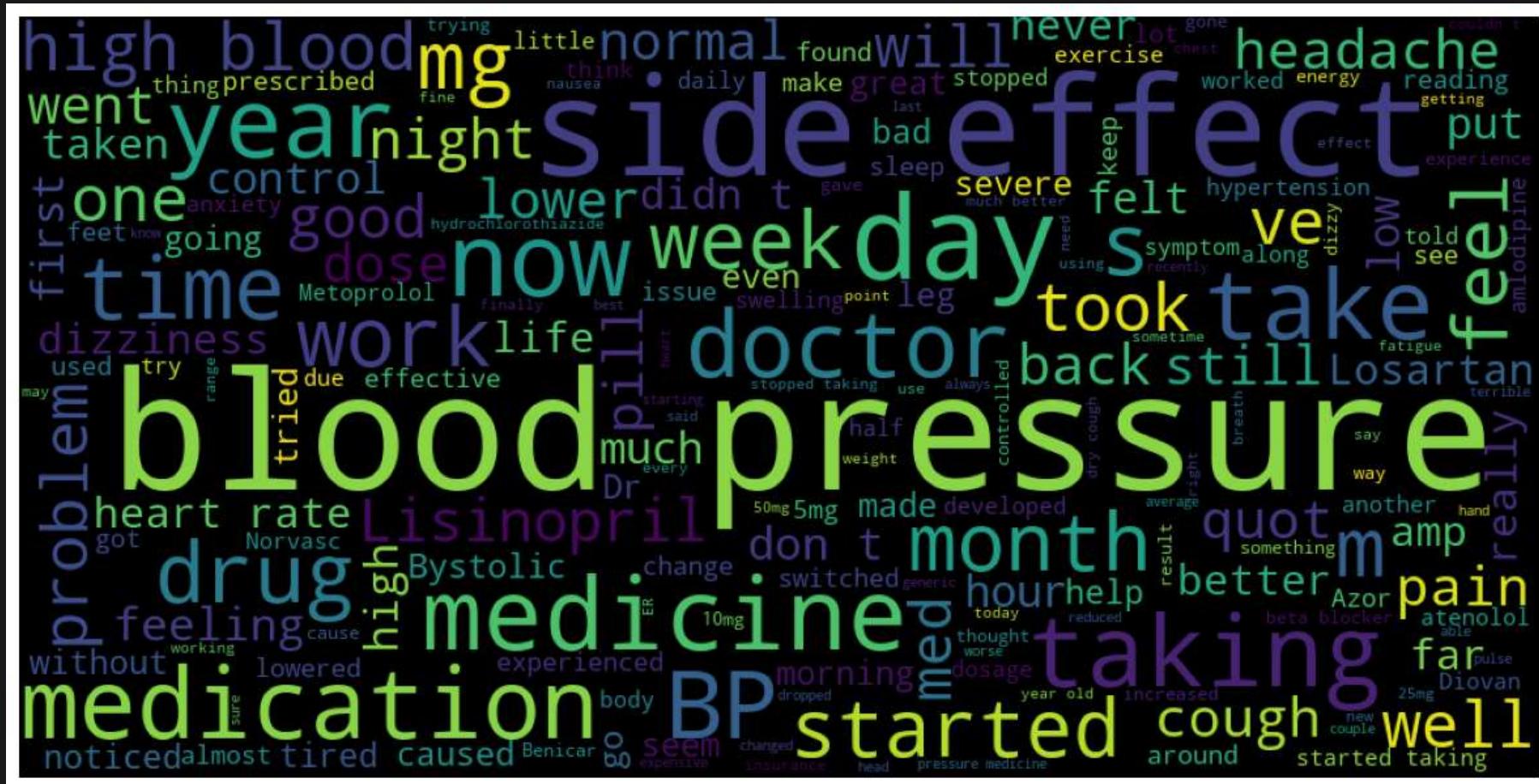
# Top 10 Drugs for High Blood Pressure



# Top 10 Least useful Drugs for High Blood Pressure based on ratings

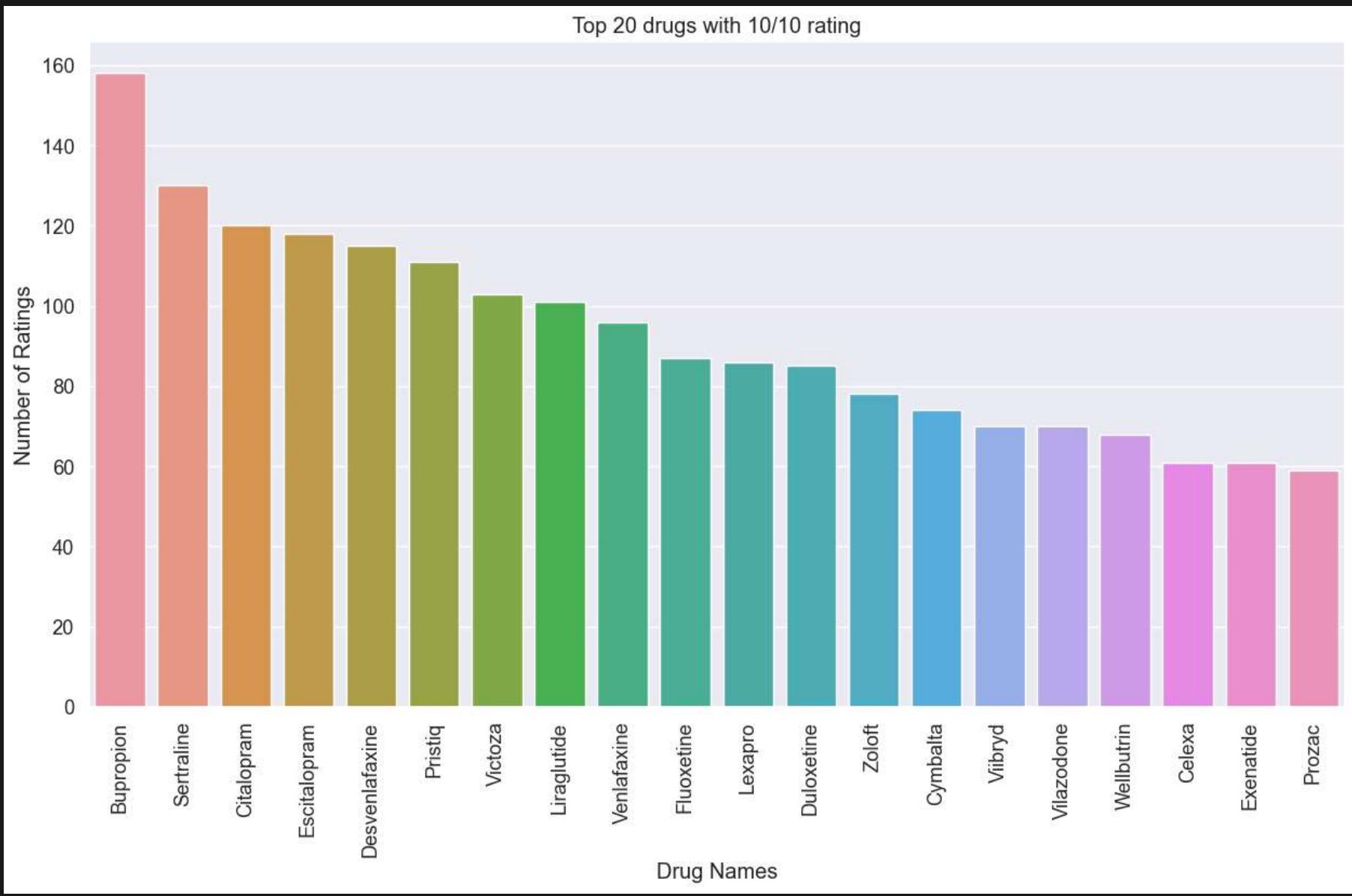


# Word cloud of reviews of High Blood Pressure

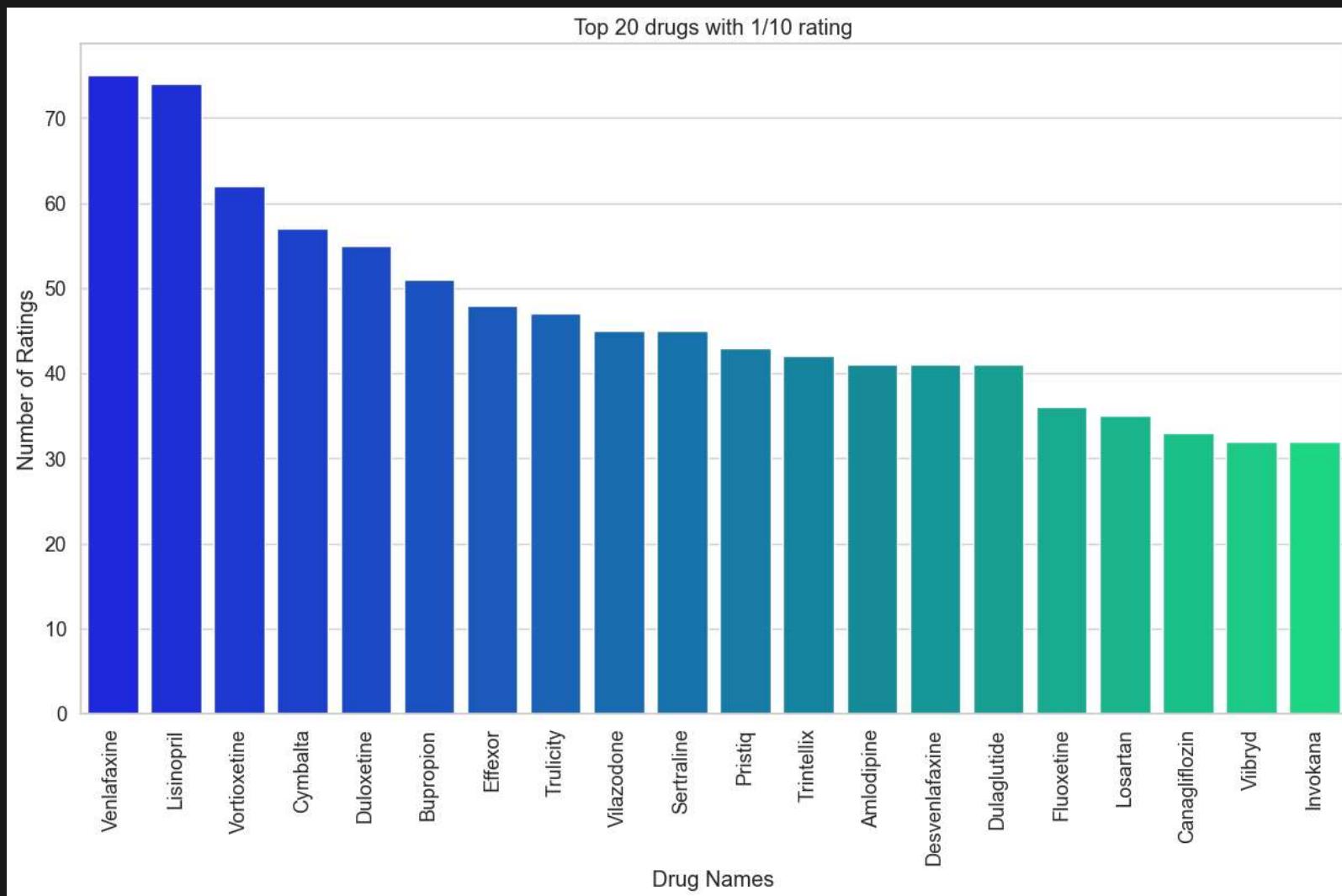


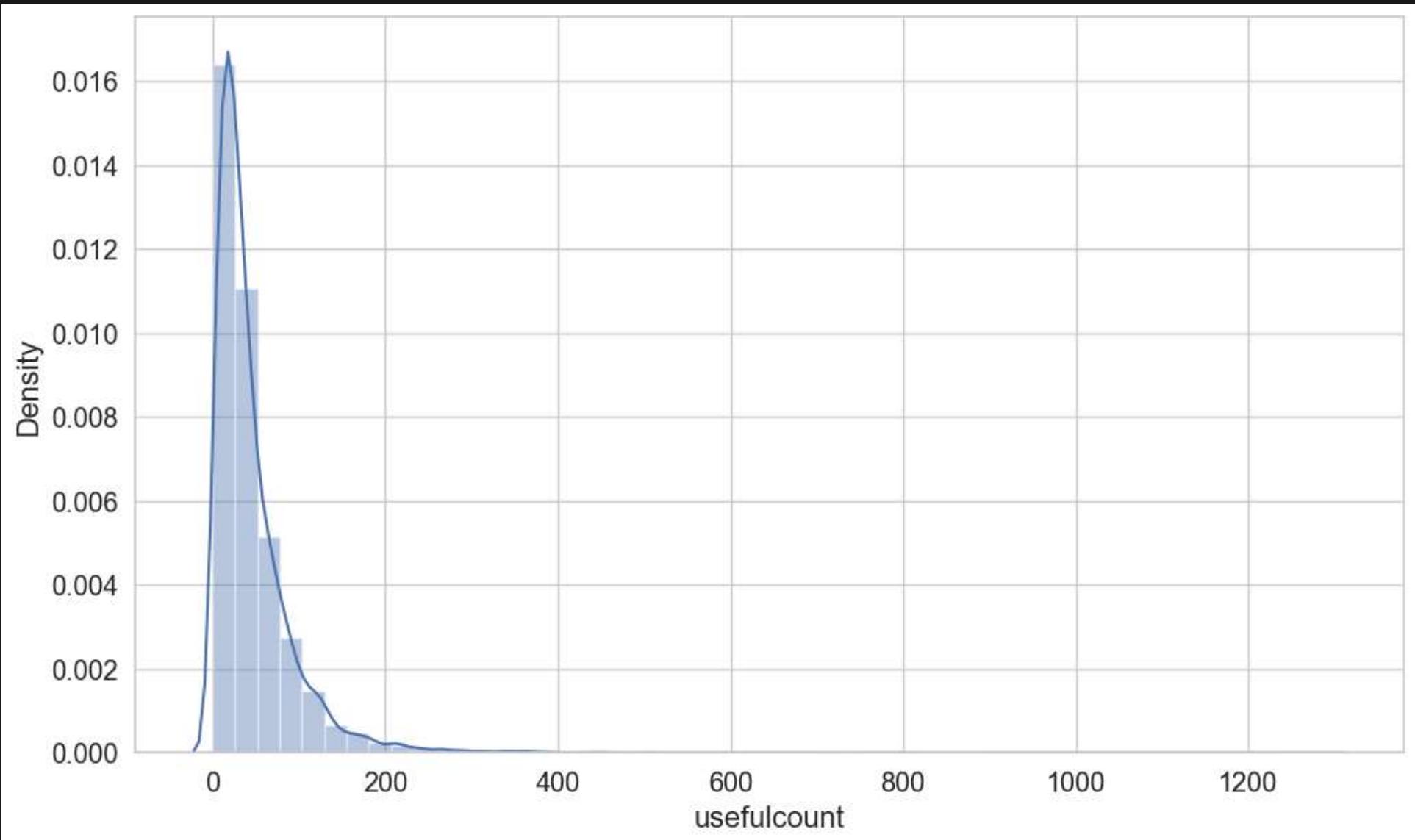
# DRUGS ANALYSIS

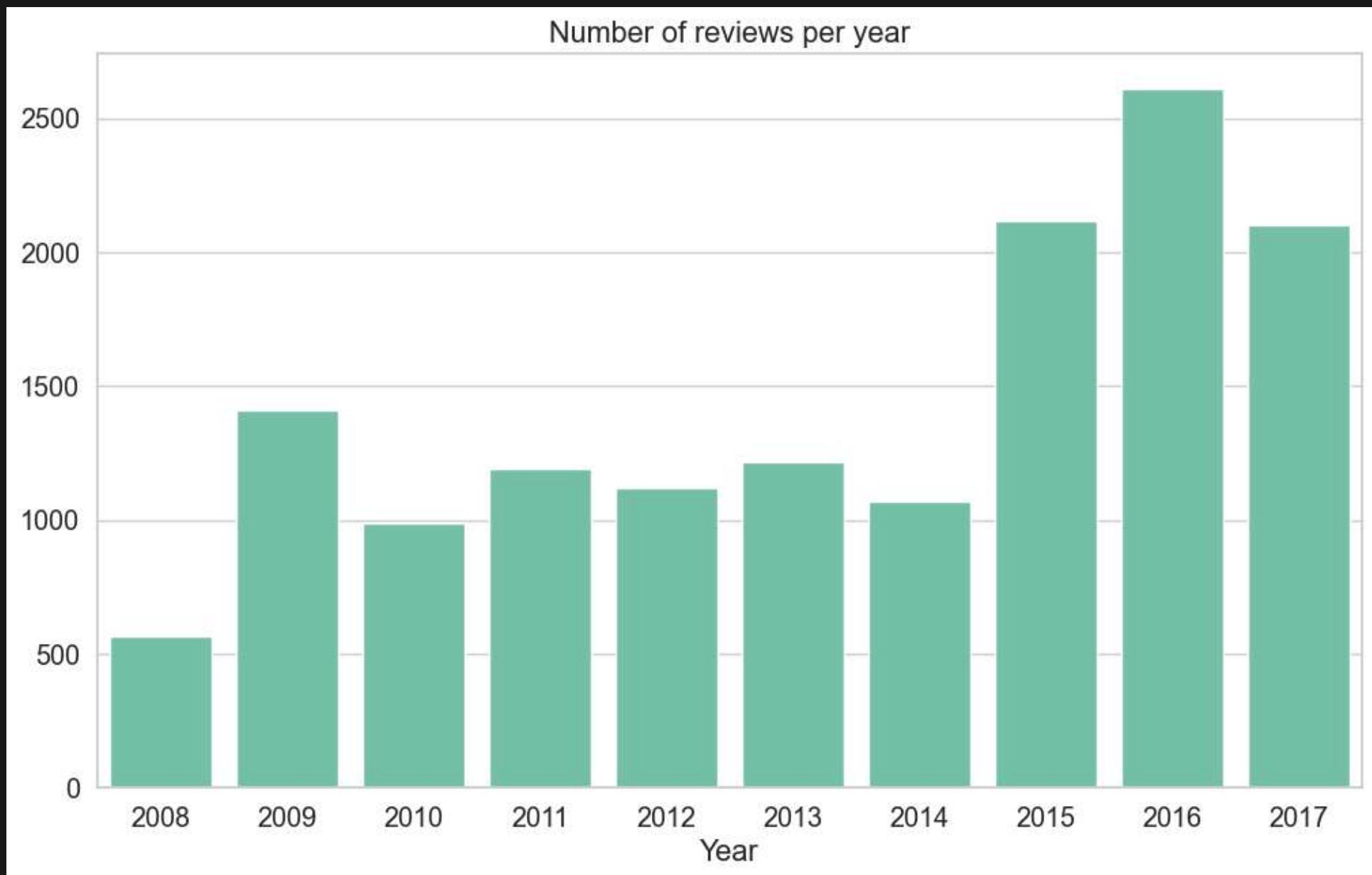
# Top 20 drugs with Maximum 10 ratings

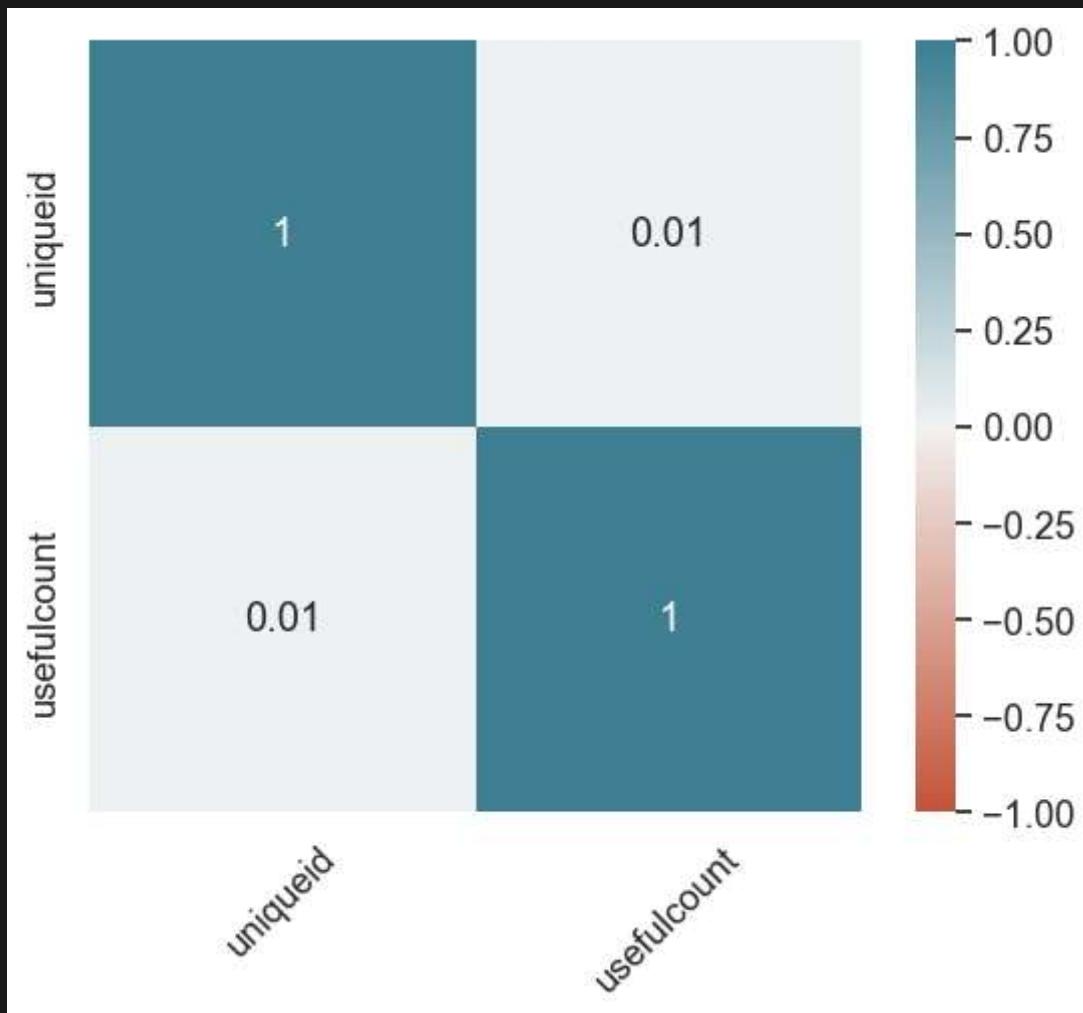


# Top 20 drugs with Maximum 1 ratings

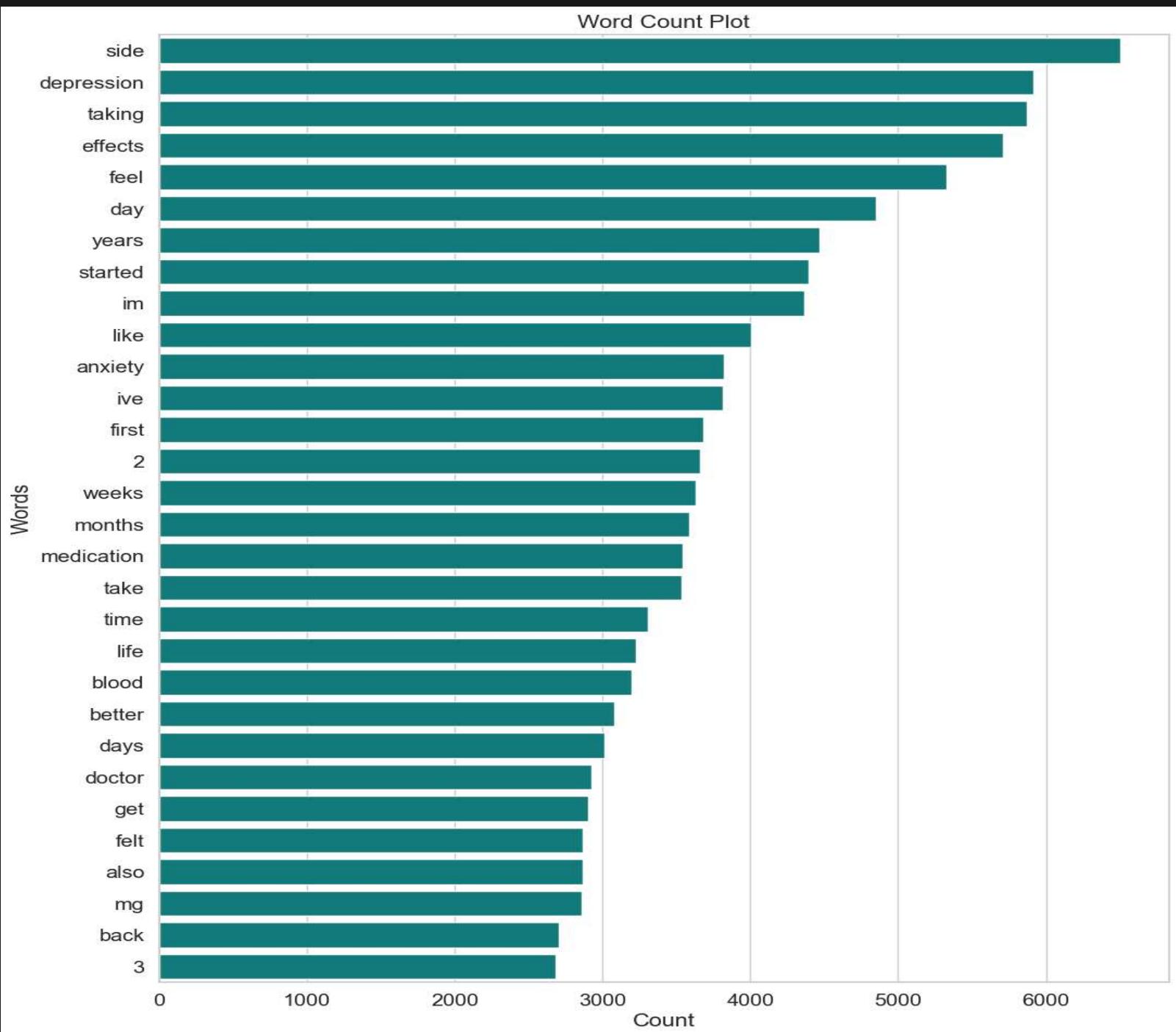




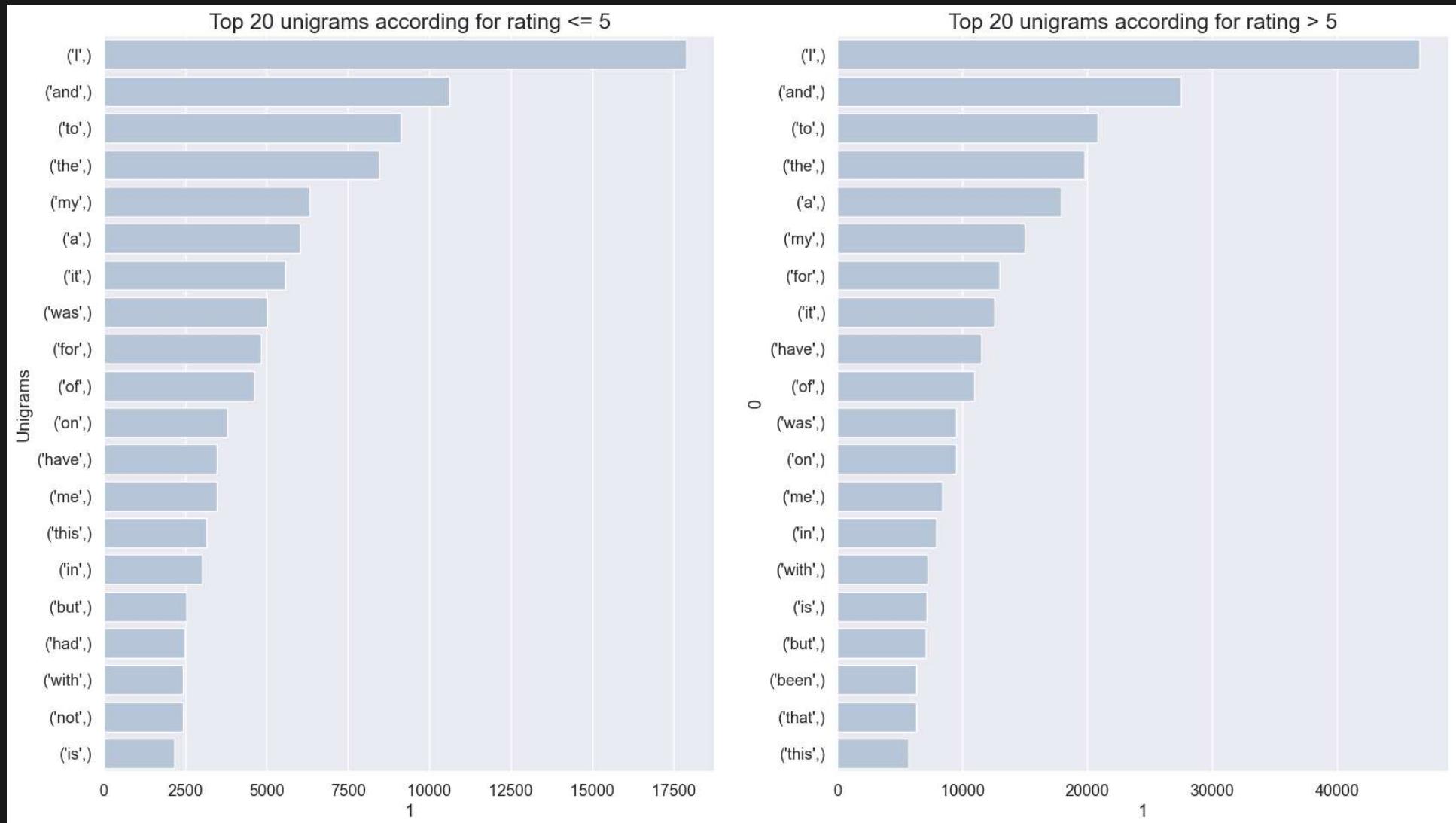




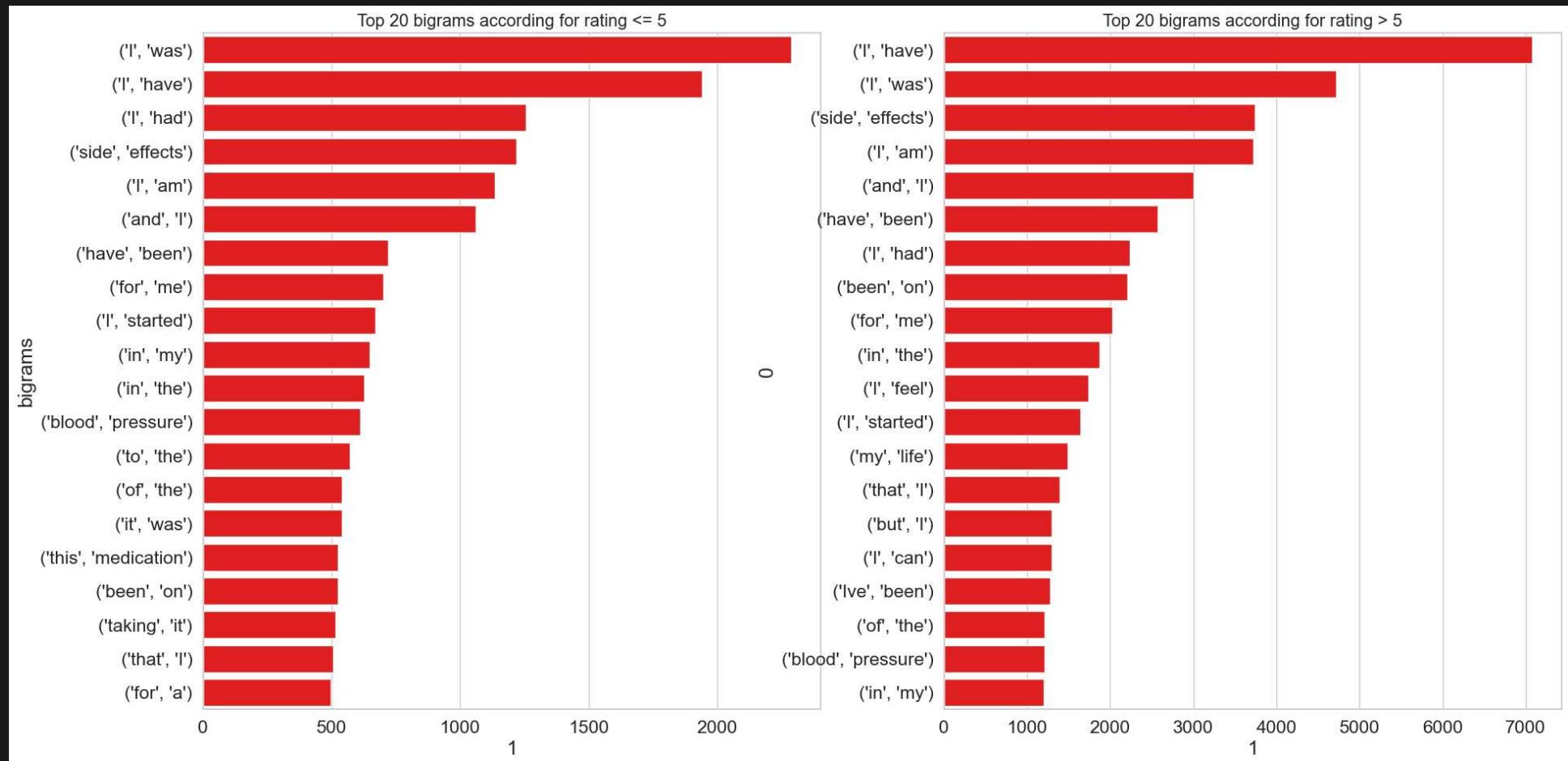
# REVIEWS WORDS ANALYSIS



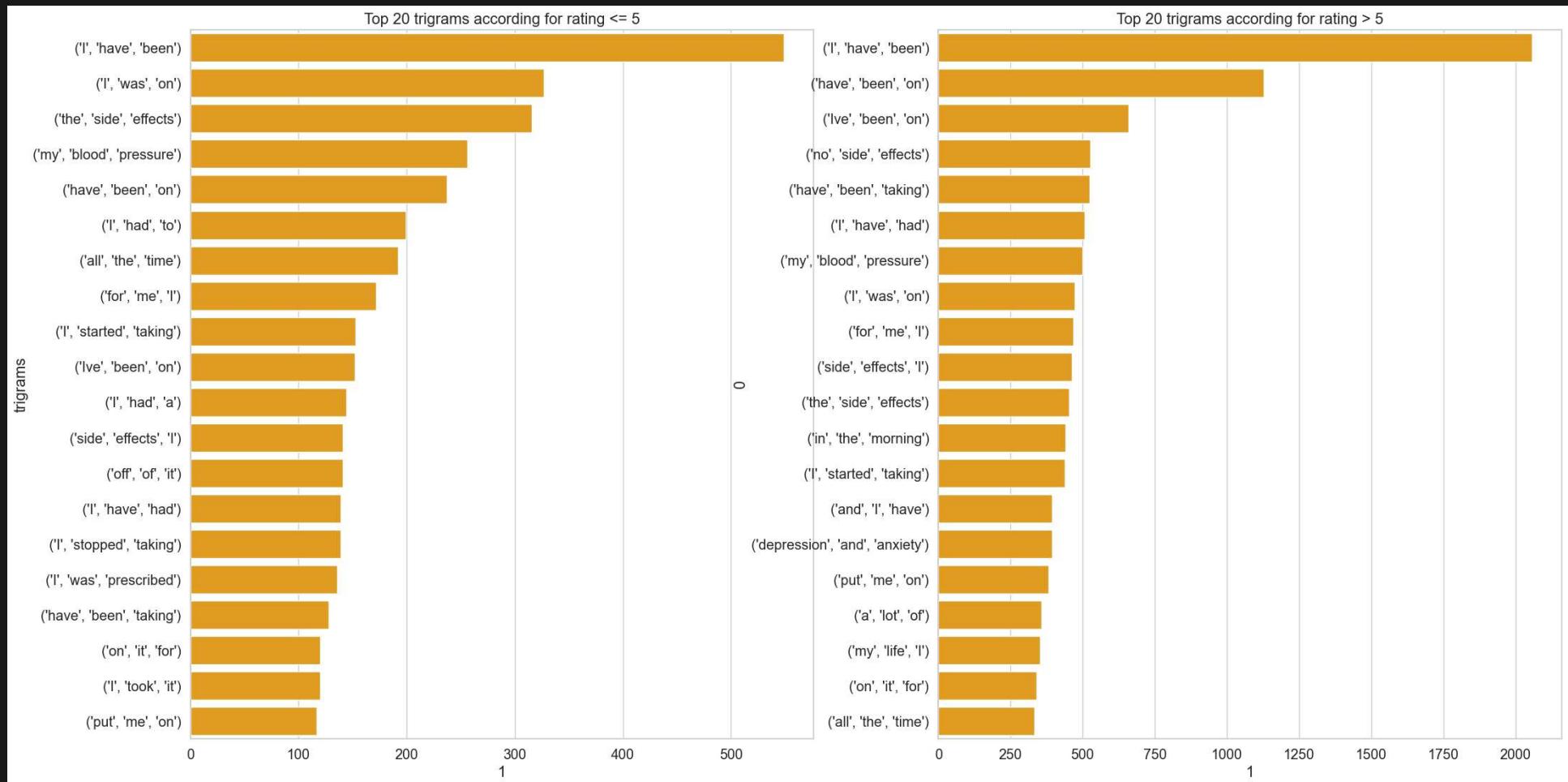
# Unigram Frequency



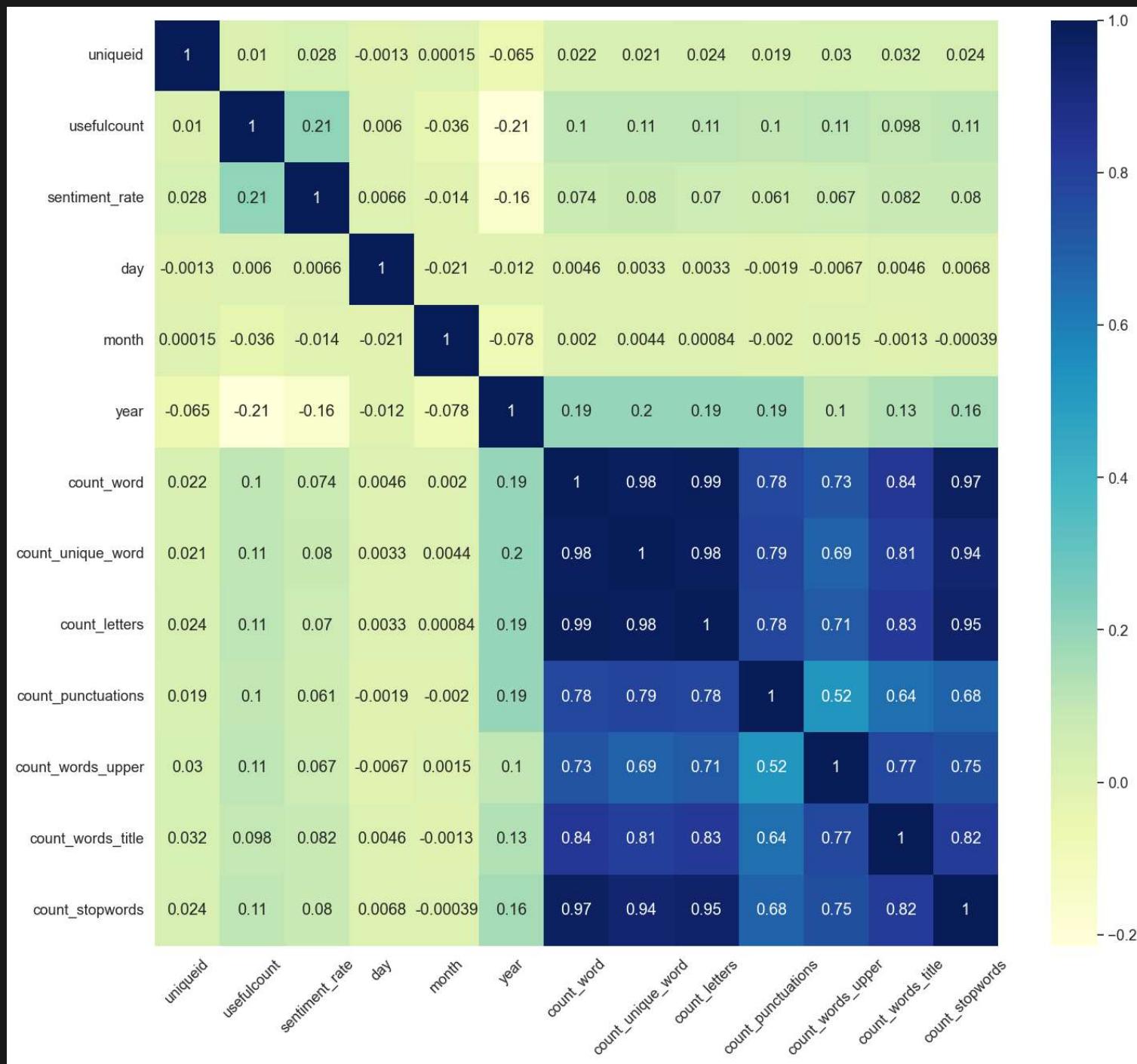
# Bigram Frequency



# Trigram Frequency



# FEATURE ENGINEERING



## BUSINESS OBJECTIVE:

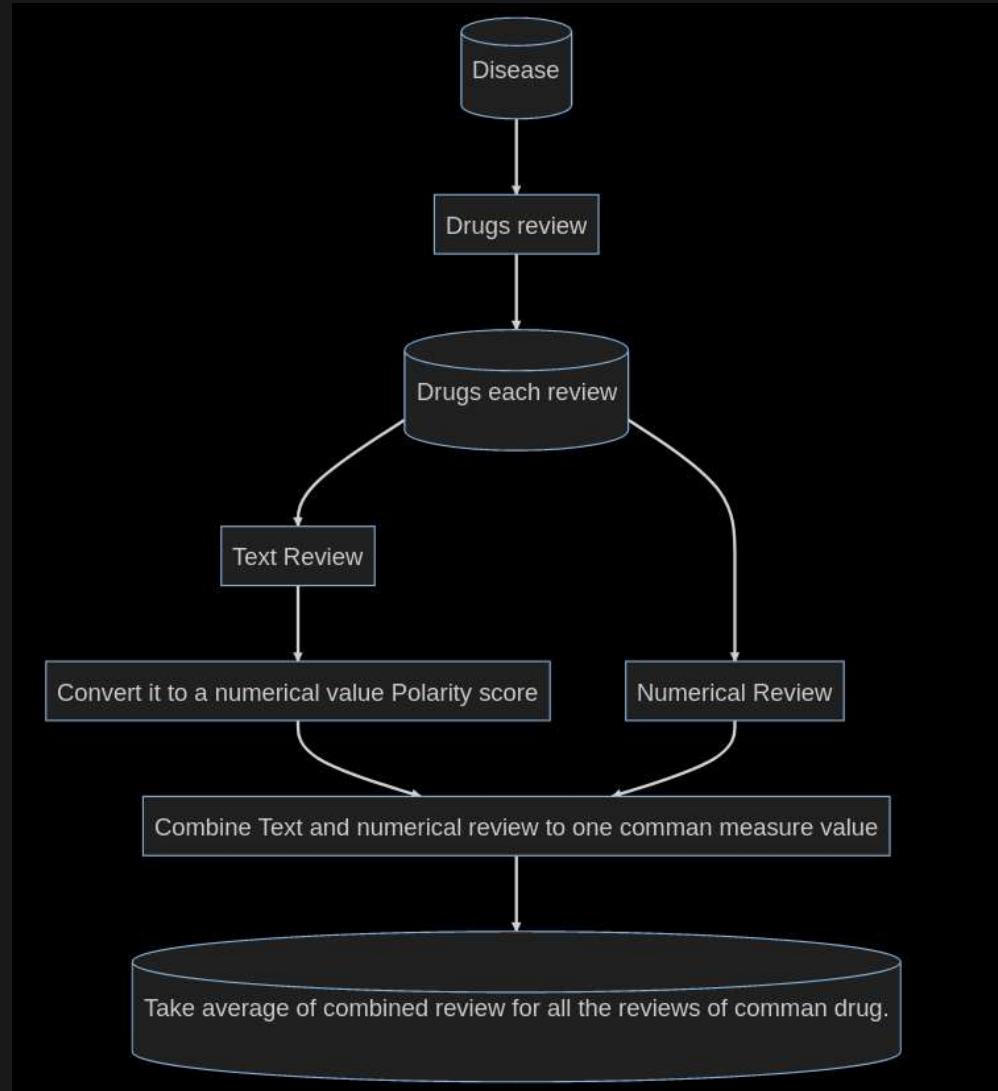
This is a sample dataset which consists of 161297 drug name, condition reviews and ratings from different patients and **our goal is to examine how patients are feeling using the drugs their positive and negative experiences so that we can recommend him a suitable drug. By analyzing the reviews, we can understand the drug effectiveness and its side effects.**

## In Review Cleaning:

- 1. changing to lower case
- 2. Replacing the repeating pattern of '
- 3. Removing all the special Characters
- 4. Removing all the non ASCII characters
- 5. Removing the leading and trailing  
Whitespaces
- 6. Replacing multiple Spaces with Single Space
- 7. Replacing Two or more dots with one

# MODEL 1 FOR ONLY DRUG RECOMMENDATION

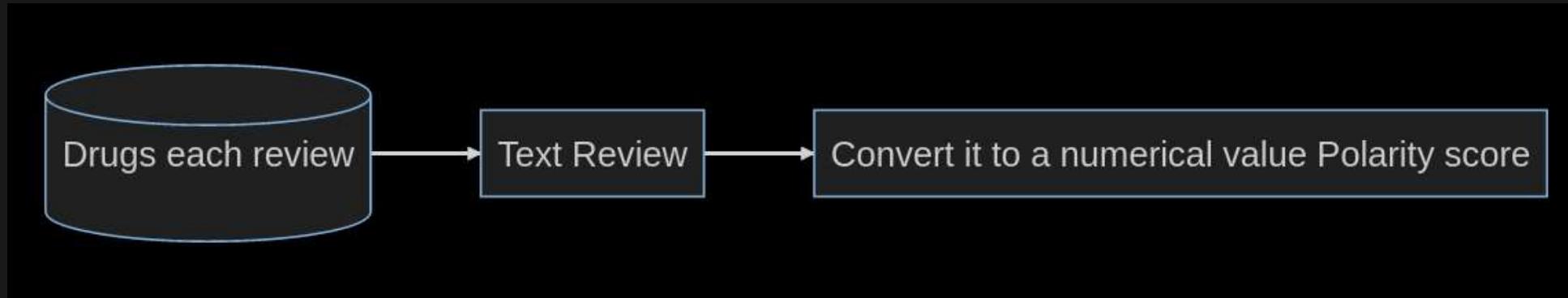
# Flow chart for drug efficacy assessment



# Polarity Score Assignment → Review



```
pol = []
for i in review:
    analysis = TextBlob(i)
    pol.append(analysis.sentiment.polarity)
return pol
```



So now assign each review a polarity score value from -1 to 1 :

- $-1 \rightarrow$  *Most - ve review* 
- $0 \rightarrow$  *Neutral review* 
- $1 \rightarrow$  *Most + ve review* 

# CONVERTING POLARITY SCORE VALUE FROM [ -1, 1 ] TO [ 1, 10 ] :

$(Polarity\ score) \times -5 + 6 = New\ score.$

# SCALING REVIEW SCORE TO RATING SCORE (NEW SCORE)

- 1 → *Most – ve review* 
- 5 → *Neutral review* 
- 10 → *Most + ve review* 

Developing a new composite score which takes both rating and polarity score weight-age equally.

$$\frac{rating + new\ score}{2} = cumulative$$

Taking the average of composite score for command drugs and base on the disorder we rank the drugs which had the highest score as the best recommendation for the patient with the disorder.

# WEB APP DRUG SUGGESTION (NO REVIEW)

In this app diagnosis is supposed to be provided by the user

## DEPLOYMENT CODE EXECUTION :

```
streamlit run {filename}.py
```

# Best Drug Suggestion for Medical disorder

Select a condition:

Depression



Enter the number of top drugs you want to know for the disorder:

5



Submit

# Best Drug Suggestion for Medical disorder

Select a condition:

Depression

Depression

Type 2 Diabetes

High Blood Pressure

Submit

# Best Drug Suggestion for Medical disorder

Select a condition:

Type 2 Diabetes



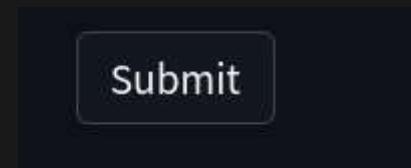
Enter the number of top drugs you want to know for the disorder:

8



Submit

BEHOLD THE GREAT SUBMIT BUTTON 🙌



Select a condition:

Type 2 Diabetes



Enter the number of top drugs you want to know for the disorder:

5

- +

Submit



	condition	drugname	cumulative
1	Diabetes, Type 2	Insulin glulisine	8.5417
2	Diabetes, Type 2	NovoLog FlexPen	9.1071
3	Diabetes, Type 2	Glimepiride / pioglitazone	9.25
4	Diabetes, Type 2	Glyset	9.25
5	Diabetes, Type 2	Humalog Mix 75 / 25	9.2625

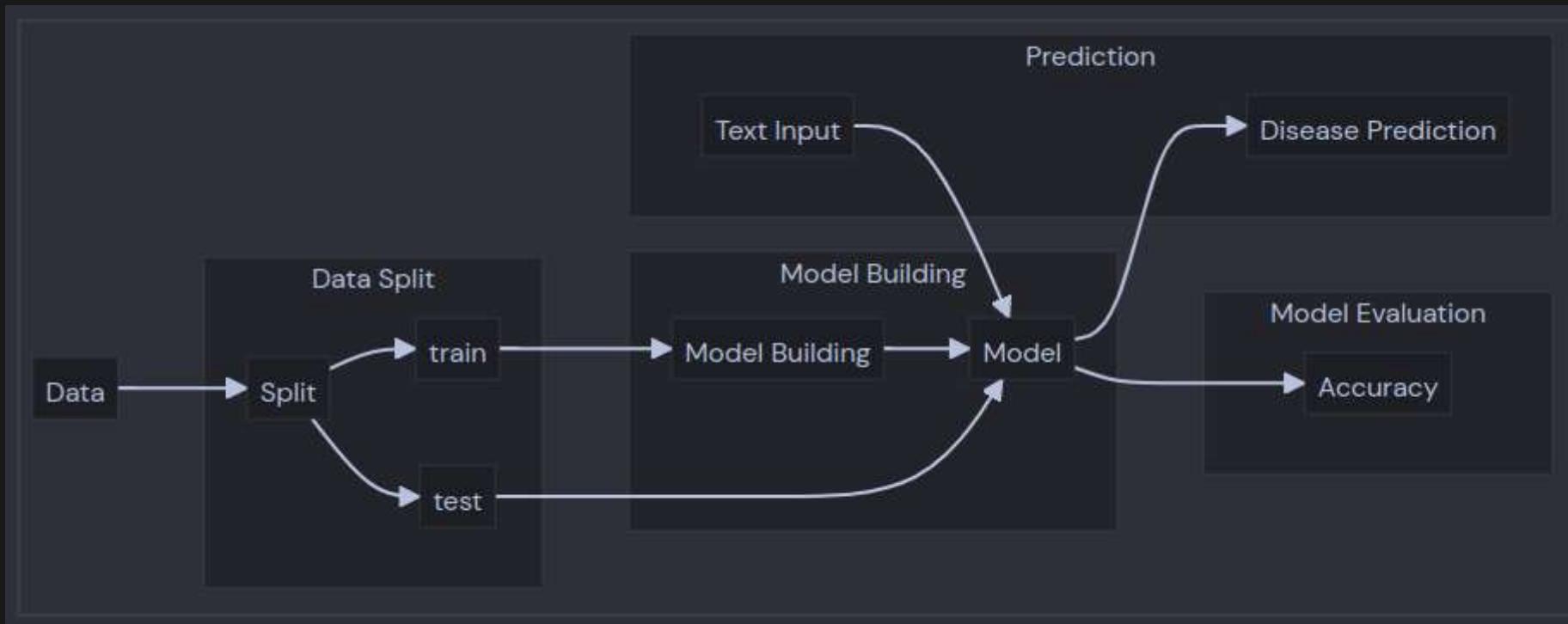
# MODEL 2 : REVIEW BASED DRUG PREDICTION



# MODELS EXPLORED

- Naives Bayes classifier
  - Count Vectorizer Model (Bag of Words)
  - TD-IDF Transformation
- Passive aggressive classifier
  - Count Vectorizer Model (Bag of Words)
  - TD-IDF Transformation
    - Monogram
    - Bigram
    - Trigram

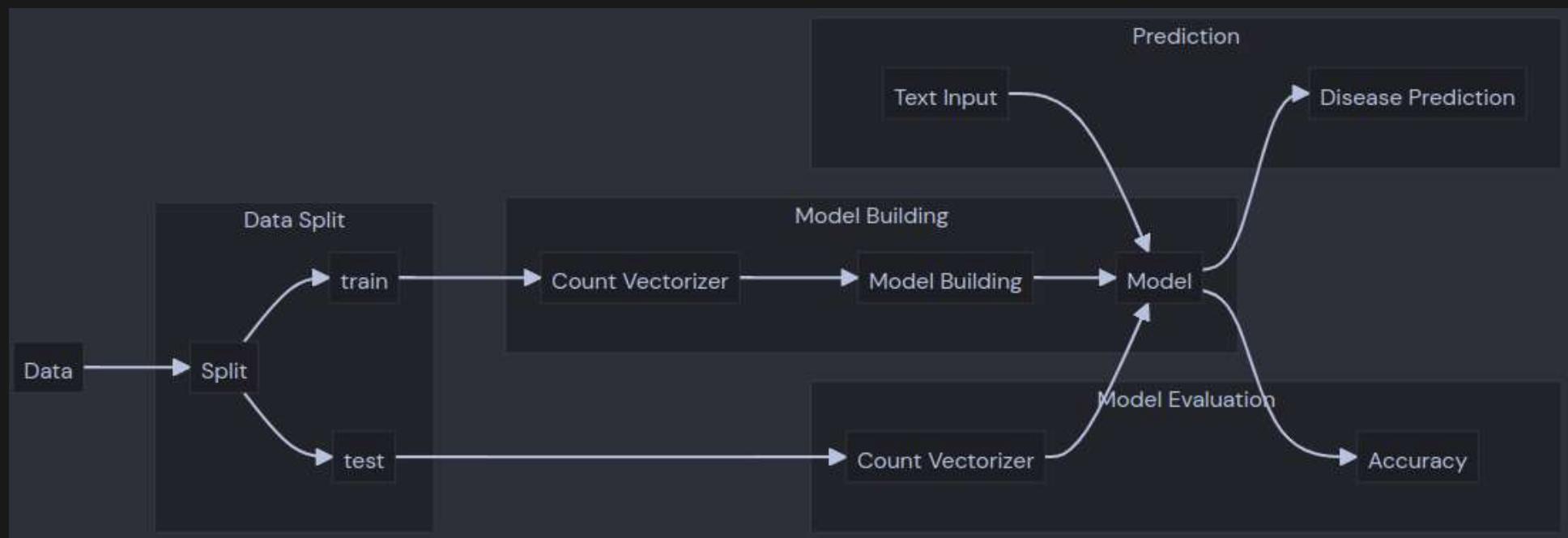
# MODEL BUILDING



## BAG OF WORDS → COUNTVECTORIZER

Document	the	cat	sat	in	hat	with
<i>the cat sat</i>	1	1	1	0	0	0
<i>the cat sat in the hat</i>	2	1	1	1	1	0
<i>the cat with the hat</i>	2	1	0	0	1	1

# Passive Aggressive Classifier & Naives Bayes Count Vectorizer:



BAG OF WORDS  → TF-IDF

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

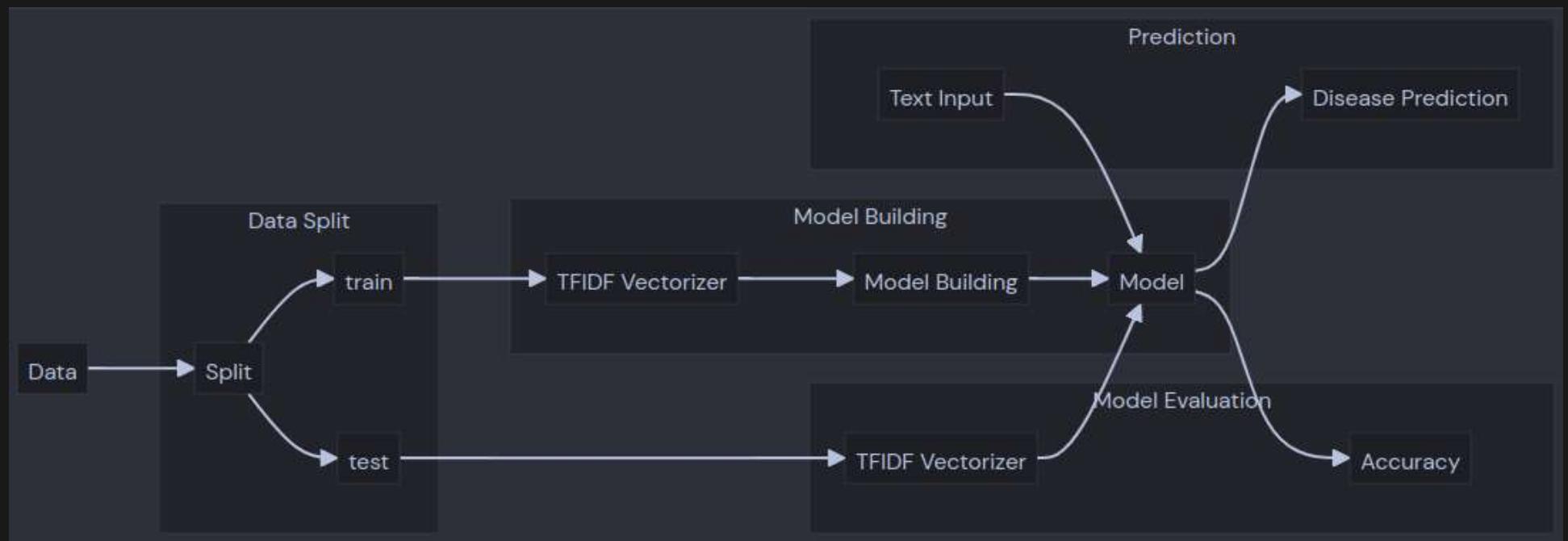
$$TF(t, d) = \frac{\text{number of times } t \text{ appears in document}}{\text{total number of words in documents}}$$

$$IDF(t) = \log \frac{\text{Total number of documents}}{\text{Number of documents that contain } t}$$

# BAG OF WORDS → TF-IDF

Word	TF			IDF	TF-IDF		
	REV1	REV2	REV3		REV1	REV2	REV3
amazing	1/8	0	0	$\ln(4/2)+1=1.7$	0.47	0	0
an	1/8	0	0	$\ln(4/2)+1=1.7$	0.47	0	0
best	0	1/8	0	$\ln(4/2)+1=1.7$	0	0.47	0
game	1/8	1/8	1/6	$\ln(4/4)+1=1$	0.28	0.28	0.32
great	0	0	1/6	$\ln(4/2)+1=1.7$	0	0	0.54
is	1/8	1/8	0	$\ln(4/4)+1=1$	0.28	0.28	0.32
of	1/8	1/8	1/6	$\ln(4/4)+1=1$	0.28	0.28	0.32
series	1/8	1/8	0	$\ln(4/3)+1=1.29$	0.36	0.36	0
so	0	0	1/6	$\ln(4/2)+1=1.7$	0	0	0.54
the	0	1/8	0	$\ln(4/2)+1=1.7$	0	0.47	0
thrones	1/8	1/8	1/6	$\ln(4/4)+1=1$	0.28	0.28	0.32
tv	0	1/8	0	$\ln(4/3)+1=1.29$	0.36	0.36	0

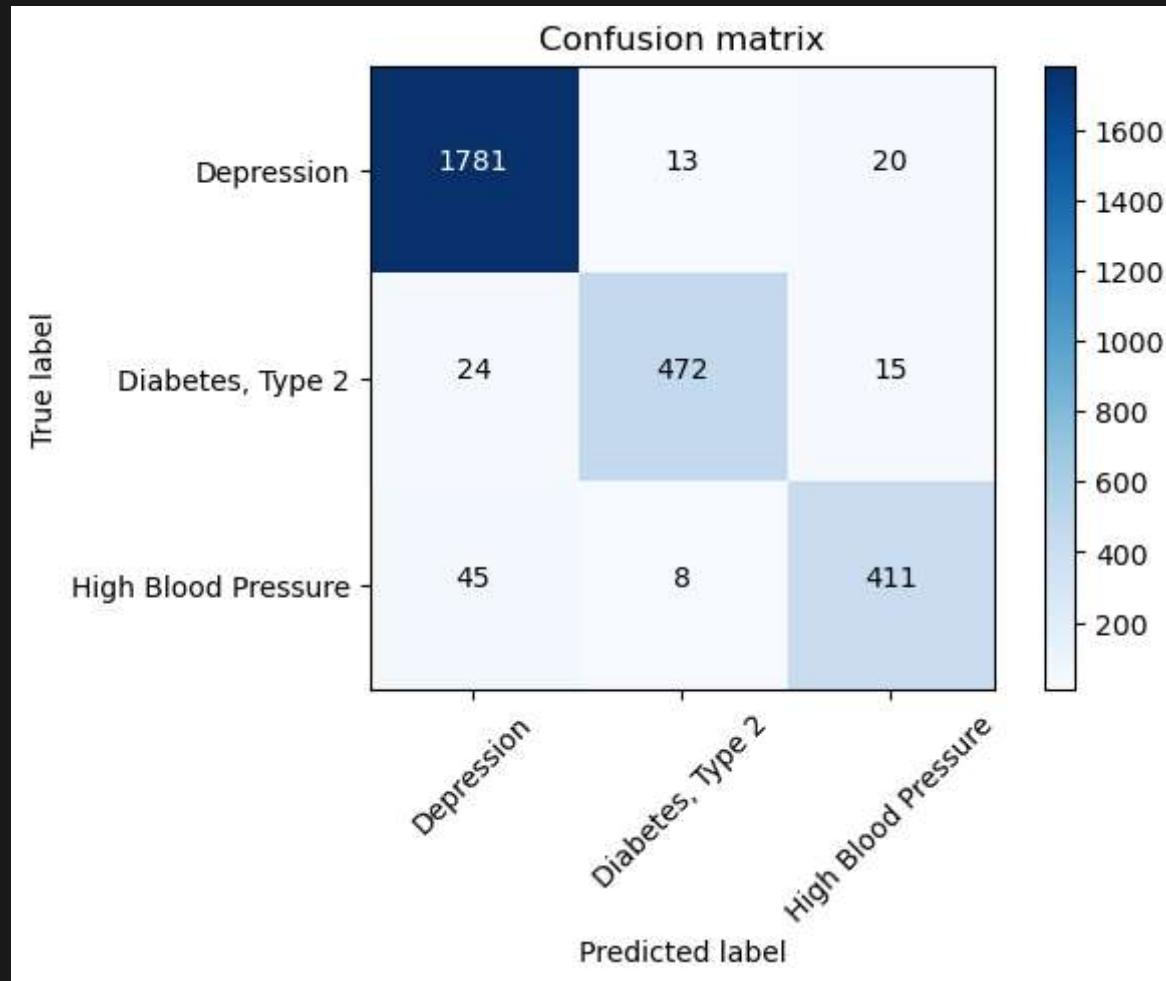
# Passive Aggressive Classifier & Naives Bayes TFIDF :



# MODEL EVALUATION

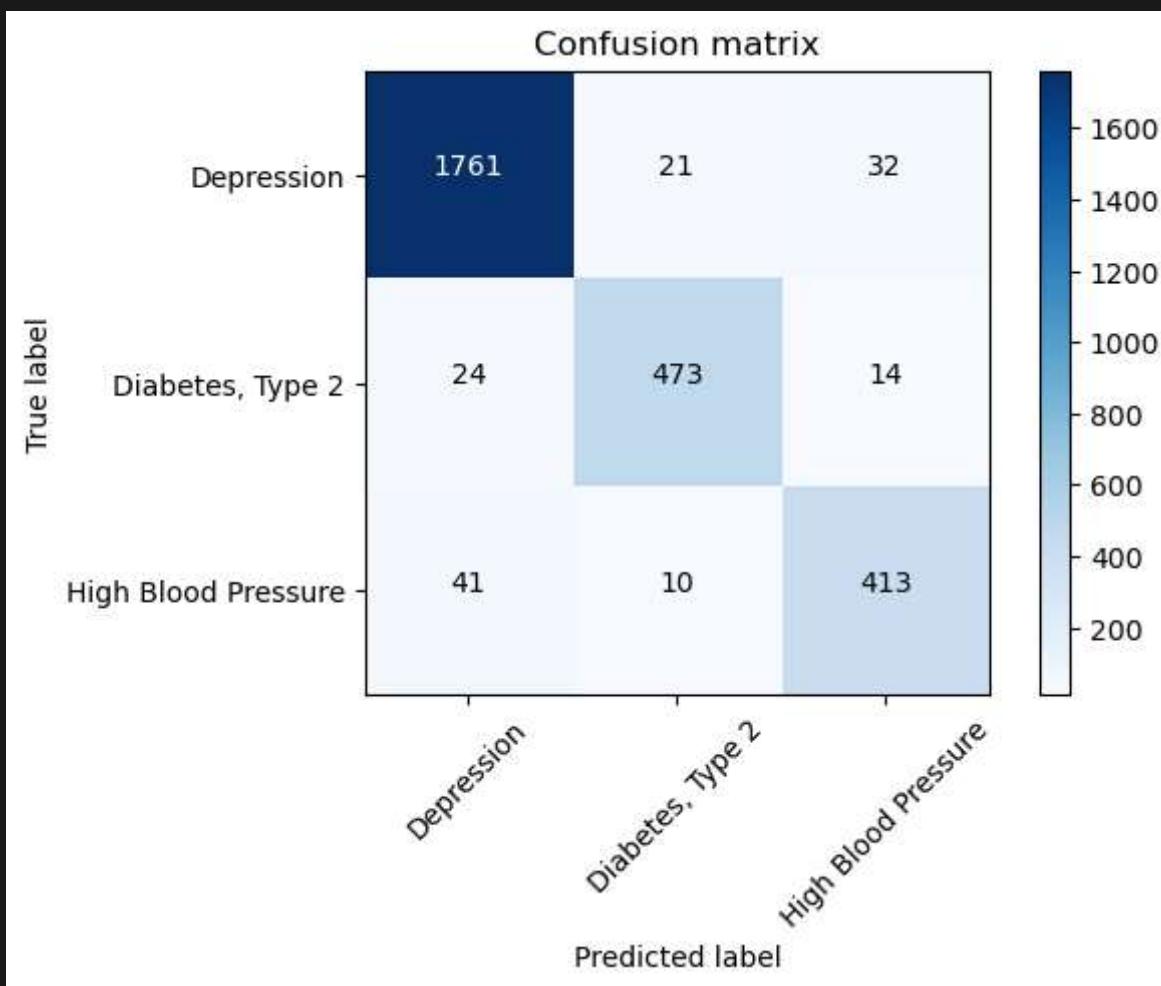


# Naive Bayes (Bag of words)



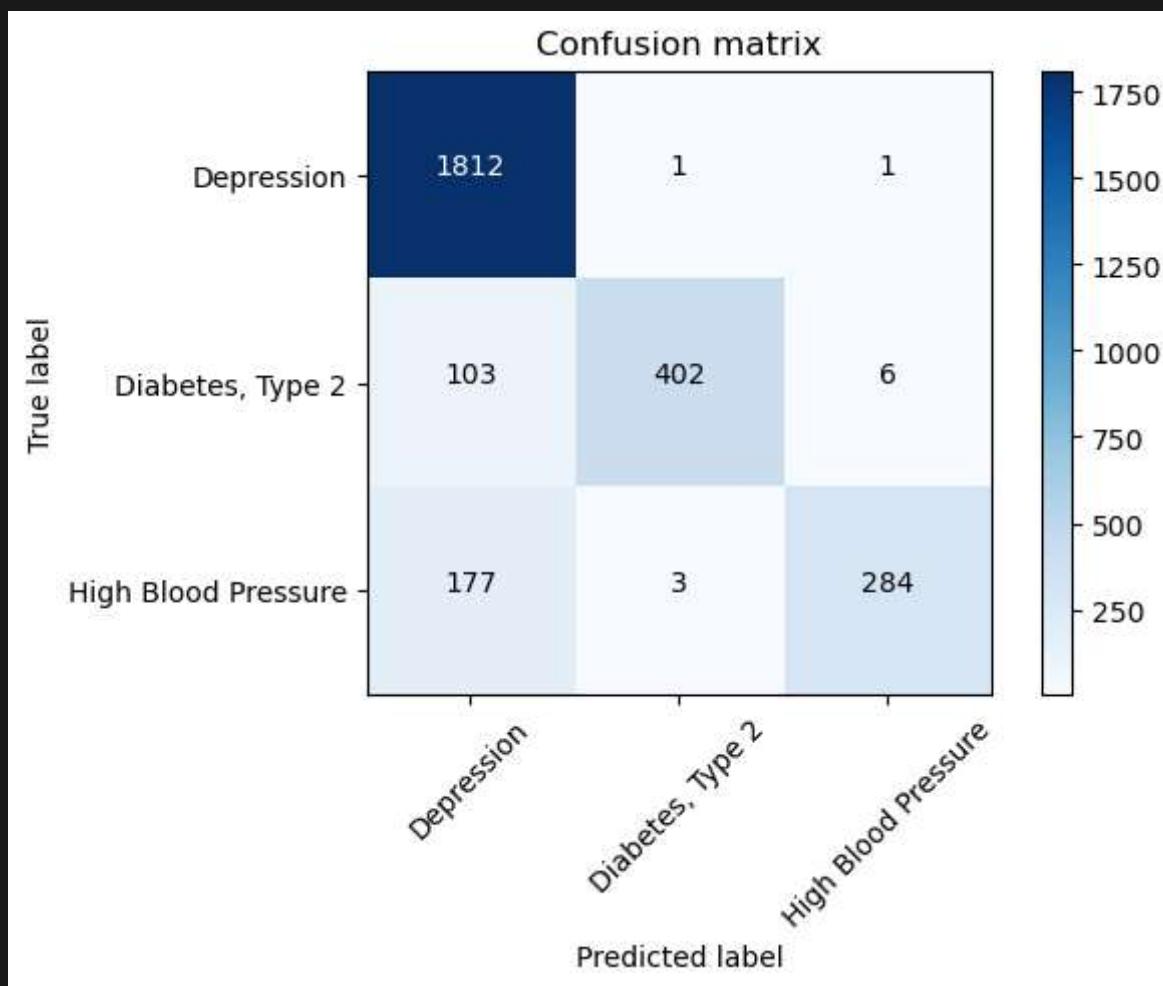
Accuracy: 0.955

# Passive Aggresive Classifier (Bag of Words )



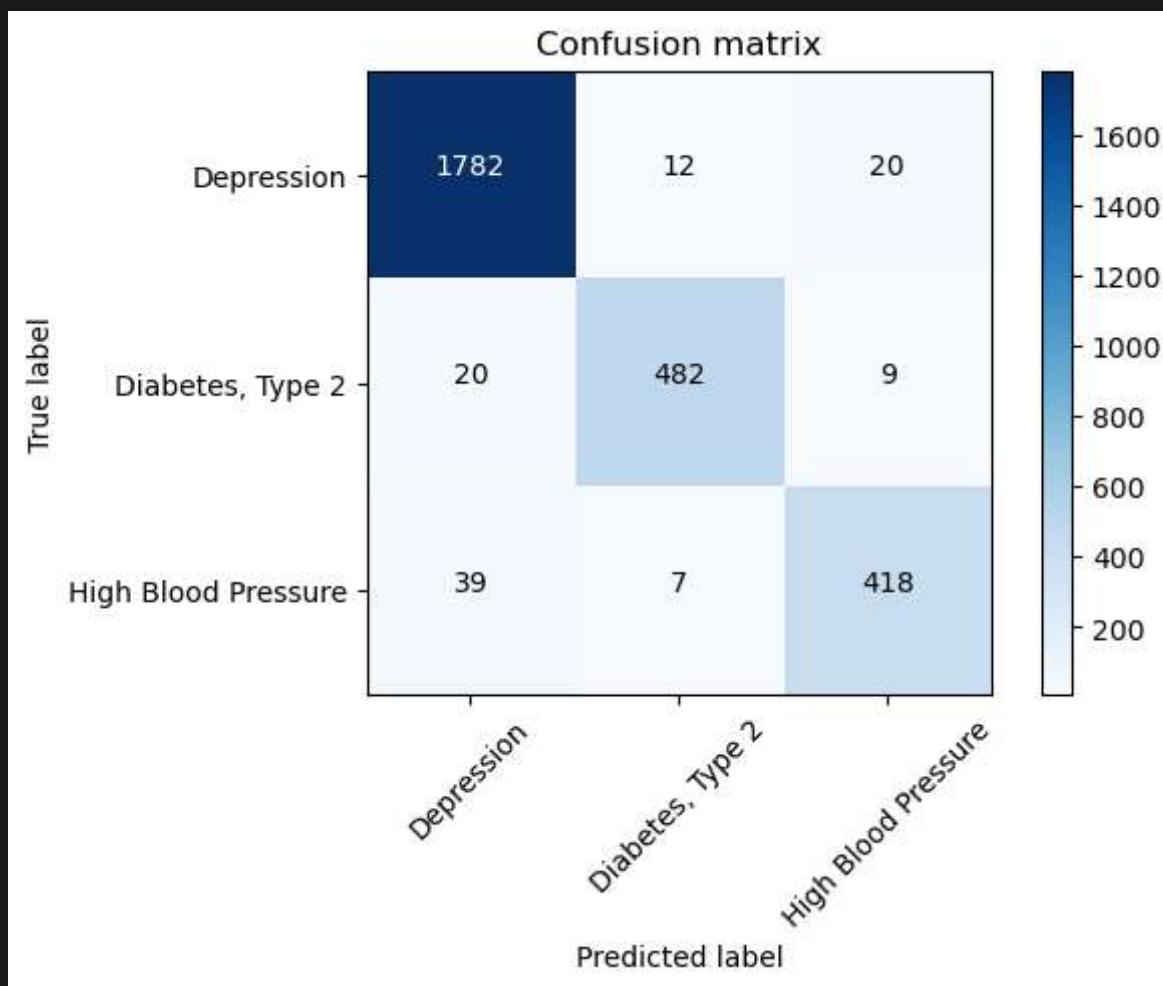
Accuracy: 0.949

# Naive Bayes (TF-IDF (Monogram))



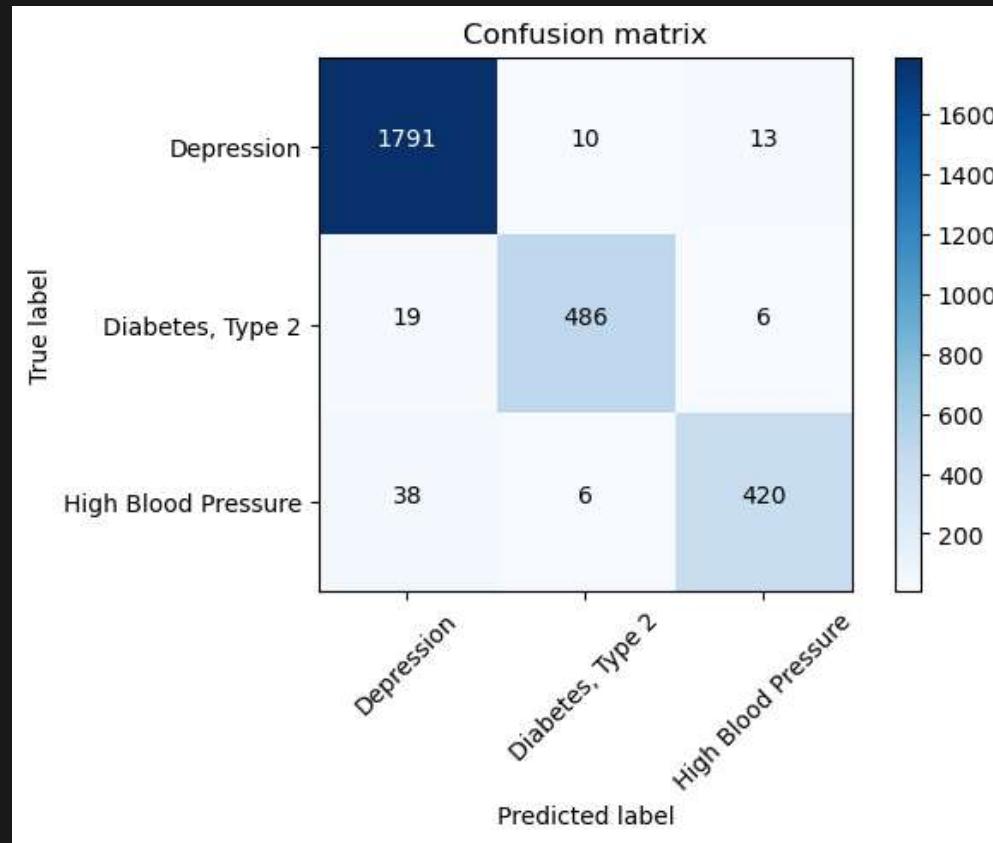
Accuracy: 0.896

# Passive Aggressive Classifier (TF-IDF (Monogram))



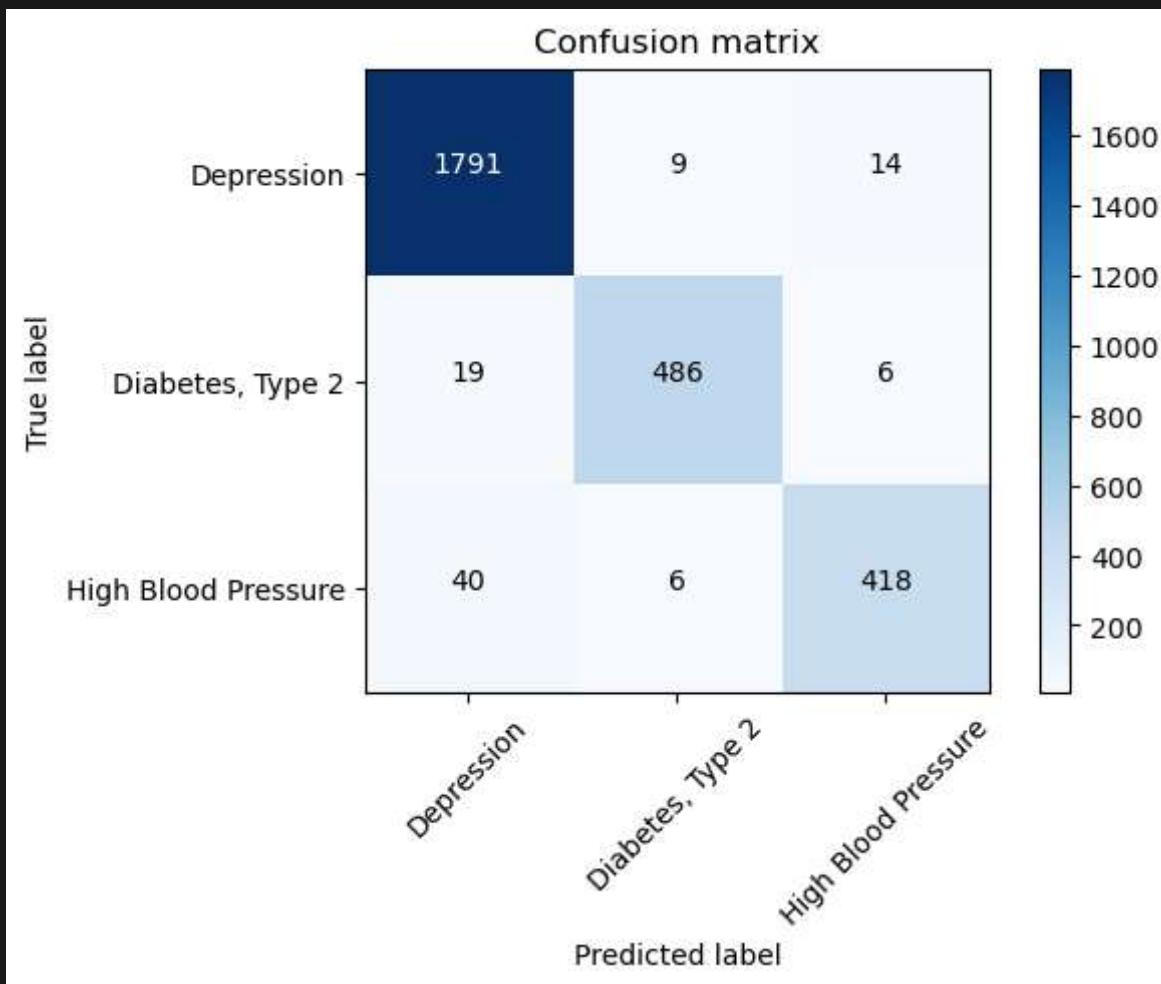
Accuracy: 0.962

# Passive Aggressive Classifier (TF-IDF (Bigram)) 😊



Accuracy: 0.967 → This is our final deployment model  
↓ Highest Accuracy 🎉

# Passive Aggressive Classifier (TF-IDF (Trigram))



Accuracy: 0.966

# WEB APP REVIEW DRUG SUGGESTION

In this app diagnosis is predicted based  
on the review submitted by the user

## DEPLOYMENT CODE EXECUTION :

```
streamlit run {filename}.py
```

# Matching Patients with Effective Medications

Welcome to our health prediction app! 🎉

Get quick and accurate results for your health condition with our app. Type in your review, click Submit, and let our advanced algorithms do the rest. We'll identify your underlying condition and provide you with the top 5 drugs for relief. Try it today and feel better 😊.

Type your review here

Submit

# Matching Patients with Effective Medications

Welcome to our health prediction app! 

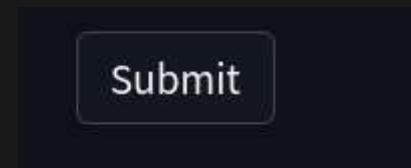
Get quick and accurate results for your health condition with our app. Type in your review, click Submit, and let our advanced algorithms do the rest. We'll identify your underlying condition and provide you with the top 5 drugs for relief. Try it today and feel better 😊.

Type your review here

This is the third med I've tried for anxiety and mild depression. Been on it for a week and I hate it so much. I am so dizzy, I have major diarrhea and feel worse than I started. Contacting my doc in the am and changing asap.

Submit

BEHOLD THE GREAT SUBMIT BUTTON 🙌



# Welcome to our health prediction app! 🙌

Get quick and accurate results for your health condition with our app. Type in your review, click Submit, and let our advanced algorithms do the rest. We'll identify your underlying condition and provide you with the top 5 drugs for relief. Try it today and feel better 😊.

Type your review here

This is the third med I've tried for anxiety and mild depression. Been on it for a week and I hate it so much. I am so dizzy, I have major diarrhea and feel worse than I started. Contacting my doc in the am and changing asap.

Submit

Based on the review given this could be your underlying condition: Depression

	condition	drugname	cumulative
1	Depression	Maprotiline	8.7765
2	Depression	Asendin	8.4583
3	Depression	Methylin ER	8.4536
4	Depression	Xanax XR	8.4375
5	Depression	Niacin	8.4059

# OVERCOMING CHALLENGES: LESSONS LEARNED IN THE PROJECT



1. Analyzing Subjective Reviews: Techniques and Best Practices
2. Bag of Words Model: Simplifying Text Analysis
3. TF-IDF (Term Frequency-Inverse Document Frequency): The Power of Term Weighting in NLP
4. Model Assessment: Evaluating and Comparing Machine Learning Models for Text Analysis
5. Streamlit Deployment: Enabling User-Friendly Access to Text Analysis Tools

# TEAM MEMBERS



- Mr. Shreyash Patil
- Mr. Anish Chichad
- Mr. Adesh Mahadev Jadhav
- Mr. Abhijeet N Patil
- Mr. Vempali Mani Ratnam
- Miss Kavya d k
- Mrs. Jeyasree J K
- Mr. Atit Naresh Dupare

# THANK YOU



