

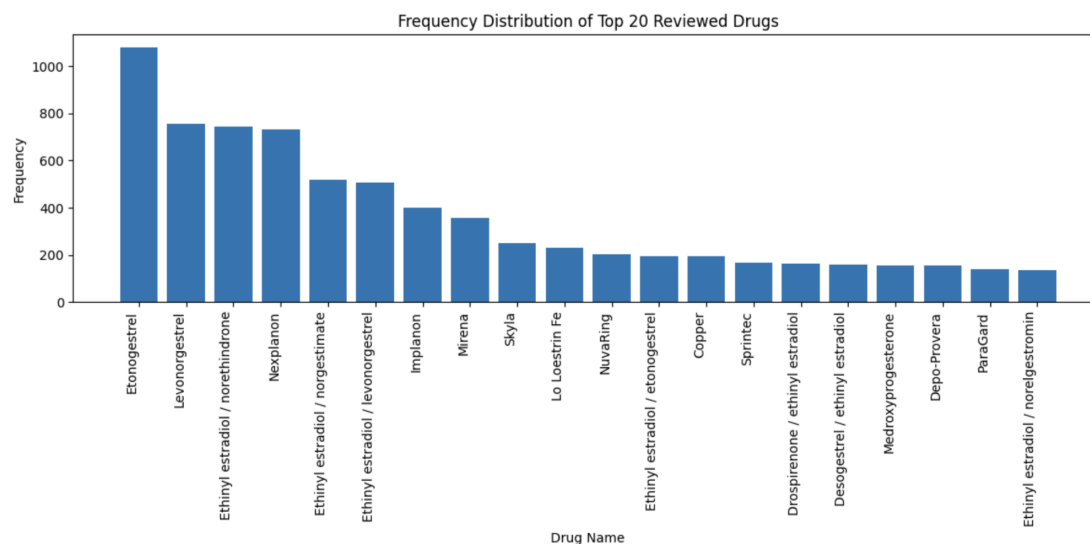
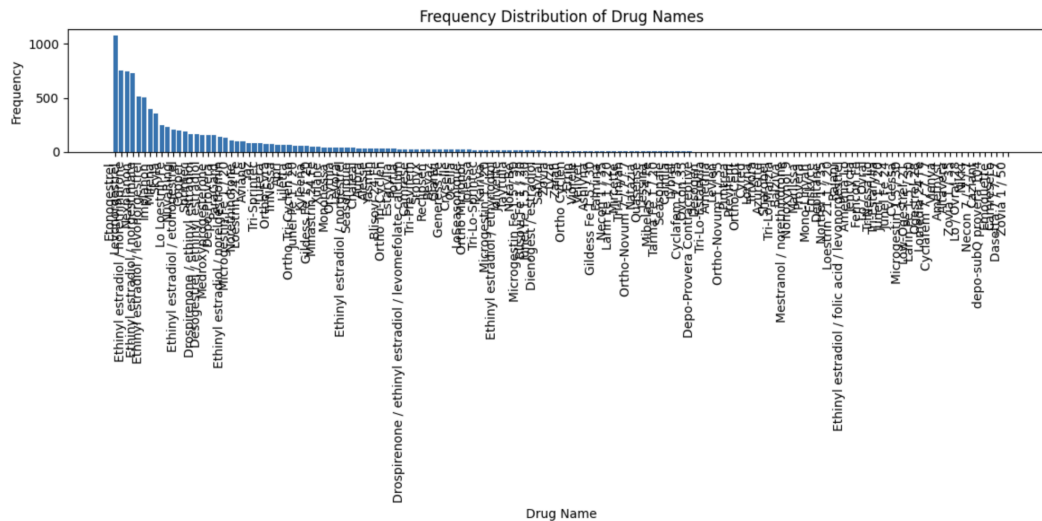
Analysis Data File 1: sentimentdf.csv

Unit of Observation: Each row in the dataset represents a single review submitted by a patient regarding their experience with a birth control drug.

1. Variable: drugName

- Definition: This variable represents the name of the birth control drug associated with the review.
- Input Data File: The variable name, drugName was taken directly from the input dataset without additional processing.
- Missing Observations: 9649(0)

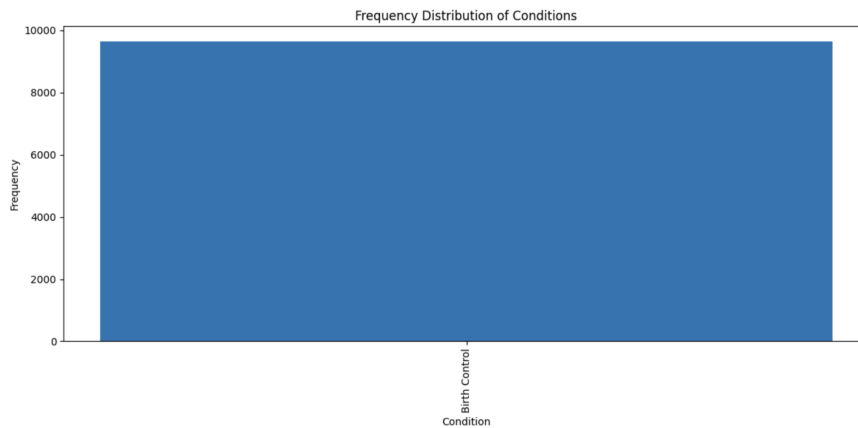
```
drugName
Etonogestrel      1080
Levonorgestrel    755
Ethinyl estradiol / norethindrone 744
Nexplanon         734
Ethinyl estradiol / norgestimate 517
...
depo-subQ provera 104
Emoquette         1
CamreseLo        1
Dasetta 7 / 7 / 7 1
Zovia 1 / 50     1
Name: count, Length: 155, dtype: int64
```



2. Variable: condition

- Definition: This variable represents the name of the medical condition that the drug is being used for
- Input Data File: The variable name, condition was filtered to only look at conditions of "Birth Control"
- Missing observations: 9649(0)

```
condition
Birth Control    9648
Name: count, dtype: int64
```



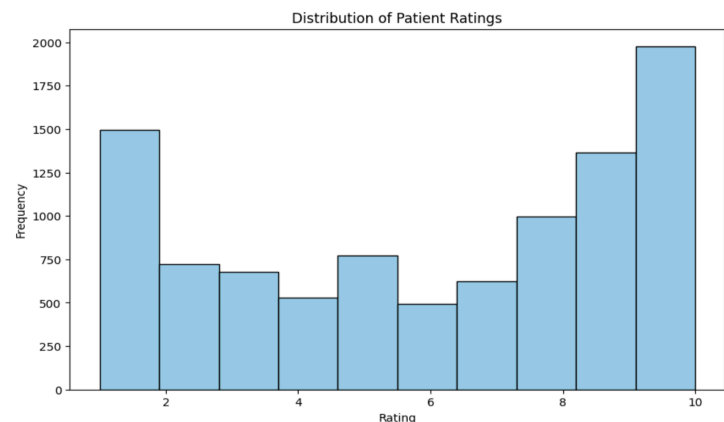
3. Variable: review

- Definition: This variable represents the comments from the patient on their overall experience with the drug. It is full, unprocessed text.
- Input data file: The variable name, review was taken directly from the input dataset without additional processing
- Missing observations: 9649(0)

4. Variable: rating

- Definition: This variable represents the patient's rating of the drug on a scale of 1 to 10. Units of measurement are any number between 1 and 10.
- Input data file: The variable name, rating was taken directly from the input dataset without additional processing
- Missing observations: 9649(0)

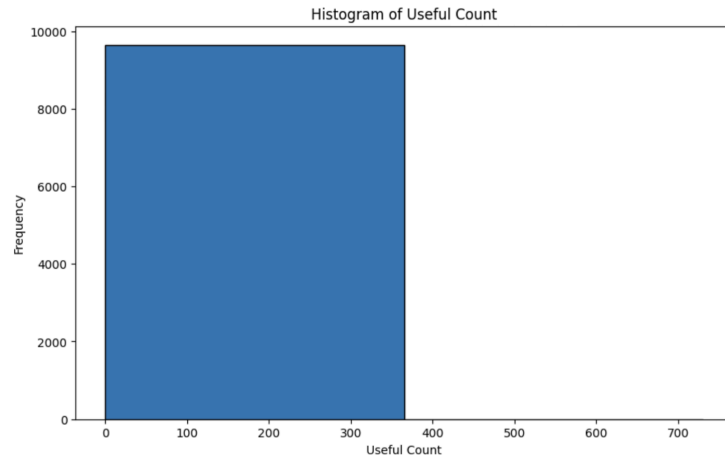
```
count    9648.000000
mean      6.042289
std       3.320338
min       1.000000
25%       3.000000
50%       7.000000
75%       9.000000
max      10.000000
Name: rating, dtype: float64
```



5. Variable: usefulCount

- Definition: This variable represents the number of times that other users have found the review useful.
- Input data file: The variable name, rating was taken directly from the input dataset without additional processing
- Missing observations: 9649(0)

```
count    9648.000000
mean       8.017828
std      17.708017
min        0.000000
25%       2.000000
50%       5.000000
75%       9.000000
max      730.000000
Name: usefulCount, dtype: float64
```



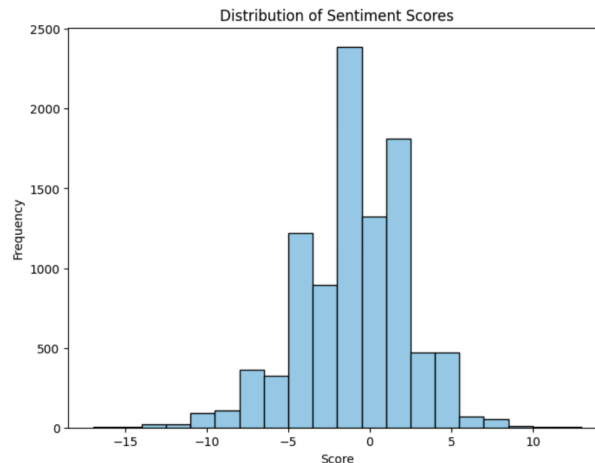
6. Variable: review.tm

- Definition: This variable represents processed text from the review column.
- Input data file: The variable name was created during sentiment analysis using the review column in the input data file. The text in the review column was processed to remove filler words, such as and, or, etc.
- Missing observations: 9649(0)

7. Variable: score

- Definition: This variable represents the overall sentiment score calculated by subtracting the negative score from the positive score.
- Input data file: This variable was created from the review column by applying a sentiment analysis model to calculate the overall sentiment of the review.
- Missing observations: 9649(0)

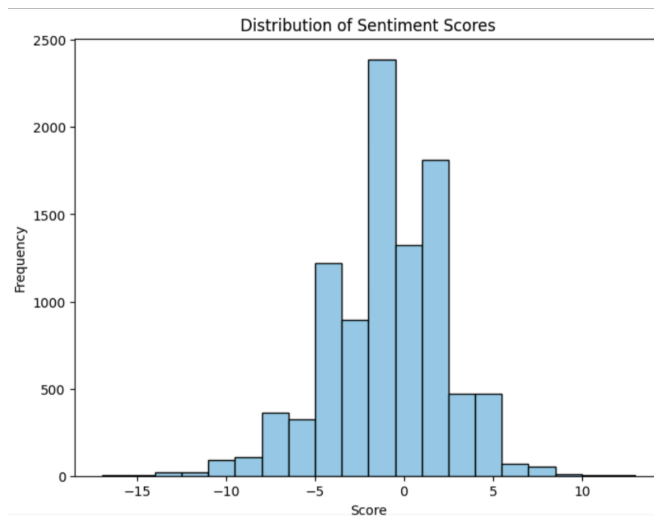
```
count    9648.000000
mean     -1.217454
std       3.315324
min     -17.000000
25%      -3.000000
50%      -1.000000
75%       1.000000
max      13.000000
Name: score, dtype: float64
```



8. Variable: positive

- Definition: This variable represents the score reflecting the degree of positivity in the processed text (review.tm)
- Input data file: The positive variable was created by processing the review column using a sentiment analysis model, which counts the number of positive words in each review.
- Missing observations: 9649(0)

```
count    9648.000000
mean      3.130390
std       2.098289
min       0.000000
25%       2.000000
50%       3.000000
75%       4.000000
max       15.000000
Name: positive, dtype: float64
```



9. Variable: negative

- Definition: This variable represents the score reflecting the degree of negativity in the processed text (review.tm)
- Input data file: The negative variable was created by processing the review column using a sentiment analysis model, which counts the number of negative words in each review.
- Missing observations: 9649(0)

```
count    9648.000000
mean      4.347844
std       2.906416
min       0.000000
25%       2.000000
50%       4.000000
75%       6.000000
max       18.000000
Name: negative, dtype: float64
```

