**Python Analysis Report**

Drug Analysis

**April 6, 2023**

Prepared By

John Smith

&

John Smith

**Table of Contents**

[1. Executive Summary 1](#_Toc251594010)

[2. Questions 3](#_Toc251594011)

[2.1. Count the number of reviews for each drug 3](#_Toc251594016)

[2.2. Find the drug with the highest rating 4](#_Toc251594017)

[2.3. Calculate the average rating for all drugs 4](#_Toc251594018)

2.4. Count the number of reviews for each condition [4](#_Toc251594019)

[2.5. Find the drug with the highest number of useful counts 5](#_Toc251594020)

[2.6. calculate the correlation between a drug's rating 5](#_Toc251594020)

[2.7. Find distribution of ratings for the drug Bactrim 5](#_Toc251594020)

[2.8. Find Median rating for drugs used to treat depression 6](#_Toc251594020)

[2.9. Find Most commonly used drug for migraine prevention 6](#_Toc251594020)

[3 Top 20 drugs with 1/10 ratings 6](#_Toc251594021)

[4 Top 10 Conditions for which people take drugs 7](#_Toc251594021)

[5 Top 10 Drugs Used for Birth Control 7](#_Toc251594021)

[6 Linear Model 7](#_Toc251594021)

[7 **Conclusion** 8](#_Toc251594021)

# 1. Executive Summary

Following an examination of the provided dataset, it is possible to establish that the dataset contains information regarding medication reviews. There are four columns in the dataset: unique ID, drug name, medical condition, review, rating, date, and helpful count.

After reading the evaluations, I discovered that individuals had shared their experiences with various drugs used to treat their medical ailments. The evaluations included topics such as depression, Crohn's disease, maintenance, and urinary tract infections.

The patient recounted their experience using Mirtazapine to treat their depression, sleeplessness, and anxiety in the first review. The patient expressed thanks for the drug, stating that it had saved their life. They also stated that they had no negative effects and even lost weight after using the drug.

The patient recounted their son's experience using Mesalamine to treat Crohn's illness in the second review. The patient claimed that their youngster had no adverse effects and improved significantly in minimizing episodes of diarrhea.

The patient recounted their experience taking Bactrim to treat urinary tract infections in the third review. The drug promptly relieved the patient's problems, according to the patient.

Overall, the dataset offered useful information about patient experiences with various drugs. The reviews can assist physicians in making educated judgements when providing drugs to their patients.

# 2. Questions

1. **Count the number of reviews for each drug and find the drug with the most?**

I discovered several missing values in the 'condition' column when analyzing the dataset. However, it has no bearing on our study because we are only interested in the amount of reviews for each medicine.

To find the medicine with the most reviews, I first verified the dataset for any null values with the code 'df.isnull().any()'. The results show that there are no null values in any column save the 'condition' column.

Then I used the function 'df.isnull().sum()' to count the number of null values in the 'condition' column. The result revealed that the 'condition' column contains 295 null entries.

I used the method 'df['drugName'].value\_counts().idxmax()' to identify the drug with the most reviews. It checks the number of reviews for each medicine and provides the drug name with the most reviews. The medicine with the most reviews is 'Levonorgestrel,' according to the results.

As a result, I can infer that 'Levonorgestrel' is the medicine with the most reviews in the supplied dataset. This information can help healthcare practitioners and researchers understand the medication's popularity and efficacy in treating a variety of medical issues.

1. **Find the drug with the highest rating?**

After analyzing the data, I discovered that Mirtazapine had the highest rating. It has a 9.1 out of 10 average rating and is often used to treat depression. The medication has received 45 reviews, with the highest rating of 10. According to the drug's evaluations, it is an excellent treatment for depression, sleeplessness, and anxiety.

Mirtazapine patients have reported no severe adverse effects and an improvement in their health. The drug's efficacy is reflected in its high rating, making it a recommended prescription for depression sufferers.

1. **Calculate the average rating for all drugs?**

I used the pandas mean() function to get the mean of the 'rating' column in our dataset to compute the average rating for all medicines. 6.9769 is the resultant value. This suggests that the medications in our dataset have an average rating of about 7.

This data may be used to better understand how patients perceive pharmaceuticals in general, as well as to discover drugs with higher or lower ratings than the norm. The average rating may also be used to compare medications, since pharmaceuticals with ratings higher than the average may be deemed more effective or well tolerated than drugs with ratings lower than the average.

1. **Count the number of reviews for each condition and find the most common one?**

I used the function 'df['condition'].value\_counts().idxmax()' to get the most prevalent medical condition for which reviews were supplied. This code counts the number of reviews for each condition and returns the condition with the most reviews.

The outcome of this code shows that the most prevalent medical condition for which the reviews were offered is 'Birth Control'. This data can help healthcare practitioners and researchers better understand the effects of birth control medicines and their efficacy in treating or preventing a variety of medical issues.

In general, studying the most frequent medical condition can assist healthcare practitioners and researchers in identifying areas of concern, developing effective therapies, and making educated judgements when giving drugs to their patients.

1. **Find the drug with the highest number of useful counts?**

I used the code 'df.loc[df['usefulCount'].idxmax(), 'drugName']' to get the drug with the most useful counts. This method finds the index of the row with the most helpful counts and returns the name of the medicine connected with that row.

The outcome of this code revealed that 'Sertraline' had the maximum amount of usable counts. This data can help healthcare practitioners and researchers evaluate the efficacy of this medicine and how it compares to other medications in treating a variety of medical issues.

In general, assessing the number of usable counts can assist healthcare practitioners and researchers in determining the efficacy of various treatments and making educated judgements when giving pharmaceuticals to their patients.

1. **calculate the correlation between a drug's rating and its usefulness count?**

I used the function 'df['rating'].corr(df['usefulCount'])' to find the relationship between a drug's rating and its usefulness count. The Pearson correlation coefficient is used to compute the correlation between the two variables in this code.

According to the results of this code, the connection between a drug's rating and its utility count is 0.2368972307481001. This data indicates a positive relationship between the two variables, implying that medications with better ratings are often linked with greater usefulness counts.

Overall, understanding the relationship between a drug's rating and its utility count can assist healthcare practitioners and researchers in evaluating the efficacy of various treatments and determining the best course of therapy for their patients.

1. **Find distribution of ratings for the drug Bactrim?**

I initially used the code 'df[df['drugName'] == 'Bactrim']' to obtain the distribution of ratings for the drug Bactrim. to generate a new dataframe containing solely Bactrim reviews. Then, using the code supplied above, I used the matplotlib package to generate a histogram of the ratings.

As a result of this code, the ratings for Bactrim are as follows: 2 (17), 4 (8), 6 (5), 8 (9), and 10 (19). This data can help healthcare practitioners and patients understand the efficacy of Bactrim in treating urinary tract infections and make educated decisions about its use.

Overall, analyzing the distribution of ratings might give useful insights into pharmaceutical efficacy and assist healthcare practitioners and patients in making educated decisions while selecting treatment options.

1. **Find Median rating for drugs used to treat depression?**

To obtain the median rating for medications used to treat depression, I first constructed a new dataframe using the code 'df[df['condition'] == 'Depression']'. Then I used the pandas'median()' method to obtain the median rating of the depressive medicines.

According to the results of this code, the median rating for medications used to treat depression is 8.0. This information can help healthcare practitioners and patients understand the efficacy of depression drugs and make educated decisions about their usage.

Overall, analyzing the median rating can give useful insights into the effectiveness of pharmaceuticals and assist healthcare practitioners and patients in making educated treatment decisions.

1. **Find Most commonly used drug for migraine prevention?**

To discover the most widely used migraine prevention medicine, I first built a new dataframe containing just the reviews for migraine prevention drugs using the code 'df[df['condition'] == 'Migraine Prevention']'. Then, using pandas' value\_counts() function, I counted the number of reviews for each medicine and sorted them in descending order.

Topiramate is the most often used medicine for migraine prophylaxis, according to the results of this code. This information can help healthcare practitioners and patients understand the most effective migraine drugs and make educated decisions about their use.

Overall, analyzing the most widely used medicine for a certain ailment can provide significant insights on pharmaceutical effectiveness and assist healthcare providers and patients in making decisions.

## 3. Top 20 drugs with 1/10 ratings

This code creates a bar chart with a rating of 1 out of 10 for the top 20 medications. We begin by filtering the dataframe to include only rows with a rating of one, and then we count the number of occurrences of each medicine and store it in a dictionary.

The seaborn library is then used to generate a horizontal bar chart with the drug names on the y-axis and the number of ratings on the x-axis. To make the x-axis labels easier to read, we changed the title, axes labels, and rotated them.

According to the statistics, Miconazole and Etonogestrel have the most one-star reviews, followed by Gabapentin and Lo loestrin fe, which are tied at number 19 & 20. This data can be used to identify medicines that may have serious adverse effects or are ineffective in treating specific illnesses.

## 4. Top 10 Conditions for which people take drugs

This code examines a dataset to discover the top ten reasons why individuals consume drugs. It begins by compiling a glossary of the medications used to treat each illness. The top ten ailments are then chosen based on the largest number of medications taken. It creates a bar graph with the top ten criteria and their counts.

The results suggest that birth control is the most often used medicine, followed by sadness, pain, anxiety, and acne. It is vital to highlight that this information can help healthcare practitioners and researchers better understand a population's health requirements and design suitable remedies.

## 5. Top 10 Drugs Used for Birth Control

The purpose of this code is to identify the top ten birth control medicines. We filtered the data using the 'Birth Control' criteria, and then used the 'value\_counts()' method to calculate the number of drugs used. We then chose the top ten medications and created a bar chart with the Seaborn library.

The resultant bar graph displays the drug names on the x-axis and the number of drugs on the y-axis. The bar chart is arranged in descending order, with the most often used birth control medication at the top. The figure shows that Etonogestrl and Levonogestrel are the top two birth control medicines, with counts of 1080 and 755, respectively.

## 6. Linear Model

This is a linear regression model that has been trained on the drug review dataset to predict the rating based on the review text. The approach converts the text input into numerical data using CountVectorizer before training a linear regression model on it. To assess the model's performance, the data is divided into training and testing sets.

The model did well on the training data, with a score of 0.804, indicating that the model can explain 80.4% of the variance in the rating. However, the model fared poorly on the testing data, with a score of -1.561, showing that the model is overfitting the training data and does not generalise well to new data.

## **Conclusion**

Finally, the dataset analysis provided useful insights into numerous areas of drugs. The investigation sought to identify the drug with the most reviews, the drug with the highest rating, the average rating of all drugs, the medical condition with the most reviews, the drug with the greatest number of useful counts, the correlation between a drug's rating and its usefulness count, and the distribution of ratings for the drug Bactrim.

Levonorgestrel had the most reviews, but Mirtazapine received the highest rating of all the medications studied. The overall average rating for all pharmaceuticals was 6.9769, suggesting that the medications in the sample had a rating of about 7. Birth Control was the most prevalent medical condition, while Sertraline had the most helpful numbers. The correlation between a drug's rating and its usefulness count was 0.2368972307481001, showing that higher-rated drugs are frequently associated with higher usefulness counts. Finally, the Bactrim rating distribution varied from 2 to 10, suggesting that the medicine had both low and high ratings.

This data can be useful to healthcare practitioners and academics in a variety of ways. For starters, it gives information about the popularity of various drugs among patients, allowing doctors to choose treatments based on patient preferences. Second, it assists in the evaluation of medicine efficacy, allowing clinicians to make educated decisions on the efficacy of a therapy. Finally, the information gleaned from this research can aid in the creation of new pharmaceuticals as well as the enhancement of existing treatments, allowing healthcare practitioners to deliver the best possible care to their patients.