**Project Report**

**Title Page**

Project\_name:- **Exploratory Data Analysis on Drugs, Side Effects, and Medical Condition**

Name:-Varsha Dudhat

Organization:- **Unified Mentor pvt.ltd**

**1. Objective**

**The goal is to analyze the relationships between drugs, their side effects, and the**

**medical conditions they treat, as well as to explore the ratings and reviews**

**associated with these drugs.**

**2. Dataset Overview**

**The dataset contains the following columns:**

**● drug\_name: Name of the drug.**

**● medical\_condition: The condition the drug is used to treat.**

**● side\_effects: Common side effects of the drug.**

**● generic\_name: The generic name of the drug.**

**● drug\_classes: The class of the drug (e.g., antibiotic, antihistamine).**

**● brand\_names: Brand names under which the drug is sold.**

**● activity: The activity of the drug (e.g., active, inactive).**

**● rx\_otc: Indicates if the drug is prescription (Rx) or over-the-counter (OTC).**

**● pregnancy\_category: The drug's pregnancy risk category.**

**● csa: Controlled Substances Act schedule, if applicable.**

**● alcohol: Interactions with alcohol.**

**● related\_drugs: Other drugs related to the primary drug.**

**● medical\_condition\_description: A brief description of the medical**

**condition.**

**● rating: Average user rating of the drug.**

**● no\_of\_reviews: Number of user reviews.**

**● drug\_link: URL link to more information about the drug.**

**● medical\_condition\_url: URL link to more information about the medical**

**condition.**

Data is having 2931 raw and 17 columns with key variables

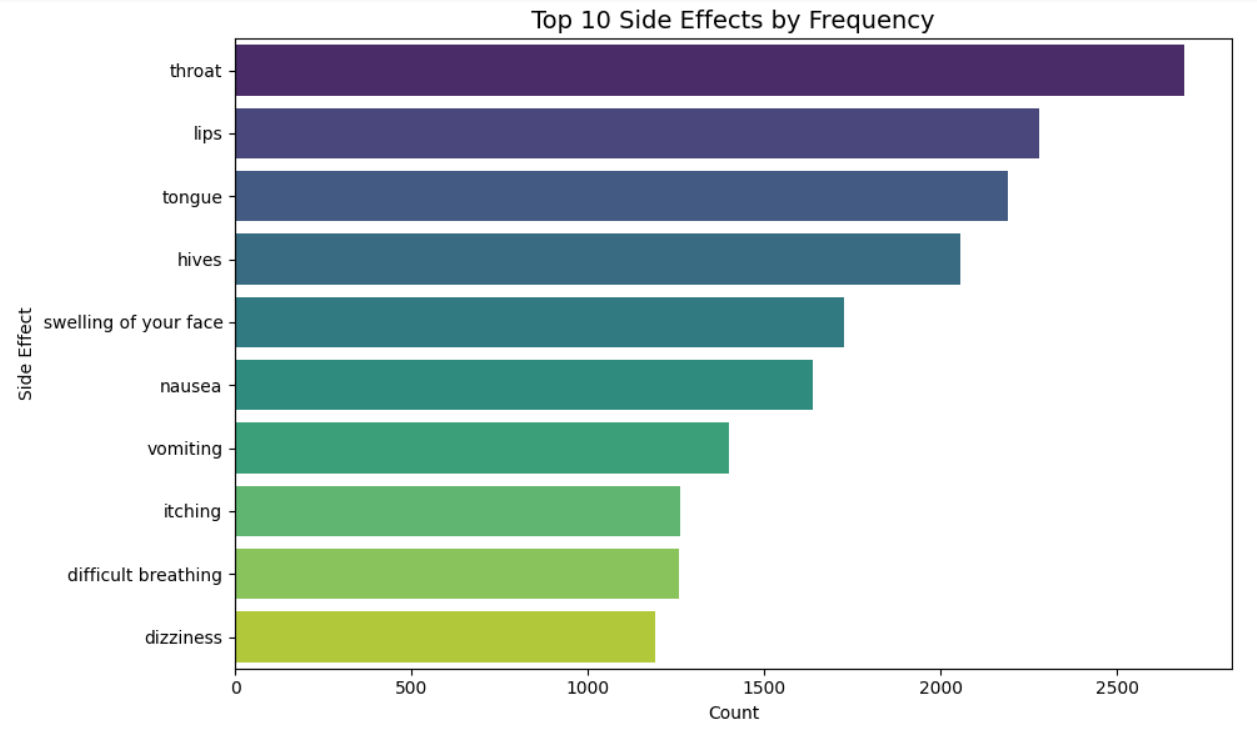
(drug\_name, medical\_condition, side\_effect) data is having around 14 % data is missing .major missing colums is rating , review ,alcohol and related durg columns.

**Data Cleaning & Preprocessing**

Filling category columns with “unknowen” and numeric data with “-1” . for easily identify missing data for the analysis with out miss leaning In analysis. All columns structure is good rathet than side effect done on it text normalization, handling of synonyms, removal of invalid entries and preparation for statistical analysis.

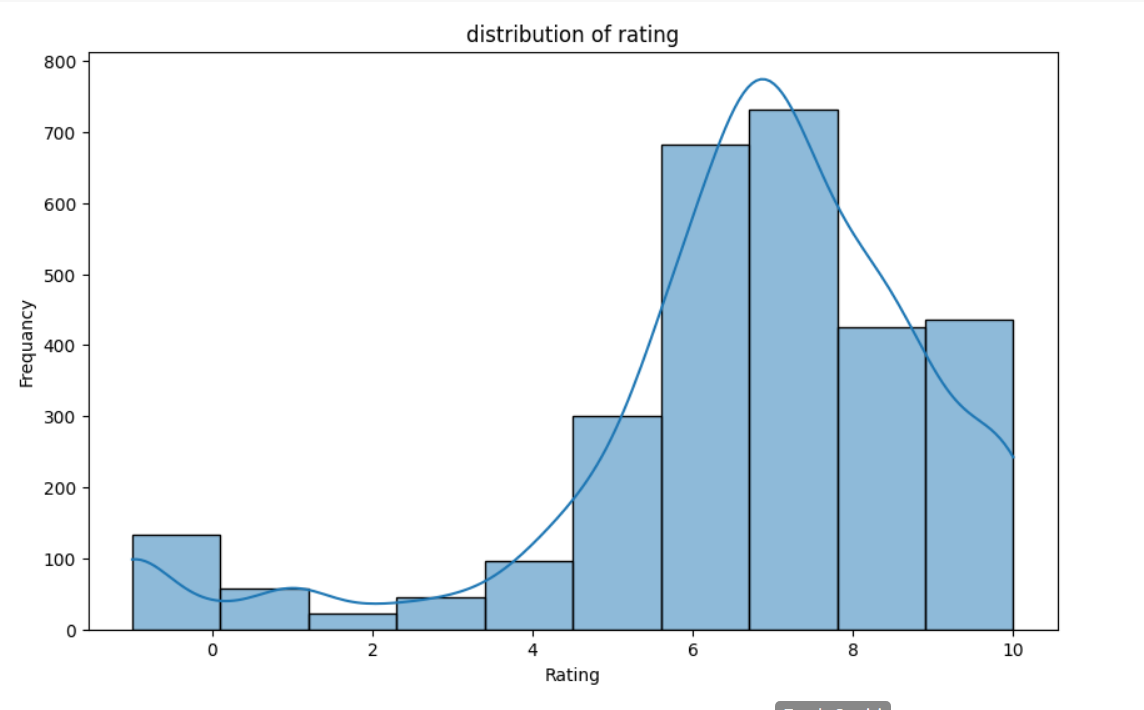
**Exploratory Data Analysis (EDA)**

**Top 10 side effect of drug**



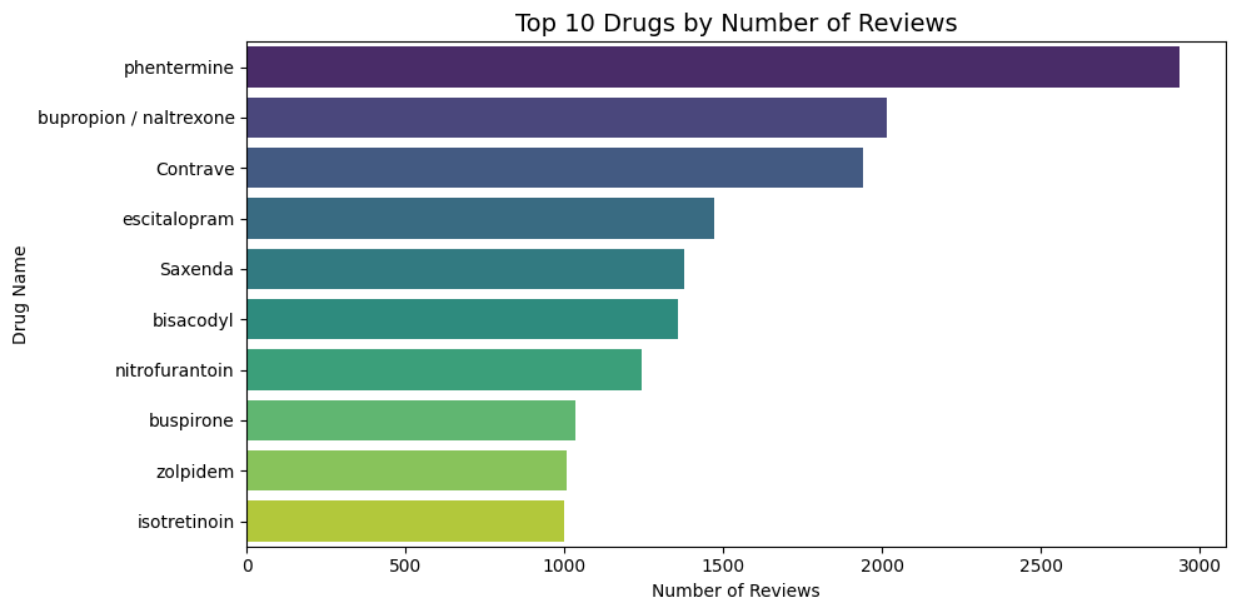
The most common side effects reported for this drug are swelling of the throat, lips, or tongue; hives; facial swelling; nausea; vomiting; itching; difficulty breathing; and dizziness.

**Histogram of rating**



The distribution of drug ratings is positively right skewed, indicating that most users reported high satisfaction (ratings 6–10), while lower or zero ratings may reflect user dissatisfaction or imputed values from missing data handling.

**Top 10 drug as per the review**



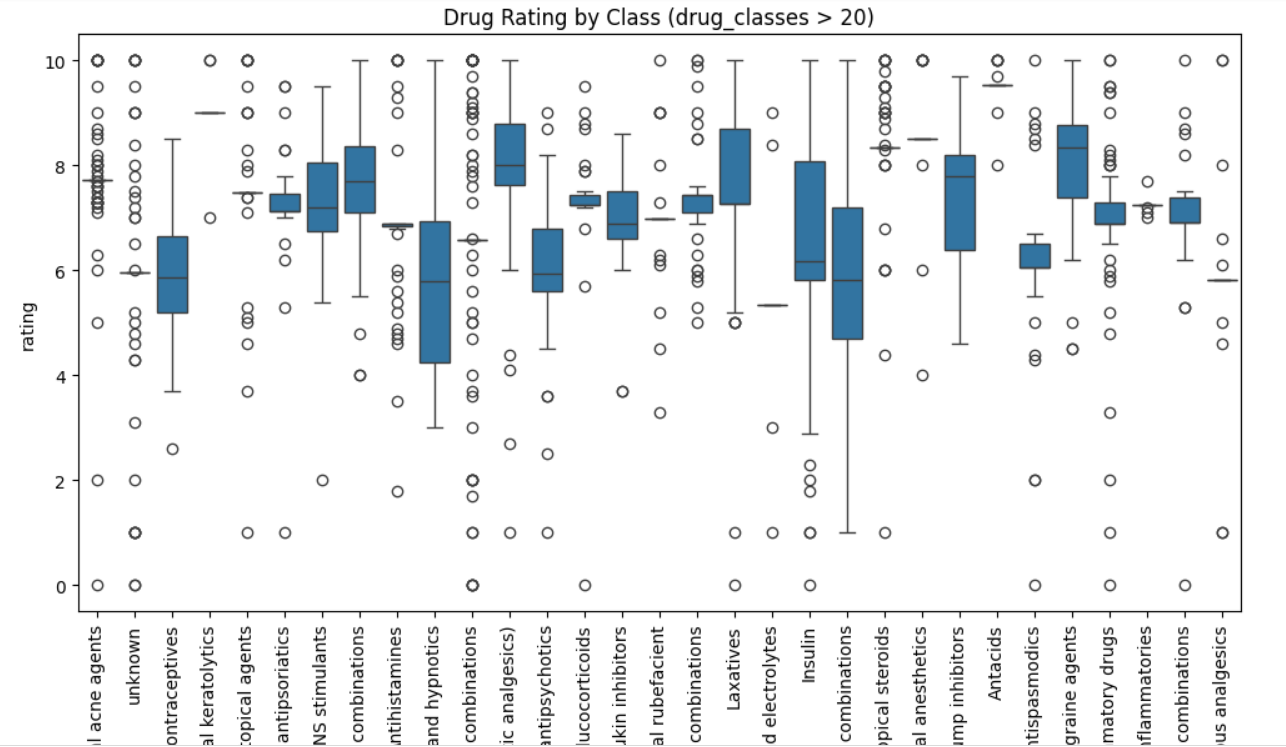
Phentermine,bupropin, contrave,escitalopram Is high received review

**Top drug as per the medical condition**

|  |  |  |
| --- | --- | --- |
| **drug\_name** | **medical\_condition** | **count** |
| acyclovir | Herpes | 2 |
| fluocinolone | Eczema | 2 |
| minocycline | Acne | 2 |
| 7T Gummy ES Chewable Tablets | Pain | 1 |
| A-G Profen | Osteoarthritis | 1 |
| A-Phedrin | Colds & Flu | 1 |
| Abatuss DMX | Colds & Flu | 1 |
| Abenol | Pain | 1 |
| Abilify | Bipolar Disorder | 1 |
| Abilify Maintena | Bipolar Disorder | 1 |

analysis shows that Acyclovir–Herpes, Fluocinolone–Eczema, and Minocycline–Acne are the most frequent drug–condition pairs, while others occur only once, indicating niche or limited usage in the dataset.

**Drug rting analysis by class**



The boxplot shows that most drug classes have median user ratings between 6 and 8, indicating generally favorable satisfaction. Classes like interleukin inhibitors, antacids, and some analgesics have high and consistent ratings, while classes like urinary antispasmodics, laxatives, and mineral supplements show lower medians and wider variability, suggesting mixed effectiveness or tolerance. Outliers in almost every category reflect individual differences in user response.

**Statistical Analysis**

Chi-Square result suggest there’s no statistically significant relationship between the drug and the medical condition. But There is an extremely strong and statistically significant association between the drug and its reported side effects.

**Key Findings**

**Distribution of Drug Ratings:**

The distribution of drug ratings is positively right skewed, indicating that most users reported high satisfaction (ratings 6–10), while lower or zero ratings may reflect user dissatisfaction or imputed values from missing data handling.

**Top Drugs for a Condition:**

analysis shows that Acyclovir–Herpes, Fluocinolone–Eczema, and Minocycline–Acne are the most frequent drug–condition pairs, while others occur only once, indicating niche or limited usage in the dataset.

**Side Effects Analysis:**

The most common side effects reported for this drug are swelling of the throat, lips, or tongue; hives; facial swelling; nausea; vomiting; itching; difficulty breathing; and dizziness."

**Limitations & Future Work**

limited collection of data per drug which limits statistical power and may hinder accurate identification of strong drug–side effect–condition association.