

Pharmaceutical Market Analysis: Drug Characteristics and Prescription Trends

Ahmed ElShafiey

December 28, 2025

Abstract

This report analyzes pharmaceutical data from [kaggle.com](https://www.kaggle.com), examining drug classifications (Rx vs. OTC), pregnancy safety categories, and patient usage patterns across medical conditions. Pain is the most common condition treated, followed by cold symptoms and acne, with pain relief showing the highest activity. Most painkillers are OTC, though potent options require prescriptions, and over 50% fall under pregnancy category C. Cold medications are predominantly OTC, with limited safety data for pregnancy. High-prevalence conditions like Hypertension and Rheumatoid Arthritis involve more restricted Category D and X drugs, while lower-prevalence conditions rely more on Category B treatments. Overall, prescription medications dominate across conditions, though some, like Hayfever and Osteoarthritis, are readily available OTC, whereas complex conditions such as Diabetes and Pneumonia are managed primarily with prescriptions.

1 Introduction

The dataset analyzed in the notebook `analysis.ipynb` contains comprehensive information regarding drug names, medical conditions, and regulatory classifications. The analysis aims to clean the raw data and visualize the market landscape for both prescription (Rx) and over-the-counter (OTC) medications. Utilizing such raw data opens our eyes to hidden patterns and trends in drug prescriptions for many common medical conditions, helping to improve drug effectiveness and prevent drug-induced adverse effects.

2 Data Description

This dataset obtained from Kaggle: <https://www.kaggle.com/datasets/jithinanievarghese/drugs-related-to-common-treatments>. The dataset includes several critical features used for pharmaceutical evaluation:

- **Activity:** A metric of site visitor engagement relative to other medications.
- **Rx/OTC:** Classification of whether a drug requires a prescription.
- **Pregnancy Category:** Safety ratings ranging from A (safe) to X (contraindicated).
- **CSA:** Controlled Substances Act schedule (potential for abuse).
- **Rating:** User-reported effectiveness on a scale of 1 to 10.

3 Data Preprocessing

3.1 Handling Missing Values

Preliminary analysis revealed significant missing data in specific columns:

- **Alcohol, Rating, and No_of_reviews:** Excluded due to nearly 50% missingness.
- **Pregnancy Category:** 6.28% missing values were imputed using the mode (**Category C**).
- **Rx/OTC:** Standardized by treating “Rx/OTC” hybrids as “OTC” for consumer accessibility analysis.

3.2 Standardization Code

The following logic was applied to ensure data integrity:

```
df['pregnancy_category'] = df['pregnancy_category'].fillna('C')
df['rx_otc'] = df['rx_otc'].replace('Rx/OTC', 'OTC')
df['rx_otc'] = df['rx_otc'].fillna('Rx')
```

```
cols_to_drop = ['alcohol', 'rating', 'no_of_reviews']
cols_existing = [col for col in cols_to_drop if col in df.columns]
df.drop(columns=cols_existing, inplace=True)
```

4 Exploratory Data Analysis

The processed dataset contains **3,959** records. The distribution of medical conditions shows that **Pain** (393 drugs), **Colds & Flu** (246 drugs), and **Acne** (238 drugs) are the most highly represented categories.

4.1 Prescription vs. OTC Distribution

The analysis confirms that a majority of the medications in the dataset (**2,702**) are prescription-only (Rx), compared to **1,257** available over-the-counter.

4.2 Analgesics Ranked by User Activity

Activity reflects relative user engagement and usage trends based on recent site interactions, rather than clinical potency. The observed rankings—particularly the unexpectedly high activity of Ibuprofen compared to more potent analgesics—suggest that this metric primarily captures frequency of use, prescription volume, or over-the-counter accessibility. Moreover, the pronounced gap between Acetaminophen and other analgesics indicates that factors beyond pharmacological strength, such as population-wide usage patterns and data aggregation methods, significantly influence the metric. These findings imply that Activity should be interpreted as a behavioral and utilization indicator, rather than a direct measure of therapeutic potency.

4.3 Analgesics OTC drugs compared to RX drugs

For analgesic prescribing patterns, OTC analgesics account for **236** medications, while POM analgesics total **157**

4.4 Distribution of Pain Drugs by Pregnancy Category

Over 50% of the analyzed painkillers fall under pregnancy category C, with no drugs in category A and only 13.5% classified as category B. This distribution emphasizes that painkillers should be prescribed during pregnancy only as a last resort, following careful evaluation. Additionally, approximately 30% of the drugs are categorized as N (not yet assessed for pregnancy safety), representing the second-largest group. The substantial proportion of unassessed medications underscores the urgent need for further evaluation to ensure the safety of painkillers during pregnancy.

4.5 Cold & Flu in each pregnancy category

Most cold and flu medications fall into Category N (unclassified) or Category C, indicating limited or uncertain data regarding pregnancy safety. In contrast, Category B drugs are relatively few, and Category D drugs are rare, reflecting their restricted use due to known fetal risks. Overall, this underscores the limited availability of clearly pregnancy-safe options and highlights the importance of cautious prescribing and patient counseling during pregnancy.

4.6 Other medical conditions – drug distribution by pregnancy category

The distribution of drugs by pregnancy category across various medical conditions shows that Acne and Hayfever have the highest total drug counts. Notably, high-count conditions such as Hypertension and Rheumatoid Arthritis exhibit a significant presence of Category D and X drugs, reflecting more restricted use, whereas lower-count conditions like UTI and Pneumonia rely more on Category B treatments. Overall, the chart indicates that, for most common conditions, available pharmaceutical options are predominantly Category C, highlighting a potential need for further clinical investigation into safer alternatives during pregnancy.

4.7 Other medical conditions – OTC vs Rx distribution

There is a strong reliance on prescription (Rx) medications across nearly all conditions. While Hayfever and Osteoarthritis exhibit notable over-the-counter (OTC) accessibility, more complex conditions such as Diabetes and Pneumonia are managed almost exclusively through regulated prescriptions.

5 Visual Analysis

The notebook utilizes a custom visual palette for professional presentation:

- **Primary Blue (#1f4e79):** Used for prescription data.
- **Light Blue (#9dc3e6):** Used for OTC data.

6 Results

The exploratory analysis of the dataset, comprising 3,959 drug records, revealed that Pain (393 drugs), Colds & Flu (246 drugs), and Acne (238 drugs) are the most represented medical conditions. Across nearly all conditions, there is a strong reliance on prescription (Rx) medications, with 2,702 drugs requiring prescriptions compared to 1,257 available over-the-counter (OTC). OTC accessibility is notable for common conditions such as Hayfever and Osteoarthritis, whereas complex conditions like Diabetes and Pneumonia are managed almost exclusively with regulated prescriptions. Analgesic user activity, reflecting engagement rather

than therapeutic potency, shows that Ibuprofen exhibits unexpectedly high usage, likely influenced by factors such as frequency of use, prescription volume, and OTC availability, while Acetaminophen demonstrates a pronounced activity gap compared to other analgesics. For analgesic prescribing patterns specifically, OTC options slightly outnumber prescription-only drugs (236 vs. 157), highlighting consumer preference for accessible pain relief.

The distribution of drugs by pregnancy category underscores limited safety data for many common medications. Over 50% of painkillers fall under Category C, only 13.5% are in Category B, none in Category A, and approximately 30% remain unassessed (Category N). Cold and flu drugs similarly cluster in Categories N and C, with few Category B options and rare Category D drugs, reflecting restricted use due to fetal risk. Across other medical conditions, high-count categories such as Hypertension and Rheumatoid Arthritis include more Category D and X drugs, while lower-count conditions like UTI and Pneumonia rely more on Category B treatments. Overall, Category C dominates most conditions, highlighting gaps in clearly pregnancy-safe options and emphasizing the importance of cautious prescribing, patient counseling, and further clinical evaluation. The findings also indicate that observed trends in drug activity and OTC accessibility are influenced not just by pharmacological properties but by usage patterns, prescription practices, and population-wide behaviors.

7 Conclusion

The analysis demonstrates that while the market for common ailments (Pain and Colds) is diverse, the majority of therapeutic options remain under regulatory prescription control. Future analysis could incorporate the excluded rating data if a more complete longitudinal dataset becomes available.

Appendix: Technical Implementation

The analysis was performed using `pandas`, `seaborn`, and `matplotlib`. Global font settings were forced to *serif* to match professional reporting standards.