This case study explores the relationship between medicines and their reported side effects, with the goal of identifying patterns that can support better understanding of drug safety and risk profiles. Using a large real-world dataset (Kaggle) of pharmaceutical products, the project applies a structured data-analytics approach to transform raw information into meaningful insights.

The analysis focuses on three main dimensions:

- the overall frequency of side effects across medicines,
- differences in side-effect burden between therapeutic categories, and
- contrasts between habit-forming and non-habit-forming medications.

To achieve this, the dataset was cleaned, reshaped, and normalized in PostgreSQL, converting wide side-effect fields into an analysis-ready format. The project follows the Google Data Analytics six-step framework (Ask, Prepare, Process, Analyze, Share, Act), ensuring a clear and professional methodology from problem definition to insight generation.

The outcome of this work is a portfolio-ready case study that demonstrates practical skills in SQL, data preparation, analytical thinking, and insight communication. The results are presented through a combination of structured analysis, visual dashboards, and documented findings, making the project relevant for roles in data analytics, business intelligence, and healthcare-related data environments.

Business Questions

The main objective of this project is to better understand how side effects are distributed across medicines and what this means for drug safety and risk awareness. To guide the analysis, the following business questions were defined:

1. **Which side effects are most commonly reported across medicines?**
   This question helps identify the most frequent adverse reactions that patients are likely to encounter, providing a baseline for safety awareness.
2. **Which therapeutic categories carry the highest side-effect burden?**
   By comparing side effects across therapeutic classes, the analysis highlights categories that may require closer monitoring and stronger patient communication.
3. **Do habit-forming medicines differ from non-habit-forming medicines in their side-effect profiles?**

This question explores whether dependency-related drugs show distinct risk patterns compared to other medications.

4. **Which individual medicines are associated with the largest number of side effects?**
   Identifying high-impact medicines helps surface potential risk hotspots and supports more targeted safety discussions.

Dataset Description

The dataset used in this project contains detailed information about a large collection of pharmaceutical products, including their substitutes, therapeutic classifications, and reported side effects. It was sourced from a publicly available medical dataset and is designed to support exploratory analysis of drug safety patterns.

Each record in the dataset represents a single medicine and includes the following key types of information:

- Basic identification: medicine ID and name.
- Substitutes: alternative medicines that can be used in place of the original drug.
- Side effects: a comprehensive list of reported adverse reactions associated with each medicine.
- Uses: the medical conditions or purposes for which the medicine is prescribed.
- Classifications:
  - *Chemical Class*
  - *Therapeutic Class*
  - *Action Class*
- Habit-forming indicator: whether the medicine has dependency potential.

Dataset size and structure

- The original dataset contains over 248,000 medicines.
- Side effects were initially stored in a wide format, with up to 42 separate columns per medicine.

- For analytical purposes, this structure was transformed into a normalized format, resulting in over 1.6 million side-effect records.

This transformation made it possible to perform meaningful aggregation, comparison, and visualization of side-effect patterns across different dimensions, such as therapeutic class and habit-forming status.

Analysis Performed

With the data fully prepared and structured, the analysis focused on identifying meaningful patterns in side effects across medicines. The goal was not only to describe the data, but to understand risk distribution and highlight areas that deserve closer attention in healthcare and pharmaceutical contexts.

The following key analyses were performed:

1. Most common side effects

The first step was to identify which side effects appear most frequently across all medicines. This provided a baseline view of the adverse reactions that patients are most likely to encounter, helping establish a general risk profile.

2. Side-effect burden by therapeutic category

Side effects were then analyzed by therapeutic class to determine which categories of medicines are associated with a higher concentration of adverse reactions. This made it possible to compare risk levels across different types of treatments and highlight high-impact categories.

3. Habit-forming vs non-habit-forming medicines

The analysis compared side-effect patterns between:

- medicines with habit-forming potential, and
- medicines without dependency risk.

This comparison explored whether these two groups differ in overall side-effect burden and in the types of adverse reactions most commonly reported.

4. Differences in side-effect profiles

To go beyond simple totals, side effects were compared across habit-forming and non-habit-forming medicines to identify reactions that show the greatest variation between the two groups. This helped uncover distinct risk characteristics.

5. Medicines with the highest number of side effects

Finally, individual medicines were ranked by the number of associated side effects. This highlighted products with particularly heavy side-effect profiles, offering concrete examples of high-risk cases.

Together, these analyses provide a comprehensive view of how side effects are distributed across medicines, therapeutic categories, and risk groups — setting the stage for clear insights and actionable recommendations.

Insights & Recommendations

Insights

The analysis reveals that **side effects are not randomly distributed across medicines**. Instead, **clear patterns emerge** that highlight differences in risk concentration, therapeutic impact, and dependency-related profiles.

1. Concentration of side effects in specific therapeutic categories

A small number of therapeutic classes account for a disproportionately large share of reported side effects. In particular, **anti-infectives**, **gastrointestinal**, and **central nervous system** (CNS) medicines show the **highest overall side-effect burden**.
This indicates that patients receiving treatments in these categories are more likely to experience adverse reactions, making these areas critical for monitoring and patient education.

2. Common side effects are consistent across medicine types

Across almost all therapeutic classes, a recurring group of side effects appears **most frequently**, including **nausea**, **headache**, **dizziness**, and **diarrhea**.
These reactions represent a baseline risk that affects a wide range of patients, regardless of the specific condition being treated. Their consistency highlights the importance of setting realistic expectations for patients and ensuring clear communication about what they may experience.

3. Habit-forming medicines show a distinct risk profile

When comparing habit-forming and non-habit-forming medicines, clear differences emerge in both the volume and the type of side effects reported. **Habit-forming medicines** tend to be associated more **strongly** with side effects

linked to the **central nervous system**, such as drowsiness, fatigue, and dizziness, while non-habit-forming medicines show higher representation of gastrointestinal and general physical reactions.
This suggests that **dependency-related drugs** carry a **more neurological side-effect profile**, which may require different monitoring strategies.

4. Some side effects clearly differentiate the two groups

The comparison between habit-forming and non-habit-forming medicines shows that certain side effects are far more common in one group than the other. These differences point to structural distinctions in how these medicines affect the body, reinforcing the idea that risk management should not follow a one-size-fits-all approach.

5. A subset of medicines carries a particularly heavy side-effect load

A small number of individual medicines stand out for having an unusually high number of associated side effects. These products represent high-impact risk cases and are especially important for clinicians and pharmacists to monitor closely.
Their profiles make them strong candidates for deeper clinical review, enhanced patient counseling, and clearer labeling.

Recommendations

Based on the insights from this analysis, the following recommendations can support improved safety awareness and decision-making in pharmaceutical and healthcare settings:

1. Prioritize high-burden therapeutic categories
   Extra attention should be given to medicines in anti-infective, gastrointestinal, and CNS categories. These areas would benefit most from enhanced patient guidance, closer follow-up, and proactive side-effect management strategies.

2. Strengthen patient communication around common side effects
   Since reactions such as nausea, headache, and dizziness are widespread across medicines, healthcare providers should standardize how these risks are communicated. Clear expectations can improve adherence and reduce anxiety when side effects occur.

3. Adopt differentiated monitoring for habit-forming medicines
   Because habit-forming drugs show a more pronounced neurological side-effect profile, monitoring protocols should emphasize symptoms like

sedation, dizziness, and fatigue. This can help reduce safety risks, particularly for patients who drive, operate machinery, or take multiple medications.

4. Flag high-risk medicines for deeper review
   Medicines with exceptionally high numbers of reported side effects should be prioritized for:
   - more detailed clinical evaluation,
   - targeted patient education materials, and
   - clearer warning labels where appropriate.
5. Use data-driven risk profiling in decision-making
   Incorporating side-effect burden analysis into prescribing and product management decisions can help shift practices from reactive to preventive, supporting better outcomes for both patients and healthcare providers.