Drug-Performance Research Project Analysis

Project Overview

This research project aimed to examine the effectiveness and tradeoffs of various drugs by combining user feedback (ratings and reviews) with medication attributes (e.g., side effects, pregnancy category, CSA scheduling) across a wide range of medical conditions. After cleaning and organizing a dataset containing drug names, associated conditions, ratings, and textual side-effect data, we employed grouping and sorting techniques to identify the highest-rated drugs within each condition. We then constructed a custom scoring framework, incorporating a severity scale for side effects, to account for both user satisfaction and potential adverse effects. These analyses enabled a more holistic ranking of treatments, illuminating how certain highly rated drugs may have undesirable side effects, while others strike a more balanced efficacy-risk profile. Ultimately, our work provides a systematic approach to inform drug-selection decisions, offering insights into the interplay between user satisfaction, side-effect severity, and broader medication characteristics.

Understanding the Dataset

Data and Columns of Interest

- Drug Name
- Medical Condition
- Rating (1–10, user-reported)
- CSA (1–5 or N, representing abuse potential)
- Pregnancy Category (A, B, C, D, X, or N)
- Rx–OTC (Rx, Rx/OTC, OTC)
- Alcohol (indicator of alcohol interaction)
- Side Effects (free-text, mapped to severity scores)

We isolated eight conditions of interest—Acne, Depression, Diabetes (Type 2), Pain, Weight Loss, ADHD, Cancer, and AIDS/HIV—to focus our deep-dive. After cleaning and shaping the data, we assigned severity scores for side effects and built a composite "drug_score" that weighs each of the above factors.

Data Analysis

Custom Drug Score Construction

1. **Rating** (Weight = 4.5): Multiplied the user rating (range 1–10) by 4.5.
2. **Rx–OTC** (Weight = 5.0):
    - Rx = 0 points
    - Rx/OTC = 1 point

    - OTC = 2 points
      After assigning points, multiplied by 5.0. This rewards over-the-counter availability.
3. **CSA (Abuse Potential)** (Weight = 3.0):
    - Schedule 1 = 5 points, …, Schedule 5 = 1 point
    - No schedule = 0
      A higher CSA value corresponds to more stringent scheduling, thus a higher numeric score (5) implies a more regulated substance. After mapping, we multiply by 3.0.
4. **Pregnancy Category** (Weight = 1.5):
    - A = 5, B = 4, C = 3, D = 2, X = 1 (and N = 0 or NaN)
      Higher means safer. Multiplied by 1.5.
5. **Alcohol Interaction** (Weight = 2.5):
    - X = 0 (known interaction)
    - NaN = 1 (no known interaction)
      Multiplied by 2.5.

Side Effects: In addition to these columns, we prepared side-effect severity scores (1–5). Although we haven't fully integrated them into this numeric total yet, they provide the foundation for future penalty-based adjustments if a drug causes severe or frequent side effects.

Top Drugs Overall

When we sorted all drugs by the new drug_score (descending), we noticed a pattern:

- Many entries with high user ratings (9–10) plus OTC status and favorable pregnancy category soared to the top of the list.
- Certain Rx-only medications with high ratings still scored well if they had a moderate CSA schedule (e.g., 3 or 4) and no alcohol interaction.
- Highly regulated substances (CSA = 2) can still appear at the top if their user ratings are perfect (10), indicating that the weighting factor for rating is very influential in the current formula.

Top 5 Drugs per Condition of Interest

We then identified the top five for each key condition. Here are some notable highlights:

1. **ADHD**
    - *ProCentra, Desoxyn, methamphetamine* rank high due to strong user ratings and presumably decent pregnancy or CSA scoring. However, these are potent stimulants (Schedule 2). It suggests a high rating outweighs the penalty of stricter scheduling in our current weighting scheme.
2. **AIDS/HIV**
    - *Reyataz, atazanavir, Tivicay PD, Epzicom, Stribild* surface as top contenders. Many have a B or C pregnancy category, so they don't get heavily penalized. Those with no alcohol interaction or no scheduling constraints also score higher.
3. **Acne**

    - *PanOxyl 10% Acne Foaming Wash, Fostex Medicated, Pernox, salicylic acid/sulfur, Acnex* appear near the top. Over-the-counter availability (OTC) combined with high ratings boosts them significantly in this system.
4. **Cancer**
    - *doxorubicin, vincristine* have a rating of 10 in the data, giving them a large boost. Others (e.g., carboplatin, ifosfamide) come in lower because of lower user rating or interactions.
5. **Depression**
    - *Forfivo XL, amoxapine, isocarboxazid, Marplan, trimipramine* rank high. They benefit from good user ratings and no known alcohol penalty, but are still prescription-based.
6. **Diabetes (Type 2)**
    - *Glyset, miglitol, chromium picolinate, Invokamet XR, Glumetza* show up top. Over-the-counter items (like chromium picolinate) had notable gains from the Rx–OTC weighting.
7. **Pain**
    - *Alfentanil, Xodol, AneCream, Bactine, Medi-Quik Spray* stand out. Interestingly, strong analgesics (alfentanil, schedule 2) remain top due to perfect user ratings (10). OTC topicals also soared to 61.00 if they had rating=10, reflecting the large impact of rating plus the 5.0 weighting for OTC.
8. **Weight Loss**
    - *Diethylpropion, phendimetrazine, Adipex-P, phentermine, benzphetamine* surface, all are stimulants with moderate to high CSA schedules, but strong user ratings. This indicates the rating factor plus partial penalty from CSA and Rx usage.

Observations and Limitations

1. **High Influence of Ratings**: Since we multiplied rating (1–10) by 4.5, a perfect rating heavily outweighs negative aspects such as Rx status or a stricter CSA schedule.
2. **OTC Advantage**: Over-the-counter products with rating=10 can achieve extremely high scores. This may reflect *accessibility* in our scoring logic, but it may overlook efficacy nuances or potential side-effect severity.
3. **Subjectivity in Weighting**: We used domain-inspired, yet somewhat **arbitrary** weights. A small tweak (e.g., lowering the rating multiplier, raising the CSA penalty) could significantly rearrange the top list.
4. **Side-Effect Severity Not Fully Integrated**: We listed side effects and assigned severity but did not yet incorporate a direct penalty into the final numeric score. A drug with extreme side effects but great user ratings could appear artificially high.

Final Piece

Future Directions

1. **Refining the Weights**: Conduct a sensitivity analysis to see how shifting the importance of each factor (rating, pregnancy category, CSA schedule, etc.) affects the top-ranked results.

2. **Integrating Side-Effect Penalties**: Implement a penalty system that subtracts points for frequent or severe side effects, ensuring that high-rating drugs with dangerous profiles drop in rank.
3. **Handling Ties and Rare Conditions**: If multiple drugs tie at the same rating, we could further differentiate them by average review count, side-effect frequency, or additional safety data.
4. **Temporal or Real-World Usage Patterns**: If the dataset includes any time trends or prescribing volumes, we can explore whether higher "real-world usage" correlates with improved or worsened user feedback over time.

Conclusions

Our custom drug_score approach offers a convenient, if simplified, method to compare medications across diverse attributes—an especially valuable starting point for broad, data-driven "best drug" discussions. By factoring in user rating, availability, pregnancy category, scheduling, and (potentially) side-effect burden, we obtain a more nuanced ranking than a one-dimensional user rating might provide.

However, caution is warranted: real-world usage and safety is complex, and each weighting or scoring threshold is subject to interpretation and bias. Future refinements could better account for side-effect severity, incorporate more detailed cost or outcome data, and address domain-specific nuances. Still, this initial analysis demonstrates how a systematic, multi-criteria scoring system can bring clarity to large drug datasets and highlight top candidates for further investigation in each medical condition.