EDA INSIGHTS:

1. After removing irrelevant columns and narrowing the data to three specific medical conditions — **Depression**, **High Blood Pressure**, and **Diabetes, Type 2** — the dataset is now:
- **Cleaner** (no redundant Unnamed: 0 column)
- **More focused** on the most common and impactful conditions
- **Easier to manage**, with a reset index and fewer categories
2. The   filtered dataset now contains:
    a. **13,944 rows** (individual drug reviews)
    b. **6 columns**: drugName, condition, review, rating, date, usefulCount
    c. You're working with **334 drugs** across **3 conditions**, with **very diverse review text**
3. **Depression** makes up **~65%** of the dataset. **Diabetes, Type 2** and **High Blood Pressure** are much smaller in comparison (~18% and ~17%, respectively).

Condition	Number of Reviews	Percentage
Depression	9,069	~65.0%
Diabetes, Type 2	2,554	~18.3%
High Blood Pressure	2,321	~16.6%

4. The rating column shows a **bimodal distribution**:
- Most reviews are **highly positive (10)**
- A significant number are **very negative (1)**
- Users tend to rate drugs **very positively (10)** or **very negatively (1)**.
- There's a **dip in the middle (ratings 4–7)**, showing fewer neutral responses.
5. The **Bupropion** leads the list, indicating it is:
- Widely prescribed
- Possibly has a strong user impact (good or bad)
- Often reviewed (maybe due to effectiveness or side effects)
6. The **Univariate Analysis Insight (for review_length):**
- Most reviews contain between **10 to 80 words**.
- The **peak frequency** is around **50–60 words**.
- Some reviews go up to **200+ words**, but these are **very rare.**
7. The **Univariate Analysis Insight (for Sentiment):**
- **Positive reviews (≈ 9,800)** heavily outweigh **negative reviews (≈ 4,100)**
- Roughly **70% of the reviews are positive**, while only about **30% are negative**
8. The **EDA Insight :**
- **Median review length** is **similar** for both positive and negative sentiments.
- **Negative reviews have more extreme outliers** — a few go above **250–300 words**, while positive reviews are mostly below **200**.
- The spread (IQR) is roughly the same: most reviews lie between **25 to 60 words** for both classes.
- **Outliers are more frequent in positive reviews**, but **longer outliers** occur more in negative reviews.
9. The **Insights from Word Frequencies by Rating:**
- ⭐ **Ratings 1–4 (Very Low Ratings):**
- Common words: day, taking, week, effect, side, medication, doctor
- Indicates focus on:
    o **Side effects** or **immediate negative reactions**
    o **Dosing or prescription issues**
- High presence of words like doctor, drug, started suggests dissatisfaction with treatment or its initiation.

- ⭐ **Ratings 5–7 (Moderate Ratings):**
- Words become more balanced between **experience** and **effectiveness**
- feel, year, depression start showing up more — mixed feedback
- User tone here is **cautious optimism** or **neutrality**
- ⭐ **Ratings 8–10 (High Ratings):**
- Clear shift in tone:
  - Words like life, feel, year, depression, effect dominate
- Reviews reflect **longer-term impact** (e.g., "this drug changed my life")
- High ratings often include reflection over **extended use** (year, month, week)
- Strong **emotional positivity** — especially with word life appearing for Rating 10
10. **Top Bigrams:**

Bigram	Count
side effect	6042
blood pressure	1951
feel like	1038
blood sugar	901
started taking	898
depression anxiety	791

- **Top Trigrams:**

Trigram	Count
high blood pressure	338
sexual side effect	229
feel much better	222
bad side effect	147

11. **Combined Insight**
- Both **positive and negative sentiments** discuss similar words like mg, week, and taking—but the **context** differs:
  - Negative reviews frame them around **issues, side effects, or improper dosage**.
  - Positive reviews describe them as **routine, manageable, and effective**.
- Words like "feel", "life", and "depression" are highly **sentiment-dependent** and are strong indicators in **positive reviews**.
12. **Depression:**
- **Most reviewed condition**, with ~9,000 reviews.
- **~28% Negative Sentiment** (~2,400–2,500 reviews).
- **~72% Positive Sentiment** (~6,300–6,500 reviews).

Despite being the most discussed, depression treatments have a **relatively high proportion of negative sentiment**, suggesting **side effects**, **limited efficacy**, or **delayed improvement** as common concerns.
Still, the majority of patients report **beneficial outcomes**, showing that many medications are **effective for a large group**.

Diabetes, Type 2:
- **~2,500 total reviews**.
- **~35–40% Negative Sentiment** (~900–1,000 reviews).
- **~60–65% Positive Sentiment** (~1,500–1,600 reviews).

Diabetes medications are **well-received overall**, with **high satisfaction rates**.
Positive feedback may reflect the **measurable, short-term impact** these drugs have on blood sugar levels.
Negative reviews likely stem from **side effects** or **difficulties in lifestyle adherence**.

High Blood Pressure:

- **~2,300–2,400 total reviews**.
- **~40–45% Negative Sentiment** (~900–1,000 reviews).
- **~55–60% Positive Sentiment** (~1,300–1,400 reviews).

Feedback is **evenly distributed**, indicating **mixed experiences** with hypertension drugs.

While many users benefit, a **sizable portion report dissatisfaction** — potentially due to **side effects like fatigue, dizziness**, or **slow symptom relief**.