Data Cleaning Questions:

1. Title

- Are there **duplicate book titles**? Should duplicates be removed or kept?
- Are there **leading/trailing spaces** or inconsistent capitalization?
- Are there any **special characters** that need cleaning?

---

2. Author

- Are there **missing author names**? How to impute them?
- Are **author names consistent** (e.g., “J.K. Rowling” vs “JK Rowling”)?
- Are there **multiple authors** listed in one cell that need splitting?

---

3. Narrator

- Any **missing narrators**? Should you fill with “Unknown”?
- Are names **formatted consistently**?
- Check for **duplicate entries** due to minor spelling differences.

---

4. Rating

- Are ratings numeric? If not, **convert to float**.
- Are there **out-of-range ratings** (e.g., >5 or <0)?
- Handle **missing ratings**: remove or impute?
- Check for **inconsistent decimal formatting** (e.g., 4.5 vs 4,5).

---

5. Reviews

- Are review counts **numeric**? If not, convert to integer.
- Handle missing reviews. Should they be 0 or NaN?
- Detect **outliers** (e.g., extremely high review counts).

---

6. Duration

- Durations may be in **text format** (e.g., "17 hrs and 5 mins").
    - Split into `hours` and `minutes`.
    - Convert total duration into **minutes** for numeric analysis.
- Handle special cases like “Less than a minute” or missing duration.

---

7. Release Date

- Convert to **datetime format**.
- Are there **missing or invalid dates**?
- Extract **release year** as a new feature.

---

8. Language

- Ensure **consistent formatting** (e.g., lowercase).
- Group similar languages if needed (e.g., “English (UK)” → “English”).
- Handle missing values: fill with “Unknown” or remove?

---

9. Price

- Convert price to **numeric format** (remove currency symbols).
- Handle missing or zero prices.
- Detect **outliers** or unreasonable values (e.g., very high or negative prices).

---

10. Genre

- Ensure **consistent category names**.
- Are there **multiple genres in one cell**? Split or choose primary genre.
- Handle missing genres appropriately.