## Book Metadata Analysis & Visualization Documentation

Project Overview:

**Task 1: Data Preparation & Cleaning**

Objective: Ensure the dataset is analysis-ready.

Steps:

Load the dataset using Pandas or Power BI.

Clean the Language column (handle multi-language entries like "Sanskrit , Hindi").

Standardize column names (e.g., lowercase, no special characters).

Convert Data into proper numeric types.

Handle missing values (e.g., fill with "Unknown", or drop rows/columns).

**IMPLEMENTATION :**

1. Load Excel Dataset

Home → Get Data → Excel

Select your file → Choose the relevant sheet → Click Transform Data

2. Clean the Language Column

Replace Values:

Replace " ," → ","

Replace ", " → ","

Format → Lowercase

3. Standardize Column Names

Right-click each column → Rename:

Use all lowercase

Remove special characters

4. Convert Data Types

Select column → Transform → Data Type

Make a Date

• There is only a year (like 2020), make a date like 1st Jan 2020

• So Power BI knows it’s a real date and can show month/week/day

5. Handle Missing Values

Fill with Default

Transform → Replace Values

Replace null or blank with "Unknown" or 0

Remove Missing

Home → Remove Rows → Remove Blank Rows

**In this Excel folder there are 4 sheets, so I just make all column names similar and remove all unnecessary columns from tables that will not be used in analysis and combine all the tables by append function by Power Query Editor.**

**Task 2: Dashboard Creation & Summary Statistics**

Objective: Understand the characteristics of the collection and Build a professional, interactive dashboard to showcase insights.

key points that should be visualized in the dashboard:

Book as Per a Day/Month/Week(If a Date column is required, please add it as appropriate).

Total Number of Books

Total Number of Pages per Book

Oldest and Newest Year of Publishing

Number of Unique Authors & Top Authors

Number of Total Publishers & Top Publisher

Total Contributors (Digitized by)

Identify the most common languages and authors.

Here, you have a Language Column that contains combinations of multiple languages in a single row, like: "English, Hindi", "Marathi, Sanskrit", "Hindi, Telugu, English", "English" etc. You want to select a single language (e.g., "Hindi") in a slicer, and filter the data to return all rows that contain that language, even if it's mixed with others.

**IMPLEMENTATION :**

**For Summary Page**

Visuals Used:

1. KPI Cards: Total books (115), publishers (44), authors (47), contributors (4), languages (345)
2. Line chart: Books published per year (1905–2017)
3. Donut Chart: Distribution of books by languages
4. Bar charts: Top 10 authors & Top publishers
5. Date slicers and dropdown filters for languages and publishers

Insights Or Results:

* Books range from year 1905 to 2017
* Most books are in Sanskrit, then Hindi
* Author Amarnath Krishna Meen has published the most books
* Yaksha Prashna is the top publisher

**Detailed Page – “Pages Details and Contributors”**

Visuals Used:

1. Table: Lists book names and their respective total pages
2. Donut Chart: Shows distribution of contributors (digitized by)

Key Results :

Top contributor: *yaksha-prashna* (44.35%)

Book “Chitraprabha” has the most pages (1380), followed by “ashvlayan shraut sutram 2” (1374).

**Main Implementation:**

To **filter rows based on a single language**, even when a cell contains **multiple languages**, like:

Hindi, English

Sanskrit, Hindi

English, Hindi, Marathi

STEPS :

**1. Duplicate the Table**

**2. Split the Language Column by Comma**

"Hindi, English" becomes:

Column1: Hindi

Column2: English

**3. Unpivot the Language Columns**

Row1 → Hindi, English

Convert into:

Row1 → Hindi

Row1 → English

So now every language is in its **own row**, which is perfect for slicers and filtering.

Result:

You get a new column like this:

Language column will be Normalized

Hindi

English

Sanskrit

Now, when you add a **slicer using this column**, it will correctly filter **any row that contains that language**, even if it originally had multiple.