**Data Collection**

**Data:**

The quantities characters’ facts or symbols on which operations are performed by a computer to convert into meaningful information.

**What is data collection?**

Data collection is the process of gathering the data from different sources and measure and analyze to reach the answers, evaluate outcomes and forecast trends and probabilities.

Data collection is used in many fields, including physical and social sciences, humanities and business.

**Data collection methods:**

* **Primary data collection**

This technique is used by researchers. This is initial information gathering step before anyone carries out any further or related research.

* **Secondary data collection:**

Secondary data is second-hand data collected by other parties and already having undergone statistical analysis.

**Which method is used to collect data for this product?**

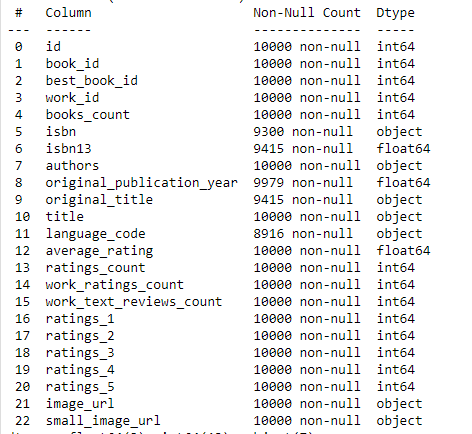
Secondary data collection method is used in this project. Because purpose of data collection is data analysis and project development not for research purpose. it is simple method of data collection because there are no specific rules.

**Books Data Collection:**

**Books Dataset:**

Books data set contains information about books data. It has more than 10000 values and 23 different columns which related to books.

**Books dataset summery**



**Columns Description**

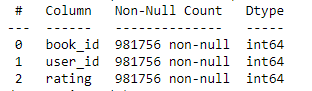
* **Book\_id column:** has unique ID for each book, integer data type, not null values. Total values are more than 10000.
* **Best book\_id column:** has unique ID for best books, but in dataset every book is best book for readers, integer data type, not null values. Total values 10000.
* **Books count:** define total number of books with same title. Each book has different numbers, integer data type, not null values. Total values 10000.
* **isbn:** has unique ISBN for each book, object data type (values are in string format), null values are present in column. Total values are 9415.
* **Isbn13:** has unique ISBN for each book, float data type, null values are present in column. Total values are 9445.

(difference between isbn and isbn13 is that isbn contains object data type values and isbn13 contains float data type values)

* **authors:** name of authors of each book, object data type (values are in string format), null values are not present in column.
* **original\_publication\_year:** in this column values are about year of publication, float data type, null values are present in column. Total values are 9979.
* **original\_title:** in this column title of each book, object data type, null values are present
* in column. Total values are 9979.
* **Average rating:** average rating of book that how many users rate this book, float data
* type, no null values in column. Total values are 10000.
* **Average rating\_count:** count rating of book that how many users rate this book, float
* data type, no null values in column. Total values are 10000.
* **Image URL:** in this column image URL of each book is present, object
* data type, no null values in column. Total values are 10000.

**Rating Dataset**

This data set has all information about users and ratings of books. It has 3 columns name as User\_id, book\_id and rating.



**Columns description**

* **Book\_id:** in this column all the unique ids of books. Values are integer, no null values.
* **user\_id:** in this column all the unique ids of users. Values are integer, no null values.
* **rating\_id:** number of ratings on each book. Values are integer, no null values.