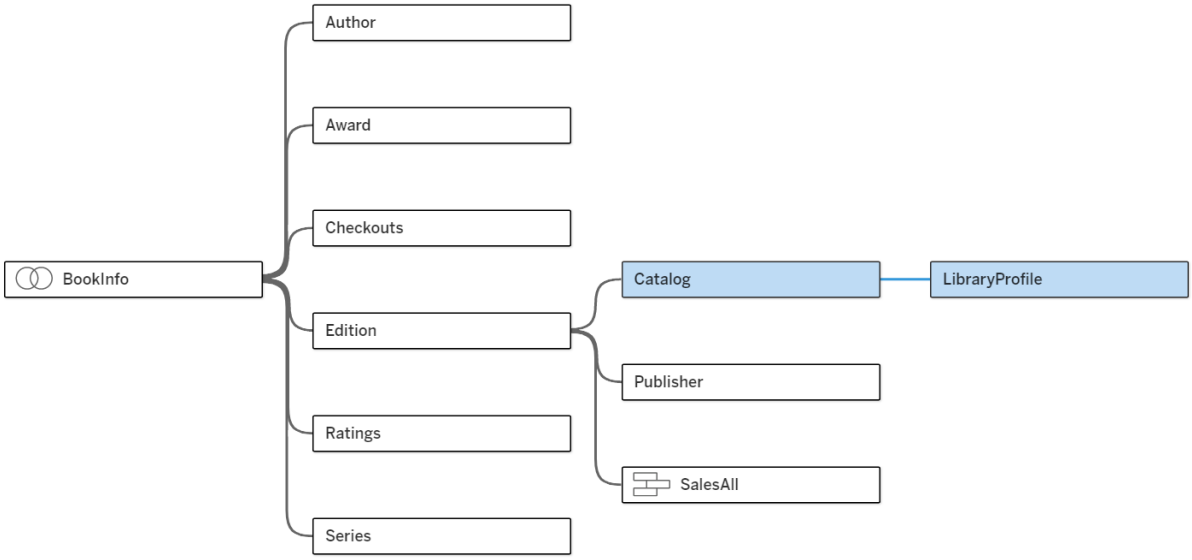
Assignment 6 Instructions

In this assignment (assignment 6), you will be creating a new Tableau workbook which will use data from two Excel files. You will need to connect the different worksheets within the two Excel files, forming the data model shown below. Then you will need to create visualizations to answer the questions further below.

Complete the instructions below to connect your data and answer the exploratory analysis questions below. We will also review this assignment in class.

1. **Let's get started**: Download the following Excel files - **A6_Bookshop.xlsx** and **A6_BookshopLibraries.xlsx** from Canvas. You will need both for this assignment.
2. Open a new Tableau workbook and Add Connections for the two Excel files.
3. Connect the tables as shown in the model below, including (a) an inner join between Book and Info; a union between Sales Q1-Sales Q4, and relationships for the remaining connections. Hint: the inner join between Book and Info will need to be specified with "BookID = [BookID1]+[BookID2]"
4. Explore the data in Tableau and answer the 14 questions below.



Exploratory Analysis

Create visualizations that will answer the questions below. One on each worksheet. Include your answer for each question below the title for each worksheet. **You can use the title or an annotation to provide your answer.**

Note: This fictitious data is set in the future. **Assume that the current date is January 1, 2194.**

1. What books are the most popular and the least popular based on sales, checkouts and ratings?
2. Who was the youngest debut author? Who was the oldest?

Hint: Create a calculated field of Birthday – Publication Day. Then create a text table of First Name, Last Name, ISBN, BookID, Publication Date and Birthday and the calculated field.
If you wanted to calculated the author's age just to confirm your understanding, use the formula Datediff('year', date("01-01-2194", [Birthday]).

3. Do some publishing houses seem to specialize in any way?
4. What was the longest time between editions of the same book?
   Hint: Create 3 calculated fields - Max of Publication Date, Min of Publication Date, DateDiff of the Max and Min. Then create a text table of Book ID, Title, Max(), Min() and DateDiff

5. Are there any seasonal trends for sales? What about checkouts?

Questions# 6 and 7 compare quantitative variables (measures). You might be thinking scatterplots but given our data model specifications and complexity, that can be challenging. Try a different approach. We'll start with author. Display author name on the row shelf, then drag measure values to columns. Measure names should appear in the Filters shelf. Click on the down arrow in the Measure names pill and select Show Filter. Next drag SalesAll(Count) to the Size card on the Marks shelf. Now we can look at the relationship between Sales and several other quantitative measures in the filter.

6. Are there any correlations between checkouts, print run size, book review ratings, and sales volume? (you may create multiple worksheets for this question if needed).
7. Do the authors who spend the most time writing have the most successful books? Do they have the highest page count?

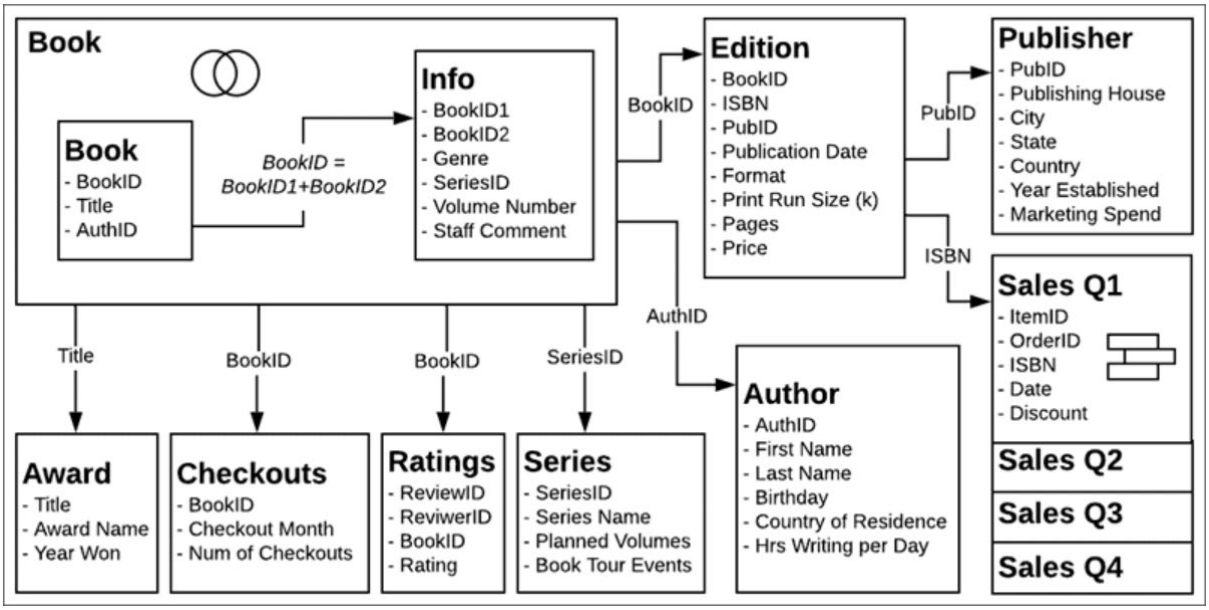
8. When are most books published? Are there any anomalies?
   Hint: Look at weekday

9. Are there any trends for genre, format, and price?
10. What sort of distributions do the ratings have? Do those distributions vary by book? By genre?
11. How would you calculate the sales price, given that there is sometimes—but not always—a discount given at the time of sale?
    Hint: Create a calculated field Sales price = [Price] – IFNULL([Price][Discount],0)*

12. Do sales approximate the Pareto principle (80/20 rule)?

    Hint: Create a table with Book title on Rows. Add SalesAll (count) to the table view. CNT(SalesAll) should be added as text to the Marks area. Now explore using Quick Table Calculations.

13. Are there any patterns in the discounts?
14. Do any tables in particular appear to have dirty data?
    Hint: Check out Author and consider the current date of January 1, 2194.


Book
Book
- BookID
- Title
- AuthID
BookID =
BookID1+BookID2
Info
- BookID1
- BookID2
- Genre
- SeriesID
- Volume Number
- Staff Comment
BookID
Edition
- BookID
- ISBN
- PubID
- Publication Date
- Format
- Print Run Size (k)
- Pages
- Price
PubID
Publisher
- PubID
- Publishing House
- City
- State
- Country
- Year Established
- Marketing Spend
ISBN
Sales Q1
- ItemID
- OrderID
- ISBN
- Date
- Discount
Sales Q2
Sales Q3
Sales Q4
AuthID
Author
- AuthID
- First Name
- Last Name
- Birthday
- Country of Residence
- Hrs Writing per Day
Title
BookID
BookID
SeriesID
Award
- Title
- Award Name
- Year Won
Checkouts
- BookID
- Checkout Month
- Num of Checkouts
Ratings
- ReviewID
- ReviwerID
- BookID
- Rating
Series
- SeriesID
- Series Name
- Planned Volumes
- Book Tour Events