

What Makes a Book a Best-Seller?

Group 6:

Thuria, Janie, Manmita and Mingming

Objective Recap

BIG QUESTION TO ANSWER -

What factors can impact whether or not a book becomes a best-seller?

INITIAL FACTORS TO EXPLORE IN DATA -

- ◆ Genre
- ◆ Audience Gender
- ◆ Audience Age
- ◆ Author
- ◆ Comparing different best-seller lists

INITIAL SUMMARY OF FINDINGS -

Age and gender information is more scarce, and comparing books across lists is tricky due to differences in formatting

Questions & Data

FINAL FACTORS TO EXPLORE IN DATA -

- ◆ Genre
- ◆ Publisher
- ◆ Length of Book
- ◆ Book Price
- ◆ Publish Date

DATA USED -

- ◆ HISTORICAL DATA - Top 100 books of all time from 1998-2010 (Nielson Books UK)
- ◆ CURRENT DATA - NY Times Best-Sellers from 2011-2018 (NY Times)
- ◆ EXTRA FACTORS - Types of books bought most frequently (GoodReads)

HISTORICAL DATA

Cleaning, Exploration, and Analysis

Cleaning and Exploration

AVAILABLE COLUMNS -

Position, ISBN, Title, Author, Imprint, Publisher Group, Volume, Value, RRP, ASP, Binding, Publ Date, Product Class

CONVERT CURRENCY -

- ◆ Strip column values of “£” sign
- ◆ Change type from string to float
- ◆ Convert pounds to dollars

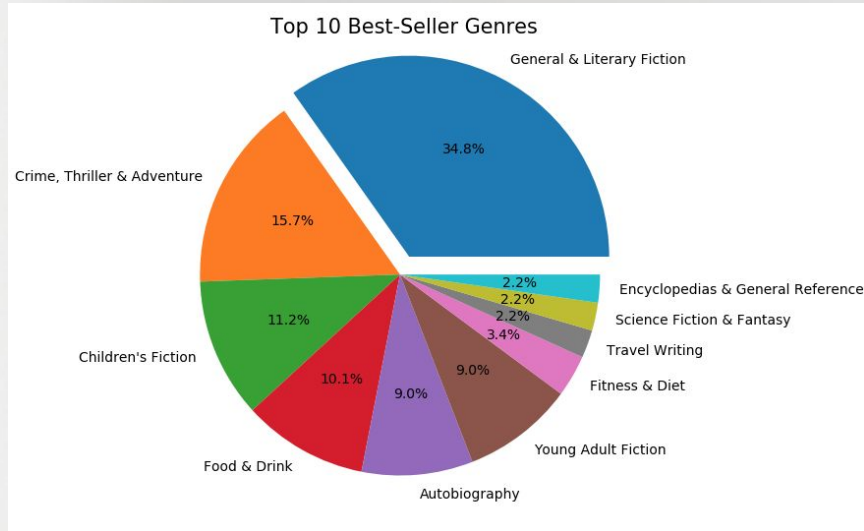
EXTRACT MONTH AND YEAR FROM DATA -

- ◆ Change date to datetime
- ◆ Create new year and month column by extracting year and month

DEFINE SEASON -

- ◆ Establish season by month value
- ◆ Loop through month column and append new column based on season

Analysis - Genre

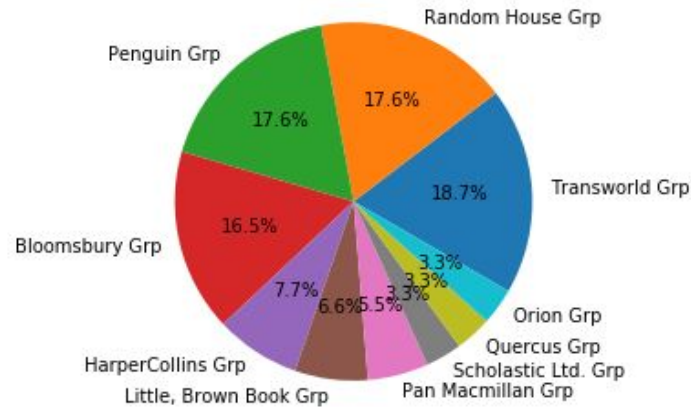


Top Genres:

- ◆ General Fiction
- ◆ Crime, Adventure and Thriller
- ◆ Children's Fiction

Analysis - Publisher

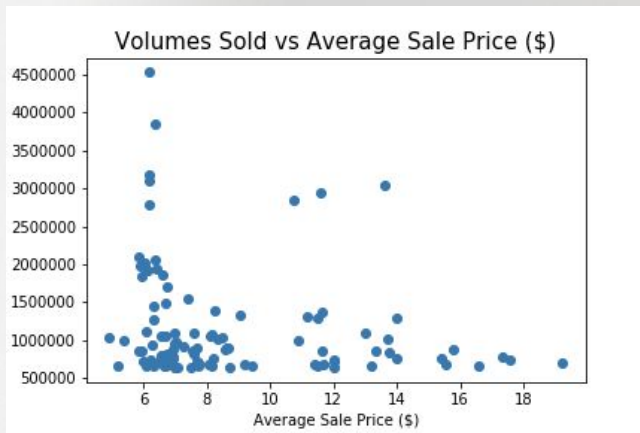
Top 10 Publishers from 1998 - 2010



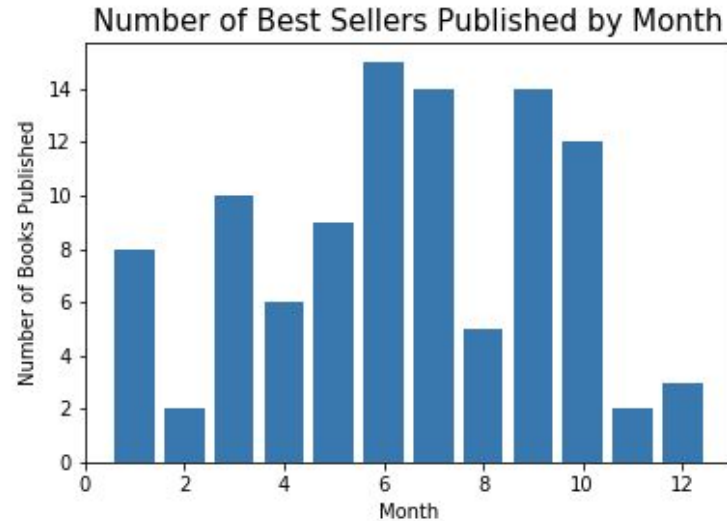
Top Publishers:

- ◆ Transworld Group
- ◆ Random House Group
- ◆ Penguin Group

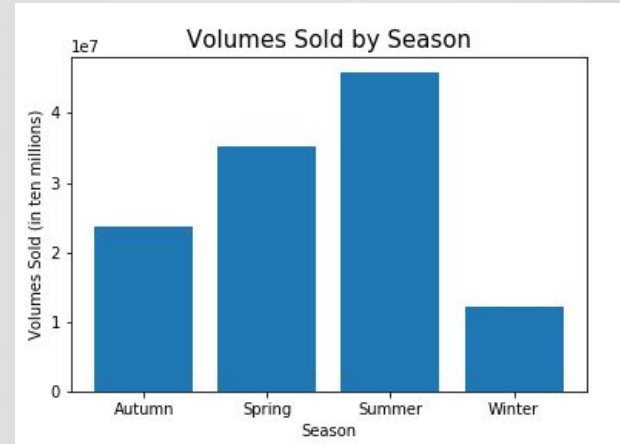
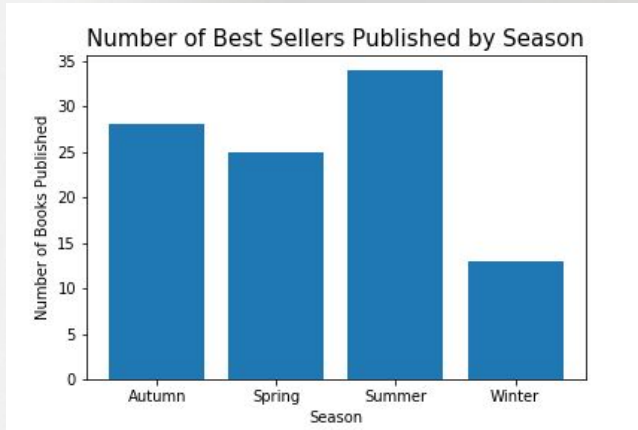
Analysis - Sale Price (\$)



Analysis - Publish by Month



Analysis - Publish by Season



CURRENT DATA

Cleaning, Exploration, and Analysis

Cleaning and Exploration

AVAILABLE COLUMNS -

Publisher, Author, Primary_isbn10, Primary_isbn13, Date, Contributor, Title, Weeks_on_list

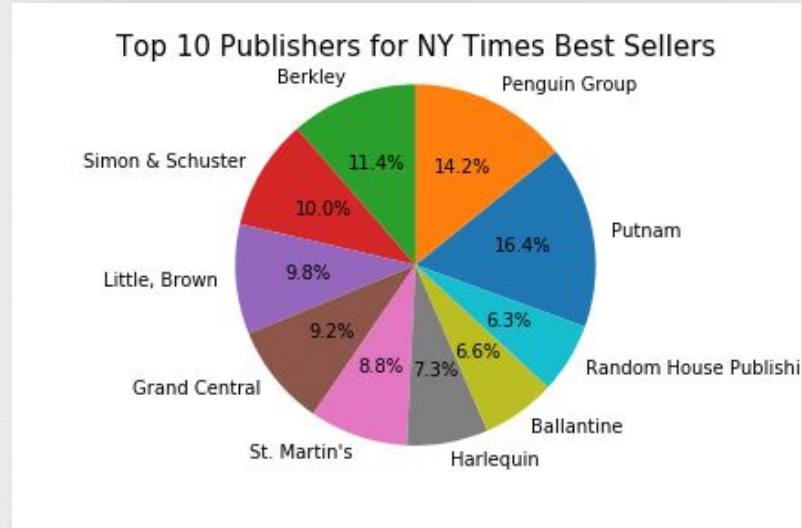
EXTRACT MONTH AND YEAR FROM DATA -

- ◆ Change date to datetime
- ◆ Create new year and month column by extracting year and month

DEFINE SEASON -

- ◆ Establish season by month value
- ◆ Loop through month column and append new column based on season

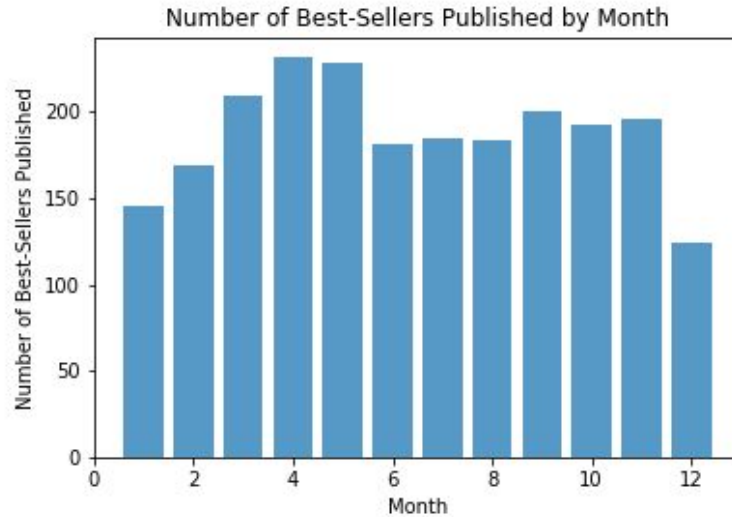
Analysis - Publisher



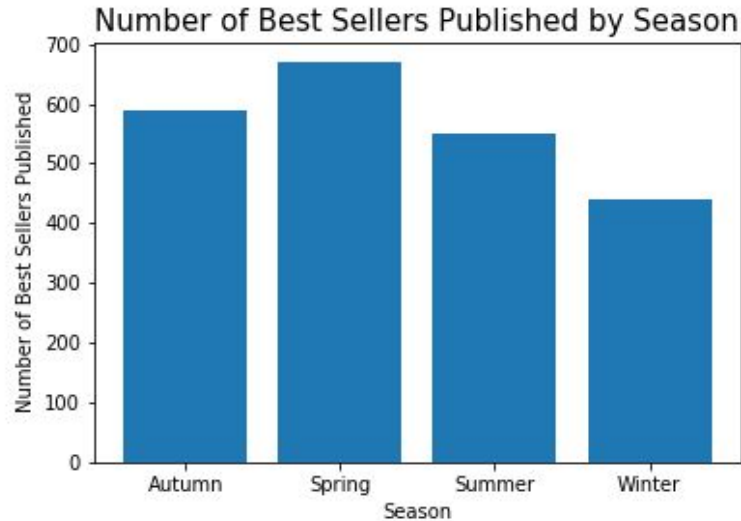
Top Publishers:

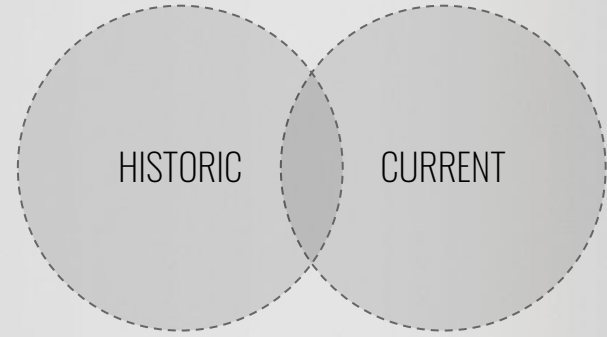
- ◆ Putnam
- ◆ Penguin Group
- ◆ Berkley

Analysis - Publish by Month



Analysis - Publish by Season



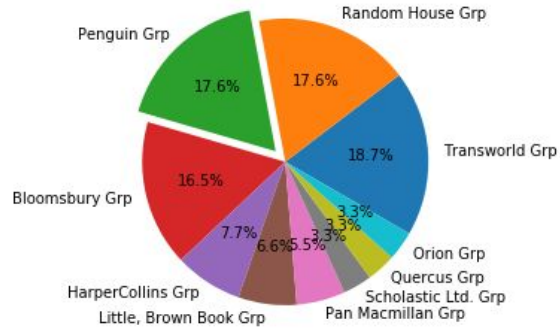


SIMILARITIES?

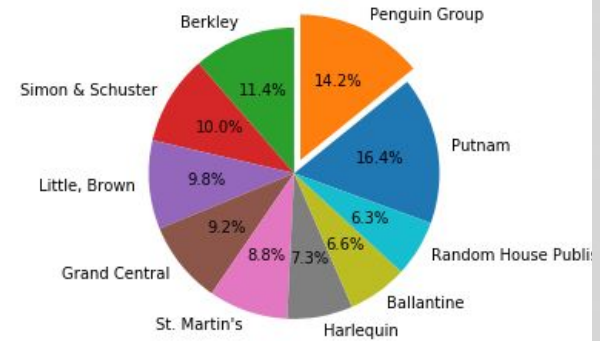
Were there similar themes with the data?

Publisher

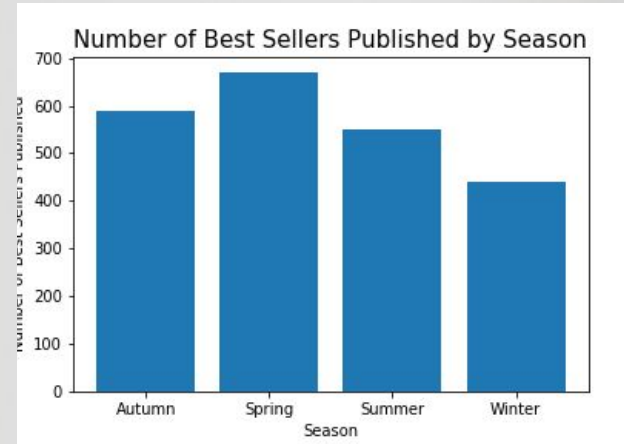
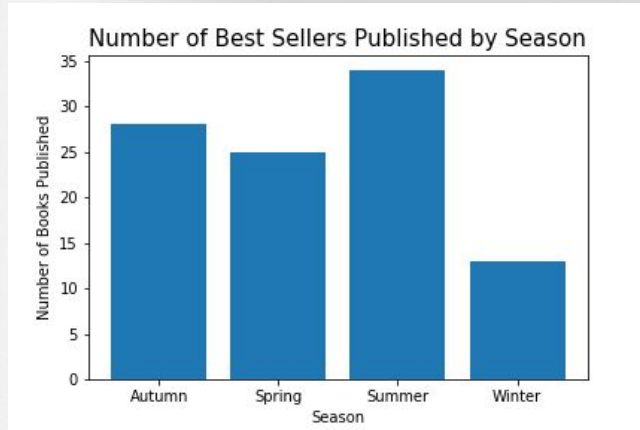
Top 10 Publishers from 1998 - 2010



Top 10 Publishers for NY Times Best Sellers



Publish Season



ONE STEP FURTHER

Cleaning, Exploration, and Analysis of Extra Book Info

Cleaning and Exploration

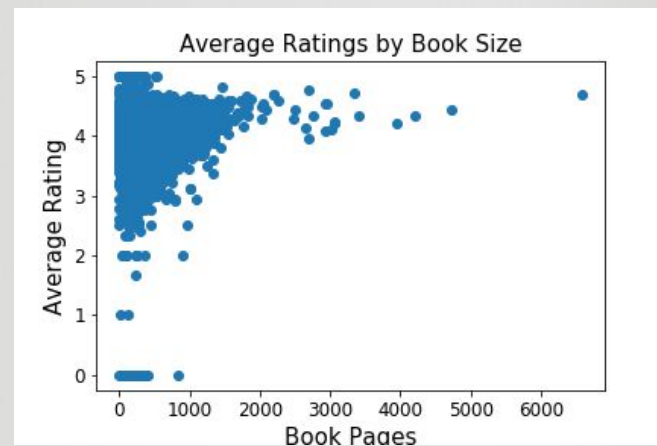
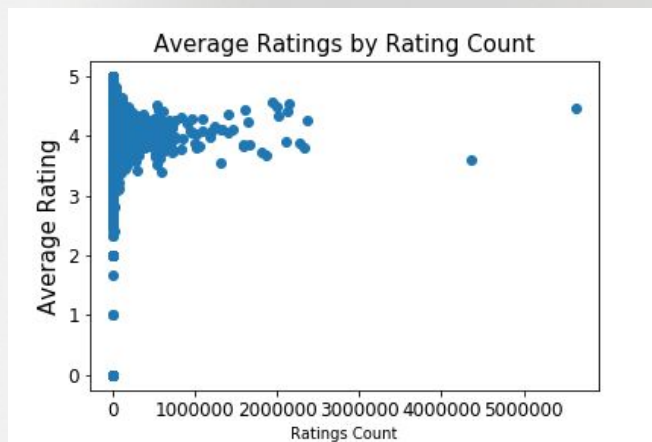
AVAILABLE COLUMNS -

BookID, Title, Authors, **Average_rating**, isbn, isbn13, Language_code, **# num_pages**, **Ratings_count**, Text_reviews_count

EXTRA TOPICS TO EXPLORE -

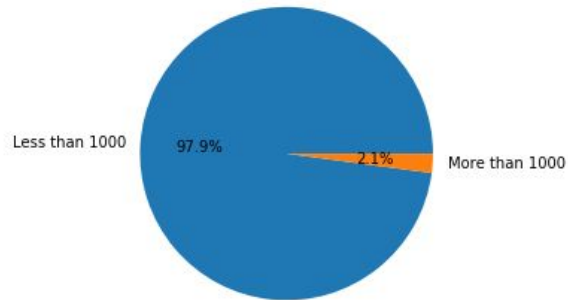
- ◆ Average rating
- ◆ Length of Book
- ◆ Number of Rating

Average Rating - 3.93 out of 5

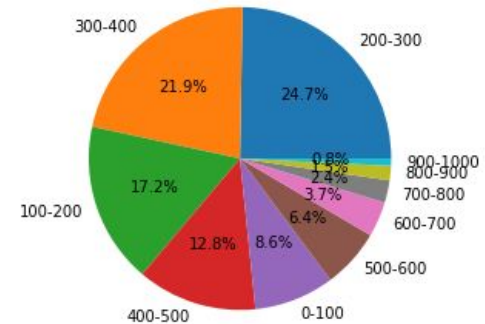


Length of Book

Percentage of Book Pages Greater or Less Than 1000



Percentage of Book Pages from 0-1000



CONCLUSION & REFLECT

What factors will increase your chances of being a best-seller?

Conclusion

FACTORS FOR A BEST SELLER -

- ◆ Genre - **General Fiction**
- ◆ Publisher - **Penguin/Transworld Group**
- ◆ Length of Book - **100-400 pages**
- ◆ Book Price - **\$10 or less**
- ◆ Publish Date - **Spring/Summer**

Reflection

WHAT IF WE HAD MORE TIME?

- ◆ Explore more statistical tests to try against the data
- ◆ Pull data from APIs
- ◆ Look into age/gender data

QUESTIONS?

Thank you for listening!