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 INFORMATION - 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
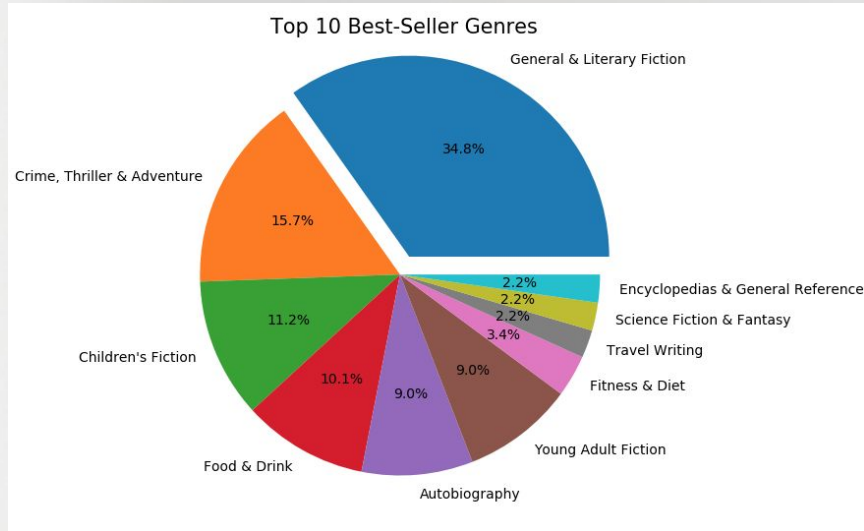- Change type from string to float
- Convert pounds to dollars

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
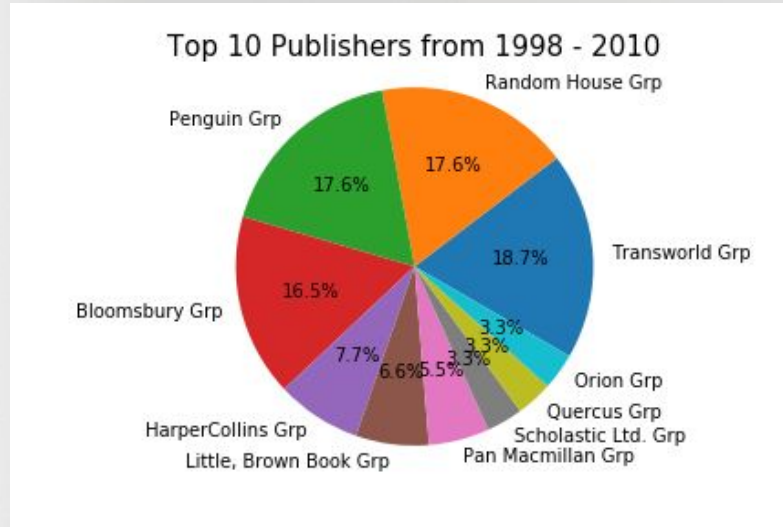- Loop through month column and append new column based on season

Analysis - Genre

Top 10 Best-Seller Genres

Top Genres:

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

Analysis - Publisher

Top 10 Publishers from 1998 - 2010

Random House Grp 17.6%
Penguin Grp 17.6%
Transworld Grp 18.7%
Orion Grp 3.3%
Quercus Grp 3.3%
Scholastic Ltd. Grp 3.3%
Pan Macmillan Grp 5.5%
Little, Brown Book Grp 6.6%
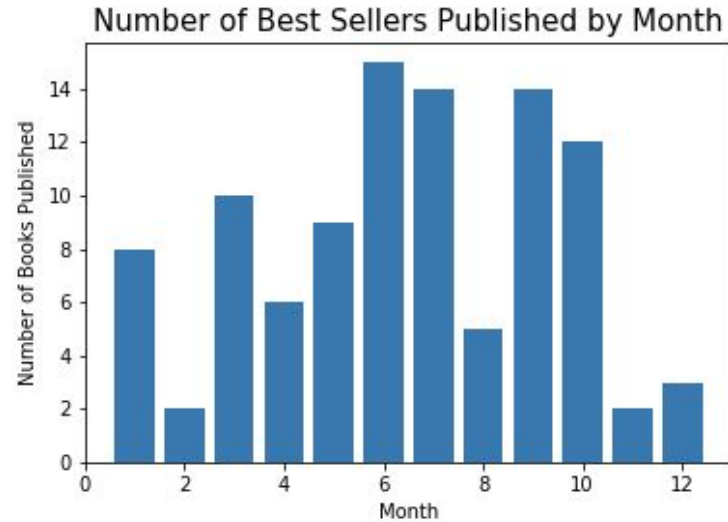HarperCollins Grp 7.7%
Bloomsbury Grp 16.5%

Top Publishers:

- Transworld Group*
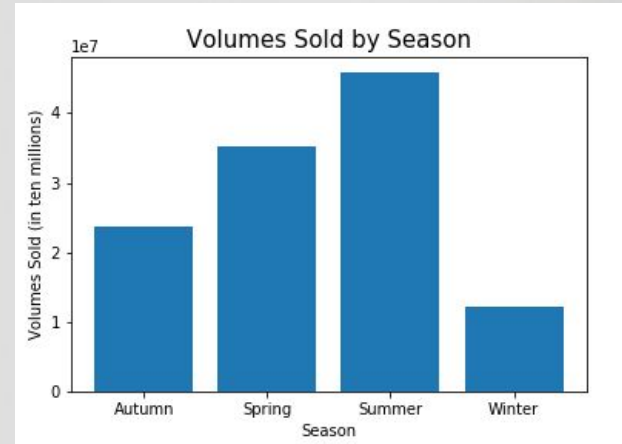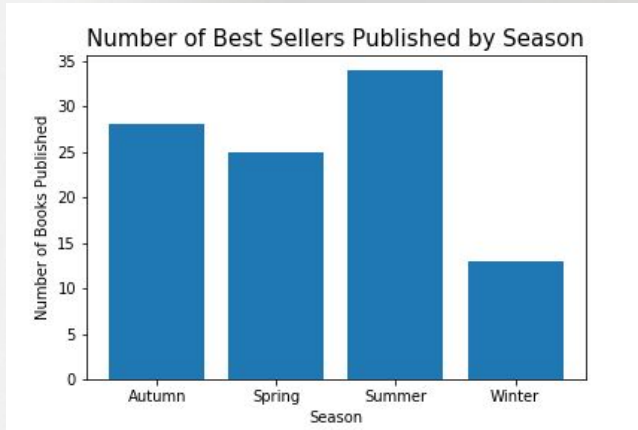- Random House Group
- Penguin Group

Analysis - Sale Price ($)
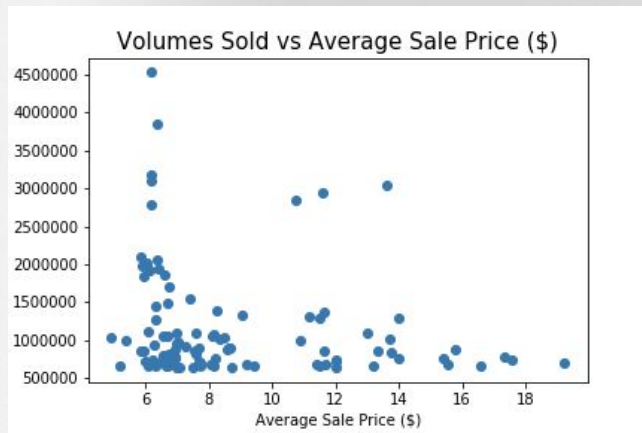
Analysis - Publish Date

Analysis - Publish Date

CURRENT DATA

Cleaning, Exploration, and Analysis

Publishers

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
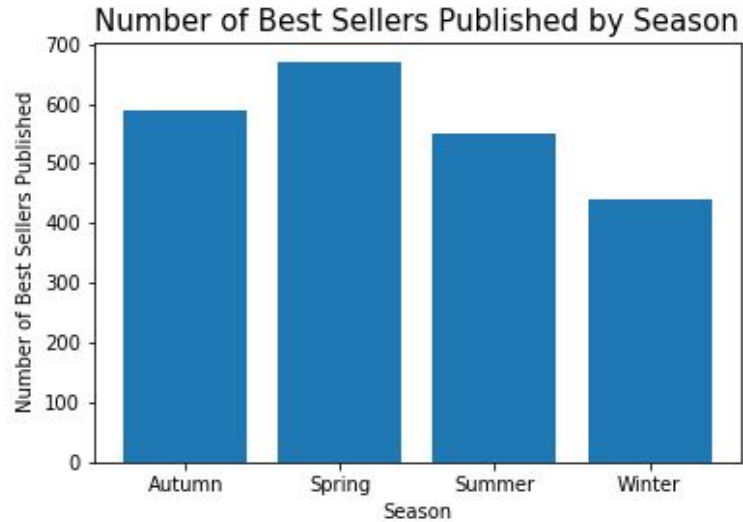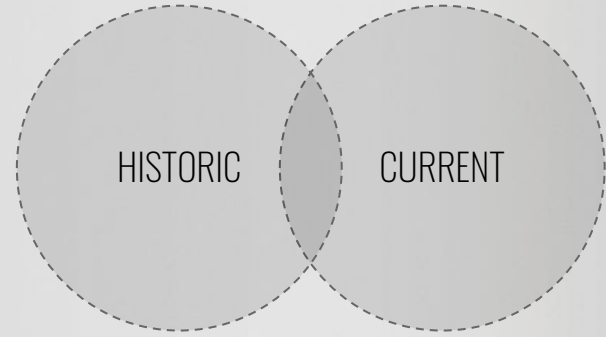
Analysis - Publisher

Top 10 Publishers for NY Times Best Sellers
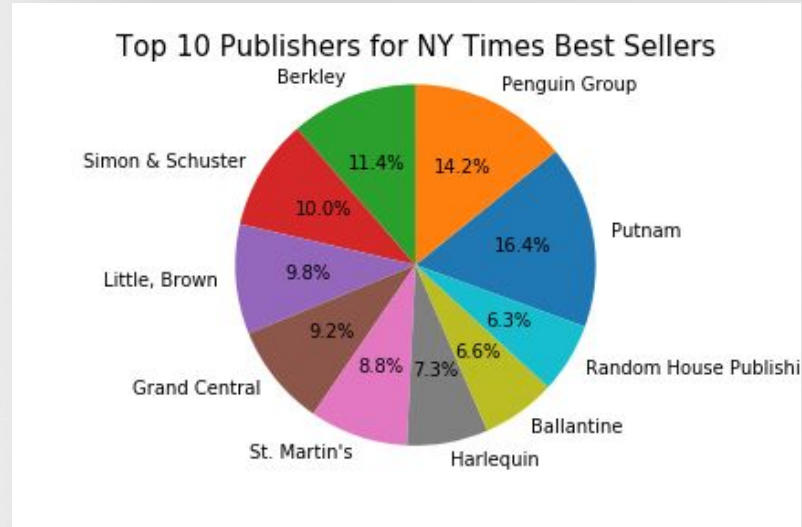
Top Publishers:

- Putnam
- Penguin Group
- Berkley

Analysis - Publish Date

Number of Best-Sellers Published by Month

Analysis - Publish Date

Number of Best Sellers Published by Season
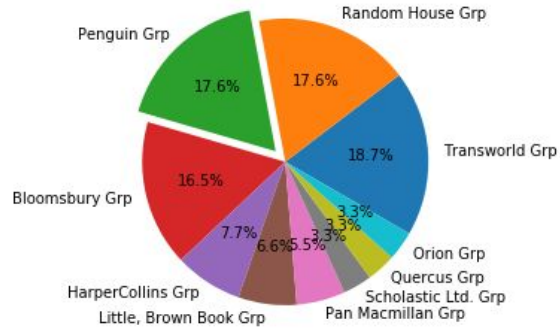
HISTORIC   CURRENT

SIMILARITIES?

Were there similar themes with the data?

Publisher

Top 10 Publishers from 1998 - 2010

- Random House Grp — 17.6%
- Penguin Grp — 17.6%
- Transworld Grp — 18.7%
- Bloomsbury Grp — 16.5%
- Orion Grp — 3.3%
- Quercus Grp — 3.3%
- Scholastic Ltd. Grp — 3.3%
- Pan Macmillan Grp — 5.5%
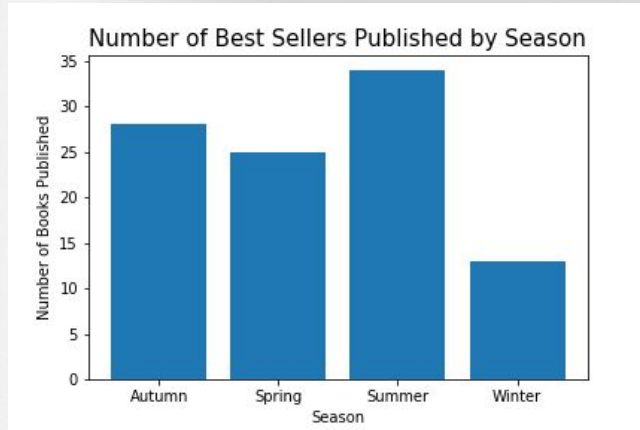- Little, Brown Book Grp — 6.6%
- HarperCollins Grp — 7.7%

Top 10 Publishers for NY Times Best Sellers

- Berkley — 11.4%
- Penguin Group — 14.2%
- Simon & Schuster — 10.0%
- Putnam — 16.4%
- Little, Brown — 9.8%
- Random House Publis — 6.3%
- Grand Central — 9.2%
- Ballantine — 6.6%
- St. Martin's — 8.8%
- Harlequin — 7.3%
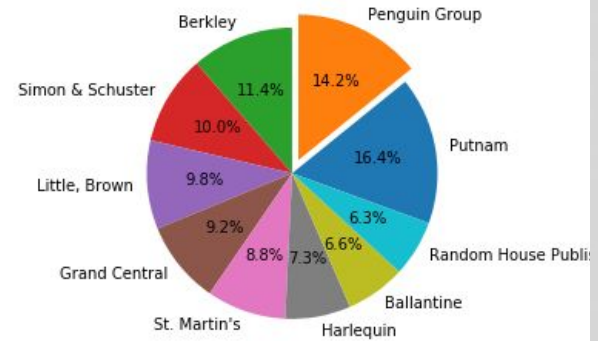
Publish Date
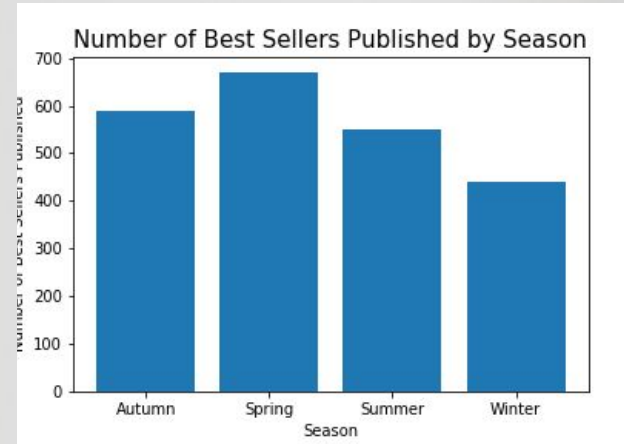
ONE STEP FURTHER

Cleaning, Exploration, and Analysis of Extra Book Info

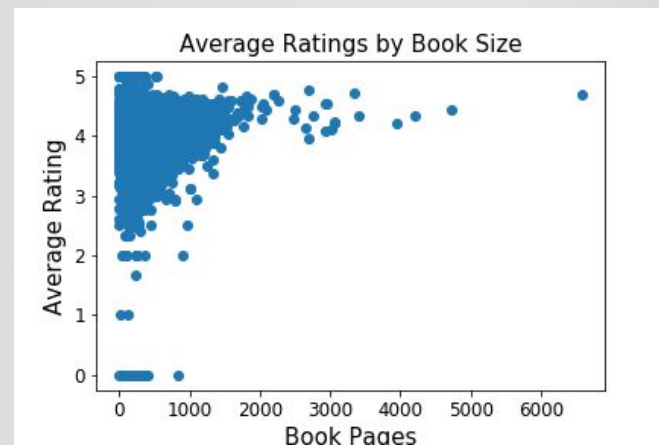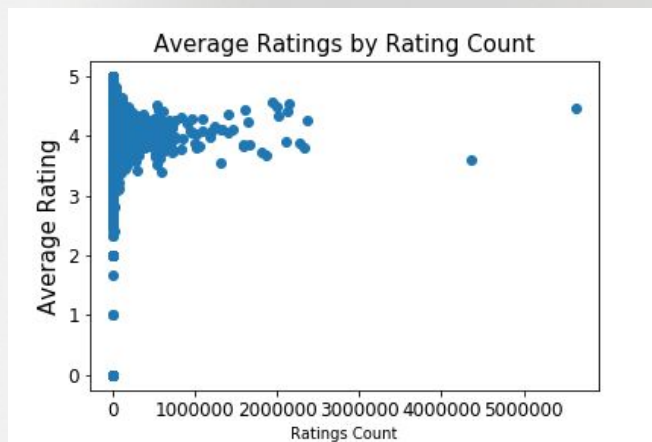# Cleaning and Exploration

AVAILABLE COLUMNS -

BookID, Title, Authors, **Average_rating,** isbn, isbn13, Language_code, **# num_pages**, **Ratings_count,** Text_reviews_count
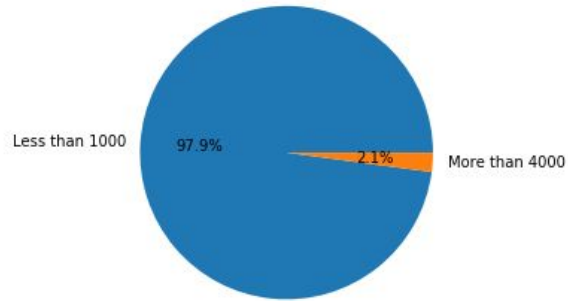
EXTRA TOPICS TO EXPLORE -

- ❖ Average rating
- ❖ Length of Book
- ❖ Number of Rating
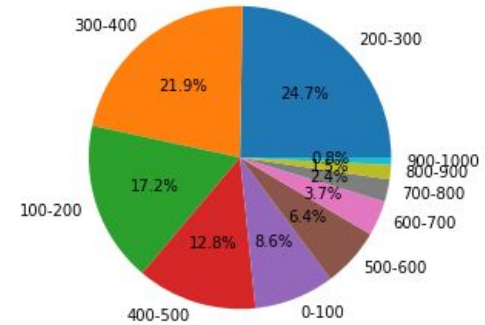
Average Rating - 3.93 out of 5

Length of Book

Percentage of Book Pages Greater or Less Than 1000

Less than 1000 — 97.9%
More than 4000 — 2.1%

Percentage of Book Pages from 0-1000

300-400 — 21.9%
200-300 — 24.7%
100-200 — 17.2%
400-500 — 12.8%
0-100 — 8.6%
500-600 — 6.4%
600-700 — 3.7%
700-800 — 2.4%
800-900 — 0.5%
900-1000 — 0.8%

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 - **$10-15**
- ❖ 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!