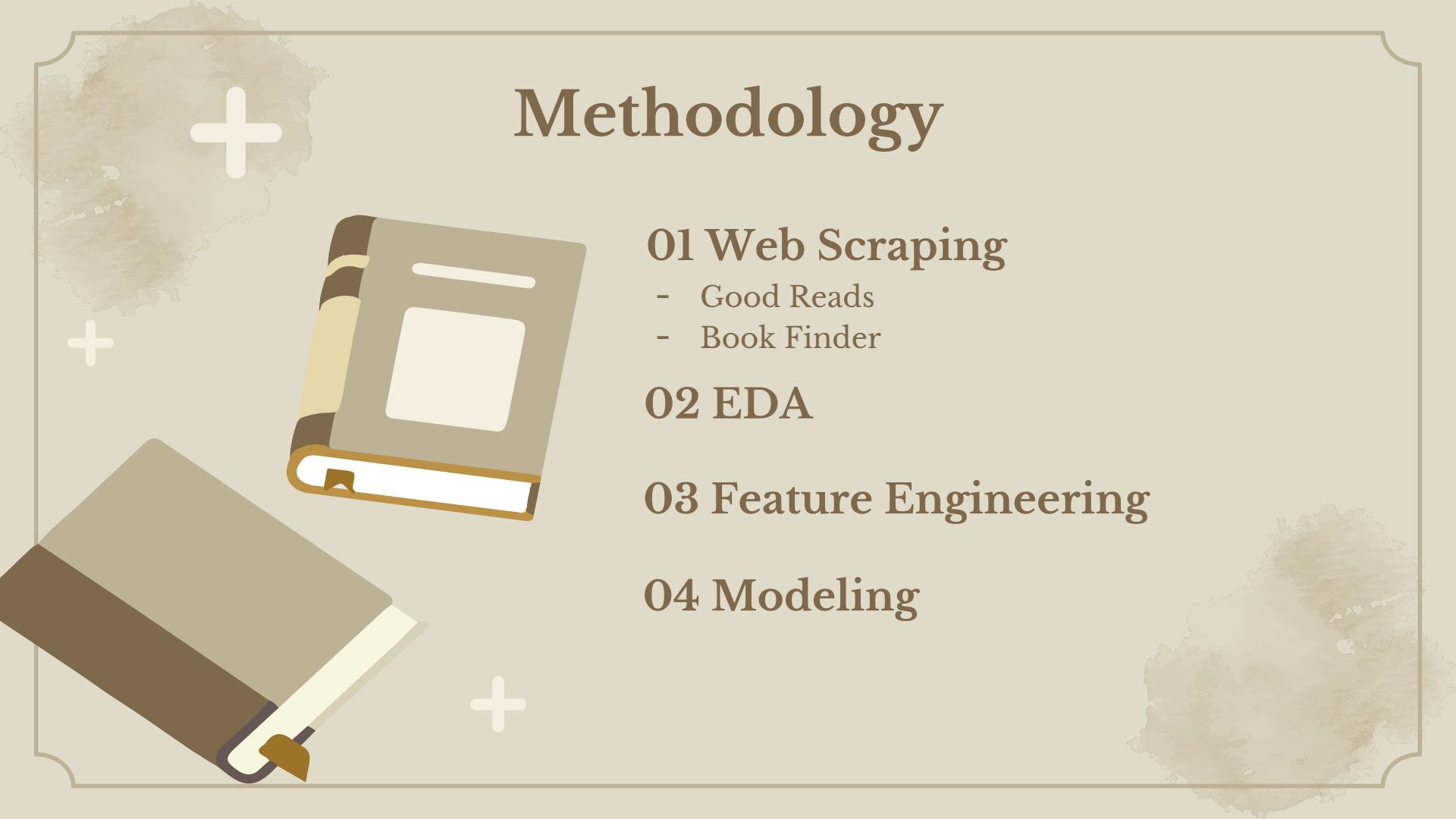


GoodReads Scraping for Book Price Prediction

Prepared by :
Sara Hawi
Taef al Khales



Methodology

01 Web Scraping

- Good Reads
- Book Finder

02 EDA

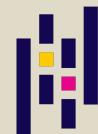
03 Feature Engineering

04 Modeling



Tools

Beautifulsoup

 **pandas** **matplotlib**
NumPy
seaborn

01

Web Scraping



goodreads

- Each Year
- Each Genre
- Each Book

goodreads Home My Books Browse Community Search books

All nominees • 50,078 VOTES TOTAL

FANTASY
VIEW RESULTS →

ROMANCE
VIEW RESULTS →

SCIENCE FICTION
VIEW RESULTS →

HORROR
VIEW RESULTS →

HUMOR
VIEW RESULTS →

NONFICTION
VIEW RESULTS →

MEMOIR & AUTOBIOGRAPHY
VIEW RESULTS →

HISTORY & BIOGRAPHY
VIEW RESULTS →

SCIENCE & TECHNOLOGY
VIEW RESULTS →

FOOD & COOKBOOKS
VIEW RESULTS →

GRAPHIC NOVELS & COMICS
VIEW RESULTS →

POETRY
VIEW RESULTS →

DEBUT NOVEL
VIEW RESULTS →

YOUNG ADULT FICTION

72,828 votes Matt Haig *The Midnight Library*
Want to Read Rate this book ★★★★☆

72,823 votes Fredrik Backman *Anxious People*
Want to Read Rate this book ★★★★☆

67,226 votes Jeanine Cummins *American Dirt*
Want to Read Rate this book ★★★★☆

59,356 votes Kiley Reid *Such a Fun Age*
Want to Read Rate this book ★★★★☆

48,664 votes Kate Elizabeth Russell *My Dark Vanessa*
Want to Read Rate this book ★★★★☆

35,614 votes Emily St. John Mandel *The Glass Hotel*
Want to Read Rate this book ★★★★☆

31,699 votes Yaa Gyasi *Transcendent Kingdom*
Want to Read Rate this book ★★★★☆

28,640 votes Abi Dare *The Girl with the Louding Voice*
Want to Read Rate this book ★★★★☆

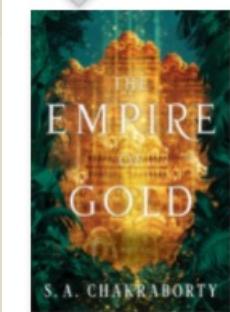
25,384 votes Ann Napolitano *Dear Edward*
Want to Read Rate this book ★★★★☆

22,008 votes Jennifer Weiner *Big Summer*
Want to Read Rate this book ★★★★☆

7,918 votes Lily King *Writers & Lovers*
Want to Read Rate this book ★★★★☆

5,607 votes Frances Chai *If I Had Your Face*
Want to Read Rate this book ★★★★☆

goodreads



Want to Read

Rate this book



The Empire of Gold

(The Daevabad Trilogy #3)



Rating details



98% of people liked it

All editions: 4.51 average rating, 23346 ratings, 3690 reviews, added by 81863 people, 49190 total reads

This edition: 4.49 average rating, 9958 ratings, 1638 reviews, added by 44085 people

Published June 30th 2020 by Harper Voyager (first published June 11th 2020)

More Details...

Edit Details

Audiobook

Fantasy > Magic

Share



[Recommend It](#) | [Stats](#) | [Recent Status Updates](#)



See similar books...

1,948 users

277 users

167 users

146 users

94 users

92 users

goodreads

Hardcover, US Edition, 766 pages

Published June 30th 2020 by Harper Voyager (first published June 11th 2020)

Original Title

The Empire of Gold

ISBN

0062678167 (ISBN13: 9780062678164)

Edition Language

English

URL

<https://www.harpercollins.com/9780062678164/the-empire-of-gold/>

Series

The Daevabad Trilogy #3

Characters

Alizayd al Qahtani, Darayavahoush e-Afshin, Nahri, Manizheh e-Nahid

Setting

Daevabad

Literary Awards

Goodreads Choice Award Nominee for Fantasy (2020)

Other Editions (28)



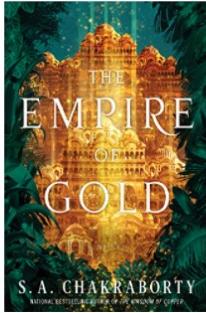
[All Editions](#) | [Add a New Edition](#) | [Combine](#)

Book Finder

Search Results

Search Author, Title or ISBN

Go



ISBN is **9780062678164** | **0062678167**

[The Empire of Gold: A Novel \(The Daevabad Trilogy, 3\)](#)

by [Chakraborty, S. A](#)

4.51 avg rating • 23,112 ratings by GoodReads

Publisher: Harper Voyager, 2020

Edition: Hardcover

Language: English

Prices INCLUDE standard shipping to Saudi Arabia

[Show prices without shipping](#)

[Change shipping destination/currency](#)

Shipping prices may be approximate. Please verify cost before checkout.

About the book:

No series since George R.R. Martin's Game of Thrones has captured the imagination of readers like The Daevabad Trilogy. In the first two books, we were introduced to the world of Daevabad, a city where the rules of society are based on intricate politics and magic. In the final chapter of the bestselling, critically acclaimed trilogy, we learn what happened to Dara, the con-woman and an idealistic djinn princess who fell in love with the fallen king of Daevabad. After a brutal conquest stripped the city of its ruler, Dara must try to rebuild her kingdom and restore the balance of power.

The final chapter in the bestselling, critically acclaimed trilogy, The Empire of Gold, follows Dara as she tries to rebuild her kingdom and restore the balance of power. After a brutal conquest stripped the city of its ruler, Dara must try to rebuild her kingdom and restore the balance of power.

Daevas have fallen.

After a brutal conquest stripped the city of its ruler, Dara must try to rebuild her kingdom and restore the balance of power. After a brutal conquest stripped the city of its ruler, Dara must try to rebuild her kingdom and restore the balance of power.

But the bloodletting and loss of life has taken its toll. Show more

Click on the price to find out more about a book.

New books: 1 - 50 of 57

#	Bookseller	Notes	Price
1.	amazon.com United States	Hardcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 Hardcover. Edition: Illustrated. 784 pages. Usually ships within 2 to 3 days. Fulfillment by Amazon.	\$31.85
2.	GreatBookPrices via United States	Hardcover, ISBN 9780062678164 Publisher: Harper Voyager, 2020 New. Hard cover. Sewn binding. Cloth over boards. 784 p. Contains: Illustrations. Daevabad Trilogy, 3. In Stock. 100% Money Back Guarantee. Brand New, Perfect Condition, allow 4-14 business days for standard shipping. To Alaska, Hawaii, U.S. protectorate, P.O. box, and APO/FPO addresses allow 4-28 business days for Standard shipping. No expedited shipping. All orders placed with expedited shipping will be cancelled. Over 3, 000, 000 happy customers.	\$35.49
3.	GreatBookPrices #2	Hardcover, ISBN 9780062678164	\$25.40

Used books: 1 - 43 of 43

#	Bookseller	Notes	Price
1.	BIBLIO .com United Kingdom	Softcover, ISBN 9780062678164 via Publisher: Harper Voyager Used - Like New: Paper	
2.	BIBLIO .co.uk United Kingdom	Softcover, ISBN 9780062678164 via Publisher: Harper Voyager Used - Like New: Paper	
3.	Seattle Goodwill via	Hardcover, ISBN 9780062678164 via Publisher: Harper Voyager Used - Good: Hardcov	

02

Data Cleaning & EDA

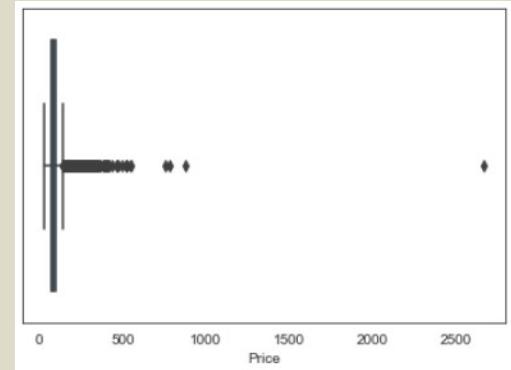


Dataset

	Title	Author	ISBN	Year	Pages	No of Rating	Avg Rating	Added by	Liked by	Reviews	Votes	Genre	Year of nomination	Price
0	And the Mountains Echoed	Khaled Hosseini	159463176X	2013	404	331300	4.05	655406	0.95	29506	24686	Fiction	2013	SAR 48.93
1	The Storyteller	Jodi Picoult	1439102767	2013	461	188749	4.25	358108	0.96	17458	21269	Fiction	2013	SAR 241.336206896552
2	Me Before You	Jojo Moyes	0670026603	2012	369	1300820	4.25	2124123	0.95	86168	19656	Fiction	2013	SAR 168.4125
3	The Goldfinch	Donna Tartt	0316055433	2013	881	791855	3.92	1534752	0.89	65185	9313	Fiction	2013	SAR 83.55
4	The Husband's Secret	Liane Moriarty	0399159347	2013	396	578835	3.95	917935	0.94	37655	8715	Fiction	2013	SAR 78.675

Data Cleaning

- Renaming columns
- Removing nulls (Mean)
- Removing the currency (SAR)
- Removing Duplicates
(Title + Author + Genre + Year of Nomination)
- Removing outliers

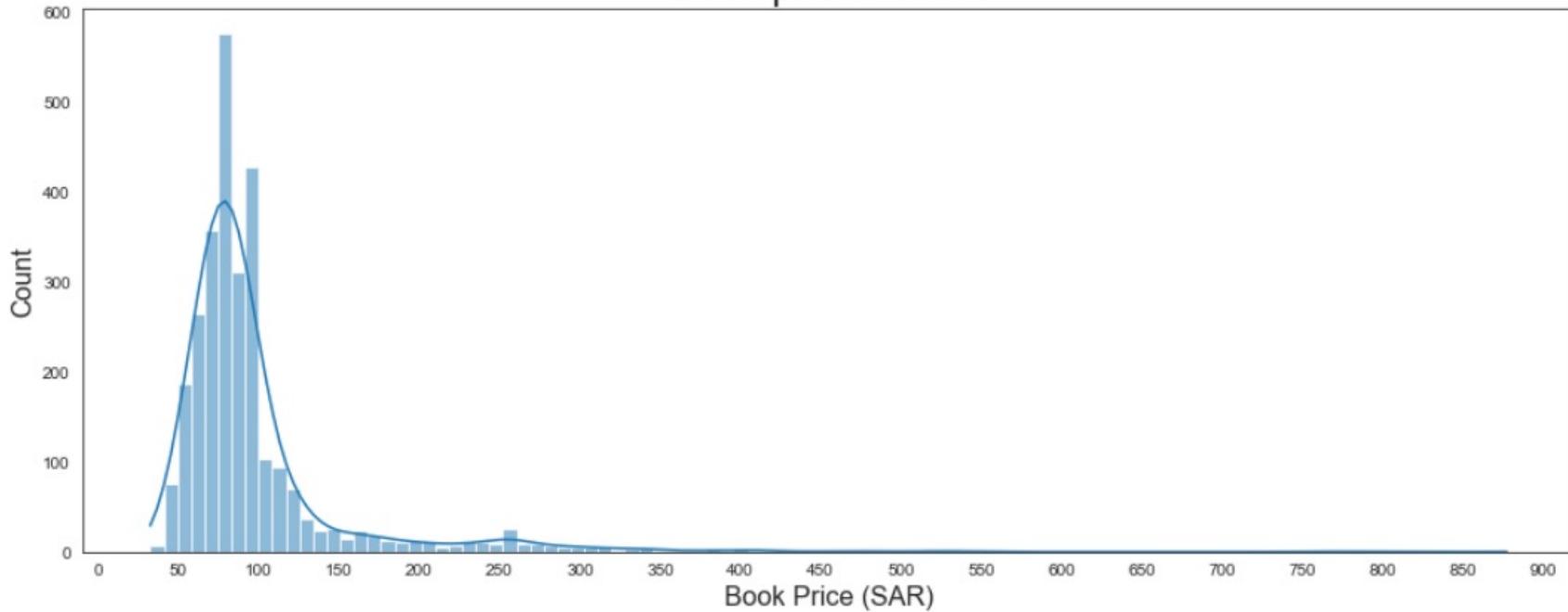


EDA plots



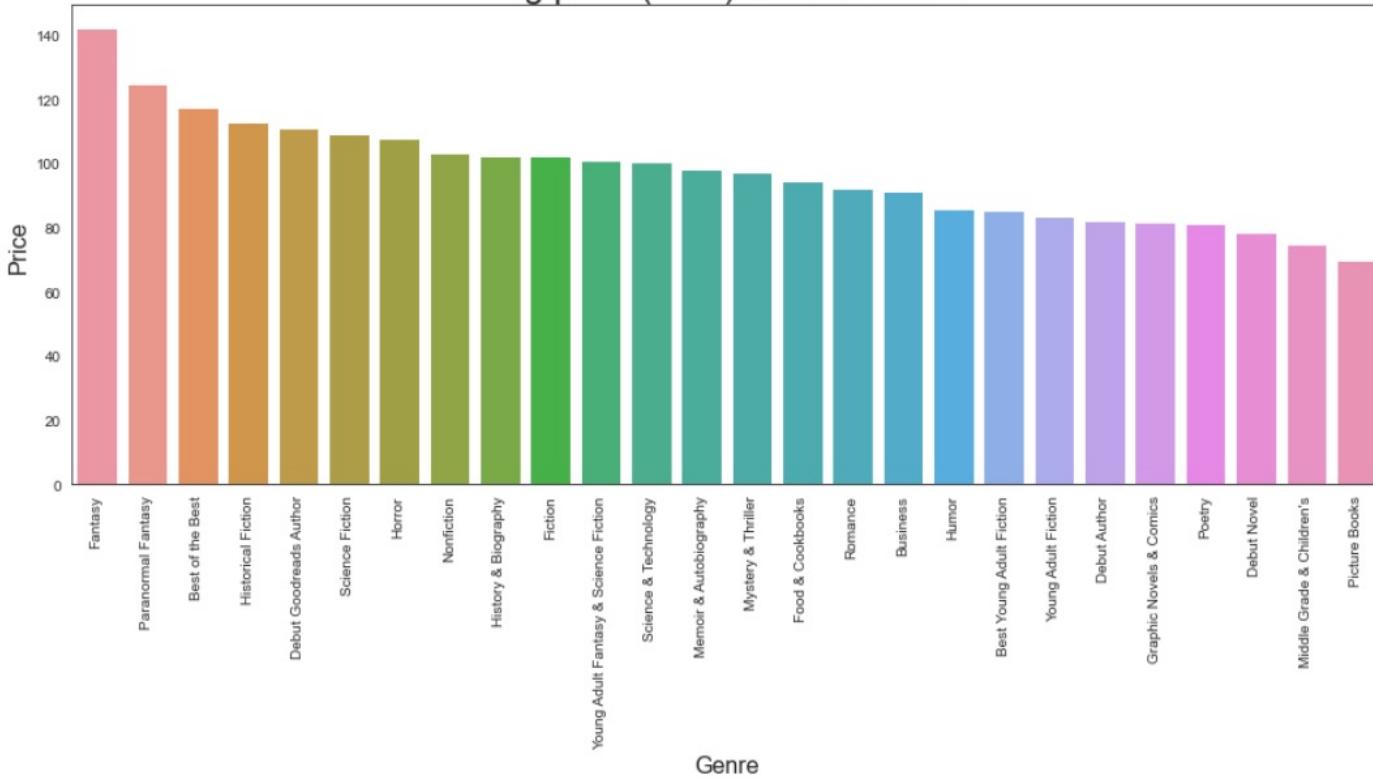
EDA plots

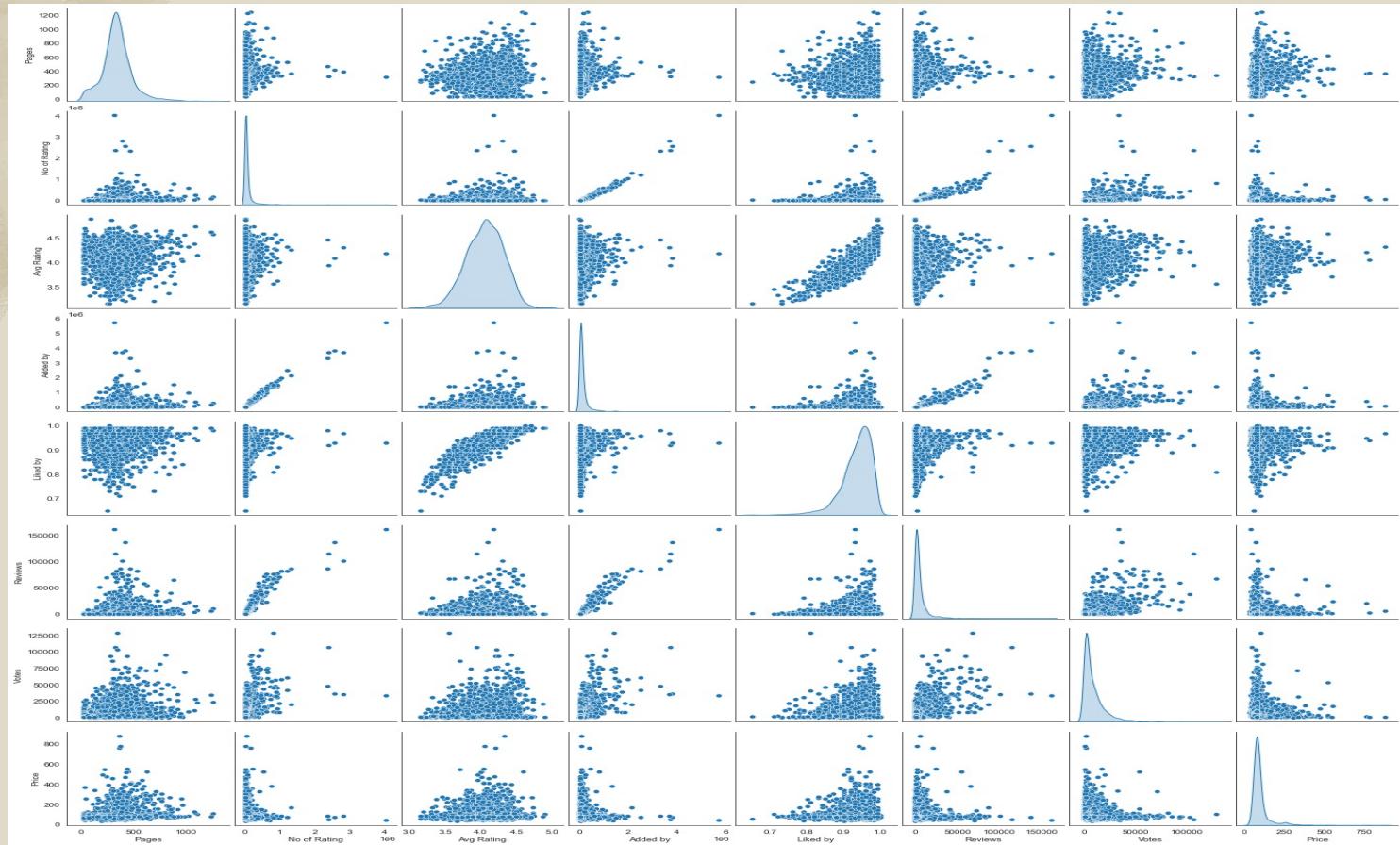
Book prices count

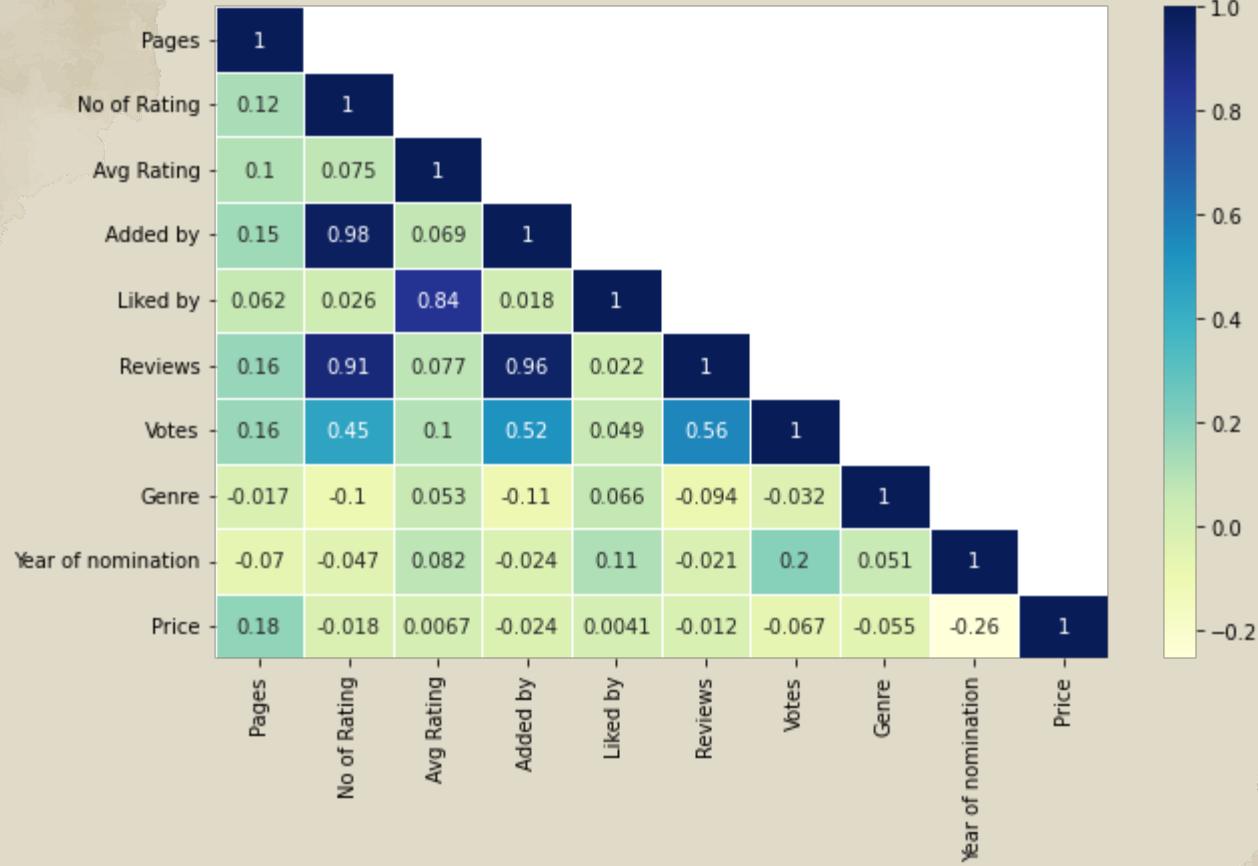


EDA plots

Avg price (SAR) of each Genre







03

Feature Engineering



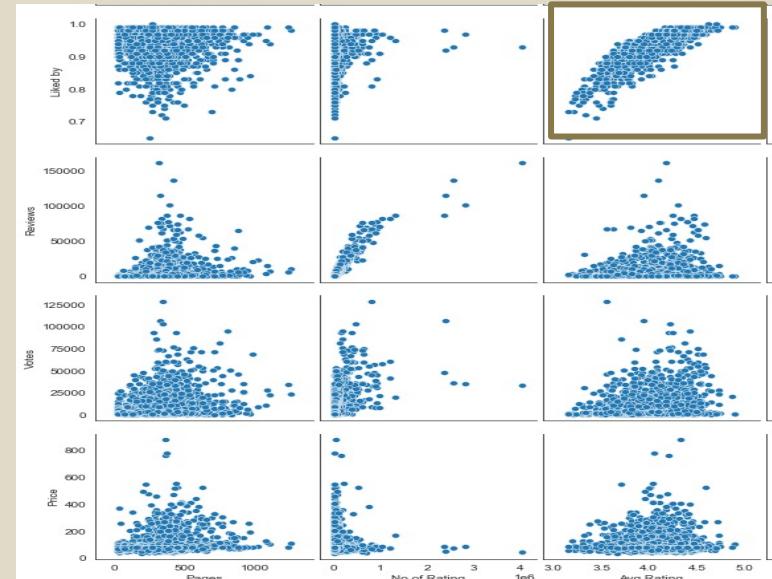
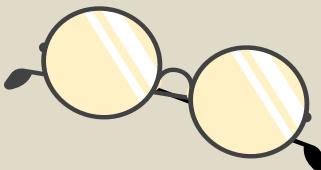
Baseline LinearResgression() model

R ² Score	MAE
0.048271	34.89

Dealing with Collinearity

- Adding Polynomial Features
- Adding Interaction Terms

MAE -2
R² Score -0.01



04

Modeling & Regularization



Regularization

Determining the best regularization hyperparameter by 5-Fold CV on Training set:

	Hyperparameter	R ² Score	MAE (SAR)
LASSO	$\alpha = 10$	0.03934	32.76118
Ridge	$\alpha = 1000$	0.03952	32.74362
ElasticNet	$\alpha = 10$	0.03969	32.75029
Polynomial Regression	Degree = 2	-0.043693	33.88504

- Scoring on Test Set for ElasticNet:
- **MAE = 34.85432 SAR**

Suggestions & Conclusion

1. Data obtained via scraping websites are not reliable enough to validate a model's performance.
2. Data might be too complex and high in dimensionality for linear regression.
3. Classification might be the better choice after labeling intervals of price (i.e., Expensive, Moderate, Affordable)



Thank you for Listening!

Feel free to ask questions

... zrak, —
ným ústaní
aviciu, skoncom boj
zdrapil za hrdo:
je už neskoro, lebo je
to vieš, bude to voja ro-
v; bol by som osval ... nebol
ti urobim čost, ja ... poznám
Há, máš klenoty, zlatú retáz
ešte, ešte ... Na ... Maš už dosť?
nohy, ktorý ho ešte stále exekútor, zvaliac sa
zni, lhočasiac, či skôr býjiceho, obstá-
loch, zápasiacich, či skôr býjiceho, a bítého;
Boulard, zadýchčaný, vylakaný, naskreže sa
ktorým ho zasypal sa uhniť, ako sa len dalo, ide-
šťastie, dozoreca konajúci služobník.
Pan Boulard vstal, bledý, vystrašený, jedno oko
opuchnuté, a nedabajúc ani, že mu spadla capica, kričal,
Srážka ... otvorie mi ... nechcem tu už ostat ani
minútku dlhšie ... Pomoc ...

priek ne-
mnoho, m-
ka... Niž co, po-
povedal Nicolas.
A teraz bysro, bys-
ta Cibulka a Germain.
ubája, hrdo mu žiaril
vane, i spokojne sa
nicou mi už zmizal
Cibulka ho sledoval
ked už dva-tri
akosi sa premie-
zdala obz-
už memohli
tikol leb-
svedčil
pripo-
a t-

ký ho očaril svojím povolením, ktorý ho predtým, keď na řom pristál pri návrate z Arménie, už predtým, keď na řom pristál pri návrate z Arménie. Tu žil v skromnom domčeku, nie ovela priestrannejšom než letná vilka, ako jednoduchý občan. Občas sa prechádzal po gymnáziu bez liktora a bez služobníka a udržiaval styky s jednoduchými Gréckmi ako so severovými. Jedného dňa ráno, keď zadeľoval práce, povedal len tak mimochodom, že by chcel navštíviť všetkých chorých v mestečku. Ti, čo boli v jeho blízkosti, tomu rozumeli inakšie a vydali príkaz priniesť všetkých chorých do verejnej kolónady a tam ich roztriediť podľa chorob. Zarazilo ho to neočakávané divadlo a dielo bol na rozpakoch, čo robí. Nakoniec však obšiel všetkých chorých a každému sa ospravedlil za to, čo sa stal, aj keď boli medzi nimi Judia nízkeho postavenia a jemu úplne neznámi. Svoju tribúnsku moc použil, ako sa zdalo, len v jednom prípade — o druhých sa aspoň nehovorí. Často navštěvoval školy a posluchárne učencov. Pri jednej takej návštěve strhla sa raz medzi antisofistami prudká hádka; Tiberius sa do nej zamiliesť a tu ho ktorí napadol a vycítil mu, že nadíza druhej strane. Tiberius sa nenápadne vyratil, o čivuľu sa znova zjavil so slúžobníkmi a hlasateľom, dal vyskakať človeka, ktorý ho urazil, pred svoj tribunál a dal ho uväzniť. Potom sa dozvedel, že jeho manželka Iulia bola pre

„aby si Tahko opustil“.
12. Zostal teda proti svojej vôle na Rode a len za pomoci matky horko-tažko dosiahol, že ho Augustus meral svojím legátom na Rode, čo malo odovzdovať jeho nepričomnosť v Rime a zakryť jeho hanbu. Od tých čias nežil už len ako jednoduchý občan, ale aj ako bojazlivý vinnik; skrýval sa na vidieku vo vnútrozemi a výhýbou sa stretňoval s mužmi, ktorí sa plavili okolo ostrova a často ho navštěvovali, pretože ani jeden úradník vincič bez toho, že by sa nezastavil na Rode, nepristípil aj iné dôvody, ktoré ho ešte väčšinou nevali. Ked totiž prišiel na Samos a navštívil tam neľastného syna Gaia, ktorý bol správcom Venuše, videl, že sa mu Gaius veľmi odcudzil, a to hlavne z hou svojho sprivedcu a vychovávateľa Marcu Lohu, ktorý ho pred ním odiernil. Okrem toho upadol do pozornenia, že posil po niektorých centúriach, ktorých sám povýšil a ktorí sa po dovolenke vrácali naspäť do tábora, viacerým osobám záhadné odkazy, ktoré, ako sa zdalo, mali každého z nich pripraviť na prevrat. Ked mu Au-