ReadMe

Book Price Prediction

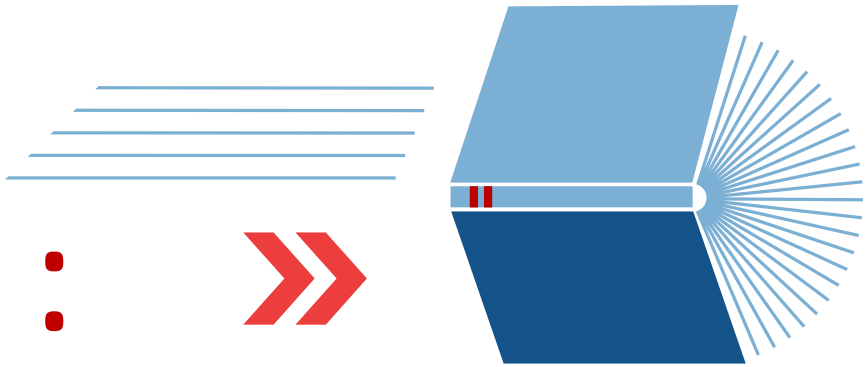
By: Batoul Alosaimi ,Norah Alqahtani and Shroaq Almutairi

ReadME

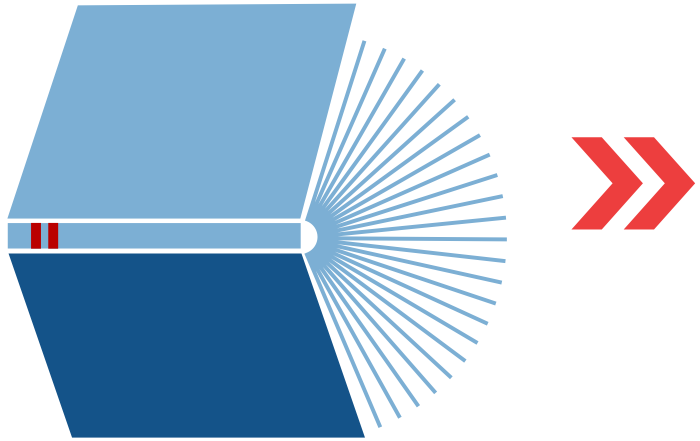
ReadMe which is an online web site that is selling books.

Methodology

Data Source :



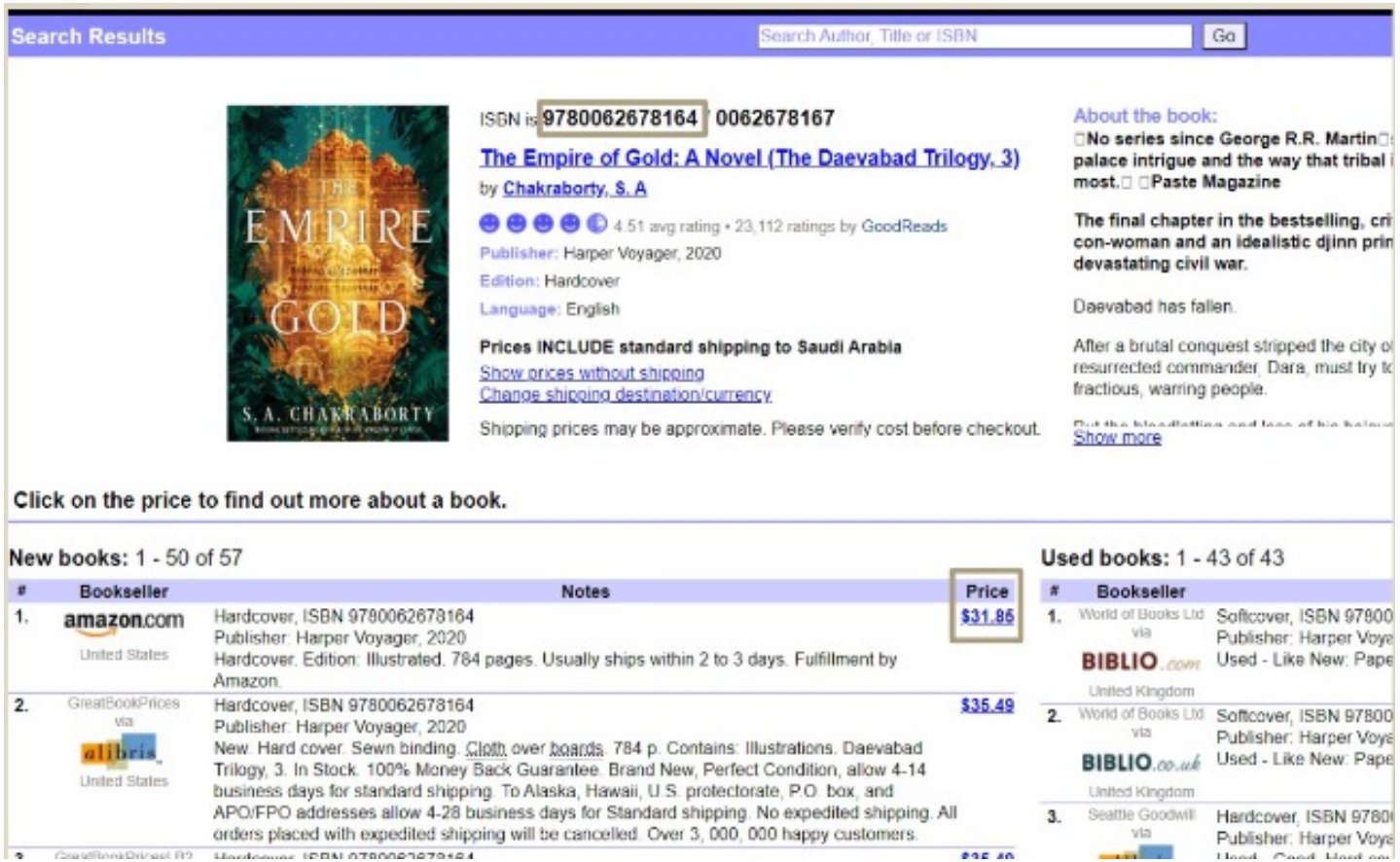
1-kaggle(CSV file)
GoodReads dataset ,
11128 books with 11 features .



READ ME

Book Finder :

Web Scraping
(Selenium)
1352 books
with 1 feature.

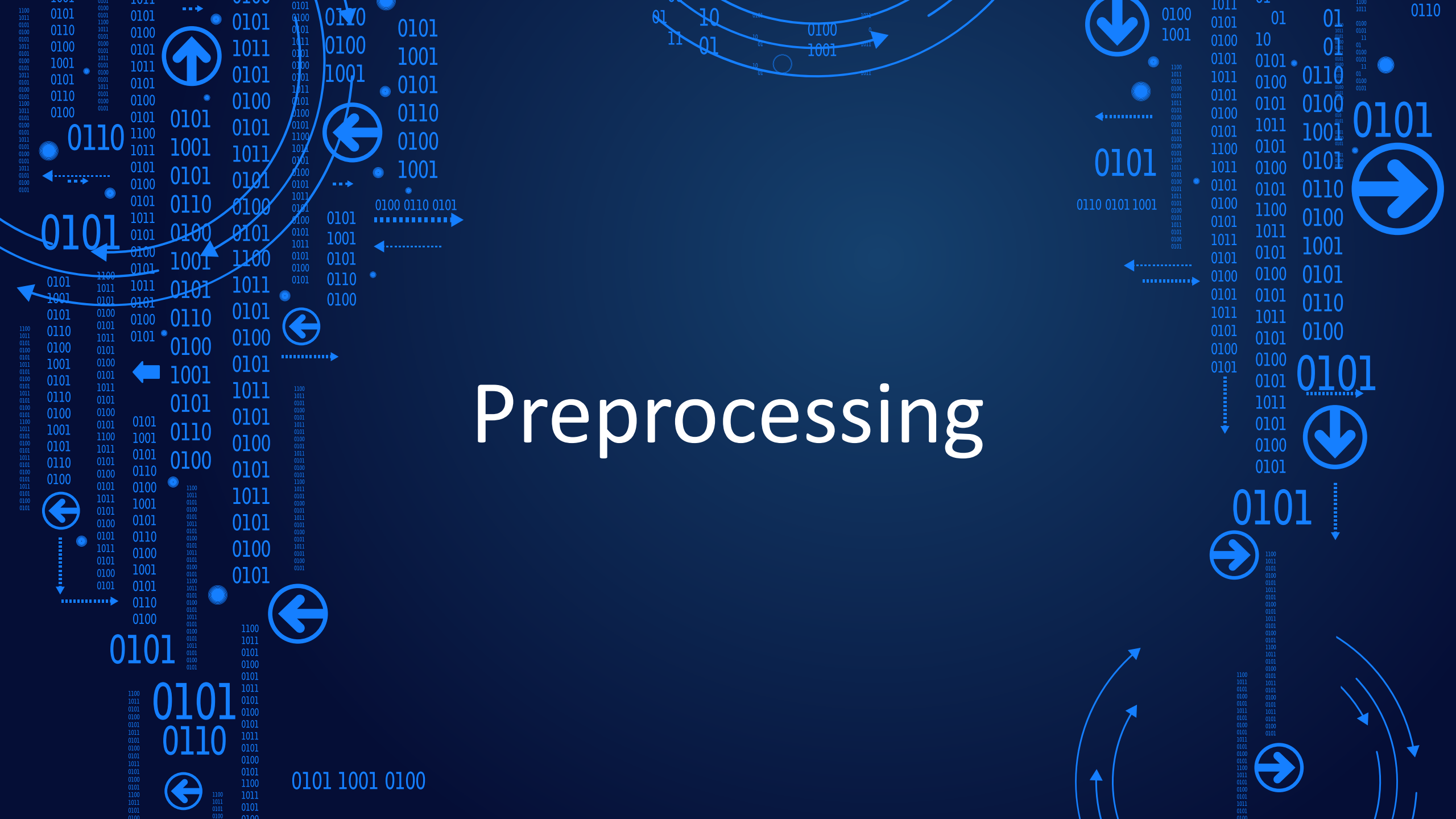


READ ME

Book dataset :

- 1298 books 12 features.

bookID	title	authors	average_rating	isbn	isbn13	language_code	num_pages	ratings_count	text_reviews_count	publication_date	publisher	Price
663	For the New Intellectual: The Philosophy of Ay...	Ayn Rand	3.68	451163087	9.780450e+12	eng	224	2750	108	12/1/1963	Signet Book	SAR 48.675
2326	Dirk Gently's Holistic Detective Agency (Dirk ...	Douglas Adams	3.98	1597770078	9.781600e+12	eng	6	58	15	9/30/2005	Phoenix Audio	Not found
3869	A Brief History of Time	Stephen Hawking	4.17	553380168	9.780550e+12	eng	212	239652	5860	9/1/1998	Bantam Books	SAR 48.675
2045	Later Novels and Other Writings: The Lady in t...	Raymond Chandler/Frank MacShane	4.47	1883011086	9.781880e+12	eng	1076	1107	48	10/1/1995	Library of America	SAR 173.732876712329


Preprocessing

Pre processing:

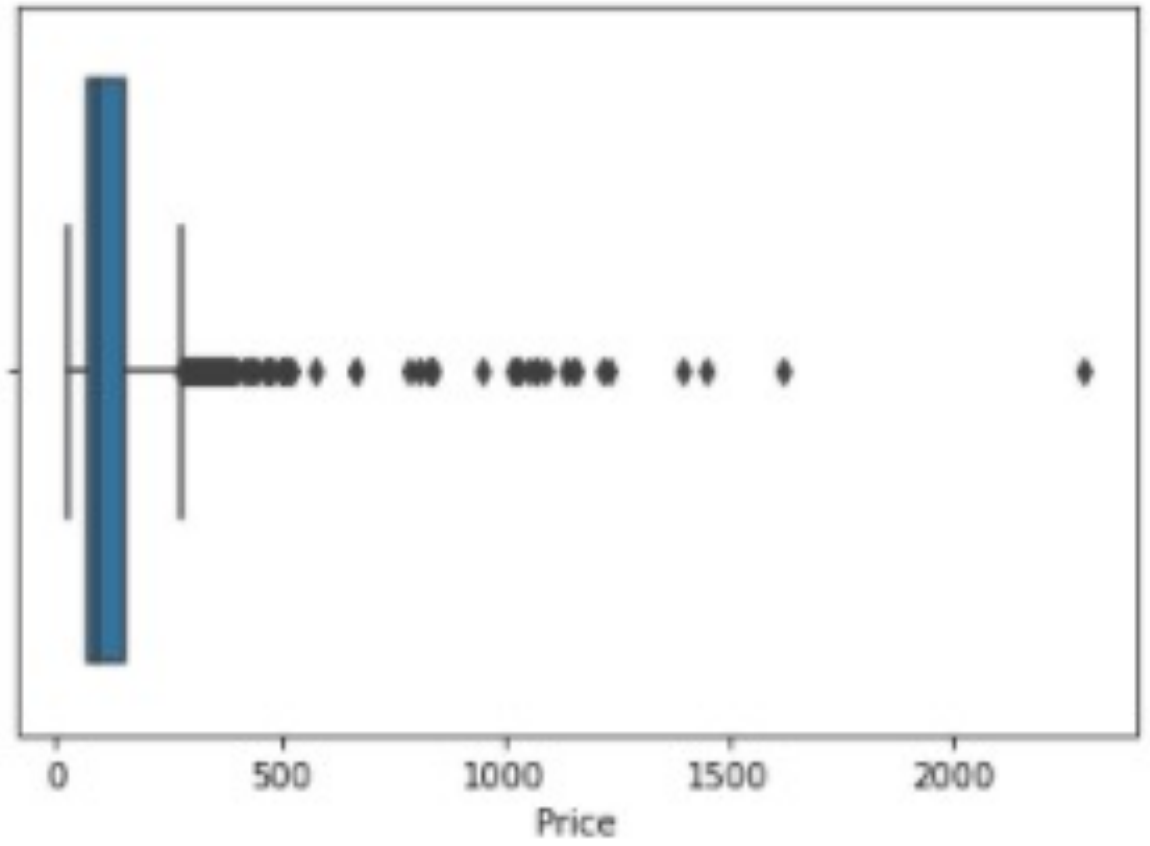
Removing nulls value.

Removing Duplicates.

Clean Price column

Pre processing:

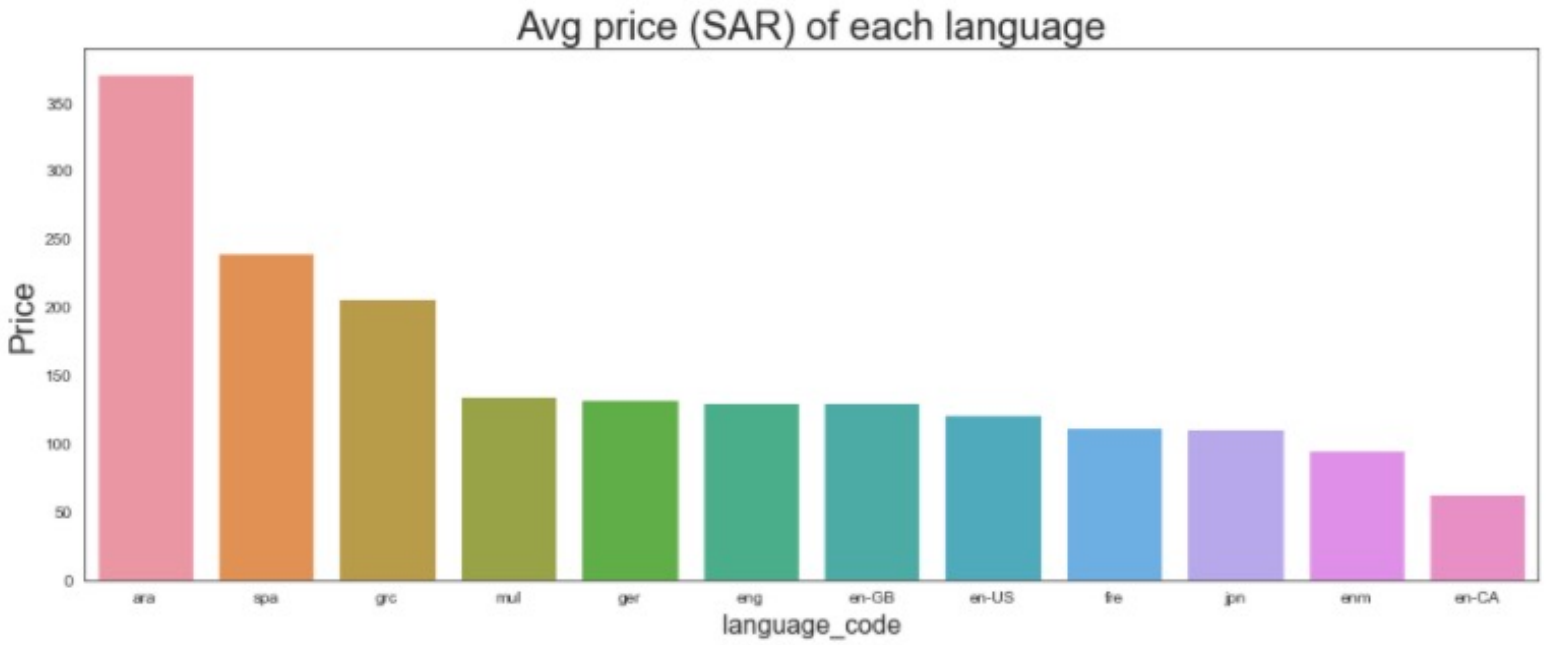
Removing outlier in the price



Exploratory Data Analysis

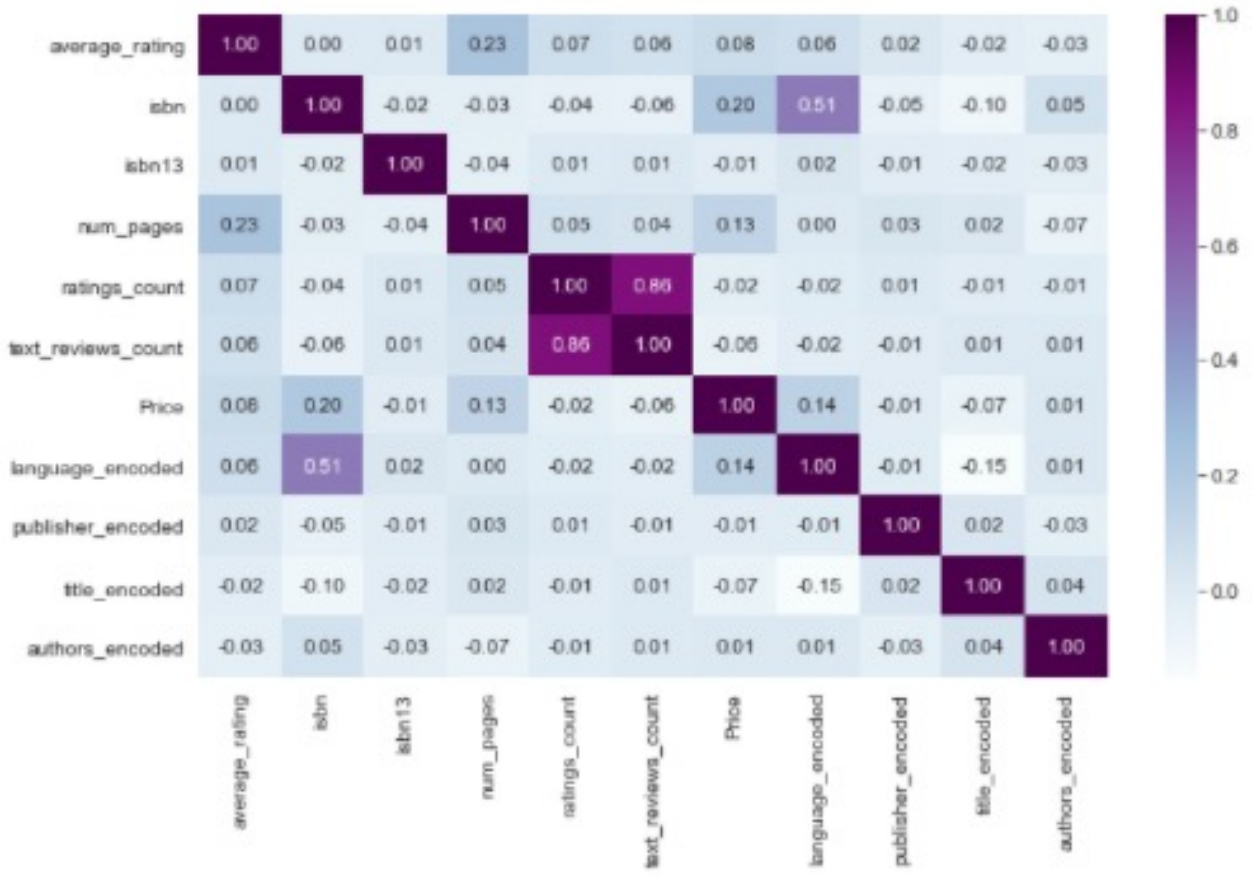
EDA:

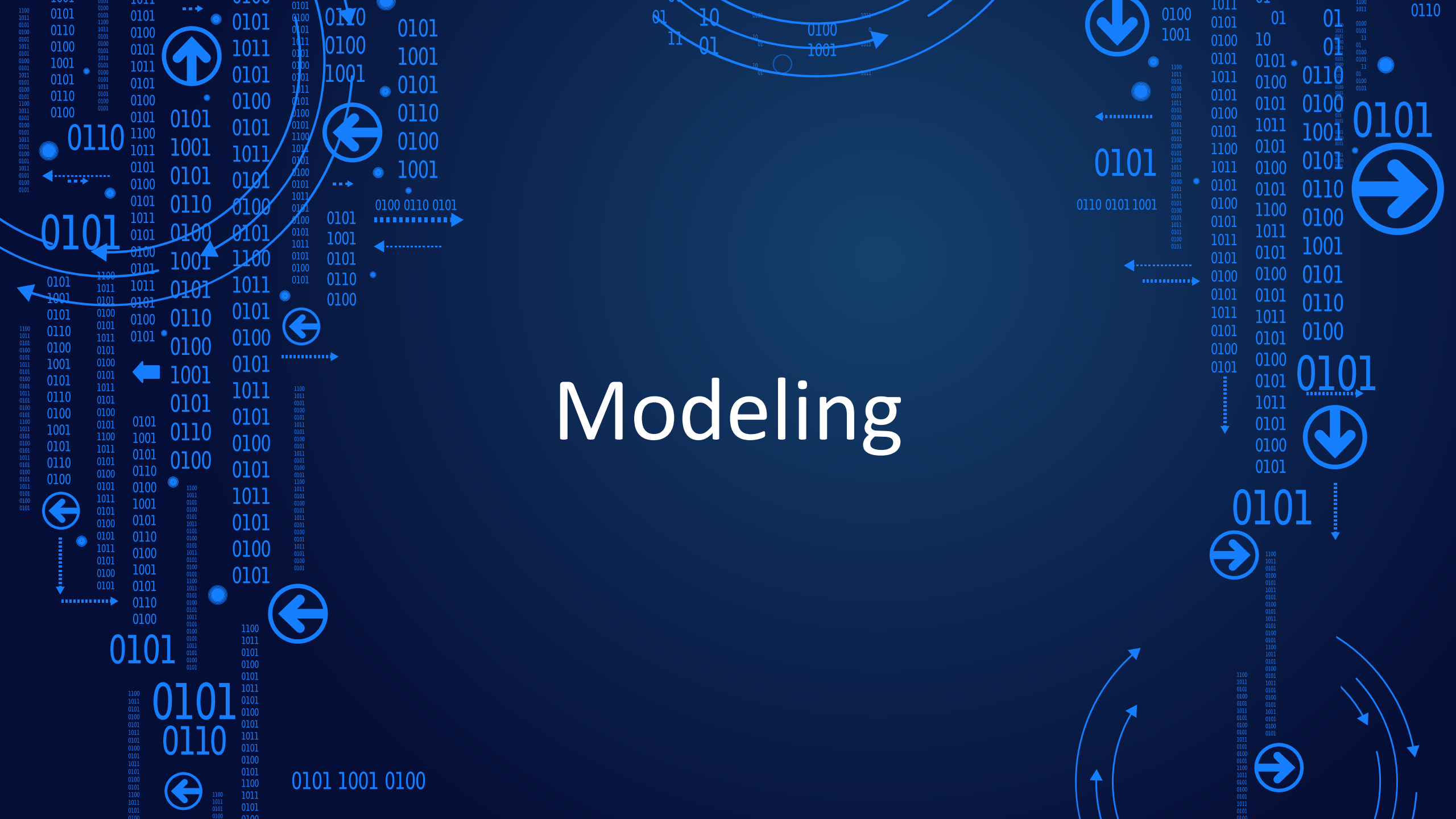
Avg price (SAR) of each language



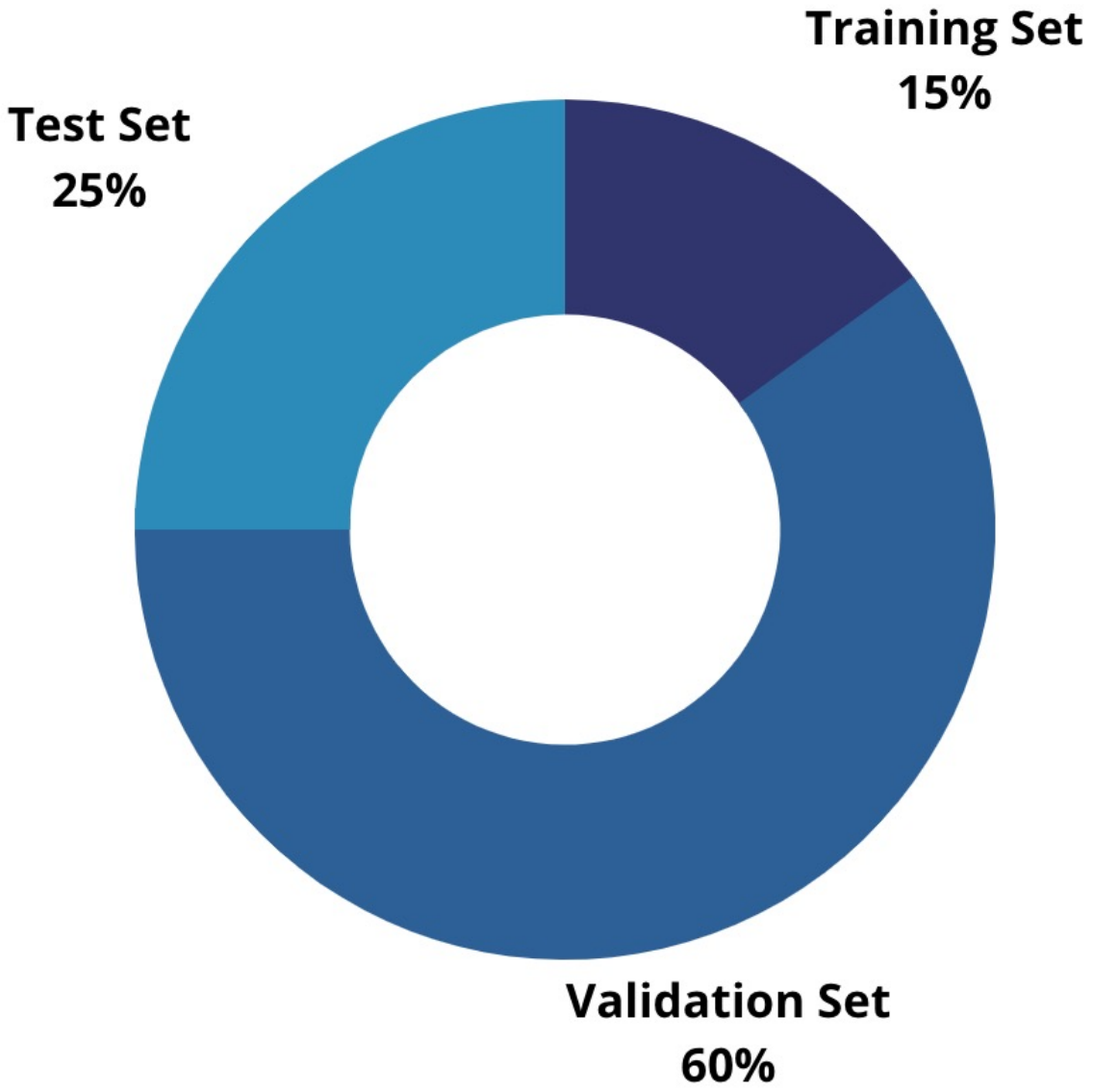
EDA:

correlation
between feature:




Modeling

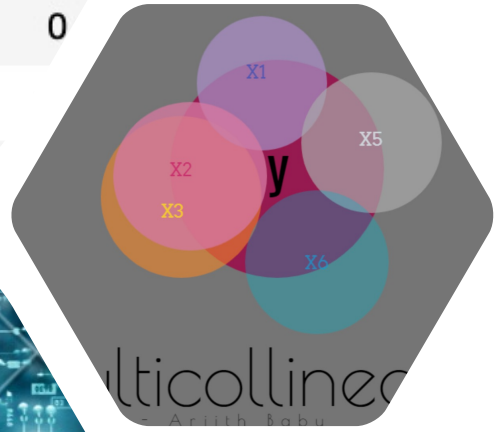
Splitting the data to Test (25%) and Train-Validation (75%) sets



Modeling :


Predicted VS. Actual Target Train
Predicted Target Train (Books price)
Actual Target Train (Books price)
300
250
200
150
100
50
0
−50
0
200
400
600
800

bookID	en-CA	en-GB	en-US	eng	enm	fre	ger	grc	jpn	mul	spa
1	0	0	0	1	0	0	0	0	0	0	
2	0	0	0	1	0	0	0	0	0		
5	0	0	0	1	0	0	0	0	0		
8	0	0	0	1	0	0	0	0			
9	0	0	1	0	0	0	0				



Feature Engineering

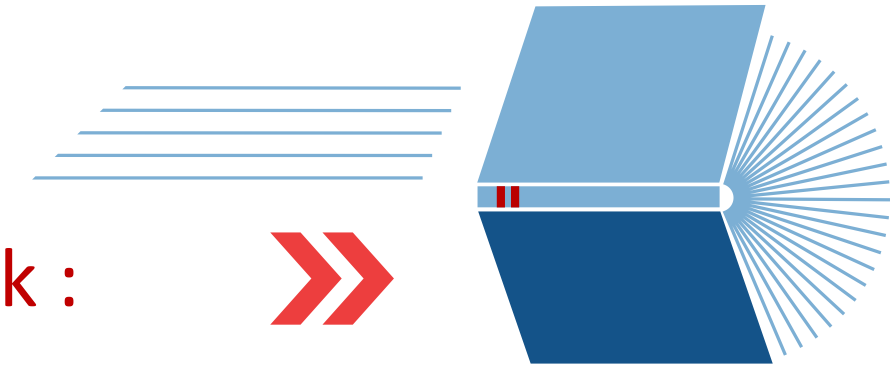
Cross Validation

Regression Algorithm	Linear Regression	Ridge Regression	Polynomial Regression (degree = 2)	Lasso Regression
Training Score	0.075843	0.044664	0.068357	0.066383
Validation Score	0.029533	0.00066	0.006979	0.002216

Modeling After feature Engineering :

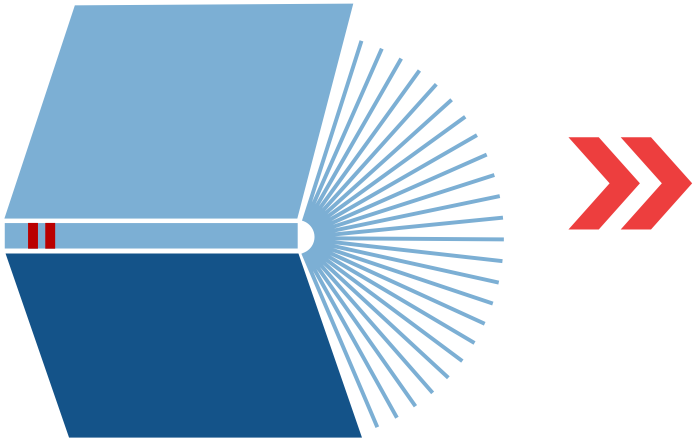
- Linear Regression Score Result:
- R^2 = 0.119

Conclusion & future work :



1. Data obtained via scraping websites are not reliable enough to validate a model's performance.

2. In future work we will co more effort to obtain more related features to reduce Bias



thank you