

## Q8\_2347116

October 10, 2023

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

file=pd.read_csv("books_dataset.csv")
```

```
[ ]: file.describe()
```

```
[ ]:
```

	price	save	pages	size	customer_reviews \
count	199.000000	185.000000	199.000000	199.000000	198.000000
mean	7.287588	7.978108	327.984925	1611.894472	379.393939
std	4.533073	7.089678	156.606451	3457.478147	648.001587
min	0.000000	0.480000	15.000000	69.000000	1.000000
25%	4.700000	3.350000	218.000000	417.000000	45.250000
50%	6.640000	6.200000	352.000000	655.000000	142.000000
75%	9.195000	10.960000	417.000000	1314.000000	459.000000
max	39.840000	61.980000	1000.000000	26789.000000	6254.000000

  

	stars
count	199.000000
mean	3.829146
std	0.542120
min	0.000000
25%	4.000000
50%	4.000000
75%	4.000000
max	5.000000

```
[ ]: file.shape
```

```
[ ]: (199, 16)
```

```
[ ]: file.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 199 entries, 0 to 198
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -

```

```

0  url                199 non-null    object
1  title              199 non-null    object
2  author             199 non-null    object
3  price              199 non-null    float64
4  save               185 non-null    float64
5  pages              199 non-null    int64
6  description        199 non-null    object
7  book_size          199 non-null    object
8  size               199 non-null    int64
9  publisher          187 non-null    object
10 language           199 non-null    object
11 text_to_speech     199 non-null    object
12 x_ray              199 non-null    object
13 lending            199 non-null    object
14 customer_reviews   198 non-null    float64
15 stars              199 non-null    int64

```

dtypes: float64(3), int64(3), object(10)

memory usage: 25.0+ KB

```
[ ]: missing_data=file.isnull()
```

	price	save	pages	size	customer_reviews \
count	187.000000	187.000000	187.000000	187.000000	187.000000
mean	7.434973	7.556471	336.855615	1637.272727	393.695187
std	4.605565	7.298142	155.361685	3527.600390	661.534456
min	0.000000	0.000000	15.000000	89.000000	1.000000
25%	4.740000	2.980000	234.000000	424.000000	51.000000
50%	6.830000	6.000000	356.000000	655.000000	174.000000
75%	9.410000	10.865000	417.000000	1335.000000	460.000000
max	39.840000	61.980000	1000.000000	26789.000000	6254.000000

	stars
count	187.000000
mean	3.834225
std	0.474404
min	1.000000
25%	4.000000
50%	4.000000
75%	4.000000
max	5.000000

```
[ ]: # display null values
```

```

null_counts = missing_data.sum()
for column, count in null_counts.items():
    print(f"Column '{column}': {count} null values")

```

Column 'url': 0 null values

Column 'title': 0 null values

```
Column 'author': 0 null values
Column 'price': 0 null values
Column 'save': 14 null values
Column 'pages': 0 null values
Column 'description': 0 null values
Column 'book_size': 0 null values
Column 'size': 0 null values
Column 'publisher': 12 null values
Column 'language': 0 null values
Column 'text_to_speech': 0 null values
Column 'x_ray': 0 null values
Column 'lending': 0 null values
Column 'customer_reviews': 1 null values
Column 'stars': 0 null values
```

```
[ ]: # replace save null values with 0

file['save'].fillna(0, inplace=True)
null_counts = file.isnull().sum()
for column, count in null_counts.items():
    print(f"Column '{column}': {count} null values")
```

```
Column 'url': 0 null values
Column 'title': 0 null values
Column 'author': 0 null values
Column 'price': 0 null values
Column 'save': 0 null values
Column 'pages': 0 null values
Column 'description': 0 null values
Column 'book_size': 0 null values
Column 'size': 0 null values
Column 'publisher': 12 null values
Column 'language': 0 null values
Column 'text_to_speech': 0 null values
Column 'x_ray': 0 null values
Column 'lending': 0 null values
Column 'customer_reviews': 1 null values
Column 'stars': 0 null values
```

```
[ ]: # use ffill on publisher

file.dropna(subset=['publisher'], inplace=True)
null_counts = file.isnull().sum()
for column, count in null_counts.items():
    print(f"Column '{column}': {count} null values")
```

```
Column 'url': 0 null values
Column 'title': 0 null values
Column 'author': 0 null values
```

```

Column 'price': 0 null values
Column 'save': 0 null values
Column 'pages': 0 null values
Column 'description': 0 null values
Column 'book_size': 0 null values
Column 'size': 0 null values
Column 'publisher': 0 null values
Column 'language': 0 null values
Column 'text_to_speech': 0 null values
Column 'x_ray': 0 null values
Column 'lending': 0 null values
Column 'customer_reviews': 0 null values
Column 'stars': 0 null values

```

```
[ ]: stars=file.groupby('stars')
     stars['price'].mean()
```

```
[ ]: stars
     1      5.820000
     3      5.778889
     4      7.651795
     5     14.495000
     Name: price, dtype: float64
```

```
[ ]: random_samples = file.sample(n=20, random_state=42)

     random_samples.set_index(['author', 'stars'], inplace=True)

     print(random_samples)
```

url

author	stars	url
Sean Platt	4	<a href="http://www.amazon.com/Publish-Repeat-No-Luck-R...">http://www.amazon.com/Publish-Repeat-No-Luck-R...</a>
Tom Clancy	4	<a href="http://www.amazon.com/Threat-Vector-Jack-Ryan-...">http://www.amazon.com/Threat-Vector-Jack-Ryan-...</a>
Lisa Gardner	4	<a href="http://www.amazon.com/Fear-Nothing-Detective-W...">http://www.amazon.com/Fear-Nothing-Detective-W...</a>
Seth Godin	4	<a href="http://www.amazon.com/Graceful-Seth-Godin-eboo...">http://www.amazon.com/Graceful-Seth-Godin-eboo...</a>
Jocelyn K. Gleib	4	<a href="http://www.amazon.com/Manage-Your-Day-Day-Crea...">http://www.amazon.com/Manage-Your-Day-Day-Crea...</a>
Anthony Dias Blue	4	<a href="http://www.amazon.com/Complete-Book-Mixed-Drin...">http://www.amazon.com/Complete-Book-Mixed-Drin...</a>
A. S. A. Harrison	3	<a href="http://www.amazon.com/The-Silent-Wife-A-Novel-...">http://www.amazon.com/The-Silent-Wife-A-Novel-...</a>
Eric Ries	4	<a href="http://www.amazon.com/The-Lean-Startup-Entrepr...">http://www.amazon.com/The-Lean-Startup-Entrepr...</a>
Rick Murcer	3	<a href="http://www.amazon.com/Caribbean-Moon-Manny-Wil...">http://www.amazon.com/Caribbean-Moon-Manny-Wil...</a>
Guy Kawasaki	4	<a href="http://www.amazon.com/Enchantment-Changing-Hea...">http://www.amazon.com/Enchantment-Changing-Hea...</a>
Karen Berman	4	<a href="http://www.amazon.com/Financial-Intelligence-E...">http://www.amazon.com/Financial-Intelligence-E...</a>
Tony Abou-Ganim	4	<a href="http://www.amazon.com/The-Modern-Mixologist-Co...">http://www.amazon.com/The-Modern-Mixologist-Co...</a>
Pat Flynn	4	<a href="http://www.amazon.com/Let-Go-Pat-Flynn-ebook/d...">http://www.amazon.com/Let-Go-Pat-Flynn-ebook/d...</a>
J. A. Jance	4	<a href="http://www.amazon.com/Name-Withheld-Beaumont-M...">http://www.amazon.com/Name-Withheld-Beaumont-M...</a>
Danny Meyer	4	<a href="http://www.amazon.com/Setting-Table-Danny-Meye...">http://www.amazon.com/Setting-Table-Danny-Meye...</a>

Sophie Delaplaine	5	<a href="http://www.amazon.com/Bartending-Your-Way-To-M...">http://www.amazon.com/Bartending-Your-Way-To-M...</a>
Drew Eric Whitman	4	<a href="http://www.amazon.com/CA-HVERTISING-Ad-Agency-...">http://www.amazon.com/CA-HVERTISING-Ad-Agency-...</a>
Alexander Osterwalder	4	<a href="http://www.amazon.com/Business-Model-Generatio...">http://www.amazon.com/Business-Model-Generatio...</a>
John Grisham	4	<a href="http://www.amazon.com/Ford-County-Stories-John...">http://www.amazon.com/Ford-County-Stories-John...</a>
Tom Clancy	4	<a href="http://www.amazon.com/Command-Authority-Jack-R...">http://www.amazon.com/Command-Authority-Jack-R...</a>

title

\		
author	stars	
Sean Platt	4	Write. Publish. Repeat. (The No-Luck-Required ...
Tom Clancy	4	Threat Vector (Jack Ryan, Jr.) [Kindle Edition]
Lisa Gardner	4	Fear Nothing: A Detective D.D. Warren Novel (D...
Seth Godin	4	Graceful [Kindle Edition]
Jocelyn K. Gleib	4	Manage Your Day-to-Day: Build Your Routine, Fi...
Anthony Dias Blue	4	The Complete Book of Mixed Drinks: Over 1,000 ...
A. S. A. Harrison	3	The Silent Wife: A Novel [Kindle Edition]
Eric Ries	4	The Lean Startup: How Today's Entrepreneurs Us...
Rick Murcer	3	Caribbean Moon (A Manny Williams Thriller, Boo...
Guy Kawasaki	4	Enchantment: The Art of Changing Hearts, Minds...
Karen Berman	4	Financial Intelligence for Entrepreneurs: What...
Tony Abou-Ganim	4	The Modern Mixologist: Contemporary Classic Co...
Pat Flynn	4	Let Go by Pat Flynn [Kindle Edition]
J. A. Jance	4	Name Withheld: A J.P. Beaumont Mystery (J. P. ...
Danny Meyer	4	Setting the Table [Kindle Edition]
Sophie Delaplaine	5	Bartending Your Way To Millions [Kindle Edition]
Drew Eric Whitman	4	CA\$HVERTISING: How to Use More than 100 Secret...
Alexander Osterwalder	4	Business Model Generation: A Handbook for Visi...
John Grisham	4	Ford County: Stories [Kindle Edition]
Tom Clancy	4	Command Authority (A Jack Ryan Novel) [Kindle ...

		price	save	pages	\
author	stars				
Sean Platt	4	5.99	10.00	479	
Tom Clancy	4	4.62	4.37	608	
Lisa Gardner	4	10.91	17.04	417	
Seth Godin	4	2.99	0.00	45	
Jocelyn K. Gleib	4	3.99	10.96	211	
Anthony Dias Blue	4	0.99	19.00	386	
A. S. A. Harrison	3	4.99	11.01	337	
Eric Ries	4	10.99	15.01	336	
Rick Murcer	3	0.99	12.00	347	
Guy Kawasaki	4	10.78	5.22	237	
Karen Berman	4	16.50	13.50	304	
Tony Abou-Ganim	4	19.25	15.75	208	
Pat Flynn	4	2.99	0.00	65	
J. A. Jance	4	5.99	4.00	400	
Danny Meyer	4	9.35	6.64	338	
Sophie Delaplaine	5	3.99	3.96	132	

Drew Eric Whitman	4	9.99	5.00	209
Alexander Osterwalder	4	19.49	15.46	288
John Grisham	4	5.99	2.00	308
Tom Clancy	4	5.99	23.96	753

description

\	author	stars	
	Sean Platt	4	Write. Publish. Repeat. The No-Luck-Required G...
	Tom Clancy	4	#1 New York Times bestselling author Tom Clanc...
	Lisa Gardner	4	In #1 New York Times bestseller Lisa Gardner's...
	Seth Godin	4	In this 32-page short eBook, New York Times be...
	Jocelyn K. Gleib	4	Stop doing busywork. Start doing your best wor...
	Anthony Dias Blue	4	With more than 100 new recipes for cocktails, ...
	A. S. A. Harrison	3	This summer's Gone Girl - I gobbled it down in...
	Eric Ries	4	Most startups fail. But many of those failures...
	Rick Murcer	3	***Updated Edition as of Sept. 26, 2012*** Sma...
	Guy Kawasaki	4	Enchantment, as defined by bestselling busines...
	Karen Berman	4	Using the groundbreaking formula they introduc...
	Tony Abou-Ganim	4	Over the course of the past two decades, Tony ...
	Pat Flynn	4	Whether success or failure, I want the satisfac...
	J. A. Jance	4	An explosive novel of betrayal and blood venge...
	Danny Meyer	4	In October 1985, at age twenty-seven, Danny Me...
	Sophie Delaplaine	5	A seasoned bartender who earned her stripes wo...
	Drew Eric Whitman	4	Barely one in a hundred businesspeople knows t...
	Alexander Osterwalder	4	Business Model Generation is a handbook for vi...
	John Grisham	4	BONUS: This edition includes an excerpt from J...
	Tom Clancy	4	The #1 New York Times-bestselling author and m...

	author	stars	book_size	size	\
	Sean Platt	4	medium	1390	
	Tom Clancy	4	big	1464	
	Lisa Gardner	4	small	1692	
	Seth Godin	4	medium	126	
	Jocelyn K. Gleib	4	big	1046	
	Anthony Dias Blue	4	small	1237	
	A. S. A. Harrison	3	big	486	
	Eric Ries	4	big	1501	
	Rick Murcer	3	big	375	
	Guy Kawasaki	4	big	1690	
	Karen Berman	4	medium	4221	
	Tony Abou-Ganim	4	big	7554	
	Pat Flynn	4	medium	1359	
	J. A. Jance	4	big	253	
	Danny Meyer	4	big	305	
	Sophie Delaplaine	5	big	201	
	Drew Eric Whitman	4	medium	714	



Tom Clancy	4	English	Enabled	Enabled	Not Enabled
------------	---	---------	---------	---------	-------------

customer\_reviews

author	stars	
Sean Platt	4	239.0
Tom Clancy	4	1719.0
Lisa Gardner	4	262.0
Seth Godin	4	41.0
Jocelyn K. Gleib	4	302.0
Anthony Dias Blue	4	11.0
A. S. A. Harrison	3	1479.0
Eric Ries	4	452.0
Rick Murcer	3	537.0
Guy Kawasaki	4	288.0
Karen Berman	4	21.0
Tony Abou-Ganim	4	33.0
Pat Flynn	4	262.0
J. A. Jance	4	51.0
Danny Meyer	4	131.0
Sophie Delaplaine	5	1.0
Drew Eric Whitman	4	101.0
Alexander Osterwalder	4	320.0
John Grisham	4	455.0
Tom Clancy	4	1262.0