

# Book Recommendation System



## Raw Data Info

- Is used the dataset from Amazon as json
- Is converted json to DataFrame
- Row: 83057
- Columns: 102
  - bool(11)
  - float64(16)
  - int64(7)
  - object(68)

# Data Cleaning

	Asin	Title	CategoryPath	Price1	PropValues	Comments	Stars
0	1781934606	Warhammer Age of Sigmar - Im Reich der Unbegr...	[Bücher, Literatur & Fiktion, Unterhaltungslit...	{'\$numberDecimal': '12.99'}	{('PropertyName': 'Authors', 'PropertyValue': ...	23	{'\$numberDecimal': '3.9'}
1	1782325956	Mein Stickerbuch: Abenteuer Raumfahrt	[Bücher, Kinderbücher, Spiel & Spaß]	{'\$numberDecimal': '6.95'}	{('PropertyName': 'Authors', 'PropertyValue': ...	10	{'\$numberDecimal': '4.9'}
2	178232707X	Mein Wisch-und-weg-Buch Schulstart: Ich schrei...	[Bücher, Kinderbücher, Kunst, Musik & Fotografie]	{'\$numberDecimal': '6.95'}	{('PropertyName': 'Authors', 'PropertyValue': ...	27	{'\$numberDecimal': '4.5'}
3	1495335216	Praxishandbuch Healing Code: Weiterentwicklung...	[Bücher, Ratgeber, Gesundheit & Medizin]	{'\$numberDecimal': '9.80'}	{('PropertyName': 'Authors', 'PropertyValue': ...	93	{'\$numberDecimal': '4.3'}
4	178232321X	Mein Mode-Stickerbuch: Hochzeitskollektion	[Bücher, Kinderbücher, Spiel & Spaß]	{'\$numberDecimal': '6.95'}	{('PropertyName': 'Authors', 'PropertyValue': ...	18	{'\$numberDecimal': '4.7'}

```
[{'PropertyName': 'Authors', 'PropertyValue': 'Horus W. Odenthal'},  
 {'PropertyName': 'Format',  
  'PropertyValue': '(Deutsch) Broschiert - 4. Oktober 2019'},  
 {'PropertyName': 'Title',  
  'PropertyValue': 'Warhammer Age of Sigmar - Im Reich der Unbegrabenen: Gotrek'},  
 {'PropertyName': 'ISBN-10', 'PropertyValue': '1781934606'},  
 {'PropertyName': 'ISBN-13', 'PropertyValue': '978-1781934609'},  
 {'PropertyName': 'Broschiert', 'PropertyValue': '448 Seiten'},  
 {'PropertyName': 'Größe und/oder Gewicht',  
  'PropertyValue': '14.8 x 2.5 x 21 cm'},  
 {'PropertyName': 'Herausgeber',  
  'PropertyValue': 'Black Library; 1. Auflage (4. Oktober 2019)'},  
 {'PropertyName': 'Sprache:', 'PropertyValue': 'Deutsch'},  
 {'PropertyName': 'Amazon Bestseller-Rang', 'PropertyValue': ''}]
```

## Selected Columns:

- **Asin**: Unique Id of book
- **Title**: Book Title
- **CategoryPath**: Genre of book
- **Price1** = Price of Book
- **Comments** = # of comments for book
- **Stars** = Average of Book rating rated by users
- **PropValues** = complicated:(



# Cleaning Steps

- Author, Publisher, Num of Page, Publication year of books are added from Propvalue columns

```
df4.drop(df4.Author[(df4.Author.isnull()) & (df4.Publisher.isnull())].index, inplace=True)
```

```
for i in df4[df4.Author.isnull()].index:  
    df4.Author[i]=df4.Publisher[i]
```

```
for i in book_recom[book_recom.Author=='unbekannt'].index:  
    book_recom.Author[i]=book_recom.Publisher[i]
```

```
for i in book_recom[book_recom.Author=='Author'].index:  
    book_recom.Author[i]=book_recom.Publisher[i]
```

```
for i in book_recom[book_recom.Author=='Klett '].index:  
    book_recom.Author[i]=book_recom.Publisher[28]
```

```
for i in book_recom[book_recom.Publisher.str.contains('Klett')].Publisher.index:  
    book_recom.Publisher[i]=book_recom.Publisher[28]
```

```
for i in book_recom[book_recom.Publisher.str.contains('PONS')].Publisher.index:  
    book_recom.Publisher[i]='PONS GmbH'
```

```
for i in book_recom[book_recom.Author.str.contains('PONS')].index:  
    book_recom.Author[i]='PONS GmbH'
```

```
for i in book_recom[book_recom.Author=="-"].index:  
    book_recom.Author[i]=book_recom.Publisher[i]
```

```
book_recom.drop([71646, 75191], axis=0, inplace=True)
```

```
book_recom.drop(book_recom[book_recom.Book_Price==0].index, axis=0, inplace=True)
```

```
book_recom.drop(book_recom[book_recom.Book_Star==0].index, axis=0, inplace=True)
```

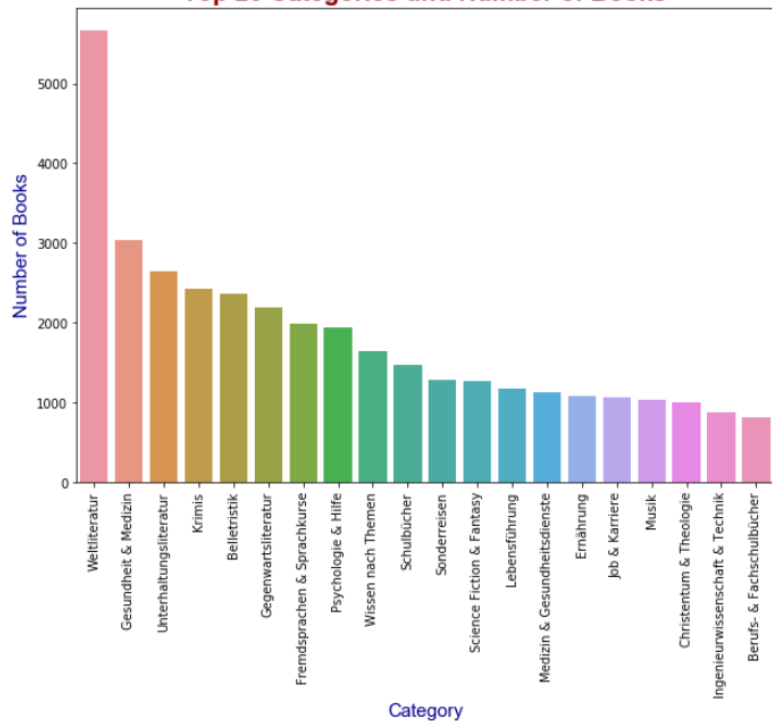
# Cleaned Data

	Asin	Title	Comments	Book_Price	Book_Star	Category	Author	Publication_year	Publisher	num_Page	Total_Rating_Score
0	1781934606	Warhammer Age of Sigmar - Im Reich der Unbegr...	23	12.99	3.9	Unterhaltungsliteratur	Horus W. Odenthal	2019	Black Library	448	89.7
1	1782325956	Mein Stickerbuch: Abenteuer Raumfahrt	10	6.95	4.9	Spiel & Spaß	Fiona Watt,Paul Nicholls	2017	Usborne Publishing	34	49.0
2	178232707X	Mein Wisch-und-weg-Buch Schulstart: Ich schrei...	27	6.95	4.5	Kunst, Musik & Fotografie	Hannah Watson,Marina Aizen	2017	Usborne Publishing	24	121.5
3	1495335216	Praxishandbuch Healing Code: Weiterentwicklung...	93	9.80	4.3	Gesundheit & Medizin	Simon Kraft	2014	CreateSpace Independent Publishing Platform	136	399.9
4	178232321X	Mein Mode-Stickerbuch: Hochzeitskollektion	18	6.95	4.7	Spiel & Spaß	Stella Baggott	2015	Usborne Publishing	34	84.6
5	178232545X	Ich weiß jetzt 100 Dinge mehr! Abenteuer Natur...	22	12.95	4.8	Wissenschaft & Natur	Alex Frith,Minna Lacey,Jerome Martin,Jonathan ...	2017	Usborne Publishing	128	105.6
6	1781934282	Die Verlorenen und die Verdammten: The Horus H...	82	15.00	4.8	Fantasy	Guy Haley	2019	Black Library	395	393.6
7	1782321675	Warum schlägt mein Herz?: 60 schlaue Fragen üb...	40	12.95	4.4	Wissenschaft & Natur	Katie Daynes,Marie-Eve Tremblay	2014	Usborne Publishing	14	176.0
8	1781933669	Warhammer 40.000 - Mephiston: Kreuzzug der Verd...	8	12.99	4.1	Fantasy	Ralph Hummel	2019	Black Library	347	32.8
9	1781932948	Warhammer 40.000 - Hereticus: Eisenhorn	17	12.99	4.5	Fantasy	Dan Abnett,Stefan Behrenbruch	2018	Black Library	400	76.5

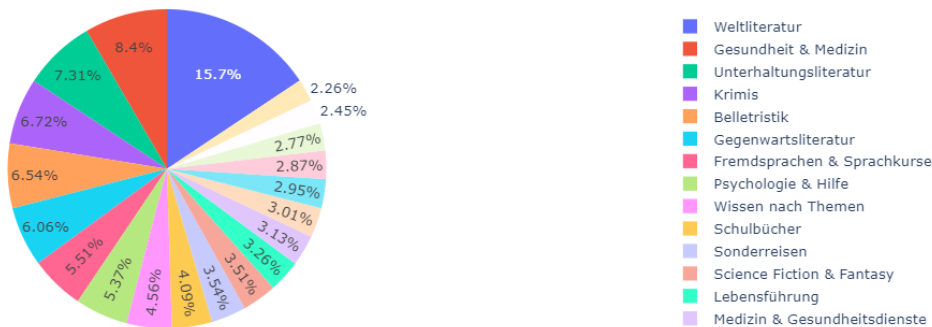
# Visualization

## Distribution of Top 20 Categories

Top 20 Categories und Number of Books



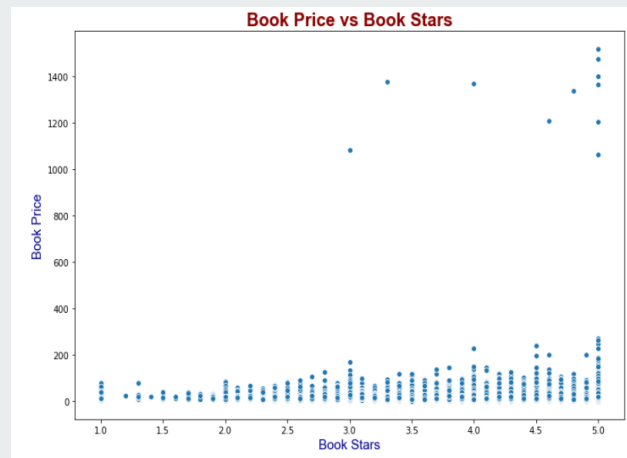
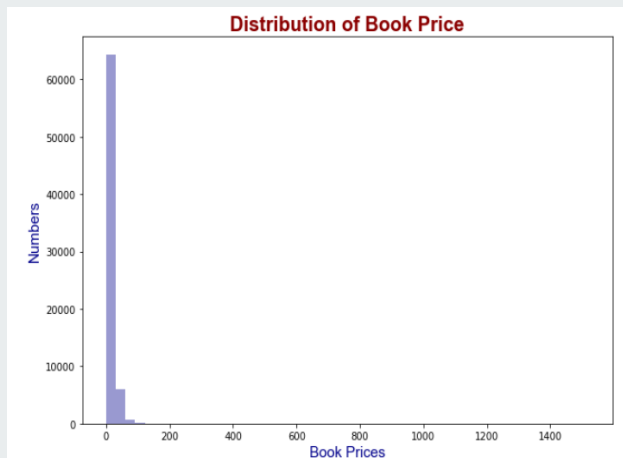
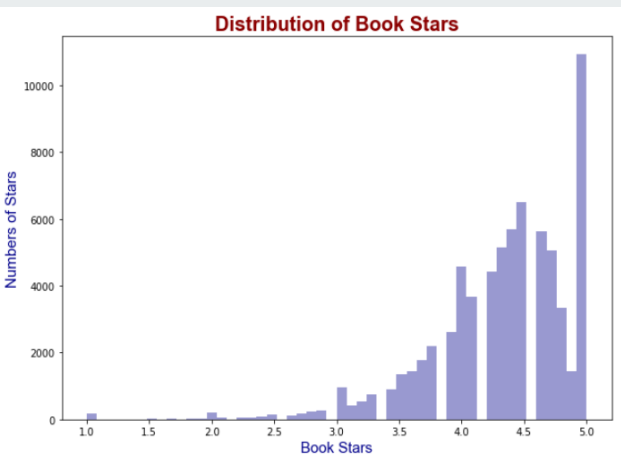
## Distribution of Top 20 Categories



# Visualization



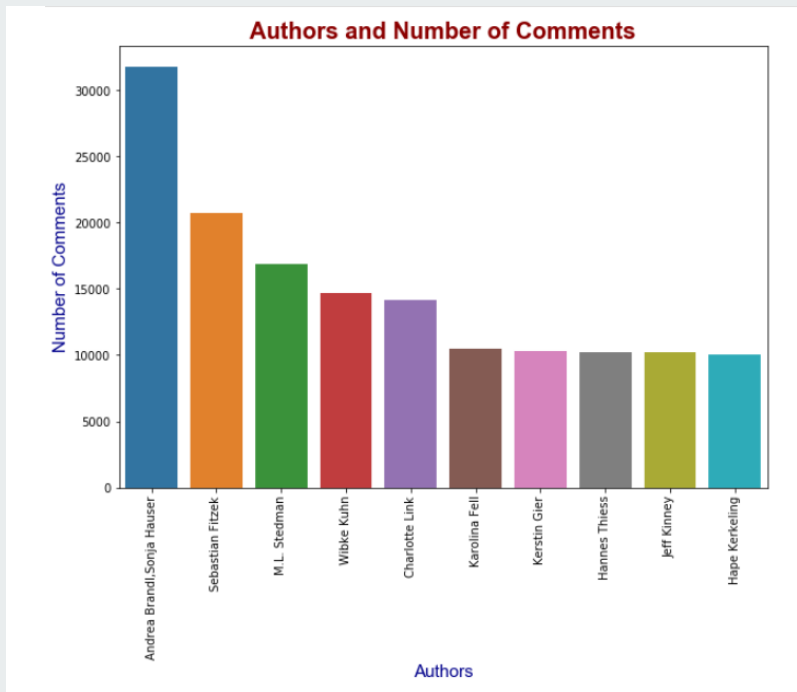
## Distribution of Books Stars and Book Price



# Visualization



## Authors and Number of Comments

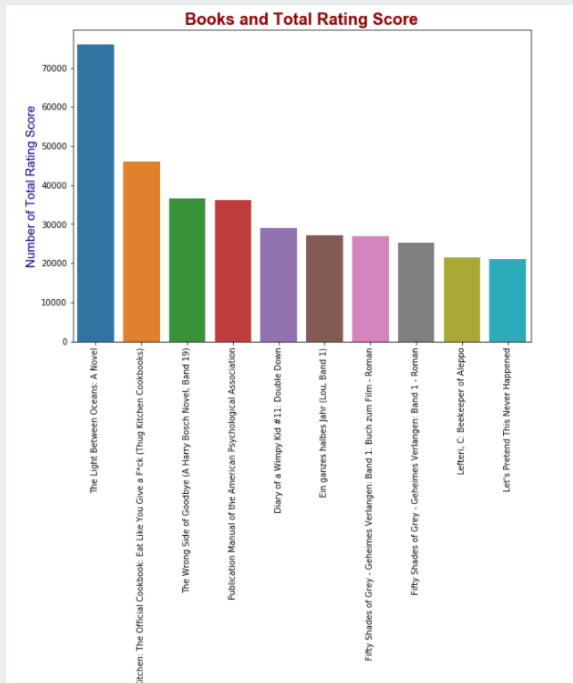




# Visualization



## Books and Number of Total Ratings Score

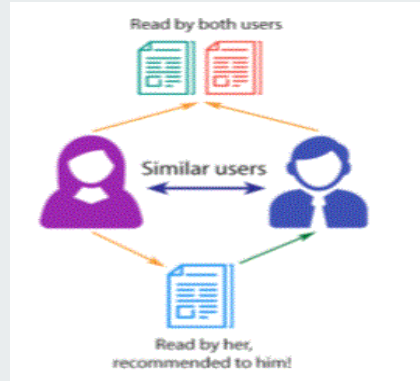


	Title	Total_Rating_Score	Comments	Book_Star
59772	The Light Between Oceans: A Novel	75946.5	16877.0	4.5
60054	Thug Kitchen: The Official Cookbook: Eat Like ...	46023.0	10005.0	4.6
59897	The Wrong Side of Goodbye (A Harry Bosch Novel...	36645.9	7797.0	4.7
51575	Publication Manual of the American Psychologic...	36256.5	8057.0	4.5
16573	Diary of a Wimpy Kid #11: Double Down	29131.2	6069.0	4.8
24509	Ein ganzes halbes Jahr (Lou, Band 1)	27153.6	5657.0	4.8
27915	Fifty Shades of Grey - Geheimes Verlangen: Ban...	26867.4	6397.0	4.2
27914	Fifty Shades of Grey - Geheimes Verlangen: Ban...	25297.0	6170.0	4.1
40749	Letferi, C: Beekeeper of Aleppo	21582.0	4796.0	4.5
41239	Let's Pretend This Never Happened	21018.8	4777.0	4.4

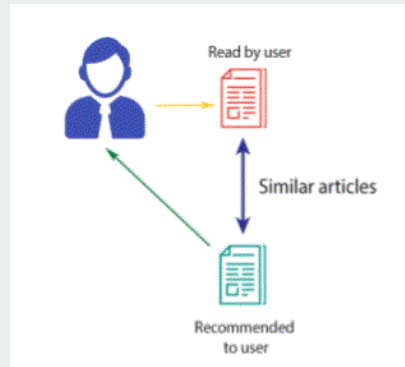
# Recommendation Systems

## Collaborative Filtering

- User based filtering
- Item based filtering



## Content Based Filtering



# Recommendation Systems



- **Book Title Column is cleaned**

- Lowercase
- Stopwords
- All except Letters
- Frequent & Rare Words Removal
- German Tokenization
- German Lemmatization

- **Word2vec**

- **Cosine Similarity Matrix**

- Row: 70K+
- Columns: 420

# Recommendation Model



## Word2vec Similarity Examples

```
print(model.similarity('stickerbuch', 'buch'))
```

```
0.49317926
```

```
print(model.similarity('langenscheidt', 'pons'))
```

```
0.83524114
```

```
print(model.similarity('vater', 'familie'))
```

```
0.8043321
```

```
print(model.similarity('wirtschafts', 'weihnachtsmarkt'))
```

```
0.31644213
```

# Model Prediction Example



Entered Book Title: “Warhammer Age of Sigmar - Im Reich der Unbegrabenen: Gotrek”

Word2vec:

```
Enter the Book Title Warhammer Age of Sigmar – Im Reich der Unbegrabenen: Gotrek
Top 5 similar books, similar to : Warhammer Age of Sigmar – Im Reich der Unbegrabenen: Gotrek are:

Broken Hearts – Gefährliche Nähe (The Original Heartbreakers)
Erzähl mir von den weißen Blüten (Die Asien-Saga, Band 1)
Anatomie eines Skandals: Du willst deinem Ehemann glauben Sie will ihn zerstören
Aus den Ruinen des Empires: Die Revolte gegen den Westen und der Wiederaufstieg Asiens
Das Stinktier kann doch nichts dafür: Die schönsten Schmunzelgeschichten aus der Arche. 3 Bücher in 1 Band
```

```
[63]: df_last[df_last.Title== "Warhammer Age of Sigmar – Im Reich der Unbegrabenen: Gotrek"]
```

	Asin	Title	Category	Comments	Book_Price	Book_Star	Publication_year	num_Page	Total_Rating_Score	Title_lower	...	90	91	92	93	94	95	96	97	98	99
0	1781934606	Warhammer Age of Sigmar - Im Reich der Unbegrab...	Unterhaltungsliteratur	-0.091684	-0.32413	-0.715066	2019	448	89.7	[warhammer, age, of, sigmar, reichen, unbegrab...	...	0.336366	-0.049011	-0.120272	0.490883	-0.295687	0.179389	-0.159696	0.025246	-0.003317	-0.291504

1 rows x 110 columns

```
[57]: df_last[df_last.Title== "Broken Hearts – Gefährliche Nähe (The Original Heartbreakers)"]
```

	Asin	Title	Category	Comments	Book_Price	Book_Star	Publication_year	num_Page	Total_Rating_Score	Title_lower	...	90	91	92	93	94	95	96	97	98	99
35820	3956492781	Broken Hearts - Gefährliche Nähe (The Original...	Gegenwartsliteratur	-0.091684	-0.474838	-0.541107	2016	448	92.0	[broken, hearts, gefährlich, nähen, the, origi...	...	0.362071	-0.2979	0.042942	0.450003	-0.472792	0.344643	-0.161712	-0.067104	-0.025165	-0.35797

1 rows x 110 columns

```
[58]: df_last[df_last.Title== "Erzähl mir von den weißen Blüten (Die Asien-Saga, Band 1)"]
```

	Asin	Title	Category	Comments	Book_Price	Book_Star	Publication_year	num_Page	Total_Rating_Score	Title_lower	...	90	91	92	93	94	95	96	97	98	99
52683	3548281931	Erzähl mir von den weißen Blüten (Die Asien-Sa...	Psychologie & Hilfe	-0.097784	-0.426611	-0.367147	2010	448	90.2	[erzähl, weiß, blüten, asien, saga]	...	0.422623	-0.298362	-0.121351	0.186727	-0.474419	-0.096226	0.028034	0.131275	0.075753	-0.437283

1 rows x 110 columns

```
[59]: df_last[df_last.Title== "Anatomie eines Skandals: Du willst deinem Ehemann glauben Sie will ihn zerstören"]
```

	Asin	Title	Category	Comments	Book_Price	Book_Star	Publication_year	num_Page	Total_Rating_Score	Title_lower	...	90	91	92	93	94	95	96	97	98	99
62020	3404177606	Anatomie eines Skandals: Du willst deinem Ehem...	Weltliteratur	-0.073382	-0.223156	-1.410905	2019	448	91.0	[anatomie, skandals, ehemann, glauben, zerstören]	...	0.146891	-0.207237	0.091626	0.149895	-0.004974	-0.015715	-0.14809	0.252491	-0.186178	-0.153408

1 rows x 110 columns

```
[61]: df_last[df_last.Title== "Das Stinktier kann doch nichts dafür: Die schönsten Schmunzelgeschichten aus der Arche. 3 Bücher in 1 Band"]
```

	Asin	Title	Category	Comments	Book_Price	Book_Star	Publication_year	num_Page	Total_Rating_Score	Title_lower	...	90	91	92	93	94	95	96	97	98	99
48554	3765541818	Das Stinktier kann doch nichts dafür: Die schö...	Grundbegriffe	-0.109985	-0.474838	0.154732	2019	448	88.0	[stinktier, schön, schmunzelgeschichten, arche...	...	0.474584	0.030513	-0.162656	0.123453	-0.388735	0.261189	0.096811	0.169889	0.042463	-0.19531