PROJECT 01

Amazon Bestselling Book Performance Report

amazon books

EXCEL Project

By:

Ankit Prajapati

1. **Data Cleaning**

2. **Exploratory Data Analysis (EDA)**

3. **Problem Statement**

4. **Data Visualization**

GitHUB

Send me Email

1. Project Goals

In this project, I'm going to clean data, run exploratory data analysis (EDA), find descriptive statistical measurements, and find the correlation among variables of a dataset. Also, I will create a dashboard to visualize the data.

2. Language, libraries, tools:

Libraries: Pivot Table, Data Analysis Tools
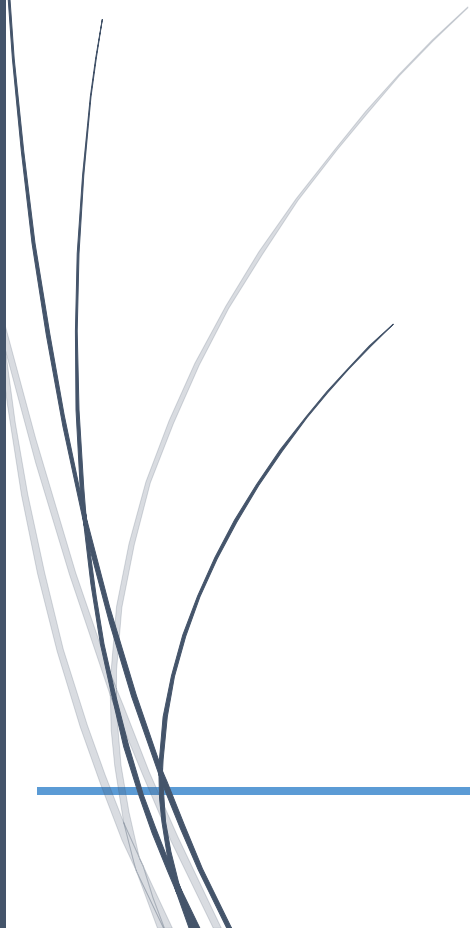
Application: Microsoft Excel

3 Data

This excel sheet that contains 7 columns with 550 records that are information about top 50 best-selling books on Amazon:

- **Name**: The Title of each book.
- **Author**: The name of the book's author.
- **User Rating**: The average rating of the book as provided by Amazon users.
- **Reviews**: The total number of reviews the book has received on Amazon.
- **Price**: The price of the book in US dollars.
- **Year**: The year the book was published.
- **Genre**: In this dataset there are two different project type: "Fiction", "Non-Fiction".

Data cleaning

Data Analysis Project: Excel Project
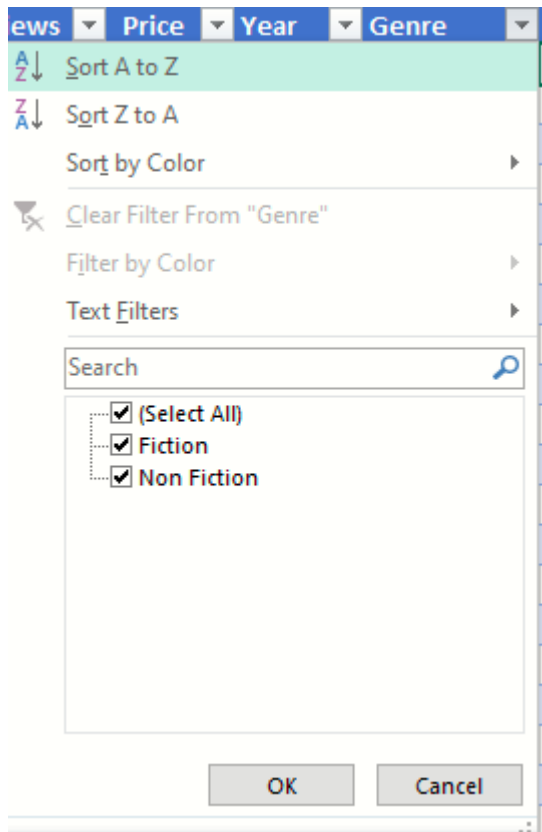
Duplicate rows:

We checked duplicate values and hopefully we don't have two same rows that contain the same values.

A	B	C	D	E	F	G	H
Name	Author	User Rating	Reviews	Price	Year	Genre	Rating
10-Day Green Smoothie Cleanse	JJ Smith	4.7	17,350	$8.0	2016	Non Fiction	Above Average
11/22/63: A Novel	Stephen King	4.6	2,052	$22.0	2011	Fiction	Average
12 Rules for Life: An Antidote to Ch	Jordan B. Peterson	4.7	18,979	$15.0	2018	Non Fiction	Above Average
1984 (Signet Classics)	George Orwell	4.7	21,424	$6.0	2017	Fiction	Above Average
5,000 Awesome Facts (About Every	National Geographic Kids	4.8	7,665	$12.0	2019	Non Fiction	Above Average
A Dance with Dragons (A Song of Ic	George R. R. Martin	4.4	12,643	$11.0	2011	Fiction	Poor
A Game of Thrones / A Clash of Kin	George				14	Fiction	Above Average
A Gentleman in Moscow: A Novel	Amor T				17	Fiction	Above Average
A Higher Loyalty: Truth, Lies, and Le	James				18	Non Fiction	Above Average
A Man Called Ove: A Novel	Fredrik				16	Fiction	Average
A Man Called Ove: A Novel	Fredrik				17	Fiction	Average
A Patriot's History of the United Sta	Larry Sc				10	Non Fiction	Average
A Stolen Life: A Memoir	Jaycee				11	Non Fiction	Average
A Wrinkle in Time (Time Quintet)	Madele				18	Fiction	Poor
Act Like a Lady, Think Like a Man: W	Steve H				09	Non Fiction	Average
Adult Coloring Book Designs: Stress	Adult C				16	Non Fiction	Poor
Adult Coloring Book: Stress Relievii	Blue St				15	Non Fiction	Average
Adult Coloring Book: Stress Relievii	Blue St				15	Non Fiction	Poor
Adult Coloring Books: A Coloring Bc	Colorin				15	Non Fiction	Poor
Alexander Hamilton	Ron Ch				16	Non Fiction	Above Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14.0	2014	Fiction	Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14.0	2015	Fiction	Average
Allegiant	Veronica Roth	3.9	6,310	$13.0	2013	Fiction	Very Poor
American Sniper: The Autobiograpl	Chris Kyle	4.6	15,921	$9.0	2015	Non Fiction	Average
And the Mountains Echoed	Khaled Hosseini	4.3	12,159	$13.0	2013	Fiction	Poor
Arguing with Idiots: How to Stop Sr	Glenn Beck	4.6	798	$5.0	2009	Non Fiction	Average
Astrophysics for People in a Hurry	Neil deGrasse Tyson	4.7	9,374	$9.0	2017	Non Fiction	Above Average
Autobiography of Mark Twain, Vol.	Mark Twain	4.2	491	$14.0	2010	Non Fiction	Poor
Baby Touch and Feel: Animals	DK	4.6	5,360	$5.0	2015	Non Fiction	Average

Remove Duplicates dialog:

To delete duplicate values, select one or more columns that contain duplicates.

Select All | Unselect All | ☑ My data has headers

Columns
- ☑ Name
- ☑ Author
- ☑ User Rating
- ☑ Reviews
- ☑ Price
- ☑ Year

OK | Cancel

Missing data:

Identify and clean any missing or inconsistent data in the "Department" column.

Note: There are not any missing or inconsistent data in the "Genre" Colum or any other column.

Data type:

In the next, I am going to check the data type of each column to make sure they are presenting the value that they were supposed to do. "Name", "Author", "Rating", and "Genre" are text. "User Rating", "Reviews, and "Year" are number and "Price" are currency.

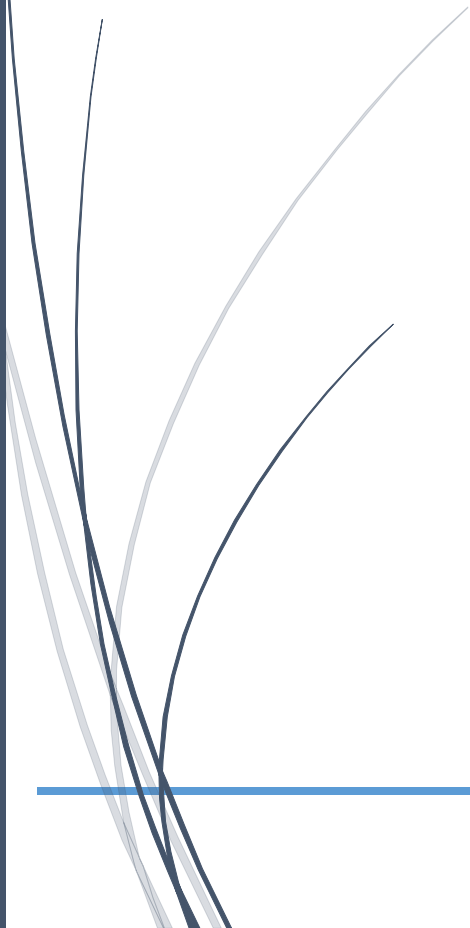# Data Analysis Project: Excel Project

Derived Columns:

To enrich data and improve the quality of data, I am going to make a calculated column. For doing this, I'm using user rating column to create a new column that shows text rating. The specific ranges used to define the text rating categories (e.g., "Exceptional", "Above Average", "Average", "Poor", "Very Poor".

Data Analysis Project: Excel Project

`2` | : | ✕ ✓ fx | =IF(C2>4.8,"Exceptional",IF(C2>4.6,"Above Average",IF(C2>4.5,"Average",IF(C2>4,"Poor","Very Poor"))))

Name	Author	User Rating	Reviews	Price	Year	Genre	Rating
10-Day Green Smoothie Cleanse	JJ Smith	4.7	17,350	$8	2016	Non Fiction	Above Average
11/22/63: A Novel	Stephen King	4.6	2,052	$22	2011	Fiction	Average
12 Rules for Life: An Antidote to Ch	Jordan B. Peterson	4.7	18,979	$15	2018	Non Fiction	Above Average
1984 (Signet Classics)	George Orwell	4.7	21,424	$6	2017	Fiction	Above Average
5,000 Awesome Facts (About Every	National Geographic Kids	4.8	7,665	$12	2019	Non Fiction	Above Average
A Dance with Dragons (A Song of Ic	George R. R. Martin	4.4	12,643	$11	2011	Fiction	Poor
A Game of Thrones / A Clash of Kin	George R. R. Martin	4.7	19,735	$30	2014	Fiction	Above Average
A Gentleman in Moscow: A Novel	Amor Towles	4.7	19,699	$15	2017	Fiction	Above Average
A Higher Loyalty: Truth, Lies, and Le	James Comey	4.7	5,983	$3	2018	Non Fiction	Above Average
A Man Called Ove: A Novel	Fredrik Backman	4.6	23,848	$8	2016	Fiction	Average
A Man Called Ove: A Novel	Fredrik Backman	4.6	23,848	$8	2017	Fiction	Average
A Patriot's History of the United Sta	Larry Schweikart	4.6	460	$2	2010	Non Fiction	Average
A Stolen Life: A Memoir	Jaycee Dugard	4.6	4,149	$32	2011	Non Fiction	Average
A Wrinkle in Time (Time Quintet)	Madeleine L'Engle	4.5	5,153	$5	2018	Fiction	Poor
Act Like a Lady, Think Like a Man: W	Steve Harvey	4.6	5,013	$17	2009	Non Fiction	Average
Adult Coloring Book Designs: Stress	Adult Coloring Book Designs	4.5	2,313	$4	2016	Non Fiction	Poor
Adult Coloring Book: Stress Relievi	Blue Star Coloring	4.6	2,925	$6	2015	Non Fiction	Average
Adult Coloring Book: Stress Relievi	Blue Star Coloring	4.4	2,951	$6	2015	Non Fiction	Poor
Adult Coloring Books: A Coloring Bc	Coloring Books for Adults	4.5	2,426	$8	2015	Non Fiction	Poor
Alexander Hamilton	Ron Chernow	4.8	9,198	$13	2016	Non Fiction	Above Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14	2014	Fiction	Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14	2015	Fiction	Average
Allegiant	Veronica Roth	3.9	6,310	$13	2013	Fiction	Very Poor
American Sniper: The Autobiograph	Chris Kyle	4.6	15,921	$9	2015	Non Fiction	Average
And the Mountains Echoed	Khaled Hosseini	4.3	12,159	$13	2013	Fiction	Poor
Arguing with Idiots: How to Stop Sn	Glenn Beck	4.6	798	$5	2009	Non Fiction	Average
Astrophysics for People in a Hurry	Neil deGrasse Tyson	4.7	9,374	$9	2017	Non Fiction	Above Average

Exploratory Data Analysis (EDA)

Data Analysis Project: Excel Project

Descriptive Measurement:

Now, let's analyze descriptive statics for quantitative variables. In this section we can find central tendency, dispersion, and shape measurements. Then we draw the distribution plot to compare with normal distribution. "Duration", "Project Cost" and "Project Benefit" are quantitative variables.

Price		
Count		550
Mean	$	13.10
Median	$	11.00
Min	$	-
Max	$	105.00
Range	$	105.00
Q1 (25%)	$	7.00
Q3 (75%)	$	16.00
St.d	$	10.84

User Rating	
Count	550
Mean	4.62
Median	4.70
Min	3.30
Max	4.90
Range	1.60
Q1 (25%)	4.50
Q3 (75%)	4.80
St.d	0.23

	Name	Author	Genre	Year
Count	550	550	550	550
Unique	350	248	2	11
Min				2009
Max				2019

Book Dataset Summary:

There are 550 entries in the dataset.

The dataset includes books from 248 unique authors and covers two genres: Fiction and Non-Fiction.

The average user rating for these books is approximately 4.62, with ratings ranging from 3.3 to 4.9.

Prices range from 0 to 105 dollar, with an average price of around $13.10

The books published between the years 2009 and 2019.

Data Analysis Project: Excel Project

Problem Statement:

Apply conditional Formatting to highlight books rating with "Average" in yellow?

Name	Author	User Rating	Reviews	Price	Year	Genre	Rating
10-Day Green Smoothie Cleanse	JJ Smith	4.7	17,350	$8	2016	Non Fiction	Above Average
11/22/63: A Novel	Stephen King	4.6	2,052	$22	2011	Fiction	Average
12 Rules for Life: An Antidote to Ch	Jordan B. Peterson	4.7	18,979	$15	2018	Non Fiction	Above Average
1984 (Signet Classics)	George Orwell	4.7	21,424	$6	2017	Fiction	Above Average
5,000 Awesome Facts (About Every	National Geographic Kids	4.8	7,665	$12	2019	Non Fiction	Above Average
A Dance with Dragons (A Song of Ic	George R. R. Martin	4.4	12,643	$11	2011	Fiction	Poor
A Game of Thrones / A Clash of King	George R. R. Martin	4.7	19,735	$30	2014	Fiction	Above Average
A Gentleman in Moscow: A Novel	Amor Towles	4.7	19,699	$15	2017	Fiction	Above Average
A Higher Loyalty: Truth, Lies, and Le	James Comey	4.7	5,983	$3	2018	Non Fiction	Above Average
A Man Called Ove: A Novel	Fredrik Backman	4.6	23,848	$8	2016	Fiction	Average
A Man Called Ove: A Novel	Fredrik Backman	4.6	23,848	$8	2017	Fiction	Average
A Patriot's History of the United Sta	Larry Schweikart	4.6	460	$2	2010	Non Fiction	Average
A Stolen Life: A Memoir	Jaycee Dugard	4.6	4,149	$32	2011	Non Fiction	Average
A Wrinkle in Time (Time Quintet)	Madeleine L'Engle	4.5	5,153	$5	2018	Fiction	Poor
Act Like a Lady, Think Like a Man: W	Steve Harvey	4.6	5,013	$17	2009	Non Fiction	Average
Adult Coloring Book Designs: Stress	Adult Coloring Book Designs	4.5	2,313	$4	2016	Non Fiction	Poor
Adult Coloring Book: Stress Relievi	Blue Star Coloring	4.6	2,925	$6	2015	Non Fiction	Average
Adult Coloring Book: Stress Relievi	Blue Star Coloring	4.4	2,951	$6	2015	Non Fiction	Poor
Adult Coloring Books: A Coloring Bc	Coloring Books for Adults	4.5	2,426	$8	2015	Non Fiction	Poor
Alexander Hamilton	Ron Chernow	4.8	9,198	$13	2016	Non Fiction	Above Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14	2014	Fiction	Average
All the Light We Cannot See	Anthony Doerr	4.6	36,348	$14	2015	Fiction	Average
Allegiant	Veronica Roth	3.9	6,310	$13	2013	Fiction	Very Poor
American Sniper: The Autobiograph	Chris Kyle	4.6	15,921	$9	2015	Non Fiction	Average
And the Mountains Echoed	Khaled Hosseini	4.3	12,159	$13	2013	Fiction	Poor
Arguing with Idiots: How to Stop Sr	Glenn Beck	4.6	798	$5	2009	Non Fiction	Average
Astrophysics for People in a Hurry	Neil deGrasse Tyson	4.7	9,374	$9	2017	Non Fiction	Above Average
Autobiography of Mark Twain, Vol.	Mark Twain	4.2	491	$14	2010	Non Fiction	Poor

Which book has the highest price?

Top 5 Most Expensive Books			
Name	Price	Year	Genre
Diagnostic and Statistical Manual of Mental Disorders	$105	2013	Non Fiction
Diagnostic and Statistical Manual of Mental Disorders	$105	2014	Non Fiction
The Twilight Saga Collection	$82	2009	Fiction
Hamilton: The Revolution	$54	2016	Non Fiction
The Book of Basketball: The NBA According to The Spc	$53	2009	Non Fiction

These books range in price from 53 to 105 dollar, with the "Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5th being the most expensive.

What is the average price of books for each year?

This chart shows prices over the years. Prices skyrocketed in 2014, with price settling in the $10 -$12 range. While there have been some ups and downs, it seems market has found its footing in recent years.

What is the correlation between the rating and the price of the books?

In this section, I'm going to see whether there is a relationship between User rating and Price and User rating and Reviews We use Pearson correlation.
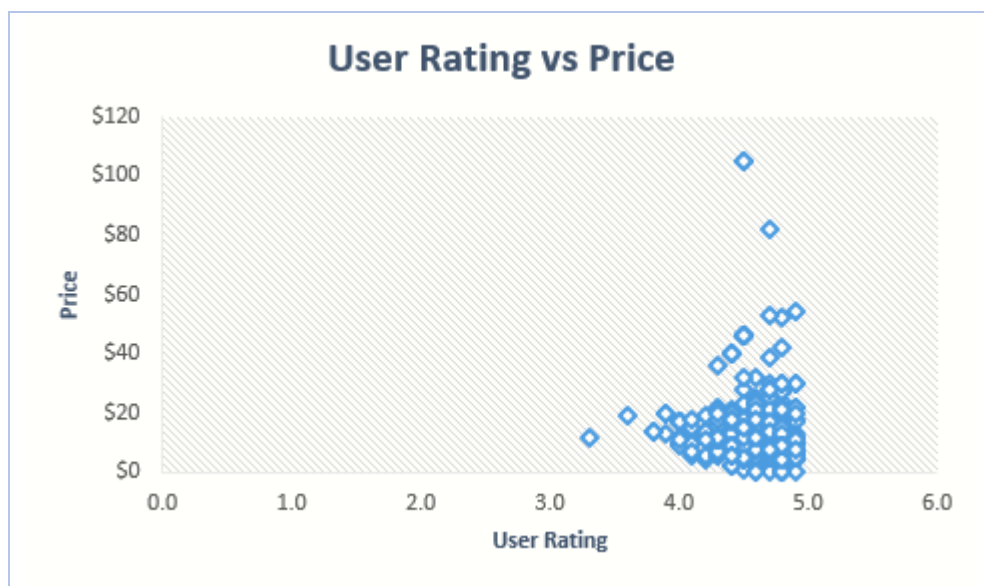
1. Covariance and Correlation

We use correlation and covariance functions in excel. The covariance shows whether we have positive or negative relationships, and correlation shows how strong that relationship is.
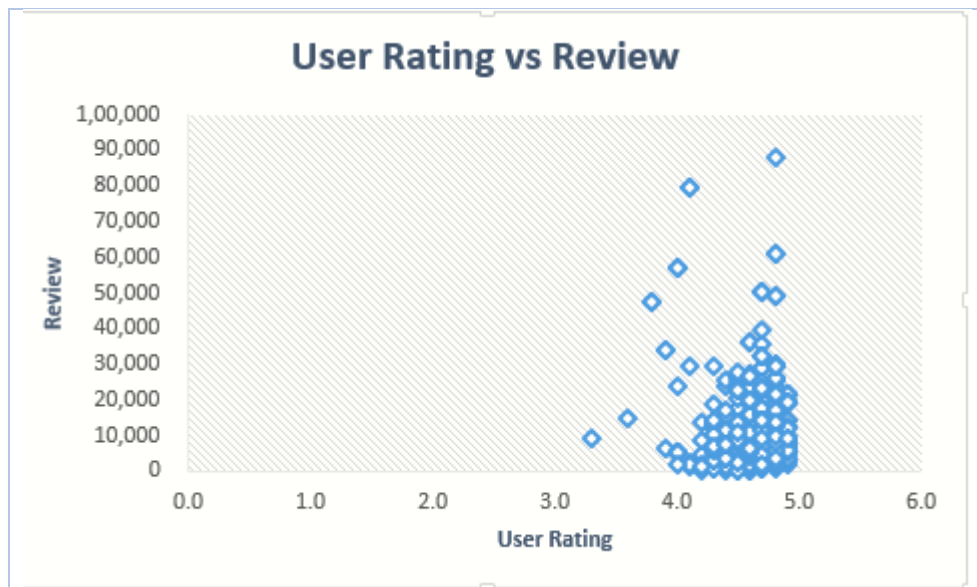
Correlation Analysis

User Rating vs Price		User Rating vs Review	
Covariance	-0.327522769	Covariance	-4.603909588
Correlation	-0.133086287	Correlation	-0.001729014

As we can see, there is a negative relationship between User Rating and Price. This means that as the price of the book increases, the user rating tends to decrease slightly, but this relationship is not strong.
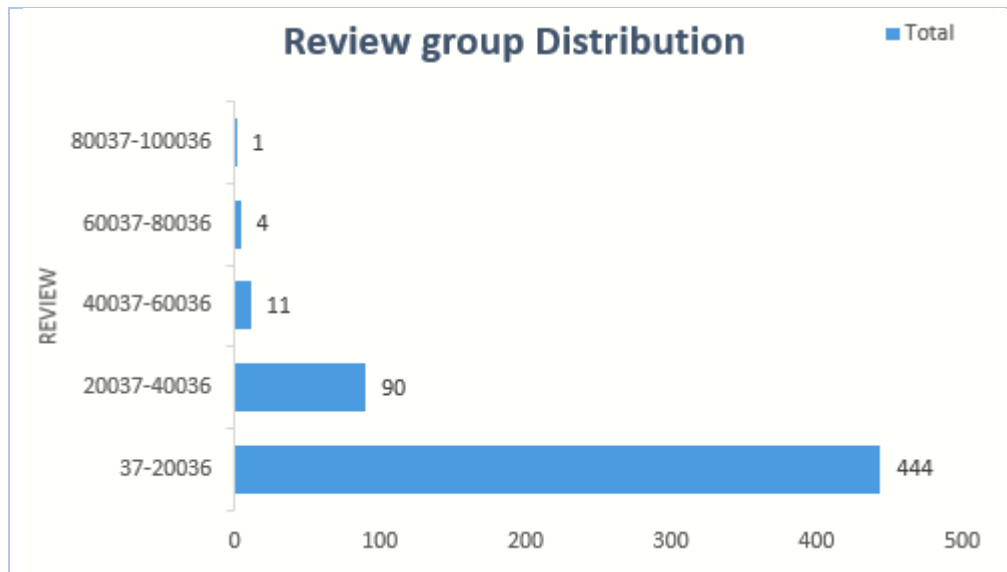
User Rating vs Price

The user rating and number of review also indicates a weak relationship. In other words, the number of reviews has almost no effect on the user rating of the books.

In both cases, the weak negative correlation indicates that the price of books and the number of reviews don't significantly influence user ratings
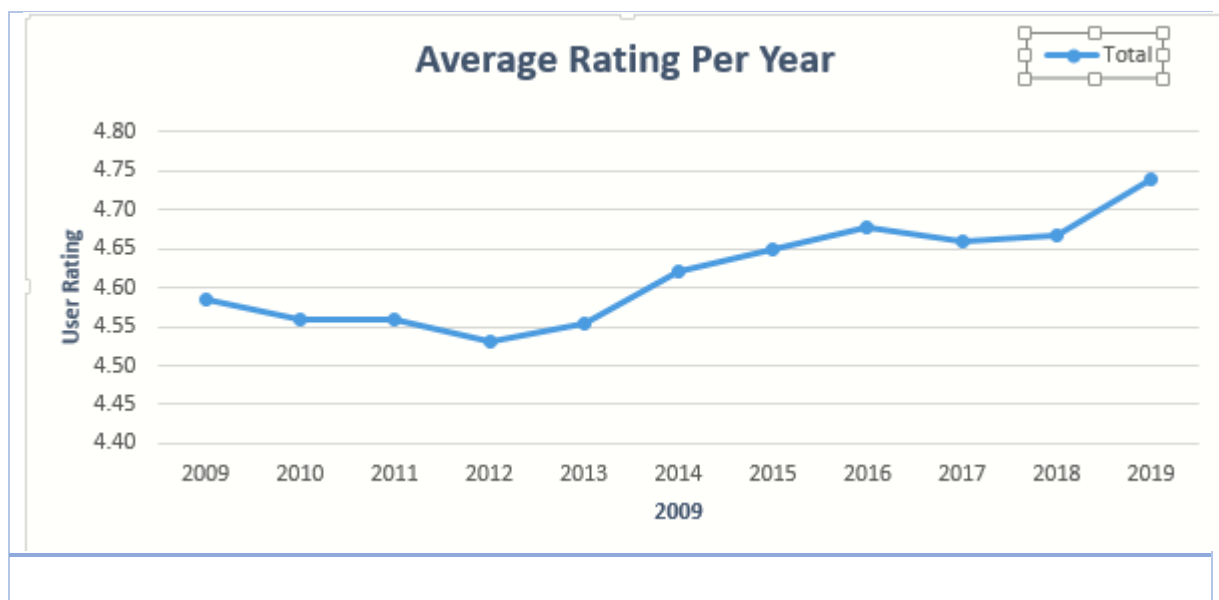
What is the distribution of the number of reviews received by the books?

Overall, the chart shows a right-skewed distribution, meaning there are more books with lower review scores than there are books with higher review scores.

What is the average rating over the year?

Looking at the chart, it seems like book rating have been pretty steady over the past decade. With a few up and downs.

Data Visualization

Data Visualization:

In this section we want to analyse data based on visualization. We use Microsoft Excel's charts to make a dashboard. To make dashboard, first I created multiple pivot table sheet and in each of them I generate the needed report based on the data and after that, I choose the best and most appropriate charts.

In left side, we see the slicers of filter that we can use them for filtering data in dashboard. They are independent and we can choose some of them or all of them at the same time. In the top middle, we see a column chart that shows average price in different year. At the bottom, we have bar chart that show the distribution of review. At the bottom middle we have line chart showing average rating for different year.

- END -