

DATA ANALYSIS PORTFOLIO

TABLE OF CONTENTS

PRIMARY REPORT

The Problem	1
The Design	2
Findings 1: VLOOKUP	3
Findings 2: Total Number of Subscribers by Category (Pie Chart)	4
Findings 3: Average number of subscribers for each subject (Bar Chart)	5
Findings 4: Average cost per subject at each level (Bar Chart)	6
Finding 5: Average content duration for each subject (Bar Chart)	7
Finding 6: Average rating per subject for each level (Column Chart)	8
Analysis	9
Conclusions	10

CAPSTONE PROJECT REPORT

The Problem	11
The Design	12
Findings 1: VLOOKUP	13
Findings 2: Sum of Reviews per Year (Stacked column chart)	14
Findings 3: Sum of User ratings per Year (3D Pie chart)	15
Findings 4: Sum of Reviews per Genre (3D Pie Chart)	16
Findings 5: Sum of User ratings per Genre (Column chart)	17
Findings 6: Sum of Price per Year (Column Chart)	18
Findings 7: Average of User ratings per Year (Scatter Chart)	19
Findings 8: Average of Reviews per Year (Column Chart)	20
Analysis	21
Conclusions	22

PRIMARY PROJECT REPORT

THE PROBLEM

Having been provided with data on courses from different topics to understand where opportunities to increase revenue may lie, I have to present the data on course revenue, then track the performance of courses.

THE DESIGN (Cleaning the Data using Excel)

1) Removing duplicates

Selected the entire data sheet data to remove duplicates.

Data > Remove Duplicates

2) Removing blank cells

Selected the entire sheet or dataset and to **Data > Create a filter**

Clicked on the Filter icon at the top of any column, then clicked on Filter by condition and select 'is empty'.

Blank cells appeared on top of the sheet and were removed.

3) Headers

Ensured I have clear and concise names for headers.

4) Find and replace

I examined the data, and I realized that the Web Development subject title was not the same as other subject titles. I used the Find and Replace function to make the Web Development subject consistent with other subjects.

Edit > Find and Replace

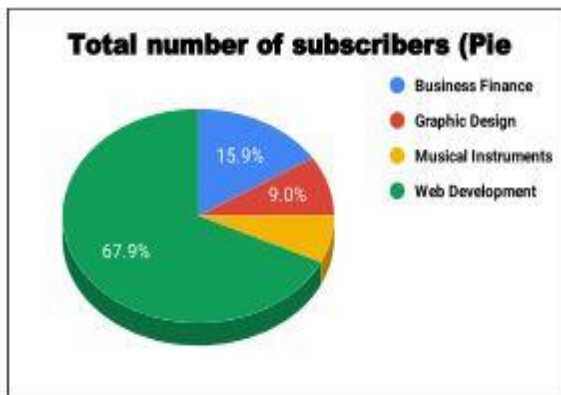
FINDINGS

Findings 1: VLOOKUP for 20 most subscribed courses, their level, free or paid, duration of course and date published.

course_id	num_subscribers	level	free or paid	duration	date established
41295	268923	Beginner Level	Free	10.5	2013-02-14
59014	161029	Expert Level	Free	3.5	2013-06-09
625204	121584	Beginner Level	Paid	43	2015-11-02
173548	120291	All Levels	Free	3	2014-04-08
764164	114512	Beginner Level	Paid	30.5	2016-03-08
19421	101154	All Levels	Free	4.5	2012-06-15
473160	98867	All Levels	Free	3	2015-04-13
94430	84897	All Levels	Paid	2	2013-10-10
130064	83737	Intermediate Level	Free	6.5	2014-07-19
364426	79612	All Levels	Paid	11.5	2015-03-12
238934	75499	Beginner Level	Paid	30	2014-08-07
756150	73783	Beginner Level	Paid	22	2016-02-11
21386	73110	All Levels	Free	2	2012-07-27
65330	72932	All Levels	Free	1	2013-09-25
405926	70773	All Levels	Free	6	2015-03-19
11174	69186	All Levels	Paid	27.5	2011-11-19
49798	65576	All Levels	Free	8	2013-04-20
314462	64128	Beginner Level	Free	1.5	2014-11-22
128946	59639	All Levels	Free	4.5	2013-12-09
289230	59361	Beginner Level	Paid	7	2014-09-24

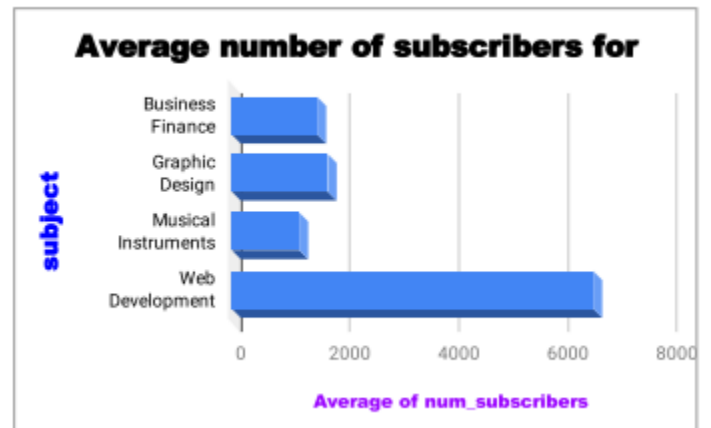
Findings 2: Total Number of Subscribers (Pie Chart)

subject	SUM of num_subscribers
Business Finance	1868711
Graphic Design	1063148
Musical Instruments	846689
Web Development	7981935



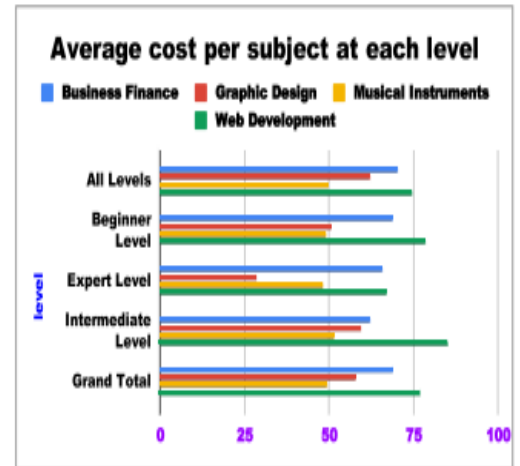
Findings 3: Average number of subscribers for each subject (Bar Chart)

subject	AVERAGE of num_subscriber
Business Finance	1569.026868
Graphic Design	1766.026578
Musical Instruments	1245.130882
Web Development	6635.024938
Grand Total	3199.260881



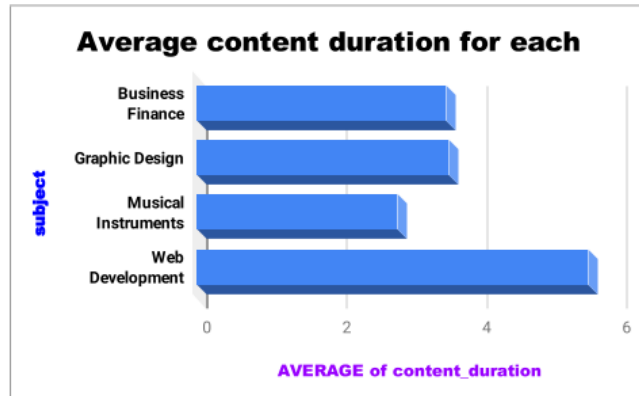
Findings 4: Average cost per subject at each level (Bar Chart)

AVERAGE of price level	subject				
	Business Finance	Graphic Design	Musical Instrume	Web Developmer	Grand Total
All Levels	70.19747235	62.11940299	49.58333333	74.54976303	66.75324675
Beginner Level	68.73433584	50.67934783	48.98496241	78.54265403	65.24390244
Expert Level	65.8	28.57142857	48.33333333	67.14285714	58.01724138
Intermediate Level	62.01492537	59.40789474	51.6025641	85.07462687	66.94312796
Grand Total	68.69437448	57.89036545	49.55882353	77.03657523	66.11534276



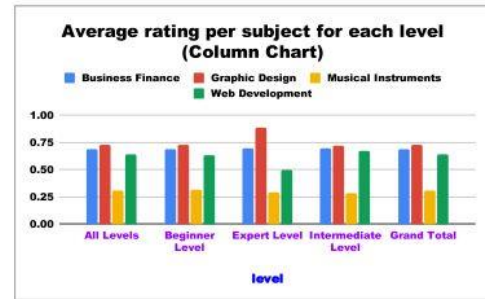
Finding 5: Average content duration for each subject (Bar Chart)

subject	AVERAGE of content_duration
Business Finance	3.557696613
Graphic Design	3.585465116
Musical Instruments	2.854387255
Web Development	5.586658354
Grand Total	4.096137106



Finding 6: Average rating per subject for each level (Column Chart)

AVERAGE of Rating	subject				
Level	Business Finance	Graphic Design	Musical Instrume	Web Developer	Grand Total
All Levels	0.6915165877	0.7309850746	0.312037037	0.6454660348	0.6193714286
Beginner Level	0.6862155388	0.7270652174	0.3138345865	0.6353791469	0.5973170732
Expert Level	0.7	0.8842857143	0.2966666667	0.4985714286	0.5901724138
Intermediate Level	0.6953731343	0.7215789474	0.281025641	0.6708955224	0.6157345972
Grand Total	0.6903526448	0.7303820598	0.3089117647	0.6430507066	0.6108677911



ANALYSIS

- Web Development has the highest percentage/ratio of subscribers, followed by Business Finance, Graphics Design and Musical Instruments courses.
- Web development has the highest average number of subscribers as well as the highest average content duration, followed by Graphics Design, Business Finance and Musical Instrument courses.
- Web Development has the most average cost per subject, followed by Business Finance, Graphics Design and Musical Instrument. Hence;
 - At All Levels and the Beginners' Level, Web Development has the highest average cost, followed by Business Finance, Graphics Design and Music Instrument.
 - At the Expert Levels, Web Development has the highest, followed by Business Finance, Musical Instruments and then Graphics Design.
 - At the Intermediate Level, Web Development has the highest, followed by Business Finance, Graphics Design and Musical Instruments.
 - Also, we have the highest average cost at Intermediate Levels, All Levels, Beginner Level and Expert Levels respectively.
- Graphics Design has the most average ratings, followed by Business Finance, Web Development and Musical Instruments.
 - Here, we have most average ratings at the All Levels, Intermediate Levels, Beginners Levels and Expert Levels respectively.

CONCLUSIONS

- Web Developments courses sell more than the other 3 courses and its price can be increased.
Likewise, its rating needs to be increased, and this can be achieved by implementing more quality contents.
- Musical Instruments courses have a very low response/turn out; and thus, needs to be marketed efficiently, incentivized and filled with premium and quality contents. Its price can as well be reduced to drive subscribers to the field.
- We have more costs at Intermediate Level, which shows that most who pay for the courses are not novices or newbies. They have a pre-knowledge about the field of their choosing.
- Graphics Design courses seem to have contents that are appealing to subscribers as they tend to give more/positive ratings to the course.

CAPSTONE PROJECT REPORT

THE PROBLEM

Having been provided with a dataset that illustrates Amazon's top 50 bestsellers from 2009 – 2019. An analysis is required to extract the following insights and visualizations:

- i. Top 20 bestsellers based on reviews, hereby presenting the Ratings, Reviews, Authors, Titles, Price, Year and Genre of each.
- ii. Sum of Reviews per Year (illustrated with a stacked column chart)
- iii. Sum of Users ratings per Year (Pie Chart)
- iv. Sum of Reviews per Genre (Pie Chart)
- v. Users Ratings per Genre (Column Chart)
- vi. Sum of Price per Year (Column Chart)
- vii. Average of Users Ratings per Year (Scatter Chart)
- viii. Average of Reviews per Year (Column Chart)

Hence, derive a conclusion based on the above analysis – expounding the books with best reactions and other necessary attributes.

THE DESIGN (Cleaning the Data using Excel)

1) Removing duplicates

Selected the entire data sheet data to remove duplicates.

Data > Remove Duplicates

2) Removing blank cells

Selected the entire sheet or dataset and to **Data > Create a filter**

Clicked on the Filter icon at the top of any column, then clicked on Filter by condition and select 'is empty'.

Blank cells appeared on top of the sheet and were removed.

3) Headers

Ensured I have clear and concise names for headers.

4) Expanding Cells to contain contents.

I highlighted the whole worksheet and hovered my cursor to the joint of two cells headers and upon the sight of a double-ended arrow, I double clicked.

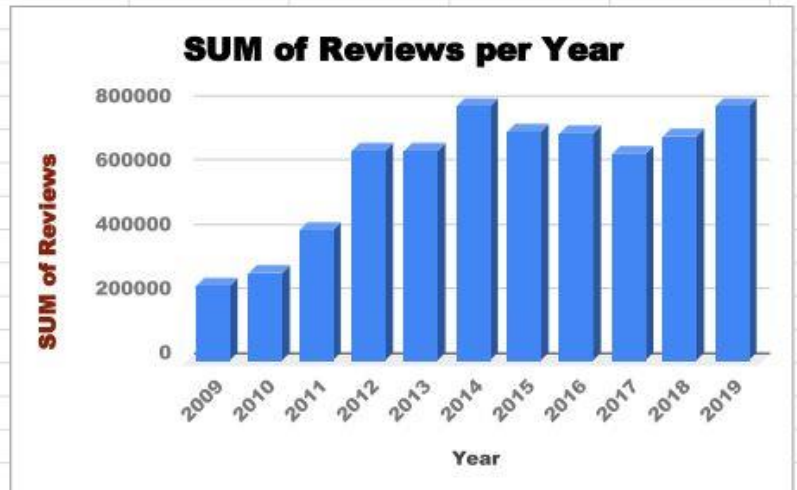
FINDINGS

Findings 1: VLOOKUP for Top 20 bestsellers based on reviews, hereby presenting the Titles, Authors, Ratings, Reviews, Price, Year and Genre of each.

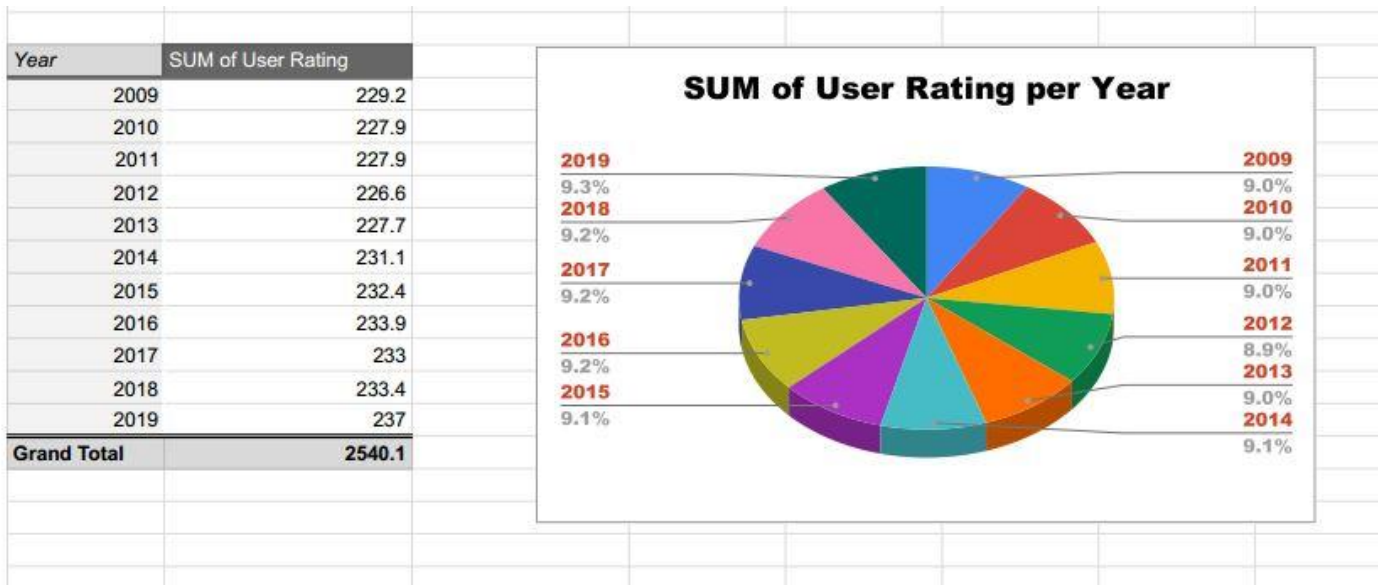
Name	Author	User Rating	Reviews	Price	Year	Genre
Where the Crawdads Sing	Delia Owens	4.8	87841	15	2019	Fiction
The Girl on the Train	Paula Hawkins	4.1	79446	7	2016	Fiction
The Girl on the Train	Paula Hawkins	4.1	79446	18	2015	Fiction
Becoming	Michelle Obama	4.8	61133	11	2018	Non Fiction
Becoming	Michelle Obama	4.8	61133	11	2019	Non Fiction
Gone Girl	Gillian Flynn	4	57271	9	2014	Fiction
Gone Girl	Gillian Flynn	4	57271	10	2012	Fiction
Gone Girl	Gillian Flynn	4	57271	10	2013	Fiction
The Fault in Our Stars	John Green	4.7	50482	7	2014	Fiction
The Fault in Our Stars	John Green	4.7	50482	13	2012	Fiction
The Fault in Our Stars	John Green	4.7	50482	13	2013	Fiction
The Fault in Our Stars	John Green	4.7	50482	13	2014	Fiction
The Nightingale: A Novel	Kristin Hannah	4.8	49288	11	2015	Fiction
The Nightingale: A Novel	Kristin Hannah	4.8	49288	11	2016	Fiction
Fifty Shades of Grey: Book One of the Fifty Shades Trilogy (Fifty Shades of Grey Series)	E L James	3.8	47265	14	2012	Fiction
Fifty Shades of Grey: Book One of the Fifty Shades Trilogy (Fifty Shades of Grey Series)	E L James	3.8	47265	14	2013	Fiction
The Martian	Andy Weir	4.7	39459	9	2015	Fiction
All the Light We Cannot See	Anthony Doerr	4.6	36348	14	2014	Fiction
All the Light We Cannot See	Anthony Doerr	4.6	36348	14	2015	Fiction

Findings 2: Sum of Reviews per Year (illustrated with a stacked column chart)

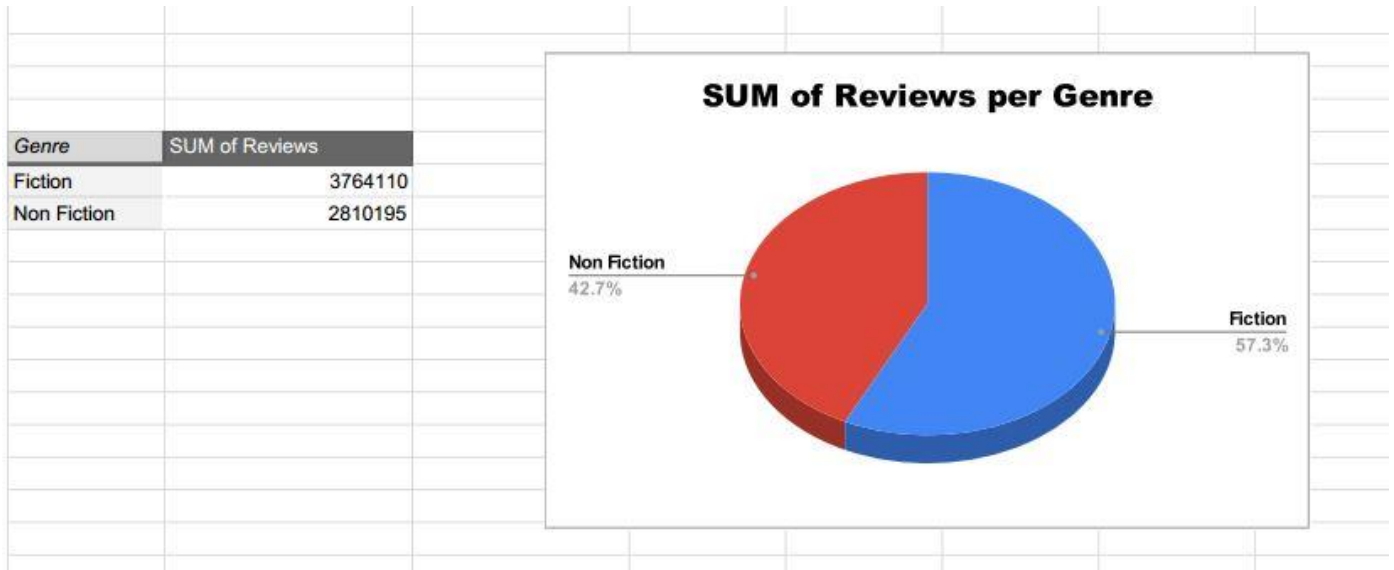
Year	SUM of Reviews
2009	235506
2010	273981
2011	405041
2012	654546
2013	654907
2014	792997
2015	711669
2016	709800
2017	644420
2018	696521
2019	794917
Grand Total	6574305



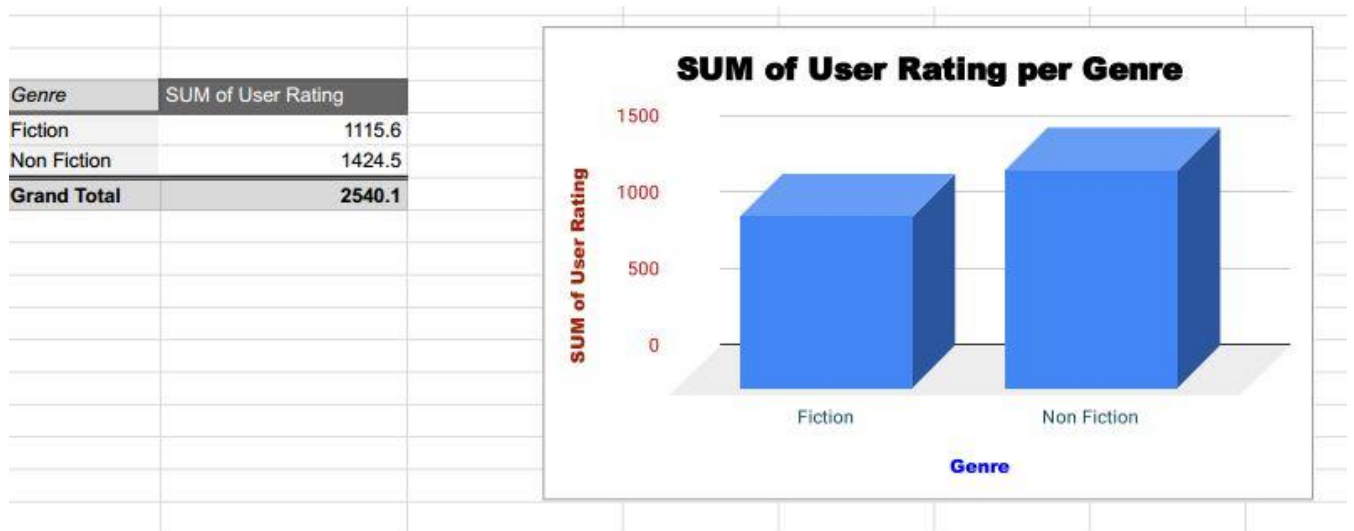
Findings 3: Sum of Users ratings per Year (Pie Chart)



Findings 4: Sum of Reviews per Genre (Pie Chart)



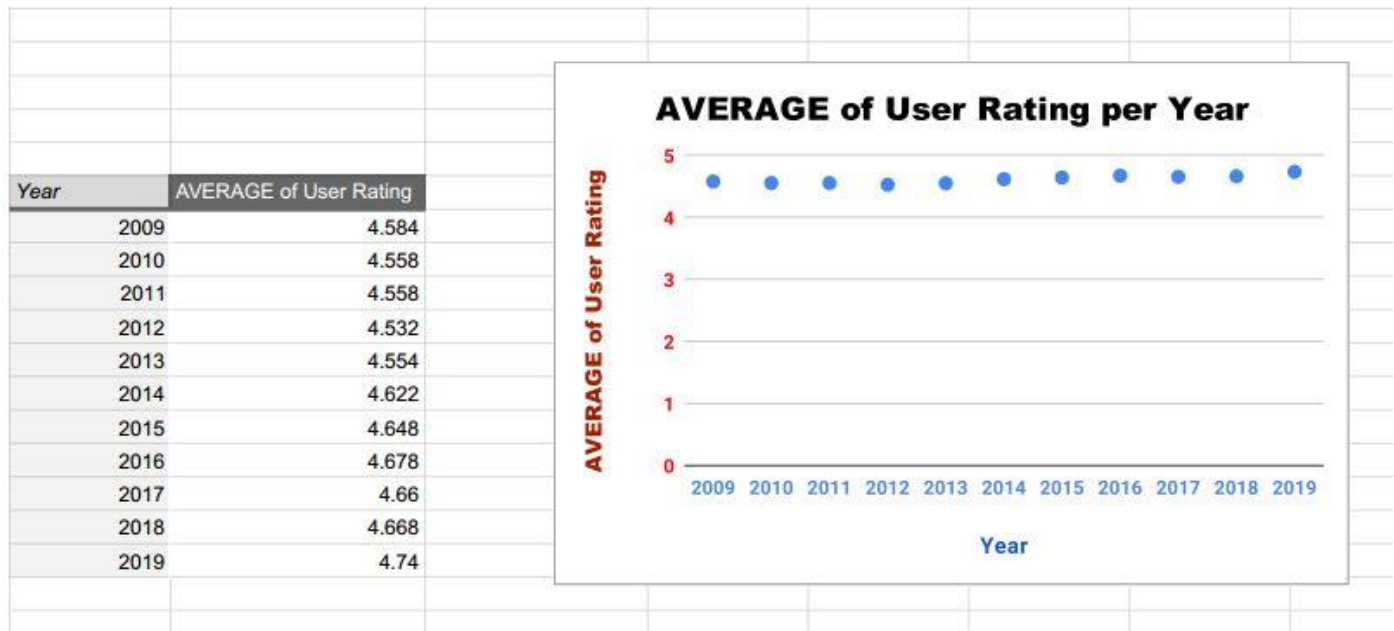
Findings 5: Users Ratings per Genre (Column Chart)



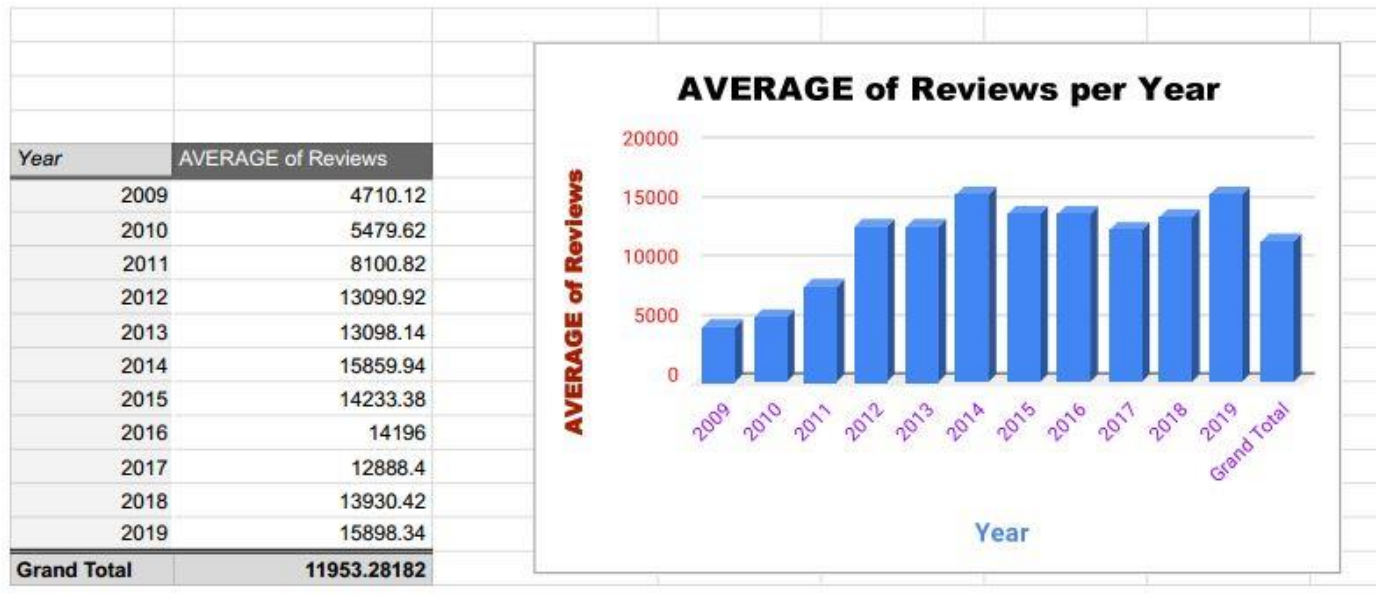
Findings 6: Sum of Price per Year (Column Chart)



Findings 7: Average Users Ratings per Year (Scatter Chart)



Findings 8: Average Reviews per Year (Column Chart)



ANALYSIS

- The Fictional 2019 book “Where the Crawdads Sing” by Delia Owens was the bestseller with highest reviews, with a rating of 4.8 and 87841 reviews.
- 2019 books have the highest reviews, followed by 2014 book; and 2009 books have the least reviews.
- 2019 books have the highest ratings followed by 2016 books; while 2012 has the least ratings.
- Fictional books have the highest reviews with 57.3% while Non-Fictional books have 42.7% reviews.
- Non-Fictional books have more user ratings than Fictional books.
- 2009 books have the most price tags, followed by 2012 and 2014 books respectively; while 2019 books have the least price tags.
- 2019 books have the most average user ratings, followed by 2016 books; while 2012 books have the least average user ratings.
- 2019 books have the highest average reviews, followed by 2014 books; while 2009 books have the least average reviews.

CONCLUSIONS

- Readers enjoyed reading 2019 books the most and assumedly got the value for their money and time, by giving the best ratings and most reviews to the books in the category.
- Readers returned with the most reviews for Fictional books than Non-Fictional books, but ended up giving the best ratings to Non-Fictional books in contrast to Fictional books.
This shows that readers enjoyed Non-Fictional books than Fictional books.
- 2009, 2012 and 2011 books are the most expensive set of books respectively, while 2019 books are the cheapest.

Conclusively, it can be assumed that the reduction in the price of books (as shown in the above analysis) yields a positive remark in readers as the year with the cheapest books appears to be the overall and all-time bestseller.