DATA ANALYSIS PORTFOLIO

FOLARIN OSUOLALE

Address: Obafemi Awolowo University, Ile-Ife, Osun State
Mobile: +2348145425241, +2348113772107
E-Mail: www.osuolalefolarin@gmail.com
LinkedIn bio: https://linkedin.com/in/folarin-osuolale
GitHub Repository: https://github.com/Folarinosuolale/my_data_science_projects

Folarin Osuolale is a Data Scientist and Machine Learning (AI) enthusiast. His goal as a Data Analyst is to provide businesses, firms, and private owners with useful data insights to help develop their businesses and works. He can also help with data processing, anomaly detection, wrangling, and visualization.

TABLE OF CONTENTS

PROFESSIONAL BACKGROUND

Folarin Osuolale uses different tools and methods to efficiently analyze data from different sources and discover useful patterns for further business decisions.

Machine learning is one of the best methods he uses to discover the data insights and listed below are some tools he uses in manipulations:

- Python
- NumPy
- Pandas
- Matplotlib
- Tableau
- and some more.

He has worked on some interesting projects which are all available in GitHub repository and can be accessed via the link provided above.

He is currently a student intern at **EntryLevel** and a Data Analyst at **AFKiT** Technologies.

He is proficient in the following skill sets:

- **Microsoft Office**
- **Web Development**
- **Public Speaking**
- **Machine Learning**
- **Excellent Verbal and Written Communication Skill**
- **Data Analysis/Data Science**
- **Python (Programming Language)**
- **People Management**
- **Problem Solving, strong analytical and numeracy skill**
- **PHP/MySQL**

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

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

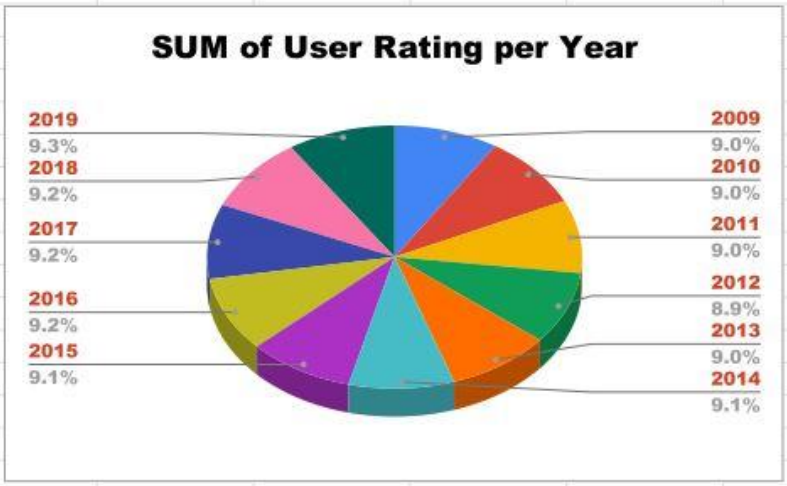
Name	Author	User Rating	Reviews	Price	Year	Genre
Where the Crawdads Sing	Delia Owens	4.8	87841	15	2019	Fiction
The Girl on the Train	Paula Hawkins	4.1	79446	7	2016	Fiction
The Girl on the Train	Paula Hawkins	4.1	79446	18	2015	Fiction
Becoming	Michelle Obama	4.8	61133	11	2018	Non Fiction
Becoming	Michelle 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

<u>Findings 2:</u> 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**

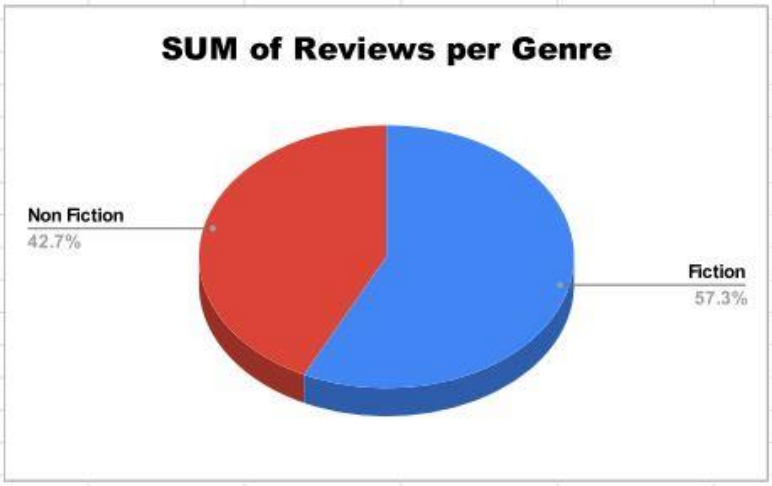
Findings 3: Sum of Users ratings per Year (Pie Chart)

Year	SUM of User Rating
2009	229.2
2010	227.9
2011	227.9
2012	226.6
2013	227.7
2014	231.1
2015	232.4
2016	233.9
2017	233
2018	233.4
2019	237
Grand Total	**2540.1**



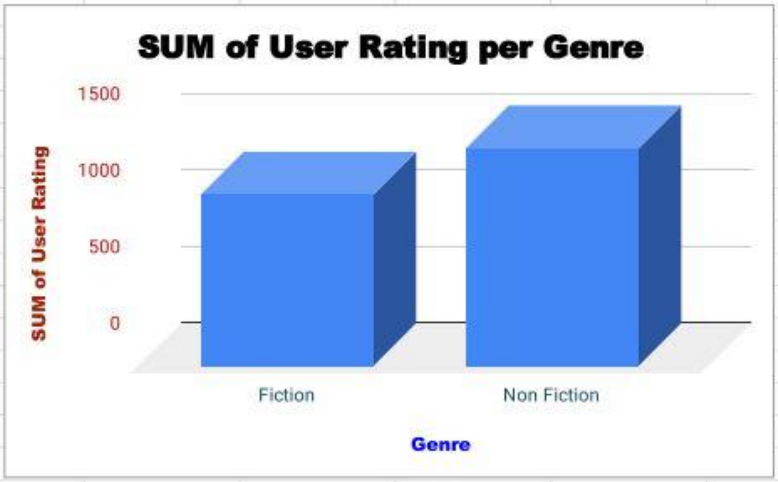
Findings 4: Sum of Reviews per Genre (Pie Chart)

Genre	SUM of Reviews
Fiction	3764110
Non Fiction	2810195



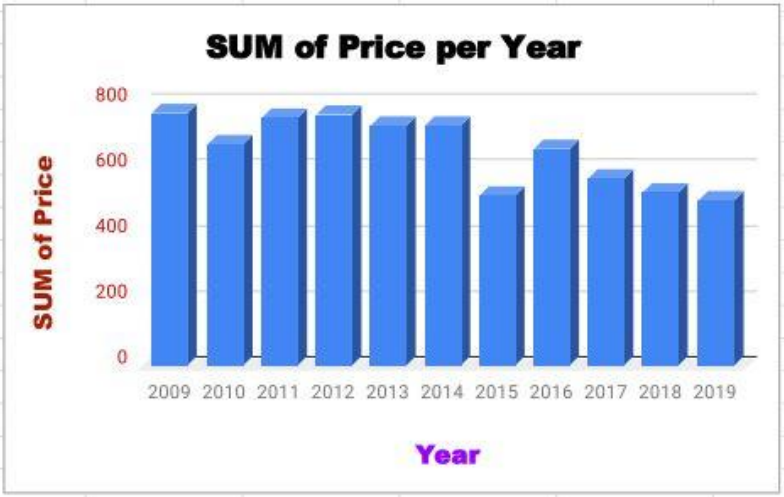
<u>Findings 5</u>: Users Ratings per Genre (Column Chart)

Genre	SUM of User Rating
Fiction	1115.6
Non Fiction	1424.5
Grand Total	**2540.1**



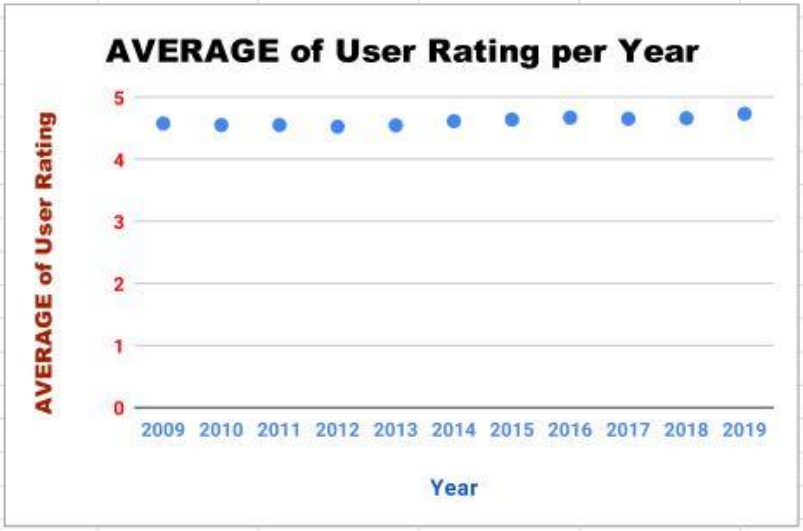
<u>Findings 6:</u> Sum of Price per Year (Column Chart)

Year	SUM of Price
2009	770
2010	674
2011	755
2012	765
2013	730
2014	732
2015	521
2016	659
2017	569
2018	526
2019	504



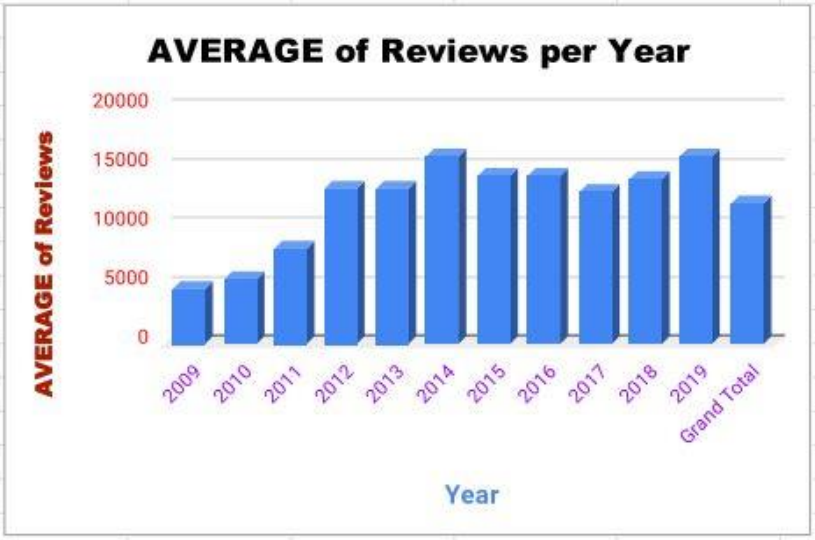
Findings 7: Average Users Ratings per Year (Scatter Chart)

Year	AVERAGE of User Rating
2009	4.584
2010	4.558
2011	4.558
2012	4.532
2013	4.554
2014	4.622
2015	4.648
2016	4.678
2017	4.66
2018	4.668
2019	4.74



<u>Findings 8:</u> Average Reviews per Year (Column Chart)

Year	AVERAGE of Reviews
2009	4710.12
2010	5479.62
2011	8100.82
2012	13090.92
2013	13098.14
2014	15859.94
2015	14233.38
2016	14196
2017	12888.4
2018	13930.42
2019	15898.34
Grand Total	**11953.28182**



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 previews.
- 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.