**Amazon Top 50 Bestselling Books**

**Project Description:**

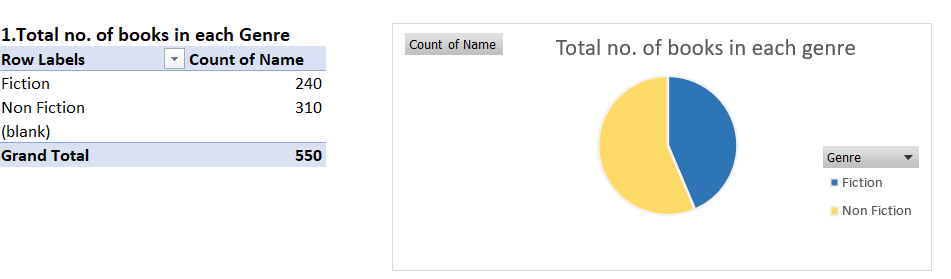
The Dataset contains Amazon's Top 50 bestselling books from 2009 to 2019 and contains 550 books, the data has been categorized into fiction and non-fiction using Goodreads. Based on the categories find the no. of books that are fiction and non-fiction and for how many years that book was under top sold books.

To have a better knowledge on which category book is sold more, the top bestselling books from 2009 to 2019 and also the authors with more books of different genre will be analyzed.

**Design:**

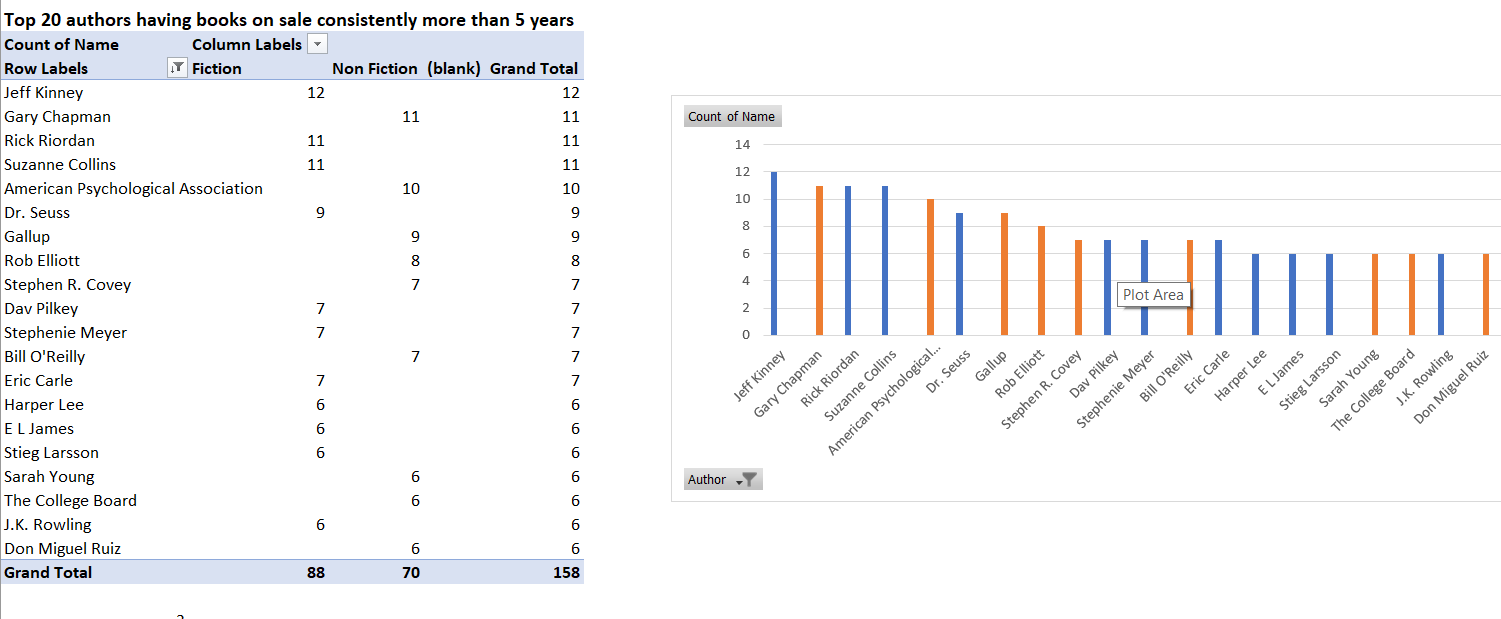
* The dataset is cleaned using excel, checked for duplicates and null values.
* The books that were sold from 2009 to 2019 consistently were found using countif function.
* The VLOOKUP table used to get the unique books that were published more than one year.
* Pivot table used to find the total no. of books in each genre, the top 20 authors published books for more than 5 years, books that were published for more than 5 years categorized based on genre.
* The analysis was performed using even SQL and the visualization was carried out using Tabelau.

**Findings**



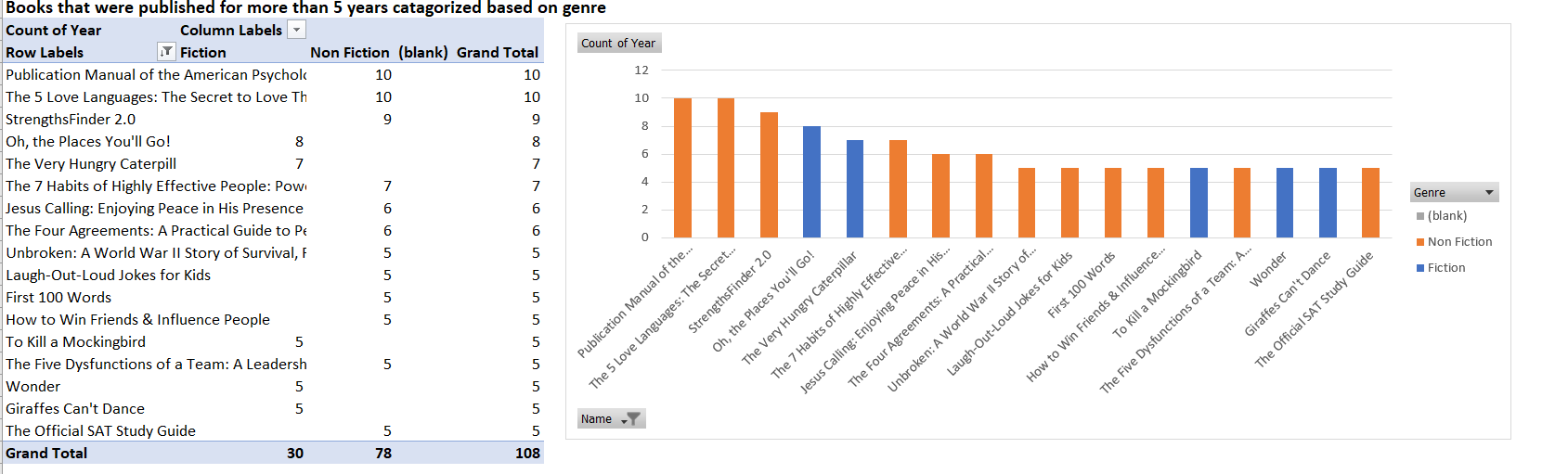
**Figure 1. Pie chart represents Total no. of books in each Genre**

The pivot table used to get the total no. of books in each genre. The pie chart shows that books from non-fiction genre is more compared to fiction.



**Figure 2. Bar graph representing top 20 authors published books in 2009 to 2019**

Top 20 authors with books were on top for more than 5 years according to the genre is shown through bar graph. From the top 20 the fiction category book is having highest number. But when considered overall the non-fiction books are on top.



**Figure 3. Bar graphs Represents books that were published more than 5 years**

From the above bar graph the books that were published consistently for more than 5 years based on genre is shown. The non-fiction books are at the peak compare to fiction.

**SQL query :**

Use PortfolioProject

Select \*

From bestsellers$

--best selling books from 2009-2019

With bestCTE As

(

Select \*,Rank() over (partition by name order by name) as rn

From bestsellers$

)

Select top 10 name, count(rn) as total

From bestCTE

Group by rn, name

Order by total desc

-- Total no of fiction and non-fiction books

Select Genre, count(\*) as counter

From bestsellers$

Group by Genre

Order by counter Desc

--Find the no. of best selling books in each genre every year

Select Genre, Year, count(\*) as counter

From bestsellers$

Group by Genre, Year

Order by Year Desc

--Total no. of books published by each author

Select Author, count(Name) as Total

From bestsellers$

Group by Author

Order by Total desc

--books with highest reviews

Select Top 10 Name,Genre, Reviews

From bestsellers$

Group by Name, Reviews, Genre

Order by Reviews Desc

--Books with highest rating

Select Top 10 Name,Genre, Rating

From bestsellers$

Group by Name, Rating, Genre

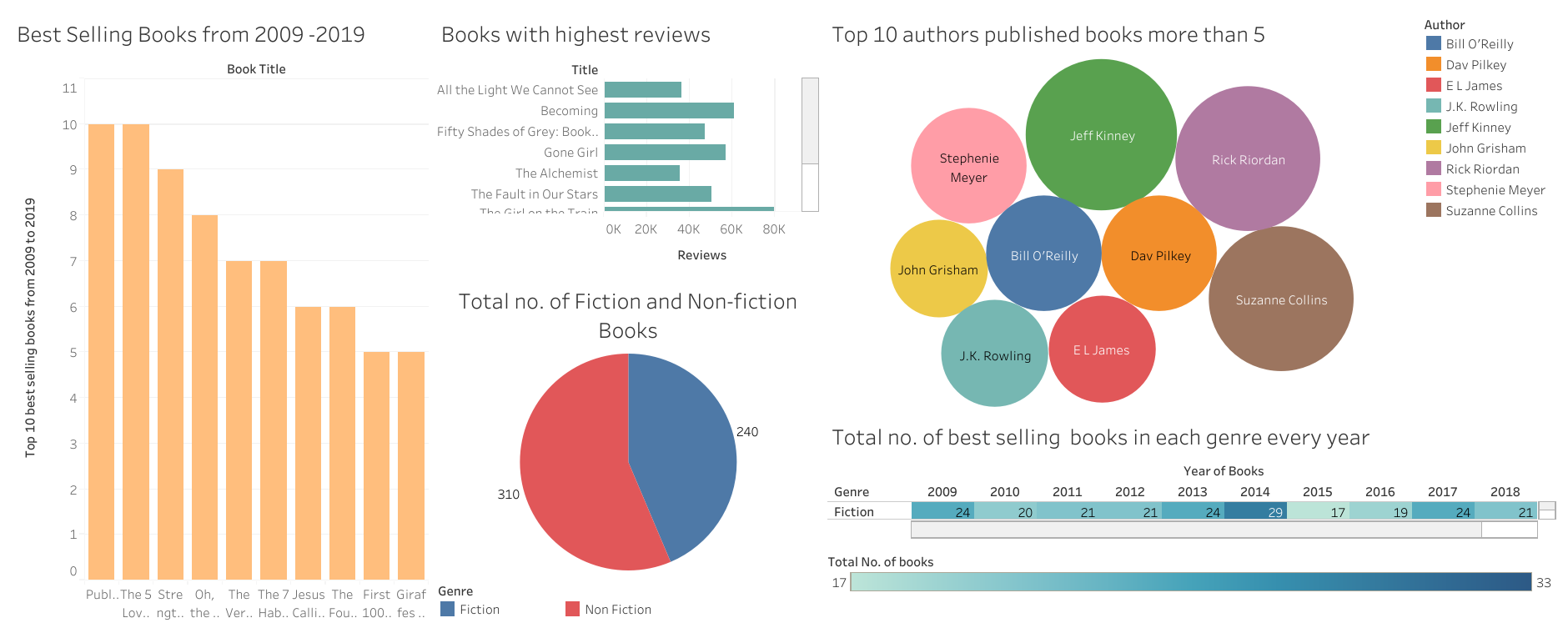
Order by Rating Desc

--Average price of books

Select avg(price) as average

From bestsellers$

SQL query is used to find bestselling books from 2009 to 2019, total no. of fiction and non-fiction books, found no. of bestselling books in each genre every year. Total no. of books published by each author is found using query and also the books with the highest reviews.



**Figure 4. Dashboard with all the visualizations for the analysis carried out**

**Analysis**

The dataset is of Amazon's Top 50 bestselling books from 2009 to 2019 and contains 550 books of two different genres. The first question that would arise is what type of book is more sold. The next question would be why the books are in top 50 and also the ratings and reviews of each book would be a next question. When analysed the non-fiction count was more than fiction and it was sold consistently from 2009 to 2019.

The ratings showed that 4.9 is being highest and the reviews of the books in general belongs to fiction genre, but when considered the bestselling books from 2009 to 2019 the non-fiction book was on top. The bestselling book was Publication Manual of the American Psychological Association, 6th Edition. And the author with more publication was Jiff Kenny.

**Conclusion**

From the analysis and visualization, it’s clearly seen that non-fiction and fiction is almost having the same race but the non-fiction slightly higher compared to fiction genre. And it’s consistently being sold for 10 years, people are more interested in non-fiction genre.

So the next action that can be implemented by giving discounts during festive seasons and make available of combo purchase with the combination of both the gen