

## Top 50 Best Selling Novels on Amazon (2009-2020)

**Aim:** The aim of the project is to identify 3 key parts which leads a novel to be in the top 50 bestselling novels on Amazon.

1. Which genre is widely popular and consistent in the top 50 over the years?
2. Does price really matter for a novel to be bestselling?
3. Do ratings and review matter?

### **Data:**

The dataset consists of columns like:

- Genre
- Year
- Reviews
- Price
- User Rating
- Author
- Name

These columns would help answer our 3 questions.

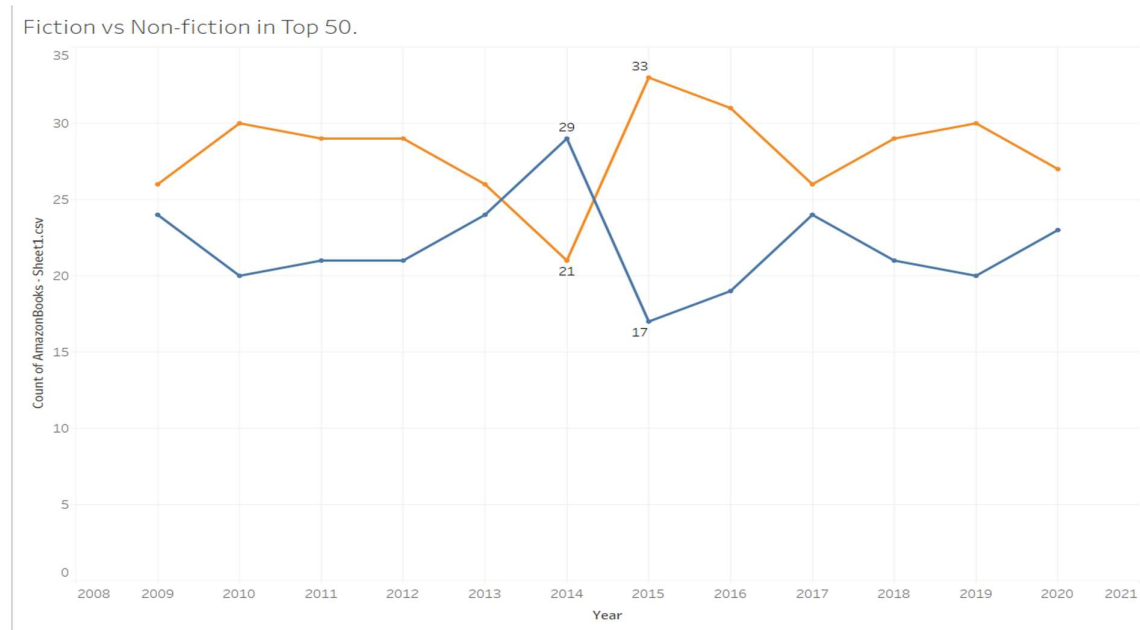
Using Tableau, we will be visualizing various plots and, by analyzing these visualizations we will be concluding whether or not ratings, reviews, price play a key role in determining the top 50 selection of the novels.

Let's have a look at these phase by phase.

**Phase 1: Determining which Genre has been consistent and widely popular over the years.**

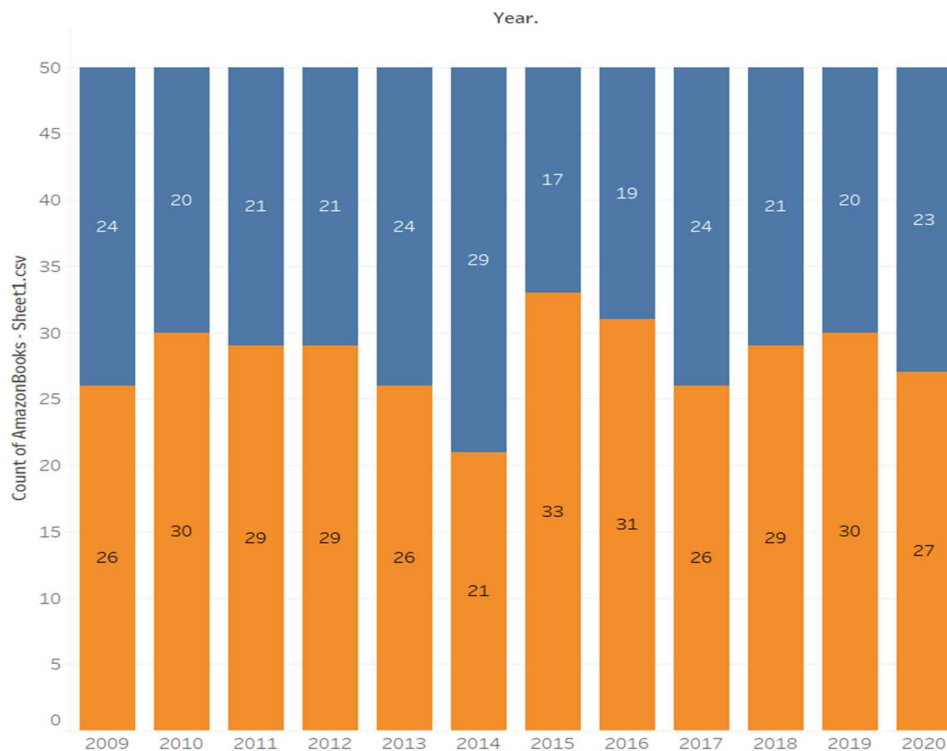
For this we would need to plot Genre against Years for the total number of books in our dataset.

**Visualization 1:**



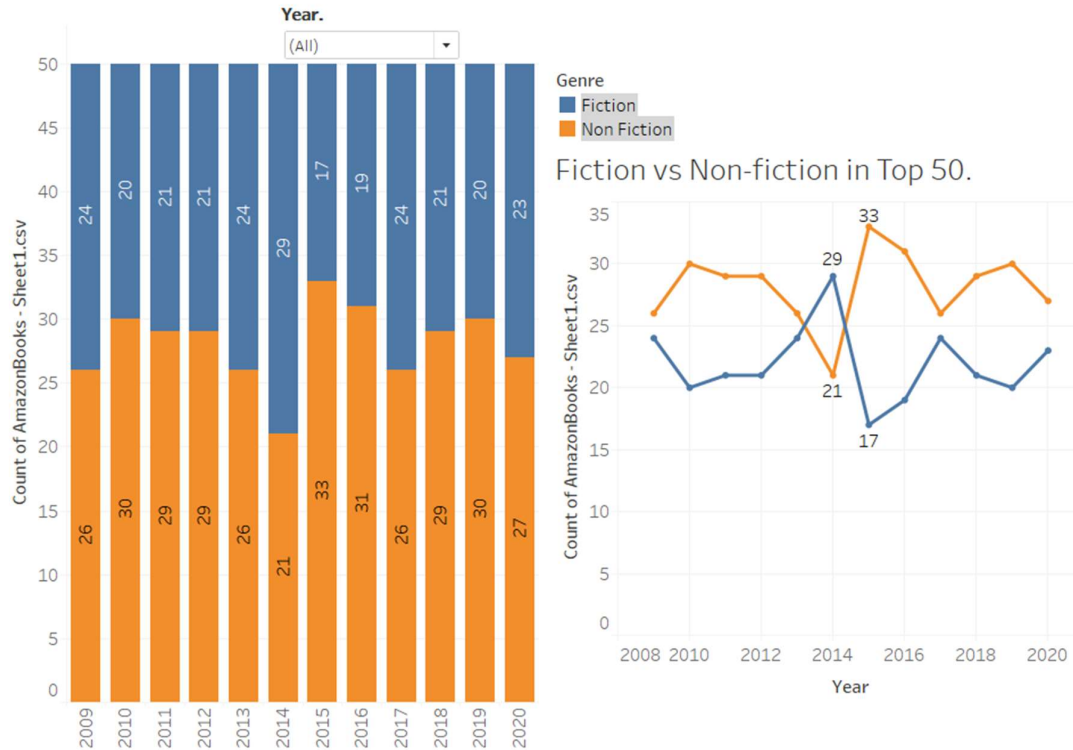
**Visualization 2:**

Share of Books by Genre for each year.



## Summary:

Share of Books by Genre for each year.



From both these graphs we can observe that the Non-Fictional Genre has been greater than the Fictional for all the years except for 2014. So, we can conclude that the Non-Fictional genre is much more popular and well-received in comparison to Fictional genre.

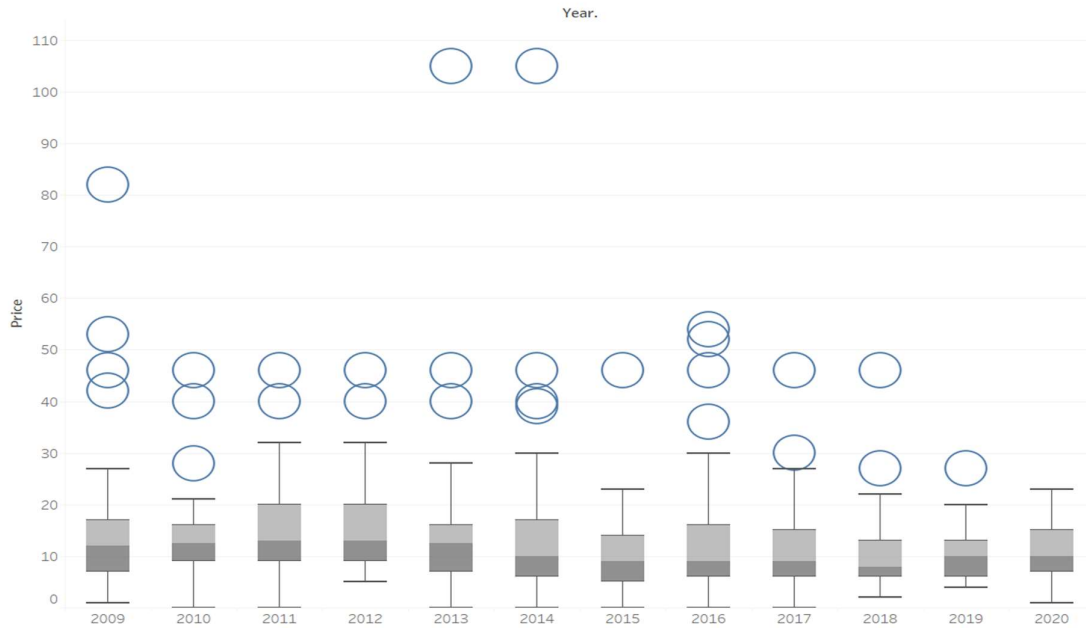
Now that we have determined which genre is more popular and consistent in the top 50, let us find out if prices, reviews and ratings play a key role in determining the top 50 novels.

## Phase 2: Price's role in determining the top 50.

To determine if Price plays a key role in shortlisting the top50, we can plot visualizations by using the average Price on Year both as a Continuous variable. This would provide us with a good line/area graph to depict how the trend has been over the years.

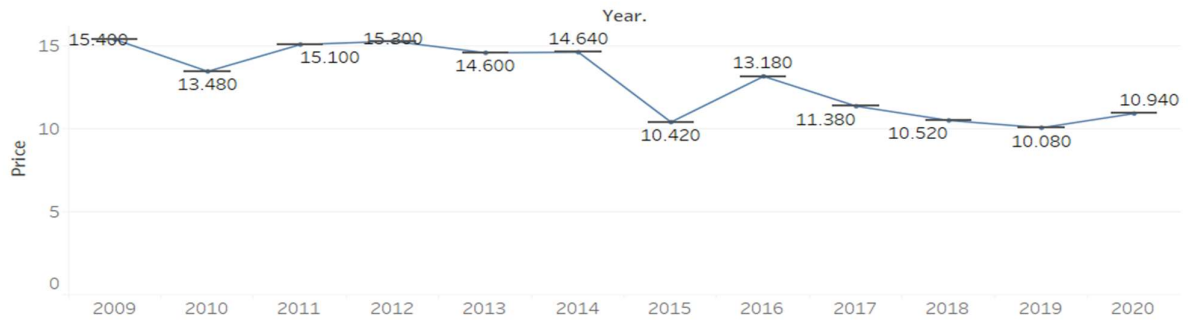
### Visualization 1:

Avg. Price by Boxplot

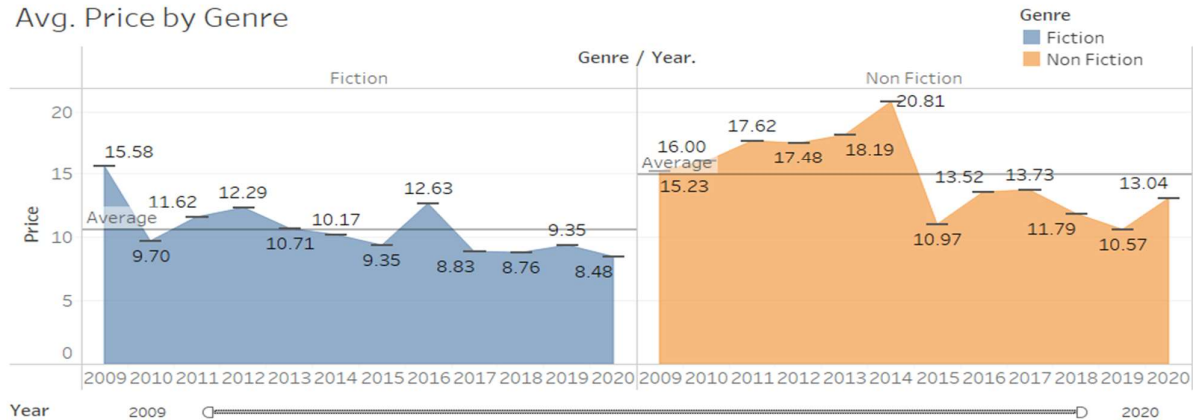


### Summary:

Average price of top50 novels over the years.



Avg. Price by Genre



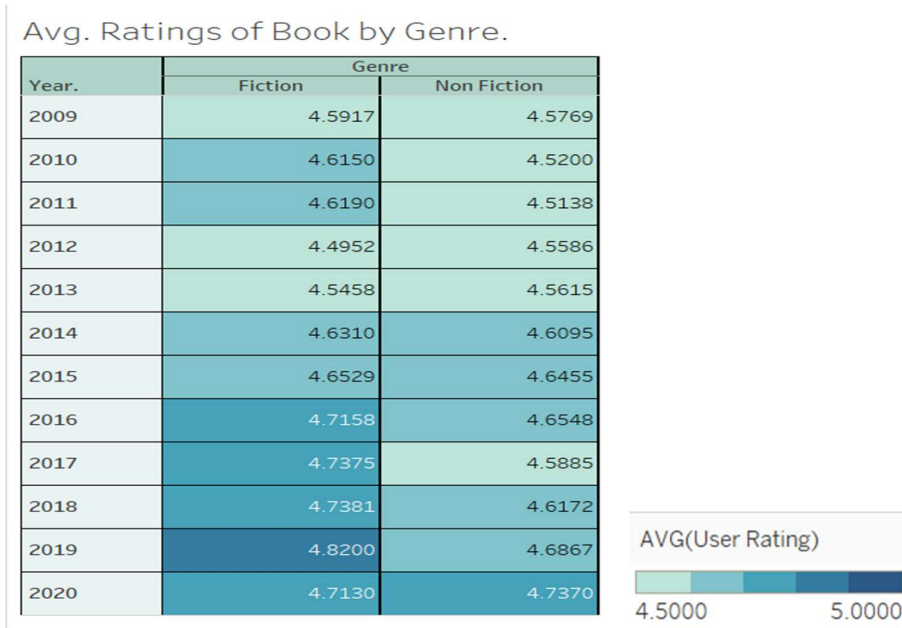
As we can see, the average price has generally been between \$10 - \$15 over the years. We can say that a lower priced novel is preferred over a novel with high price. So, low price becomes a selling point for a novel. Also, if we look at the graph carefully, in the initial years from 2009- 2014 the average price was around \$14 while after 2014 the average price drops down to about \$10-\$11. This indicates that in the recent years low priced novels are preferred over high-priced ones. So, if the novel falls between the average price there are chances of that novel to be in the top 50.

If we consider the average price by genre, we can see that there is not much variation between the price range. For fictional, the average price is about \$10-\$10.5 while for non-fictional it is about \$15.

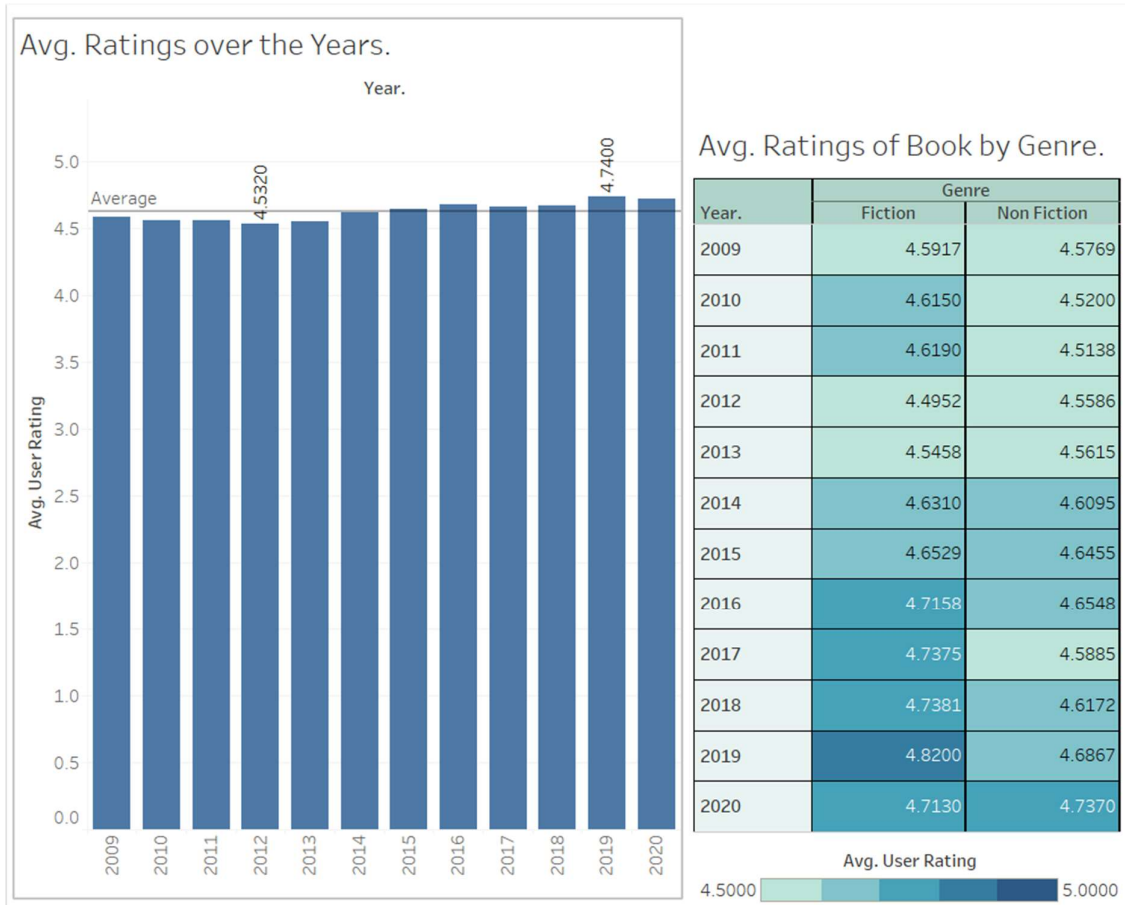
### Phase 3: What are the average ratings of top 50 novels?

We need to find out what the average ratings have been for the past years. To visualize this, we need to plot User ratings as a continuous variable on Year as a discrete variable. This would give us a simple bar graph to identify the average ratings over the years.

#### Visualization 1:



#### Summary:



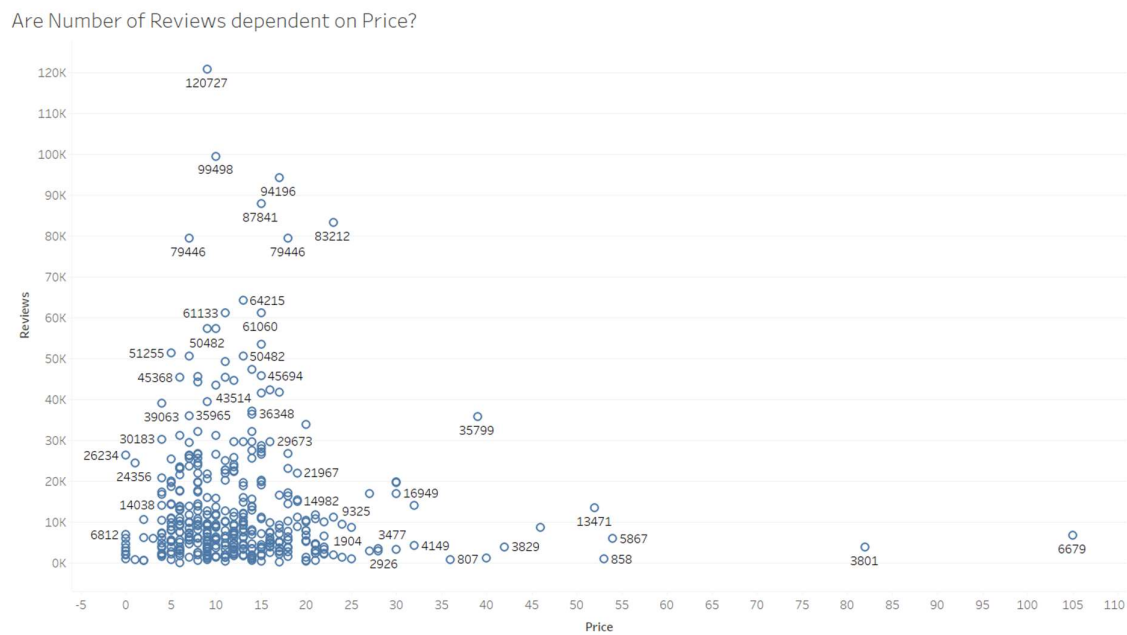
We can see that the average ratings for a novel have always been above 4.5 regardless of the Genre or Year. The average ratings for all the years combined is about 4.6. We can say good ratings are a key component for a novel to be shortlisted within the top 50.

Now to have a good rating, the novel should ideally have many reviews as well. So to identify the overall components dependency let us have a look at the reviews.

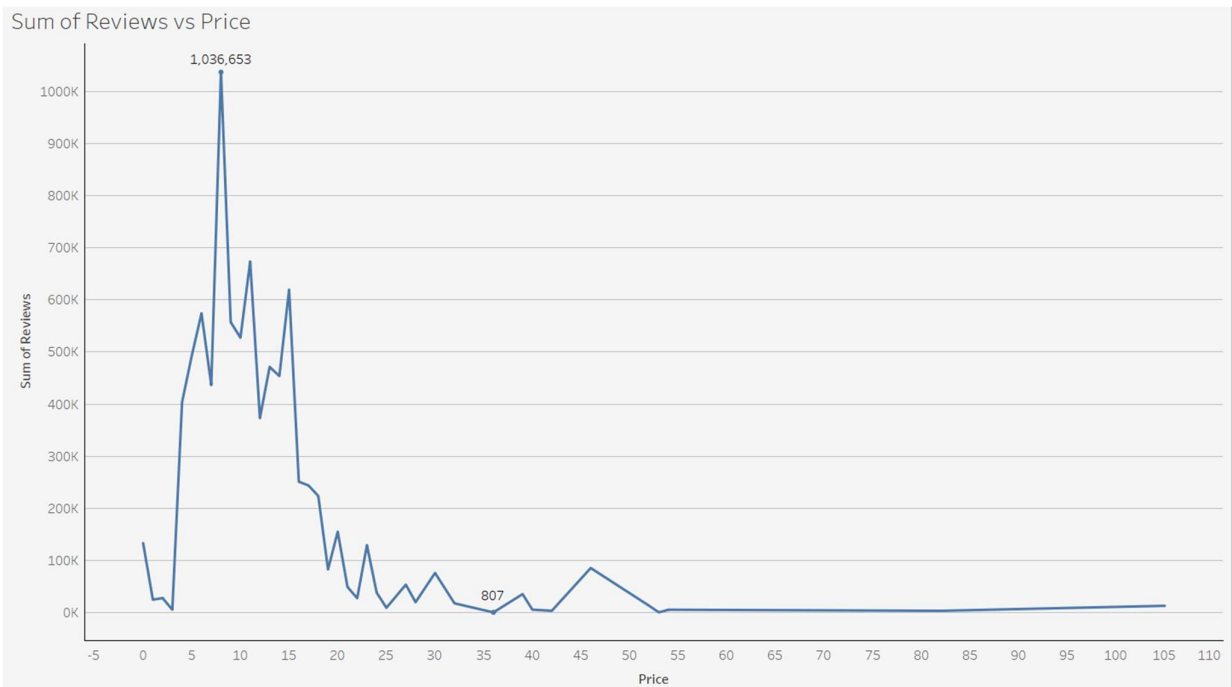
Phase 3: Effect of Reviews

Number of reviews can only be increased if a certain book is bought by multiple users. As we saw previously that price plays a key role and the avg. price over the years has been around \$10-\$15. Let us see first see if price has any correlation with the no. of reviews.

Visualization 1:

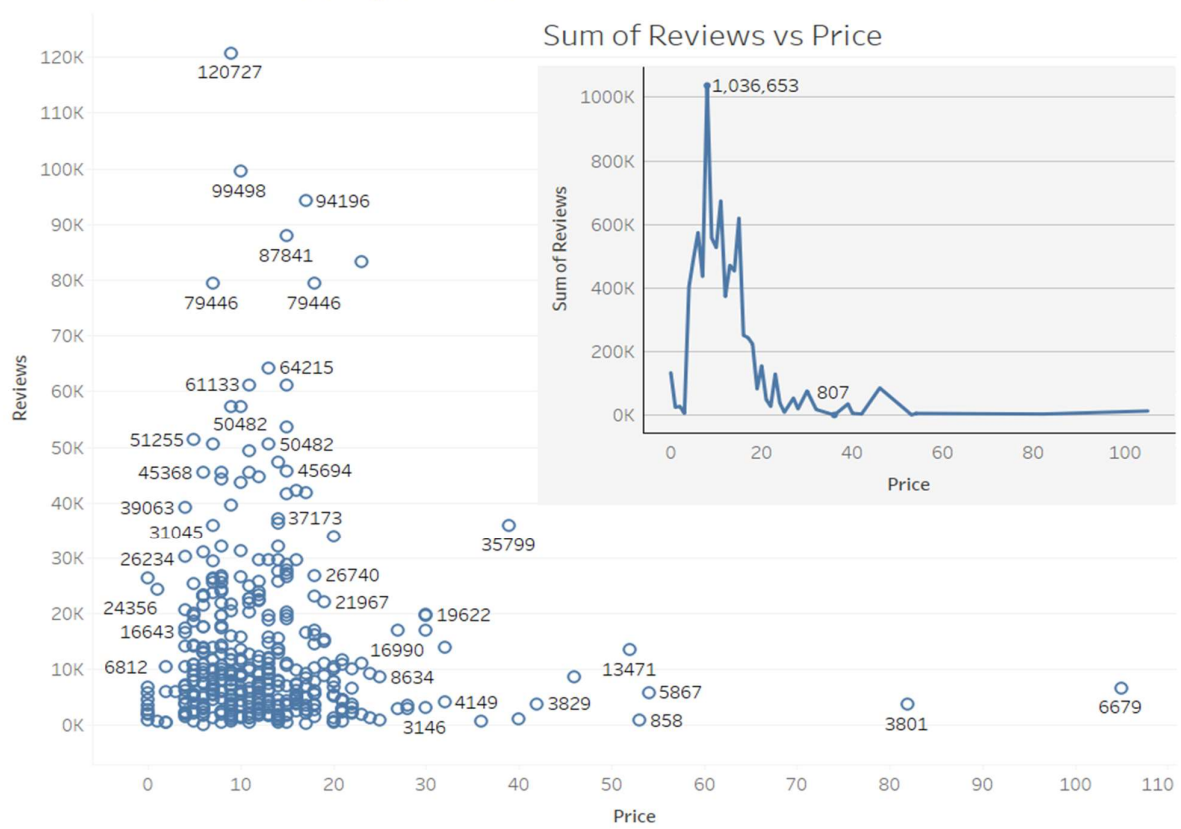


Visualization 2:





Are Number of Reviews dependent on Price?

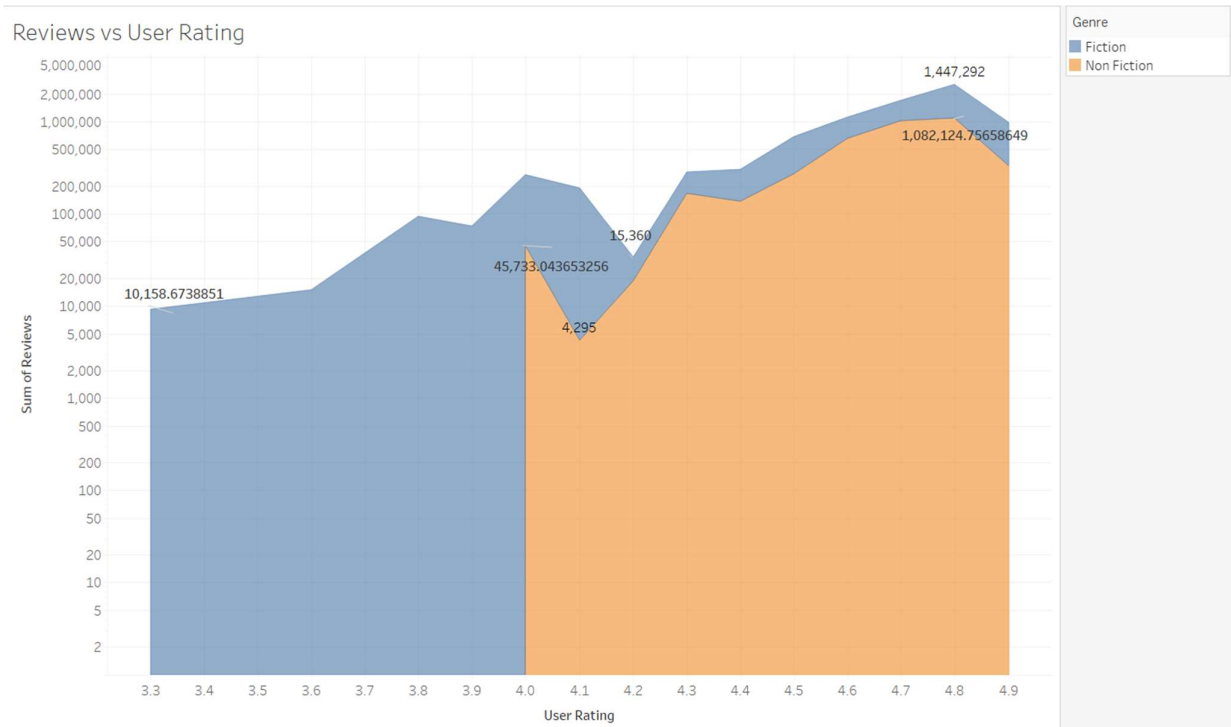
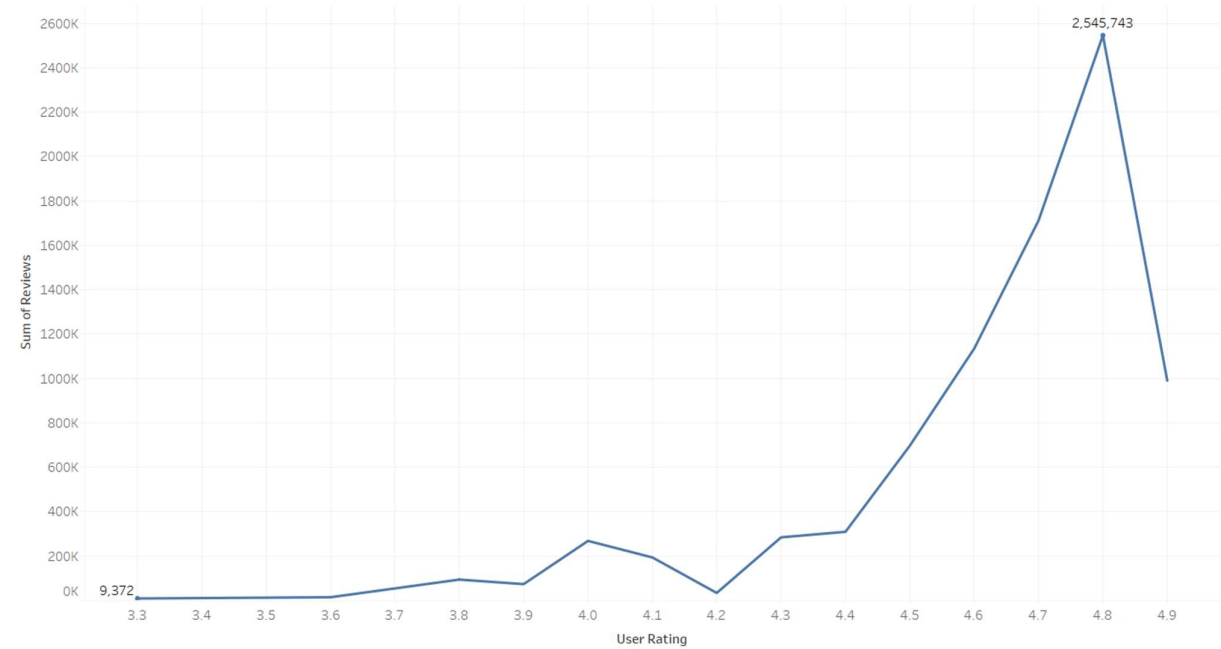


We can see that a high amount of review's concentration lies within the \$5-\$20 range which indicates that price does have a correlation with the number of reviews received for a novel.

Now, let's see if a similar correlation exists for user-ratings vs reviews. To identify this, we can sum up the total no. of reviews and plot that sum against the user ratings and see what kind of graph we get.

**Visualizations:**

Reviews vs User Ratings.

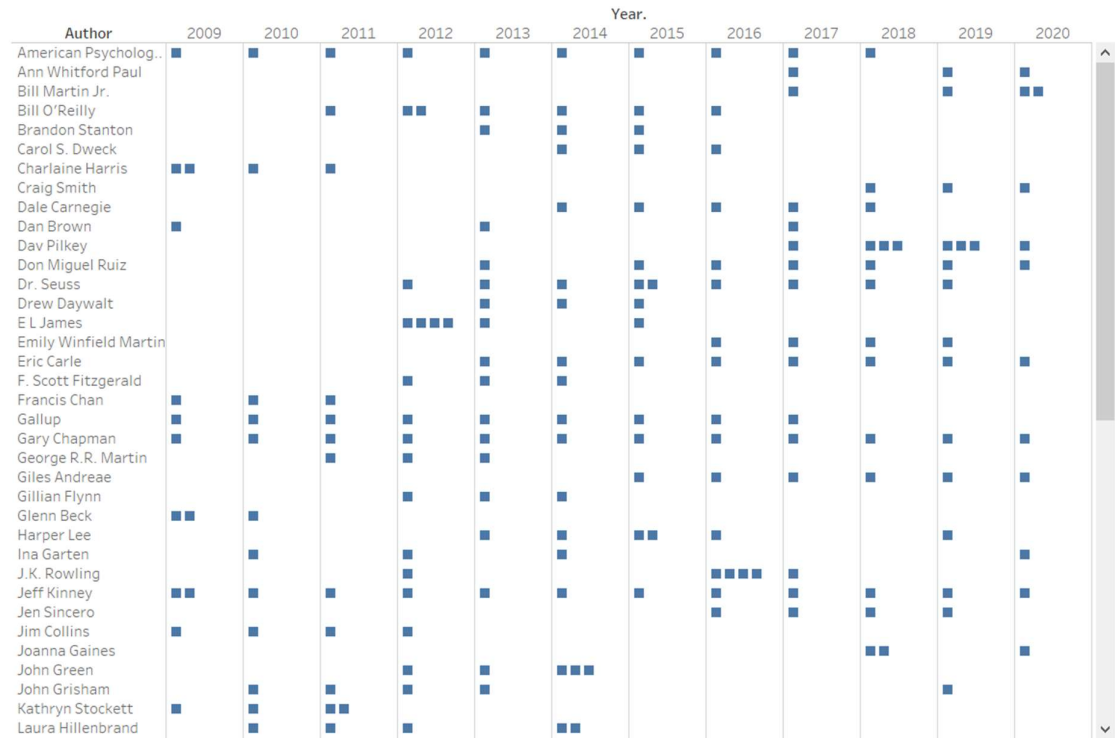


As we can see, as the sum of reviews increases the ratings start increasing. We have a significant correlation between the no. of reviews received for a novel and the ratings that are given to a novel.

Now, that we have concluded that Price, Ratings and Reviews matter in the shortlisting of Top 50 category, let us have a look at visualizations that provide insights about consistency of author's and books in the Top 50 bestselling novels.

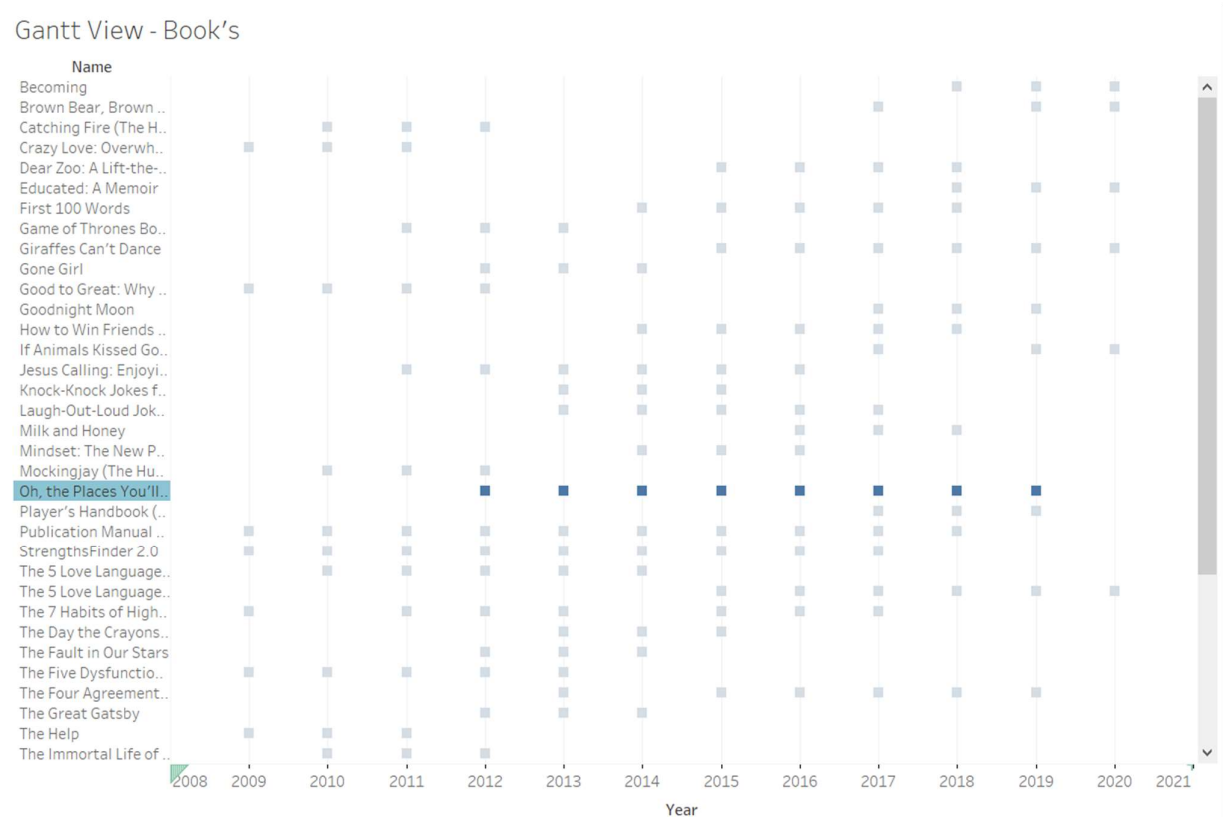
**Visualization:** Author's whose books have been consistent in the top 50 at least 3 years.

Author's Books in Top50.



In this Gantt-view, we can see the list of authors who have featured consistently for more than 3 years with either same or various novels. Multiple blocks within the same year indicates that more than one novel of the author was featured in the top 50 for the year.

**Visualization:** Books that have been consistent in the top 50 bestselling for at least 3 years.



In this Gantt view, we can see the Novel's that have been consistently in the top 50 for at least 3 years.