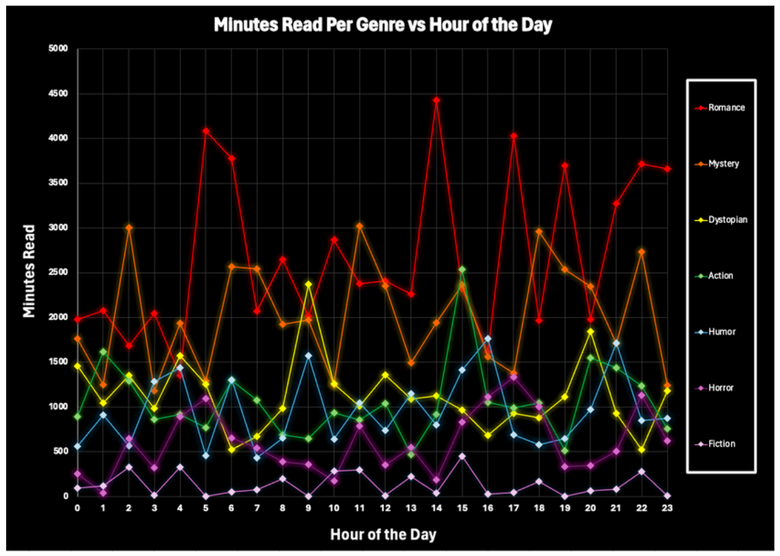
Kyle Krebs
Ansley Orrell
Rylie Platt
Tristan Miller
Daniel Oleynikov

# The Story Safari Data Analysis

R·I·T

**Figure. 6:** Connected scatter plot representing how many minutes read each genre accounts for through each hour of the day. It should be noted that this graph only contains data from authors that have at least 5 ratings.




**Fitted Normal Distribution**
**Goodness-of-Fit Test**

| | W | Prob<W |
|---|---|---|
| Shapiro-Wilk | 0.9678296 | 0.6138 |

| | A2 | Simulated p-Value |
|---|---|---|
| Anderson-Darling | 0.2775998 | 0.6436 |

**Figure. 8:** Goodness-of-Fit test outcomes that provide insight into the likelihood of Figure. 7 being a normal distribution.

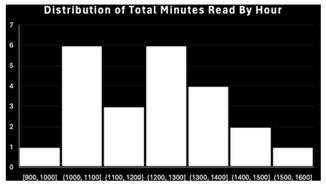



**Figure. 7:** Distribution of total minutes read per hour separated by 7 bins each having a range of 1000 minutes.

This connected scatter plot, based on data from four tables—client, genre, read, and book—illustrates how users engage with different book genres at various times of the day. It maps the relationship between the hour of the day and total minutes read per genre, with each hour representing a full 60-minute interval starting at the hour. The analysis reveals that romance books dominate reading across all hours, suggesting a potentially larger female audience given the genre's popularity among women. Conversely, fiction sees the least engagement, indicating a possible shift in future content strategy to align more with audience preferences.

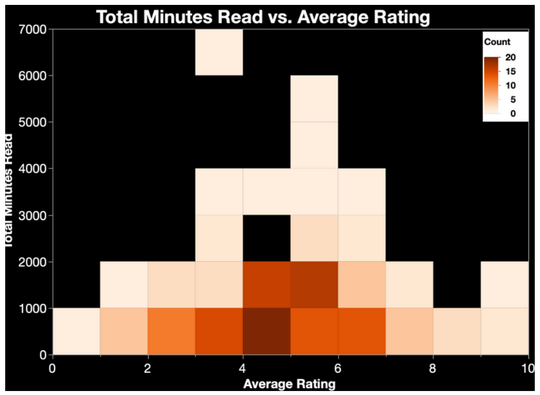



**Figure. 1:** A heat map representation of authors' total minutes read vs their average ratings. All data is collected directly from user ratings and users read time.

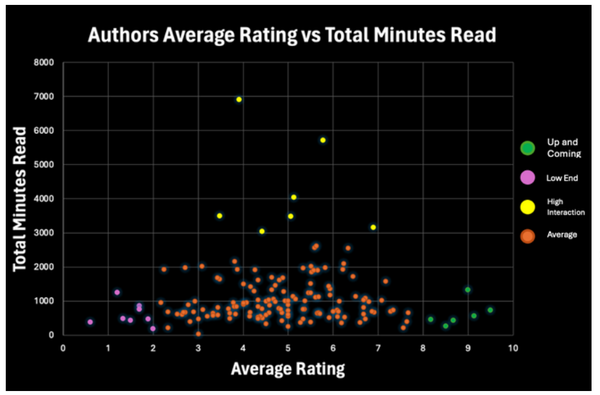



**Figure. 2:** A scatter plot representation of authors' total minutes read vs their average ratings. Classifications of four different groups were created.

Figures 1 and 2 display data drawn from three tables—writes, read, and rate—illustrating the relationship between an author's total minutes read and their average rating. The data categorizes authors into four clusters. The "Average" cluster comprises authors whose popularity aligns with the majority in our database. The "High Interaction" cluster includes authors who have amassed a significant amount of reading time but maintain middle-of-the-road ratings. The "Up and Coming" cluster represents authors with limited total reading time but exceptionally high ratings. Lastly, the "Low End" cluster consists of authors who struggle with both low ratings and minimal reading time.

| Author ID | Author Name | Total Minutes | Average Rating |
|---|---|---|---|
| 250 | Sandi Jones | 456 | 8.17 |
| 189 | Mark Millar | 264 | 8.5 |
| 205 | Clive Cussler and Boyd Morrison | 432 | 8.67 |
| 246 | Naomi Alderman | 1328 | 9 |
| 120 | Ernest Cline | 558 | 9.13 |
| 143 | Lois Lowry | 728 | 9.5 |

**Figure. 5:** List of data points from Figure. 2 that correspond with the “Up and Coming” cluster.

**Figure. 4:** List of data points from Figure. 2 that correspond with the “High Interaction” cluster.

| Author ID | Author Name | Total Minutes | Average Rating |
|---|---|---|---|
| 136 | Neil Gaiman | 3492.5 | 3.48 |
| 159 | Robert Jordan | 3040 | 4.42 |
| 279 | Herman Melville | 3476 | 5.06 |
| 134 | Samantha Shannon | 6900.5 | 3.91 |
| 123 | Cormac McCarthy | 4041 | 5.13 |
| 133 | Erin Morgenstern | 5711.5 | 5.78 |
| 1 | Suzanne Collins | 3152 | 6.89 |

| Author ID | Author Name | Total Minutes | Average Rating |
|---|---|---|---|
| 215 | Delia Owens | 384 | 0.6 |
| 131 | Orson Scott Card | 486 | 1.33 |
| 177 | Herman Melville | 1248 | 1.2 |
| 147 | M. Scott Peck | 474 | 1.89 |
| 172 | Cormac McCarthy | 860.5 | 1.69 |
| 200 | Robert Cinnamon | 759.5 | 1.69 |
| 222 | Heather Morris | 432 | 1.5 |

**Figure. 3:** List of data points from Figure. 2 that correspond with the “Low End” cluster.

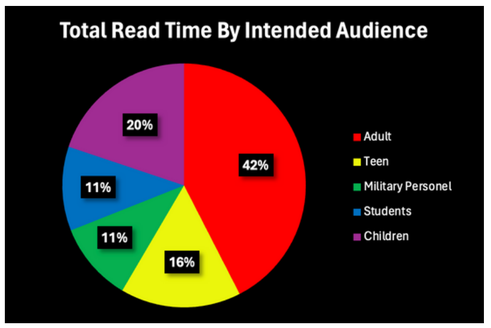



**Figure. 9:** Pie chart illustrating the distribution of our database's total read time across our five main intended audiences of books.

This pie chart, sourced from three tables—book, audience, and read—shows the distribution of total read time across different audience segments in our book database. Adults lead with 42%, followed closely by Children and Teens, each at 20% and 16% respectively. Military Personnel and Students each hold 11%. This data suggests two strategic paths: we could concentrate on catering to adults, our largest audience segment, or diversify our offerings to broaden our appeal across different audience groups.