**David Ramos**

**Tableau Project**

**Data Analytics**

**10/23/2023**

**Goodreads Book Reviews**

**Project Purpose:**

Goodreads is an American social cataloging website and subsidiary of Amazon.com that allows enthusiast readers to search its database of books, annotations, quotes and reviews. Users of this website and application can sign up and register books to produce library catalogs of different books and reading lists. The website launched in 2007.

The purpose of this project is to create a data visualization dashboard within Tableau that will showcase data on various books and detailed information about each book from the Goodreads API (Application Programming Interface).

I retrieved this Goodreads dataset from Kaggle.com which presents a comprehensive list of books listed on Goodreads. The dataset used in this project is “clean”.

I organized the details of the Goodreads data set into categories to better understand and visualize the data within this project. The categories are as follows:

* Authors by Total Books
* Publishers by Total Books
* Titles by Rating Count
* Books by Publication Date
* Top 10 Language Codes
* Authors
* Language Code
* Publication Date
* Publisher
* Ratings Count
* Number of Pages
* Average Rating
* Text Reviews

**Project Questions:**

* **Which authors wrote the most books?**

Stephen King, P.G. Wodehouse and Rumiko Takahashi are top three most active authors in the Goodreads books dataset with a total of 40 and 39 books published between them.

* **What books have the highest number of ratings?**

The top three books have massive amounts of ratings per book. *The Hobbit or There and Back Again* by J.R.R Tolkien and Alan Lee had a total rating count of 2,530,894, *The Catcher in the Rye* by J.D. Salinger achieved a substantial rating count of 2,457,092 and *Animal Farm* by George Orwellacquired a total rating count of 2,111,750.

* **What are the average ratings of the most popular books?**

The most popular books in the Goodreads dataset have differing average ratings according to the data. The average rating of *Harry Potter and the Order of the Phoenix* by J.K. Rowling is 4. *The Da Vinci Code* by Dan Brown has an average rating of 15, *The Alchemist* by Paulo Coelho has an average rating of 8 and *Charlotte’s Web* by E.B. White has an average rating of 21.

* **Which publishers published the most books?**

The top 6 publishers of the Goodreads books in the dataset are:

1. Vintage
2. Penguin Books
3. Penguin Classics
4. Ballantine Books
5. Mariner Books
6. HarperCollins

* **When were the most books published?**

The most books were published in 2006 with a total amount of 1,700 books.

* **What language codes were most popular amongst all the books in the dataset?**

The language codes are used within this dataset as an abbreviation for the language that each book was written and published in. The top 6 language codes are and the number of books in those codes are as follows:

1. Language Code- eng, # of Books- 8908
2. Language Code- en-US, # of Books- 1408
3. Language Code- spa, # of Books- 218
4. Language Code- en-GB, # of Books- 214
5. Language Code- fre, # of Books- 144
6. Language Code- ger, # of Books- 99