PROJECT PROPOSAL

ABSTRACT

The goal of this project is to do basic EDA and visualization , to do Analyzing and Recommending Books , 'prediction of book rating' and use regression to predict that the most entertaining thing you do in your free time.

other :
 -How many reviews on books
- Most author published book
- Most language written by
- Most rated books
- Highly rated author

DESIGN

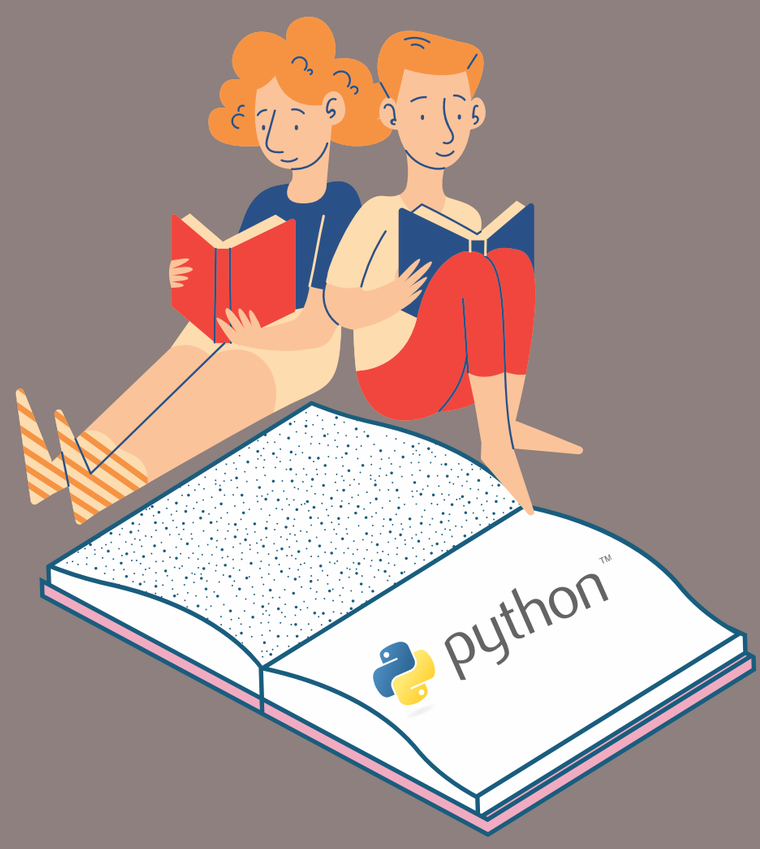
based on the data from Goodreads-books
https://www.kaggle.com/jealousleopard/goodreadsbooks
I well answer to these questions
- How many reviews on books ?
- Who is the most author that published books ?
- What is the most language written by ?
- What is the most rated books ?
-Who is the Highly rated author ?

DATA

To achieve the goal of this study the dataset **Goodreads-books** will be used. This dataset can be found at
https://www.kaggle.com/jealousleopard/goodreadsbooks

the dataset contains information about books, it contains 11123 rows , 12 column (feature)
 Let's take a look :

- bookID - Contains the unique ID for each book/series
- title - contains the titles of the books
- authors
- average_rating
- ISBN - the International Standard Book Number.



- language_code - Helps understand what is the primary language of the book
- Num_pages
- Ratings_count - Total number of ratings the book received.
- text_reviews_count - Total number of written text reviews the book received

TOOLS

numpy & pandas for data manipulation (EDA)
matplotlib & seaborn for plotting
Scikit-learn for modeling

ALGORITHM

sklearn : Linear Regression & may be logistic Regression

MVP

the goal of this project is to predict book rating by using regression to predict that