Minor Project #1:

Fundamentals of Data Analytics.(EDA)

Thorough analysis of accurate, complete data serves as the foundation of your data story. Analyzing data using descriptive, diagnostic, predictive, and prescriptive analysis can enable you to understand its full picture.

Before you start working with any dataset, it is important to explore the data using descriptive statistics. Descriptive statistics enable you to compare various measures across the different variables. These include mean, mode, standard deviation, etc. There are many kinds of graphical summary methods such as histograms, bar graphs, pie charts etc.

For this part of the project, you will be analyzing some data points listing some details about the books irrespective of the language and publication and all of that . You will analyze the data and familiarize yourself with it, and then answer some of the questions that are listed below. Basically using this dataset we are trying to analyze and recommend good books to read to readers of all types.

So as you work your way through the project instructions, keep the following questions in mind. Jotting down some notes related to each of them will help you to synthesize your findings into your project report.

What you need to show in your project:

Discuss Overall Findings and Answer the Questions given below.

1. What is the most popular book?
2. Are books with fewer pages rated higher than those with large page counts?
3. What is the most popular book of the 60s?
4. Who wrote the most pages?
5. What's an author's average page count?

6. How many books have been written with less than 200 pages?
7. What is Houghton Mifflin Harcourt's most popular book?
8. Display the most popular book written by each author.
9. What is the least popular book of the 90s?
10. What is the highest-rated book with over 500 pages?
11. Create graphs to better understand the existence of any correlations in the dataset.
12. If you can think of more questions that could be relevant to this analysis you can list and answer them too.(efforts appreciated)

Descriptive Statistics

1. What are typical values in this dataset?
2. How widely do values in the dataset vary?(st.dv.)
3. Are there any unusually high or low values in this dataset? (outliers)
4. What is the size of the dataset?
5. Is any observation repeating more frequently than others?
6. What is the central value of different columns?
7. What is the most popular book?
8. Create some storyline and relevant graphs to present your findings from the analysis that you have performed.

Use tools like *pandas* and *matplotlib* to complete the above tasks, and create a compelling data story(verbal + visual) around it.

Link to the Data Set:

https://www.kaggle.com/datasets/jealousleopard/goodreadsbooks?select=books.csv