The project is completed in Python using Jupyter Notebook.

Web scraping is achieved using Beautiful Soup library in python. Amazon website data is used for data.

**Beautiful soup:**

Beautiful soup is a python library used for extracting data from website. It parses data from HTML and XML files. It saves programmers hours of work using lxml and html5lib.

**Steps :**

1. Install the beautiful soup library.
2. Import the necessary libraries like pandas, NumPy, matplotlib etc. to work with data.
3. Scripting the Amazon Best Selling Books from the below URL.

<https://www.amazon.in/gp/bestsellers/books/ref=zg_bs_pg_'+str(pageNo)+'?ie=UTF8&pg='+str(pageNo)>

1. Connect to the URL and fetch the HTML contents.
2. Book name, Author, Rating, Customer Rated and Price are retrieved.
3. Save the output list to a data frame.
4. Save the data frame as a csv file.
5. Import the csv file to interpret the data.
6. There are 100 rows and 5 columns in the csv.
7. Preprocess the data by removing excess columns, replacing zero values and removing the null values.
8. Authors highest priced book is retrieved.
9. Finally, the most customer rated Author and Boks are plotted.
10. The plot shows scatter plot of authors who achieved highest customer ratings vs the actual rating.

The project is based on web scraping using beautiful soup and how we can sense out information extracted by visualizing it using bokeh plotting library.