## Domain Insight:

E-commerce is filled with user based reviews and recommendation based on the same. The data given for the Amazon’s books review is unstructured data that has been pre-prosed and structured.

As per the given information, lower the Sales Rank books should be highly recommended as they tend to increase the sales volume.

Furthermore, Higher “AvgRatings” implies higher popularity among the crowd and have higher possibility to be bought by a customer.

Additionally, the number of the cases the book been cross-sold with other ones is reflected in its degree of centrality. Also, clustering coefficient strengths the cross-sell metrics for the neighbors of a particular book.

## Normalizing Data:

Data like Degree of centrality and Sales Rank needs to be scaled before using it as a metric to recommend a book to the customer.

Also, to make sure that the scaled value should be positive value.

Hence we add the output by 4. Y = SS(Z) = Z+4

## Metrics Creation:

There are two metrics created based on aforementioned metrics. The top 5 books are recommended based on **higher score** of the sum of the two metrics.

### Metrics 1:

* It is the **product** of **clustering coefficient** and **degree of centrality**.
* It gives takes care that a book with high clustering coefficient but low degree of centrality and vis-verse are penalized due to low score of the another.
* This ensures that a case like high clustering coefficient for books with low degree of centrality aren’t recommended over the ones with equally high clustering coefficient but with better degree of centrality.

### Metrics 2:

* It is the **ratio of average ratings to sales rank**.
* The metrics takes average ratings i.e. higher the better metrics and divides it by sales rank i.e. lower the better.
* Finally, a book with high average rating and low sales rank isn’t recommended as it would be highly unlikely to result in sales of the book.