Zalora's Data Science Technical test.

Introduction & Objective:

Zalora is a online fashion retailer. We have hundreds of thousands of products accross all of our ventures - Singapore, Indonesia, Malaysia, Hong Kong, Thailand, Philippines, Vietnam, Brunei, Australia and New Zealand.

One of our challenging problems, is to determine which products are better than others, essentially a venture-wide ranking for all of its products. The ranking can be different for each different product category / subcategory.

The result of this ranking serves a lot of purposes:

- Displaying best selling products on our homepage.
- Sorting our catalogs / search results whenever users surf our homepage.
- Sending better, more profitable email campaigns.

Dataset:

(Attached as a single products.csv file) Luckily, our Data Engineering team has done a lot of data cleaning and aggregation to produce everything in a clearly formatted table. The given data set contains 4000 Indonesia products and their metadatas that are currently available on our store, sorted by a random order.

[Notice: a large part of this was randomized since it contains confidential information]

Questions:

1) Give us your suggestions on how we could make our data set better / more useful.

2) With the given dataset, can you come up with a scientific approach and model for our ranking?

3) How would you test, train, and evaluate your model?