

Rent the Runway – Customer and Item segmentation

Problem Statement:

The objective of this study is to segment the similar user and similar items into clusters. Understanding the common characteristic of each group will help the organization to promote their business.

The datasets used in this work contain measurements of clothing fit from RentTheRunway.

RentTheRunWay is a unique platform that allows women to rent clothes for various occasions; we collected data from several categories. These datasets contain self-reported fit feedback from customers as well as other side information like reviews, ratings, product categories, catalog sizes, customers' measurements (etc.). In this datasets, fit feedback belongs to one of three classes: 'Small,' 'Fit,' and 'Large.'

FILE DESCRIPTION

These datasets contain measurements of clothing fit from RentTheRunway.

1. Number of users: 105,508
2. Number of items: 5,850
3. Number of transactions: 192,544

Metadata:

1. ratings and reviews
2. fit feedback (small/fit/large etc.)
3. user/item measurements
4. category information

Data Citation:

Decomposing fit semantics for product size recommendation in metric spaces

Rishabh Misra, Mengting Wan, Julian McAuley

RecSys, 2018

We have altered this dataset by deriving some new columns which has been computed by applying text analytics strategies on the review column of data.

Appropriate features need to be selected for user segmentation(clustering) and Item Segmentation (clustering) process.

Expected Approach/Outcomes:

Objective: To segment the similar user and similar items in to clusters and understanding the common characteristic of each group will help the organization to promote their business.

Techniques used:

- **Min Max scalar** used to scale the data
- Implemented **PCA** for dimensionality reduction.
- **K Means Algorithm** used to cluster the data

Outcome: Separate cluster were built for user and items separately. Most important features w.r.t users (K=2 is the optimal for the user data) and w.r.t items (the ratings, review polarity, subjectivity and the positive sentiment score are high).

Key skills: Data Preparation | Clustering Model Building.