

Dataquest Challenge 2025: Recommender Systems

Data Dictionary

This document shows a brief overview and explanation of the dataset that is used for the Dataquest challenge. A screenshot of the top set of rows are shown in Figure 1.

idcol	interaction	int_date	item	page	tod	item_type	item_descrip	segment	beh_segment	active_ind
755	DISPLAY	17-Jan-23	NONE	Screen1	Afternoon	ALL		segment3	B01	Semi Active
4521	DISPLAY	27-Feb-23	NONE	Screen1	Afternoon	ALL		segment1	B07	Semi Active
4521	DISPLAY	18-Feb-23	NONE	Screen1	Afternoon	ALL		segment1	B07	Semi Active
4521	DISPLAY	30-Jan-23	NONE	Screen1	Morning	ALL		segment1	B07	Semi Active
4521	CLICK	5-Feb-23	IBAB	Screen1	Afternoon	INSURE	GENERIC MESSAGE	segment1	B07	Semi Active
4521	CHECKOUT	5-Feb-23	IBAB	Screen1	Afternoon	INSURE	GENERIC MESSAGE	segment1	B07	Semi Active
6145	DISPLAY	26-Feb-23	NONE	Screen1	Evening	ALL		segment3	B01	Cold Start
6145	DISPLAY	27-Jan-23	NONE	Screen1	Early	ALL		segment3	B01	Cold Start
6145	DISPLAY	10-Feb-23	NONE	Screen1	Morning	ALL		segment3	B01	Cold Start
6145	DISPLAY	10-Jan-23	NONE	Screen1	Afternoon	ALL		segment3	B01	Cold Start
7125	DISPLAY	18-Feb-23	NONE	Screen1	Morning	ALL		segment3	B01	Cold Start
8469	DISPLAY	5-Feb-23	NONE	Screen1	Morning	ALL		segment1	B01	Semi Active
8469	DISPLAY	14-Jan-23	NONE	Screen1	Afternoon	ALL		segment1	B01	Semi Active
8469	DISPLAY	21-Feb-23	NONE	Screen1	Morning	ALL		segment1	B01	Semi Active
13768	DISPLAY	2-Feb-23	NONE	Screen1	Morning	ALL		segment3	B01	Cold Start
14454	DISPLAY	9-Feb-23	NONE	Screen1	Morning	ALL		segment2	B01	Active
14454	CLICK	8-Feb-23	CAFM	Screen2	Morning	TRANSACT	FUSION MIGRATION	segment2	B01	Active
14454	CHECKOUT	8-Feb-23	CAFM	Screen2	Morning	TRANSACT	FUSION MIGRATION	segment2	B01	Active

Figure 1 Screenshot of the top rows of the dq_ps_challenge_v2.csv dataset.

Each of the features columns are grouped into these types of feature-types, to fit into the user-item interaction matrix as in **Figure 2**.

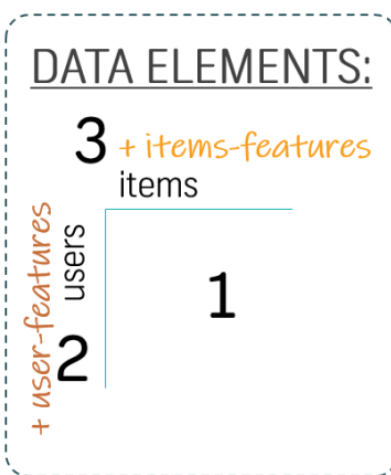


Figure 2 The user-item-interaction matrix, and related features.

The feature type and description of each column in the dataset, are detailed in and it's contents is depicted in Table 1.

Table 1 The Data Dictionary for the Personalised Solutions Dataquest challenge.

Column	Feature-type	Description
Idcol	User-identifier (2)	The id-column that is a unique identifier per customer
interaction	Interaction (1)	Three interactions: DISPLAY: this refers to a customer logging in, but no specific click or checkout of a item CLICK: a click-interaction, showing that a customer have clicked on one of the items that were displayed. This shows some initial interest in a product, but is not the final CHECKOUT interaction yet. CHECKOUT: this is a final step to take-up a product that were displayed to a customer.
int_date	Interaction-date (1)	The date on which the interaction occurred
item	Item-identifier (3)	The item-code with which the customer interacted with. The items are grouped into the <i>item_type</i> column, and more details of what the code refers to is in the <i>item_description</i> column. When there was only a DISPLAY interaction (i.e. no CLICK or CHECKOUT interaction), then the item is the NONE class.
page	context	There are two pages on the App via which the customer can interact
tod	context	Time-of-day: The time of each interaction was grouped into 4 timeslots: Early: 00:00 -<06:00 Morning: 06:00-<12:00 Afternoon: 12:00-<18:00 Evening: 18:00-<00:00
item_type	Item-feature (3)	Item_type: All of the items are grouped into the following groups: TRANSACT, LEND, INVEST, INSURE, CONNECT and LIFESTYLE When the interaction is DISPLAY, the item_type reflects as ALL
item_descrip	Item-feature (3)	A description of each item
segment	user-feature (2)	4 broad customer segments, broadly linked to income, with segment1 to segment 4 ranging from lower to higher income
beh_segment	user-feature (2)	50 detailed customers segments. The exact definition is not disclosed, but can be seen as 50 different customer segments.
active_ind	user-feature (2)	Active: customers who interact regularly on this platform Semi Active: customers who interact occasionally on this platform Cold Start: customers who interact very rarely on this platform, or not at all. <i>This definition was applied as at the start of the data time-window (1 January 2023).</i>