Unique IDs vs Journeys Report

Data file: C:\Users\samvi\Downloads\dat_train1.csv
Partitions folder: C:\Users\samvi\Downloads\m148_eda_work_full_fast\partitions
Partition files: 9,472
Events processed: 54,960,961
Journeys: 2,020,239
Inactivity threshold: 60 days
Success event: order_shipped
Dataset window: 2020-11-03 03:31:30+00:00 to 2023-01-23 12:29:56+00:00

Summary: IDs and multi-journey duplication

metric	value
journeys_total	2.02e+06
unique_ids_customer_account	1.43e+06
avg_journeys_per_id	1.412
median_journeys_per_id	1
ids_with_multiple_journeys	4.862e+05
extra_journeys_beyond_1_per_id	5.898e+05
unique_customers	1.391e+06
customers_with_multiple_journeys	4.976e+05
journeys_with_multiple_order_shipped	139

Journeys per (customer_id, account_id)

metric	journeys_per_id
count	1.43e+06
mean	1.412
std	0.6269
min	1
p01	1
p05	1
p25	1
p50	1
p75	2
p95	3
p99	3
max	5

Journeys per customer_id

metric	journeys_per_customer
count	1.391e+06
mean	1.452
std	0.6778
min	1
p01	1
p05	1
p25	1
p50	1
p75	2
p95	3
p99	3
max	8

Outcomes

outcome	count	share
INACTIVITY	1582548	0.7833
SUCCESS	279363	0.1383
CENSORED	158328	0.07837

Timing

Journey duration (days)

metric	duration_days
count	2.02e+06
mean	62.69
std	73.15
min	0
p01	0
p05	0
p25	1.908
p50	35
p75	98.55
p95	214.8
p99	297.8
max	437.1

Inter-action gaps (days)

metric	gap_days
count	5.294e+07
mean	2.392
std	7.609
min	0
p01	0
p05	0
p25	0.0001505
p50	0.0007986
p75	0.0314
p95	19.02
p99	40.79
max	60

Figures

Journey outcomes

Journey duration (days)

Inter-action gaps (days)