Execution Report

Results

The following is the resulting data that was successfully migrated to the production table for Users:

In total, 28 rows were successfully processed out of the original 34 given in the sample data.

	ID	UserID	FullName	Age	Email	RegistrationDate	LastLoginDate	PurchaseTotal	RecordLastUpdated
1	1	101	John Doe	30	johndoe@example.com	2023-01-10	2023-03-01	350	2024-07-26 09:45:41.070
2	2	102	Jane Smith	25	janesmith@example.com	2020-05-15	2023-02-25	200	2024-07-26 09:45:41.070
3	3	103	Emily Johnson	22	emilyjohnson@example.com	2019-03-23	2023-01-30	120.5	2024-07-26 09:45:41.070
4	4	104	Michael Brown	35	michaelbrown@example.com	2018-07-18	2023-02-20	300.75	2024-07-26 09:45:41.070
5	5	105	Jessica Garcia	28	jessicagarcia@example.com	2022-08-05	2023-02-18	180.25	2024-07-26 09:45:41.070
6	6	106	David Miller	40	davidmiller@example.com	2017-12-12	2023-02-15	220.4	2024-07-26 09:45:41.070
7	7	107	Sarah Martinez	33	sarahmartinez@example.com	2018-11-30	2023-02-10	140.6	2024-07-26 09:45:41.070
8	8	108	James Taylor	29	jamestaylor@example.com	2019-06-22	2023-02-05	210	2024-07-26 09:45:41.070
9	9	109	Linda Anderson	27	lindaanderson@example.com	2021-04-16	2023-01-25	165.95	2024-07-26 09:45:41.070
10	10	110	Robert Wilson	31	robertwilson@example.com	2020-02-20	2023-01-20	175	2024-07-26 09:45:41.070
11	11	112	Bob Marley	27	bobmarley@example.com	2021-05-20	2023-02-28	0	2024-07-26 09:45:56.620
12	12	114	Derek Nowak	29	dereknowak@example.com	2020-03-17	2023-02-24	180.75	2024-07-26 09:45:56.620
13	13	116	Frank Poe	34	frankpoe@example.com	2019-11-13	2023-02-22	190.4	2024-07-26 09:45:56.620
14	14	117	George Kay	28	georgekay@example.com	2021-06-07	2023-02-21	170	2024-07-26 09:45:56.620
15	15	118	Hanna Lux	32	hannalux@example.com	2019-02-18	2023-02-20	220.15	2024-07-26 09:45:56.620
16	16	119	Ian Volt	26	ianvolt@example.com	2020-07-23	2023-02-19	130	2024-07-26 09:45:56.620
17	17	120	Julia Nex	24	julianex@example.com	2022-01-12	2023-02-18	145.6	2024-07-26 09:45:56.620
18	18	122	Lana Molt	31	lanamolt@example.com	2020-09-09	2023-02-16	155.3	2024-07-26 09:45:56.620
19	19	123	Mike Dolt	33	mikedolt@example.com	2018-05-05	2023-02-15	205	2024-07-26 09:45:56.620
20	20	124	Nina Colt	27	ninacolt@example.com	2019-07-29	2023-02-14	160.75	2024-07-26 09:45:56.620
21	21	125	Oscar Holt	29	oscarholt@example.com	2022-11-11	2023-02-13	175.45	2024-07-26 09:45:56.620
22	22	126	Patty Jolt	31	pattyjolt@example.com	2018-12-13	2023-02-12	185.99	2024-07-26 09:45:56.620
23	23	127	Quincy Molt	34	quincymolt@example.com	2020-04-17	2023-02-11	195.25	2024-07-26 09:45:56.620
24	24	128	Rita Bolt	36	ritabolt@example.com	2021-08-21	2023-02-10	210.5	2024-07-26 09:45:56.620
25	25	129	Steve Jolt	38	stevejolt@example.com	2019-01-01	2023-02-09	225.75	2024-07-26 09:45:56.620
26	26	131	Special Characters Name	34	special$$name@example.com	2021-04-17	2023-02-07	200.5	2024-07-26 09:45:56.620
27	27	132	Old Format	30	oldformat@example.com	2019-07-05	2023-02-06	150.5	2024-07-26 09:45:56.620
28	28	138	Duplicate ID	27	duplicateid@example.com	2019-03-15	2023-01-31	200	2024-07-26 09:45:56.620

Excluded Records

The following table is the data that was moved into the “Invalids” table for review/correction.

	UserID	FullName	Age	Email	RegistrationDate	LastLoginDate	PurchaseTotal
1	113	Cathy Smith	null	cathysmith@example.com	10/12/2019	2/27/2023	210.50
2	115	Eva Green	null	evagreen@example.com	8/22/2018	2/23/2023	NULL
3	121	Kevin Yolt	abc	kevinyolt@example.com	3/14/2021	2/17/2023	NULL
4	130	Invalid Date	29	invaliddatetest@example.com	2022-02-30	2/8/2023	180.00
5	134	Negative Age	-1	negativeage@example.com	8/15/2021	2/4/2023	130.00
6	136	Extra Large Total	27	extralargetotal@example.com	10/10/2020	2/2/2023	"1,000,000.00"
7	NULL	Null User	25	NULL	2020-12-01	2023-02-25	100
8	137	Incorrect Email	28	notanemail	2021-11-11	2023-02-01	190
9	135	Very Old Date	90	veryolddate@example.com	1920-01-01	2023-02-03	100
10	111	Alice Johnson	NULL	alicejohnson@example.com	2022-07-15	NULL	NULL
11	133	Future Date	25	futuredate@example.com	2024-01-01	2025-01-02	160
12	101	John Doe	30	johndoe@example.com	2024-01-10	2023-03-01	250

A majority of data was migrated to this table due to data type mismatches that arose during the data conversion stage of the ETL process (e.g. age being a string saying “null”). The rest of the data in this table was data that made it to the staging table before being moved to Invalids due to semantic errors. These reasons included:

- NULL values (for any column except PurchaseTotal)
- FullName not containing a space, therefore not being a full name (e.g. “Johnny”)
- Emails not containing “@”
- RegistrationDates that were older than the user themself
- LastLoginDate dating earlier than RegistrationDate
- LastLoginDate being in the future (i.e. a later date than current day)

All of these cases depict values in columns that would be incorrect without exception. All other conditions that could have moved more rows to Invalids were thought to be case-by-case and therefore not assumed to be an error.

Challenges

The main challenge that I faced was deciding if rows were semantically incorrect enough to exclude or deal with and what exactly the bounds of these conditions would be. Though conditions like not having an “@” in the email string were obvious, it was difficult to make a decision on what would be semantically wrong for RegistrationDates without exception, for example. Even though it doesn’t seem right for the registration date to be the same year the user was when they were two years old, this is technically possible and could not be outright ruled out of consideration (e.g. maybe someone else made them the account). Ultimately, the hardest part of trying to make sense of what to do with the data came down to the fact that it is hard to understand for sure what the data was even about. This made making semantic decisions much harder.