

NoSQL

September 1, 2025

```
[1]: import boto3, os
import pandas as pd
from dotenv import load_dotenv
import psycopg2
import psycopg2.extras as pgx
import json
import sys
import praw
import time, datetime as dt
```

```
[2]: load_dotenv("/home/ubuntu//deds2025b_proj/opt/reddit_pipeline/.env")
bucket_name = os.environ["LAKE_BUCKET"]
s3 = boto3.client("s3")
```

```
[5]: secrets = boto3.client('secretsmanager')
```

```
[3]: def dynamodb_client():
    return boto3.resource("dynamodb", region_name="us-east-1")
```

```
[6]: def reddit_client():
    cfg = json.loads(secrets.
↳get_secret_value(SecretId=REDDIT_SECRET_ARN)['SecretString'])
    return praw.Reddit(
        client_id=cfg['client_id'],
        client_secret=cfg['client_secret'],
        user_agent='aws:batch-ec2:1.0 (by u/Entire-Success-5370)'
    )
```

```
[8]: REDDIT_SECRET_ARN = os.environ['REDDIT_SECRET_ARN']
```

```
[10]: rd = reddit_client()
dy = dynamodb_client()
table = dy.Table("reddit")
response = table.scan()
```

```
[11]: df = pd.DataFrame(response['Items'])
pd.set_option("display.max_rows", None)
df
```

[11]:	SK	PK subscribers \
0	2025-08-31T06:00:00Z	SUBREDDIT:cancer 77848
1	2025-08-31T07:00:00Z	SUBREDDIT:cancer 77849
2	2025-08-31T08:00:00Z	SUBREDDIT:cancer 77850
3	2025-08-31T09:00:00Z	SUBREDDIT:cancer 77851
4	2025-08-31T10:00:00Z	SUBREDDIT:cancer 77850
5	2025-08-31T11:00:00Z	SUBREDDIT:cancer 77852
6	2025-08-31T16:00:00Z	SUBREDDIT:cancer 77862
7	2025-08-31T17:00:00Z	SUBREDDIT:cancer 77865
8	2025-08-31T18:00:00Z	SUBREDDIT:cancer 77865
9	2025-08-31T19:00:00Z	SUBREDDIT:cancer 77866
10	2025-08-31T20:00:00Z	SUBREDDIT:cancer 77867
11	2025-08-31T21:00:00Z	SUBREDDIT:cancer 77868
12	2025-08-31T22:00:00Z	SUBREDDIT:cancer 77867
13	2025-08-31T23:00:00Z	SUBREDDIT:cancer 77868
14	2025-09-01T00:00:00Z	SUBREDDIT:cancer 77869
15	2025-09-01T04:00:00Z	SUBREDDIT:cancer 77878
16	2025-09-01T05:00:00Z	SUBREDDIT:cancer 77878
17	2025-09-01T10:00:00Z	SUBREDDIT:cancer 77884
18	2025-09-01T11:00:00Z	SUBREDDIT:cancer 77884
19	2025-09-01T12:00:00Z	SUBREDDIT:cancer 77884
20	2025-09-01T13:00:00Z	SUBREDDIT:cancer 77882
21	2025-09-01T14:00:00Z	SUBREDDIT:cancer 77886
22	2025-08-31T06:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
23	2025-08-31T07:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
24	2025-08-31T08:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
25	2025-08-31T09:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
26	2025-08-31T10:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
27	2025-08-31T11:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
28	2025-08-31T16:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
29	2025-08-31T17:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
30	2025-08-31T18:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
31	2025-08-31T19:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
32	2025-08-31T20:00:00Z	SUBREDDIT:IndustrialPharmacy 1234
33	2025-08-31T21:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
34	2025-08-31T22:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
35	2025-08-31T23:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
36	2025-09-01T00:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
37	2025-09-01T04:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
38	2025-09-01T05:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
39	2025-09-01T10:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
40	2025-09-01T11:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
41	2025-09-01T12:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
42	2025-09-01T13:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
43	2025-09-01T14:00:00Z	SUBREDDIT:IndustrialPharmacy 1235
44	2025-08-31T06:00:00Z	SUBREDDIT:breastcancer 38049
45	2025-08-31T07:00:00Z	SUBREDDIT:breastcancer 38048

46	2025-08-31T08:00:00Z	SUBREDDIT:breastcancer	38048
47	2025-08-31T09:00:00Z	SUBREDDIT:breastcancer	38048
48	2025-08-31T10:00:00Z	SUBREDDIT:breastcancer	38047
49	2025-08-31T11:00:00Z	SUBREDDIT:breastcancer	38048
50	2025-08-31T16:00:00Z	SUBREDDIT:breastcancer	38059
51	2025-08-31T17:00:00Z	SUBREDDIT:breastcancer	38059
52	2025-08-31T18:00:00Z	SUBREDDIT:breastcancer	38060
53	2025-08-31T19:00:00Z	SUBREDDIT:breastcancer	38061
54	2025-08-31T20:00:00Z	SUBREDDIT:breastcancer	38063
55	2025-08-31T21:00:00Z	SUBREDDIT:breastcancer	38066
56	2025-08-31T22:00:00Z	SUBREDDIT:breastcancer	38068
57	2025-08-31T23:00:00Z	SUBREDDIT:breastcancer	38071
58	2025-09-01T00:00:00Z	SUBREDDIT:breastcancer	38071
59	2025-09-01T04:00:00Z	SUBREDDIT:breastcancer	38074
60	2025-09-01T05:00:00Z	SUBREDDIT:breastcancer	38075
61	2025-09-01T10:00:00Z	SUBREDDIT:breastcancer	38075
62	2025-09-01T11:00:00Z	SUBREDDIT:breastcancer	38076
63	2025-09-01T12:00:00Z	SUBREDDIT:breastcancer	38076
64	2025-09-01T13:00:00Z	SUBREDDIT:breastcancer	38076
65	2025-09-01T14:00:00Z	SUBREDDIT:breastcancer	38078
66	2025-08-31T06:00:00Z	SUBREDDIT:LivingWithMBC	3684
67	2025-08-31T07:00:00Z	SUBREDDIT:LivingWithMBC	3683
68	2025-08-31T08:00:00Z	SUBREDDIT:LivingWithMBC	3683
69	2025-08-31T09:00:00Z	SUBREDDIT:LivingWithMBC	3683
70	2025-08-31T10:00:00Z	SUBREDDIT:LivingWithMBC	3683
71	2025-08-31T11:00:00Z	SUBREDDIT:LivingWithMBC	3683
72	2025-08-31T16:00:00Z	SUBREDDIT:LivingWithMBC	3684
73	2025-08-31T17:00:00Z	SUBREDDIT:LivingWithMBC	3684
74	2025-08-31T18:00:00Z	SUBREDDIT:LivingWithMBC	3684
75	2025-08-31T19:00:00Z	SUBREDDIT:LivingWithMBC	3684
76	2025-08-31T20:00:00Z	SUBREDDIT:LivingWithMBC	3684
77	2025-08-31T21:00:00Z	SUBREDDIT:LivingWithMBC	3683
78	2025-08-31T22:00:00Z	SUBREDDIT:LivingWithMBC	3682
79	2025-08-31T23:00:00Z	SUBREDDIT:LivingWithMBC	3682
80	2025-09-01T00:00:00Z	SUBREDDIT:LivingWithMBC	3682
81	2025-09-01T04:00:00Z	SUBREDDIT:LivingWithMBC	3681
82	2025-09-01T05:00:00Z	SUBREDDIT:LivingWithMBC	3681
83	2025-09-01T10:00:00Z	SUBREDDIT:LivingWithMBC	3682
84	2025-09-01T11:00:00Z	SUBREDDIT:LivingWithMBC	3682
85	2025-09-01T12:00:00Z	SUBREDDIT:LivingWithMBC	3682
86	2025-09-01T13:00:00Z	SUBREDDIT:LivingWithMBC	3682
87	2025-09-01T14:00:00Z	SUBREDDIT:LivingWithMBC	3683
88	2025-08-31T06:00:00Z	SUBREDDIT:BRCA	5332
89	2025-08-31T07:00:00Z	SUBREDDIT:BRCA	5332
90	2025-08-31T08:00:00Z	SUBREDDIT:BRCA	5333
91	2025-08-31T09:00:00Z	SUBREDDIT:BRCA	5333
92	2025-08-31T10:00:00Z	SUBREDDIT:BRCA	5333

93	2025-08-31T11:00:00Z	SUBREDDIT:BRCA	5333
94	2025-08-31T16:00:00Z	SUBREDDIT:BRCA	5334
95	2025-08-31T17:00:00Z	SUBREDDIT:BRCA	5334
96	2025-08-31T18:00:00Z	SUBREDDIT:BRCA	5335
97	2025-08-31T19:00:00Z	SUBREDDIT:BRCA	5336
98	2025-08-31T20:00:00Z	SUBREDDIT:BRCA	5336
99	2025-08-31T21:00:00Z	SUBREDDIT:BRCA	5336
100	2025-08-31T22:00:00Z	SUBREDDIT:BRCA	5337
101	2025-08-31T23:00:00Z	SUBREDDIT:BRCA	5338
102	2025-09-01T00:00:00Z	SUBREDDIT:BRCA	5338
103	2025-09-01T04:00:00Z	SUBREDDIT:BRCA	5339
104	2025-09-01T05:00:00Z	SUBREDDIT:BRCA	5339
105	2025-09-01T10:00:00Z	SUBREDDIT:BRCA	5339
106	2025-09-01T11:00:00Z	SUBREDDIT:BRCA	5340
107	2025-09-01T12:00:00Z	SUBREDDIT:BRCA	5340
108	2025-09-01T13:00:00Z	SUBREDDIT:BRCA	5340
109	2025-09-01T14:00:00Z	SUBREDDIT:BRCA	5340
110	2025-08-31T06:00:00Z	SUBREDDIT:ovariancancer_new	939
111	2025-08-31T07:00:00Z	SUBREDDIT:ovariancancer_new	939
112	2025-08-31T08:00:00Z	SUBREDDIT:ovariancancer_new	939
113	2025-08-31T09:00:00Z	SUBREDDIT:ovariancancer_new	939
114	2025-08-31T10:00:00Z	SUBREDDIT:ovariancancer_new	939
115	2025-08-31T11:00:00Z	SUBREDDIT:ovariancancer_new	939
116	2025-08-31T16:00:00Z	SUBREDDIT:ovariancancer_new	939
117	2025-08-31T17:00:00Z	SUBREDDIT:ovariancancer_new	939
118	2025-08-31T18:00:00Z	SUBREDDIT:ovariancancer_new	939
119	2025-08-31T19:00:00Z	SUBREDDIT:ovariancancer_new	939
120	2025-08-31T20:00:00Z	SUBREDDIT:ovariancancer_new	939
121	2025-08-31T21:00:00Z	SUBREDDIT:ovariancancer_new	939
122	2025-08-31T22:00:00Z	SUBREDDIT:ovariancancer_new	939
123	2025-08-31T23:00:00Z	SUBREDDIT:ovariancancer_new	939
124	2025-09-01T00:00:00Z	SUBREDDIT:ovariancancer_new	939
125	2025-09-01T04:00:00Z	SUBREDDIT:ovariancancer_new	940
126	2025-09-01T05:00:00Z	SUBREDDIT:ovariancancer_new	940
127	2025-09-01T10:00:00Z	SUBREDDIT:ovariancancer_new	940
128	2025-09-01T11:00:00Z	SUBREDDIT:ovariancancer_new	940
129	2025-09-01T12:00:00Z	SUBREDDIT:ovariancancer_new	940
130	2025-09-01T13:00:00Z	SUBREDDIT:ovariancancer_new	940
131	2025-09-01T14:00:00Z	SUBREDDIT:ovariancancer_new	940

	active_users	timestamp
0	25	2025-08-31 06:42:46.127680+00:00
1	23	2025-08-31 07:00:02.606803+00:00
2	18	2025-08-31 08:00:02.139939+00:00
3	23	2025-08-31 09:00:02.449542+00:00
4	29	2025-08-31 10:00:02.834977+00:00
5	27	2025-08-31 11:00:02.831949+00:00

6	46	2025-08-31	16:00:02.925712+00:00
7	33	2025-08-31	17:05:03.451432+00:00
8	41	2025-08-31	18:00:03.433051+00:00
9	33	2025-08-31	19:05:03.702767+00:00
10	29	2025-08-31	20:00:02.326490+00:00
11	34	2025-08-31	21:00:02.657450+00:00
12	37	2025-08-31	22:05:03.010660+00:00
13	39	2025-08-31	23:00:02.512052+00:00
14	27	2025-09-01	00:00:02.759697+00:00
15	33	2025-09-01	04:25:32.988053+00:00
16	39	2025-09-01	05:00:08.533048+00:00
17	68	2025-09-01	10:35:16.318005+00:00
18	49	2025-09-01	11:50:29.137169+00:00
19	33	2025-09-01	12:00:03.052861+00:00
20	64	2025-09-01	13:00:08.197750+00:00
21	31	2025-09-01	14:00:08.801532+00:00
22	3	2025-08-31	06:42:46.200511+00:00
23	3	2025-08-31	07:00:02.674375+00:00
24	2	2025-08-31	08:00:02.214563+00:00
25	4	2025-08-31	09:00:02.539536+00:00
26	1	2025-08-31	10:00:02.916954+00:00
27	3	2025-08-31	11:00:02.905288+00:00
28	5	2025-08-31	16:00:03.004101+00:00
29	0	2025-08-31	17:05:03.535636+00:00
30	3	2025-08-31	18:00:03.524403+00:00
31	4	2025-08-31	19:05:03.783356+00:00
32	3	2025-08-31	20:00:02.406943+00:00
33	2	2025-08-31	21:00:02.753041+00:00
34	3	2025-08-31	22:05:03.084054+00:00
35	4	2025-08-31	23:00:02.592045+00:00
36	2	2025-09-01	00:00:02.855585+00:00
37	2	2025-09-01	04:25:33.058805+00:00
38	0	2025-09-01	05:00:08.624048+00:00
39	3	2025-09-01	10:35:16.400130+00:00
40	0	2025-09-01	11:50:29.223653+00:00
41	6	2025-09-01	12:00:03.147496+00:00
42	1	2025-09-01	13:00:08.394494+00:00
43	4	2025-09-01	14:00:08.902389+00:00
44	30	2025-08-31	06:42:45.839972+00:00
45	26	2025-08-31	07:00:02.383408+00:00
46	38	2025-08-31	08:00:01.930434+00:00
47	43	2025-08-31	09:00:02.115848+00:00
48	47	2025-08-31	10:00:02.629762+00:00
49	32	2025-08-31	11:00:02.616111+00:00
50	45	2025-08-31	16:00:02.677535+00:00
51	54	2025-08-31	17:05:03.249436+00:00
52	43	2025-08-31	18:00:03.081090+00:00

53	44	2025-08-31	19:05:03.469958+00:00
54	56	2025-08-31	20:00:02.092629+00:00
55	42	2025-08-31	21:00:02.420423+00:00
56	60	2025-08-31	22:05:02.772278+00:00
57	55	2025-08-31	23:00:02.258043+00:00
58	39	2025-09-01	00:00:02.513321+00:00
59	37	2025-09-01	04:25:32.751736+00:00
60	52	2025-09-01	05:00:08.288050+00:00
61	40	2025-09-01	10:35:16.090620+00:00
62	45	2025-09-01	11:50:28.897945+00:00
63	53	2025-09-01	12:00:02.720035+00:00
64	51	2025-09-01	13:00:07.225343+00:00
65	52	2025-09-01	14:00:08.448540+00:00
66	8	2025-08-31	06:42:45.717972+00:00
67	2	2025-08-31	07:00:02.222649+00:00
68	6	2025-08-31	08:00:01.829329+00:00
69	3	2025-08-31	09:00:02.007281+00:00
70	1	2025-08-31	10:00:02.518445+00:00
71	2	2025-08-31	11:00:02.503584+00:00
72	4	2025-08-31	16:00:02.552233+00:00
73	4	2025-08-31	17:05:03.143406+00:00
74	1	2025-08-31	18:00:02.946081+00:00
75	2	2025-08-31	19:05:03.354390+00:00
76	0	2025-08-31	20:00:01.967666+00:00
77	3	2025-08-31	21:00:02.297266+00:00
78	0	2025-08-31	22:05:02.661400+00:00
79	5	2025-08-31	23:00:02.112783+00:00
80	5	2025-09-01	00:00:02.363494+00:00
81	8	2025-09-01	04:25:32.648505+00:00
82	2	2025-09-01	05:00:08.143659+00:00
83	5	2025-09-01	10:35:15.983544+00:00
84	1	2025-09-01	11:50:28.784461+00:00
85	3	2025-09-01	12:00:02.606834+00:00
86	1	2025-09-01	13:00:06.849073+00:00
87	6	2025-09-01	14:00:08.128283+00:00
88	6	2025-08-31	06:42:46.067357+00:00
89	4	2025-08-31	07:00:02.542111+00:00
90	4	2025-08-31	08:00:02.071969+00:00
91	0	2025-08-31	09:00:02.380233+00:00
92	2	2025-08-31	10:00:02.774319+00:00
93	3	2025-08-31	11:00:02.769471+00:00
94	8	2025-08-31	16:00:02.843050+00:00
95	2	2025-08-31	17:05:03.386711+00:00
96	0	2025-08-31	18:00:03.347057+00:00
97	6	2025-08-31	19:05:03.637909+00:00
98	6	2025-08-31	20:00:02.237451+00:00
99	6	2025-08-31	21:00:02.582515+00:00

100	6	2025-08-31 22:05:02.928461+00:00
101	5	2025-08-31 23:00:02.429047+00:00
102	6	2025-09-01 00:00:02.682314+00:00
103	5	2025-09-01 04:25:32.907527+00:00
104	6	2025-09-01 05:00:08.460041+00:00
105	3	2025-09-01 10:35:16.247352+00:00
106	6	2025-09-01 11:50:29.065935+00:00
107	6	2025-09-01 12:00:02.905802+00:00
108	12	2025-09-01 13:00:07.722122+00:00
109	2	2025-09-01 14:00:08.677195+00:00
110	0	2025-08-31 06:42:46.004000+00:00
111	2	2025-08-31 07:00:02.463003+00:00
112	4	2025-08-31 08:00:02.001346+00:00
113	1	2025-08-31 09:00:02.210272+00:00
114	3	2025-08-31 10:00:02.711659+00:00
115	4	2025-08-31 11:00:02.700617+00:00
116	6	2025-08-31 16:00:02.758061+00:00
117	1	2025-08-31 17:05:03.316897+00:00
118	3	2025-08-31 18:00:03.177066+00:00
119	2	2025-08-31 19:05:03.561384+00:00
120	5	2025-08-31 20:00:02.167089+00:00
121	3	2025-08-31 21:00:02.500386+00:00
122	3	2025-08-31 22:05:02.847135+00:00
123	3	2025-08-31 23:00:02.344390+00:00
124	4	2025-09-01 00:00:02.599052+00:00
125	1	2025-09-01 04:25:32.822231+00:00
126	2	2025-09-01 05:00:08.374056+00:00
127	5	2025-09-01 10:35:16.166400+00:00
128	4	2025-09-01 11:50:28.985679+00:00
129	2	2025-09-01 12:00:02.805774+00:00
130	4	2025-09-01 13:00:07.577840+00:00
131	2	2025-09-01 14:00:08.552598+00:00