Exploratory & Post-ETL Data Quality Assessment Report

Project: Data Integration & Master Table Validation

Prepared by Team 2

Date: 15/06/2025

Step 1: Exploratory Data Assessment (Pre-ETL)

Dataset Structures and Purposes

Dataset	Purpose	Primary Key(s)
Learner_Raw	Learner demographics and academic background	learner_id
Opportunity_Raw	Details about learning/growth opportunities	opportunity
CohortRaw	Information about cohorts	cohort_id, cohort_code
ad_campaign_data	Ad campaign performance metrics	ad_account_name, campaign_name
LearnerOpportunity_Raw	Mapping of learners to opportunities and cohorts	enrollment_id, learner_id, assigned_cohort
staging_users	User profiles and registration details	user_id, email

Relationships Between Datasets

- Learner_Raw.learner_id ↔ LearnerOpportunity_Raw.learner_id
- LearnerOpportunity_Raw.assigned_cohort ↔ CohortRaw.cohort_id
- staging_users.email ↔ Learner_Raw.email

Step 2: Identify Data Quality Issues (Pre-ETL)

Missing Values

➢ Learner_Raw: 'degree', 'institution', and 'major' had "null" (as text).
➢ staging_users: 'gender', 'birthdate', 'state', 'zip' contained blank or invalid entries.
➢ LearnerOpportunity_Raw: Missing 'assigned_cohort' or 'apply_date'.

Duplicate Records

➢ Duplicate learner_id in Learner_Raw
➢ Duplicate (enrollment_id, learner_id) in LearnerOpportunity_Raw
➢ Duplicate user_id or email in staging_users

Inconsistent Formats

➢ Gender variations: 'male', 'Male', 'M'
➢ Country names: inconsistent casing or spelling
➢ Campaign naming: variations in delivery_status or result_type

Orphan Records

➢ assigned_cohort not matching any CohortRaw
➢ learner_id in LearnerOpportunity_Raw not found in Learner_Raw
➢ email in staging_users missing in Learner_Raw

Step 3: Transformation Plan for ETL

➢ Convert 'null', 'none', '' (empty strings) to actual NULL
➢ Normalize text casing (e.g., gender, country, result_type)
➢ Remove duplicates using ROW_NUMBER() or DISTINCT
➢ Join and validate foreign key references
➢ Correct data types for dates, numerics, and categories

Step 4: Post-ETL Validation of Master Table

Data Quality Checks Performed:

Check	Purpose	Result
Record Count Validation	Ensured row counts match across source and Master Table	🟩 Matches confirmed
Duplicate Checks	No duplicates in enrollment_id or learner_id	🟩 No duplicates found
Missing Data Review	Ensured fields like degree, gender, birthdate are handled	🟩 Cleaned and NULLs used properly
Foreign Key Integrity	All references (learner, cohort, opportunity) are valid	🟩 Keys matched
Data Type Verification	Confirmed all fields are in correct formats	🟩 Correct data types

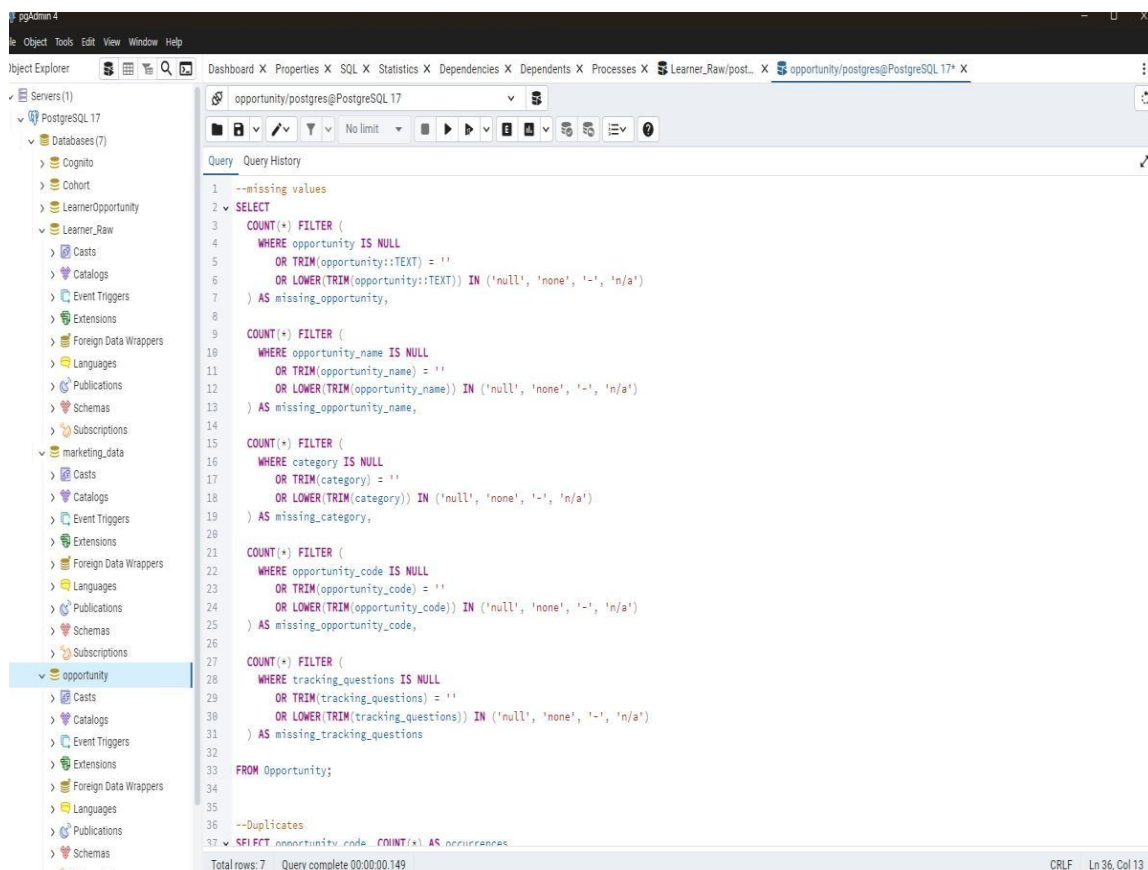## Step 5: ETL Refinement (If Issues Found)

Example Refinements

Issue	Fix Applied
Text 'null' as string	Converted to actual NULL values
Inconsistent gender values	Standardized to 'Male', 'Female', 'Other'
Missing cohort link	Used left join fallback and validation rules

Step 6: Final Assessment

The Master Table has passed all quality checks and is ready for use in downstream analysis. All source data was successfully integrated, cleaned, and validated.

Pictures indicate the queries identifying but not modifying missing values, duplicates, inconsistencies and Master Tables.

Object Explorer | Dashboard X Properties X SQL X Statistics X Dependencies X Dependents X Processes X Learner_Raw/postgres@PostgreSQL 17 X

Learner_Raw/postgres@PostgreSQL 17

Query Query History

```sql
3      WHERE LOWER(degree) = 'null';
4
5
6  v  SELECT
7        COUNT(*) FILTER (WHERE learner_id IS NULL OR TRIM(learner_id) = '') AS missing_learner_id,
8        COUNT(*) FILTER (WHERE LOWER(degree) = 'null' OR degree IS NULL OR TRIM(degree) = '') AS missing_degree,
9        COUNT(*) FILTER (WHERE LOWER(institution) = 'null' OR institution IS NULL OR TRIM(institution) = '') AS missing_institution,
10       COUNT(*) FILTER (WHERE LOWER(major) = 'null' OR major IS NULL OR TRIM(major) = '') AS missing_major,
11       COUNT(*) FILTER (WHERE LOWER(country) = 'null' OR country IS NULL OR TRIM(country) = '') AS missing_country
12     FROM Learner;
13
14     -- Duplicate
15  v  SELECT learner_id, COUNT(*) FROM Learner
16     GROUP BY learner_id
17     HAVING COUNT(*) > 1;
18     -- Inconsistent
19  v  SELECT DISTINCT LOWER(TRIM(degree)) AS cleaned_degree, COUNT(*)
20     FROM Learner
21     GROUP BY LOWER(TRIM(degree))
22     ORDER BY COUNT(*) DESC;
```

Data Output Messages Notifications

Showing rows: 1 to 8   Page No: 1   of 1

	cleaned_degree text	count bigint
1	null	52693
2	graduate student	31806
3	undergraduate student	30709
4	not in education	6319
5	high school student	4109
6	other professional	2997
7	teacher/educator	562
8	parent of student	64

Total rows: 8   Query complete 00:00:00.422   CRLF   Ln 14,

Object Explorer | Dashboard X Properties X SQL X Statistics X Dependencies X Dependents X Processes X Learner_Raw/post... X opportunity/postgres@PostgreSQL 17* X

opportunity/postgres@PostgreSQL 17

Query Query History

```sql
35
36     --Duplicates
37  v  SELECT opportunity_code, COUNT(*) AS occurrences
38     FROM Opportunity
39     GROUP BY opportunity_code
40     HAVING COUNT(*) > 1;
41
42     --inconsistencies
43  v  SELECT LOWER(TRIM(category)) AS normalized_category, COUNT(*) AS count
44     FROM Opportunity
45     GROUP BY LOWER(TRIM(category))
46     ORDER BY count DESC;
47
48
49     |
50
51
52
53
54
```

Data Output Messages Notifications

Showing rows: 1 to 7   Page No: 1   of 1

	normalized_category text	count bigint
1	internship	43
2	competition	41
3	event	41
4	career	23
5	course	18
6	masterclass	11
7	engagement	10

Total rows: 7   Query complete 00:00:00.149   CRLF

File Object Tools Edit View Window Help

Object Explorer | Dashboard X Properties X SQL X Statistics X Dependencies X Dependents X Processes X Learner_Raw/post... X opportunity/postgr... X marketing_data/postgres@PostgreSQL 17* X

marketing_data/postgres@PostgreSQL 17

Query | Query History

```sql
1  SELECT
2    COUNT(*) FILTER (
3      WHERE ad_account_name IS NULL
4        OR TRIM(ad_account_name) = ''
5        OR LOWER(TRIM(ad_account_name)) IN ('null', 'none', '-', 'n/a')
6    ) AS missing_ad_account_name,
7
8    COUNT(*) FILTER (
9      WHERE campaign_name IS NULL
10       OR TRIM(campaign_name) = ''
11       OR LOWER(TRIM(campaign_name)) IN ('null', 'none', '-', 'n/a')
12   ) AS missing_campaign_name,
13
14   COUNT(*) FILTER (
15     WHERE delivery_status IS NULL
16       OR TRIM(delivery_status) = ''
17       OR LOWER(TRIM(delivery_status)) IN ('null', 'none', '-', 'n/a')
18   ) AS missing_delivery_status,
19
20   COUNT(*) FILTER (
21     WHERE reach IS NULL
```

Data Output | Messages | Notifications

missing_ad_account_name bigint	missing_campaign_name bigint	missing_delivery_status bigint	missing_reach bigint	missing_outbound_clicks bigint
7	8	7	7	9

Showing rows: 1 to 1 | Page No: 1 of 1

Total rows: 1   Query complete 00:00:00.373   CRLF   Ln 1, Col 1

File Object Tools Edit View Window Help

Object Explorer | Dashboard X Properties X SQL X Statistics X Dependencies X Dependents X Processes X Learner_Raw/post... X opportunity/postgr... X marketing_data/postgres@PostgreSQL 17* X

marketing_data/postgres@PostgreSQL 17

Query | Query History

```sql
33
34   -- Duplicate
35   SELECT campaign_name, COUNT(*) AS occurrences
36   FROM ad_campaign_data
37   GROUP BY campaign_name
38   HAVING COUNT(*) > 1;
39
40   --Inconsistent
41   SELECT LOWER(TRIM(delivery_status)) AS normalized_status, COUNT(*) AS count
42   FROM ad_campaign_data
43   GROUP BY LOWER(TRIM(delivery_status))
44   ORDER BY count DESC;
```

Data Output | Messages | Notifications

normalized_status text	count bigint
completed	86
inactive	37
active	10
[null]	7
not_delivering	4
archived	3
recently_completed	1

Showing rows: 1 to 7 | Page No: 1 of 1

Total rows: 7   Query complete 00:00:00.188   CRLF   Ln 4

Cohort/postgres@PostgreSQL 17

Query   Query History

```sql
1  SELECT
2    COUNT(*) FILTER (
3      WHERE cohort_id IS NULL
4        OR TRIM(cohort_id::TEXT) = ''
5        OR LOWER(TRIM(cohort_id::TEXT)) IN ('null', 'none', '-', 'n/a')
6    ) AS missing_cohort_id,
7
8    COUNT(*) FILTER (
9      WHERE cohort_code IS NULL
10       OR TRIM(cohort_code) = ''
11       OR LOWER(TRIM(cohort_code)) IN ('null', 'none', '-', 'n/a')
12   ) AS missing_cohort_code,
13
14   COUNT(*) FILTER (
15     WHERE start_date IS NULL
16       OR TRIM(start_date::TEXT) = ''
17       OR LOWER(TRIM(start_date::TEXT)) IN ('null', 'none', '-', 'n/a')
18   ) AS missing_start_date,
19
20   COUNT(*) FILTER (
21     WHERE end_date IS NULL
22       OR TRIM(end_date::TEXT) = ''
23       OR LOWER(TRIM(end_date::TEXT)) IN ('null', 'none', '-', 'n/a')
24   ) AS missing_end_date,
25
```

Data Output   Messages   Notifications

Showing rows: 1 to 1   Page No: 1   of 1

missing_cohort_id bigint	missing_cohort_code bigint	missing_start_date bigint	missing_end_date bigint	missing_size bigint
2	2	2	2	2

Total rows: 1   Query complete 00:00:00.204   CRLF   Ln 33, Col 1

LearnerOpportunity/postgres@PostgreSQL 17

Query   Query History

```sql
1  SELECT
2    COUNT(*) FILTER (
3      WHERE enrollment_id IS NULL
4        OR TRIM(enrollment_id::TEXT) = ''
5        OR LOWER(TRIM(enrollment_id::TEXT)) IN ('null', 'none', '-', 'n/a')
6    ) AS missing_enrollment_id,
7
8    COUNT(*) FILTER (
9      WHERE learner_id IS NULL
10       OR TRIM(learner_id::TEXT) = ''
11       OR LOWER(TRIM(learner_id::TEXT)) IN ('null', 'none', '-', 'n/a')
12   ) AS missing_learner_id,
13
14   COUNT(*) FILTER (
15     WHERE assigned_cohort IS NULL
16       OR TRIM(assigned_cohort::TEXT) = ''
17       OR LOWER(TRIM(assigned_cohort::TEXT)) IN ('null', 'none', '-', 'n/a')
18   ) AS missing_assigned_cohort,
19
20   COUNT(*) FILTER (
21     WHERE apply_date IS NULL
22       OR TRIM(apply_date::TEXT) = ''
23       OR LOWER(TRIM(apply_date::TEXT)) IN ('null', 'none', '-', 'n/a')
24   ) AS missing_apply_date,
25
```

Data Output   Messages   Notifications

Showing rows: 1 to 1   Page No: 1   of 1

missing_enrollment_id bigint	missing_learner_id bigint	missing_assigned_cohort bigint	missing_apply_date bigint	missing_status bigint
0	0	13318	188	186

Total rows: 1   Query complete 00:00:01.336   CRLF   Ln 25, C

Screenshot 1 — LearnerOpportunity/postgres@PostgreSQL 17

```sql
--Duplicate
SELECT enrollment_id, COUNT(*) AS occurrences
FROM public.staging_learner_opportunities
GROUP BY enrollment_id
HAVING COUNT(*) > 1;

--Inconsistent
SELECT LOWER(TRIM(status)) AS normalized_status, COUNT(*) AS count
FROM public.staging_learner_opportunities
GROUP BY LOWER(TRIM(status))
ORDER BY count DESC;
```

	normalized_status (text)	count (bigint)
1	1070	76109
2	1030	12236
3	1055	11471
4	1120	9048

Showing rows: 1 to 11 — Page No: 1 of 1
Total rows: 11   Query complete 00:00:00.338

Screenshot 2 — Cognito/postgres@PostgreSQL 17

```sql
SELECT
  COUNT(*) FILTER (
    WHERE user_id IS NULL
      OR TRIM(user_id::TEXT) = ''
      OR LOWER(TRIM(user_id::TEXT)) IN ('null', 'none', '-', 'n/a')
  ) AS missing_user_id,

  COUNT(*) FILTER (
    WHERE email IS NULL
      OR TRIM(email) = ''
      OR LOWER(TRIM(email)) IN ('null', 'none', '-', 'n/a')
  ) AS missing_email,

  COUNT(*) FILTER (
    WHERE gender IS NULL
      OR TRIM(gender) = ''
      OR LOWER(TRIM(gender)) IN ('null', 'none', '-', 'n/a')
  ) AS missing_gender,

  COUNT(*) FILTER (
    WHERE user_create_date IS NULL
      OR TRIM(user_create_date::TEXT) = ''
      OR LOWER(TRIM(user_create_date::TEXT)) IN ('null', 'none', '-', 'n/a')
  ) AS missing_user_create_date,
```

	missing_user_id	missing_email	missing_gender	missing_user_create_date	missing_user_last_modified_date	missing_birthdate	missing_city	missing_zip	missing_state
1	0	0	42862	0	0	42862	42871	42947	42973

Showing rows: 1 to 1 — Page No: 1 of 1
Total rows: 1   Query complete 00:00:01.866

Screenshot 3 — Cognito/postgres@PostgreSQL 17

```sql
--Duplicate
SELECT email, COUNT(*) AS occurrences
FROM public.staging_users
GROUP BY email
HAVING COUNT(*) > 1
ORDER BY occurrences DESC;

--_inconsistent
SELECT DISTINCT TRIM(LOWER(gender)) AS cleaned_gender, COUNT(*) AS count
FROM public.staging_users
GROUP BY cleaned_gender
ORDER BY count DESC;
```

	cleaned_gender (text)	count (bigint)
1	male	49344
2	[null]	42862
3	female	36515
4	don%27t want to specify	361

Showing rows: 1 to 5 — Page No: 1 of 1
Total rows: 5   Query complete 00:00:00.884

	gender character varying (50) 🔒	count bigint 🔒
1	Don't want to specify	43223
2	Female	36515
3	Male	49344
4	Other	96

Data Output   Messages   Notifications   Showing rows: 1 to 1000   Page No: 1   of 130

	learner_id [PK] text	country text	degree text	institution text
1	Learner#00011c80-0c5c-4601-9696-b3ca787e26...	Bangladesh	Not specified	Institution not provided
2	Learner#000141a7-4c82-401f-a2e6-dd12b4b260...	Nigeria	Not specified	Institution not provided
3	Learner#0acf3501-57a8-4585-91e3-6bbb89d3c8...	Ghana	Not specified	Institution not provided
4	Learner#0001ca2c-7bec-4a33-833c-b844a29f4d...	Nigeria	Graduate Student	Nasarawa State University, Keff
5	Learner#0003bed9-d9d9-49a7-a755-a9562aaa0d...	Pakistan	Graduate Student	CTTI College KP Campus
6	Learner#0004295c-717e-4953-b3bf-fff2daf0e903	Nigeria	Undergraduate Student	Caleb University
7	Learner#00049a81-94a9-4b25-92ed-62d017f3b6...	Philippines	Undergraduate Student	Our lady of fatima university
8	Learner#0005243a-868e-41c4-8d53-3a61649236...	Pakistan	Graduate Student	PGMI
9	Learner#00064ab2-0e32-4b50-8d8c-5c6b0d0993...	India	Not specified	Institution not provided
10	Learner#000687a6-b0e3-4f21-9df8-98ad0db2ff7e	India	Not specified	Institution not provided
11	Learner#000693c5-fafb-4d03-867d-ac6c825db75a	India	Not specified	Institution not provided
12	Learner#00070ea9-aa88-4018-8af2-aec93139ff21	Nigeria	Not specified	Institution not provided
13	Learner#00075a4d-e755-4ea6-9ef2-e4119495ad...	Nigeria	Graduate Student	Durham University
14	Learner#00084381-3917-4d03-9639-0a501aa68c...	India	High School Student	Saket International School

Query   Query History

```sql
CREATE TABLE learners (
    learner_id   TEXT PRIMARY KEY,
    country      TEXT,
    degree       TEXT,
    institution  TEXT,
    major        TEXT
);
select * from "learners"

UPDATE learners
SET
   degree = COALESCE(degree, 'Not specified'),
   institution = COALESCE(institution, 'Institution not provided'),
   major = COALESCE(major, 'Major not specified');
```

```sql
UPDATE learners
SET country = COALESCE(country, 'City not specified');
select * from "learners"


UPDATE learners
SET country = 'Country is not specified'
WHERE country = 'City not specified';
```

Query Query History

```sql
1  SELECT
2      COUNT(*) FILTER (WHERE opportunity_id IS NULL)      AS null_opportunity_id,
3      COUNT(*) FILTER (WHERE opportunity_name IS NULL)    AS null_opportunity_name,
4      COUNT(*) FILTER (WHERE category IS NULL)            AS null_category,
5      COUNT(*) FILTER (WHERE opportunity_code IS NULL)    AS null_opportunity_code,
6      COUNT(*) FILTER (WHERE tracking_questions IS NULL)  AS null_tracking_questions
7  FROM opportunity;
8
9  UPDATE opportunity
10 SET tracking_questions = 'Don''t know now'
11 WHERE tracking_questions IS NULL;
12
```

Data Output Messages Notifications

Showing rows: 1 to 1   Page No: 1   of 1

	null_opportunity_id bigint	null_opportunity_name bigint	null_category bigint	null_opportunity_code bigint	null_tracking_questions bigint
1	0	0	0	0	0

	cohort_key [PK] text	size integer	start_date date	end_date date
65	Cohort#_BMB7W...	1000	2023-07-22	2023-11-15
66	Cohort#_BTNDYST	1500	2025-02-20	2025-02-20
67	Cohort#_BOP8K...	800	2024-03-09	2024-03-09
68	Cohort#_B7FTJOZ	200	2023-07-22	2023-07-22
69	Cohort#_BPTM9...	800	2024-03-09	2024-03-09
70	Cohort#_BZM76...	1000	2024-10-27	2026-02-02
71	Cohort#_BW0A3...	800	2024-07-03	2024-07-03
72	Cohort#_BKCDFG7	800	2024-07-03	2024-07-03
73	Cohort#_BGW5V...	500	2024-07-03	2024-10-27
74	Cohort#_B6QHKJ0	1600	2024-10-27	2025-02-20
75	Cohort#_BWRW...	1000	2023-11-15	2024-03-09
76	Cohort#_BKQ872R	800	2024-10-27	2024-10-27
77	Cohort#_B6XIIFV	1500	2024-07-03	2024-10-27

```sql
Query   Query History
 1 v  ALTER TABLE cohortraw
 2      ALTER COLUMN start_date TYPE DECIMAL;
 3
 4 v  ALTER TABLE cohortraw
 5      ALTER COLUMN end_date TYPE DECIMAL;
 6
 7 v  Select * from "cohortraw"
 8
 9      ALTER TABLE cohortraw
10      ADD COLUMN cohort_key TEXT;
11
12 v  UPDATE cohortraw
13      SET cohort_key = cohort_id || '_' || cohort_code;
14
15 v  ALTER TABLE cohortraw
16      DROP CONSTRAINT IF EXISTS cohortraw_pkey;
17
18 v  ALTER TABLE cohortraw
19      ADD CONSTRAINT cohortraw_pkey PRIMARY KEY (cohort_key);
20
21 v  ALTER TABLE cohortraw
22      DROP COLUMN cohort_id,
23      DROP COLUMN cohort_code;
```

```sql
   ALTER TABLE cohortraw ADD COLUMN start_date_converted TIMESTAMP;


   ALTER TABLE cohortraw ADD COLUMN end_date_converted TIMESTAMP;

v  UPDATE cohortraw
   SET start_date_converted = to_timestamp(start_date / 1000),
       end_date_converted = to_timestamp(end_date / 1000);

   ALTER TABLE cohortraw DROP COLUMN start_date;
   ALTER TABLE cohortraw DROP COLUMN end_date;
   ALTER TABLE cohortraw RENAME COLUMN start_date_converted TO start_date;
   ALTER TABLE cohortraw RENAME COLUMN end_date_converted TO end_date;

   ALTER TABLE cohortraw RENAME COLUMN size TO size_old;
   ALTER TABLE cohortraw ADD COLUMN size INTEGER;
   UPDATE cohortraw SET size = size_old;
   ALTER TABLE cohortraw DROP COLUMN size_old;

v  ALTER TABLE cohortraw
   ADD COLUMN start_date_only DATE,
   ADD COLUMN end_date_only DATE;
```

```
   > Domain   Query   Query History
   > FTS Co    20 v  SELECT
   > FTS Dic   21        (SELECT ad_account_name
   > A* FTS Pa 22          FROM public.marketing_campaigns
   > FTS Te    23          GROUP BY ad_account_name
   > Foreign   24          ORDER BY COUNT(*) DESC
   > Functic   25          LIMIT 1) AS mode_ad_account_name,
   > Materia   26
   > Operati   27        (SELECT campaign_name
   > Proced    28          FROM public.marketing_campaigns
   v 1-3 Sequer 29          GROUP BY campaign_name
   v Tables   30          ORDER BY COUNT(*) DESC
     > cogr   31          LIMIT 1) AS mode_campaign_name,
     v coho   32
       v C    33        (SELECT delivery_status
              34          FROM public.marketing_campaigns
              35          GROUP BY delivery_status
              36          ORDER BY COUNT(*) DESC
              37          LIMIT 1) AS mode_delivery_status,
              38
              39        (SELECT delivery_level
              40          FROM public.marketing_campaigns
     > C      41          GROUP BY delivery_level
              42          ORDER BY COUNT(*) DESC
```

```
66    WHERE
67       delivery_status IS NULL OR
68       delivery_level IS NULL OR
69       result_type IS NULL;
70  v UPDATE public.marketing_campaigns
71    SET result_type = 'Website applications submitted'
72    WHERE result_type IS NULL;
73
74  v SELECT
75       AVG(reach) AS avg_reach,
76       AVG(outbound_clicks) AS avg_outbound_clicks,
77       AVG(landing_page_views) AS avg_landing_page_views,
78       AVG(cost_per_result) AS avg_cost_per_result,
79       AVG(cpc) AS avg_cpc
80    FROM public.marketing_campaigns;
81
82  v UPDATE public.marketing_campaigns
83    SET reach = 140000.0
84    WHERE reach IS NULL;
85
86  v UPDATE public.marketing_campaigns
87    SET outbound_clicks = 3000.0
88    WHERE outbound_clicks IS NULL;
```

	opportunity_name text	category character varying (50)	opportunity_code character varying (20)	tracking_questions text
1	Business Consulting Early Internship	Internship	IBLCQ1D	{code:QAGLZJ1,is_required_for_badge_i
2	Green Technology Solutions Workshop	Event	EZI9BVU	{serial_number:1,is_required_for_badge_
3	Project Management Early Internship	Internship	I448557	{serial_number:11,code:QK2Q8MR,is_rec
4	Secrets to Operational Excellence Course	Course	URMUHH7	{serial_number:1,is_required_for_badge_
5	Project Management Associate Early Internship	Internship	IP5EYAL	{code:Q34K35A,is_required_for_badge_i
6	Data Visualization Early Internship	Internship	I352815	{serial_number:7,code:QN95HXN,is_requ
7	Work 2030: Adapting to Tomorrow%27s Workplace	Event	EBOGN8J	{serial_number:1,is_required_for_badge_
8	Jump Start: Developing your Emotional Intelligence	Course	URFOJWP	{is_required_for_badge_award:true,code
9	CloudCraft: Mastering Cloud Solutions	Event	E8JV0Q5	{serial_number:1,is_required_for_badge_
10	Project Management Early Internship	Internship	I448557	{serial_number:11,code:QK2Q8MR,is_rec

Master table Query--

```
1  v CREATE TABLE master_learners AS
2    SELECT
3         l.learner_id,
4         l.country,
5         l.degree,
6         l.institution,
7         l.major,
8         lo.opportunity_id,
9         lo.assigned_cohort,
10        lo.status,
11        lo.apply_date
12   FROM learners l
13   LEFT JOIN learner_opportunity lo
14     ON l.learner_id = lo.learner_id;
```

```
15
16  v CREATE TABLE learner_learneropp_opportunity AS
17    SELECT
18         ml.*,
19         o.opportunity_name,
20         o.category,
21         o.opportunity_code,
22         o.tracking_questions
23    FROM master_learners ml
24    LEFT JOIN opportunity o
25      ON ml.opportunity_id::TEXT = o.opportunity_id::TEXT;
26
```

Table 1: learner table

learner_id (text)	country (text)	degree (text)	institution (text)
Learner#eca1b97a-7760-407f-b501-9eb109eb84...	Nigeria	Undergraduate Student	Nnamdi Azikiwe university
Learner#eca4e2ce-c9c3-4755-9abb-8a57858e48...	Pakistan	Graduate Student	GC University Lahore Pakistan
Learner#eca551bb-88b0-4dbc-ace8-d0b60c8c99...	Ghana	Undergraduate Student	University of Health And Allied Sciences
Learner#ecad8f47-e2de-45c5-b6ea-4efc0c0a4417	Nigeria	Graduate Student	Anambra State University, Awka
Learner#ecaefbd7-2c89-446c-919e-0a9d31faf122	Kenya	Graduate Student	Kisii University
Learner#ecaf330d-fea2-4cfc-ba9c-bcf53e0d63b4	India	Graduate Student	Illinois Institute of Technology
Learner#ecb0bac7-8863-4b56-9fae-283fa1b822...	Kenya	Graduate Student	JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TI

Table 2: pgAdmin 4 — Data Output

Showing rows: 1 to 1000 — Page No: 1 of 58

	major (text)	opportunity_id (text)	assigned_cohort (text)	status (integer)	apply_date (text)
1	Agriculture and Forestry	Opportunity#00000000101BRJY9TVWC6XG9BH	BP6VD9E	1070	2024-03-30
2	Bachelor of Business Administration	Opportunity#0000000010AQBQRE1W8XBYDP...	B8NALTQ	1070	2025-02-10
3	Medical Laboratory Science	Opportunity#000000000GHB4N83QX9KJM48K2	BMYSZAZ	1070	2024-09-19
4	Sociology	Opportunity#00000000102S9XBS7SM4KB0Q57	BLJK6ZC	1070	2024-12-10
5	Project Management	Opportunity#0000000010SAZXDAE05AN2CGAN	BIYM5IR	1070	2025-01-18
6	Computer Science	Opportunity#000000000GWQAXC5X45C2MHJ...	B8L5EQI	1120	2023-04-28
7	Architecture	Opportunity#0000000010E5BC31XKZK44K880	BZLDX4E	1070	2024-12-09
8	Marketing	Opportunity#00000000100JPM3A0WJ85T68M5	BJZ6HXM	1070	2024-12-25
9	Radiography	Opportunity#0000000010Q3NK8RDGVXQPW9...	BFEG6ER	1070	2024-11-16
10	Project Management	Opportunity#000000000GHB4N83QX9KJM48K2	BQH2HGO	1070	2024-05-09
11	Civil Engineering	Opportunity#0000000010SAZXDAE05AN2CGAN	BIYM5IR	1070	2024-12-28
12	Bachelor of Technology	Opportunity#00000000106YSVYZ2WNJ29PZZN	B7PFXPN	1070	2025-02-06
13	Finance and Banking	Opportunity#0000000010CD3CAK3FZK9VXKRM	BULL239	1070	2024-03-30
14	Advertising	Opportunity#0000000010CD3CAK3FZK9VXKRM	BLETLXU	1070	2025-02-10

Total rows: 57965 — Query complete 00:00:00.386 — CRLF — Ln 1, Col 1

Table 3: opportunity table

	opportunity_name (text)	category (character varying (50))	opportunity_code (character varying (20))	tracking_questions (text)
1	Business Consulting Early Internship	Internship	IBLCQ1D	{code:QAGLZJ1,is_required_for_badge_a
2	Green Technology Solutions Workshop	Event	EZI9BVU	{serial_number:1,is_required_for_badge_
3	Project Management Early Internship	Internship	I448557	{serial_number:11,code:QK2Q8MR,is_re
4	Secrets to Operational Excellence Course	Course	URMUHH7	{serial_number:1,is_required_for_badge_
5	Project Management Associate Early Internship	Internship	IP5EYAL	{code:Q34K35A,is_required_for_badge_a
6	Data Visualization Early Internship	Internship	I352815	{serial_number:7,code:QN95HXN,is_requ
7	Work 2030: Adapting to Tomorrow%27s Workplace	Event	EBOGN8J	{serial_number:1,is_required_for_badge_
8	Jump Start: Developing your Emotional Intelligence	Course	URFOJWP	{is_required_for_badge_award:true,code
9	CloudCraft: Mastering Cloud Solutions	Event	E8JV0Q5	{serial_number:1,is_required_for_badge_
10	Project Management Early Internship	Internship	I448557	{serial_number:11,code:QK2Q8MR,is_re

	column_name name	data_type character varying	is_nullable character varying (3)
1	learner_id	text	YES
2	country	text	YES
3	degree	text	YES
4	institution	text	YES
5	major	text	YES
6	opportunity_id	text	YES
7	assigned_cohort	text	YES
8	status	integer	YES
9	apply_date	text	YES
10	opportunity_name	text	YES
11	category	character varying	YES
12	opportunity_code	character varying	YES
13	tracking_questions	text	YES

Total rows: 13   Query complete 00:00:00.109

```sql
SELECT
    COUNT(*) AS total_rows,
    COUNT(*) - COUNT(learner_id) AS null_learner_id,
    COUNT(*) - COUNT(country) AS null_country,
    COUNT(*) - COUNT(degree) AS null_degree,
    COUNT(*) - COUNT(institution) AS null_institution,
    COUNT(*) - COUNT(major) AS null_major,
    COUNT(*) - COUNT(opportunity_id) AS null_opportunity_id,
    COUNT(*) - COUNT(status) AS null_status,
    COUNT(*) - COUNT(apply_date) AS null_apply_date,
    COUNT(*) - COUNT(opportunity_name) AS null_opportunity_name,
    COUNT(*) - COUNT(category) AS null_category,
    COUNT(*) - COUNT(opportunity_code) AS null_opportunity_code,
    COUNT(*) - COUNT(tracking_questions) AS null_tracking_questions
FROM learner_learneropp_opportunity;
```

Data Output   Messages   Notifications

Showing rows: 1 to 1   Page No: 1   of 1

null_learner_id bigint	null_country bigint	null_degree bigint	null_institution bigint	null_major bigint	null_opportunity_id bigint	null_status bigint	null_apply_date bigint	null_opportunity_name bigint
0	0	0	0	0	0	0	0	0

Showing rows: 1 to 1   Page No: 1   of 1

null_opportunity_name bigint	null_category bigint	null_opportunity_code bigint	null_tracking_questions bigint
0	0	0	0