

Data Visualization Associate powered by **Excelerate platform**

Objective

The primary objective of this project is to analyze and integrate multiple datasets—including user data and marketing data—from the Excelerate learning platform, perform an in-depth data quality assessment, and develop a fully functional dashboard using PostgreSQL for data processing and Looker Studio for visualization.

Team Members and Roles

- ***Team Lead***

Abigail Sunday Udo

- Represents the team to the sponsor on email and calls
- Minimizes communication errors and leads team coordination

- ***Project Manager***

Siddharth Ramgundam

- Provides team guidance and draws insights from members
- Keeps project execution on track

- ***Project Scribe***

Adithya P K

- Records meeting minutes and shares notes and assignments
- Assists Team Lead in drafting emails and sponsor communication

- ***Project Lead***

Adithya Thirumala Reddy Vajrала

- Holds the team accountable for deadlines
- Ensures all project deliverables are completed on time

- ***Data Strategist***

Abhay Rathore

- Conducts data analysis and strategy planning
- Collaborates with the team on analytical execution

Data Quality Exploratory Data Analysis (EDA)

Dataset Overview

- **Learner Data** – User profiles, education background, and enrollment status
- **Cohort Data** – Details of cohort-based learning programs
- **Marketing Campaigns** – Campaign performance and spending metrics
- **Tracking & Content Data** – Learner interactions and engagement
- **User Demographics** – Gender, location, birthdate, and other attributes

Key Data Quality Issues Identified

General Data Quality Observations

- Undocumented status codes present in learner records
- Duplicate entries based on learner ID and application date
- Learners assigned to multiple cohorts simultaneously
- Inconsistent NULL handling: both blank and literal "NULL" values
- Malformed data rows and inconsistent or corrupted ID formats

Inconsistencies & Duplicates

- Repeated campaign names (e.g., "Copy 3", "Copy 4") complicate performance tracking
- Demographic gaps: missing gender, birthdates, and postal codes
- Tracking data fields contain irregular naming and missing question values

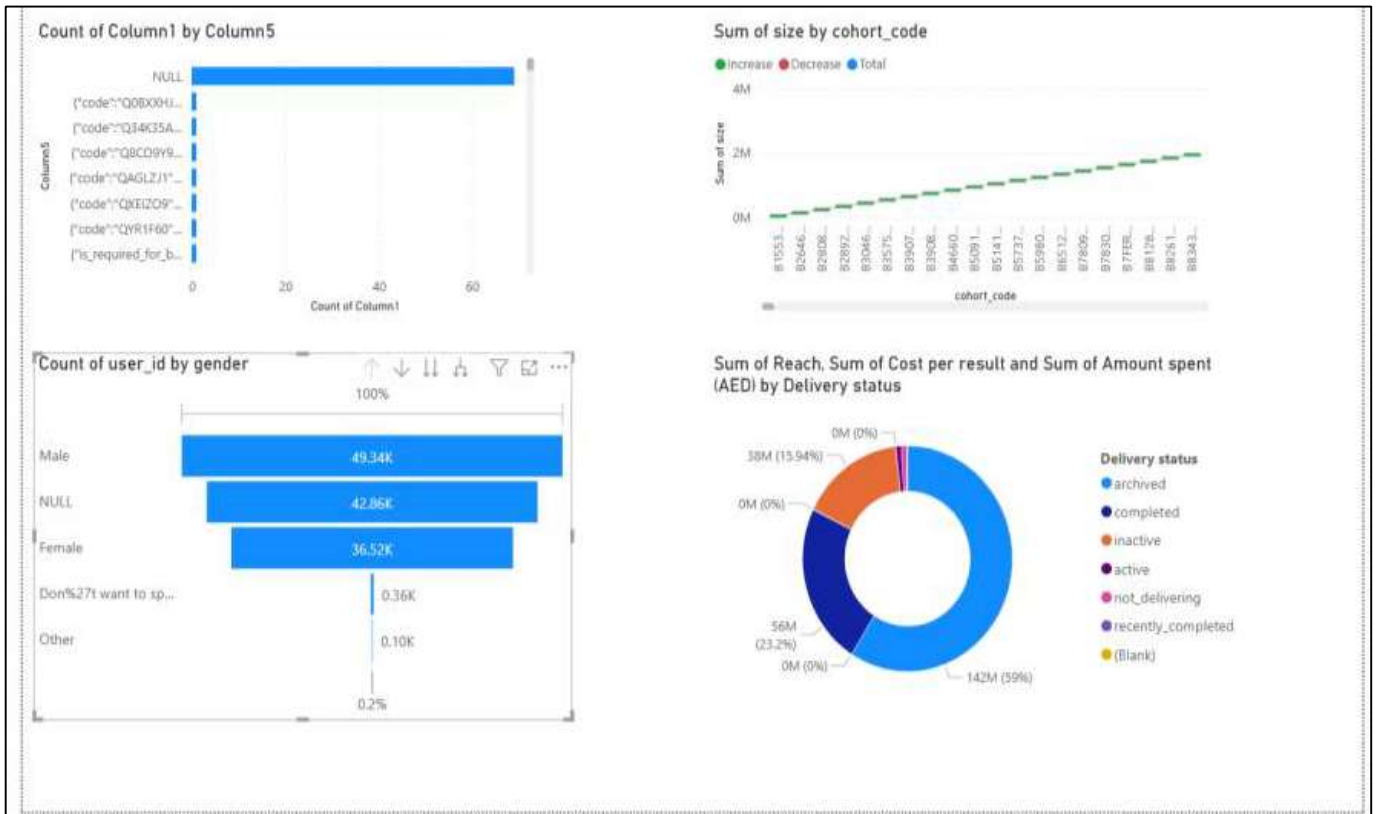
Outliers & Anomalies

- Cohorts with unrealistically large sizes or durations of 0 days
- Campaigns with high spending but minimal engagement results
- Erratic cost-per-result values across campaigns
- Epoch-based timestamps requiring human-readable conversion

Key Insights

- **Improve Data Governance:** Apply rules to check data quality, remove duplicates, and use consistent naming.
- **Optimize Processes:** Set clear checks for data, manage marketing campaigns better, and fix import errors.

- **Enhance Data Quality:** Clean and standardize demographic data and other fields for



```

File Object Tools Edit View Window Help
Object Explorer
  Functions
  Materialized Views
  Operators
  Procedures
  Sequences
  Tables (6)
    Cognito_Raw2
    cohortraw
    learneropportunity
    learners
    marketing_campaigns
      Columns
      Constraints
      Indexes
      RLS Policies
      Rules
      Triggers
    opportunities
      Columns
      Constraints
      Indexes
      RLS Policies
      Rules
  Execelerate/postgres@PostgreSQL 17
    The session is idle and there is no current transaction.
    Query Query History
    Scratch Pad X
    1 --mode for each column
    2 SELECT
    3   -- Mode of learner_id
    4   (SELECT learner_id
    5    FROM learners
    6    WHERE learner_id IS NOT NULL AND learner_id <> ''
    7    GROUP BY learner_id
    8    ORDER BY COUNT(*) DESC
    9    LIMIT 1) AS mode_learner_id,
    10
    11   -- Mode of country
    12   (SELECT country
    13    FROM learners
    14    WHERE country IS NOT NULL AND country <> ''
    15    GROUP BY country
    16    ORDER BY COUNT(*) DESC
    17    LIMIT 1) AS mode_country,
    18
    19   -- Mode of degree
    20   (SELECT degree
    21    FROM learners
    22    WHERE degree IS NOT NULL AND degree <> ''
    23    GROUP BY degree
  
```

Execelate/postgres@PostgreSQL 17

Query Query History Scratch Pad X

```
1 -- Replace NULL and blank assigned_cohort
2 UPDATE learneropportunity
3 SET assigned_cohort = (
4     SELECT assigned_cohort
5     FROM learneropportunity
6     WHERE assigned_cohort IS NOT NULL AND TRIM(assigned_cohort) <> ''
7     GROUP BY assigned_cohort
8     ORDER BY COUNT(*) DESC
9     LIMIT 1
10 )
11 WHERE assigned_cohort IS NULL OR TRIM(assigned_cohort) = '';
12
13 -- Replace NULL status
14 UPDATE learneropportunity
15 SET status = (
16     SELECT status
17     FROM learneropportunity
18     WHERE status IS NOT NULL
19     GROUP BY status
20     ORDER BY COUNT(*) DESC
21     LIMIT 1
22 )
23 WHERE status IS NULL;
```

Execelate/postgres@PostgreSQL 17

Query Query History Scratch Pad X

```
1 --count the no. of null values in each column
2 SELECT
3     COUNT(*) FILTER (WHERE enrollment_id IS NULL) AS null_enrollment_id,
4     COUNT(*) FILTER (WHERE learner_id IS NULL) AS null_learner_id,
5     COUNT(*) FILTER (WHERE assigned_cohort IS NULL OR TRIM(assigned_cohort) = '') AS null_assigned_cohort,
6     COUNT(*) FILTER (WHERE apply_date IS NULL) AS null_apply_date,
7     COUNT(*) FILTER (WHERE status IS NULL) AS null_status
8 FROM learneropportunity;
```

Data Output Messages Notifications

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

	null_enrollment_id bigint	null_learner_id bigint	null_or_blank_assigned_cohort bigint	null_apply_date bigint	null_status bigint
1	0	0	13318	188	186

Exploratory & Post-ETL Data Quality Assessment Report

Exploratory Data Assessment (Pre-ETL):

✓ Dataset Overview & Purpose

Dataset Name	Purpose Description
Learner_Raw	Learner demographics and academic background
Opportunity_Raw	Learning/growth opportunity details
CohortRaw	Cohort schedules and details
ad_campaign_data	Ad campaign performance metrics
LearnerOpportunity_Raw	Mapping of learners to opportunities and cohorts
staging_users	User registration and profile info

Dataset Relationships

Learner_Raw.learner_id ↔ LearnerOpportunity_Raw.learner_id

LearnerOpportunity_Raw.assigned_cohort ↔ CohortRaw.cohort_id

staging_users.email ↔ Learner_Raw.email

Pre-ETL Data Issues Identified:

Missing Values

- Text-based "null" in degree, institution, major
- Blank or invalid gender, birthdate, state, zip
- Missing assigned_cohort and apply_date

Duplicate Records

- Duplicate learner_id
- Duplicate (enrollment_id, learner_id) combinations
- Duplicate user_id or email in staging_users

Inconsistent Formats

- Gender: 'male', 'Male', 'M'
- Country names: inconsistent casing
- Campaign data: inconsistent delivery_status, result_type

Orphan Records

- assigned_cohort not found in CohortRaw
- learner_id missing from Learner_Raw
- email in staging_users not matched in Learner_Raw

ETL Transformation Plan:

Convert "null", "none", and blanks to NULL

Normalize casing (e.g., gender, country)

Remove duplicates using ROW_NUMBER() or DISTINCT

Enforce foreign key validation via joins

Correct data types (dates, numeric, categories)

✔ Post-ETL Validation:

Check Type	Purpose	Result
Record Count Validation	Ensure source and master table match	✔ Match confirmed
Duplicate Checks	Eliminate enrollment_id, learner_id dupes	✔ No duplicates found
Missing Data Review	Confirm clean and proper NULL usage	✔ Cleaned successfully
Foreign Key Integrity	Validate cross-table references	✔ All keys matched
Data Type Verification	Ensure correct formats for all fields	✔ Verified

Execelerate/postgres@PostgreSQL 17

Query Query History

```
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

Query Query History

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



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

```
ALTER TABLE cohortraw ADD COLUMN end_date_converted TIMESTAMP;
```

▼ 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;
```

▼ ALTER TABLE cohortraw

```
ADD COLUMN start_date_only DATE,
```

```
ADD COLUMN end_date_only DATE;
```

The screenshot shows a SQL IDE interface. On the left is a table browser with a tree view containing categories like Domain, FTS Co, FTS Dir, FTS Pa, FTS Te, Foreign, Function, Material, Operat, Proced, Sequer, Tables, cogr, coh, C, and C. The main area is a query editor with a 'Query' tab selected, displaying a SQL query. The query is a multi-part SELECT statement. The first part selects 'ad_account_name' from 'public.marketing_campaigns', grouped by 'ad_account_name', ordered by count descending, and limited to 1, aliased as 'mode_ad_account_name'. The second part selects 'campaign_name' from 'public.marketing_campaigns', grouped by 'campaign_name', ordered by count descending, and limited to 1, aliased as 'mode_campaign_name'. The third part selects 'delivery_status' from 'public.marketing_campaigns', grouped by 'delivery_status', ordered by count descending, and limited to 1, aliased as 'mode_delivery_status'. The fourth part selects 'delivery_level' from 'public.marketing_campaigns', grouped by 'delivery_level', ordered by count descending. The query is numbered from 20 to 42. A 'Scratch Pad' tab is visible on the right.

```
20 SELECT  
21 (SELECT ad_account_name  
22 FROM public.marketing_campaigns  
23 GROUP BY ad_account_name  
24 ORDER BY COUNT(*) DESC  
25 LIMIT 1) AS mode_ad_account_name,  
26  
27 (SELECT campaign_name  
28 FROM public.marketing_campaigns  
29 GROUP BY campaign_name  
30 ORDER BY COUNT(*) DESC  
31 LIMIT 1) AS mode_campaign_name,  
32  
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  
41 GROUP BY delivery_level  
42 ORDER BY COUNT(*) DESC
```


Domain
FTS Co
FTS Dic
FTS Pa
FTS Te
Foreignr
Functio
Materiali
Operati
Proced
1.3 Sequer
Tables
cogr
coh
C
C
C
C
C
C

Query
Query History
Scratch Pad

```

66 WHERE
67     delivery_status IS NULL OR
68     delivery_level IS NULL OR
69     result_type IS NULL;
70 UPDATE public.marketing_campaigns
71 SET result_type = 'Website applications submitted'
72 WHERE result_type IS NULL;
73
74 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 UPDATE public.marketing_campaigns
83 SET reach = 140000.0
84 WHERE reach IS NULL;
85
86 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,
2	Green Technology Solutions Workshop	Event	EZ19BVU	{serial_number:1,is_required_for_badge,
3	Project Management Early Internship	Internship	I448557	{serial_number:11,code:QK2Q8MR,is_rei
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_reqi
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_rei

Data Mapping & Dashboard Wireframe

Target Field (Master Table)	Source Table	Source Column Name	Transformation/Logic	Purpose for Dashboard
enrollment_id	staging_learner_opportunities	enrollment_id	None	Unique ID for each enrollment
learner_id	learner	learner_id	None	Join key across learner-related tables
country	learner	country	Clean whitespace / set NULL for 'null' string	Demographics grouping
degree	learner	degree	Clean whitespace / set NULL for 'null' string	Academic profile analysis
institution	learner	institution	Clean whitespace	Filter by school
major	learner	major	Standardize text	Academic interest clustering
apply_date	staging_learner_opportunities	apply_date	Parse to DATE format	Trend line by application date
status	staging_learner_opportunities	status	Convert to lowercase / trim	Funnel chart stage
assigned_cohort	staging_learner_opportunities	assigned_cohort	None	FK to cohort table
cohort_code	cohort_data	cohort_code	None	Filter and label cohorts
cohort_start_date	cohort_data	start_date	Convert from epoch/Excel to DATE	Timeline / Gantt chart
cohort_end_date	cohort_data	end_date	Convert from epoch/Excel to DATE	Duration analysis
opportunity_code	staging_learner_opportunities	opportunity_code	None	FK to opportunity
opportunity_name	opportunity	opportunity_name	None	Labeling charts and tables
opportunity_category	opportunity	category	Standardize values	Group/filter opportunity types
tracking_question	opportunity	tracking_question	None	Qualitative feedback display
gender	staging_users	gender	Lowercase, remove blanks	Demographic grouping
birthdate	staging_users	birthdate	Convert from DD/MM/YYYY to DATE	Calculate age distribution
state	staging_users	state	Standardize text	Geo filter
Ad Account Name	Marketing Campaign Data	Ad Account Name	Standardised Capital Letter	getting Account holder
campaign_name	marketing_data	campaign_name	None	Display campaign performance
Delivery_Status	marketing_data	Delivery_Status	None	Display Delivery Status
amount_spent_aed	marketing_data	amount_spent_aed	Cast to NUMERIC	Budget tracking
results	marketing_data	results	Ensure numeric format	Outcome metric
cost_per_result	marketing_data	cost_per_result	Ensure numeric format	Show the cost per result
cost_per_click	marketing_data	Cost_per_link_click	Ensure numeric format	Show the cost per link click
reach	marketing_data	reach	Ensure numeric format	Awareness metric
landing_page_views	marketing_data	landing_page_views	Numeric, set blanks to 0	Website engagement
outbound_clicks	marketing_data	outbound_clicks	Numeric, treat blanks as 0	CTR metric
result_type	marketing_data	result_type	Standardize text	Conversion type categorization
reporting_starts	marketing_data	reporting_starts	Convert to DATE	Time-based comparison

Wireframe Overview

Total Enrollments

Unique Learners

Total Campaign Reach

Total Results

Cost per Result

Click Through Rate


- Filters:
- Country
 - State
 - Gender
 - Institution
 - Degree
 - Campaign Name
 - Opportunity Type
 - Time Period

 Application Trend Over Time (Line Chart)


 Enrollments by Institution (Bar Chart)

 Funnel by Status (Funnel Chart)

 Demographics by Gender/Country (Pie Chart)

 Opportunities by Category (Bar Chart)

 Budget Over Time (Line Chart)

 Detailed Table View: enrollment_id, learner_id, degree, institution, country, state, status, results, etc.

Dashboard Overview

