



WEEK-3

Development of Data Visualizations in
Looker Studio

EXELERATE
DVA | TEAM 12

Mapping Table Development

Purpose of the Mapping Table

The Mapping Table serves as the foundation for aligning and connecting key data fields across multiple raw datasets. It ensures consistency, accuracy, and completeness when preparing data for visualization. By defining relationships, applying transformation logic, and normalizing values, the Mapping Table enables smooth integration into the Master Table and Looker Studio dashboard.

Data Sources Involved

The Mapping Table pulls from the following datasets:

- learner_raw
- learneropportunity_raw
- opportunity_raw
- cohort_raw

Each source contributes distinct but related fields, and their proper mapping ensures referential integrity and meaningful joins.

Keys & Relationships

Relationship	Type	Condition Applied
• learner_id ↔ enrollment_id	• Inner Join	• enrollment_id in learneropportunity_raw = learner_id in learner_raw
• learner_id ↔ opportunity_id	• Inner Join	• opportunity_id = learner_id in learneropportunity_raw
• assigned_cohort ↔ cohort_code	• Inner Join	• Matching assigned_cohort with cohort_code

Primary Key of the Master Table: **learner_id**
Joins: **INNER JOINS** were used to maintain data quality and filter out unmatched or invalid records.

Column-Level Mappings

Master Column	Source Table	Transformation / Logic Applied
• learner_id	• learner_raw	• TRIM(LOWER(...)) for normalization
• country	• learner_raw	• INITCAP(TRIM(...)) for consistent casing
• degree	• learner_raw	• Standardized terms (e.g., "Unknown" handled)
• enrollment_id	• learneropportunity_raw	• Included only if matched with learner_id from learner_raw
• opportunity_id	• opportunity_raw	• Matched with learner_id from learneropportunity_raw
• opportunity_name	• opportunity_raw	• INITCAP(TRIM(...))
• category	• opportunity_raw	• INITCAP(TRIM(...))
• assigned_cohort	• learneropportunity_raw	• Only values matching cohort_code from cohort_raw
• start_date	• cohort_raw	• Converted from epoch → TO_TIMESTAMP(start_date / 1000)
• end_date	• cohort_raw	• Same conversion as start_date
• size	• cohort_raw	• Used as-is, NULLs handled in filtering

Data Normalization & Cleaning Rules

Key transformations applied during the mapping include:

- **Text Normalization:** All string fields are trimmed of whitespace and standardized to consistent case (e.g., INITCAP for names).

- **Null Handling:** Records with null or critical missing values (e.g., missing `learner_id`, `opportunity_id`, or `assigned_cohort`) were excluded from the Master Table.
- **Date Conversion:** Timestamps in `start_date` and `end_date` were originally stored in epoch format and were converted to readable datetime using `TO_TIMESTAMP(...)`.
- **ID Matching:** Joins were only performed where keys matched exactly to preserve data integrity.

Issue	Resolution
<ul style="list-style-type: none">• Mixed casing & extra spaces	<ul style="list-style-type: none">• Used <code>TRIM()</code> and <code>INITCAP()</code> or <code>LOWER()</code>
<ul style="list-style-type: none">• Null or missing values	<ul style="list-style-type: none">• Excluded rows with NULLs in critical fields
<ul style="list-style-type: none">• Epoch timestamps in <code>cohort_raw</code>	<ul style="list-style-type: none">• Converted using <code>TO_TIMESTAMP(start_date / 1000)</code>
<ul style="list-style-type: none">• Irregular mappings	<ul style="list-style-type: none">• Only included rows with confirmed matches across datasets

Business Logic for Visualization

The mapping logic supports the dashboard in the following ways:

- Links each learner to their **opportunity**, **cohort**, and **application status**.
- Ensures only valid, complete, and meaningful records are included.
- Provides clean dimensions for grouping/filtering in visual elements (e.g., by *degree*, *category*, or *cohort*).
- Lays groundwork for accurate KPIs like total applications, participation rates by category, and time-based trends.

Creating a Dashboard in Looker Studio from PostgreSQL Dataset

Objective:

To visualize data stored in a PostgreSQL database using Google Looker Studio, highlighting key performance indicators (KPIs) using charts such as pie, bar, line, and cards.

Step 1: Connecting Looker Studio to PostgreSQL Database

- Open Looker Studio: Go to <https://lookerstudio.google.com> and sign in with your Google Account.
- Create a New Report: Click on 'Blank Report' or '+ Create > Report'.
- Connect to Data Source: Choose '+ Add data' > Search for 'PostgreSQL'.
- Enter PostgreSQL Credentials:
 - Host: Your database IP/domain (e.g., db.example.com)
 - Port: Default PostgreSQL port is 5432
 - Database Name
 - Username and Password
 - SSL: Enable if your DB requires secure connection
- Authorize Access: Allow Looker Studio to connect to PostgreSQL. Select the desired table/view and click 'Add', then 'Add to Report'.

Step 2: Understanding the Dataset (Fields)

Main Columns:

- learner_id
- country
- degree
- enrollment_id
- opportunity_id
- opportunity_name
- category
- assigned_cohort
- start_date
- end_date
- size

Step 3: Identifying KPIs and Visualization Types

• KPI / Metric	• Column(s)	• Chart Type	• Description
• Total Enrollments	• learner_id	• Scorecard / Card	• Count of distinct learners
• Total Opportunities	• opportunity_id	• Scorecard / Card	• Unique count of opportunities
• Opportunities by Category	• category	• Pie Chart / Bar Chart	• Share of Internships, Courses, Careers
• Learners by Country	• country	• Bar / Geo Chart	• Learner distribution across countries
• Opportunities Over Time	• start_date	• Time Series / Line Chart	• Trend of opportunities started per month/year
• Average Program Size	• size	• Scorecard	• Mean of the size column
• Program Size by Category	• category, size	• Bar Chart	• Compare sizes across categories
• Learners by Degree	• degree	• Pie / Bar Chart	• Distribution of Graduate vs Undergraduate
• Duration of Opportunities	• start_date, end_date	• Table / Calculated Field	• Compute and display duration
• Cohorts and Assignment Count	• assigned_cohort	• Bar Chart	• Number of learners per cohort

<ul style="list-style-type: none"> Average Duration per Category 	<ul style="list-style-type: none"> start_date, end_date, category 	<ul style="list-style-type: none"> Bar Chart 	<ul style="list-style-type: none"> Highlights which program types last longest.
<ul style="list-style-type: none"> Learner Engagement (Total Days) 	<ul style="list-style-type: none"> learner_id, start_date, end_date 	<ul style="list-style-type: none"> Table/Histogram 	<ul style="list-style-type: none"> Total exposure time per learner.
<ul style="list-style-type: none"> Concurrent Programs per Learner 	<ul style="list-style-type: none"> learner_id, start_date, end_date 	<ul style="list-style-type: none"> Bar Chart 	<ul style="list-style-type: none"> Detects learners enrolled in overlapping programs.
<ul style="list-style-type: none"> Opportunity Load by Country 	<ul style="list-style-type: none"> country, opportunity_id 	<ul style="list-style-type: none"> Bar/TreeMap 	<ul style="list-style-type: none"> How many opportunities learners from each country engage in.
<ul style="list-style-type: none"> Repeat Learner Analysis 	<ul style="list-style-type: none"> learner_id 	<ul style="list-style-type: none"> Pie/Bar Chart 	<ul style="list-style-type: none"> Who enrolled in multiple programs.
<ul style="list-style-type: none"> Monthly Enrollment Trend 	<ul style="list-style-type: none"> start_date, learner_id 	<ul style="list-style-type: none"> Line Chart 	<ul style="list-style-type: none"> Tracks monthly signups.
<ul style="list-style-type: none"> Average Size by Country 	<ul style="list-style-type: none"> country, size 	<ul style="list-style-type: none"> Bar Chart 	<ul style="list-style-type: none"> How program sizes vary across regions.
<ul style="list-style-type: none"> Category Preference by Degree 	<ul style="list-style-type: none"> degree, category 	<ul style="list-style-type: none"> Stacked Bar 	<ul style="list-style-type: none"> Undergrad vs Grad preferences.
<ul style="list-style-type: none"> Enrollment Gap per Learner 	<ul style="list-style-type: none"> start_date, end_date, learner_id 	<ul style="list-style-type: none"> Table/Line Chart 	<ul style="list-style-type: none"> Time between program completions.

Step 4: Additional Calculations and Fields in Looker Studio

- Duration (in days): DATE_DIFF(END_DATE, START_DATE)
- Year of Start Date: YEAR(START_DATE)
- Month-Year Field for Trend: FORMAT_DATE("%Y-%m", START_DATE)
- Unique Learners Count: COUNT_DISTINCT(learner_id)

Step 5: Dashboard Layout Tips

- Top Section: Cards showing total enrollments, total opportunities, average size, etc.
- Left Column: Filters – category, country, degree, year
- Middle Section: Time Series + Bar Charts (e.g., enrollments over time)
- Right Column: Pie charts (category, degree distribution)
- Bottom: Detailed table view with opportunity name, start/end, size.

