***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 Vajrala**  
➤ 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

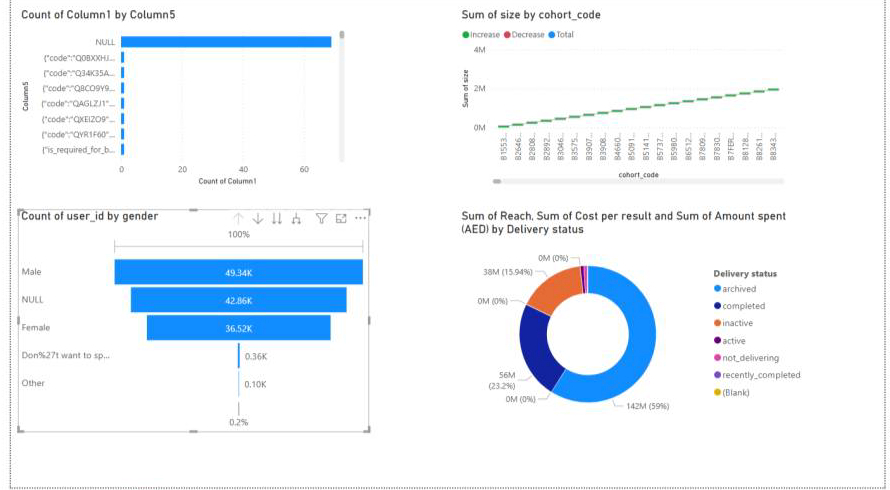
**Epoch-based timestamps** requiring human-readable conversion

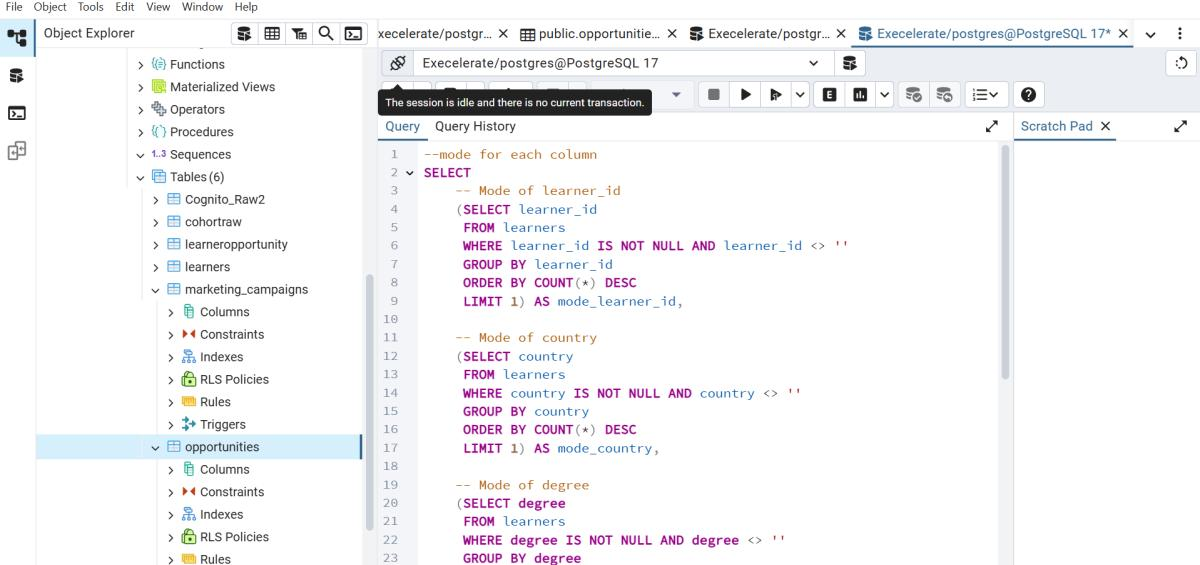
### Key Insights

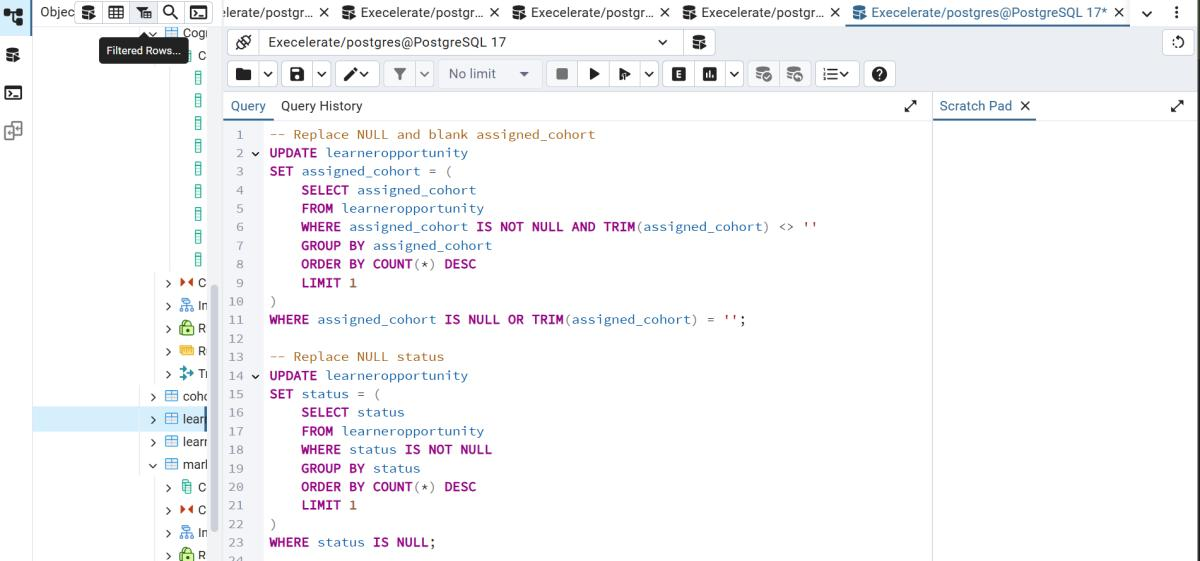
**Improve Data Governance:** Apply rules to check data quality, remove duplicates, and use consistent naming.

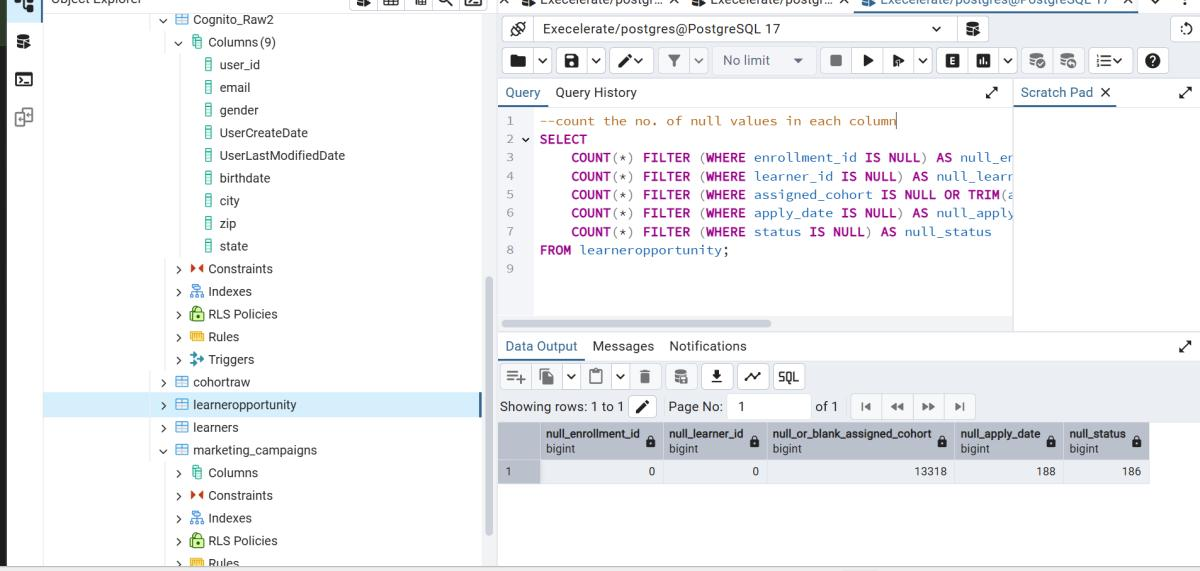
**Optimize Processes:** Set clear checks for data, manage marketing campaigns better, and fix import errors.

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

******







## 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

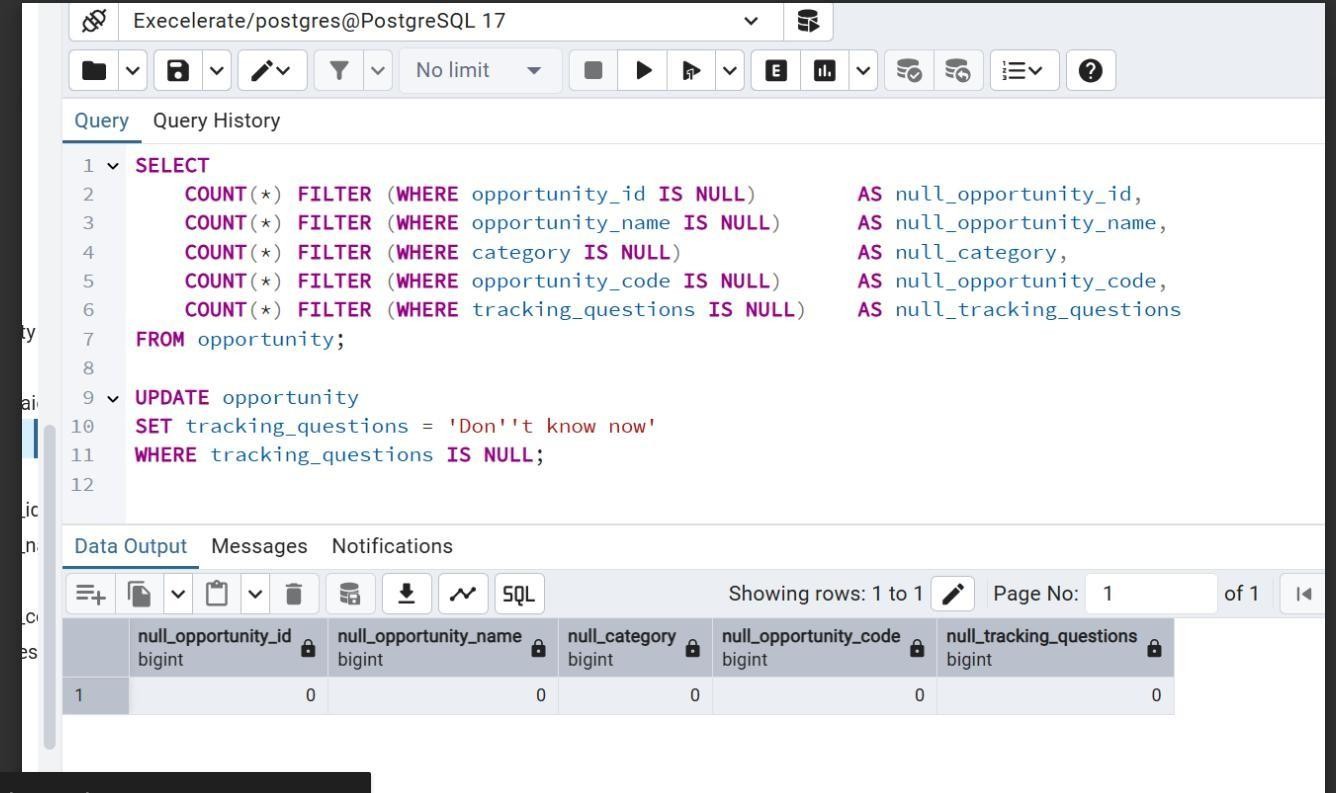
Enforce foreign key validation via joins

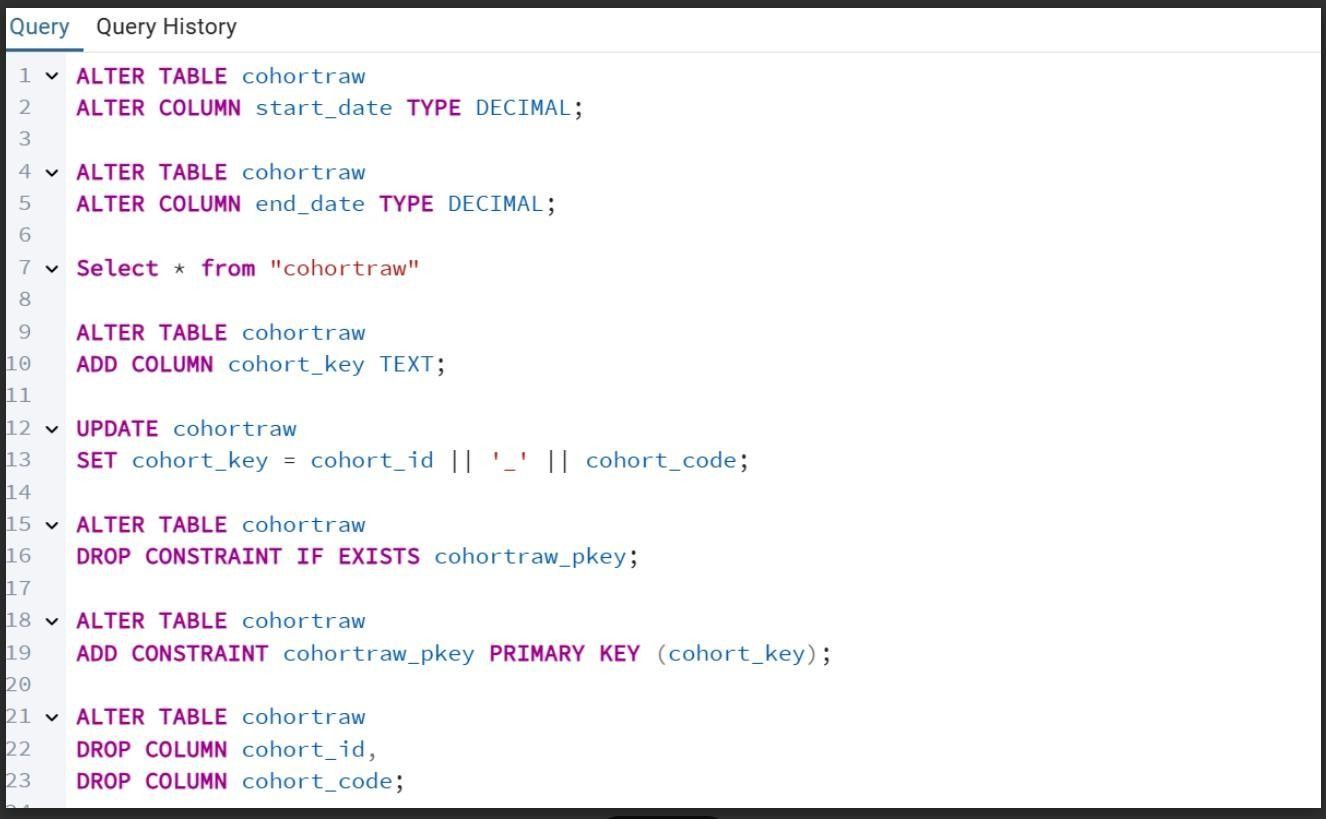
Correct data types (dates, numeric, categories)

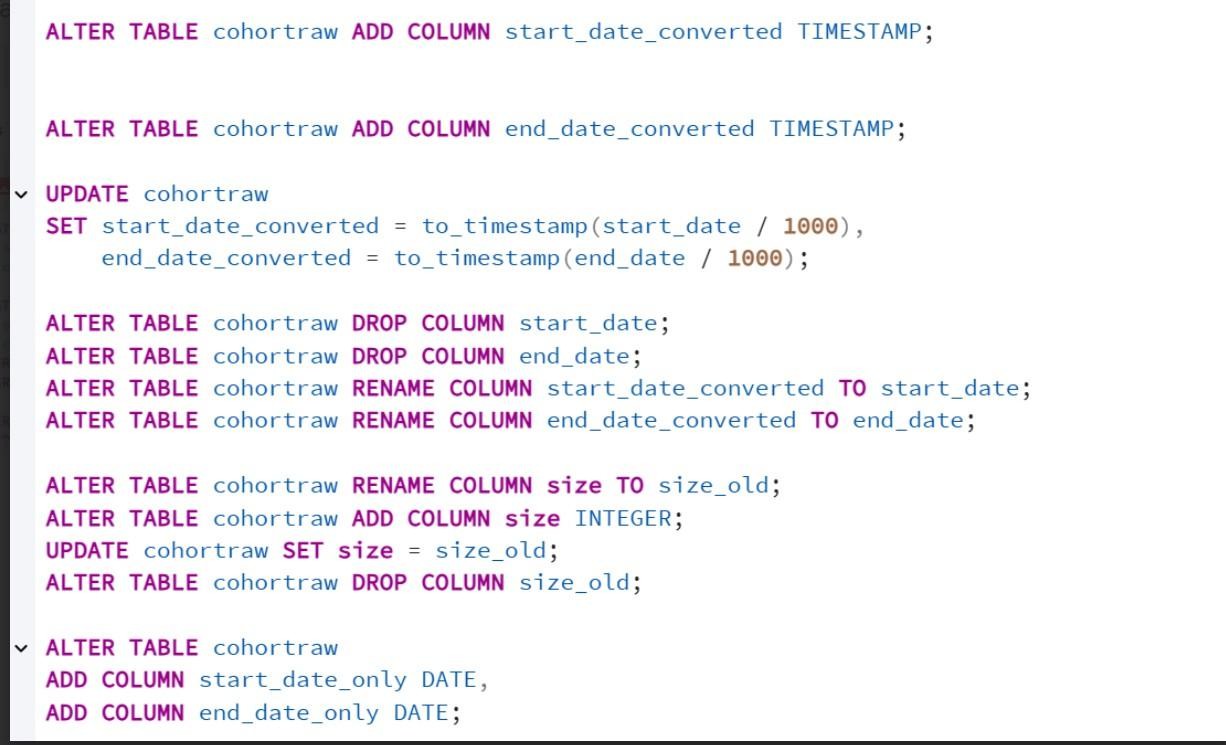
### ✅ Post-ETL Validation:

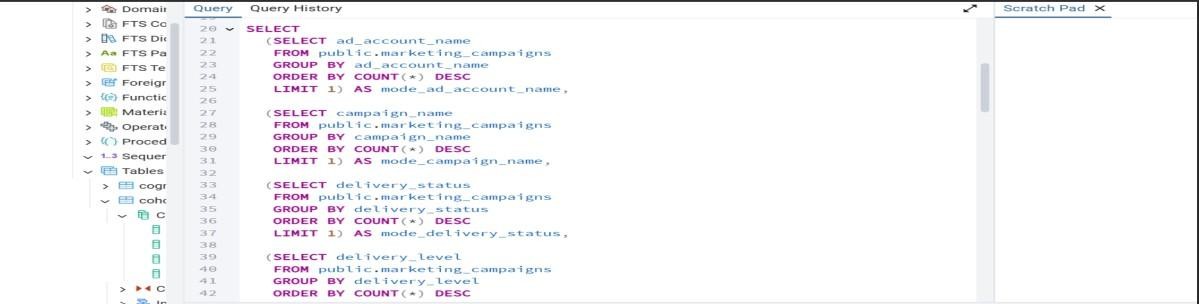
| **Check Type** | **Purpose** | **Result** |
| --- | --- | --- |
| Record Count Validation | Ensure source and master table match | ✅ Match confirmed |
| Duplicate Checks | Eliminate enrollment\_id, learner\_id dupes | ✅ No duplicates found |
| Missing Data Review | Confirm clean and proper NULL usage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

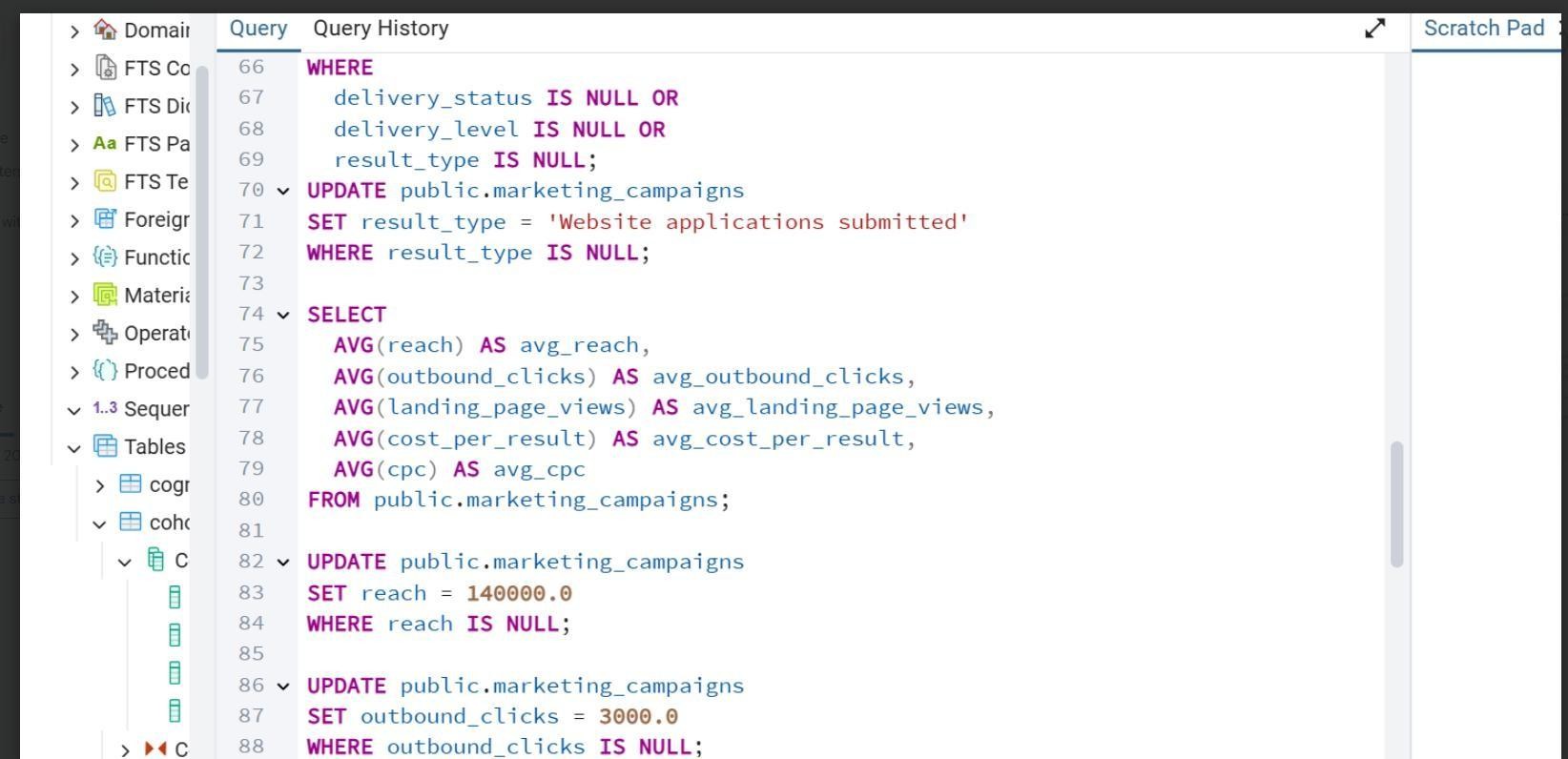






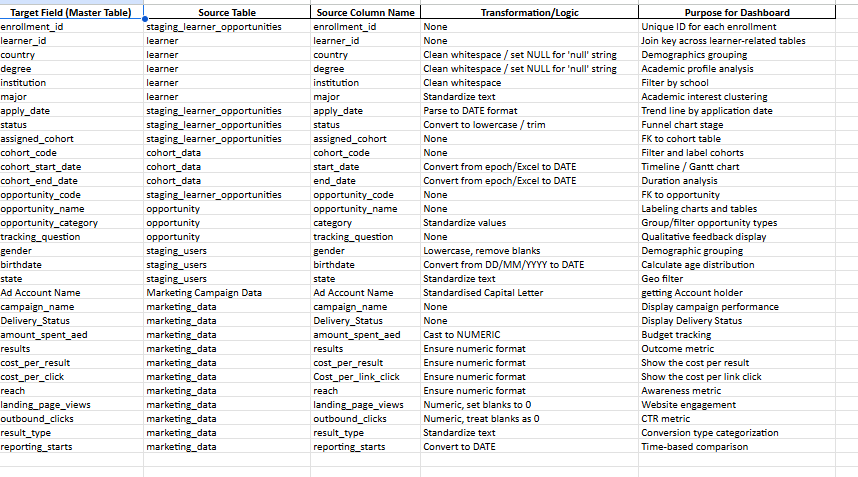








**Data Mapping & Dashboard Wireframe**

******

