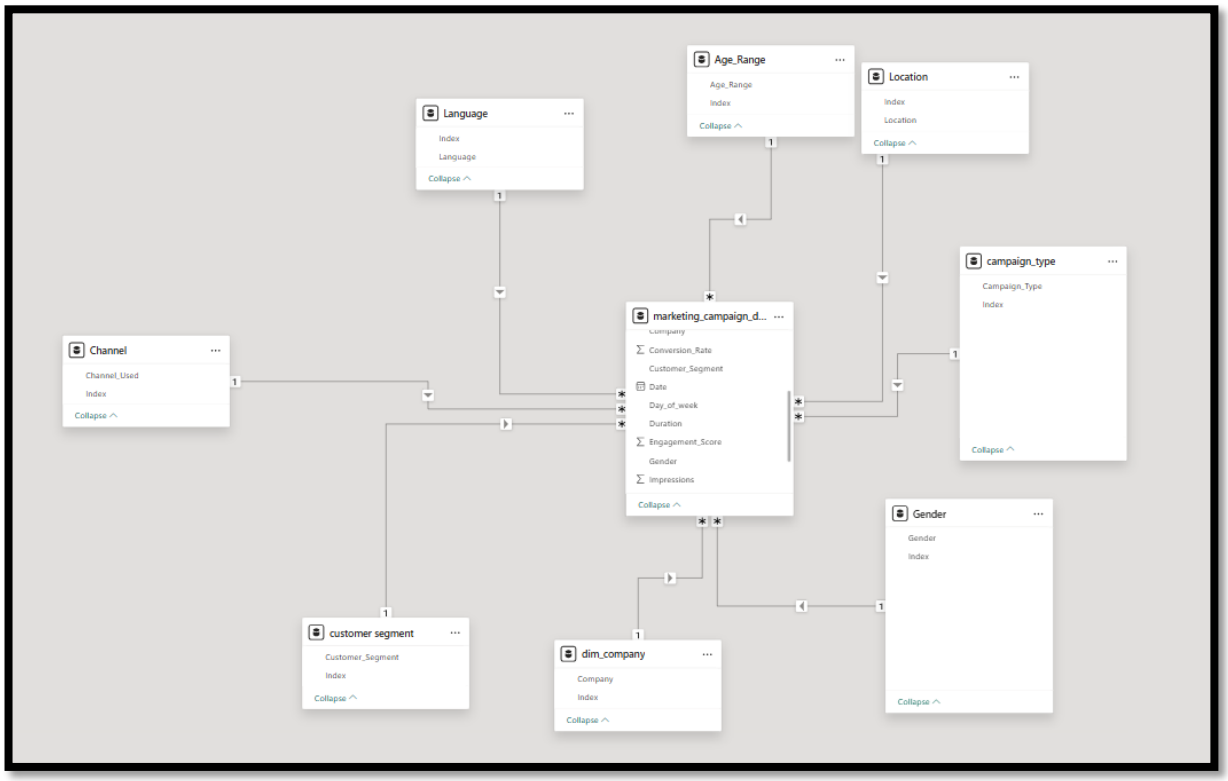
4. System Analysis & Design

Database Design & Data Modeling

For this project, the Entity-Relationship Diagram is implemented as a **Star Schema** in the Power BI Model View. This design places the central marketing_campaign_dataset table (containing all metrics) in the middle, surrounded by the descriptive Dimension tables (used for filtering).

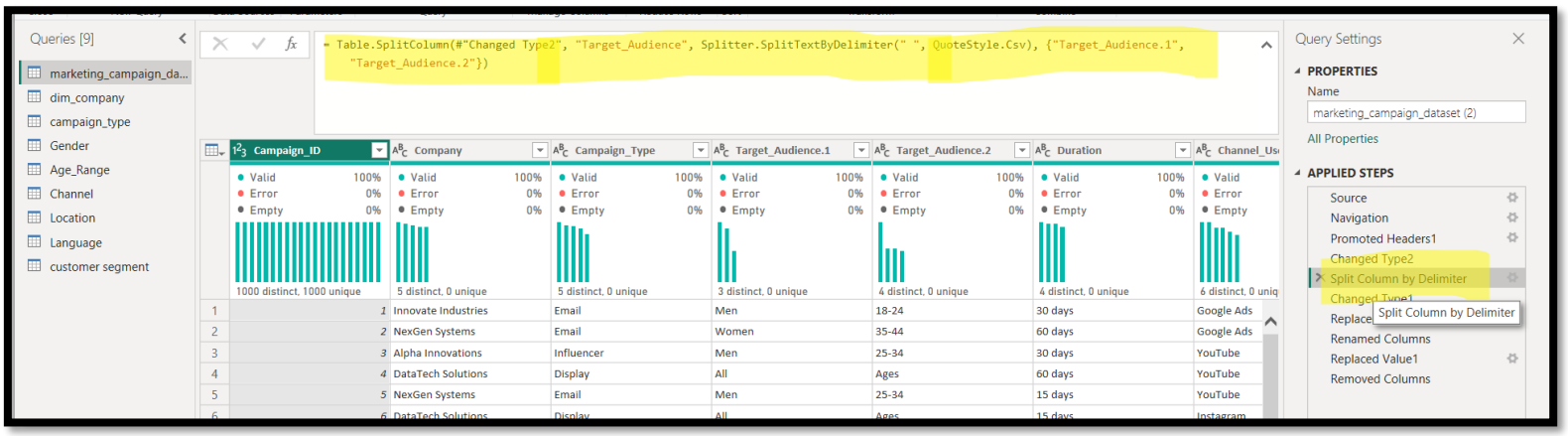
The star schema:

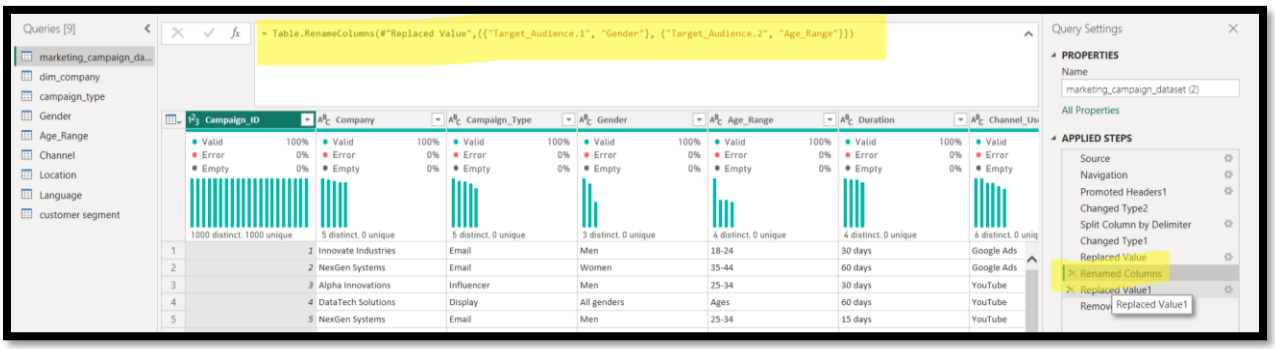


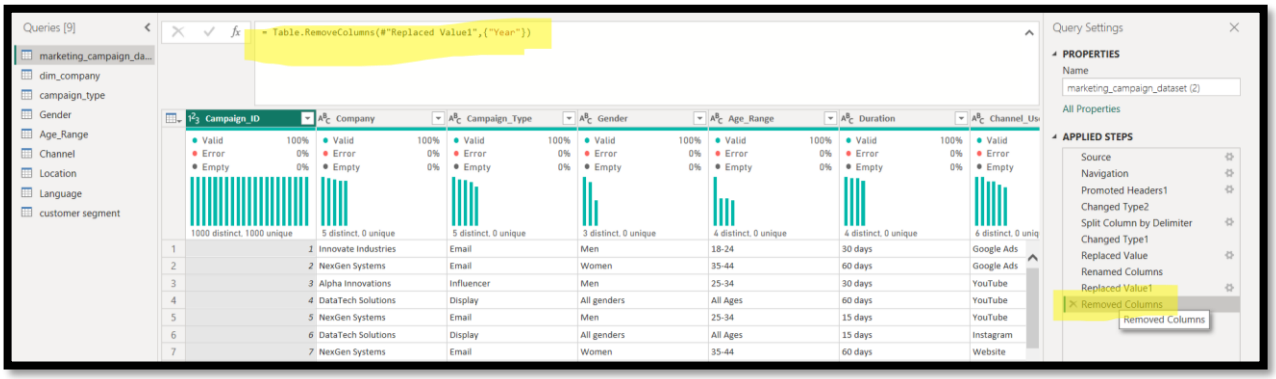
The star schema steps:

1 Load Data: The raw marketing_campaign_dataset Excel file was loaded into the Power Query Editor. This table served as our initial Fact table.

2 Transform Main Tables: The Target_audience column is split into the Gender and Age _range column and the Year column is dropped as all the data fall in the same year.

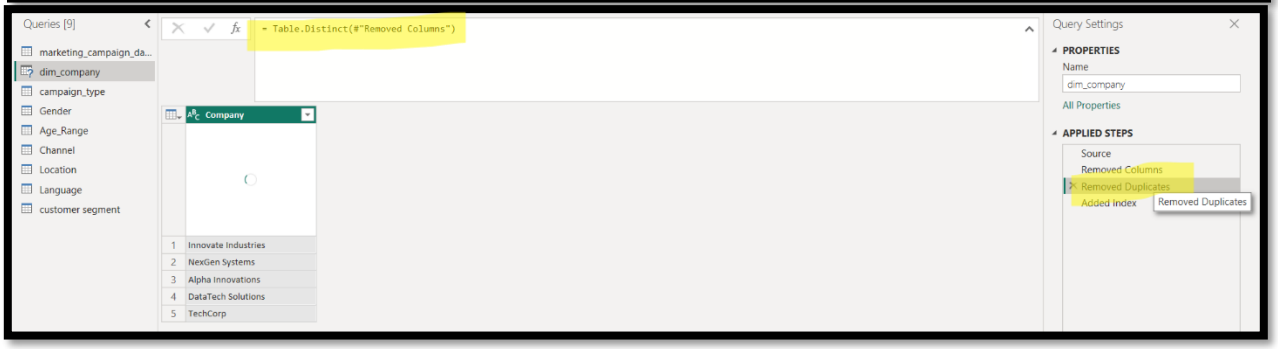






3 Create Dimension Tables:

- To normalize the data, we created separate lookup tables (Dimensions) for each descriptive category.
- For each dimension (e.g., Dim_Company, Dim_Channel, Dim_Location, Dim_Language, Dim_CustomerSegment), we performed the following actions in Power Query:
    1. Referenced the main marketing_campaign_dataset query.
    2. Kept only the single relevant column (e.g., Company).
    3. Used "Remove Duplicates" to create a clean, unique list of values.
    4. Added Index column to the table



4 Handle Complex Dimensions:

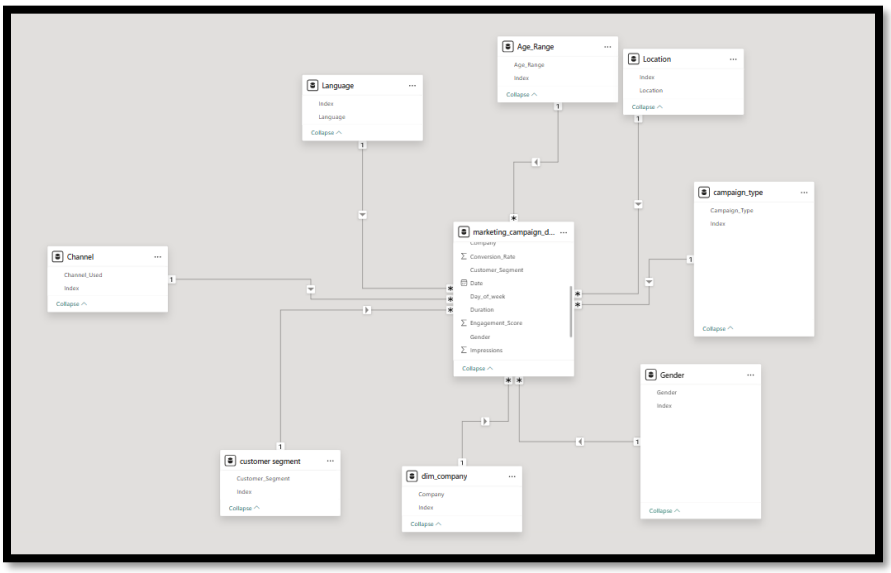
- For the Dim_TargetAudience table, we applied more advanced logic *inside* that specific query to split the Target_Audience column (e.g., "Men 18-24") into two new, clean columns: Gender and Age_Range.

5 Load All Tables to Model:

- Instead of merging tables in Power Query (which was causing "ghost" errors), we adopted the standard DAX/Model-View approach.
- We ensured all tables (the marketing_campaign_dataset table and all new Dim_ tables) had "Enable load" checked.
- We clicked "Close & Apply" to load all tables into the Power BI data model.

6 Create Relationships Manually (Model View):

- We navigated to the "Model" view in Power BI.
- To build the Star Schema, we created relationships by dragging and dropping the matching columns between the tables.
- Example: We dragged the Company column from the marketing_campaign_dataset table and dropped it onto the Company_Name column in the Dim_Company table.
- This process was repeated for all dimensions (Channel, Location, Date, etc.), visually creating the "wires" that define the Star Schema.



Logical & Physical Schema

1. Normalization Considerations

Our design strategy involves transforming the source data from a denormalized state into a normalized **Star Schema**.

- **Initial State (Denormalized):** The source marketing_campaign_dataset is a single, "flat" table. This design is inefficient, causing massive data redundancy (e.g., the text string "Innovate Industries" is repeated thousands of times). This redundancy wastes memory and severely slows down report filtering, as visuals must scan all 200,000 rows to find matching text.
- **End State (Normalized Star Schema):** We normalized the model by:
    1. **Creating Dimension Tables:** We extracted all repeating, descriptive text (like Company, Location, Channel_Used) into separate lookup tables (e.g., Dim_Company).
    2. **Applying 3rd Normal Form (3NF):** Each dimension table is in 3NF—it contains only unique values (the Primary Key) and its related attributes (like Gender and Age_Range in Dim_TargetAudience). This eliminates all data redundancy.
    3. **Defining Relationships:** We connected these small, fast dimension tables to the large marketing_campaign_dataset table.
- **Benefit (Performance):** This schema is the standard for business intelligence. Power BI's engine is highly optimized to filter the small dimension tables (e.g., finding 1 company out of 5) rather than scanning the 200,000-row fact table, resulting in a significantly faster and more responsive report.

2. Physical Schema (Tables, Attributes, & Keys)

The model consists of one Fact Table and eight Dimension Tables.

- **Key Definitions:**
    - **Primary Key (PK):** The unique identifier column in a Dimension table.
    - **Foreign Key (FK):** The matching column in the Fact table that links to a Primary Key.

Dimension Tables (The "Lookups")

Table: Dim_Company

- Company_Name (Text) - **[PK]**

Table: Dim_Channel

- Channel_Name (Text) - **[PK]**

Table: Dim_CampaignType

- Campaign_Type (Text) - **[PK]**

Table: Dim_Location

- Location (Text) - **[PK]**

Table: Dim_Language

- Language (Text) - **[PK]**

Table: Dim_CustomerSegment

- Segment_Name (Text) - **[PK]**

Table: Dim_Gender

- Gender (Text) - **[PK]**

Table: Dim_Gender

- Age_Range (Text) - **[PK]**

Fact Table (The "Data")

Table: Fact_Marketing (Contains 200,000+ rows)

- **Foreign Keys (Links to Dimensions)**
    - Company (Text) - [Links to Dim_Company[Company_Name]]
    - Channel_Used (Text) - [Links to Dim_Channel[Channel_Name]]
    - Campaign_Type (Text) - [Links to Dim_CampaignType[Campaign_Type]]
    - Location (Text) - [Links to Dim_Location[City]]
    - Language (Text) - [Links to Dim_Language[Language]]

    - Customer_Segment (Text) - [Links to Dim_CustomerSegment[Segment_Name]]
    - Target_Audience (Text) - [Links to Dim_TargetAudience[Target_Audience]]
    - Date (Date/Time) - [Links to Dim_Date[Date]]
- **Attributes (Measures)**
    - Campaign_ID (Whole Number)
    - Duration (Text)
    - Conversion_Rate (Decimal Number)
    - Acquisition_Cost (Fixed Decimal / Currency)
    - ROI (Decimal Number)
    - Clicks (Whole Number)
    - Impressions (Whole Number)
    - Engagement_Score (Whole Number)
    - Click_Through_Rate (CTR) (Decimal Number)