Project Documentation: Student Course Data Transformation

Project Overview

This project utilizes Power Query to clean, transform, and analyze student course registration data from an Excel spreadsheet. The goal is to streamline data processing by removing duplicates, handling missing values, and creating new calculated fields such as Student ID and Fee Status. The outcome is a clean dataset loaded into Excel, accompanied by a pivot table for easy analysis of course registrations and fees paid.

Steps

1. Connect to the Data:
   - Load the raw Excel file and open it in Power Query Editor.

2. Data Cleaning Steps:
   a. Remove Unnecessary Columns:
      - Keep only the relevant columns: Name, Email, Course, Registration Date, Fee Paid.
   b. Remove Unnecessary Rows:
      - Delete the first three rows that contain null or irrelevant data.
   c. Promote Headers:
      - Set the first row as headers for all columns.
   d. Change Data Types:
      - Assign suitable data types to each column.
   e. Filter Blanks:
      - Remove records with null values in Registration Date and blanks in Course.
   f. Remove Duplicates:

Project Documentation: Student Course Data Transformation

  - Remove duplicate entries based on the Email column.

3. Data Transformation Steps:

  a. Generate Student ID:

    - Create a new column combining the first name and the first letter of the last name.

  b. Determine Term Start Date:

    - Create a conditional column for Term Start Date based on Registration Date.

  c. Calculate Fee Status:

    - Create a new column categorizing Fee Paid into Not Paid, Fully Paid, and Partly Paid.

4. Load Data into Excel:

  - Rename the query as 'Students Courses' and load the cleaned data into Excel.

5. Post-Load Adjustments:

  - Change the Term Start Date format from General to Date in Excel.

6. Create Pivot Table for Analysis:

  - Insert a pivot table showing Course, total Fee Paid, and student count.