

Project Documentation: Student Course Data Transformation using Power Query

Table of Contents

1. Introduction
 2. Project Overview
 3. Data Preparation
 4. Power Query Steps
 1. 4.1 Connecting to Data
 2. 4.2 Data Cleaning
 3. 4.3 Data Transformation
 5. Loading Data into Excel
 6. Pivot Table Creation
 7. Conclusion
-

1. Introduction

This document outlines the process of using Power Query in Excel to clean, transform, and analyze student course registration data. The goal is to streamline data handling and provide meaningful insights through analysis.

2. Project Overview

The project focuses on an Excel dataset containing student registration details with columns such as Name, Email, Course, Registration Date, and Fee Paid. The steps involve cleaning the data to remove errors and duplicates, transforming it to derive new fields, and preparing it for analysis.

3. Data Preparation

Raw Data Structure

- **Name:** Student's full name
- **Email:** Student's email address
- **Course:** Enrolled course name
- **Registration Date:** Date of registration
- **Fee Paid:** Amount paid by the student

4. Power Query Steps

4.1 Connecting to Data

1. Open Excel and navigate to **Data > Get Data > From File > From Workbook**.
2. Select the raw Excel file and click **Import** to open it in Power Query Editor.

4.2 Data Cleaning

1. **Remove Unnecessary Columns:** Delete columns not needed for analysis, retaining only Name, Email, Course, Registration Date, and Fee Paid.
2. **Remove Unnecessary Rows:** Delete the first three rows that contain null values and irrelevant data.
3. **Promote Headers:** Set the first row as headers for the dataset.
4. **Change Data Types:**
 - Name: Text
 - Email: Text
 - Course: Text
 - Registration Date: Date
 - Fee Paid: Whole Number
5. **Check for Blanks:**
 - Filter out null values in the Registration Date column.
 - Filter out blank entries in the Course column.
6. **Remove Duplicates:** Right-click on the Email column and select **Remove Duplicates** to ensure unique entries.

4.3 Data Transformation

1. Create Student ID:

- Use Add Column > Column from Examples to create a Student ID by combining the first name with the first letter of the last name (e.g., "Hailey Antcliff" becomes "HaileyA").

2. Determine Term Start Date:

- Use Add Column > Conditional Column to create a Term Start Date based on the Registration Date:
 - Before December 15, 2022: February 1, 2023
 - On or after December 15, 2022: April 1, 2023

3. Calculate Fee Status:

- Create a new column using conditional logic:
 - Null: "Not Paid"
 - 500: "Fully Paid"
 - Below 500: "Partly Paid"

5. Loading Data into Excel

- Rename the query to **Students Courses**.
- Close the Power Query Editor and load the cleaned data into Excel.

6. Pivot Table Creation

1. Insert a pivot table in a new worksheet.
2. Set the **Rows** to Course.
3. For **Values**, add:
 - Fee Paid (to sum total fees).
 - Email (to count the number of students per course).

7. Conclusion

The project successfully cleans and transforms student course registration data using Power Query. The final output includes a structured dataset and a pivot table, enabling efficient analysis of course registrations and associated fees.