**HR dataset documentation**

**Project: Employees Data Analysis**

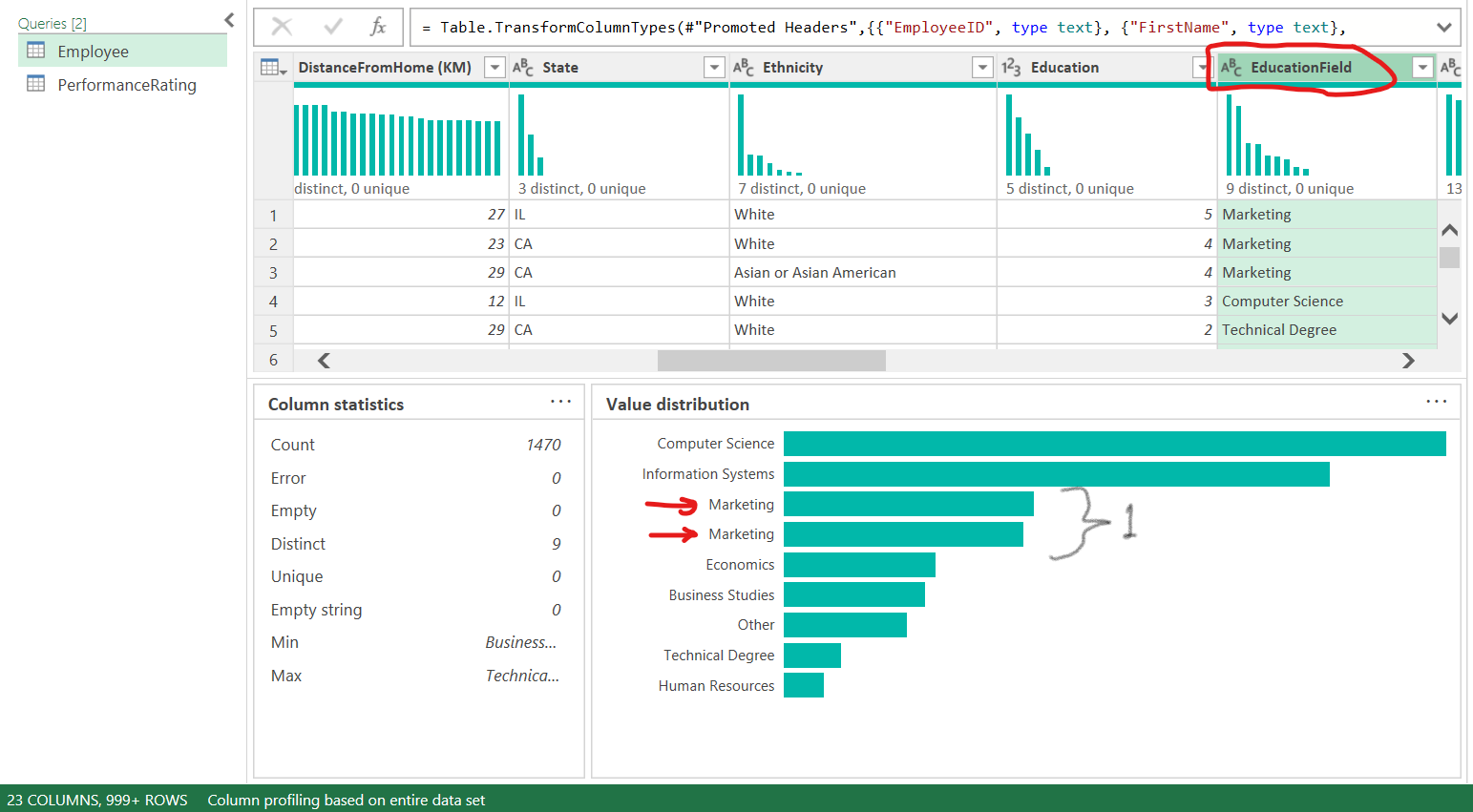
**Date:** 1/10/2024  
**Prepared by:** Team 3 (Youssef, Nadeem, Rana, Islam)  
**Objective:** Clean and transform the dataset to ensure data consistency and accuracy for analysis.

**Overview**

This documentation outlines the steps taken to clean and transform the HR dataset. The focus was on standardizing values in the **EducationField** column, correctly transforming the **HireDate** column to the appropriate date format, and creating a calculated column, **Actual years at company**, with decimal precision. Each transformation step is described in detail below.

**1. Standardizing the "Marketing" Values in the EducationField Column**

* **Column:** EducationField
* **Issue Identified:** The **EducationField** column contained two similar but inconsistent values for "Marketing"—one with an extra trailing space.
  + Example: "Marketing " and "Marketing"



* **Cleaning Action:**
  + The value "Marketing " (with an extra space) was replaced with "Marketing" to ensure consistency across the dataset.
* **Approach:**
  + Applied a transformation to replace "Marketing " with "Marketing".

A screenshot of a computer

Description automatically generated

* **Code**:

= Table.ReplaceValue(#"Previous Step", "Marketing ", "Marketing", Replacer.ReplaceText, {"EducationField"})

**2. Correcting the "HireDate" Format**

* **Column:** HireDate
* **Issue Identified:** The **HireDate** column had dates in the **MM/DD/YYYY** format, which could not be directly converted to the **DD/MM/YYYY** format needed for the analysis.
* **Cleaning Action:**
  + The **HireDate** column was split into **three separate columns** for the month, day, and year using the “/” delimiter.
  + After splitting, the **order** of the columns was rearranged to **DD/MM/YYYY** format.
  + Finally, the three columns were **merged back** together into a single date column in the correct format.

A screenshot of a computer

Description automatically generated

* **Steps Taken:**
  + **Split the HireDate** column by the / delimiter to create three separate columns: Month, Day, and Year.
  + **Rearranged the columns** from Month-Day-Year to Day-Month-Year.
  + **Merged the columns back** into one column in the DD/MM/YYYY format.
  + Transformed the merged column to the **Date** data type.
* **Code**:

// Step 1: Split the column

Table.SplitColumn(#"Previous Step", "HireDate", Splitter.SplitTextByDelimiter("/", QuoteStyle.Csv), {"Month", "Day", "Year"})

// Step 2: Reorder and merge the columns

Table.CombineColumns(#"Previous Step", {"Day", "Month", "Year"}, Combiner.CombineTextByDelimiter("/", QuoteStyle.Csv), "HireDate")

// Step 3: Convert to Date type

Table.TransformColumnTypes(#"Previous Step", {{"HireDate", type date}})

**3. Creating the "Actual years at company" Column**

* **Calculated Column:** Actual years at company
* **Objective:** To calculate the exact tenure (in decimal years) for employees with no attrition, while retaining the value from the **YearsAtCompany** column for employees with attrition.

A screenshot of a computer

Description automatically generated

* **Steps Taken:**
  1. Created a new calculated column, **"Actual years at company"**, based on the following logic:
     + If **Attrition** = "No", calculate the number of years from **HireDate** to the current date (31/12/2022).
     + If **Attrition** = "Yes", use the existing value from the **YearsAtCompany** column.
  2. Formula used for this calculation:

Table.AddColumn(#"Change Type", "Actual years at company", each if [Attrition] = "No" then Duration.Days(#date(2022, 12, 31) - [HireDate]) / 365 else [YearsAtCompany])

**4. Rounding "Actual years at company" and Changing Data Type**

* **Objective:** Round the **Actual years at company** column to **2 decimal places** for consistency and readability, and then change its type to **Decimal Number**.
* **Steps Taken:**
  1. Applied rounding to the **Actual years at company** column to limit the values to 2 decimal places.
  2. Changed the column type to **Decimal Number** for accurate numerical representation and calculations.

**Code:**

Table.TransformColumns(#"Previous Step", {{"Actual years at company", each Number.Round(\_, 2), type number}})

A screenshot of a computer

Description automatically generated

**Summary of Transformations:**

* **Standardized** the "Marketing" values in the **EducationField** column by replacing "Marketing " with "Marketing".
* **Corrected** the **HireDate** format by splitting, rearranging, and merging date components to follow the **DD/MM/YYYY** structure.
* **Created** a new calculated column, **"Actual years at company"**, to represent employees' tenure in decimal years.
* **Rounded** the calculated values to 2 decimal places and **converted** the column type to **Decimal Number** for precision.

**Data modeling:A screenshot of a computer

Description automatically generated**

**Next Steps:**

With the dataset now cleaned and standardized, further analysis and insights can be conducted based on accurate and reliable data.

**Analysis and insights:**

1. **Attrition and Retention**

* **What is the attrition rate by department, job role, and education field?**
  + Understanding which areas have higher turnover can help target retention efforts.
* **What are the key factors (e.g., job satisfaction, distance from home, salary) influencing employee attrition?**
* **Is there a correlation between tenure (Actual years at company) and attrition?**
  + Investigate whether longer tenure is associated with lower attrition rates.
* **Do employees with higher salaries have lower attrition rates?**
  + Examine if competitive compensation reduces turnover.

1. **Performance and productivity vs salary and promotion**

* **Is there a correlation between salary and performance ratings?**
  + Understand if higher salaries lead to higher performance.
* **Are employees with higher performance closer to have promotions?**
  + Search the relation between performance rating and promotions.
* **Do employees promoted within the last year perform better or worse than those who haven’t been promoted?**
  + Examine how promotions affect employee performance.

1. **Departments comparison**

* **What are the differences in job satisfaction across different departments and roles?**
  + Compare satisfaction levels to identify areas needing attention.
* **What are the differences between departments across average salaries, average employees rating and education levels?**