**D599 - Part 1 - Data Profiling  
Andrew Marchese**

Contents

[A. Profile Data 1](#_Toc175400335)

[B. Indicate the data type and data subtype for each variable. 1](#_Toc175400336)

[C. Provide a sample of observable values for each variable. 2](#_Toc175400337)

## A. Profile Data

a. Describe the general characteristics of the initial dataset (e.g., rows, columns).  
The data dictionary for this dataset describes a table containing over 30 different columns, each representing another characteristic about the employee. Each employee is recorded on a different row, capturing many aspects relating to that employee, such as education level, pay rate, department, job satisfaction, and tenure with the company. This dataset contains almost all demographic information, job-related factors, and work-life balance metrics that may influence employee turnover. This structured data is of immense help while analyzing the trends and patterns of retention and turnover of employees in the firm. Informed decision-making with regard to human resource strategies would therefore be possible.

## B. Indicate the data type and data subtype for each variable.

**Age**: Integer (Numeric)

**Turnover**: Categorical (String - Binary)

**BusinessTravel**: Categorical (String - Binary)

**DailyRate**: Integer (Numeric)

**Department**: Categorical (String - Nominal)

**DistanceFromHome**: Integer (Numeric)

**Education**: Ordinal (Numeric - Categorical)

**EducationField**: Categorical (String - Nominal)

**EmployeeCount**: Constant (Integer)

**EmployeeNumber**: Identifier (Integer)

**EnvironmentSatisfaction**: Ordinal (Numeric - Categorical)

**Gender**: Categorical (String - Binary)

**HourlyRate**: Integer (Numeric)

**JobInvolvement**: Ordinal (Numeric - Categorical)

**JobLevel**: Ordinal (Numeric - Categorical)

**JobRole**: Categorical (String - Nominal)

**JobSatisfaction**: Ordinal (Numeric - Categorical)

**MaritalStatus**: Categorical (String - Nominal)

**MonthlyIncome**: Integer (Numeric)

**MonthlyRate**: Integer (Numeric)

**NumCompaniesWorked**: Integer (Numeric)

**Over18**: Categorical (String - Binary)

**OverTime**: Categorical (String - Binary)

**PercentSalaryHike**: Integer (Numeric)

**PerformanceRating**: Ordinal (Numeric - Categorical)

**RelationshipSatisfaction**: Ordinal (Numeric - Categorical)

**StandardHours**: Integer (Numeric - Constant)

**StockOptionalLevel**: Ordinal (Numeric - Categorical)

**TotalWorkingYears**: Integer (Numeric)

**TrainingTimesLastYear**: Integer (Numeric)

**WorkLifeBalance**: Ordinal (Numeric - Categorical)

**YearsAtCompany**: Integer (Numeric)

**YearsInCurrentRole**: Integer (Numeric)

**YearsSinceLastPromotion**: Integer (Numeric)

**YearsWithCurrManager**: Integer (Numeric)

## C. Provide a sample of observable values for each variable.

**Age**: 33, 35, 27, 44, 56

**Turnover**: Yes, Yes, Yes, No, No

**BusinessTravel**: Non-Travel, Non-Travel, Travel\_Frequently, Travel\_Rarely, Travel\_Rarely

**DailyRate**: 241, 679, 359, 1133, 118

**Department**: Hardware, Support, Hardware, Software, Software

**DistanceFromHome**: 16, 7, 50, 12, 43

**Education**: 3, 2, 1, 5, 2

**EducationField**: Technical Degree, Life Sciences, Life Sciences, Life Sciences, Human Resources

**EmployeeCount**: 1 (constant across all samples)

**EmployeeNumber**: 3505, 1129, 6305, 4595, 7203

**EnvironmentSatisfaction**: 1, 3, 4, 2, 2

**Gender**: Female, Male, Female, Female, Female

**HourlyRate**: 67, 122, 199, 150, 115

**JobInvolvement**: 3, 2, 1, 2, 3

**JobLevel**: 3, 5, 1, 4, 1

**JobRole**: Manufacturing Director, Research Director, Sales Representative, Research Director, Developer

**JobSatisfaction**: 1, 2, 3, 3, 2

**MaritalStatus**: Married, Single, Single, Divorced, Married

**MonthlyIncome**: 36809, 1690, 50883, 11166, 42537

**MonthlyRate**: 294472, 32110, 865011, 89328, 1063425

**NumCompaniesWorked**: 1, 6, 8, 5, 6

**Over18**: Y, Y, Y, Y, Y

**OverTime**: Yes, Yes, Yes, No, Yes

**PercentSalaryHike**: 18, 5, 7, 34, 49

**PerformanceRating**: 1, 4, 2, 3, 2

**RelationshipSatisfaction**: 1, 1, 4, 4, 2

**StandardHours**: 80 (constant across all samples)

**StockOptionLevel**: 4, 1, 4, 3, 1

**TotalWorkingYears**: 35, 5, 10, 19, 3

**TrainingTimesLastYear**: 4, 1, 4, 4, 3

**WorkLifeBalance**: 4, 1, 2, 1, 4

**YearsAtCompany**: 13, 4, 4, 14, 1

**YearsInCurrentRole**: 2, 3, 2, 7, 1

**YearsSinceLastPromotion**: 8, 3, 4, 12, 1

**YearsWithCurrManager**: 11, 4, 2, 2, 1

B.