Mid term project

Manikya Deepika Eadara

2023-10-08

employee\_data <- read\_excel("C:\\Users\\manik\\OneDrive\\Employee Data Formatted - GBUS 738 Fall 2023 (1100 Records) f.xlsx")  
head(employee\_data, 10)

## # A tibble: 10 × 13  
## left\_company departm…¹ job\_l…² salary weekl…³ busin…⁴ yrs\_a…⁵ yrs\_s…⁶ previ…⁷  
## <chr> <chr> <chr> <dbl> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Yes Sales Direct… 1.19e5 56 Rarely 6 0 5  
## 2 No Sales Senior… 8.56e4 42 Freque… 10 1 5  
## 3 Yes Product … Associ… 4.62e4 56 Rarely 0 0 6  
## 4 No IT and A… Direct… 1.17e5 50 Freque… 8 3 1  
## 5 No Sales Associ… 3.66e4 46 Rarely 2 2 2  
## 6 No Marketing Senior… 8.35e4 48 Freque… 7 3 1  
## 7 No Marketing Senior… 8.86e4 44 Rarely 1 0 3  
## 8 No Sales Direct… 1.22e5 47 Rarely 1 0 3  
## 9 No Finance … Senior… 9.46e4 50 Freque… 9 1 3  
## 10 No Product … Direct… 1.25e5 51 Rarely 7 7 3  
## # … with 4 more variables: job\_satisfaction <chr>, performance\_rating <chr>,  
## # marital\_status <chr>, miles\_from\_home <dbl>, and abbreviated variable names  
## # ¹​department, ²​job\_level, ³​weekly\_hours, ⁴​business\_travel, ⁵​yrs\_at\_company,  
## # ⁶​yrs\_since\_promotion, ⁷​previous\_companies

## [1] 1100 13

## tibble [1,100 × 13] (S3: tbl\_df/tbl/data.frame)  
## $ left\_company : chr [1:1100] "Yes" "No" "Yes" "No" ...  
## $ department : chr [1:1100] "Sales" "Sales" "Product Development" "IT and Analytics" ...  
## $ job\_level : chr [1:1100] "Director" "Senior Manager" "Associate" "Director" ...  
## $ salary : num [1:1100] 118681 85576 46236 117227 36635 ...  
## $ weekly\_hours : num [1:1100] 56 42 56 50 46 48 44 47 50 51 ...  
## $ business\_travel : chr [1:1100] "Rarely" "Frequently" "Rarely" "Frequently" ...  
## $ yrs\_at\_company : num [1:1100] 6 10 0 8 2 7 1 1 9 7 ...  
## $ yrs\_since\_promotion: num [1:1100] 0 1 0 3 2 3 0 0 1 7 ...  
## $ previous\_companies : num [1:1100] 5 5 6 1 2 1 3 3 3 3 ...  
## $ job\_satisfaction : chr [1:1100] "Very High" "Medium" "High" "High" ...  
## $ performance\_rating : chr [1:1100] "Meets Expectations" "Exceeds Expectations" "Minimally Effective" "Exceptional" ...  
## $ marital\_status : chr [1:1100] "Single" "Married" "Single" "Married" ...  
## $ miles\_from\_home : num [1:1100] 1 8 2 3 2 2 3 24 23 27 ...

## # A tibble: 6 × 13  
## left\_…¹ depar…² job\_l…³ salary weekl…⁴ busin…⁵ yrs\_a…⁶ yrs\_s…⁷ previ…⁸ job\_s…⁹  
## <chr> <chr> <chr> <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <chr>   
## 1 Yes Sales Direct… 1.19e5 56 Rarely 6 0 5 Very H…  
## 2 No Sales Senior… 8.56e4 42 Freque… 10 1 5 Medium   
## 3 Yes Produc… Associ… 4.62e4 56 Rarely 0 0 6 High   
## 4 No IT and… Direct… 1.17e5 50 Freque… 8 3 1 High   
## 5 No Sales Associ… 3.66e4 46 Rarely 2 2 2 Medium   
## 6 No Market… Senior… 8.35e4 48 Freque… 7 3 1 Very H…  
## # … with 3 more variables: performance\_rating <chr>, marital\_status <chr>,  
## # miles\_from\_home <dbl>, and abbreviated variable names ¹​left\_company,  
## # ²​department, ³​job\_level, ⁴​weekly\_hours, ⁵​business\_travel, ⁶​yrs\_at\_company,  
## # ⁷​yrs\_since\_promotion, ⁸​previous\_companies, ⁹​job\_satisfaction

## Rows: 1,100  
## Columns: 13  
## $ left\_company <chr> "Yes", "No", "Yes", "No", "No", "No", "No", "No", …  
## $ department <chr> "Sales", "Sales", "Product Development", "IT and A…  
## $ job\_level <chr> "Director", "Senior Manager", "Associate", "Direct…  
## $ salary <dbl> 118680.74, 85576.44, 46235.79, 117226.84, 36634.73…  
## $ weekly\_hours <dbl> 56, 42, 56, 50, 46, 48, 44, 47, 50, 51, 46, 48, 48…  
## $ business\_travel <chr> "Rarely", "Frequently", "Rarely", "Frequently", "R…  
## $ yrs\_at\_company <dbl> 6, 10, 0, 8, 2, 7, 1, 1, 9, 7, 5, 9, 5, 2, 4, 10, …  
## $ yrs\_since\_promotion <dbl> 0, 1, 0, 3, 2, 3, 0, 0, 1, 7, 0, 0, 4, 1, 0, 8, 0,…  
## $ previous\_companies <dbl> 5, 5, 6, 1, 2, 1, 3, 3, 3, 3, 1, 3, 2, 2, 2, 2, 3,…  
## $ job\_satisfaction <chr> "Very High", "Medium", "High", "High", "Medium", "…  
## $ performance\_rating <chr> "Meets Expectations", "Exceeds Expectations", "Min…  
## $ marital\_status <chr> "Single", "Married", "Single", "Married", "Married…  
## $ miles\_from\_home <dbl> 1, 8, 2, 3, 2, 2, 3, 24, 23, 27, 16, 15, 26, 19, 2…

## left\_company department job\_level salary   
## Length:1100 Length:1100 Length:1100 Min. : 30488   
## Class :character Class :character Class :character 1st Qu.: 69896   
## Mode :character Mode :character Mode :character Median : 88378   
## Mean : 93727   
## 3rd Qu.:116290   
## Max. :212135   
## weekly\_hours business\_travel yrs\_at\_company yrs\_since\_promotion  
## Min. :40.00 Length:1100 Min. : 0.000 Min. : 0.00   
## 1st Qu.:47.00 Class :character 1st Qu.: 3.000 1st Qu.: 0.00   
## Median :49.00 Mode :character Median : 5.000 Median : 1.00   
## Mean :50.09 Mean : 7.081 Mean : 2.23   
## 3rd Qu.:52.00 3rd Qu.: 9.000 3rd Qu.: 3.00   
## Max. :66.00 Max. :40.000 Max. :15.00   
## previous\_companies job\_satisfaction performance\_rating marital\_status   
## Min. :1.00 Length:1100 Length:1100 Length:1100   
## 1st Qu.:2.00 Class :character Class :character Class :character   
## Median :3.00 Mode :character Mode :character Mode :character   
## Mean :3.26   
## 3rd Qu.:4.00   
## Max. :7.00   
## miles\_from\_home   
## Min. : 1.000   
## 1st Qu.: 2.000   
## Median : 7.000   
## Mean : 9.033   
## 3rd Qu.:13.000   
## Max. :29.000

Data summary

|  |  |
| --- | --- |
| Name | employee\_data |
| Number of rows | 1100 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| numeric | 6 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| left\_company | 0 | 1 | 2 | 3 | 0 | 2 | 0 |
| department | 0 | 1 | 5 | 22 | 0 | 6 | 0 |
| job\_level | 0 | 1 | 7 | 14 | 0 | 5 | 0 |
| business\_travel | 0 | 1 | 4 | 10 | 0 | 3 | 0 |
| job\_satisfaction | 0 | 1 | 3 | 9 | 0 | 4 | 0 |
| performance\_rating | 0 | 1 | 11 | 20 | 0 | 5 | 0 |
| marital\_status | 0 | 1 | 6 | 8 | 0 | 3 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| salary | 0 | 1 | 93727.33 | 37761.80 | 30488.15 | 69896.09 | 88377.68 | 116290.4 | 212134.7 | ▃▇▃▁▁ |
| weekly\_hours | 0 | 1 | 50.09 | 4.90 | 40.00 | 47.00 | 49.00 | 52.0 | 66.0 | ▂▇▃▂▁ |
| yrs\_at\_company | 0 | 1 | 7.08 | 6.29 | 0.00 | 3.00 | 5.00 | 9.0 | 40.0 | ▇▂▁▁▁ |
| yrs\_since\_promotion | 0 | 1 | 2.23 | 3.30 | 0.00 | 0.00 | 1.00 | 3.0 | 15.0 | ▇▁▁▁▁ |
| previous\_companies | 0 | 1 | 3.26 | 1.59 | 1.00 | 2.00 | 3.00 | 4.0 | 7.0 | ▇▇▁▂▃ |
| miles\_from\_home | 0 | 1 | 9.03 | 8.08 | 1.00 | 2.00 | 7.00 | 13.0 | 29.0 | ▇▅▁▂▂ |



     