

Final Project

Ettore Falde, Samoussa Fofana, Federico Basaglia

11/29/2021

Introduction

In this report we consider that the new CEO of a specific IT company has contacted us because she wants us to **analyze the current Human Resources status** of the company. She has just sent a data set with all available employee information. As we can see, the **company has two locations**: the first one in **London**, and the second one in **Barcelona**.

The new CEO is concerned about several issues. She truly believes in gender equality in organizations as it implies a signal to society. On the other hand, she is concerned that the offices in Barcelona do not follow a similar structure to the one in London. **In her opinion, the structure of the Barcelona offices should tend towards the London structure.** In her meeting with us, she also told us that she would like to know the attitudes (e.g., satisfaction) of the employees across the different departments and if anything could be done to improve them. Finally, she commented that she is very concerned about the company's succession strategy and in particular some positions in certain departments.

Let's consider that the new CEO of a specific IT company has contacted us because she wants us to analyze the current Human Resources status of the company. She has just sent a data set with all available employee information. This information is in the attached data set. As we can see, the company has two locations: the first one in London, and the second one in Barcelona.

The new CEO is concerned about several issues. She truly believes in gender equality in organizations as it implies a signal to society. On the other hand, she is concerned that the offices in Barcelona do not follow a similar structure to the one in London. In her opinion, the structure of the Barcelona offices should tend towards the London structure. In her meeting with us, she also told us that she would like to know the attitudes (e.g., satisfaction) of the employees across the different departments and if anything could be done to improve them. Finally, she commented that she is very concerned about the company's succession strategy and in particular some positions in certain departments.

Based on this information, we need to carry out an exploratory data analysis and prepare a technical report (with Rmarkdown) and a technical presentation (5-10 minutes).

Note: It is highly recommended to seek external sources of information (either in dataset or report formats) for the analysis and the reporting.

Based on this information, we will to carry out an exploratory data analysis and prepare a technical report (with Rmarkdown) and a technical presentation (5-10 minutes).

Setup the software

The software used for the development of the study and the writing of the report is R[1]. The first step is to define the work directory and to load the libraries needed:

```
library(tidyverse)
library(ggplot2)
library(broom)
library(janitor)
library(caTools)
library(ROCR)
library(corrplot)
```

Importing Data

The first step is to load the dataset in the system, and check the names of the variables.

```
mydb <- read.csv2("dataset.csv")
names(mydb)
```

```
## [1] "Age" "Attrition"
## [3] "BusinessTravel" "DailyRate"
## [5] "Department" "DistanceFromHome"
## [7] "Education" "EducationField"
## [9] "EmployeeCount" "EmployeeNumber"
## [11] "EnvironmentSatisfaction" "Gender"
## [13] "HourlyRate" "JobInvolvement"
## [15] "JobLevel" "JobRole"
## [17] "JobSatisfaction" "MaritalStatus"
## [19] "MonthlyIncome" "MonthlyRate"
## [21] "NumCompaniesWorked" "Over18"
## [23] "OverTime" "PercentSalaryHike"
## [25] "PerformanceRating" "RelationshipSatisfaction"
## [27] "StandardHours" "StockOptionLevel"
## [29] "TotalWorkingYears" "TrainingTimesLastYear"
## [31] "WorkLifeBalance" "YearsAtCompany"
## [33] "YearsInCurrentRole" "YearsSinceLastPromotion"
## [35] "YearsWithCurrManager" "City"
```

```
# web_db <- read.csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
# names(web_db)
```

We got the dataset from the website of Atenea, it is composed by 1506 observations of 36 variables. The variables selected for this dataset are:

1. **Age:** Variable that represent the age of the employee
2. **Attrition:** variable that represent the departure of employees from the organization for any reason
3. **BusinessTravel:** Represent how often an employee travel for work purpose
4. **DailyRate:** The amount of money the employees are paid per day
5. **Department:** Department of the company at which the employee belong
6. **DistanceFromHome:** Employee home distance from the workplace
7. **Education:** Educational level of the employee (1=Below College, 2=College, 3=Bachelor, 4=Master, 5= Doctor)
8. **EducationField:** Education field of employee (Human Resources, Life Sciencies, Marketing, Medical, Technical Degree, Other)
9. **EmployeeCount:** Coolumn all equal to 1 to count the total number of employee in the data set

10. **EmployeeNumber**: unique number to identify the employee
11. **EnvironmentSatisfaction**: level of environment satisfaction (1=Low, 2=Medium, 3=High, 4=Very High)
12. **Gender**: Gender of the employee (Male, Female)
13. **HourlyRate**: The amount of money the employees are paid per hour
14. **JobInvolvement**: Level of involvement of the employee (1=Low, 2=Medium, 3=High, 4=Very High)
15. **JobLevel**: Is a category of authority in the company (1=low, 5=High)
16. **JobRole**: Represent the role cover by the employee (Sales Executive, Research Scientist, Laboratory Technician, Manufacturing Director, Healthcare Representative, Manager, Sales Representative, Research Director, Human Resources)
17. **JobSatisfaction**: Level of satisfaction of the employee (1=Low, 2=Medium, 3=High, 4=Very High)
18. **MaritalStatus**: Marital status of the employee (Divorced, Married, Single)
19. **MonthlyIncome**: Monthly income of the employee
20. **MonthlyRate**: Monthly rate of employee
21. **NumCompaniesWorked**: Number of companies for which the employee worked
22. **Over18**: If the age of the employee is higher than 18 (Y = yes, N = no)
23. **OverTime**: If the employee perform over time (Yes, No)
24. **PercentSalaryHike**: Represent the percentage increase of a salary
25. **PerformanceRating**: Performance rating of the employee (1=Low, 2=Good, 3=Excellent, 4=Outstanding)
26. **RelationshipSatisfaction**: Relationship satisfaction of the employee (1=Low, 2=Medium, 3=High, 4=Very High)
27. **StandardHours**: Standard working hour per **week?** (80 for everyone)
28. **StockOptionLevel**: Stock option level
29. **TotalWorkingYears**: Total years of working
30. **TrainingTimesLastYear**: Training hours of the last year
31. **WorkLifeBalance**: the amount of time you spend doing your job compared with the amount of time you spend with your family and doing things you enjoy (1=Bad, 2=Good, 3=Better, 4=Best)
32. **YearsAtCompany**: Total years of working at the company
33. **YearsInCurrentRole**: Total years spent in the current position
34. **YearsSinceLastPromotion**: How many year ago the employee had the last promotion
35. **YearsWithCurrManager**: How many years the employee is with the actual manager
36. **City**: where the employee works (London, Barcelona)

Cleaning Data

Names In this sub-point we are going to change the names of the variables in order to have all the names of the variables with the same layout.

```
mydb <- mydb %>% clean_names(., "snake")
```

Dimensions First of all, we are going to check the actual dimension of our dataset. Hence, from the following code we can understand that there are 36 variables in total and

```
mydb %>% dim()
```

```
## [1] 1506 36
```

```
mydb %>% nrow()
```

```
## [1] 1506
```

```
mydb %>% ncol()
```

```
## [1] 36
```

Head and Tail Here, we are going to check the first 10 elements at the beginning and at the end of the dataset. Consecutively, we are going to check the top and the bottom values of the main relevant variables to catch some errors.

```
mydb %>% head(10)
```

```
##      age attrition  business_travel daily_rate      department
## 1    41      Yes    Travel_Rarely    1102      Sales
## 2    49     No Travel_Frequently    279 Research & Development
## 3    37      Yes    Travel_Rarely    1373 Research & Development
## 4    33     No Travel_Frequently    1392 Research & Development
## 5    27     No    Travel_Rarely    591 Research & Development
## 6    32     No Travel_Frequently    1005 Research & Development
## 7    59     No    Travel_Rarely    1324 Research & Development
## 8    30     No    Travel_Rarely    1358 Research & Development
## 9    38     No Travel_Frequently    216 Research & Development
## 10   36     No    Travel_Rarely    1299 Research & Development
## distance_from_home education education_field employee_count employee_number
## 1             1             2 Life Sciences             1             1
## 2             8             1 Life Sciences             1             2
## 3             2             2 Other                  1             4
## 4             3             4 Life Sciences             1             5
## 5             2             1 Medical                  1             7
## 6             2             2 Life Sciences             1             8
## 7             3             3 Medical                  1            10
## 8            24             1 Life Sciences             1            11
## 9            23             3 Life Sciences             1            12
## 10           27             3 Medical                  1            13
## environment_satisfaction gender hourly_rate job_involvement job_level
## 1             2 Female          94             3             2
## 2             3 Male           61             2             2
## 3             4 Male           92             2             1
## 4             4 Female          56             3             1
## 5             1 Male           40             3             1
## 6             4 Male           79             3             1
## 7             3 Female          81             4             1
## 8             4 Male           67             3             1
## 9             4 Male           44             2             3
## 10            3 Male           94             3             2
##      job_role job_satisfaction marital_status monthly_income
## 1 Sales Executive          4      Single      5993
## 2 Research Scientist        2      Married     5130
## 3 Laboratory Technician      3      Single     2090
## 4 Research Scientist        3      Married     2909
## 5 Laboratory Technician      2      Married     3468
## 6 Laboratory Technician      4      Single     3068
## 7 Laboratory Technician      1      Married     2670
## 8 Laboratory Technician      3      Divorced     2693
```

## 9	Manufacturing Director	3	Single	9526
## 10	Healthcare Representative	3	Married	5237
##	monthly_rate num_companies_worked over18 over_time percent_salary_hike			
## 1	19479 8 Y Yes	11		
## 2	24907 1 Y No	23		
## 3	2396 6 Y Yes	15		
## 4	23159 1 Y Yes	11		
## 5	16632 9 Y No	12		
## 6	11864 0 Y No	13		
## 7	9964 4 Y Yes	20		
## 8	13335 1 Y No	22		
## 9	8787 0 Y No	21		
## 10	16577 6 Y No	13		
##	performance_rating relationship_satisfaction standard_hours			
## 1	3 1 80			
## 2	4 4 80			
## 3	3 2 80			
## 4	3 3 80			
## 5	3 4 80			
## 6	3 3 80			
## 7	4 1 80			
## 8	4 2 80			
## 9	4 2 80			
## 10	3 2 80			
##	stock_option_level total_working_years training_times_last_year			
## 1	0 8 0			
## 2	1 10 3			
## 3	0 7 3			
## 4	0 8 3			
## 5	1 6 3			
## 6	0 8 2			
## 7	3 12 3			
## 8	1 1 2			
## 9	0 10 2			
## 10	2 17 3			
##	work_life_balance years_at_company years_in_current_role			
## 1	1 6 4			
## 2	3 10 7			
## 3	3 0 0			
## 4	3 8 7			
## 5	3 2 2			
## 6	2 7 7			
## 7	2 1 0			
## 8	3 1 0			
## 9	3 9 7			
## 10	2 7 7			
##	years_since_last_promotion years_with_curr_manager city			
## 1	0 5 London			
## 2	1 7 London			
## 3	0 0 London			
## 4	3 0 London			
## 5	2 2 London			
## 6	3 6 Barcelona			
## 7	0 0 London			

```
## 8          0          0    London
## 9          1          8    London
## 10         7          7    London
```

```
mydb %>% tail(10)
```

```
##      age attrition  business_travel daily_rate      department
## 1497  40         No   Travel_Rarely      898   Human Resources
## 1498  65         No   Non-Travel      530         Sales
## 1499  58         No Travel_Frequently  1216 Research & Development
## 1500  39         No   Travel_Rarely      835 Research & Development
## 1501  29         No   Travel_Rarely      592 Research & Development
## 1502  26         No Travel_Frequently  1096 Research & Development
## 1503  34         No Travel_Frequently   735 Research & Development
## 1504  54         No   Travel_Rarely      543 Research & Development
## 1505  13         No Travel_Frequently   688 Research & Development
## 1506   6         No   Travel_Rarely      628 Research & Development
##      distance_from_home education  education_field employee_count
## 1497          6          2      Medical          1
## 1498         16          3    Life Sciences          1
## 1499         15          4 Technical Degree          1
## 1500         19          4      Other          1
## 1501          7          3    Life Sciences          1
## 1502          6          3      Other          1
## 1503         22          4 Technical Degree          1
## 1504          1          4    Life Sciences          1
## 1505          4          2    Life Sciences          1
## 1506          8          3      Medical          1
##      employee_number environment_satisfaction gender hourly_rate
## 1497          1550          3 Male          38
## 1498          1681          3 Female          36
## 1499          1837          1 Male          87
## 1500          1871          4 Male          41
## 1501          1883          4 Male          59
## 1502          1918          3 Female          61
## 1503          1932          3 Male          86
## 1504          2012          1 Male          83
## 1505          2025          4 Female          97
## 1506          2068          2 Male          82
##      job_involvement job_level      job_role job_satisfaction
## 1497          3          4      Manager          4
## 1498          3          2    Sales Executive          4
## 1499          3          4    Research Director          3
## 1500          3          2    Research Scientist          4
## 1501          3          1 Laboratory Technician          1
## 1502          4          1 Laboratory Technician          4
## 1503          2          2    Research Scientist          4
## 1504          3          1 Laboratory Technician          4
## 1505          3          2 Manufacturing Director          2
## 1506          4          2 Laboratory Technician          3
##      marital_status monthly_income monthly_rate num_companies_worked over18
## 1497      Single      16437      17381          1      Y
## 1498    Divorced      5368      16130          1      Y
## 1499    Married      15787      21624          2      Y
```

##	1500	Divorced	3902	5141	8	Y
##	1501	Single	2062	19384	3	Y
##	1502	Married	2544	7102	0	Y
##	1503	Married	5747	26496	1	Y
##	1504	Married	2406	4060	8	Y
##	1505	Divorced	5131	9192	7	Y
##	1506	Married	4404	10228	2	Y
##		over_time	percent_salary_hike	performance_rating	relationship_satisfaction	
##	1497	Yes	21	4		4
##	1498	Yes	25	4		3
##	1499	Yes	14	3		2
##	1500	No	14	3		2
##	1501	No	14	3		2
##	1502	No	18	3		1
##	1503	Yes	15	3		2
##	1504	No	19	3		3
##	1505	No	13	3		2
##	1506	No	12	3		1
##		standard_hours	stock_option_level	total_working_years		
##	1497	80	0	21		
##	1498	80	1	7		
##	1499	80	0	23		
##	1500	80	3	7		
##	1501	80	0	11		
##	1502	80	1	8		
##	1503	80	0	16		
##	1504	80	2	8		
##	1505	80	3	18		
##	1506	80	0	6		
##		training_times_last_year	work_life_balance	years_at_company		
##	1497	2	3	21		
##	1498	2	3	6		
##	1499	3	3	2		
##	1500	2	3	2		
##	1501	2	3	2		
##	1502	3	3	7		
##	1503	3	3	15		
##	1504	3	2	1		
##	1505	3	3	4		
##	1506	3	4	4		
##		years_in_current_role	years_since_last_promotion	years_with_curr_manager		
##	1497	7	7	7		
##	1498	5	1	2		
##	1499	2	2	2		
##	1500	2	2	2		
##	1501	3	1	2		
##	1502	7	7	7		
##	1503	10	6	11		
##	1504	0	0	0		
##	1505	2	0	2		
##	1506	3	1	2		
##		city				
##	1497	London				
##	1498	Barcelona				

```
## 1499 London
## 1500 London
## 1501 London
## 1502 London
## 1503 Barcelona
## 1504 London
## 1505 Barcelona
## 1506 London
```

```
mydb %>% arrange(desc(age)) %>% top_n(10, age)
```

```
##      age attrition  business_travel daily_rate      department
## 1  210      No Travel_Frequently    1297 Research and Development
## 2  148      No   Travel_Rarely        446 Research & Development
## 3  141      No   Travel_Rarely    1283 Research & Development
## 4  136      No   Travel_Rarely    429 Research & Development
## 5  125      No   Travel_Rarely    855 Research & Development
## 6   98      No   Travel_Rarely    121 Research & Development
## 7   65      No   Non-Travel      530 Sales
## 8   60      No   Travel_Rarely    422 Research & Development
## 9   60      No Travel_Frequently    1499 Sales
## 10  60      No   Travel_Rarely    1179 Sales
## 11  60      No   Travel_Rarely    696 Sales
## 12  60      No   Travel_Rarely    370 Research & Development
##      distance_from_home education education_field employee_count employee_number
## 1           1           4      Medical           1           1922
## 2           1           4 Life Sciences           1           635
## 3           5           5      Medical           1          1448
## 4           2           4 Life Sciences           1          1294
## 5           4           3 Life Sciences           1          1363
## 6           2           4      Medical           1           804
## 7          16           3 Life Sciences           1          1681
## 8           7           3 Life Sciences           1           549
## 9          28           3 Marketing            1           573
## 10          16           4 Marketing            1           732
## 11           7           4 Marketing            1          1233
## 12           1           4      Medical           1          1697
##      environment_satisfaction gender hourly_rate job_involvement job_level
## 1           2 Male           44           3           2
## 2           2 Female          65           3           2
## 3           2 Male           90           4           1
## 4           3 Female          53           3           2
## 5           4 Male           54           3           3
## 6           3 Female          86           2           1
## 7           3 Female          36           3           2
## 8           1 Female          41           3           5
## 9           3 Female          80           2           3
## 10          1 Male           84           3           2
## 11          2 Male           52           4           2
## 12          3 Male           92           1           3
##      job_role job_satisfaction marital_status monthly_income
## 1 Healthcare Representative           3      Single          5399
## 2 Manufacturing Director             2      Married          6447
## 3 Research Scientist                 3      Married          2127
```


## 4	Manufacturing Director	2	Single	5410	
## 5	Manufacturing Director	4	Single	7898	
## 6	Research Scientist	1	Single	4381	
## 7	Sales Executive	4	Divorced	5368	
## 8	Manager	1	Married	19566	
## 9	Sales Executive	1	Married	10266	
## 10	Sales Executive	1	Single	5405	
## 11	Sales Executive	4	Divorced	5220	
## 12	Healthcare Representative	4	Divorced	10883	
##	monthly_rate	num_companies_worked	over18	over_time	percent_salary_hike
## 1	14511	4	Y	No	12
## 2	15701	6	Y	No	12
## 3	5561	2	Y	Yes	12
## 4	2323	9	Y	Yes	11
## 5	18706	1	Y	No	11
## 6	7530	1	Y	No	11
## 7	16130	1	Y	Yes	25
## 8	3854	5	Y	No	11
## 9	2845	4	Y	No	19
## 10	11924	8	Y	No	14
## 11	10893	0	Y	Yes	18
## 12	20467	3	Y	No	20
##	performance_rating	relationship_satisfaction	standard_hours		
## 1	3		3	80	
## 2	3		2	80	
## 3	3		1	80	
## 4	3		4	80	
## 5	3		3	80	
## 6	3		3	80	
## 7	4		3	80	
## 8	3		4	80	
## 9	3		4	80	
## 10	3		4	80	
## 11	3		2	80	
## 12	4		3	80	
##	stock_option_level	total_working_years	training_times_last_year		
## 1	0	12	3		
## 2	1	8	2		
## 3	0	7	5		
## 4	0	18	2		
## 5	0	11	2		
## 6	0	6	3		
## 7	1	7	2		
## 8	0	33	5		
## 9	0	22	5		
## 10	0	10	1		
## 11	1	12	3		
## 12	1	19	2		
##	work_life_balance	years_at_company	years_in_current_role		
## 1	3	4	2		
## 2	2	6	5		
## 3	2	4	2		
## 4	3	16	14		
## 5	3	10	9		

```

## 6          3          6          5
## 7          3          6          5
## 8          1         29          8
## 9          4         18         13
## 10         3          2          2
## 11         3         11          7
## 12         4          1          0
##   years_since_last_promotion years_with_curr_manager    city
## 1              0              3 Barcelona
## 2              4              3   London
## 3              0              3   London
## 4              5             12   London
## 5              0              8   London
## 6              1              3 Barcelona
## 7              1              2 Barcelona
## 8             11             10   London
## 9             13             11 Barcelona
## 10             2              2   London
## 11             1              9   London
## 12             0              0   London

```

```
mydb %>% arrange(age) %>% top_n(-10, age)
```

```

##   age attrition  business_travel daily_rate    department
## 1    6        No    Travel_Rarely    916 Research & Development
## 2    6        No    Travel_Rarely    628 Research & Development
## 3    7        No    Travel_Rarely    232 Research & Development
## 4   13        No  Travel_Frequently    688 Research & Development
## 5   18        Yes    Travel_Rarely    230 Research & Development
## 6   18        No    Travel_Rarely    812           Sales
## 7   18        Yes  Travel_Frequently   1306           Sales
## 8   18        No      Non-Travel    287 Research & Development
## 9   18        Yes      Non-Travel    247 Research & Development
## 10  18        No      Non-Travel   1124 Research & Development
## 11  18        Yes  Travel_Frequently    544           Sales
## 12  18        No      Non-Travel   1431 Research & Development
##   distance_from_home education  education_field employee_count employee_number
## 1              17         2    Life Sciences            1            347
## 2              8         3      Medical            1           2068
## 3             19         3  Technical Degree            1            611
## 4              4         2    Life Sciences            1           2025
## 5              3         3    Life Sciences            1            405
## 6             10         3      Medical            1            411
## 7              5         3    Marketing            1            614
## 8              5         2    Life Sciences            1           1012
## 9              8         1      Medical            1           1156
## 10             1         3    Life Sciences            1           1368
## 11             3         2      Medical            1           1624
## 12             14         3      Medical            1           1839
##   environment_satisfaction gender hourly_rate job_involvement job_level
## 1              4 Female          82          4          2
## 2              2  Male          82          4          2
## 3              4  Male          34          3          2
## 4              4 Female          97          3          2

```

## 5	3	Male	54	3	1
## 6	4	Female	69	2	1
## 7	2	Male	69	3	1
## 8	2	Male	73	3	1
## 9	3	Male	80	3	1
## 10	4	Female	97	3	1
## 11	2	Female	70	3	1
## 12	2	Female	33	3	1
##	job_role job_satisfaction marital_status monthly_income				
## 1	Research Scientist	1	Single	6545	
## 2	Laboratory Technician	3	Married	4404	
## 3	Manufacturing Director	4	Divorced	4262	
## 4	Manufacturing Director	2	Divorced	5131	
## 5	Laboratory Technician	3	Single	1420	
## 6	Sales Representative	3	Single	1200	
## 7	Sales Representative	2	Single	1878	
## 8	Research Scientist	4	Single	1051	
## 9	Laboratory Technician	3	Single	1904	
## 10	Laboratory Technician	4	Single	1611	
## 11	Sales Representative	4	Single	1569	
## 12	Research Scientist	3	Single	1514	
##	monthly_rate num_companies_worked over18 over_time percent_salary_hike				
## 1	23016	3	Y	Yes	13
## 2	10228	2	Y	No	12
## 3	22645	4	Y	No	12
## 4	9192	7	Y	No	13
## 5	25233	1	Y	No	13
## 6	9724	1	Y	No	12
## 7	8059	1	Y	Yes	14
## 8	13493	1	Y	No	15
## 9	13556	1	Y	No	12
## 10	19305	1	Y	No	15
## 11	18420	1	Y	Yes	12
## 12	8018	1	Y	No	16
##	performance_rating relationship_satisfaction standard_hours				
## 1	3		3	80	
## 2	3		1	80	
## 3	3		2	80	
## 4	3		2	80	
## 5	3		3	80	
## 6	3		1	80	
## 7	3		4	80	
## 8	3		4	80	
## 9	3		4	80	
## 10	3		3	80	
## 11	3		3	80	
## 12	3		3	80	
##	stock_option_level total_working_years training_times_last_year				
## 1	0	10		1	
## 2	0	6		3	
## 3	2	8		2	
## 4	3	18		3	
## 5	0	0		2	
## 6	0	0		2	

```

## 7          0          0          3
## 8          0          0          2
## 9          0          0          0
## 10         0          0          5
## 11         0          0          2
## 12         0          0          4
##   work_life_balance years_at_company years_in_current_role
## 1          3          3          2
## 2          4          4          3
## 3          4          3          2
## 4          3          4          2
## 5          3          0          0
## 6          3          0          0
## 7          3          0          0
## 8          3          0          0
## 9          3          0          0
## 10         4          0          0
## 11         4          0          0
## 12         1          0          0
##   years_since_last_promotion years_with_curr_manager    city
## 1              0              2 Barcelona
## 2              1              2   London
## 3              1              2   London
## 4              0              2 Barcelona
## 5              0              0   London
## 6              0              0   London
## 7              0              0   London
## 8              0              0 Barcelona
## 9              0              0   London
## 10             0              0   London
## 11             0              0   London
## 12             0              0   London

```

Removing In this part of the data cleaning we are going to remove all the blank rows, the duplicates and strange values that may affect our analysis.

```

# Remove blank rows and columnsn
mydb <- mydb %>% remove_empty(c("rows", "cols"))

# Removing entries with too high and too low age
mydb <- mydb %>% filter(age <= 80 & age >= 16)
mydb <- mydb %>% filter(job_involvement <= 4)
mydb <- mydb %>% filter(num_companies_worked >= 0)

```

Therefore, as we can see, this line of code did not affected our dataset. So, this mean that there are no rows or columns that are empty.

Now, we are going to pass to the study of duplicates, by the *employee_number* variable that we suggest it is the key.

```

# Duplicates removal
mydb %>% get_dupes(employee_number)

```

```

##   employee_number dupe_count age attrition    business_travel daily_rate

```

## 1	4	2 37	Yes	Travel_Rarely	1373
## 2	4	2 37	Yes	Travel_Rarely	1373
## 3	86	2 45	No	Travel_Rarely	1339
## 4	86	2 45	No	Travel_Rarely	1339
## 5	94	2 29	No	Travel_Rarely	1328
## 6	94	2 29	No	Travel_Rarely	1328
## 7	165	2 58	Yes	Travel_Rarely	147
## 8	165	2 58	Yes	Travel_Rarely	147
## 9	182	2 39	No	Travel_Rarely	1329
## 10	182	2 39	No	Travel_Rarely	1329
## 11	194	2 30	No	Travel_Rarely	438
## 12	194	2 30	No	Travel_Rarely	438
## 13	204	2 35	No	Travel_Rarely	662
## 14	204	2 35	No	Travel_Rarely	662
## 15	243	2 19	Yes	Travel_Rarely	303
## 16	243	2 19	Yes	Travel_Rarely	303
## 17	247	2 34	No	Travel_Rarely	629
## 18	247	2 34	No	Travel_Rarely	629
## 19	308	2 31	No	Non-Travel	979
## 20	308	2 31	No	Non-Travel	979
## 21	346	2 29	No	Travel_Rarely	665
## 22	346	2 29	No	Travel_Rarely	665
## 23	468	2 42	No	Travel_Rarely	810
## 24	468	2 42	No	Travel_Rarely	810
## 25	469	2 29	No	Non-Travel	746
## 26	469	2 29	No	Non-Travel	746
## 27	470	2 42	No	Travel_Rarely	544
## 28	470	2 42	No	Travel_Rarely	544
## 29	497	2 27	No	Travel_Rarely	1469
## 30	497	2 27	No	Travel_Rarely	1469
## 31	498	2 28	No	Travel_Rarely	304
## 32	498	2 28	No	Travel_Rarely	304
## 33	1175	2 28	Yes	Travel_Rarely	1485
## 34	1175	2 28	Yes	Travel_Rarely	1485
## 35	1201	2 53	No	Travel_Rarely	1223
## 36	1201	2 53	No	Travel_Rarely	1223
## 37	1235	2 32	No	Travel_Frequently	1316
## 38	1235	2 32	No	Travel_Frequently	1316
## 39	1310	2 39	Yes	Travel_Rarely	360
## 40	1310	2 39	Yes	Travel_Rarely	360
## 41	1355	2 38	No	Travel_Rarely	1009
## 42	1355	2 38	No	Travel_Rarely	1009
## 43	1481	2 44	No	Travel_Frequently	383
## 44	1481	2 44	No	Travel_Frequently	383
## 45	1533	2 23	No	Travel_Rarely	507
## 46	1533	2 23	No	Travel_Rarely	507
## 47	1549	2 33	No	Travel_Rarely	589
## 48	1549	2 33	No	Travel_Rarely	589
## 49	1550	2 40	No	Travel_Rarely	898
## 50	1550	2 40	No	Travel_Rarely	898
## 51	1681	2 33	No	Non-Travel	530
## 52	1681	2 65	No	Non-Travel	530
## 53	1837	2 58	No	Travel_Frequently	1216
## 54	1837	2 58	No	Travel_Frequently	1216

## 55	1871	2	39	No	Travel_Rarely	835
## 56	1871	2	39	No	Travel_Rarely	835
## 57	1883	2	29	No	Travel_Rarely	592
## 58	1883	2	29	No	Travel_Rarely	592
## 59	1918	2	26	No	Travel_Frequently	1096
## 60	1918	2	26	No	Travel_Frequently	1096
## 61	1932	2	34	No	Travel_Frequently	735
## 62	1932	2	34	No	Travel_Frequently	735
## 63	2012	2	40	No	Travel_Rarely	543
## 64	2012	2	54	No	Travel_Rarely	543
## 65	NA	3	25	No	Travel_Rarely	622
## 66	NA	3	34	No	Travel_Rarely	1320
## 67	NA	3	31	No	Travel_Rarely	196
##	department	distance_from_home	education	education_field		
## 1	Research & Development		2	2	Other	
## 2	Research & Development		2	2	Other	
## 3	Research & Development		7	3	Life Sciences	
## 4	Research & Development		7	3	Life Sciences	
## 5	Research & Development		2	3	Life Sciences	
## 6	Research & Development		2	3	Life Sciences	
## 7	Research & Development		23	4	Medical	
## 8	Research & Development		23	4	Medical	
## 9	Sales		4	4	Life Sciences	
## 10	Sales		4	4	Life Sciences	
## 11	Research & Development		18	3	Life Sciences	
## 12	Research & Development		18	3	Life Sciences	
## 13	Sales		1	5	Marketing	
## 14	Sales		1	5	Marketing	
## 15	Research & Development		2	3	Life Sciences	
## 16	Research & Development		2	3	Life Sciences	
## 17	Research & Development		27	2	Medical	
## 18	Research & Development		27	2	Medical	
## 19	Research & Development		1	4	Medical	
## 20	Research & Development		1	4	Medical	
## 21	Research & Development		15	3		
## 22	Research & Development		15	3	Technical Degree	
## 23	Research & Development		23	5	Life Sciences	
## 24	Research & Development		23	5	Life Sciences	
## 25	Sales		2	3	Life Sciences	
## 26	Sales		2	3	Life Sciences	
## 27	Human Resources		2	1	Technical Degree	
## 28	Human Resources		2	1	Technical Degree	
## 29	Research & Development		1	2	Medical	
## 30	Research & Development		1	2	Medical	
## 31	Sales		9	4	Life Sciences	
## 32	Sales		9	4	Life Sciences	
## 33	Research & Development		12	1	Life Sciences	
## 34	Research & Development		12	1	Life Sciences	
## 35	Research & Development		7	2	Medical	
## 36	Research & Development		7	2	Medical	
## 37	Research & Development		2	2	Life Sciences	
## 38	Research & Development		2	2	Life Sciences	
## 39	Research & Development		23	3	Medical	
## 40	Research & Development		23	3	Medical	

## 41	Sales	2	2	Life Sciences	
## 42	Sales	2	2	Technical Degree	
## 43	Sales	1	5	Marketing	
## 44	Sales	1	5	Marketing	
## 45	Research & Development	20	1	Life Sciences	
## 46	Research & Development	20	1	Technical Degree	
## 47	Research & Development	28	4	Life Sciences	
## 48	Research & Development	28	4	Life Sciences	
## 49	Human Resources	6	2	Medical	
## 50	Human Resources	6	2	Medical	
## 51	Sales	16	3	Life Sciences	
## 52	Sales	16	3	Life Sciences	
## 53	Research & Development	15	4	Life Sciences	
## 54	Research & Development	15	4	Technical Degree	
## 55	Research & Development	19	4	Other	
## 56	Research & Development	19	4	Other	
## 57	Research & Development	7	3	Life Sciences	
## 58	Research & Development	7	3	Life Sciences	
## 59	Research & Development	6	3	Other	
## 60	Research & Development	6	3	Other	
## 61	Research & Development	22	4	Other	
## 62	Research & Development	22	4	Technical Degree	
## 63	Research & Development	1	4	Life Sciences	
## 64	Research & Development	1	4	Life Sciences	
## 65	Sales	13	1	Medical	
## 66	Research & Development	20	3	Technical Degree	
## 67	Sales	29	4	Marketing	
##	employee_count	environment_satisfaction	gender	hourly_rate	job_involvement
## 1	1	4	Male	92	2
## 2	1	4	Male	92	2
## 3	1	2	Male	59	3
## 4	1	2	Male	59	3
## 5	1	3	Male	76	3
## 6	1	3	Male	76	3
## 7	1	4	Female	94	3
## 8	1	4	Female	94	3
## 9	1	4	Female	47	2
## 10	1	4	Female	47	2
## 11	1	1	Female	75	3
## 12	1	1	Female	75	3
## 13	1	3	Male	94	3
## 14	1	3	Male	94	3
## 15	1	2	Male	47	2
## 16	1	2	Male	47	2
## 17	1	4	Female	95	3
## 18	1	4	Female	95	3
## 19	1	3	Male	90	1
## 20	1	3	Male	90	1
## 21	1	3	Male	60	3
## 22	1	3	Male	60	3
## 23	1	1	Female	44	3
## 24	1	1	Female	44	3
## 25	1	4	Male	61	3
## 26	1	4	Male	61	3

## 27	1	3	Male	52	3
## 28	1	3	Male	52	3
## 29	1	4	Male	82	3
## 30	1	4	Female	82	3
## 31	1	2	Male	92	3
## 32	1	2	Male	92	3
## 33	1	3	Female	79	3
## 34	1	3	Female	79	3
## 35	1	4	Female	50	3
## 36	1	4	Female	50	3
## 37	1	NA	Female	NA	3
## 38	1	NA	Female	NA	3
## 39	1	3	Male	93	3
## 40	1	3	Female	93	3
## 41	1	2	Female	31	3
## 42	1	2	Female	31	3
## 43	1	1	Female	79	3
## 44	1	1	Female	79	3
## 45	1	1	Male	97	3
## 46	1	1	Male	97	3
## 47	1	2	Male	79	3
## 48	1	2	Male	79	3
## 49	1	3	Male	38	3
## 50	1	3	Male	38	3
## 51	1	3	Female	36	3
## 52	1	3	Female	36	3
## 53	1	1	Male	87	3
## 54	1	1	Male	87	3
## 55	1	4	Male	41	3
## 56	1	4	Male	41	3
## 57	1	4	Male	59	3
## 58	1	4	Male	59	3
## 59	1	3	Male	61	4
## 60	1	3	Female	61	4
## 61	1	3	Male	86	2
## 62	1	3	Male	86	2
## 63	1	1	Male	83	3
## 64	1	1	Male	83	3
## 65	1	2	Male	40	3
## 66	1	3	Female	89	4
## 67	1	1		91	2
##	job_level	job_role	job_satisfaction	marital_status	
## 1	1	Laboratory Technician	3	Single	
## 2	1	Laboratory Technician	3	Single	
## 3	3	Research Scientist	1	Divorced	
## 4	3	Research Scientist	1	Divorced	
## 5	1	Research Scientist	2	Married	
## 6	1	Research Scientist	2	Married	
## 7	3	Healthcare Representative	4	Married	
## 8	3	Healthcare Representative	4	Married	
## 9	2	Sales Executive	3	Married	
## 10	2	Sales Executive	3	Married	
## 11	1	Research Scientist	3	Single	
## 12	1	Research Scientist	3	Single	

## 13	3	Sales Executive	2	Married
## 14	3	Sales Executive	2	Married
## 15	1	Laboratory Technician	4	Single
## 16	1	Laboratory Technician	4	Single
## 17	1	Research Scientist	2	Single
## 18	1	Research Scientist	2	Single
## 19	2	Manufacturing Director	3	Married
## 20	2	Manufacturing Director	3	Married
## 21	1	Research Scientist	4	Single
## 22	1	Research Scientist	4	Single
## 23	4	Research Director	4	Single
## 24	4	Research Director	4	Single
## 25	2	Sales Executive	3	Married
## 26	2	Sales Executive	3	Married
## 27	1	Human Resources	3	Divorced
## 28	1	Human Resources	3	Divorced
## 29	1	Laboratory Technician	2	Divorced
## 30	1	Laboratory Technician	2	Divorced
## 31	2	Sales Executive	4	Single
## 32	2	Sales Executive	4	Single
## 33	1	Laboratory Technician	4	Married
## 34	1	Laboratory Technician	4	Married
## 35	5	Manager	3	Divorced
## 36	5	Manager	3	Divorced
## 37	2	Research Scientist	3	Single
## 38	2	Research Scientist	3	Single
## 39	1	Research Scientist	1	Single
## 40	1	Research Scientist	1	Single
## 41	2	Sales Executive	1	Divorced
## 42	2	Sales Executive	1	Divorced
## 43	2	Sales Executive	3	Married
## 44	2	Sales Executive	3	Married
## 45	2	Laboratory Technician	3	Single
## 46	2	Laboratory Technician	3	Single
## 47	2	Laboratory Technician	3	Married
## 48	2	Laboratory Technician	3	Married
## 49	4	Manager	4	Single
## 50	4	Manager	4	Single
## 51	2	Sales Executive	4	Divorced
## 52	2	Sales Executive	4	Divorced
## 53	4	Research Director	3	Married
## 54	4	Research Director	3	Married
## 55	2	Research Scientist	4	Divorced
## 56	2	Research Scientist	4	Divorced
## 57	1	Laboratory Technician	1	Single
## 58	1	Laboratory Technician	1	Single
## 59	1	Laboratory Technician	4	Married
## 60	1	Laboratory Technician	4	Married
## 61	2	Research Scientist	4	Married
## 62	2	Research Scientist	4	Married
## 63	1	Laboratory Technician	4	Married
## 64	1	Laboratory Technician	4	Married
## 65	1	Sales Representative	3	Married
## 66	1	Research Scientist	3	Married

## 67	2	Sales Executive	4		
##	monthly_income	monthly_rate	num_companies_worked	over18	over_time
## 1	2090	2396	6	Y	Yes
## 2	2090	2396	6	Y	Yes
## 3	9724	18787	2	Y	No
## 4	9724	18787	2	Y	No
## 5	2703	4956	0	Y	No
## 6	2703	4956	0	Y	No
## 7	10312	3465	1	Y	No
## 8	10312	3465	1	Y	No
## 9	5902	14590	4	Y	No
## 10	5902	14590	4	Y	No
## 11	2632	23910	1	Y	No
## 12	2632	23910	1	Y	No
## 13	7295	11439	1	Y	No
## 14	7295	11439	1	Y	No
## 15	1102	9241	1	Y	No
## 16	1102	9241	1	Y	No
## 17	2311	5711	2	Y	No
## 18	2311	5711	2	Y	No
## 19	4345	4381	0	Y	No
## 20	4345	4381	0	Y	No
## 21	2340	22673	1	Y	No
## 22	2340	22673	1	Y	No
## 23	15992	15901	2	Y	No
## 24	15992	15901	2	Y	No
## 25	4649	16928	1	Y	No
## 26	4649	16928	1	Y	No
## 27	2696	24017	0	Y	Yes
## 28	2696	24017	0	Y	Yes
## 29	3816	17881	1	Y	No
## 30	3816	17881	1	Y	No
## 31	5253	20750	1	Y	No
## 32	5253	20750	1	Y	No
## 33	2515	22955	1	Y	Yes
## 34	2515	22955	1	Y	Yes
## 35	18606	18640	3	Y	No
## 36	18606	18640	3	Y	No
## 37	4998	2338	4	Y	Yes
## 38	4998	2338	4	Y	Yes
## 39	3904	22154	0	Y	No
## 40	3904	22154	0	Y	No
## 41	6893	19461	3	Y	No
## 42	6893	19461	3	Y	No
## 43	4768	9282	7	Y	No
## 44	4768	9282	7	Y	No
## 45	2272	24812	0	Y	No
## 46	2272	24812	0	Y	No
## 47	5207	22949	1	Y	Yes
## 48	5207	22949	1	Y	Yes
## 49	16437	17381	1	Y	Yes
## 50	16437	17381	1	Y	Yes
## 51	5368	16130	1	Y	Yes
## 52	5368	16130	1	Y	Yes

## 53	15787	21624	2	Y	Yes
## 54	15787	21624	2	Y	Yes
## 55	3902	5141	8	Y	No
## 56	3902	5141	8	Y	No
## 57	2062	19384	3	Y	No
## 58	2062	19384	3	Y	No
## 59	2544	7102	0	Y	No
## 60	2544	7102	0	Y	No
## 61	5747	26496	1	Y	Yes
## 62	5747	26496	1	Y	Yes
## 63	2406	4060	8	Y	No
## 64	2406	4060	8	Y	No
## 65	2096	26376	1	Y	No
## 66	2585	21643	0	Y	No
## 67	5468	13402	1	Y	No
##	percent_salary_hike	performance_rating	relationship_satisfaction		
## 1	15	3		2	
## 2	15	3		2	
## 3	17	3		3	
## 4	17	3		3	
## 5	23	4		4	
## 6	23	4		4	
## 7	12	3		4	
## 8	12	3		4	
## 9	14	3		3	
## 10	14	3		3	
## 11	14	3		3	
## 12	14	3		3	
## 13	13	3		1	
## 14	13	3		1	
## 15	22	4		3	
## 16	22	4		3	
## 17	15	3		4	
## 18	15	3		4	
## 19	12	3		4	
## 20	12	3		4	
## 21	19	3		1	
## 22	19	3		1	
## 23	14	3		2	
## 24	14	3		2	
## 25	14	3		1	
## 26	14	3		1	
## 27	11	3		3	
## 28	11	3		3	
## 29	11	3		2	
## 30	11	3		2	
## 31	16	3		4	
## 32	16	3		4	
## 33	11	3		4	
## 34	11	3		4	
## 35	18	3		2	
## 36	18	3		2	
## 37	14	3		NA	
## 38	14	3		NA	

## 39	13	3	1
## 40	13	3	1
## 41	15	3	4
## 42	15	3	4
## 43	12	3	3
## 44	12	3	3
## 45	14	3	2
## 46	14	3	2
## 47	12	3	2
## 48	12	3	2
## 49	21	4	4
## 50	21	4	4
## 51	25	4	3
## 52	25	4	3
## 53	14	3	2
## 54	14	3	2
## 55	14	3	2
## 56	14	3	2
## 57	14	3	2
## 58	14	3	2
## 59	18	3	1
## 60	18	3	1
## 61	15	3	2
## 62	15	3	2
## 63	19	3	3
## 64	19	3	3
## 65	11	3	3
## 66	NA	3	4
## 67	14	3	1
##	standard_hours	stock_option_level	total_working_years
## 1	80	0	7
## 2	80	0	7
## 3	80	1	25
## 4	80	1	25
## 5	80	1	6
## 6	80	1	6
## 7	80	1	40
## 8	80	1	40
## 9	80	1	17
## 10	80	1	17
## 11	80	0	5
## 12	80	0	5
## 13	80	2	10
## 14	80	2	10
## 15	80	0	1
## 16	80	0	1
## 17	80	0	9
## 18	80	0	9
## 19	80	1	6
## 20	80	1	6
## 21	80	0	6
## 22	80	0	6
## 23	80	0	16
## 24	80	0	16

## 25	80	1	4
## 26	80	1	4
## 27	80	1	4
## 28	80	1	4
## 29	80	1	5
## 30	80	1	5
## 31	80	0	7
## 32	80	0	7
## 33	80	0	1
## 34	80	0	1
## 35	80	1	26
## 36	80	1	26
## 37	80	0	10
## 38	80	0	10
## 39	80	0	6
## 40	80	0	6
## 41	80	1	11
## 42	80	1	11
## 43	80	1	11
## 44	80	1	11
## 45	80	0	5
## 46	80	0	5
## 47	80	1	15
## 48	80	1	15
## 49	80	0	21
## 50	80	0	21
## 51	80	1	7
## 52	80	1	7
## 53	80	0	23
## 54	80	0	23
## 55	80	3	7
## 56	80	3	7
## 57	80	0	11
## 58	80	0	11
## 59	80	1	8
## 60	80	1	8
## 61	80	0	16
## 62	80	0	16
## 63	80	2	8
## 64	80	2	8
## 65	80	NA	7
## 66	80	0	NA
## 67	80	2	13
##	training_times_last_year	work_life_balance	years_at_company
## 1	3	3	0
## 2	3	3	0
## 3	2	3	1
## 4	2	3	1
## 5	3	3	5
## 6	3	3	5
## 7	3	2	40
## 8	3	2	40
## 9	1	4	15
## 10	1	4	15

## 11	4	2	5
## 12	4	2	5
## 13	3	3	10
## 14	3	3	10
## 15	3	2	1
## 16	3	2	1
## 17	3	3	3
## 18	3	3	3
## 19	2	3	5
## 20	2	3	5
## 21	NA	3	6
## 22	NA	3	6
## 23	2	3	1
## 24	2	3	1
## 25	3	2	4
## 26	3	2	4
## 27	5	3	3
## 28	5	3	3
## 29	2	3	5
## 30	2	3	5
## 31	1	3	7
## 32	1	3	7
## 33	4	2	1
## 34	4	2	1
## 35	6	3	7
## 36	6	3	7
## 37	2	3	8
## 38	2	3	8
## 39	2	3	5
## 40	2	3	5
## 41	3	3	7
## 42	3	3	7
## 43	4	2	1
## 44	4	2	1
## 45	2	3	4
## 46	2	3	4
## 47	3	3	15
## 48	3	3	15
## 49	2	3	21
## 50	2	3	21
## 51	2	3	6
## 52	2	3	6
## 53	3	3	2
## 54	3	3	2
## 55	2	3	2
## 56	2	3	2
## 57	2	3	2
## 58	2	3	2
## 59	3	3	7
## 60	3	3	7
## 61	3	3	15
## 62	3	3	15
## 63	3	2	1
## 64	3	2	1

## 65	1	3	7
## 66	5	2	1
## 67	3	3	12
##	years_in_current_role	years_since_last_promotion	years_with_curr_manager
## 1	0	0	0
## 2	0	0	0
## 3	0	0	0
## 4	0	0	0
## 5	4	0	4
## 6	4	0	4
## 7	10	15	6
## 8	10	15	6
## 9	11	5	9
## 10	11	5	9
## 11	4	0	4
## 12	4	0	4
## 13	8	0	6
## 14	8	0	6
## 15	0	1	0
## 16	0	1	0
## 17	2	1	2
## 18	2	1	2
## 19	4	1	4
## 20	4	1	4
## 21	5	1	5
## 22	5	1	5
## 23	0	0	0
## 24	0	0	0
## 25	3	0	2
## 26	3	0	2
## 27	2	1	0
## 28	2	1	0
## 29	2	0	4
## 30	2	0	4
## 31	5	0	7
## 32	5	0	7
## 33	1	0	0
## 34	1	0	0
## 35	7	4	7
## 36	7	4	7
## 37	7	0	7
## 38	7	0	7
## 39	2	0	3
## 40	2	0	3
## 41	7	1	7
## 42	7	1	7
## 43	0	0	0
## 44	0	0	0
## 45	3	1	2
## 46	3	1	2
## 47	14	5	7
## 48	14	5	7
## 49	7	7	7
## 50	7	7	7

## 51	5	1	2
## 52	5	1	2
## 53	2	2	2
## 54	2	2	2
## 55	2	2	2
## 56	2	2	2
## 57	3	1	2
## 58	3	1	2
## 59	7	7	7
## 60	7	7	7
## 61	10	6	11
## 62	10	6	11
## 63	0	0	0
## 64	0	0	0
## 65	4	0	6
## 66	0	0	0
## 67	7	5	7
##	city		
## 1	London		
## 2	London		
## 3	London		
## 4	London		
## 5	London		
## 6	London		
## 7	London		
## 8	London		
## 9	London		
## 10	London		
## 11	London		
## 12	London		
## 13	London		
## 14	London		
## 15	London		
## 16	London		
## 17	London		
## 18	London		
## 19	London		
## 20	London		
## 21	Barcelona		
## 22	Barcelona		
## 23	London		
## 24	London		
## 25	London		
## 26	London		
## 27	London		
## 28	London		
## 29	London		
## 30	London		
## 31	London		
## 32	London		
## 33	London		
## 34	London		
## 35	London		
## 36	London		


```
## 37    London
## 38    London
## 39    London
## 40    London
## 41    London
## 42    London
## 43 Barcelona
## 44 Barcelona
## 45    London
## 46    London
## 47    London
## 48    London
## 49    London
## 50    London
## 51 Barcelona
## 52 Barcelona
## 53    London
## 54    London
## 55    London
## 56    London
## 57    London
## 58    London
## 59    London
## 60    London
## 61 Barcelona
## 62 Barcelona
## 63    London
## 64    London
## 65    London
## 66    London
## 67 Barcelona
```

```
mydb <- mydb %>% distinct(employee_number, .keep_all= TRUE)
```

Hence, now it is time to check the n's.

```
## attrition      n    percent
##           No 1221 0.8374486
##           Yes  237 0.1625514
```

```
## business_travel      n    percent
##                3 0.002057613
##           Non-Travel 150 0.102880658
## Travel_Frequently  273 0.187242798
## Travel_Rarely 1032 0.707818930
```

```
##                department      n    percent
##                3 0.002061856
##                HR    3 0.002061856
##           Human Resources  61 0.041924399
## Research & Development  931 0.639862543
## Research and Development  14 0.009621993
##                Sales  443 0.304467354
```

```

## education_field  n      percent
##                3 0.002066116
## Human Resources  27 0.018595041
## Life Sciences   600 0.413223140
## Marketing       157 0.108126722
## Medical         456 0.314049587
## Other           81 0.055785124
## Technical Degree 128 0.088154270

## gender  n      percent
##        1 0.0006901311
## female  10 0.0069013112
## Female  564 0.3892339545
## female   3 0.0020703934
## male    14 0.0096618357
## Male    857 0.5914423741

##                job_role  n      percent
##                3 0.002071823
## Healthcare Representative 129 0.089088398
## Human Resources          52 0.035911602
## Laboratory Technician    256 0.176795580
## Manager                  101 0.069751381
## Manufacturing Director   140 0.096685083
## Research Director        78 0.053867403
## Research Scientist       284 0.196132597
## Sales Executive          322 0.222375691
## Sales Representative     83 0.057320442

## marital_status  n      percent
##                1 0.0006920415
## Divorced        325 0.2249134948
## Marrie           3 0.0020761246
## Married         656 0.4539792388
## Single          454 0.3141868512
## Soltero         6 0.0041522491

## over18  n      percent
##         5 0.003462604
## N        7 0.004847645
## Y       1432 0.991689751

## city  n      percent
## Barcelona  262 0.182070883
## Barcelone   1 0.000694927
## London    1176 0.817234190

```

To conclude, the cleaning of the dataset, we are going to remove every line with at least one empty gap.

```
mydb <- mydb %>% drop_na()
```

From now on we can easily proceed with our analysis.

Analysis

...

Conclusions

...

Referemces

1. *M., L. (2004). Moneyball: The art of winning an unfair game. New york: John Wiley and sons.*
2. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
3. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
4. David Robinson, Alex Hayes and Simon Couch (2021). broom: Convert Statistical Objects into Tidy Tibbles. R package version 0.7.9. <https://CRAN.R-project.org/package=broom>