

HR data analysis and contract termination prediction

Gabriel Scozzarro

1.0 Introduction

2.0 Preparation and reading data

Data was imported from the set contained inside the file HRDataset_v14.csv

Feature	Description	Type
i..Employee_Name	Employee's full name	Text
EmpID	Employee ID is unique to each employee	Text
MarriedID	Is the person married (1 or 0 for yes or no)	Binary
MaritalStatusID	Marital status code that matches the text field MaritalDesc	Integer
GenderID	Gender ID that matches Sex	Binary
EmpStatusID	Employment status code that matches text field EmploymentStatus	Integer
DeptID	Department ID code that matches the department the employee works in	Integer
PerfScoreID	Performance Score code that matches the employee's most recent performance score	Integer
FromDiversityJobFairID	Was the employee sourced from the Diversity job fair? 1 or 0 for yes or no	Binary
Salary	The person's annual pay rate	Float
Termd	Has this employee been terminated - 1 or 0	Binary
PositionID	An integer indicating the person's position	Integer
Position	The text name/title of the position the person has	Text
State	The state that the person lives in	Text
Zip	The zip code for the employee	Text
DOB	Date of Birth for the employee	Date
Sex	Sex - M or F	Text
MaritalDesc	The marital status of the person (divorced, single, widowed, separated, etc)	Text
CitizenDesc	Label for whether the person is a Citizen or Eligible NonCitizen	Text
HispanicLatino	Yes or No field for whether the employee is Hispanic/Latino	Text
RaceDesc	Description/text of the race the person identifies with	Text
DateofHire	Date the person was hired	Date
DateofTermination	Date the person was terminated, only populated if, in fact, Termd = 1	Date
TermReason	A text reason / description for why the person was terminated	Text
EmploymentStatus	A description/category of the person's employment status. Anyone currently working full time = Active	Text
Department	Name of the department that the person works in	Text
ManagerName	The name of the person's immediate manager	Text
ManagerID	A unique identifier for each manager	Integer
RecruitmentSource	The name of the recruitment source where the employee was recruited from	Text
PerformanceScore	Performance Score text/category (Fully Meets, Partially Meets, PIP, Exceeds)	Text
EngagementSurvey	Results from the last engagement survey, managed by our external partner	Float
EmpSatisfaction	A basic satisfaction score between 1 and 5, as reported on a recent employee satisfaction survey	Integer
SpecialProjectsCount	The number of special projects that the employee worked on during the last 6 months	Integer
LastPerformanceReview_Date	The most recent date of the person's last performance review	Date
DaysLateLast30	The number of times that the employee was late to work during the last 30 days	Integer
Absences	The number of times the employee was absent from work	Integer

i..Employee_Name	EmpID	MarriedID	MaritalStatusID	GenderID	EmpStatusID	DeptID	PerfScoreID	FromDiversityJobFairID
Length:311	Min. :10001	Min. :0.0000	Min. :0.0000	Min. :0.0000	Min. :1.000	Min. :1.000	Min. :1.000	Min. :0.00000
Class :character	1st Qu.:10078	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:1.000	1st Qu.:5.000	1st Qu.:3.000	1st Qu.:0.00000
Mode :character	Median :10156	Median :0.0000	Median :1.0000	Median :0.0000	Median :1.000	Median :5.000	Median :3.000	Median :0.00000
NA	Mean :10156	Mean :0.3987	Mean :0.8103	Mean :0.4341	Mean :2.392	Mean :4.611	Mean :2.977	Mean :0.09325
NA	3rd Qu.:10234	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:0.00000
NA	Max. :10311	Max. :1.0000	Max. :4.0000	Max. :1.0000	Max. :5.000	Max. :6.000	Max. :4.000	Max. :1.00000
NA	NA	NA	NA	NA	NA	NA	NA	NA

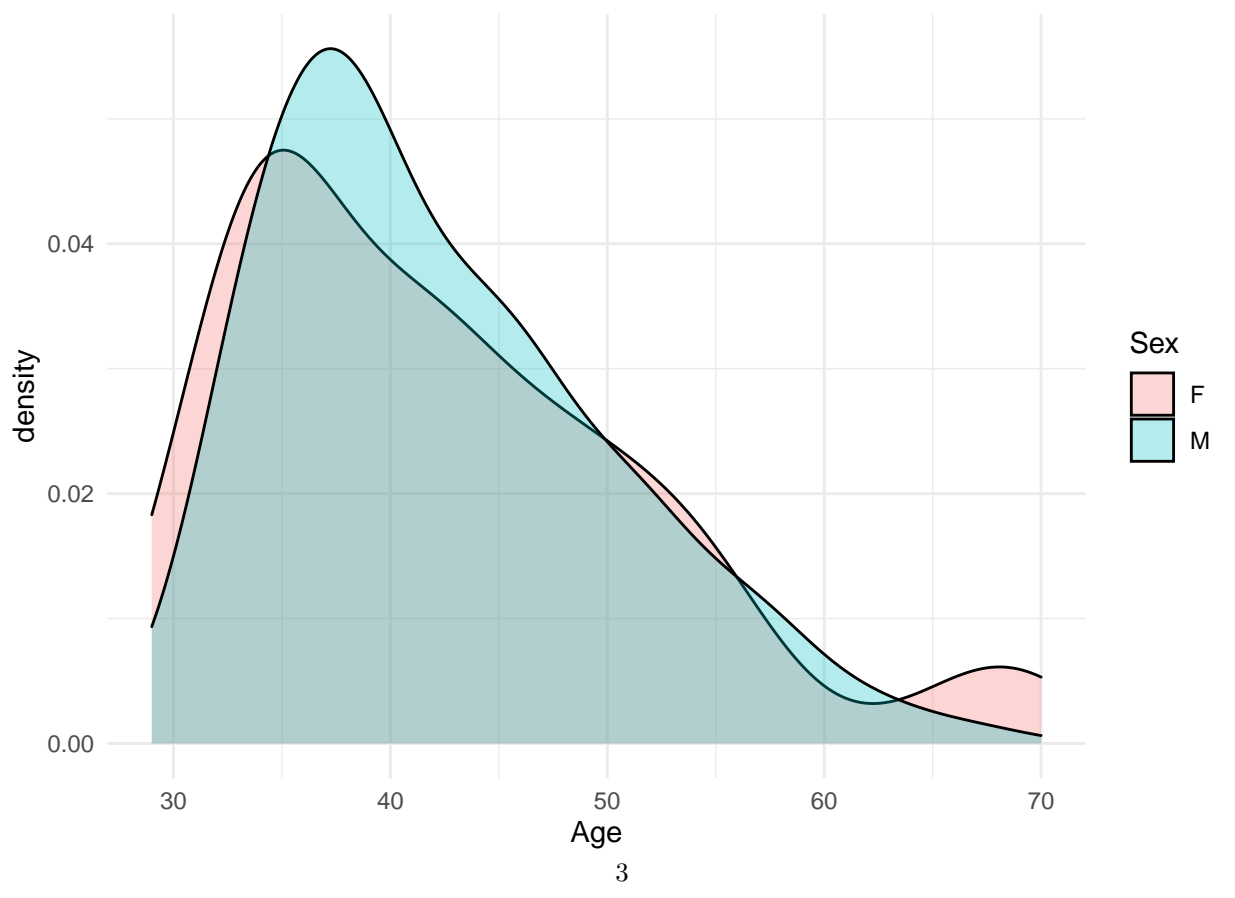
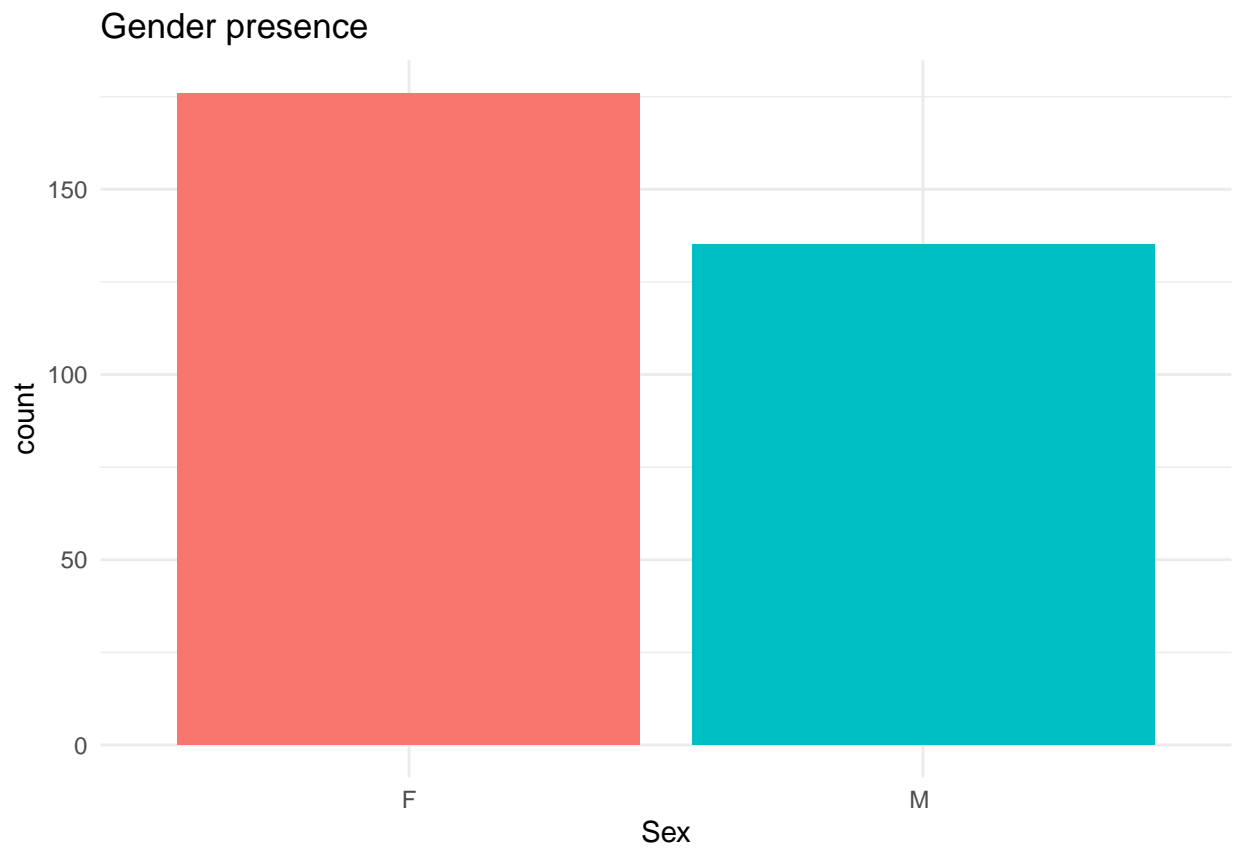
Salary	Termd	PositionID	Position	State	Zip	DOB	Sex	MaritalDesc
Min. : 45046	Min. :0.0000	Min. : 1.00	Length:311	Length:311	Min. : 1013	Length:311	Length:311	Length:311
1st Qu.: 55502	1st Qu.:0.0000	1st Qu.:18.00	Class :character	Class :character	1st Qu.: 1902	Class :character	Class :character	Class :character
Median : 62810	Median :0.0000	Median :19.00	Mode :character	Mode :character	Median : 2132	Mode :character	Mode :character	Mode :character
Mean : 69021	Mean :0.3344	Mean :16.85	NA	NA	Mean : 6555	NA	NA	NA
3rd Qu.: 72036	3rd Qu.:1.0000	3rd Qu.:20.00	NA	NA	3rd Qu.: 2355	NA	NA	NA
Max. :250000	Max. :1.0000	Max. :30.00	NA	NA	Max. :98052	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA

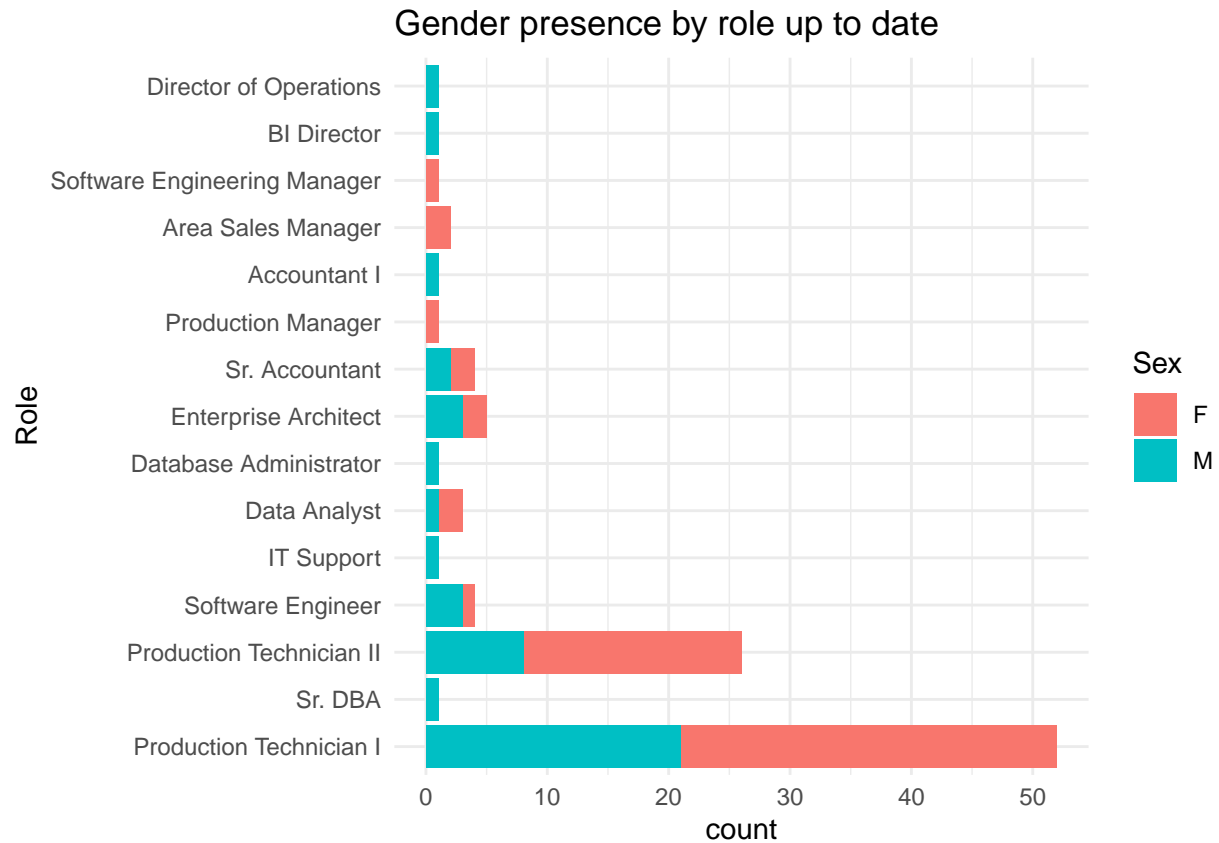
CitizenDesc	HispanicLatino	RaceDesc	DateofHire	DateofTermination	TermReason	EmploymentStatus	Department	ManagerName
Length:311	Length:311	Length:311	Length:311	Length:311	Length:311	Length:311	Length:311	Length:311
Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character
NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA

ManagerID	RecruitmentSource	PerformanceScore	EngagementSurvey	EmpSatisfaction	SpecialProjectsCount	LastPerformanceReview _Date	DaysLateLast30	Absences
Min. : 1.00	Length:311	Length:311	Min. :1.12	Min. :1.000	Min. :0.000	Length:311	Min. :0.0000	Min. : 1.00
1st Qu.:10.00	Class :character	Class :character	1st Qu.:3.69	1st Qu.:3.000	1st Qu.:0.000	Class :character	1st Qu.:0.0000	1st Qu.: 5.00
Median :15.00	Mode :character	Mode :character	Median :4.28	Median :4.000	Median :0.000	Mode :character	Median :0.0000	Median :10.00
Mean :14.57	NA	NA	Mean :4.11	Mean :3.891	Mean :1.219	NA	Mean :0.4148	Mean :10.24
3rd Qu.:19.00	NA	NA	3rd Qu.:4.70	3rd Qu.:5.000	3rd Qu.:0.000	NA	3rd Qu.:0.0000	3rd Qu.:15.00
Max. :39.00	NA	NA	Max. :5.00	Max. :5.000	Max. :8.000	NA	Max. :6.0000	Max. :20.00
NA's :8	NA	NA	NA	NA	NA	NA	NA	NA

The total amount of employees are 311 and 36 features was collected for each of them.

EDA





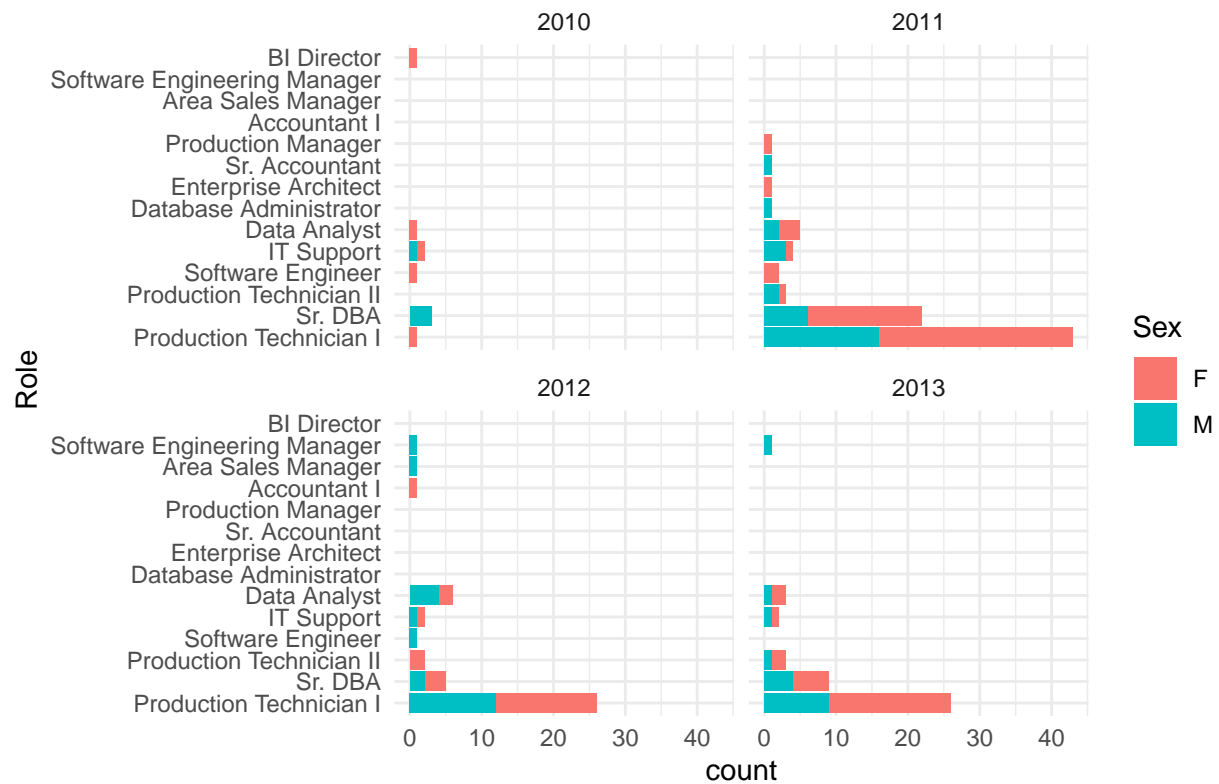
The company has a majority presence of women (176) compare to men (135) with a very close average age of respectively 42.5 for women and 42.3 for men. Regardless the average age we can see that the age distribution is shift to the left for both men and women meaning that the the majority of the employees are younger compare to the average age.

The company has a large an prevalent presence of women employed in the production as Product Technician I and Product Technician II. Managerial roles also has a 100% prevalence of women guiding fundamental aspects such as software engineering, sales and production.

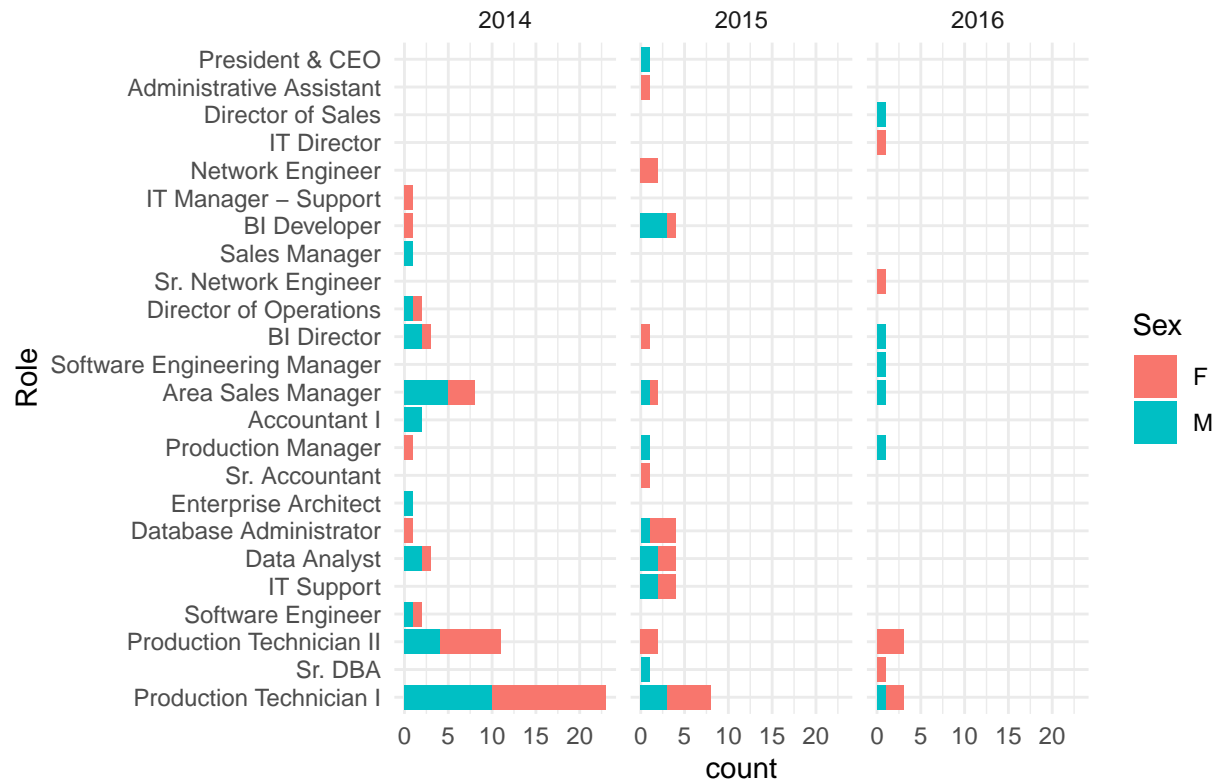
Enrolment by role and gender 2006 – 2009



Enrolment by role and gender 2010 – 2013



Enrolment by role and gender 2014 – 2016



Enrolment by role and gender 2017 – 2018

