

Human Resource Analytics

Pankaj Shah

5/29/2019

Human Resource Analytics

Out of curiousity I took this datasets from one of the Medium article. My small step effort to understand the HR analytics part of it.Thank you Nafee Afshin for such a beautiful blog and various R community Kagglers and Coders for breaking it down and make easy to understand various elements of Human Resource.

We all know that there is job opening in every company every now and then. Depends upon size of human resource how often they have their position open. It's not easy for both employer and employee to find the right working conditions, satisfaction level and many other things to consider for a perfect match. Apart from finding perfect match both the epmloyer and employee need to invest both time and money to train within the companies policy and working condition, if there is new tool to learn they have to invest the time to learn. CRISP-DM stands for CRoss Industry Standard Process for Data Mining.

When we apply for the job or start working for any company we should be familiar with Human Resource Department of the organisation. In some big organization Human resource (HR) departments seems increasingly turning to analytics department by having Business Intelligence people working for them to drive company hiring and workforce management more smoothly and within budget. As with every nooks and corner of the company (e.g. marketing, finance, etc), HR departments also do require analytics solutions which could be very unique to their domain and relational database between departments.

We might think in particular, HR departments would usually have 'small' datasets in relation to other departments which might not be true because HR department is very crucial department for growth of the company. They are more interpretation driven than predictive based model driven.

General skills can be transferred from organisation to organisation but how to apply those tools depends upon the goal of the company and the projects everyone is involved. It all depends upon what is the motive behind working. The goal of most of the company is to utilise its manpower in right direction so that it can propel its ship smoothly. There is always bumps on the way, someone leaving the company for better future, family and better offer when someone has skills & proper talent. It is definitely not in favorable condition for the company to lose hardworking and good ethics employee. But wouldn't it be ideal for the company to invest time to find the reason What makes employer leave the company. If it just because of current Manager,Employee Working Satisfaction, distance to commute, Working Environment etc. There are various factors if one need to consider what makes Employee/ Employee happy. It is balance of lot of things which play critical role. I am trying to explore some of this with this small datasets . We all love Revenue turnover whether we are employee or employee but no one likes Employee Churn over ratio or Attrition. Lets dive in and see what can we find from this datasets.

Attrition: It is basically the turnover rate of employees inside an organization.

The goal of this particular project would be find if there is any relationship between the variables and if we could find the insights for Employee Churn over the period of time. If possible I would love to do some Machine learning to predict the Churn ratio over period of time.

Just to make things little clear Human Analytics is not only gathering data about employee and increasing efficiency but also its aim is to return better ROI for the company and keep employee happy with the fulfillment of need and happier workplace environment. If we could better understand the patterns and take right steps we all will be benefitting from each other.

So lets start diving into our datasets by importing necessary libraries.

Library

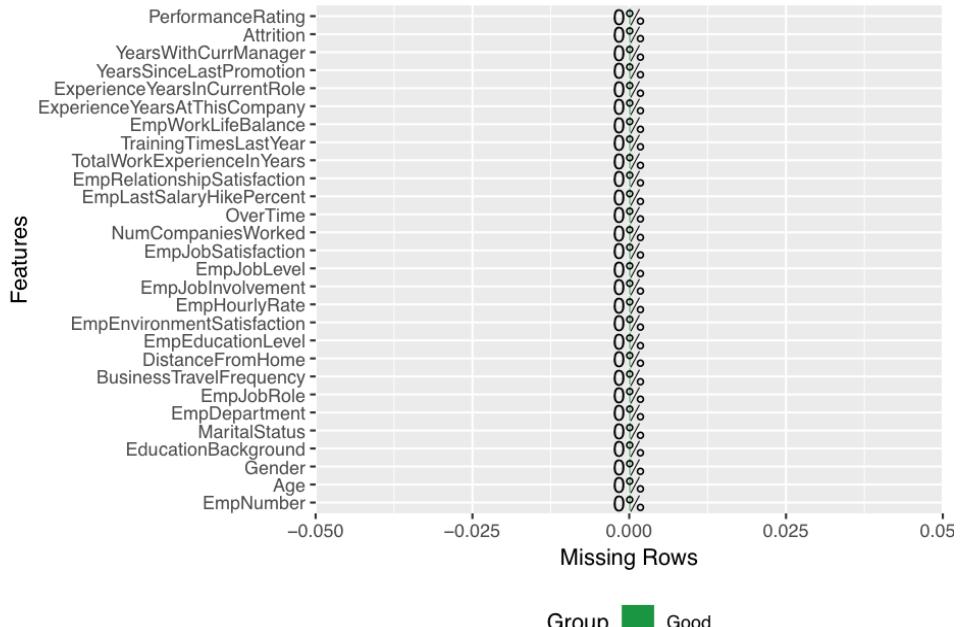
DATA INSIGHTS:

Lets have a glimpse of our datasets.

```
## Observations: 1,200
## Variables: 28
## $ EmpNumber <fct> E1001000, E1001006, E1001007, E10...
## $ Age <int> 32, 47, 40, 41, 60, 27, 50, 28, 3...
## $ Gender <fct> Male, Male, Male, Male, Male, Mal...
## $ EducationBackground <fct> Marketing, Marketing, Life Scienc...
## $ MaritalStatus <fct> Single, Single, Married, Divorced...
## $ EmpDepartment <fct> Sales, Sales, Sales, Human Resour...
## $ EmpJobRole <fct> Sales Executive, Sales Executive, ...
## $ BusinessTravelFrequency <fct> Travel_Rarely, Travel_Rarely, Tra...
## $ DistanceFromHome <int> 10, 14, 5, 10, 16, 10, 8, 1, 8, 1...
## $ EmpEducationLevel <int> 3, 4, 4, 4, 2, 4, 2, 3, 3, 3, ...
## $ EmpEnvironmentSatisfaction <int> 4, 4, 2, 1, 4, 4, 1, 1, 3, 1, ...
## $ EmpHourlyRate <int> 55, 42, 48, 73, 84, 32, 54, 67, 6...
## $ EmpJobInvolvement <int> 3, 3, 2, 2, 3, 3, 3, 1, 4, 3, 1, ...
## $ EmpJobLevel <int> 2, 2, 3, 5, 2, 3, 1, 1, 3, 3, 1, ...
## $ EmpJobSatisfaction <int> 4, 1, 1, 4, 1, 1, 2, 2, 1, 3, 3, ...
## $ NumCompaniesWorked <int> 1, 2, 5, 3, 8, 1, 7, 7, 9, 4, 2, ...
## $ OverTime <fct> No, No, Yes, No, No, No, Yes, ...
## $ EmpLastSalaryHikePercent <int> 12, 12, 21, 15, 14, 21, 15, 13, 1...
## $ EmpRelationshipSatisfaction <int> 4, 4, 3, 2, 4, 3, 4, 4, 1, 4, 3, ...
## $ TotalWorkExperienceInYears <int> 10, 20, 20, 23, 10, 9, 4, 10, 10, ...
## $ TrainingTimesLastYear <int> 2, 2, 2, 1, 4, 2, 4, 2, 4, 5, ...
## $ EmpWorkLifeBalance <int> 2, 3, 3, 2, 3, 2, 3, 3, 3, 4, 3, ...
## $ ExperienceYearsAtThisCompany <int> 10, 7, 18, 21, 2, 9, 2, 7, 8, 1, ...
## $ ExperienceYearsInCurrentRole <int> 7, 7, 13, 6, 2, 7, 2, 7, 7, 0, 2, ...
## $ YearsSinceLastPromotion <int> 0, 1, 1, 12, 2, 1, 2, 3, 0, 0, 1, ...
## $ YearsWithCurrManager <int> 8, 7, 12, 6, 2, 7, 2, 7, 5, 0, 4, ...
## $ Attrition <fct> No, No, No, No, No, No, Yes, ...
## $ PerformanceRating <int> 3, 3, 4, 3, 3, 4, 3, 3, 3, 3, ...
```

Exploratory Data Analysis

```
## [1] "Missing Data"
```



```
## $data.frame
##   name    size
## 1 hr_df 0.2 Mb
##
## $dimensions
##   rows columns
## 1 1200      28
##
## $column.details
##                                column class unique.values missing.count
## 1                  EmpNumber factor        1200            0
## 2                 EmpHourlyRate integer       71            0
## 3                   Age integer        43            0
## 4 TotalWorkExperienceInYears integer       40            0
## 5 ExperienceYearsAtThisCompany integer       37            0
## 6           DistanceFromHome integer       29            0
## 7             EmpJobRole factor        19            0
## 8 ExperienceYearsInCurrentRole integer       19            0
## 9          YearsWithCurrManager integer       18            0
## 10     YearsSinceLastPromotion integer       16            0
## 11    EmpLastSalaryHikePercent integer       15            0
## 12      NumCompaniesWorked integer       10            0
## 13     TrainingTimesLastYear integer        7            0
## 14      EducationBackground factor        6            0
## 15        EmpDepartment factor        6            0
## 16      EmpEducationLevel integer        5            0
## 17      EmpJobLevel integer        5            0
```

```

## 18 EmpEnvironmentSatisfaction integer      4      0
## 19          EmpJobInvolvement integer      4      0
## 20          EmpJobSatisfaction integer      4      0
## 21 EmpRelationshipSatisfaction integer      4      0
## 22          EmpWorkLifeBalance integer      4      0
## 23          MaritalStatus factor          3      0
## 24 BusinessTravelFrequency factor          3      0
## 25          PerformanceRating integer      3      0
## 26          Gender factor                2      0
## 27          OverTime factor               2      0
## 28          Attrition factor              2      0
##   missing.pct
## 1      0
## 2      0
## 3      0
## 4      0
## 5      0
## 6      0
## 7      0
## 8      0
## 9      0
## 10     0
## 11     0
## 12     0
## 13     0
## 14     0
## 15     0
## 16     0
## 17     0
## 18     0
## 19     0
## 20     0
## 21     0
## 22     0
## 23     0
## 24     0
## 25     0
## 26     0
## 27     0
## 28     0

```

We have almost 1200 rows of data with 28 variables. Total size of our datasets is 0.2MB.

Sanity Check:

Look for any missing Rows and the graph shows no missing rows.Aww thats so sweet we don't have any missing data at all. It seems like we have clean datasets which happens rarely in Data Analytics field and especially in People Analytics as lot of information are personal and because of privacy, confidentiality lots of data are not provided publicly.

Head

We will analyse most of the individuals variables at first and then try to link dependent and independent variables for both machine learning and Exploratory Data Analysis.

Lets start by Exploratory Data Analysis. Lets look at the top of the table to get sense of datasets.

```
##   EmpNumber Age Gender EducationBackground MaritalStatus EmpDepartment
## 1 E1001000 32 Male      Marketing      Single      Sales
## 2 E1001006 47 Male      Marketing      Single      Sales
##           EmpJobRole BusinessTravelFrequency DistanceFromHome
## 1 Sales Executive          Travel_Rarely            10
## 2 Sales Executive          Travel_Rarely            14
##   EmpEducationLevel EmpEnvironmentSatisfaction EmpHourlyRate
## 1                      3                         4                  55
## 2                      4                         4                  42
##   EmpJobInvolvement EmpJobLevel EmpJobSatisfaction NumCompaniesWorked
## 1                      3                         2                  4                  1
## 2                      3                         2                  1                  2
##   OverTime EmpLastSalaryHikePercent EmpRelationshipSatisfaction
## 1        No                   12                  4
## 2        No                   12                  4
##   TotalWorkExperienceInYears TrainingTimesLastYear EmpWorkLifeBalance
## 1                     10                  2                  2
## 2                     20                  2                  3
##   ExperienceYearsAtThisCompany ExperienceYearsInCurrentRole
## 1                           10                  7
## 2                           7                  7
##   YearsSinceLastPromotion YearsWithCurrManager Attrition PerformanceRating
## 1             0                   8      No            3
## 2             1                   7      No            3
```

Sample

By looking at the top of table we should usually get the sense of what kind of datasets we are dealing with. What does the column and value represents. But going one further step I usually love to randomly sample my datasets to find something new. It doesnt take long time but saves much time if I could catch something fishy in the beginning of the analysis.

If I have small set of data I usually use `sample_n` to pass around and get sense. We could also use `sample_frac` to see more random fraction of data.

```
##   EmpNumber Age Gender EducationBackground MaritalStatus EmpDepartment
## 1 1188    E100977 36 Male      Medical      Divorced Development
## 2 1148    E100898 35 Female     Medical      Single      Sales
##           EmpJobRole BusinessTravelFrequency DistanceFromHome
## 1 1188 Developer          Travel_Rarely            9
## 2 1148 Sales Executive          Travel_Rarely            1
##   EmpEducationLevel EmpEnvironmentSatisfaction EmpHourlyRate
## 1 1188                      2                         2                  48
## 2 1148                      3                         1                  80
##   EmpJobInvolvement EmpJobLevel EmpJobSatisfaction NumCompaniesWorked
## 1 1188                      2                         2                  2
## 2 1148                      3                         2                  3                  1
##   OverTime EmpLastSalaryHikePercent EmpRelationshipSatisfaction
## 1 1188 Yes                   11                  3
## 2 1148 No                    16                  4
##   TotalWorkExperienceInYears TrainingTimesLastYear EmpWorkLifeBalance
## 1 1188                     13                  2                  3
## 2 1148                     5                   3                  3
```

```

##      ExperienceYearsAtThisCompany ExperienceYearsInCurrentRole
## 1188                  3                      2
## 1148                  5                      4
##      YearsSinceLastPromotion YearsWithCurrManager Attrition
## 1188                  0                      2        No
## 1148                  0                      3        No
##      PerformanceRating
## 1188                  3
## 1148                  3

```

Data Lookup

At first glance, by running Random Sample all the datasets looks normal. Lets see how many unique value each column has so that we can get more gist of it.

If we would like to get more sense of data. I usually run the unique code in each column to know which field are interesting enough to diagnose. If any column have handful of unique values it is easy to look into without wasting lot of times.

```

## [1] Male   Female
## Levels: Female Male
## [1] Single Married Divorced
## Levels: Divorced Married Single

```

Duplicate data

```

## No duplicate combinations found of: EmpNumber
## # A tibble: 0 x 29
## # ... with 29 variables: EmpNumber <fct>, dupe_count <int>, Age <int>,
## #   Gender <fct>, EducationBackground <fct>, MaritalStatus <fct>,
## #   EmpDepartment <fct>, EmpJobRole <fct>, BusinessTravelFrequency <fct>,
## #   DistanceFromHome <int>, EmpEducationLevel <int>,
## #   EmpEnvironmentSatisfaction <int>, EmpHourlyRate <int>,
## #   EmpJobInvolvement <int>, EmpJobLevel <int>, EmpJobSatisfaction <int>,
## #   NumCompaniesWorked <int>, Overtime <fct>,
## #   EmplastSalaryHikePercent <int>, EmpRelationshipSatisfaction <int>,
## #   TotalWorkExperienceInYears <int>, TrainingTimesLastYear <int>,
## #   EmpWorkLifeBalance <int>, ExperienceYearsAtThisCompany <int>,
## #   ExperienceYearsInCurrentRole <int>, YearsSinceLastPromotion <int>,
## #   YearsWithCurrManager <int>, Attrition <fct>, PerformanceRating <int>
## [1] 1200    28

```

Range

If we would like to drill down each column then we can do this by following way knowing min and max value of each column which give you sense of how spread your data is. Sometimes you can catch the outliers easily instead of plotting boxplot everytime.

```

##                               [,1]          [,2]
## EmpNumber                 "E1001000"    "E100998"
## Age                       "18"           "60"

```

```

## Gender           "Female"      "Male"
## EducationBackground "Human Resources" "Technical Degree"
## MaritalStatus    "Divorced"     "Single"
## EmpDepartment    "Data Science"  "Sales"
## EmpJobRole       "Business Analyst" "Technical Lead"
## BusinessTravelFrequency "Non-Travel"   "Travel_Rarely"
## DistanceFromHome "1"           "29"
## EmpEducationLevel "1"           "5"
## EmpEnvironmentSatisfaction "1"           "4"
## EmpHourlyRate     "30"          "100"
## EmpJobInvolvement "1"           "4"
## EmpJobLevel       "1"           "5"
## EmpJobSatisfaction "1"           "4"
## NumCompaniesWorked "0"           "9"
## OverTime          "No"          "Yes"
## EmpLastSalaryHikePercent "11"          "25"
## EmpRelationshipSatisfaction "1"           "4"
## TotalWorkExperienceInYears "0"           "40"
## TrainingTimesLastYear   "0"           "6"
## EmpWorkLifeBalance    "1"           "4"
## ExperienceYearsAtThisCompany "0"           "40"
## ExperienceYearsInCurrentRole "0"           "18"
## YearsSinceLastPromotion "0"           "15"
## YearsWithCurrManager   "0"           "17"
## Attrition          "No"          "Yes"
## PerformanceRating   "2"           "4"

```

Recode

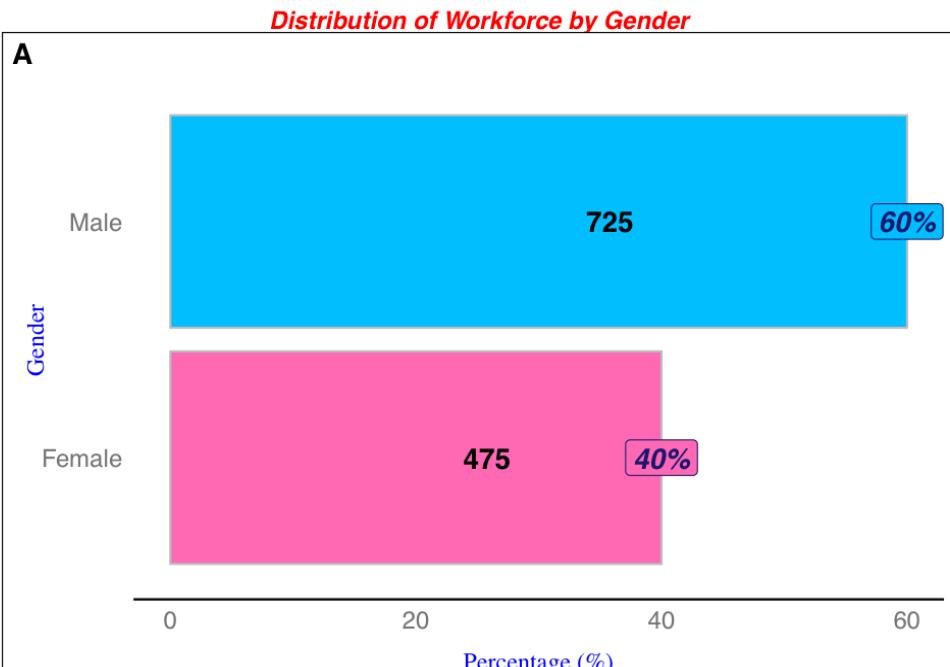
Lets Recode some of our attributes for quick analysis as well as to fit better in our graphical representation.

- i. EmpDepartment
- ii. Education Level
- iii. Age_bin
- iv. Generation
- v. Hike Percentage
- vi. Years With manager
- vii. Emp HourlyRate

Gender

Lets start to dig in with the Gender distribution.

- i. Employee Distribution Count by Gender
- ii. How is Male and Female distributed Across the Department ?



Gender distribution across the Department



A

- 60% Male and 40% Female.

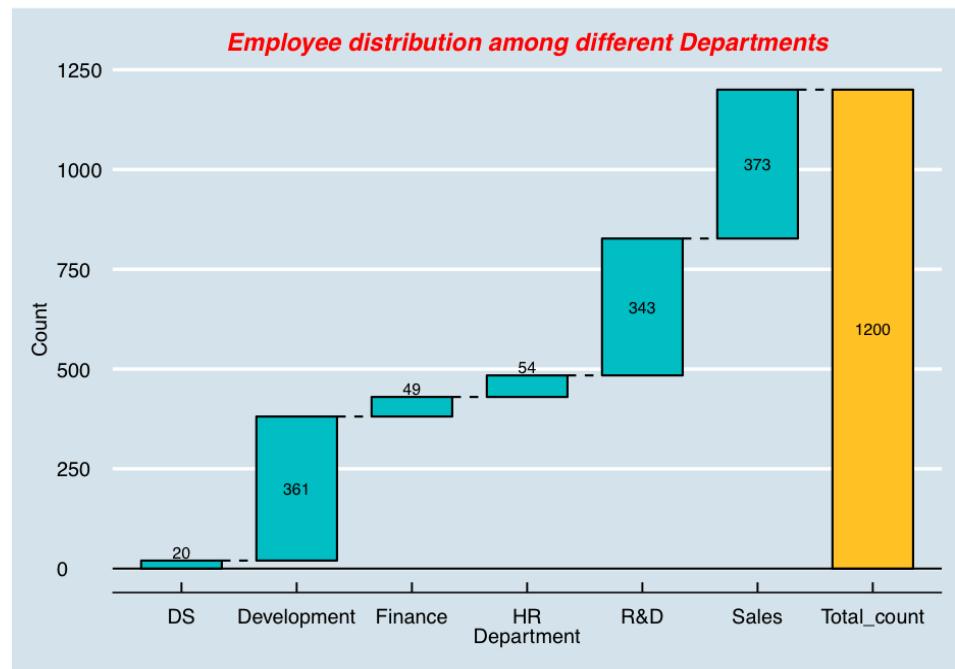
B

Above chart shows how Female and Male population are distributed among Department. Male population are equally distributed among the top three branches of the organisation whereas we can see slight deviation in Female department. Females seems to love the Sales department out of all brances.

EmpDepartment

- i. Waterfall Chart
- ii. Pie chart
- iii. Donut Chart

Employee distributed among the Organization



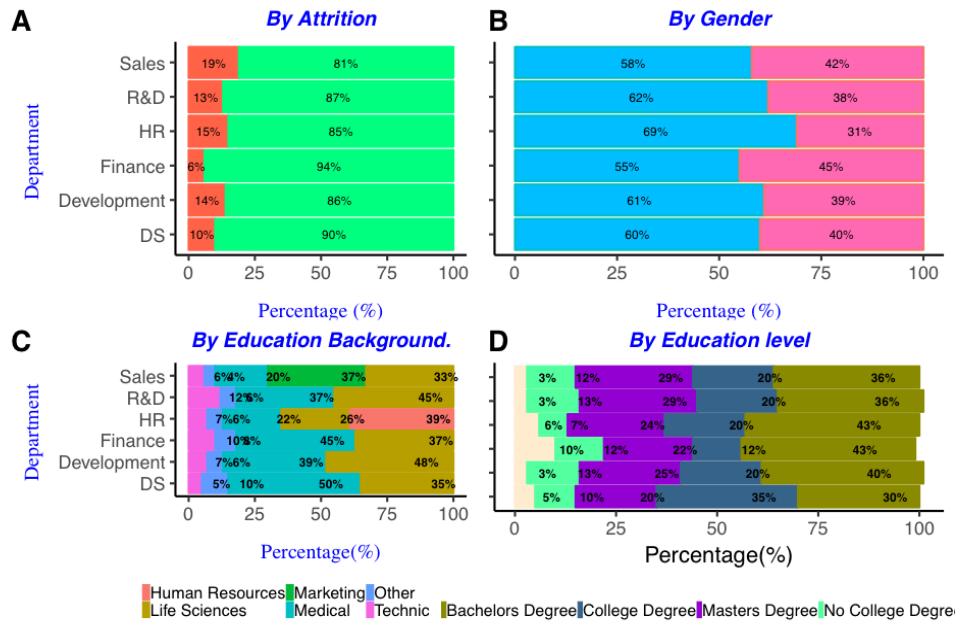
Three biggest Department are: Sales, Development & R&D. Distribution are

1. Sales : 31.1%
2. Development : 30.6%
3. R&D : 28.6%
4. HR : 4.5%
5. Finance : 4.08%
6. DS : 1.67 %

EmpDepartment & Gender

- A. Workforce breakdown among different Department
- B. Workforce breakdown among different Department by Gender
- C. EmpDepartment and Educational_Background
- D. EmpDepartment and Educational_Levels

Workforce breakdown across Department



A

- More Employee working in Sales department tends to leave compare to rest.
- HR being the second.

B

- It seems like most of the department is splitted between 60% Male population and 40% Female population except Human Resource department where there are 70% Male and 30% Female population.

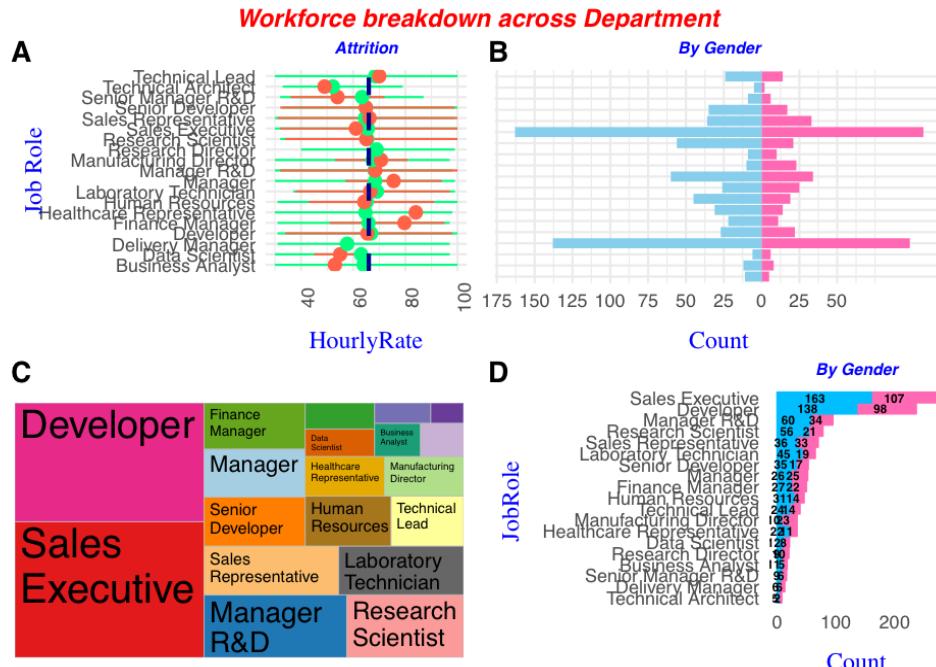
C

- Major Dominant in all Departments is Life Science Field except in Human Resource department where background field knowledge is Human Resource and Sales department where the person need to have Marketing background knowledge.

D

- In all field major Education level is Bachelor Degree.
- Except Data Science where major education level is College Degree(35%).
- Finance: Compare to all other departments Finance have 10% Employee with PhD degree.
- R&D: R&D have more Master Degree Employee compare to other departments
- HR and Finance: HR and Finance have 43% population with bachelor degree.
- HR: HR has less employee with no college degree compare to rest of department.

Employee Job Role



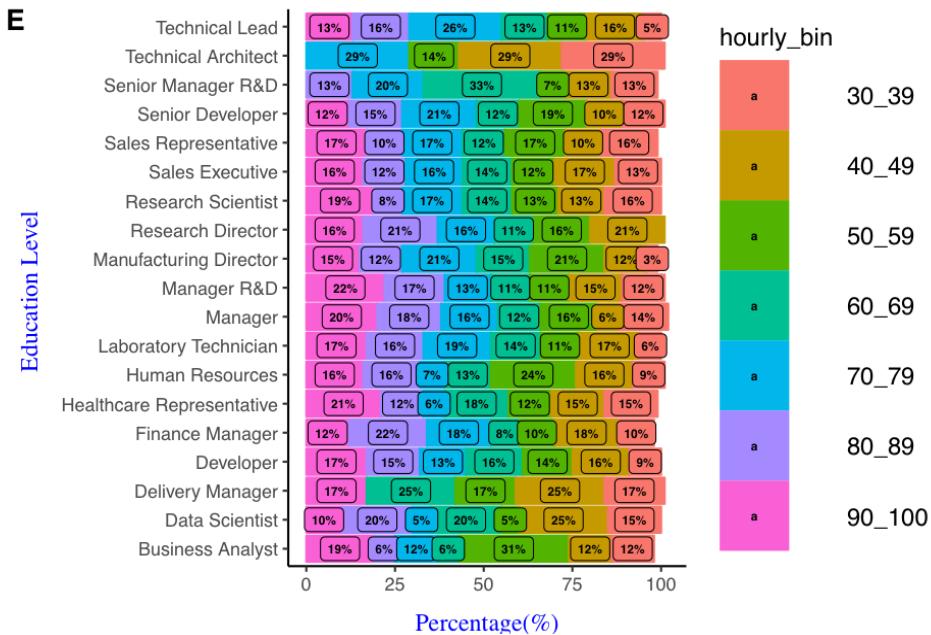
- A**

 - We can see how does the Hourly Rate and Job Role impacts the attrition Rate.

B,C,D

 - Sales Executive and Developer are two pillars of whole organization. There are more Male population compare to Female Population in both field as it is clear from all three charts.
 - Attrition of Job Role Employee by Hourly Rate. We will dig into analysis part later down below.

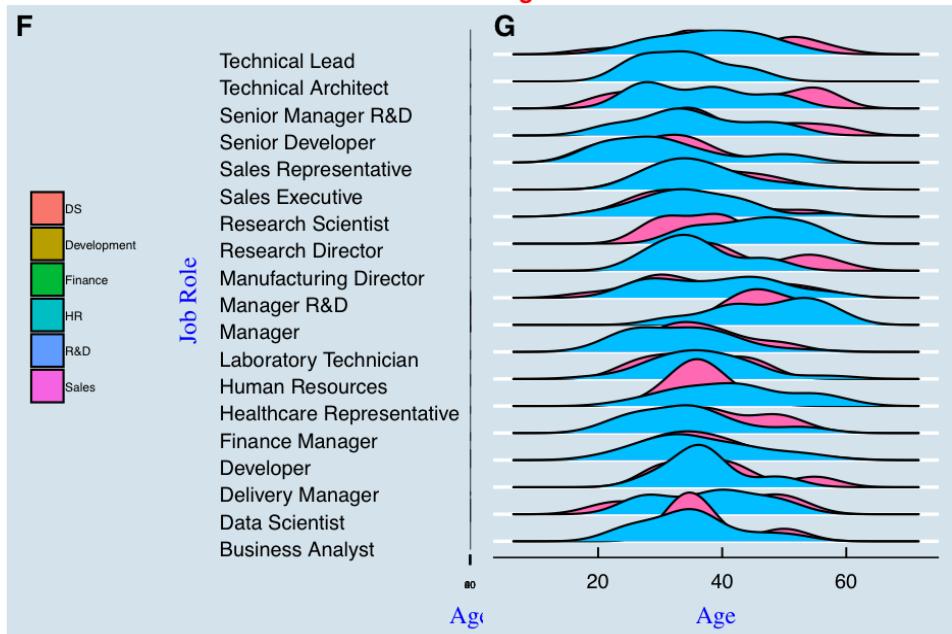
Proportion of Hourly Rate for the specific Job Role



E

- Except Technical Architect and Senior Manager in R&D all other JobRole Employee could make 90-100\$ an hour(colored PinK label).
- 22% Manager R&D make almost \$ 90-100 an hour.
- Technical Architecture has less layers of Hourly distribution (Only 4)
- Most of the Technical Architect fall between
 - \$30-\$39 : 29%
 - \$40-\$49 : 29%
 - \$50-\$59 : 14%
 - \$70-\$79 : 29%
- Senior Manager R&D Payscale lies mostly between \$60-\$69 about 33%.
- Most of the business Analysr make about \$50-\$59 an hour about 31%
- Research Director Starting hourly Rate is 40–\$49.
- Only 10% Data Scientist make \$90-\$100 an hour.

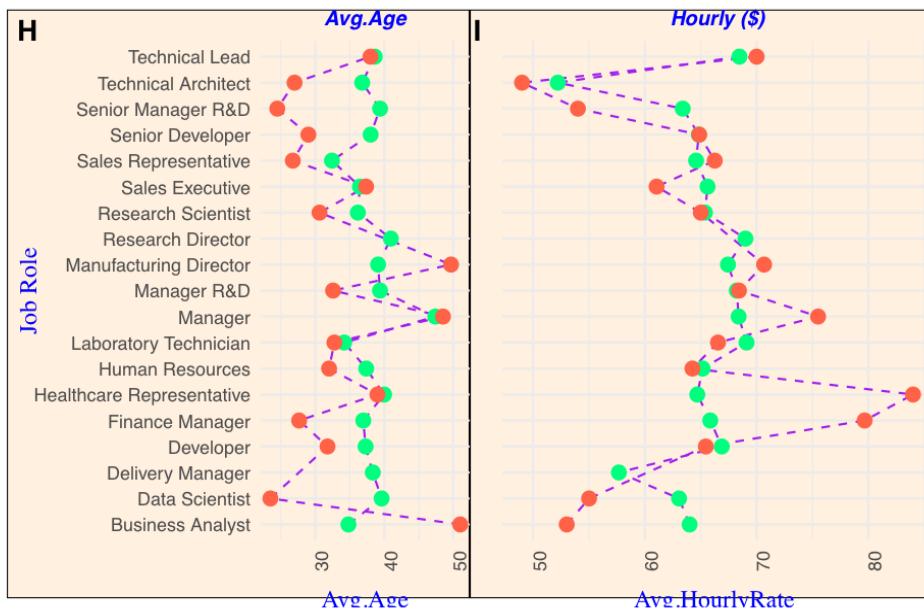
Job Role & Gender Age Distribution



F & G

- Business Analyst and Healthcare Representative Job Role are preferred to Younger Female Candidate.
- Research Director before age 40 goes mostly to Females than to Males.
- Technical Lead, Senior Manager R&D , Senior Developer, Manufacturing Director have more Senior Female than Senior Male.
- Delivery Managers before Age 40 goes to Male than to Female.

Job Role & Attrition based on Age and Hourly Rate

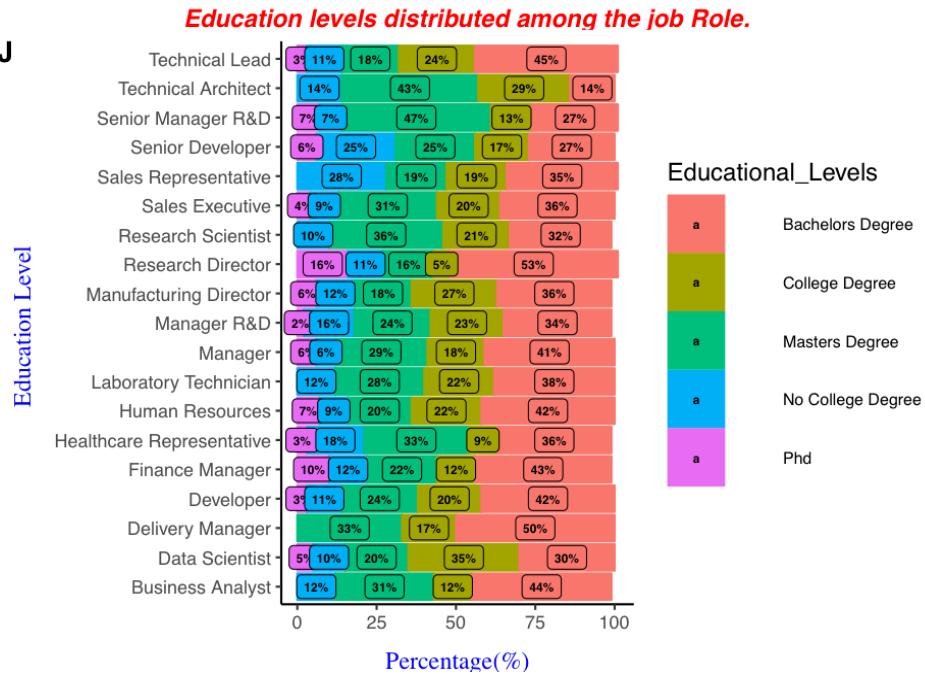


H

- We can see most of the employee around average age 30-35 are attributed more regardless of Job Role.
- Technical Lead , Sales Executive and Healthcare Representative have same average age of attrition.
- Manufacturing Director, Manager and Business Analyst also have same age of attrition.

I

- We could spot the average pay difference between Employee who tends to leave the company and who stays.
- The average difference for each some jobs are pretty close whereas for some jobs there is huge difference.



J

PhD: - We can also Note every Job Role requires PhD so the Education background is limited according to Job Role.

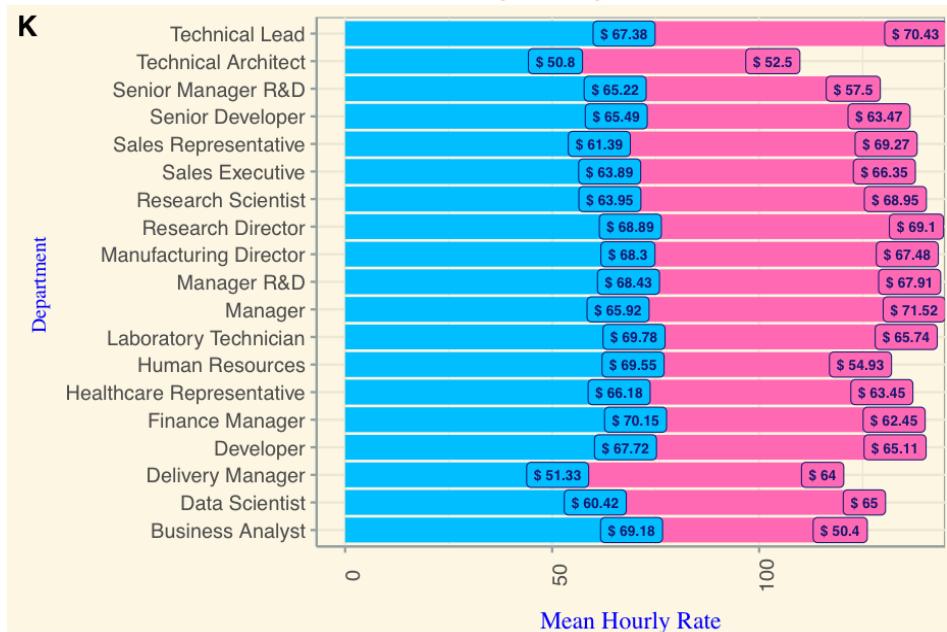
Master Degree: - Senior Manager R&D have more Master degree.

Bachelor Degree: - Almost 53% Research Director have Bachelor Degree.

College Degree: - Most of Data Scientist have only College Degree almost 35%

No College Degree: - 28% Sales Representative have no College Degree.

Job Role Mean Hourly Rate by Gender.

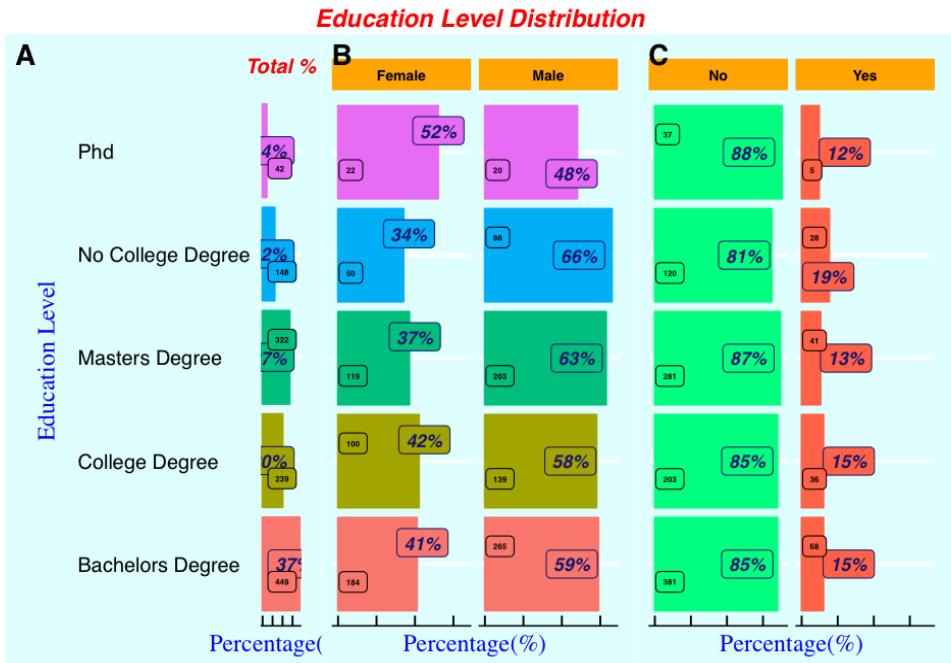


K

- In Some Job Role Female Tends to make more money than Male do.
- In Some the difference is Minimal and in some there is huge difference which could be because of Education levels and which could be because of Experience in Years.

Education level

- i. Education level by Gender
- ii. Education level & Attrition.
- iii. Education level & Gender.
- iv. Education level, Attrition and Department



A

As we can see that Most of our workforce have Bachelor degree i.e 37% followed by Master degree i.e 27%. As we aspect Phd are very less comprising 4%.

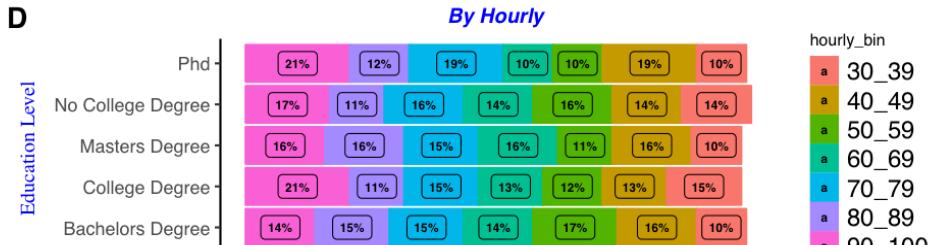
B

when we break it down by Gender it seems like there are equal number of Phd candidates.42 Employee have done Phd and when we even break it down by gender it seems like we have equal number of Male and Female Employee.

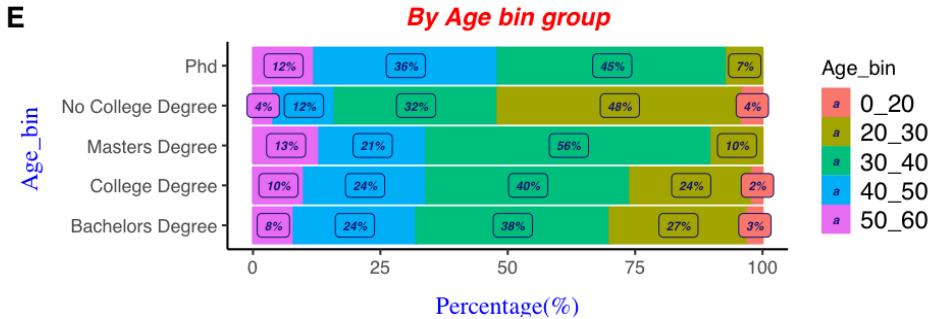
C

Employee with no college degree seems to leave the company compare to rest of the education field either could be for getting higher education degree or for lack of growth inside the organization.

Workforce breakdown based on their Education Level



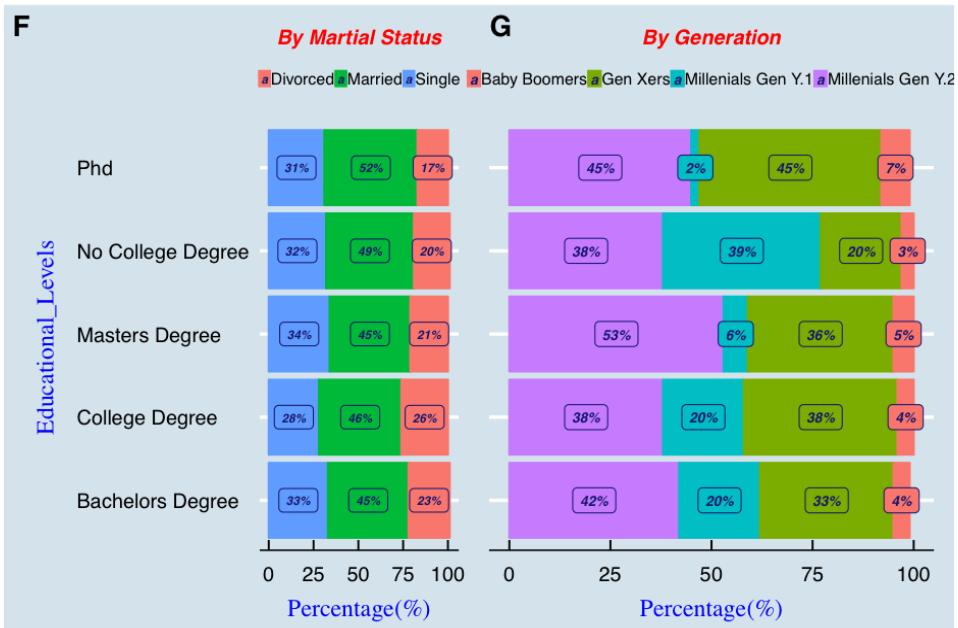
By Age bin group



D

- Education level doesn't play any role in How much Money you make. Education level is independent of How much money you could make.

Workforce breakdown based on their Education Level

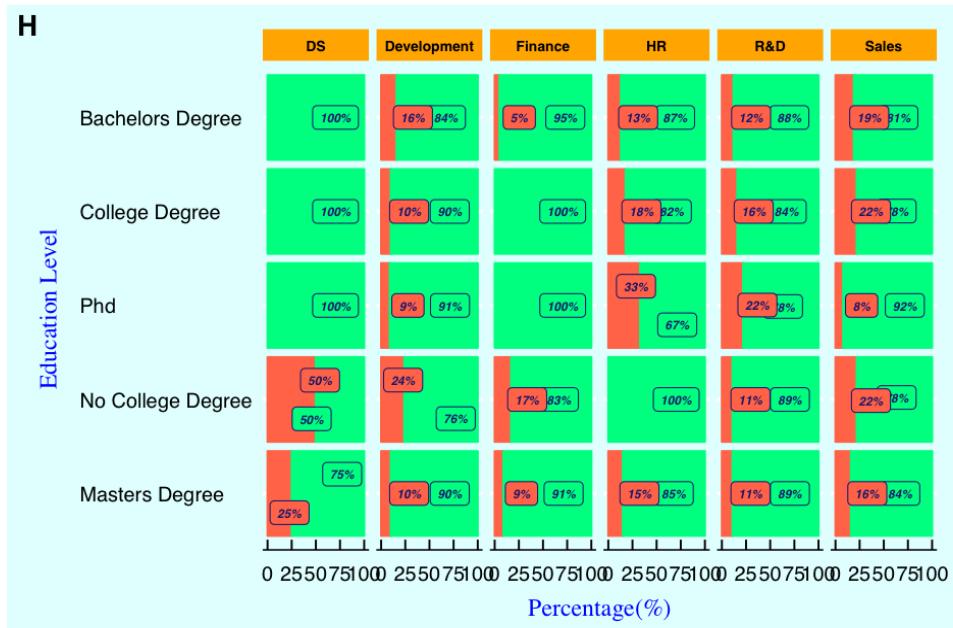


E, F & G

- Phd:
 - 7% Employee are aged between 20-30.
 - 45% Employee Aged between 30-40.
 - 52% Married Employee have PhD.
- Master Degree:
 - 10 % Employee are aged between 20-30
 - 56 % Employee are aged between 30-40.
 - 45 % Married Employee have Master Degree.
- Bachelor Degree:
 - 3 % have Employee either have already completed or have Bachelor degree under Age 20.
 - 38% Employee are aged between 30-40
 - 45 % Married Employee have Bachelor Degree.
- College Degree
 - 2% are Aged Under 20.
 - 40 % Employee are aged between 30-40
 - 26% Married Employee have College Degree.
- No College Degree:
 - 4% Employee are aged under 20

- 4% Employee are aged 50-60
- 48% Employee between age 20-30
- 49% Married Employee have No College Degree.
- 73% Millenials have No College Degree.
- 3% Baby Boomers have No college Degree.
- 1 % Silent Gen who have No college Degree.

Attrition across Department based on Education Levels



H

- Phd:
 - 33% Attrition in Human Resource Department.
 - 0% Attrition in Finance.
- Master Degree:
 - 25 % Attrition in Data Science
 - 9 % Attrition in Finance.
- Bachelor Degree:
 - 19% Attrition in Sales Department.
 - 5% Attrition in Finance Department.
- College Degree:
 - 22% Attrition in Sales Department.

- 0% Attrition in Data Science.
- No College Degree:
- 50 % Attrition in Data Science Department.
- 0 % Attrition in Human Resource.

Age

Lets look at the Age distribution in our datasets.

- i. Density Age Distribution
 - ii. Barplot Age Distribution
 - iii. Mean Age Distribution
 - iv. Percentage of Male & Female Distribution
- Highest Age of Employment in the organisation is Age: 60.

```
##      EmpNumber Age Gender EducationBackground MaritalStatus EmpDepartment
## 5     E1001010 60   Male          Marketing       Single        Sales
## 549   E1001975 60   Male          Medical       Divorced       R&D
##      EmpJobRole BusinessTravelFrequency DistanceFromHome
## 5           Sales Executive            Travel_Rarely      16
## 549  Healthcare Representative      Travel_Rarely       1
##      EmpEducationLevel EmpEnvironmentSatisfaction EmpHourlyRate
## 5             4                      1              84
## 549            4                      3              92
##      EmpJobInvolvement EmpJobLevel EmpJobSatisfaction NumCompaniesWorked
## 5             3                      2                  1              8
## 549            1                      3                  4              3
##      OverTime EmpLastSalaryHikePercent EmpRelationshipSatisfaction
## 5           No                     14                 4
## 549          No                     20                 3
##      TotalWorkExperienceInYears TrainingTimesLastYear EmpWorkLifeBalance
## 5             10                     1                 3
## 549            19                     2                 4
##      ExperienceYearsAtThisCompany ExperienceYearsInCurrentRole
## 5                   2                     2
## 549                1                     0
##      YearsSinceLastPromotion YearsWithCurrManager Attrition
## 5                   2                     2      No
## 549                0                     0      No
##      PerformanceRating Educational_Levels Age_bin Generation Hike_Pct
## 5             3      Masters Degree 50_60 Baby Boomers 14-16
## 549            4      Masters Degree 50_60 Baby Boomers 21-23
##      CatYearsManager hourly_bin
## 5           1-2 years Old    80_89
## 549 Recently Hired Newbie    90_100
##      EmpNumber Age Gender EducationBackground MaritalStatus EmpDepartment
## 160   E1001290 18   Male          Life Sciences     Single       R&D
## 243   E1001434 18   Male          Medical       Single       R&D
##      EmpJobRole BusinessTravelFrequency DistanceFromHome
## 160  Research Scientist            Non-Travel      5
## 243 Laboratory Technician        Non-Travel      8
```

```

##      EmpEducationLevel EmpEnvironmentSatisfaction EmpHourlyRate
## 160              2                      2                  73
## 243              1                      3                  80
##      EmpJobInvolvement EmpJobLevel EmpJobSatisfaction NumCompaniesWorked
## 160              3                      1                  4                  1
## 243              3                      1                  3                  1
##      OverTime EmpLastSalaryHikePercent EmpRelationshipSatisfaction
## 160        No                   15                  4
## 243        No                   12                  4
##      TotalWorkExperienceInYears TrainingTimesLastYear EmpWorkLifeBalance
## 160                 0                   2                  3
## 243                 0                   0                  3
##      ExperienceYearsAtThisCompany ExperienceYearsInCurrentRole
## 160                 0                   0
## 243                 0                   0
##      YearsSinceLastPromotion YearsWithCurrManager Attrition
## 160                 0                   0          No
## 243                 0                   0          Yes
##      PerformanceRating Educational_Levels Age_bin           Generation
## 160                 3    College Degree   0_20 Millenials Gen Y.1
## 243                 3  No College Degree   0_20 Millenials Gen Y.1
##      Hike_Pct      CatYearsManager hourly_bin
## 160  14-16 Recently Hired Newbie    70_79
## 243  11-13 Recently Hired Newbie    80_89

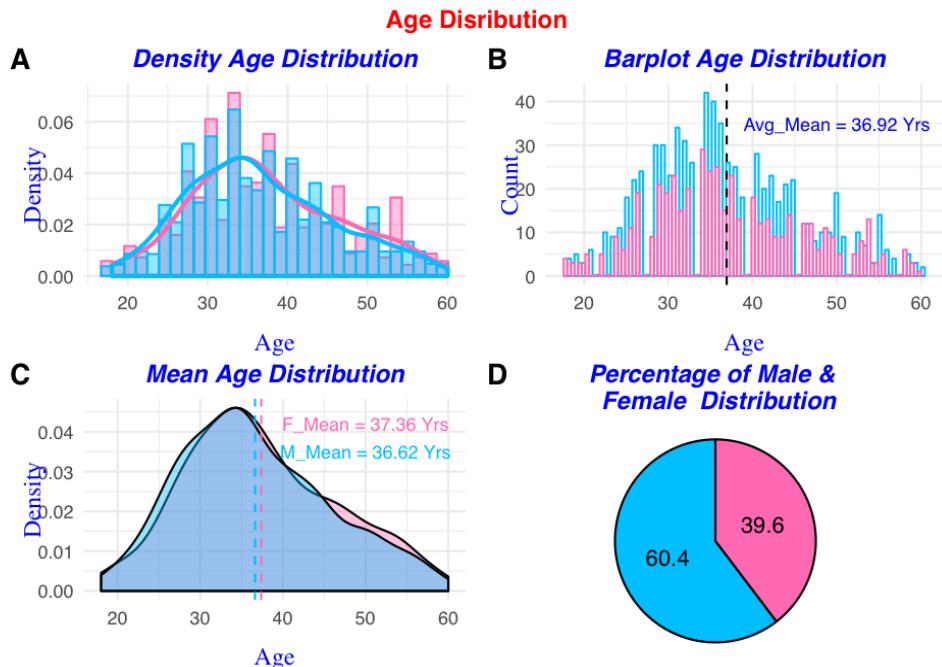
```

Oldest Employee

- 3 Employees
- We have Sales Executive who is at Age 60 (Retirement) who is still Single and makes \$84 an hour.

Youngest Employee

- 8 Newbies
- We have 18 year old Female in R&D Department who is Manager and makes \$97 an hour.

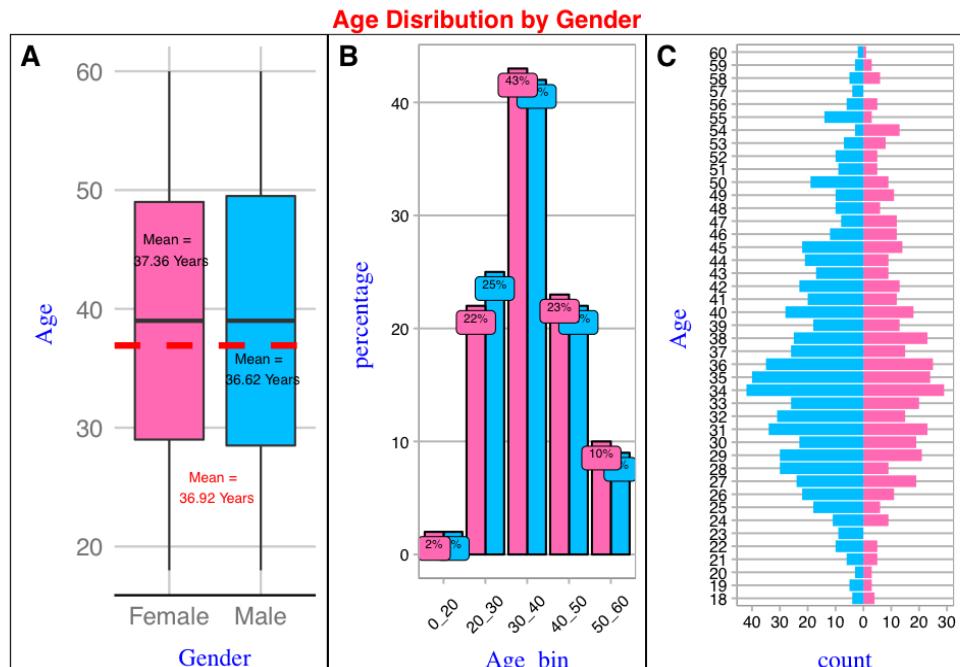


- Age is Normal Distributed.
- In our Datasets there are 39.6% Female poulation and 60.4% Male Population.
- Overall Average Age of our population datasets is 36.92
- Average Age of Male Population is 36.62
- Average Age of Female Population is 37.36

Age_bin

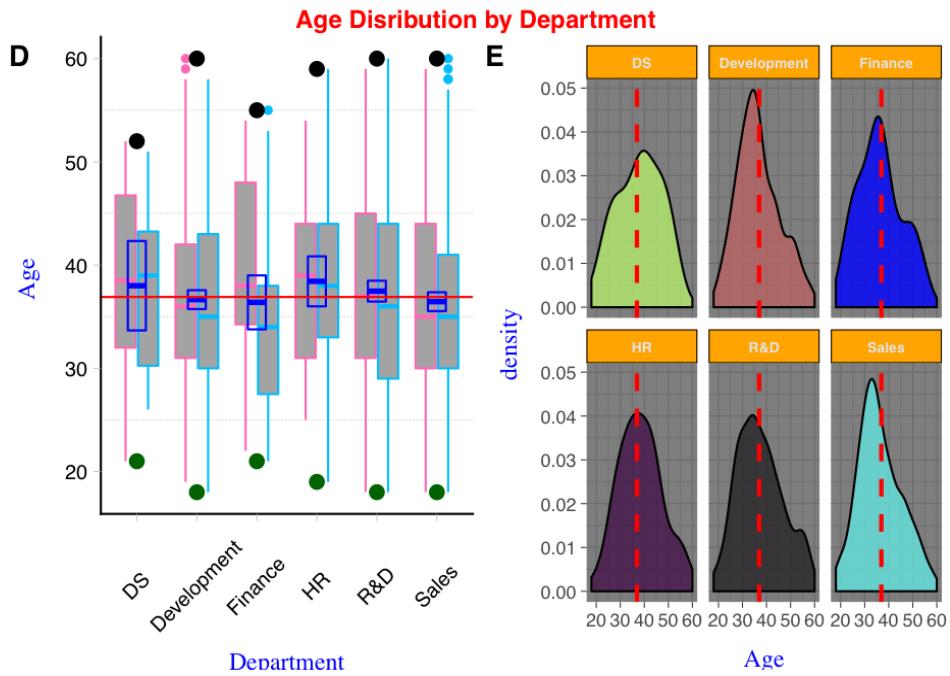
Age column has been spread from Minimum Age 18 to max Age 60. Lets bin all of them into decades. By doing so we will be able to reduce the noise and outliers and be more focused on insights without being biased.

- i. Age distribution by Gender
- ii. Age distribution by Department



A, B & C

Our datasets when we broke down by the agebin group, it seems like we have balance datasets between male and Female across all age groups.

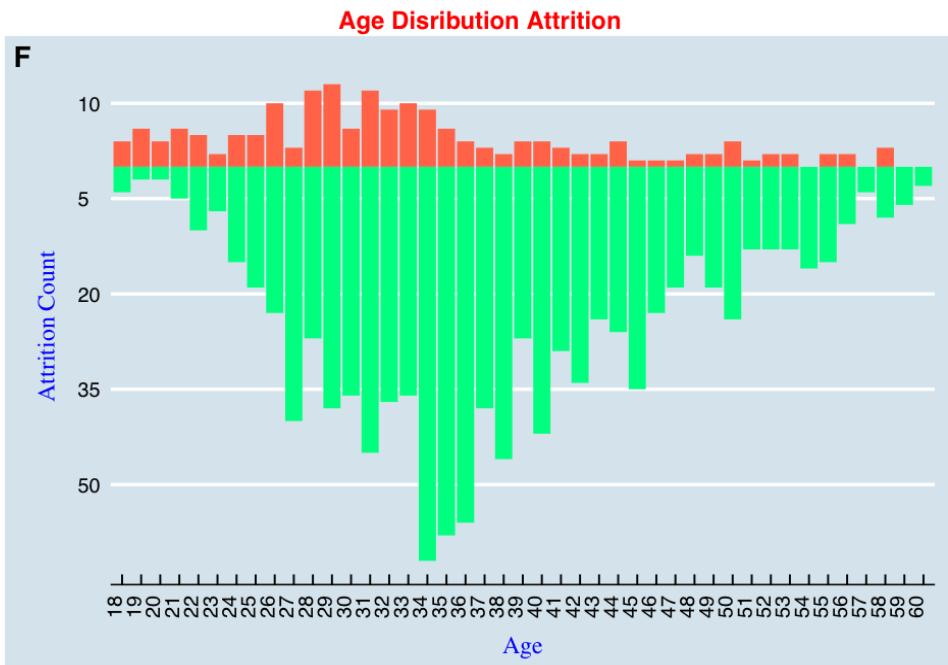


D

- In Finance we can see there is difference between Male and Female Age Group.
- We can see Male Outliers in Sales and Females Outliers in Development Department.
- In Data Science Department also we can see some difference between male and Female Age Group.
- There is less difference in R&D and HR department between Male and Female Age.

E

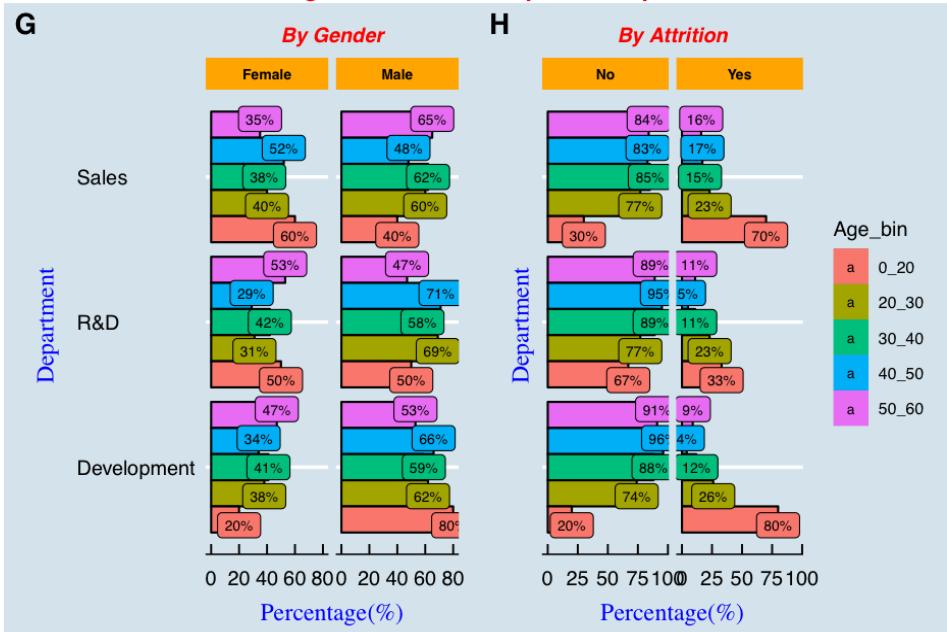
- The peaks of a Density Plot help display where values are concentrated over the interval.
- we can illustrate how the distribution of a Age change over time.
- In sales we can see lot of Young People followed by Development and then by R & D before their mean age.



F

- We can see Age and Attrition of the Employee. Seems like Normal Distribution in both who leaves the company and who stays throughout.

Age Distribution in top three dept.



G

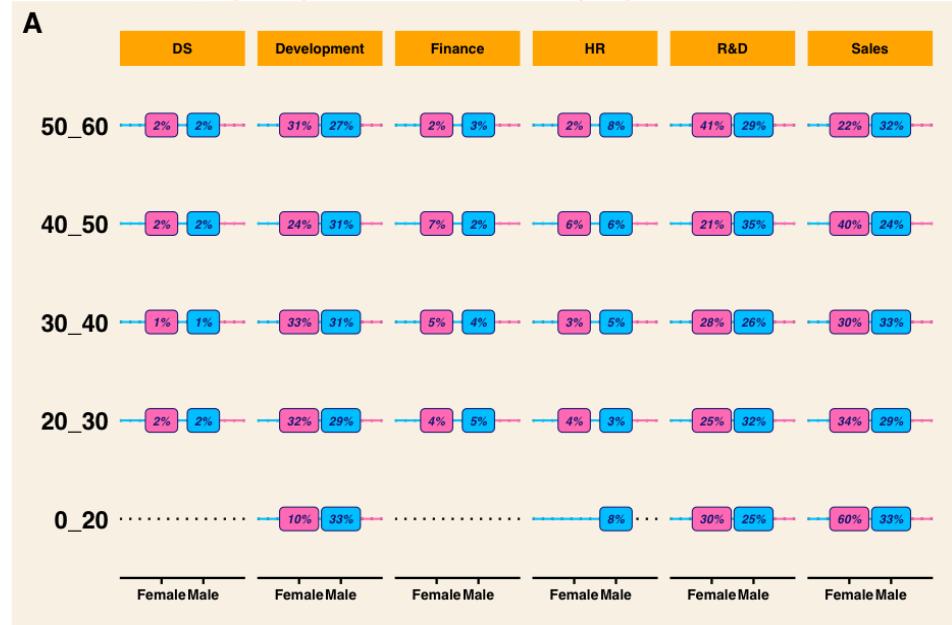
- 60 % Female in Sales department of Age under 20 compare to Male which is just 40%.
- 80% Male compare to 20% Female in Development Department of Age under 20.
- Rest every Age group is distributed between 60-40.

H

- 80% in Development and 70% in Sales Department Employee Attrition under 20.
- 4 % in Development and 5% in R&D of Age 40-50 Attrition.

Age | Department | Gender | Job Role

Employee Age distribution Percentage by Gender & Dept



A

0-20: - In Finance and Data Science there is no Employee which is under age 20. - In Human Resource there are 8% Male Employee

20-30: - 34% Female in Sales and 32% Male in R&D.

30-40: - 33 % Female in Development & 33% Male in Sales

40-50: - 40% Female in Sales & 31 % Male in Development

50-60: - 41% Female in R&D & 32% Male in Sales.

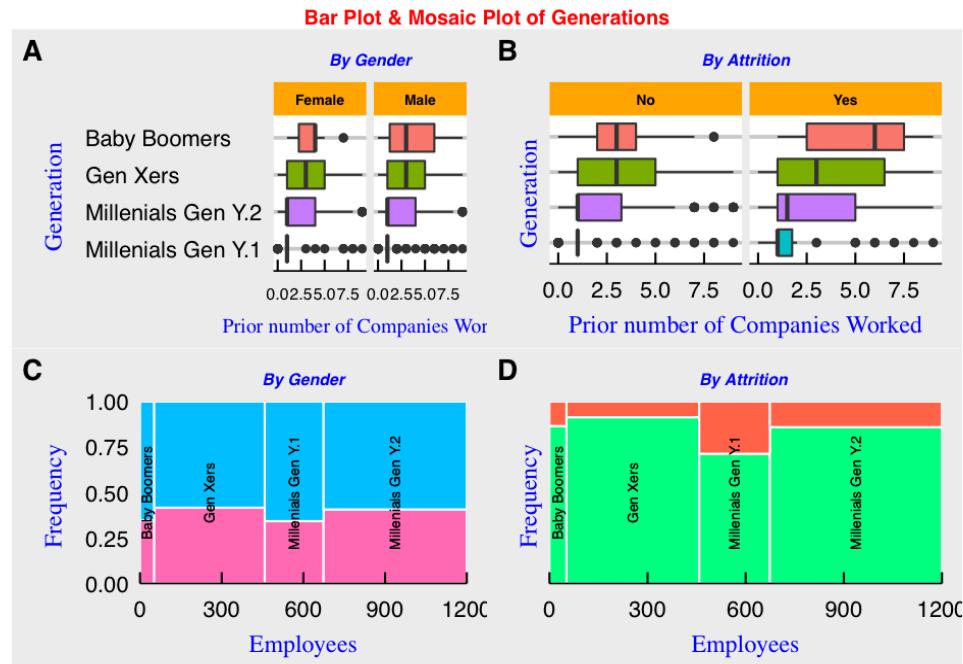
Generation

We can even classify all the Employee based on their age in four different category. As of dated today 2019.

- Baby Boomers: Baby boomers were born between 1944 and 1964. They're current between 55-75 years old (76 million in U.S.)
- Gen X: Gen X was born between 1965 - 1979 and are currently between 40-54 years old (82 million people in U.S.)
- Gen Y: Gen Y, or Millennials, were born between 1980 and 1994. They are currently between 25-39 years old.
- Gen Y.1 = 25-29 years old (31 million people in U.S.)
- Gen Y.2 = 29-39 (42 million people in U.S.)

- Gen Z: Gen Z is the newest generation to be named and were born between 1995 and 2015. They are currently between 4-24 years old (nearly 74 million in U.S.)

As we don't have many Gen Z in our datasets we will combine Gen Z with Gen Y.2. Otherwise we will have barplots with couple of dots.



A

- We Don't have much data for Millenials Gen Y.1.
- We can see some outliers in Millenials Gen Y.2 in both Male and Female
- We can also spot outliers in Baby Boomers in Female.
- Male Baby Boomers have more experienced than Female.

B

- There are more outliers in Millenials Gen Y.1 in both leaving the company as well as staying.
- As the prior number of companies worked experience increases we can see Employee tend to leave the company also increases than those who stay.

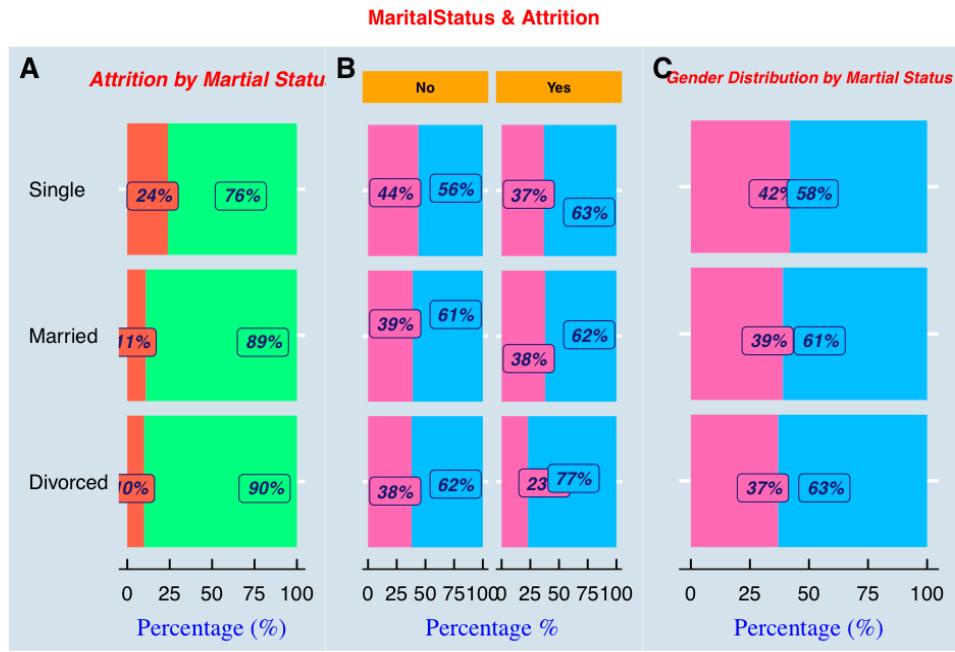
C

- We can see the proportion of Malea and Female Employee by Generations.
- More Male Millenials Gen Y.1 and Baby Boomers than Female.

D

- Millenials Gen Y.1 (Age 25-29) tends to leave the company more than other Generations.

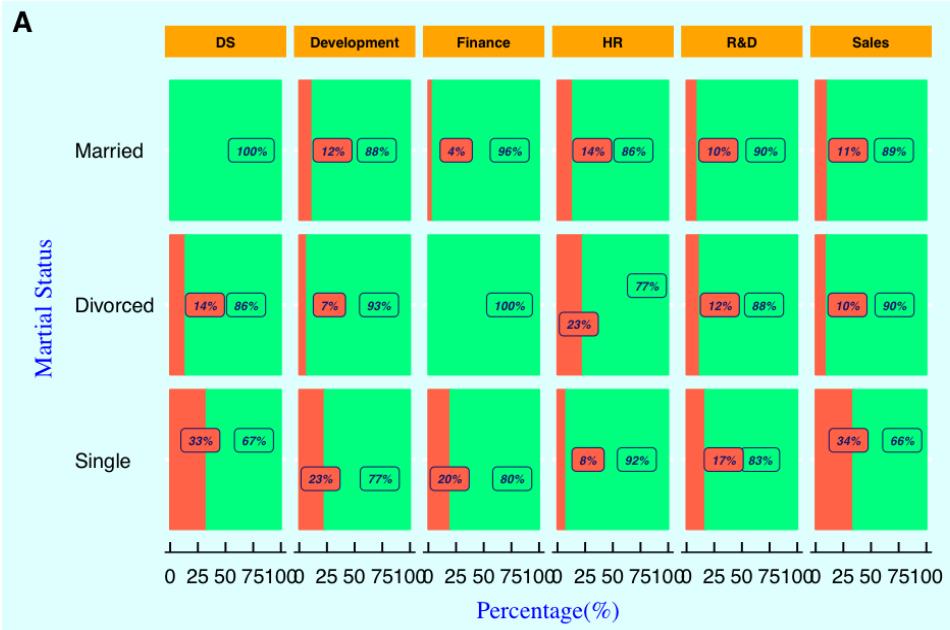
Martial Status



A, B, C

- Single tends to leave than Married and Divorced roughly about 24%.
- Almost the Ratio of Male and Female by martial Status is 60:40(M:F)
- Attrition Rate in Male population is 60:40(M:F) in Single and Married whereas in Divorced it is almost 80:20(M:F)
- Employee who stays beside of their Martial Status is almost 60:40(M:F)

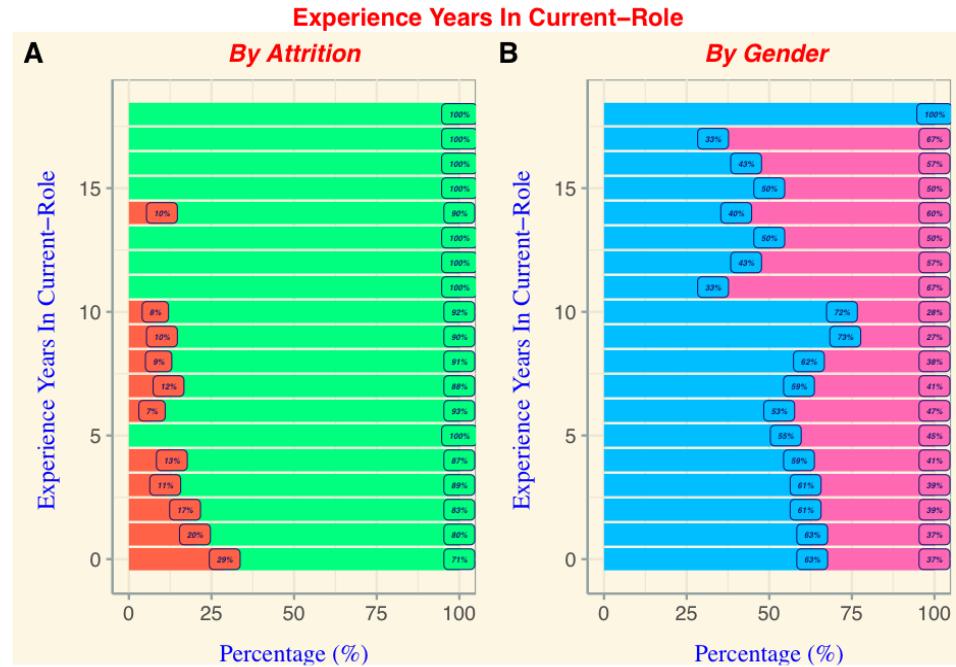
Attrition across Department based on Martial Status



D

- No Attrition in Finance Department if person is Divorced.
- No Attrition in Data Science Department if person is Married.
- High Attrition if the employee is Single in Sales and Data Science Department.

Experience Years In CurrentRole & Attrition



A

- As you gain more experience in current Role you are less likely to leave.
- If you have 10 years of experience it is less likely you will leave until and unless there is major change. As we can see there is one outlier in 14 years of experience in current role.
- Percentage of Attrition keeps on decreasing as once Employee have more experience in current role.

B

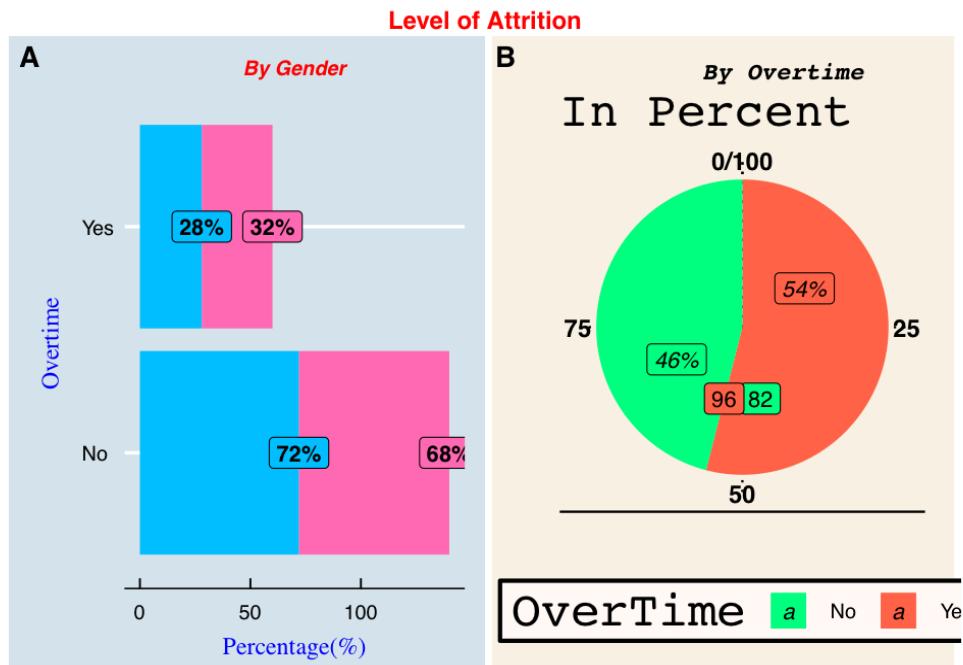
- There are less Male with more than 10 years of experience in current role than Female except one exception at 18 years of experience.
- Female tends to be more stable as we move up in experience except year 9 and 10 where ratio changes from 60:40(M:F) to 70:30(M:F) then as we move up ratio seems 30:70(M:F)

Overtime

```
##
##      Cell Contents
## |-----|
## |           N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
```

```

## |      N / Table Total |
## |-----|
## 
## 
## Total Observations in Table: 1200
## 
## 
##          | hr_df$OverTime
## hr_df$Attrition |      No |      Yes | Row Total |
## -----|-----|-----|-----|
##      No |    765 |    257 |    1022 |
##      | 2.640 | 6.334 |    |
##      | 0.749 | 0.251 | 0.852 |
##      | 0.903 | 0.728 |    |
##      | 0.637 | 0.214 |    |
## -----|-----|-----|-----|
##      Yes |     82 |     96 |    178 |
##      | 15.157 | 36.368 |    |
##      | 0.461 | 0.539 | 0.148 |
##      | 0.097 | 0.272 |    |
##      | 0.068 | 0.080 |    |
## -----|-----|-----|-----|
## Column Total |    847 |    353 |    1200 |
##      | 0.706 | 0.294 |    |
## -----|-----|-----|-----|
## 
## 
```

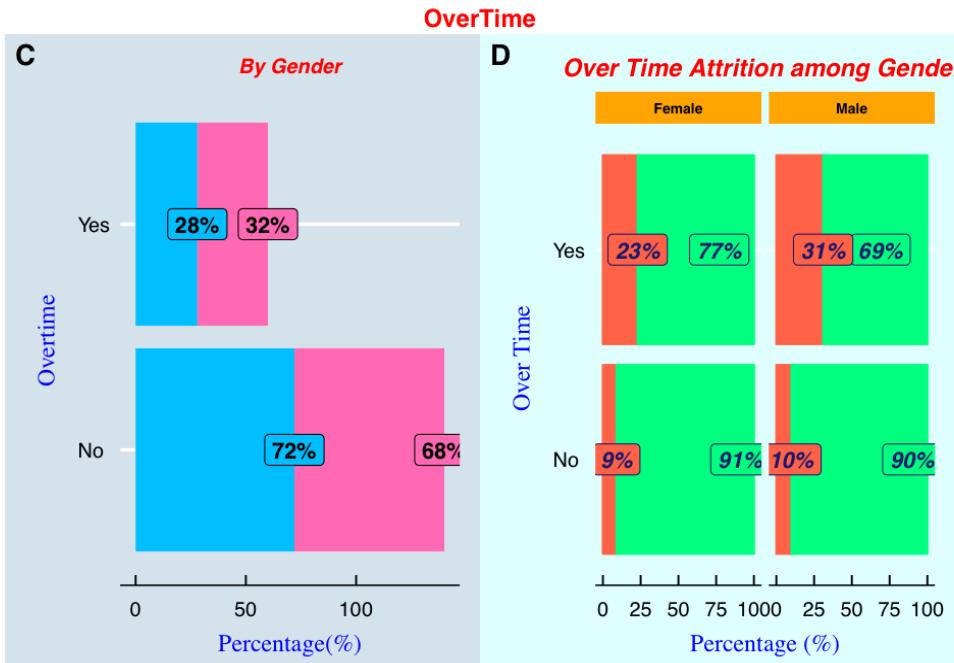


A

- Equal distribution on Overtime By Gender.
- 28% Male and 32% Female does Overtime. Female seems do to Overtime Compare to Male.

B

- 54 % People do leave the company if they are doing Overtime.In our datasets we have 96 people left compare to 82 who did stay beside doing overtime.



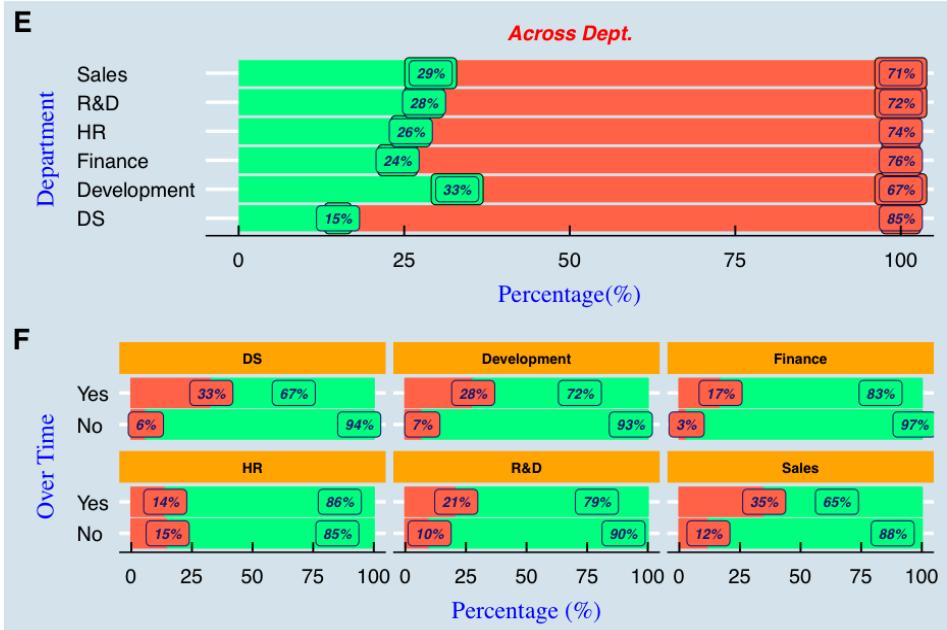
C

- 32% Female Employee seems to do Overtime compare to 28% Male Employee.

D

- When we groupby by OverTime, Gender and then by Attrition we see that :
- 31% Male Seems to leave the company compare to 23% Female who works OverTime.
- When we groupby by OverTime, Gender and then by Attrition we see that :
- 31% Male Seems to leave the company compare to 23% Female who works OverTime.

OverTime Attrition



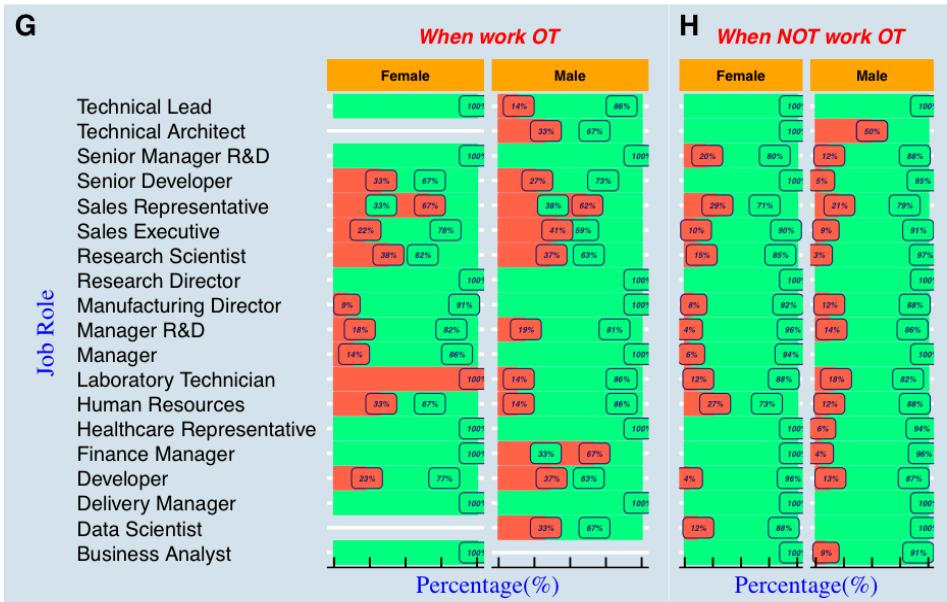
E

- In Data Science Employee seems to do more OverTime compare to rest of Other Department.

F

- Sales and Data Science Department have more Attrition when they work OverTime than rest of Department.

Job Role & Attrition



G

- Female Laboratory Technician tends to leave 100% when Work OverTime but if its Male Laboratory Technician then case seems to be completely opposite.
- Female Sales Representative tends to leave 67% of time.
- There is No Female Data Scientist & Female Technical Architect who have worked OverTime.
- There is No Male Business Analyst who have worked OverTime.
- 67% of Male Finance Manager seems to leave the company if they work OverTime.

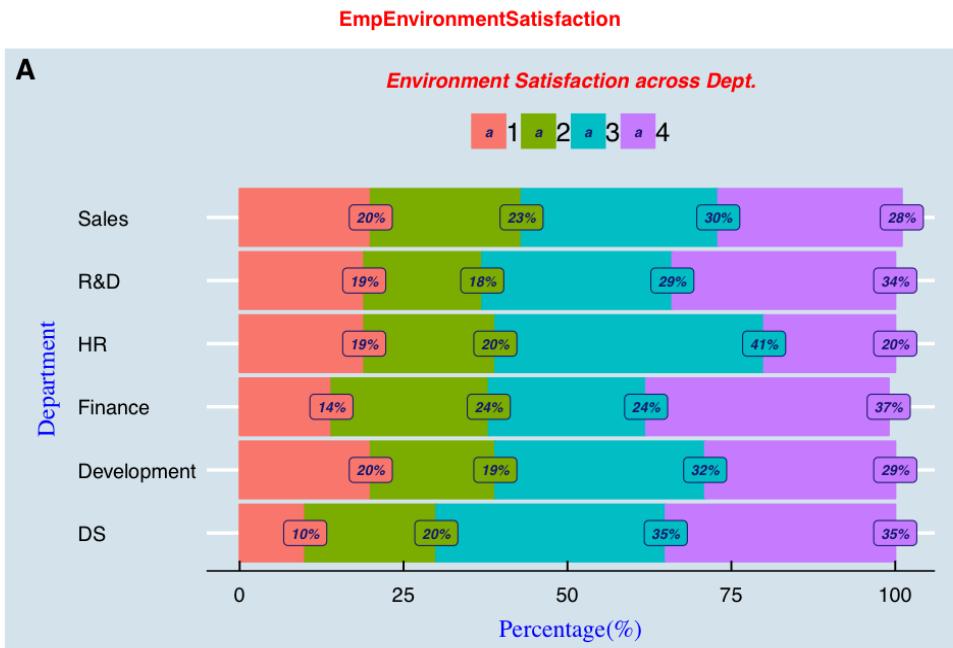
H

- Male Technical Architect only seems to leave the company even Not working Overtime.

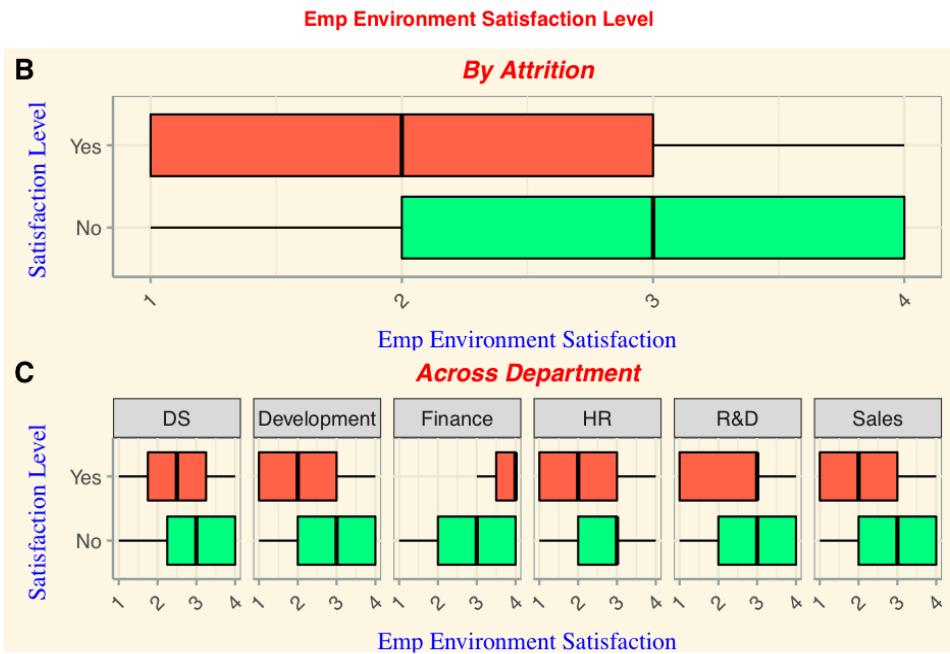
Rest of them are splitted more or less.

Employee Environment Satisfaction

- Employee Environment Satisfaction vs. Attrition.
- Employee Environment Satisfaction vs. Attrition vs. EmpDepartment.
- Employee Environment Satisfaction vs. Job Role vs. Attrition.
- Age vs. Job Role vs. Attrition.
- Hourly Rate vs. Job Role vs. Attrition.



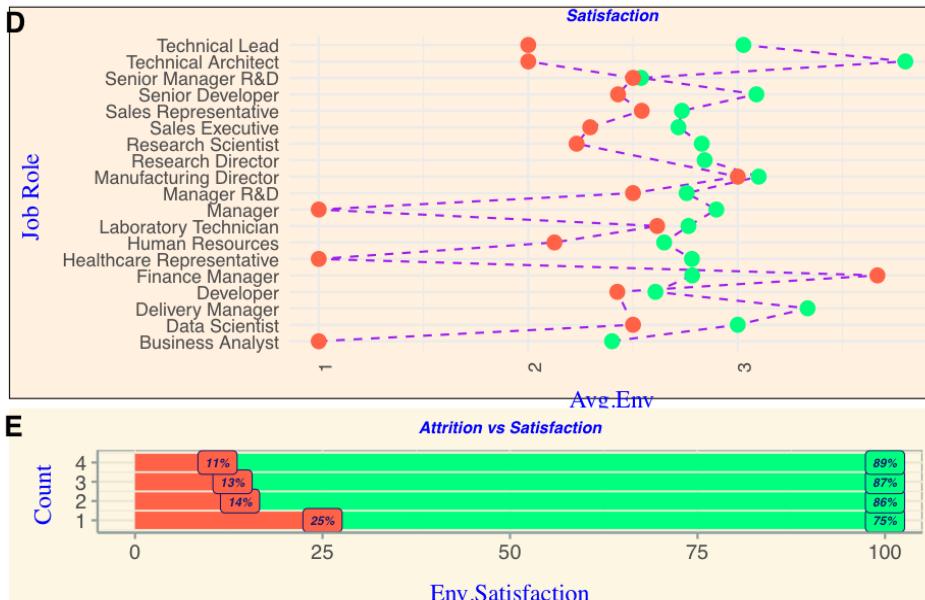
A - Employee working in top 3 Department have low satisfaction score roughly around 20%. - HR Employee have satisfaction score around 3 for 41% of Employees also only 20% of them have fully score of 4 as Employee Satisfaction Score.



B - As we can see if Employee are less satisfied with the Environment that you work with. You are more prone to leave the company than to stay.

C - In Finance even if the Employee have greater Environment Satisfaction they seems to leave. So In Finance department Satisfaction Level has no role to keep or leave the employee. - In Sales, Development, R&D, HR we can see there is effect of Emp Environment Satisfaction on Attrition.

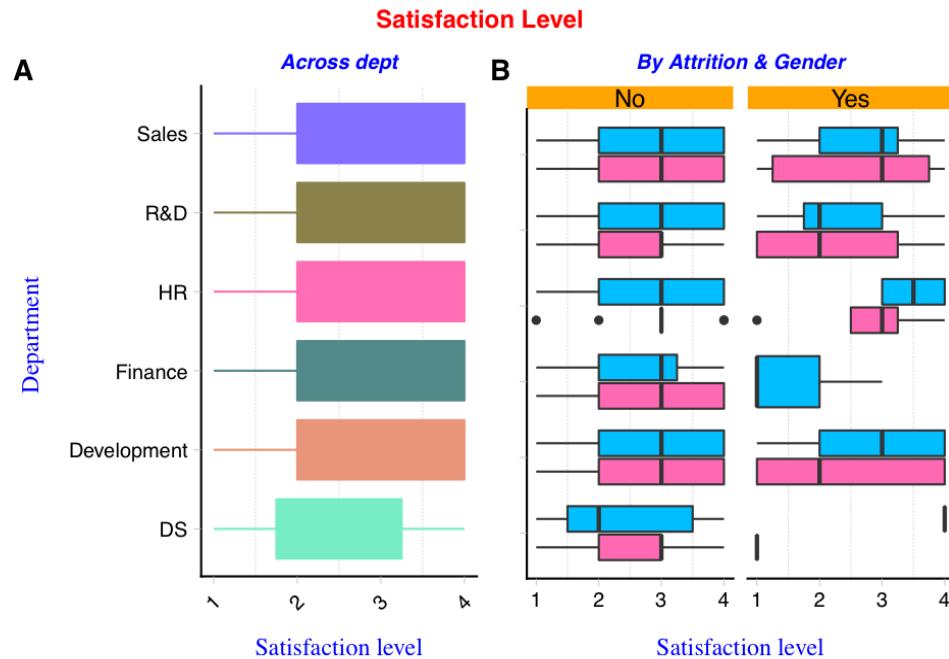
Emp Environment Satisfaction Level



D - Any Employee who has Satisfaction Score around 2.5 or less have all left the company. - There is one exception for the Finance Manager who has Satisfaction score for 4 and have left the company.

E - We have 25% of Employee whose Satisfaction score is 1 and have left the company.

Employee Relationship Satisfaction



A - Only Data Science field deviate towards less than 2 on average Employee Relationship Satisfaction compare to all other Department.

B

- When we break it down by Gender and Attrition although Data Science field have less score on Employee Relationship Satisfaction they haven't left the company. Males shows more less Employee Relationship Satisfaction score than Females.
- In Human Resource Male with high Employee Relationship Satisfaction Score seems to leave than Female.
- In Finance Male with low Employee Relationship Satisfaction definetly leaves the company.

Years with Manager

```
##      CatYearsManager Pro-Veteran 1-2 years Old 2-5 Years hired 5-10 Years Vet Recently Hired Newb
##  Gender
##  Female          2.25       10.92        7.08      12.42           6.9
##  Male            2.50       18.08       10.75      18.08          11.0
##
##                                     Attrition   No    Yes
##  Gender CatYearsManager
##  Female  Pro-Veteran          2.17  0.08
##          1-2 years Old          9.58  1.33
##          2-5 Years hired         6.58  0.50
```

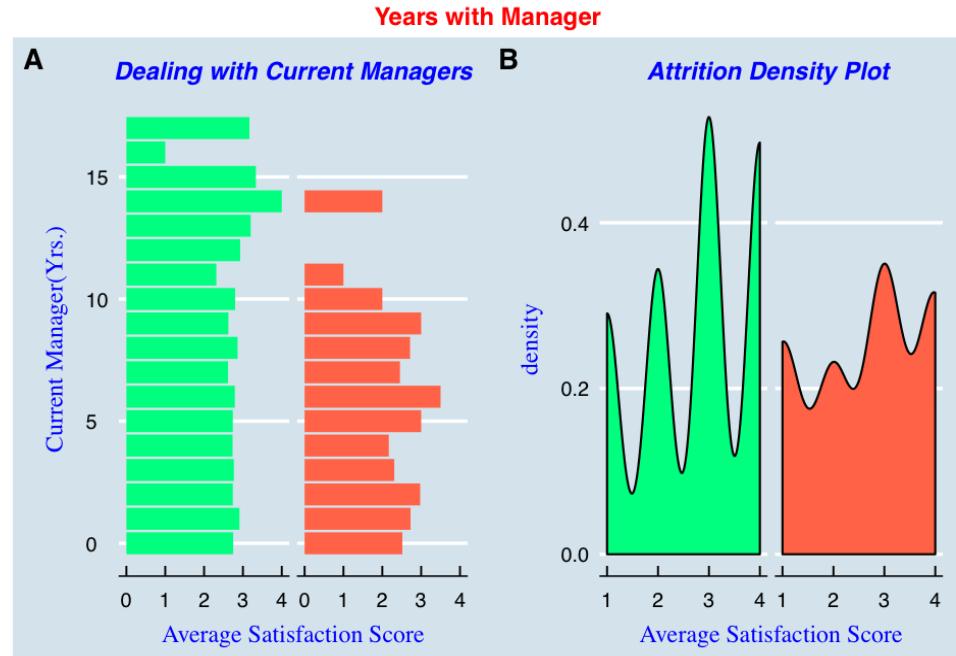
```

##      5-10 Years Vet          10.92  1.50
## Recently Hired Newbie     5.08   1.83
## Male Pro-Veteran          2.42   0.08
## 1-2 years Old             15.25  2.83
## 2-5 Years hired            9.42   1.33
## 5-10 Years Vet             16.25  1.83
## Recently Hired Newbie     7.50   3.50

##           EmpDepartment    DS Development Finance    HR    R&D Sales
## Gender Attrition
## Female No                  0.58    10.83   1.83  1.08  9.42 10.58
## Yes                         0.08     1.00   0.00  0.33  1.33  2.50
## Male No                     0.92    15.00   2.00  2.75 15.50 14.67
## Yes                        0.08     3.25   0.25  0.33  2.33  3.33

## Scale for 'fill' is already present. Adding another scale for 'fill',
## which will replace the existing scale.

```

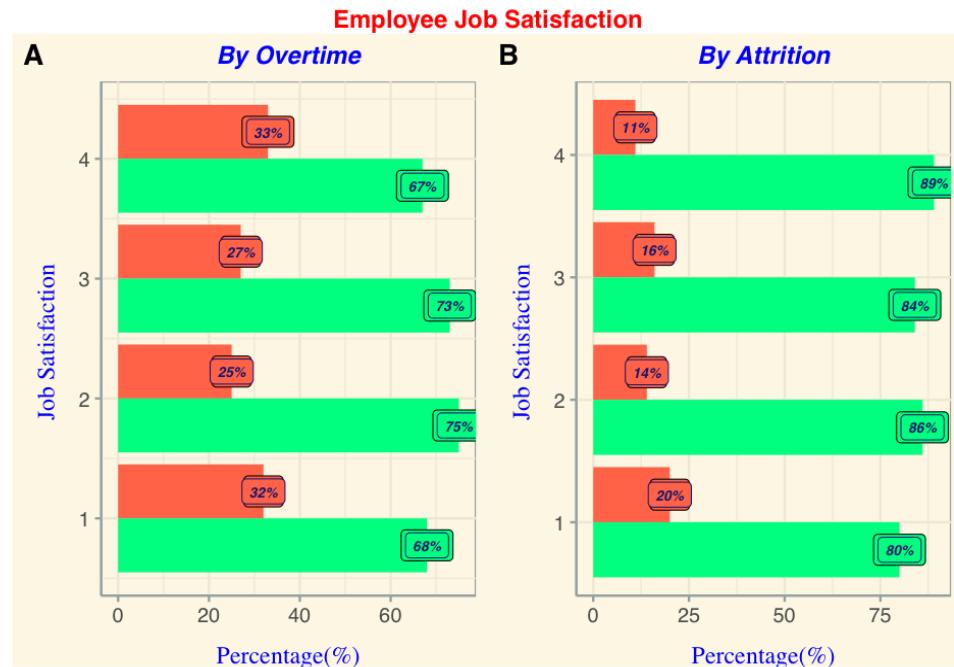


A

- If Employee have dealt mostly with current Managers for really long time they tend to stay. As year increases Attrition decreases.
- On an average Employee dealing with current Managers have Average Satisfaction score of 2.5

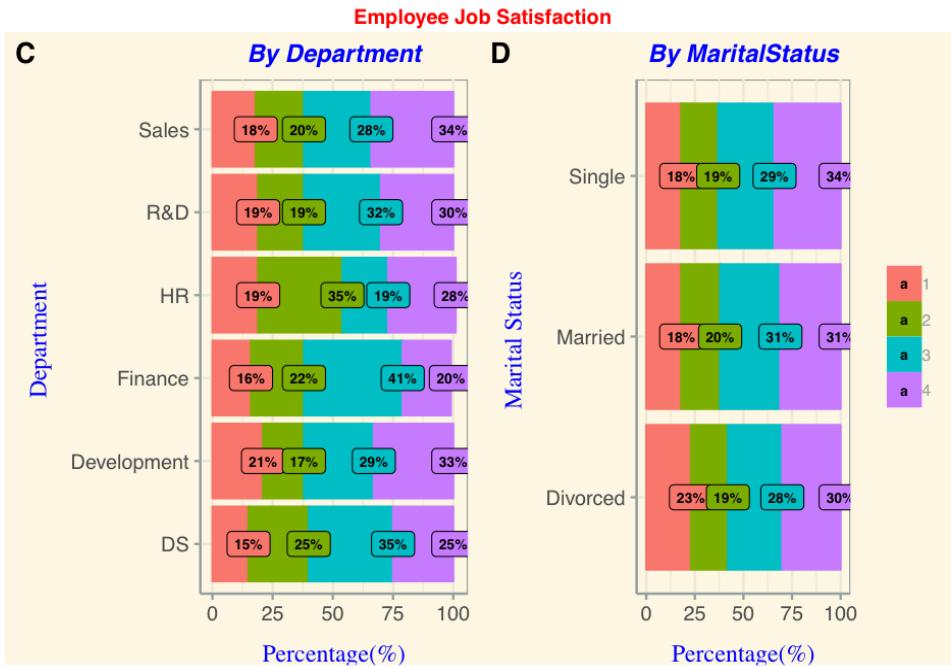
B - We can see the density plot of Employee Leaving Average Satisfaction score and Employee staying Average Satisfaction score. The peak value defines the difference in thier Average Satisfaction score.

Employee Job Satisfaction



A - Employee Job Satisfaction score with OverTime is splitted 70:30 (No:Yes)

B - Low Job Satisfaction Score tends to leave the company more than than high Employee Job Satisfaction.



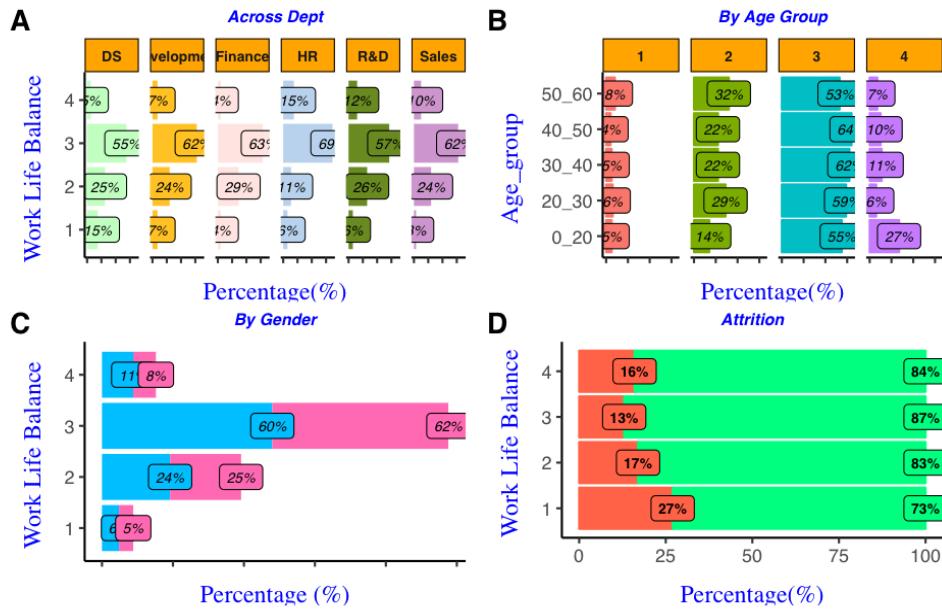
C - Same top 3 department have low Employee Job Satisfaction score.

D - Martial Status seems independent of Employee Job Satisfaction. Roughly equally distributed.

Worklife Balance

WorkLifeBalance. From Data Dictionary 1: 'Bad' 2: 'Good' 3: 'Better' 4: 'Best'

Work Life Balance



A

- 15% Employee Working in Data Science Department have worst Work Life Balance.
- 15% Employee Working in Human Resource Department have best Work Life Balance.
- Most of the Department have Work Life Balance score of 3.

B

- Most of the Employee have Better Work Life Balance at different Age group.
- 27% of Employee of Age 0-20 have Work Life Balance of 4.

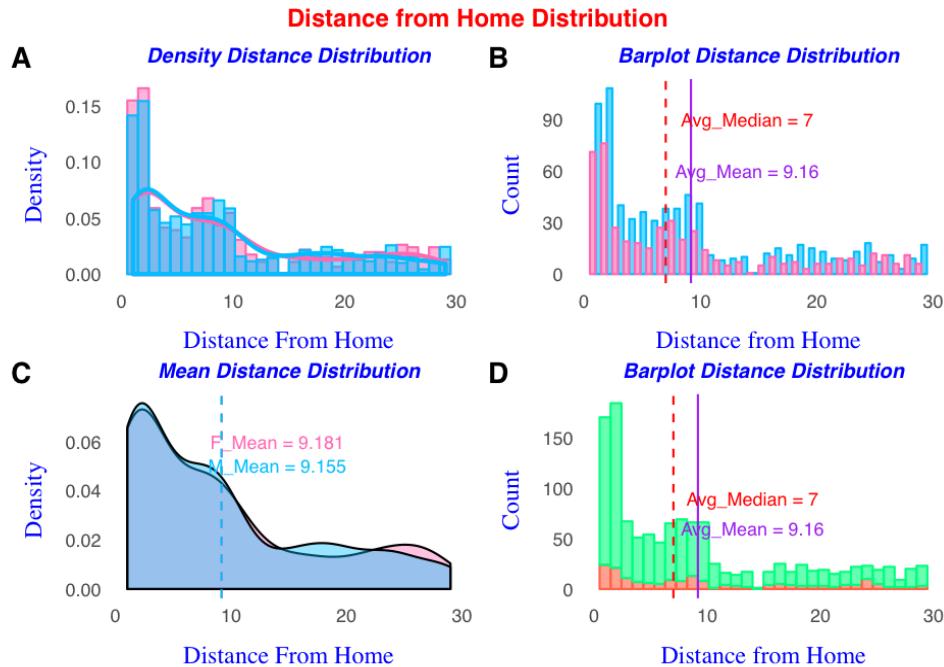
C

- Most of the Male and Female Employee have work Life Balance Score equally splitted between all scores.

D

- WorkLife Balance of Score 1 are more likely to Leave as we can see there 27% Attrition.

Distance from Home



A

- Most of the Employee lives pretty close to Home.
- Both Male and Female employee lives within 10 Miles Radius.

B

- Average Median Distance for all Employee is 7.
- Averageg Mean Distance for all Employee is 9.16.

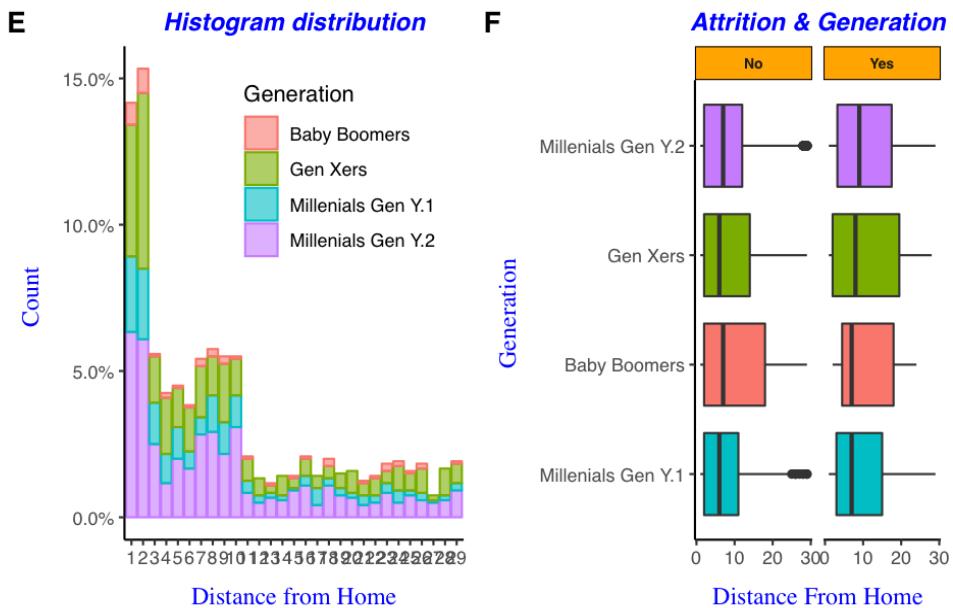
C

- Mean Distance for All Female Employee is 9.181
- Mean Distance for All Male Employee is 9.155.

D

- Distance from Home is less likely factor for Attrition as we see very less Attrition just based on Distance from Home.

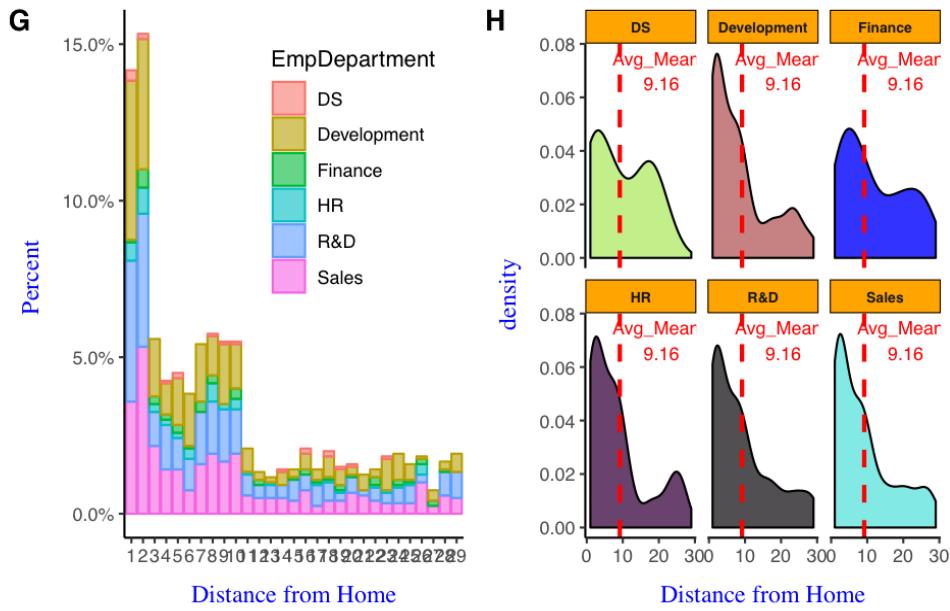
Distance from Home



E - Lot of Millenials Gen Y.2 within 10 Mile Radius. - 30% Employee lived around 2 Miles Radius.

F - Distance from Home doesn't seems like one of the attrition features among different Generations.

Distance from Home Across Department

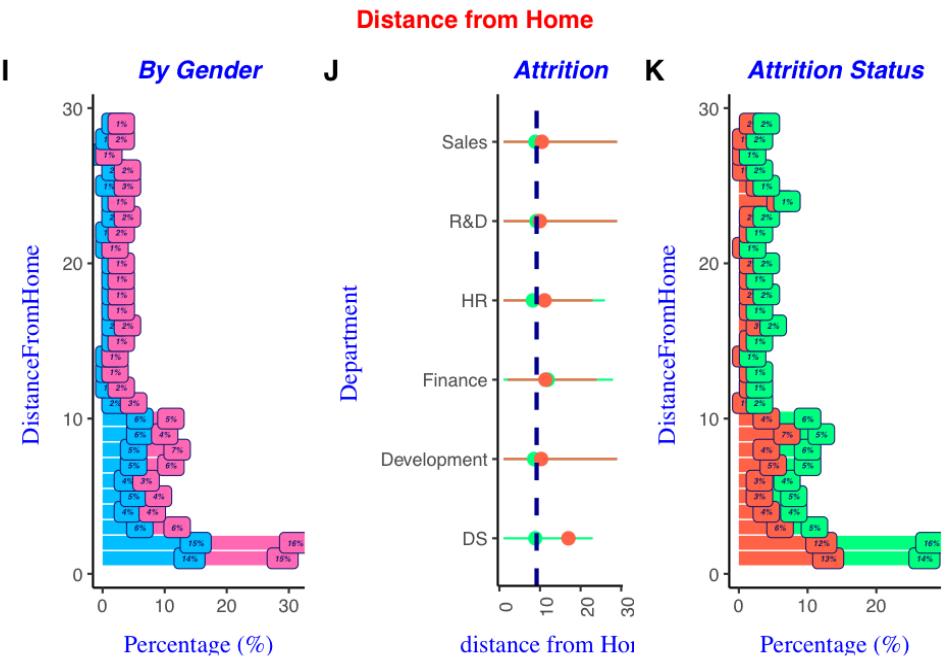


G

- When We break it down by the department we don't see patterns of Attrition could be distance from Home.
- Three big departments have attrition more than any other departments.

H

- Except Data Science and Finance all departments employee have same pattern of living.



I

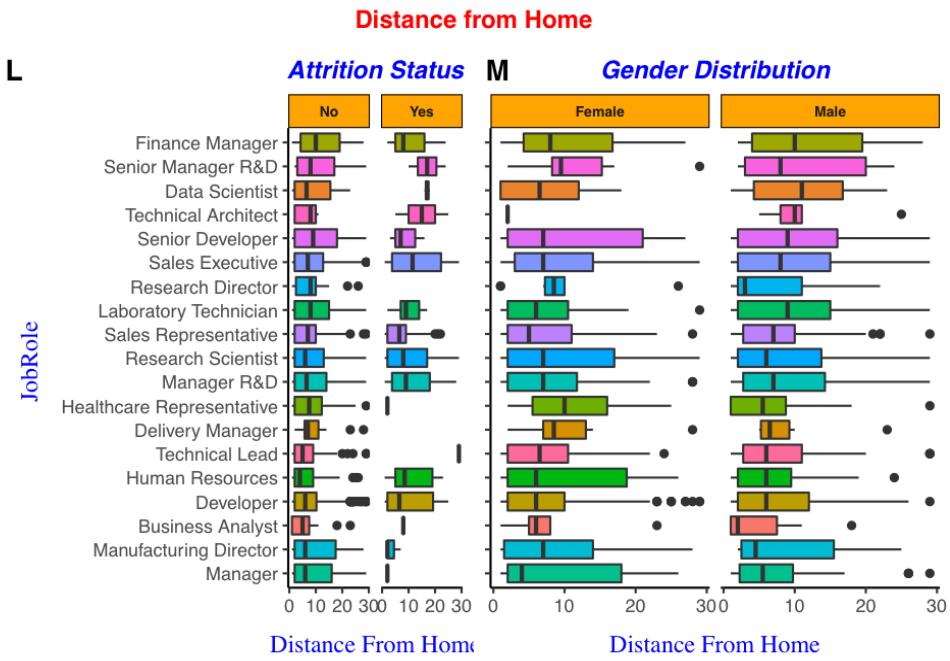
- Male and Female are equally distributed how far do they live from the Workplace.

J

- Most of them fall right in middle.

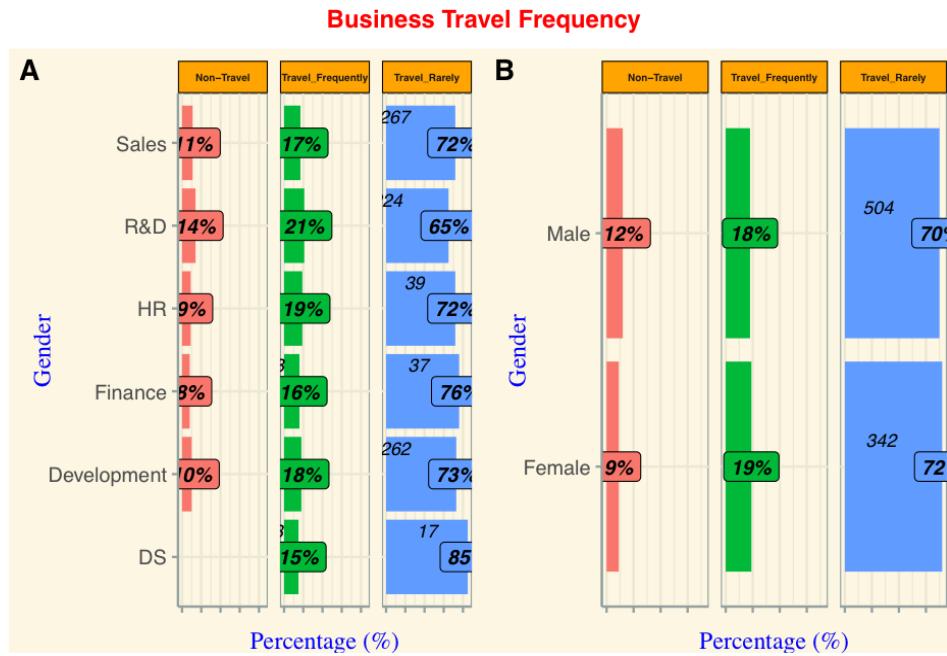
K

- Distance from the Home seems like Independent factor for Attrition.
- We can see there are some outliers based on their position but all outliers didn't leave the company just because of commuting distance.



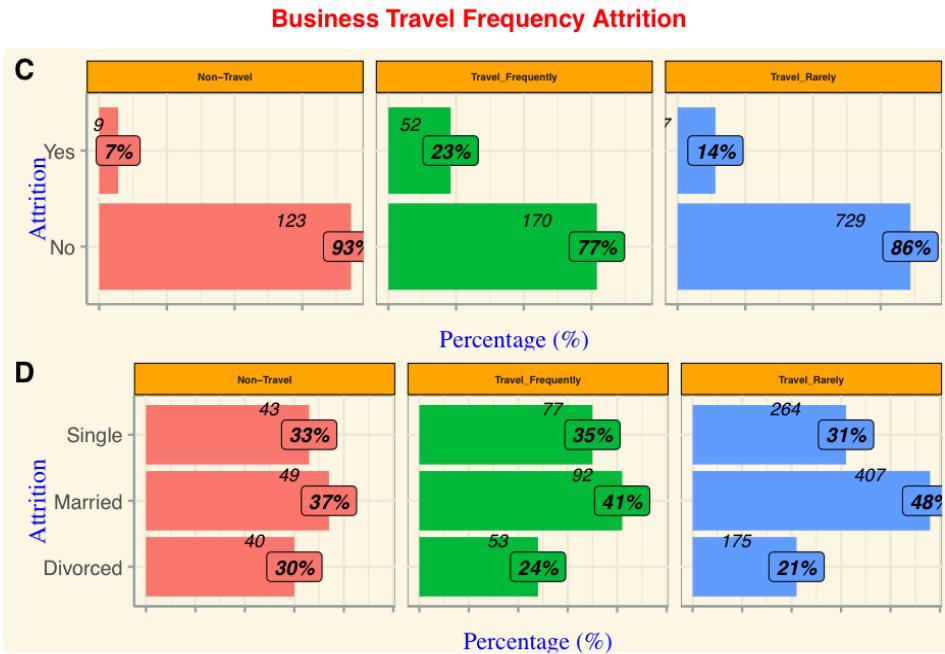
L & M - One can make lot of Interpretation based on Which Position Employee leave how far from the Workplace. - And Also if the Employee is Male and Female and what's the trend look like based on Gender.

Business Travel



A - Seems like in each department most of the Employee do travel rarely.

B - When we break it down by Male and Female we see same proportion in all 3 groups.



C - Travel Frequently groups tend to leave the company compare to rarely and non travel.

D - Travel Rarely comprises mostly Married Person and least comprise of Divorced Employee similar is the case with Travel Frequently.

Hourly Income

i. Hourly Income | Gender | Department

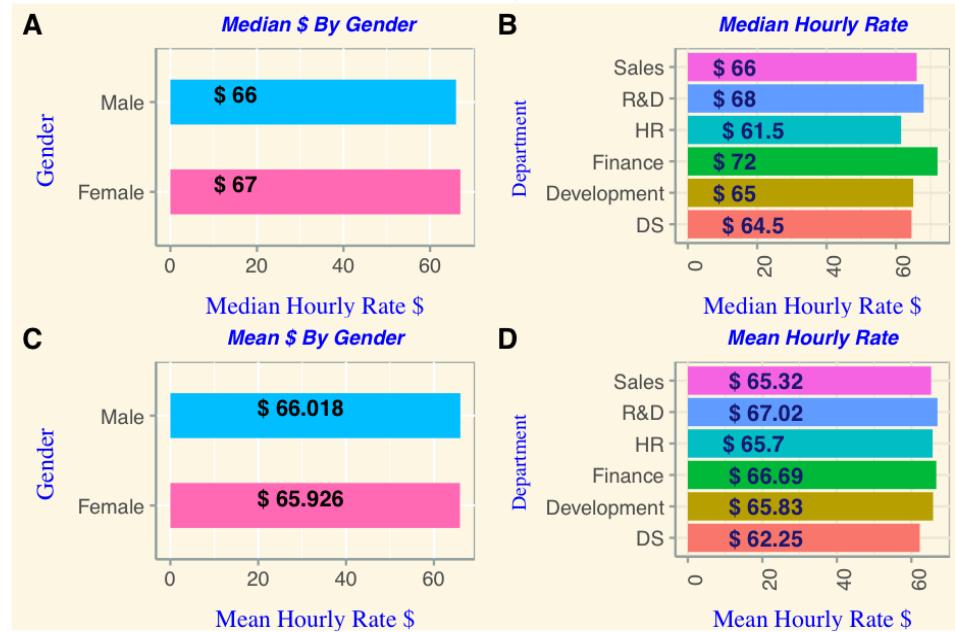
```
## # A tibble: 10 x 4
## # Groups:   EmpHourlyRate, EmpJobRole [9]
##   EmpHourlyRate EmpJobRole   Gender count
##       <int>     <fct>      <fct>  <int>
## 1        100 Business Analyst  Male     1
## 2        100 Developer        Male     2
## 3        100 Human Resources Male     2
## 4        100 Manager R&D    Female   1
## 5        100 Research Scientist Male     1
## 6        100 Sales Executive Female   3
## 7        100 Sales Executive  Male     1
## 8        100 Sales Representative Female  1
## 9        100 Senior Developer  Male     1
## 10       100 Technical Lead   Male     1
## 
## hr_df$EmpHourlyRate    N= 1200
## 
## +-----+-----+-----+
```

```

## |           |N|hr_df$EmpHourlyRate|
## +-----+-----+
## |hr_df$Gender|Female| 475|65.92632
## |           |Male   | 725|66.01793
## +-----+-----+
## |Overall    |      |1200|65.98167
## +-----+-----+

```

Mean and Median Hourly Rate Distribution



A

- Is it True that Male tends to make less compare to Female. There is a dollar difference in an Median hourly Rate?
- On Average Male and Female throughout the organisation makes more or less around \$67 an hour.

B

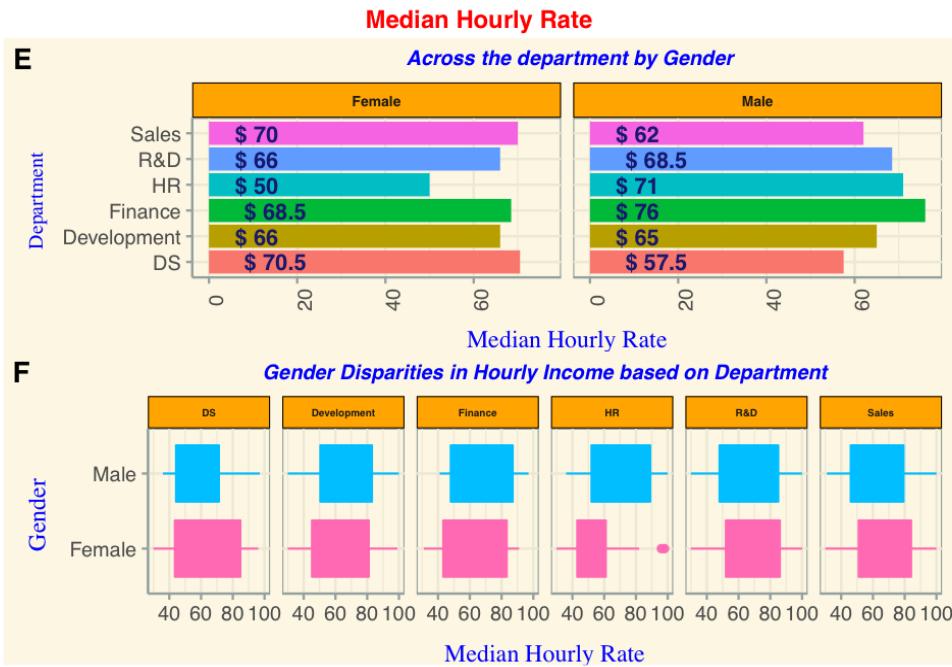
- When we break it down by the Department we can see their is variation.
- Finance Employee on Average tends to make more and HR employee tends to make Less per hour if we look Median Hourly Rate.

C

- If we took Overall Mean we see Male making \$66.02 compare that to Female \$65.93. Seems like there is barely any difference.

D

- DS Employee seems to be on low end compare to Rest of the Department.

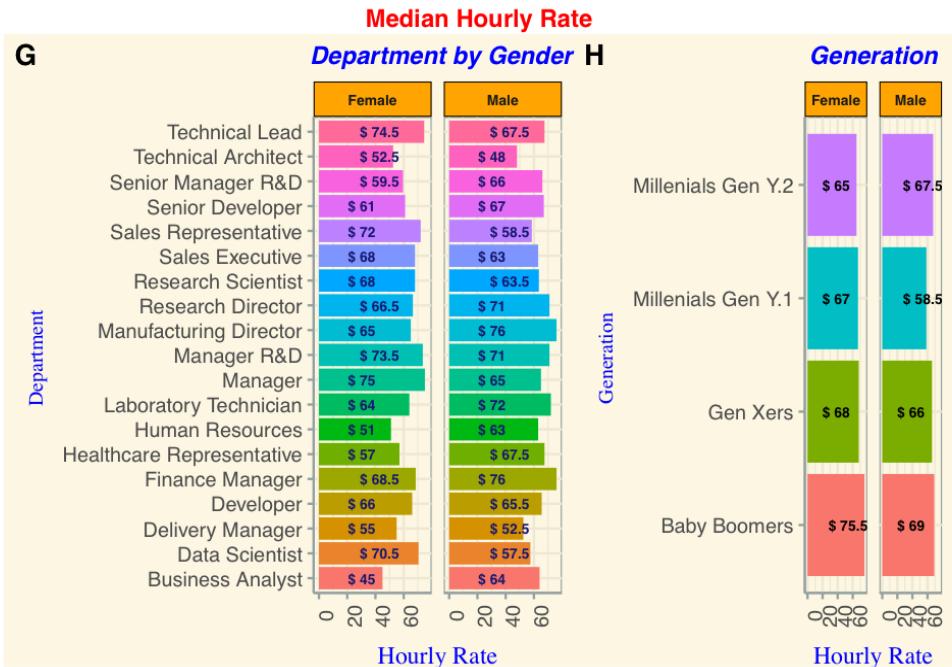


E

- When we even break it down by Department and Gender we can see some disparities.
- Taking Median Hourly Income is not affected by Outliers which we can see in Human Resource Department.
- When we break it down by the Gender and Department now we can see the Wage disparities.
- There is Huge difference in Human Resource Department, Male making around \$71/hr compare to Female \$50/hr.
- In Sales Female tends to make \$70/hr compare to Male \$62/hr
- In Finance Female tends to make \$68.5/hr compare to Male \$76/hr
- But it won't be True until and unless we consider same Job Role position and Work Experience to make that judgement.

F

- In box plot we can see only outlier which is present in Human Resource Department but there are no any outliers in any other department to make that call.

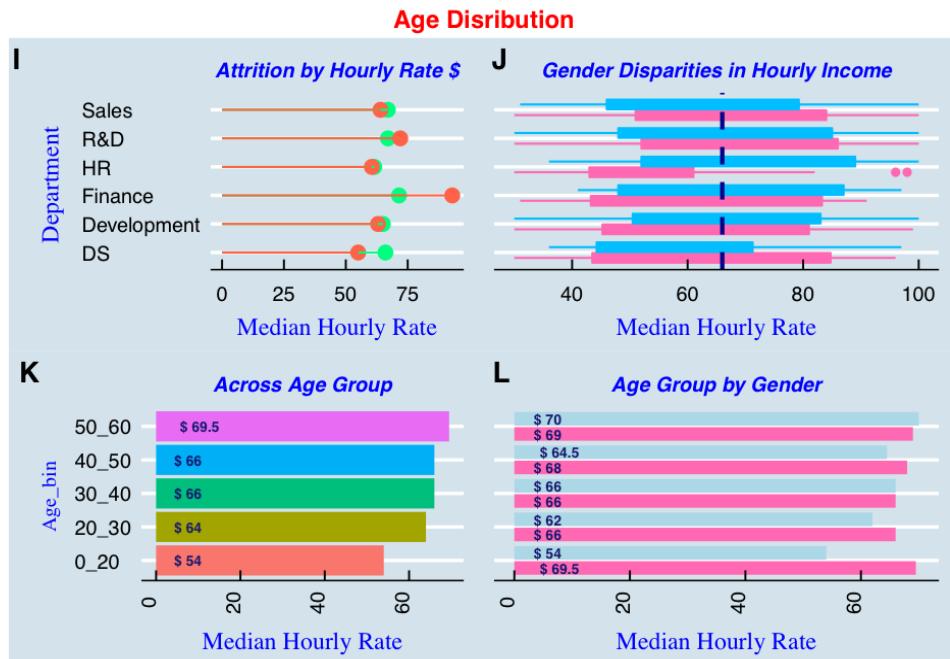


G

- When considering breaking it down by the Job Role it makes more sense as we can see the variation among the Gender.
- In Some of the Role Female tends to make more than Male and in Some Field Male tends to make more than Female.
- Another factor that need to be reconsider before we make assumption is Years of Experience and their Education level.
- Its harder to compare and reach out to conclusion until and unless we compare apples to apples.
- Variability factor could be Years of Experience either in current Job or Past Experience, Education level.

H

- Baby Boomers(Age:) makes more money which does make sense.
- Male Millenials Y.1 makes less money even compare to Female Millenials Y.1.



I

- Is Hourly Rate a Reason for Employees to Leave Department?
- In Some it could be the case but in some it's Not. In Finance Employee doesn't seem to live just by Hourly Rate. In others there could be possibility.

J

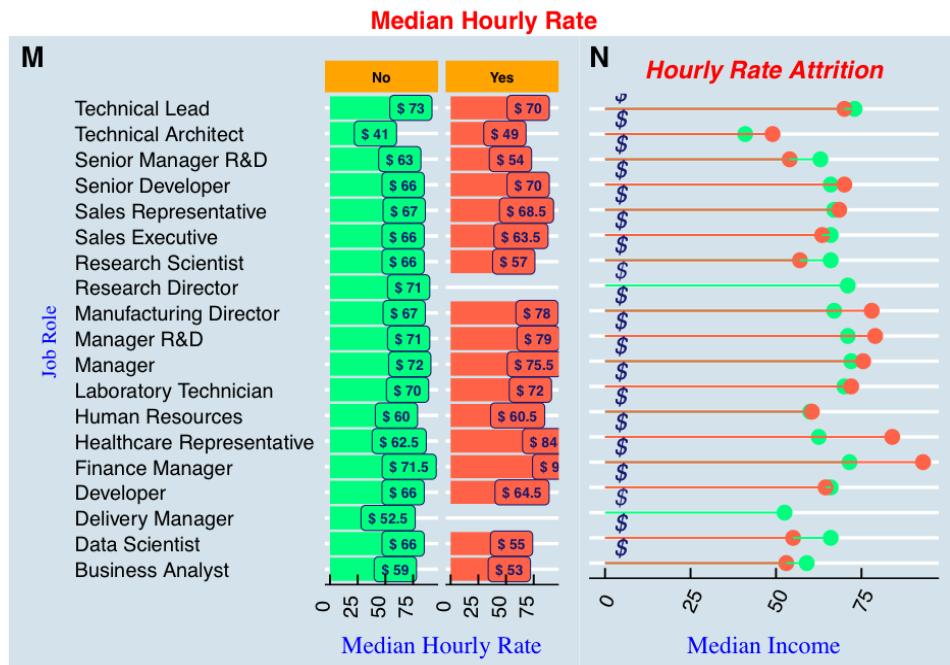
- We can see Gender Disparities between Genders across Department.

K

- When we break it down by Age group as Employee Age they tend to get paid better, which could be because of Senior Role or Experience.

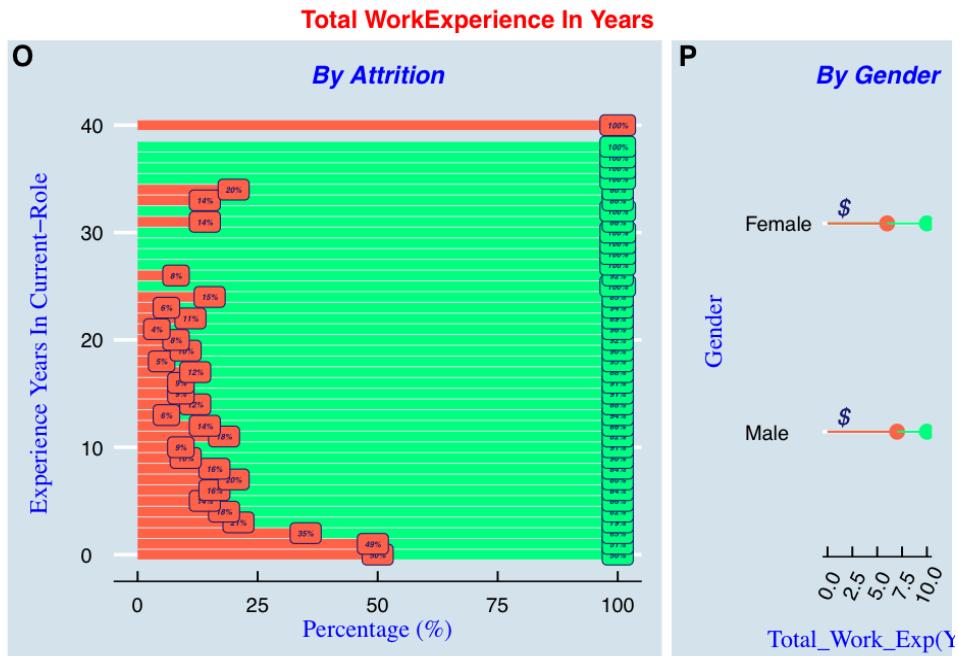
L

- When we break those age group by Gender we can see the Male of Age below 20 seems to be paid badly compared to Female of same Age group.



M & N

- The difference between Employee Leaving the company to staying is very Minimal.
- As this is calculated in hourly Rate, it could rack up huge when you calculated yearly.



O

- One can generalise how Work Experience might count on Employee to either leave or stay.
- We can see that as we move up Experience in years we see more green except one outlier which could possible be retired or death or could left for better opportunities and money.
- But overall as someone has more work experience they seem to stay rather than leave.

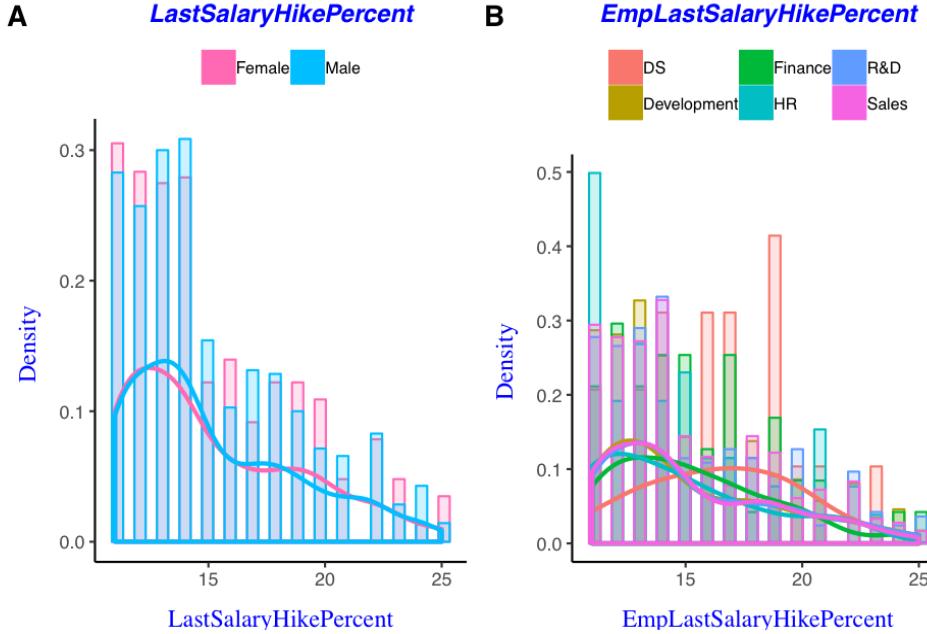
P

- In both gender we can see that more you spent gaining more experience you are more likely to not leave the company.

Salary Hike Percent

- Salary Hike Percent across Department
- Salary Hike Percent across age & Dept.
- Salary Hike Percent segregated by Gender
- Salary Hike Percent segregated by Attrition
- Percent Hike and attrition between gender (iii + v)

Comparison of Salary Hike Percent by Department & Gender

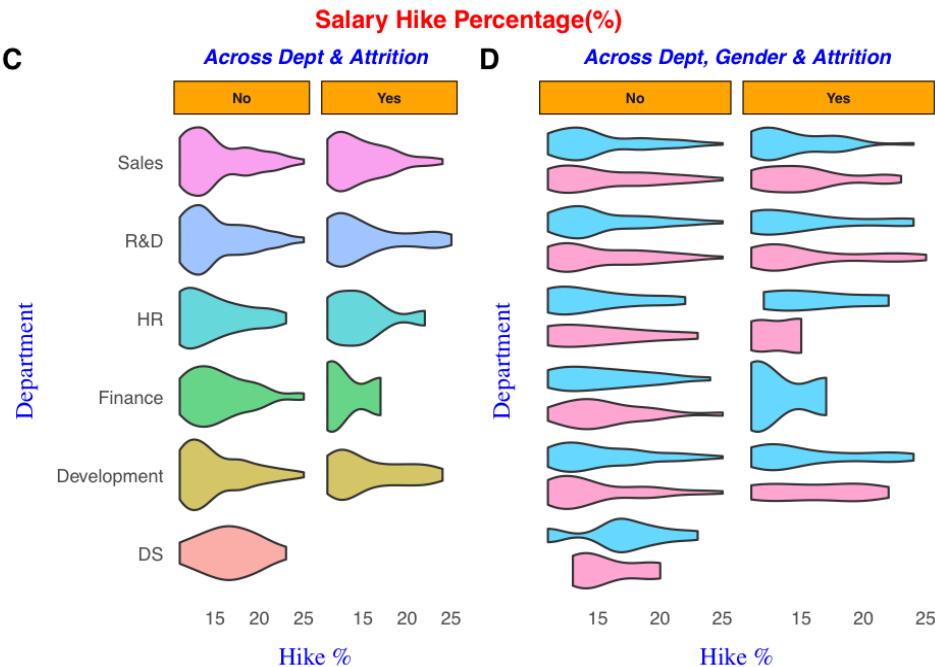


A

- We have Salary Hike Percentage ranging from 11% to 25%. If you see its Left Skewed meaning there is more probability that you will have 11-14% of Salary Hike percentage in the beginning and then it slows down.
- If you see the Department we can see Data Science Department(Blue line) stands out than other departments.

B

- There is no bias in Salary Hike Percentage based on Gender.



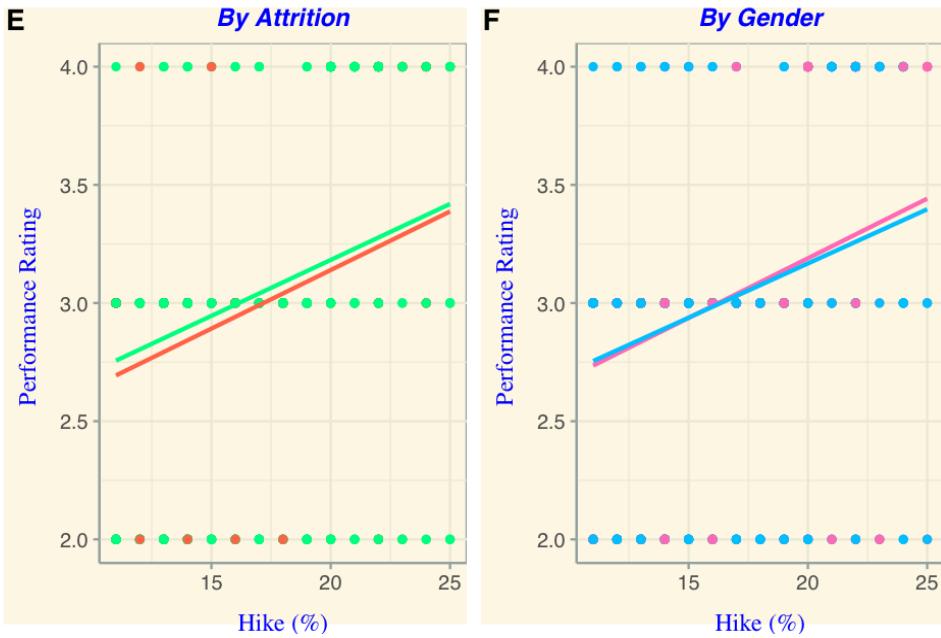
C

- How does Salary Hike Percentage helps Employee to either stay or leave the company.
- When we break it down by Department we can see in Finance more than 15% and less than 20% have kept the employee to work.
- In Data Science Department Increase Salary has kept employee to not leave the company.

D

- In Finance department Salary Hike Percentage more than 15 and less than 20 % has kept all to work.
- In Data Science field Employee doesn't seem to leave if there is Salary Hike which can be seen in both Gender.
- In Finance all the employee who left beside Salary Hike Percent are all Male.
- In Finance may be if there is Salary Hike Percent, we can keep most of the Female employee from leaving the company.

Comparison of Salary Hike Percent to Performance Rating

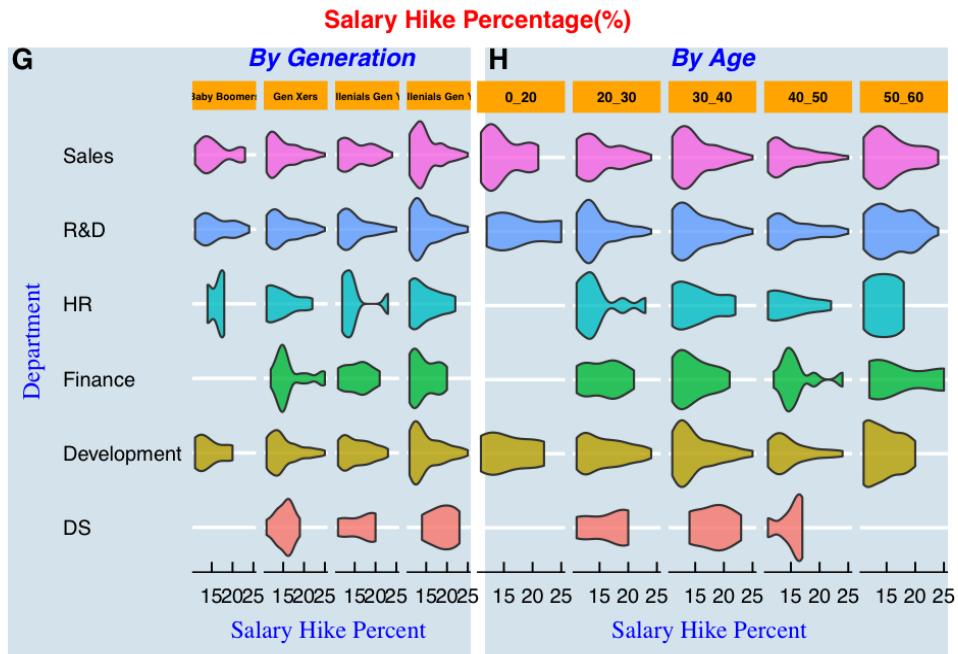


E

- Overall as performance rating increases the Salary Hike Percentage does increase.

F

- Highest Performance Rating does have edge over Hike Percentage those who stays and those who leave.

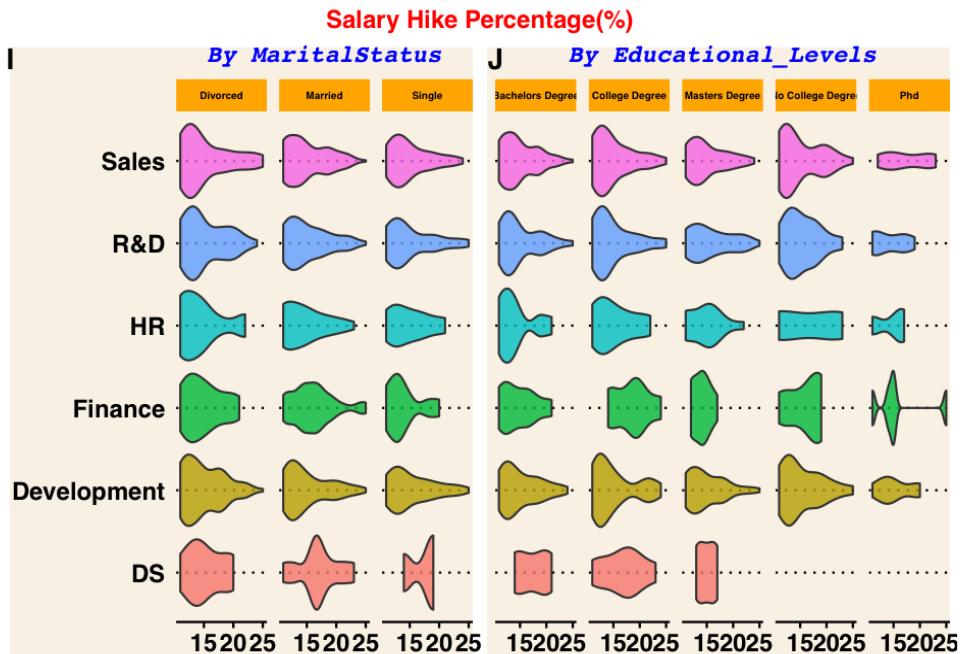


G

- Baby Boomers(Age:55 -76) in Data Science and Finance field doesn't get Hike % compare to rest of the department.
- We can see segregation in HR Millenials Gen Y.2.

H

- We can see some of the interesting features in HR for Employee Age between 20_30 and In Finance between Age 40_50.

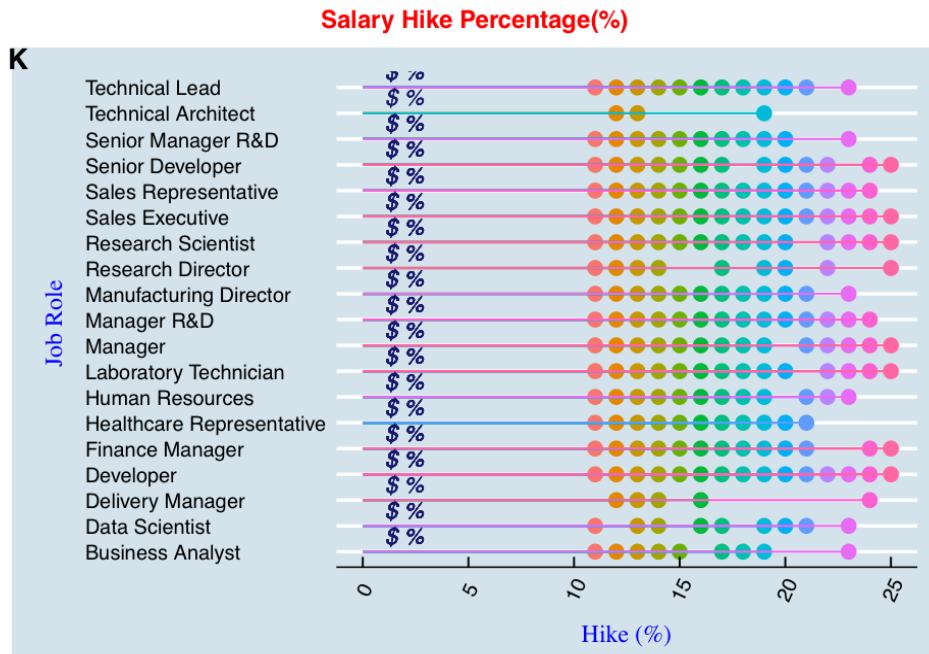


I

- Three big department doesn't care so much about Salary Hike based on our Martial Status.
- In Data Science field we can see Martial Status has some influence. Single are restricted to 20% Hike. Married Employee gets 15% to 20% Hike and flattens out. Divorced are restricted to 20 %.
- In Finance Singles are restricted to 20% Salary Hike.

J

- If we look how does the education level have helped employee to get more salary hike percent, we can see again interesting pattern in Data Science department.
- No college Degree and PhD employee doesn't get any salary increment.
- In Finance we can see segregation in PhD Employee. There are employee who get Salary Hike Percent around 15% and then there are very few Employee who gets around 25%.
- One can say it might depend upon Job Role and experience than education which we can look by Job Role.



K

- Median hike percentage.
- All the job Roles start at 11% Hike except Delivery manager and Technical Architect who gets 12% Hike.
- There are some jobs which gets more Hike % than others.
- Developer & Sales Executive have variation in Hike % more than any other jobs.



L

- When we try to break it down by Department we can see Interesting patterns
- Data Science Employees have performance Rating of only 3.
- Their Salary Hike Percentage is maximum upto 20%.
- Either Low or High Performance Rating Data Scinetist doesnt get any Salary Hike Percentage.
- For rest of the Department even if you have low Performance rating you still can get upto 25% Salary hike.
- If you have Performance Rating of 4 majority of the employee gets more than 20% Salary Hike.

M

- As we have more Male Employee we can see Attrition has been condensed in Male Employee whereas it has been spread out in Female.

Show 10 entries Search:

| | Gender | Age_bin | Hike_Pct | Freq |
|----|--------|---------|----------|------|
| 1 | Female | 0_20 | 11-13 | 0.17 |
| 2 | Male | 0_20 | 11-13 | 0.33 |
| 3 | Female | 20_30 | 11-13 | 1.75 |
| 4 | Male | 20_30 | 11-13 | 4.33 |
| 5 | Female | 30_40 | 11-13 | 5.42 |
| 6 | Male | 30_40 | 11-13 | 6.17 |
| 7 | Female | 40_50 | 11-13 | 2.83 |
| 8 | Male | 40_50 | 11-13 | 3.67 |
| 9 | Female | 50_60 | 11-13 | 1.08 |
| 10 | Male | 50_60 | 11-13 | 1.25 |

Showing 1 to 10 of 50 entries Previous 2 3 4 5 Next Search:

Show 10 entries Search:

| | EmpDepartment | Hike_Pct | Freq |
|----|---------------|----------|-------|
| 1 | DS | 11-13 | 0.17 |
| 2 | Development | 11-13 | 8.25 |
| 3 | Finance | 11-13 | 1.00 |
| 4 | HR | 11-13 | 1.50 |
| 5 | R&D | 11-13 | 7.50 |
| 6 | Sales | 11-13 | 8.58 |
| 7 | DS | 14-16 | 0.42 |
| 8 | Development | 14-16 | 10.50 |
| 9 | Finance | 14-16 | 1.42 |
| 10 | HR | 14-16 | 1.50 |

Showing 1 to 10 of 30 entries Previous 2 3 Next

```
##   Gender Age_bin Hike_Pct Freq
## 8   Male   30_40    14-16 9.92
```

Show 10 entries Search: |

| | Gender | Age_bin | Hike_Pct | Freq |
|----|--------|---------|----------|------|
| 1 | Male | 0_20 | 11-13 | 0.33 |
| 2 | Male | 20_30 | 11-13 | 4.33 |
| 3 | Male | 30_40 | 11-13 | 6.17 |
| 4 | Male | 40_50 | 11-13 | 3.67 |
| 5 | Male | 50_60 | 11-13 | 1.25 |
| 6 | Male | 0_20 | 14-16 | 0.33 |
| 7 | Male | 20_30 | 14-16 | 5.17 |
| 8 | Male | 30_40 | 14-16 | 9.92 |
| 9 | Male | 40_50 | 14-16 | 4.92 |
| 10 | Male | 50_60 | 14-16 | 1.92 |

Showing 1 to 10 of 25 entries

Previous 1 2 3 Next

```
##   Gender Age_bin Hike_Pct Freq
## 3 Female  30_40    11-13 5.42
```

Show 10 entries Search: |

| | Gender | Age_bin | Hike_Pct | Freq |
|----|--------|---------|----------|------|
| 1 | Female | 0_20 | 11-13 | 0.17 |
| 2 | Female | 20_30 | 11-13 | 1.75 |
| 3 | Female | 30_40 | 11-13 | 5.42 |
| 4 | Female | 40_50 | 11-13 | 2.83 |
| 5 | Female | 50_60 | 11-13 | 1.08 |
| 6 | Female | 0_20 | 14-16 | 0.33 |
| 7 | Female | 20_30 | 14-16 | 3.42 |
| 8 | Female | 30_40 | 14-16 | 5.33 |
| 9 | Female | 40_50 | 14-16 | 2.67 |
| 10 | Female | 50_60 | 14-16 | 1.17 |

Showing 1 to 10 of 25 entries

Previous 1 2 3 Next

Select by Freq

- In our datasets about 9.92% of Employee of Male Age between 30-40 have got 14-16% Salary Hike.
- In our datasets about 5.42% of Employee of Female Age between 30-40 have got 11-13% Salary Hike as we don't have equal number of Males and Female in our datasets it makes sense. It doesn't mean that there is any biases as we have shown in bar plot chart that they are treated equally in terms of hike Percentage.
- In our datasets about 11.17% of Employee from Sales Department got 14-16% Salary Hike which is the biggest.
- In our datasets about 10.50% of Employee from Development Department got 14-16% Salary Hike which is the biggest.

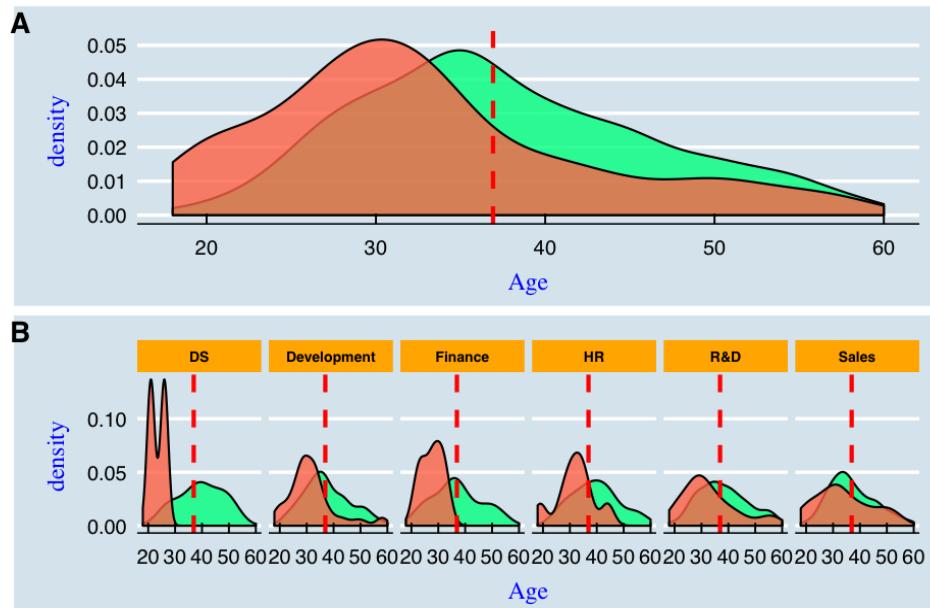
- In our datasets about 10.17% of Employee from R&D Department got 14-16% Salary Hike which is the biggest.

ATTRITION

Our main question for doing all this Exploratory data analysis beside finding the Relationship between the variables is also to understand the Attrition Rates and diagnose the cause and find useful insights. Lets look at the Attrition Number of our datasets.

Attrition by Age

Relationship between Age and Attrition



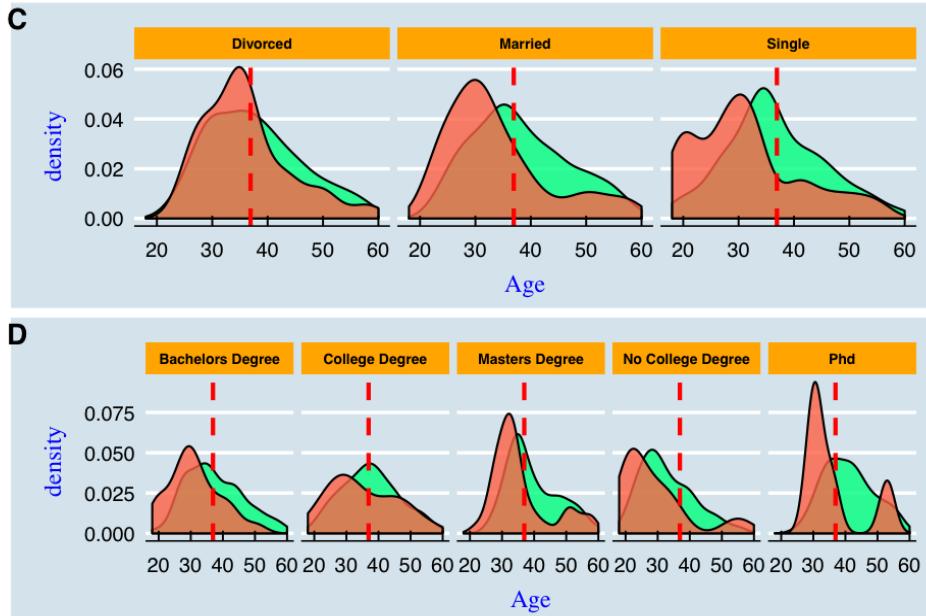
A

- As we have look into all the components, now its time to see how is the attrition distributed across various variables.
- In Finance department we can see the younger Employee seems to leave the company.
- In Data Science also we can see clear differentiation.
- In Human Resource we can also early retention but after mean age they are less likely to leave the company.
- Looking through out the department.

B

-Younger Employee tends to leave the company than to stay. - After Age 35 we can see more Employees tends to stay than to leave the company could be because of Financial Stability and years of experience to settle down.

Relationship between Age and Attrition



C

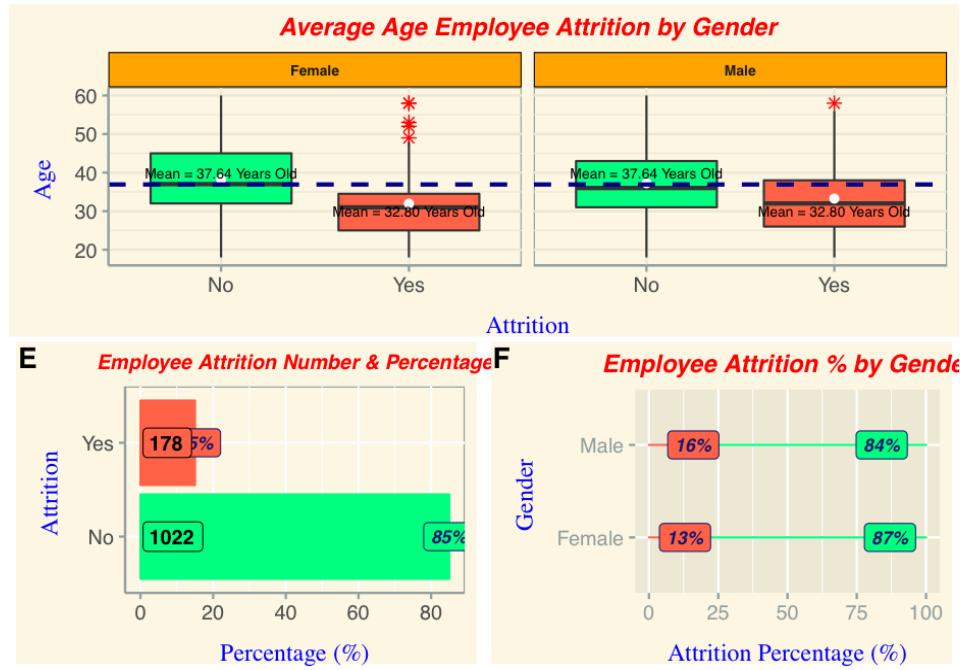
- Interesting Features in education level PhD. Rest of the education levels is somewhat more or less similar.

D

- Martial status does have some effect on Attrition. Just look at geom density between Single and rest.

Attrition Number & Percentage by Gender

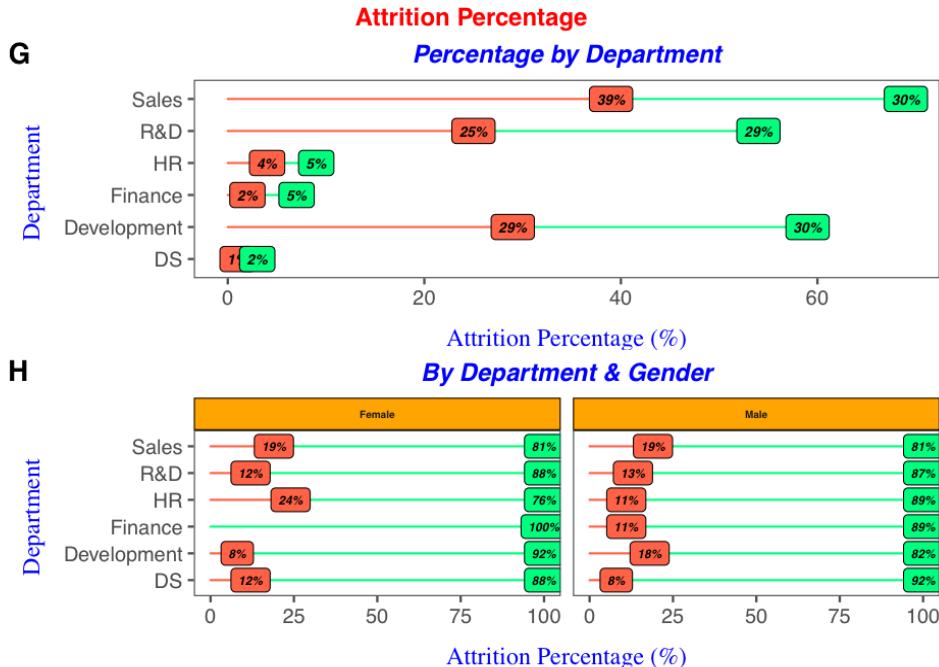
```
##   Gender Attrition count att_gender_percent
## 1 Female      No    412           87
## 2 Female     Yes     63            13
## 3  Male      No    610           84
## 4  Male     Yes    115            16
```



E & F

- 15% Leave the company while 85% still do work there.
- Out of 15 % who leave the company 16% are Male and 13% are Female.
- Average age of Employee who doesn't leave the company falls in Mean age of Entire Population of the workforce. Mean Age of the Employee who leave the company have less mean age in both Gender. Red Asterick Represents the outliers.

Attrition by Dept. & Gender



G

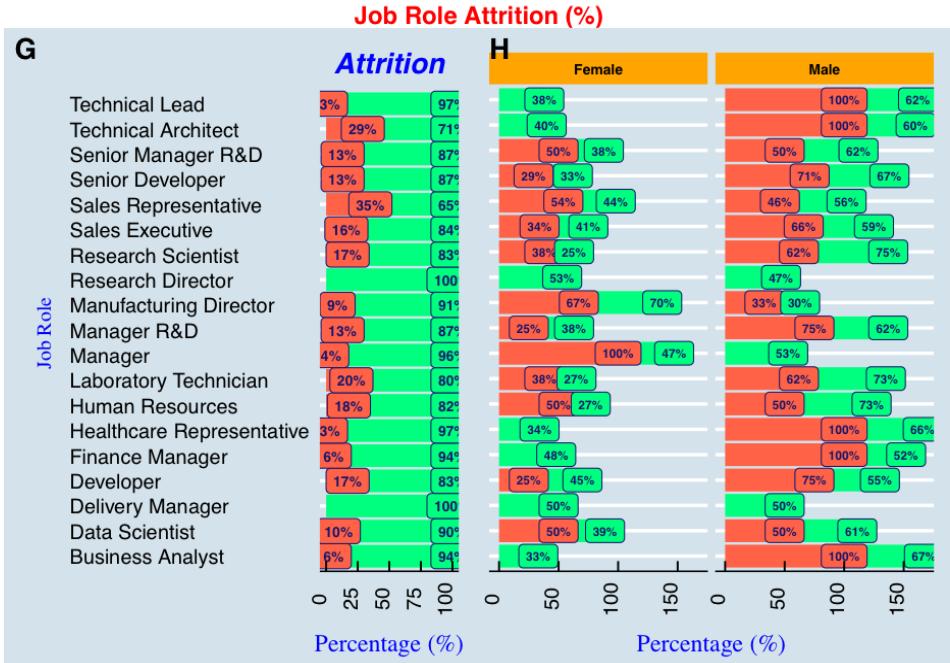
- As guessed we have biggest number of attrition coming from 3 big department Sales leading the more attrition than any other department.

H

- When we break it down by the Gender we can see that Males and Females leave the company at the same rate.
- In Finance we saw 2% Employee leaving the company and its all Male. None of the Female left the company.
- Most of the Female Working in Human Resource department seems to leave the company. Out of 4% of total attrition 24% of the Employee seems to be Female which is double than Male 11% leaving the company.

Attrition Number by Employee Job Role

Lets look if we could find some insights in the job Role. May be this could give us some insights on how the Attrition factor plays any role.



- Sales Representative contribute more for the attrition. Almost about 35% of them leave the company.
- Out of 35% Male and Female almost participate roughly equally. 54% are Female and 44% are Male.
- 3% Technical Lead are all Male.
- Seconds is Technical Architect. About 29% of Technical Architect leave the company and all of them who leave the company are Male.
- Senior Manager R&D. 13% Employee leave out of which half of them are Male and other half is Female.
- Research Director never leaves the company at all.
- 4% Manager who leaves the company are all Female.
- 3% Healthcare Representative who leave the company are all Male.
- 6% Finance Manager who leaves the company are all Male.
- Delivery Managers never leave the company at all.
- 10 % Data scientist who leave the company. Half of them are female.
- 6% Business Analyst who leave the company are all Male.

Looking at above insights company can see the patterns coming out of particular Gender and then decide what factors might result them to leave the company.

- I have subset trying to find if there is any insights among any of the variables.
- There are in total 8 Employee who have left the company who are all Male.

```
##   EmpNumber Age Gender EducationBackground MaritalStatus EmpDepartment
## 1 E1001405  39   Male      Life Sciences     Divorced        R&D
## 2 E1001882  28   Male       Medical      Married        Finance
## 3 E1001962  23   Male       Medical     Single        Finance
## 4 E1001970  32   Male      Life Sciences    Single        Finance
## 5 E100333   26   Male      Life Sciences    Single   Development
## 6 E100343   28   Male  Technical Degree    Single   Development
## 7 E100457   51   Male      Life Sciences    Single   Development
```

```

## 8 E100560 38 Male Medical Married Development
##                           EmpJobRole BusinessTravelFrequency DistanceFromHome
## 1 Healthcare Representative Travel_Frequently 2
## 2 Finance Manager      Travel_Rarely   24
## 3 Finance Manager      Travel_Rarely   8
## 4 Finance Manager      Travel_Rarely   2
## 5 Technical Architect  Travel_Rarely   25
## 6 Technical Architect  Travel_Rarely   5
## 7 Business Analyst     Travel_Frequently 8
## 8 Technical Lead       Travel_Rarely   29
##   EmpEducationLevel EmpEnvironmentSatisfaction EmpHourlyRate
## 1            3                  1             84
## 2            3                  3             51
## 3            1                  4             93
## 4            4                  4             95
## 5            3                  1             48
## 6            4                  3             50
## 7            4                  1             53
## 8            1                  2             70
##   EmpJobInvolvement EmpJobLevel EmpJobSatisfaction NumCompaniesWorked
## 1            3                  4             4              7
## 2            3                  1             2              1
## 3            2                  1             3              1
## 4            3                  1             2              1
## 5            1                  1             3              1
## 6            3                  1             3              1
## 7            1                  3             4              2
## 8            3                  2             1              7
##   OverTime EmpLastSalaryHikePercent EmpRelationshipSatisfaction
## 1      No           11                  4
## 2    Yes           17                  3
## 3    Yes           11                  1
## 4      No           12                  1
## 5      No           12                  3
## 6    Yes           13                  3
## 7      No           15                  4
## 8    Yes           19                  2
##   TotalWorkExperienceInYears TrainingTimesLastYear EmpWorkLifeBalance
## 1            21                  4             3
## 2            1                  3             3
## 3            5                  2             3
## 4            1                  2             3
## 5            1                  2             2
## 6            2                  3             2
## 7            18                 2             3
## 8            17                 2             3
##   ExperienceYearsAtThisCompany ExperienceYearsInCurrentRole
## 1            18                  7
## 2            1                  1
## 3            5                  4
## 4            1                  0
## 5            1                  0
## 6            2                  2
## 7            4                  2

```

```

## 8          1          0
##   YearsSinceLastPromotion YearsWithCurrManager Attrition PerformanceRating
## 1           11          5    Yes      2
## 2            0          0    Yes      3
## 3            1          2    Yes      3
## 4            0          0    Yes      3
## 5            0          1    Yes      3
## 6            2          2    Yes      3
## 7            0          3    Yes      4
## 8            0          0    Yes      3
##   Educational_Levels Age_bin       Generation Hike_Pct
## 1 Bachelors Degree  30_40        Gen Xers  11-13
## 2 Bachelors Degree  20_30        Millenials Gen Y.1 17-20
## 3 No College Degree 20_30        Millenials Gen Y.1 11-13
## 4 Masters Degree   30_40        Millenials Gen Y.2 11-13
## 5 Bachelors Degree  20_30        Millenials Gen Y.1 11-13
## 6 Masters Degree   20_30        Millenials Gen Y.1 14-16
## 7 Masters Degree   50_60        Gen Xers  14-16
## 8 No College Degree 30_40        Millenials Gen Y.2 17-20
##   CatYearsManager hourly_bin
## 1 2-5 Years hired   80_89
## 2 Recently Hired Newbie 50_59
## 3 1-2 years Old     90_100
## 4 Recently Hired Newbie 90_100
## 5 1-2 years Old     40_49
## 6 1-2 years Old     50_59
## 7 2-5 Years hired   50_59
## 8 Recently Hired Newbie 70_79

```

ANALYSIS

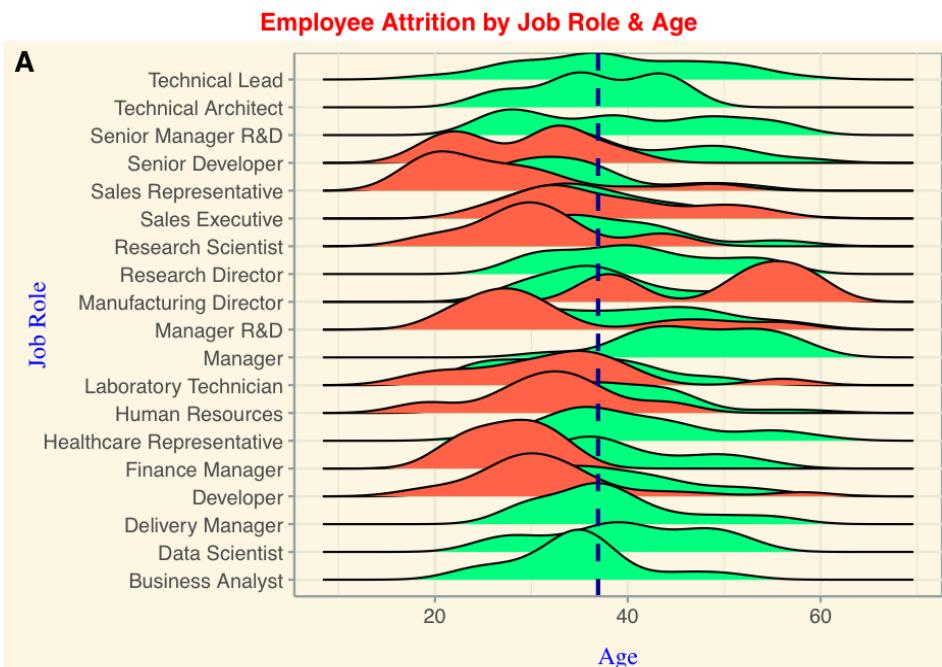
- 5 out of 8 are Single.
- 4 Works in Development, 3 works in Finance and 1 in R&D
- 3 are Finance Manager, 2 are Technical Architect, 1 Technical Lead, 1 Business Analyst and 1 is Healthcare Representative.
- 6 out of 8 Travel Rarely.
- 3 out of 8 leave very far from Home.
- 5 out of 8 have Job Involvement category 3.
- 5 out of 8 have Job level 1.
- 5 out of 8 have worked only one previous company.
- 4 have worked OverTime and 3 have Not.
- Salary Hike Percent is from 11-19.
- 3 of them have Workexperience of more than 17 years.
- 6 out of 8 have WorkLife Balance score of 3 which is Better.
- 4 out of them have just spent about a year in this company.
- 5 out of 8 have been promoted last year.
- One of them is working for 5 years with Current Managers while rest of them have either 0,1,2,3 years with current Manager. Could this be one of the crucial factor for them to leave.

- 6 out of 8 have performance rating of 3.
- 3 have Bachelor Degree, 3 have Master Degree and 2 of them have No college Degree.
- one of them makes Hourly between 40-49, 3 of them makes between 50-59, 2 of them makes between 90-100\$ an hour.
- We have seen Single leaving the company so Single creates High Volatility.
- Did 3 of them who have Bachelor degree left to get Higher Education Degree.
- All of them have Salary Hike percent.
- Was Current Managers somehow responsible for Employee to leave.
- We could deep digger to find out if cuurent Managers was somehow responsible for Employee to leave by categorising Employee who stayed under Current Manager and who left.

| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|------------------------------|-----|----|----|----|----|----|---|---|----|----|----|----|----|----|----|----|----|----|----|
| ## | | | | | | | | | | | | | | | | | | | |
| ## | | | | | | | | | | | | | | | | | | | |
| ## Business Analyst | No | 4 | 1 | 2 | 2 | 2 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Data Scientist | No | 5 | 1 | 2 | 0 | 3 | 0 | 0 | 3 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Delivery Manager | No | 2 | 1 | 2 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | |
| ## | Yes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Developer | No | 34 | 8 | 52 | 20 | 11 | 8 | 3 | 16 | 17 | 10 | 4 | 4 | 3 | 1 | 0 | 3 | 0 | |
| ## | Yes | 15 | 6 | 5 | 3 | 2 | 1 | 2 | 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Finance Manager | No | 3 | 2 | 15 | 5 | 3 | 0 | 1 | 7 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | |
| ## | Yes | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Healthcare Representative | No | 5 | 3 | 5 | 1 | 2 | 0 | 2 | 7 | 0 | 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Human Resources | No | 3 | 2 | 11 | 4 | 6 | 0 | 1 | 5 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 3 | 0 | 4 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Laboratory Technician | No | 8 | 2 | 19 | 6 | 3 | 1 | 1 | 6 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 7 | 0 | 3 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Manager | No | 9 | 4 | 4 | 0 | 4 | 0 | 2 | 8 | 3 | 4 | 3 | 1 | 3 | 1 | 1 | 0 | 1 | |
| ## | Yes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| ## Manager R&D | No | 10 | 3 | 18 | 10 | 7 | 4 | 2 | 15 | 4 | 3 | 1 | 1 | 1 | 2 | 0 | 0 | 1 | |
| ## | Yes | 6 | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Manufacturing Director | No | 1 | 1 | 7 | 3 | 4 | 0 | 0 | 5 | 5 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | |
| ## | Yes | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Research Director | No | 4 | 0 | 2 | 0 | 0 | 0 | 0 | 4 | 3 | 3 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | |
| ## | Yes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Research Scientist | No | 13 | 6 | 17 | 8 | 3 | 0 | 1 | 8 | 4 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | |
| ## | Yes | 6 | 0 | 2 | 2 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Sales Executive | No | 24 | 13 | 37 | 21 | 23 | 9 | 7 | 47 | 17 | 13 | 3 | 4 | 4 | 3 | 0 | 0 | 1 | |
| ## | Yes | 9 | 2 | 9 | 5 | 2 | 1 | 0 | 9 | 3 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | |
| ## Sales Representative | No | 15 | 1 | 22 | 1 | 0 | 1 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 13 | 0 | 9 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Senior Developer | No | 7 | 3 | 10 | 1 | 4 | 0 | 2 | 8 | 8 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Senior Manager R&D | No | 0 | 1 | 5 | 2 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Technical Architect | No | 1 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| ## Technical Lead | No | 3 | 3 | 12 | 4 | 2 | 0 | 0 | 7 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |
| ## | Yes | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

- When I break it with Current Managers I didn't find any concurrent evidence.

Attrition by Job Role And Age_bin



- We can see bell shape curve in both people who stays and work and people who leaves the company.
- Senior Developer, Sales Representative, Manager R&D, Healthcare Representative have early turnover before the mean age of Entire workforce.
- Blue dotted lines represent Mean Age of Workforce.
- Only Manufacturing Directors who tend to leave the company around the retirement Age.

Job Satisfaction & Attrition

How does the Job Satisfaction level play a role in terms of Attrition.



How does the Salary Increment affects the Attrition Rate Differentiated by the Gender.

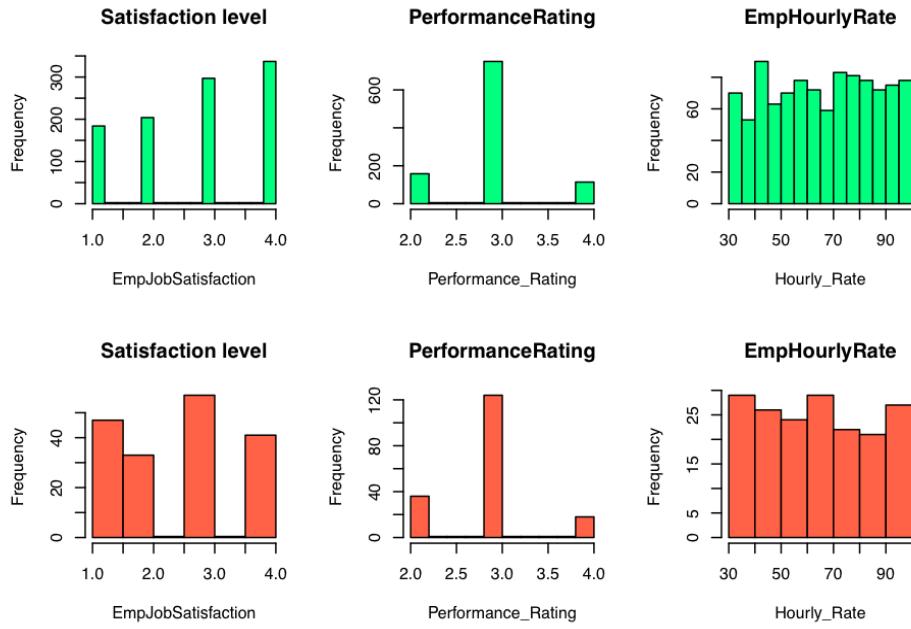
Hourly Income by Department & Attrition

How is Hourly Rate distributed across various Department?



Satisfaction level | PerformanceRating | EmpHourlyRate & Attrition

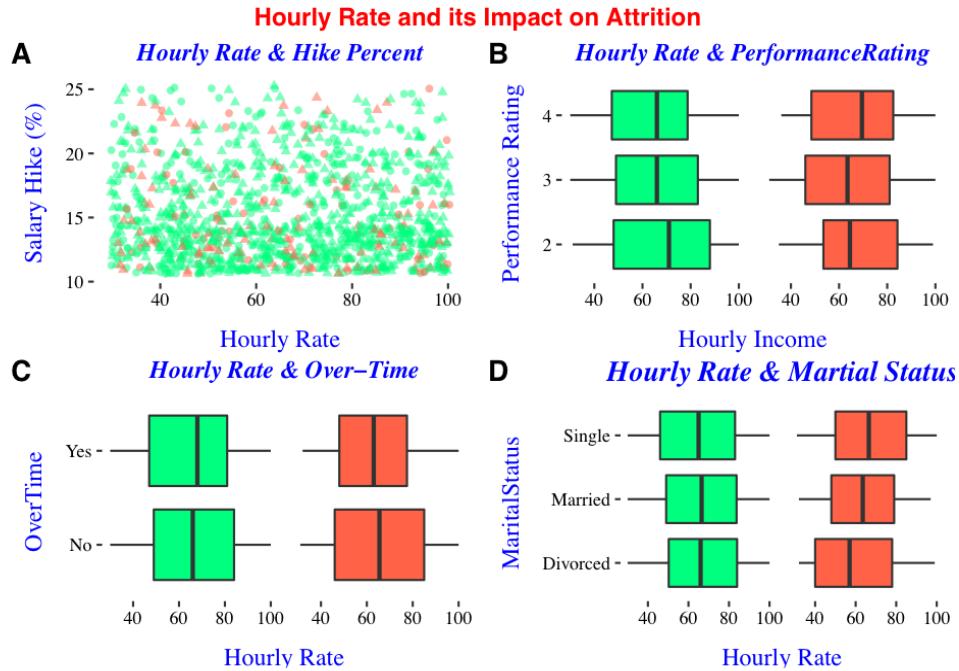
Let's see how does the Attrition factor is influenced by Satisfaction level, PerformanceRating, EmpHourlyRate.



Attrition Employee have low satisfaction level as well as low Hourly Rate. They have similar performance rating as that of Non Attrition Employee.

As we can see that there is definitely a reason for Employee to leave when it comes to Leaving or staying in the company. If the job satisfaction level is one or two, you are more likely to leave even increase in dollar per hour didn't save them. But if your satisfaction level is 3 or 4 and there is increase in hourly rate by \$5 then you are more likely to stay.

Hourly Rate and its Impact on Attrition



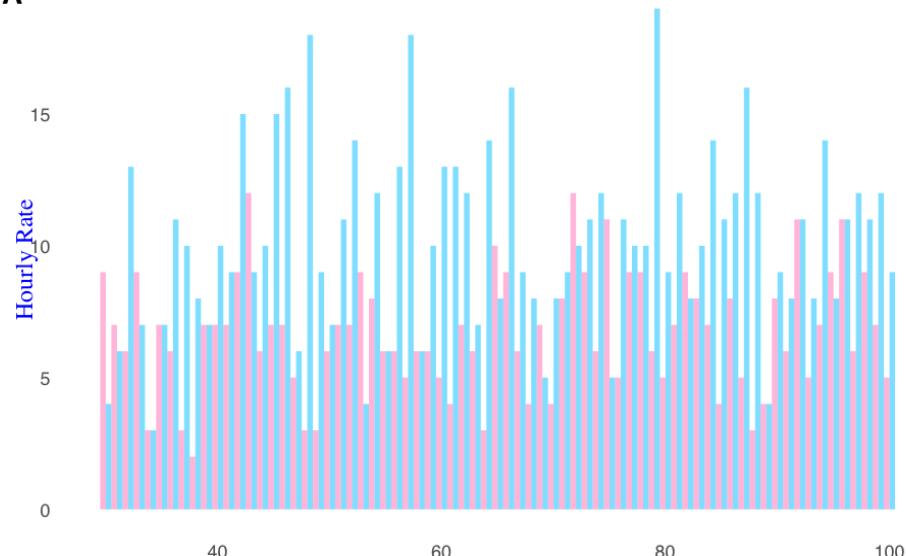
We will see some more variables and how does it affect the attrition level. Some of the factors like overtime, Martial Status, Performance Rating and Last Salary hike percent can be considered in finding some interesting observations.

- We can observe that Higher Salary/hour person are less likely to leave if the salary Hike percent is more than 15. - We can see some cluster at the bottom and spread out at the top.
- Overtime Employee that are paid less hourly leaves the company than that are well paid.
- low performance rating who tends to leave the company are less paid averaging somewhere around \$61-\$65
- People who leave the company: Single Status Employee make more hourly than Married and divorced
- But incase of people who stays seems to have same average hourly income.

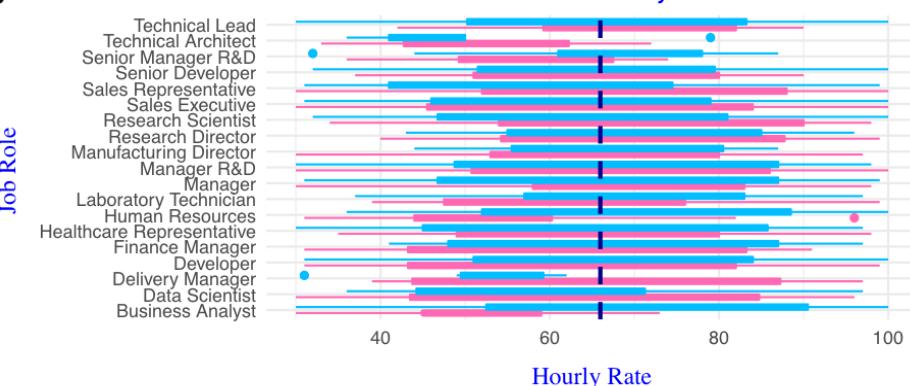
Wage disparities among Department | Gender | Attrition

Overall Hourly Income distribution by Gender

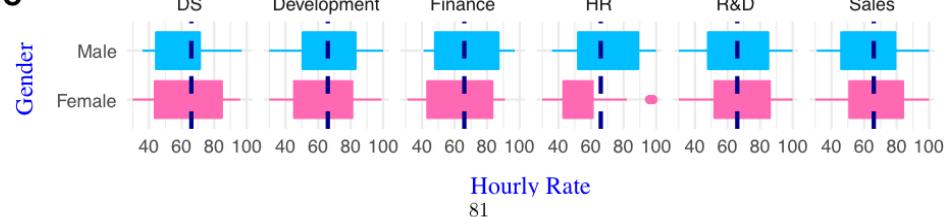
A



B



C



- As we can see it's not true that Females do make less money hourly in same position. It depends upon the Job Role.
- In some Job Role they seems to make more money hourly than Male as we can see that in Sales Executive Role, Research Scientist Role, Technical Architecture and many more.
- As we can see Females are paid less Hourly compare to Male after we grouped by Same job Role.
- As we can see it to be True in Human Resource mostly, Development, Finance.
- But incase of Sales that is not true that could be reason we have seen more Female in Sales Department than Men.
- In the field of Data Science Female tends to make more hourly than Male.
-

Let's group by EmpJobRole, TotalWorkExperienceInYears, ExperienceYearsInCurrentRole, EmpHourlyRate, Education level & Gender and see if this still holds true.

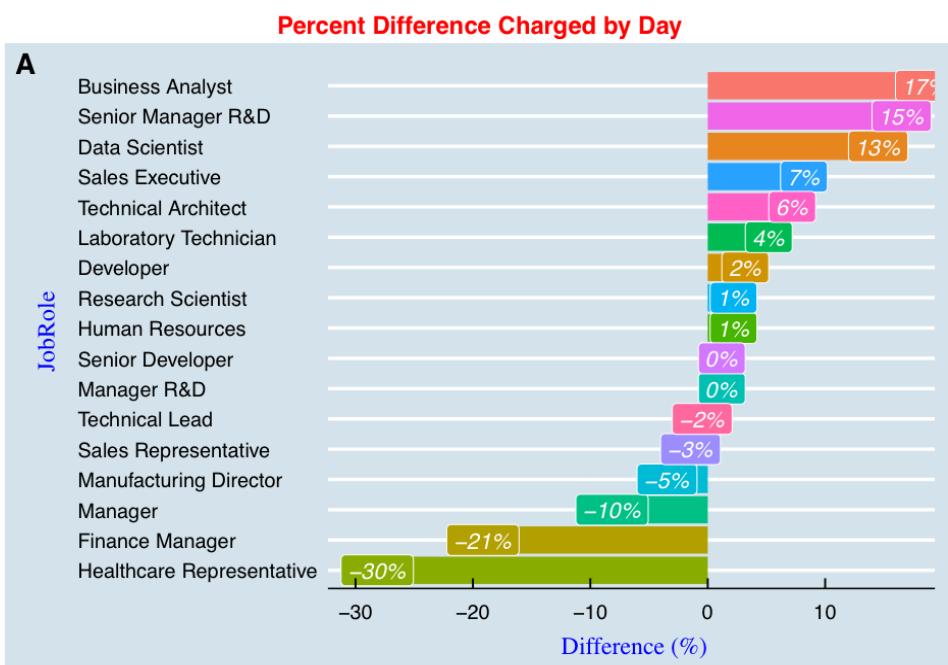
As we can see yes there is some wage disparities among male and Female although they have same or almost similar level of experience in certain roles. - It can be mostly seen in Development, Finance, HR department. - In HR there are some outliers as well in Female Population. - It is also true that Men are paid quite less in Sales, R&D department. - In DS field womens are paid quite more for same job role and experience than Male even after considering all the possible factors.

Education level & Attrition

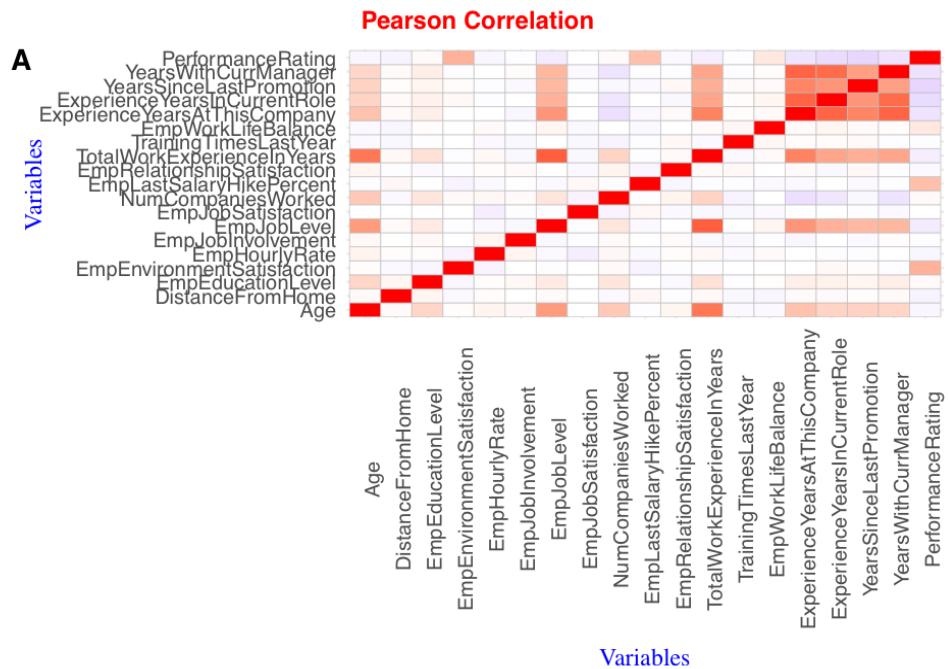
How does the Level of Education play role in leaving the company.

In terms of ranking who are more likely to leave are - 19% Workforce who doesn't have any college degree. - 15% Workforce who have only college degree - 15% Workforce who have only Bachelor degree - 13 % Workforce who have only Master Degree. - 12 % Workforce who have highest degree Phd. - When we break down by gender we can see Male population who have Master degree(76%) are more likely to leave and on Female side (~60%) who have Phd.

Percent Difference (%)



Correlational Matrix.



- Can see the pearson Correlation between the variables.