HR DATA ANALYSIS

--IBM HR DATASET

MSBA 304 Gurui Huang & Maoyi Song

INTRODUCTION

- General data description and data investigation on IBM HR dataset by using SQL in R and R for data visualization; suggestions are made based on data interpretation, inferences only, models and tests are needed for further conclusion or decision, which will be not included in this project.

TABLE OF CONTENT

HR DATA OVERALL

- The data "IBM HR. csv" is a sample of IBM human resource information data with 33 variables and 1470 observations:

```
names(hr)
```

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

HR DATA OVERALL

- Noticed that some variables scaled by number 1 to 5, let's discovering the meaning of all scales: 1 is the lowest level where 5 is the highest level.

- Data scales description

levels	Education	EnvironmentSatisfaction
1	Below College	Low
2	College	Medium
3	Bachelor	High
4	Master	Very High
5	Doctor	No such level

HR DATA OVERALL

- Continue

levels	JobInvolvement	JobSatisfaction	PerformanceRating
1	Low	Low	Low
2	Medium	Medium	Good
3	High	High	Excellent
4	Very High	Very High	Outstanding
5	No such level	No such level	No such level

HR DATA OVERALL

- Continue

levels	RelationshipSatisfaction	WorkLifeBalance
1	Low	Bad
2	Medium	Good
3	High	Better
4	Very High	Best
5	No such level	No such level

COMPANY ORGNIZATIONAL STRUCTURE

- **How many employee working for different department?**

Employee Distribution by Department
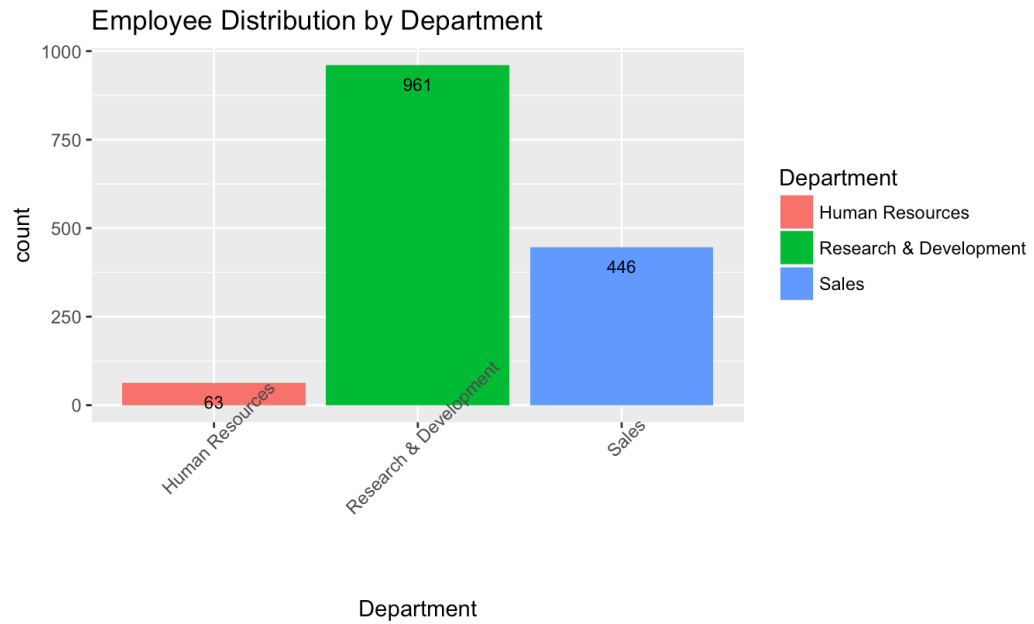
COMPANY ORGNIZATIONAL STRUCTURE

- Department Overview (continued)

```
##                        Department Employee_Count
## 1            Human Resources                63
## 2 Research & Development                    961
## 3                      Sales                446
```

- According to Santiago C.D, swswati L., and Gerard s. (2016), a typical but not operating model for product based and high-tech industry company, which most employees are in product team. However, this data sample may only contains only general framework of departments.

COMPANY ORGNIZATIONAL STRUCTURE

- Code: R for data visualization SQL for table

```
# plot 'Employee Distribution by Department'
ggplot(hr, aes(x = Department, fill = Department)) +
  geom_bar() +
  geom_text(stat = 'count', aes(label = ..count..), vjust = 2, size =
  ggtitle('Employee Distribution by Department') +
  theme(plot.margin=unit(rep(1,4),'lines'), axis.text.x = element_tex
```

```
# for employee count table:
# SELECT Department, COUNT(EmployeeNumber) AS Employee_Count
# FROM hr
# GROUP BY Department
```

EMPLOYMENT SATISFACTION

- Which role has the highest satisfaction level?

```
                    JobRole count(JobSatisfaction)
1 Healthcare Representative                      43
2            Human Resources                     13
3      Laboratory Technician                     80
4                    Manager                     33
5      Manufacturing Director                    38
6          Research Director                     22
7          Research Scientist                    95
8            Sales Executive                    112
9       Sales Representative                     23
```

- Which role has the lowest satisfaction level?

```
                    JobRole count(JobSatisfaction)
1 Healthcare Representative                      26
2            Human Resources                     10
3      Laboratory Technician                     56
4                    Manager                     21
5      Manufacturing Director                    26
6          Research Director                     15
7          Research Scientist                    54
8            Sales Executive                     69
9       Sales Representative                     12
```

EMPLOYMENT SATISFACTION

- Total Employee Number by Role

```
                    JobRole sum (EmployeeCount)
1 Healthcare Representative                131
2           Human Resources                52
3       Laboratory Technician               259
4                   Manager               102
5       Manufacturing Director              145
6           Research Director               80
7           Research Scientist              292
8             Sales Executive               326
9          Sales Representative             83
```

- Code:

- 1. select JobRole, count(JobSatisfaction) from HR where JobSatisfaction=4 group by JobRole

- 2. select JobRole, count(JobSatisfaction) from HR where JobSatisfaction=1 group by JobRole

- 3. select JobRole, sum (EmployeeCount) from HR group by JobRole

EMPLOYMENT SATISFACTION

Highest Satisfaction level 4

Job Role	Percent (%)
Healthcare Representative	32.83
Human Resources	25
Laboratory Technician	30.89
Manager	32.35
Manufacturing Director	26.21
Research Director	27.5
Research Scientist	32.53
Sales Executive	34.36
Sales Representative	27.71

Lowest Satisfaction level 1

Job Role	Percent (%)
Healthcare Representative	19.85
Human Resources	19.23
Laboratory Technician	21.62
Manager	20.59
Manufacturing Director	17.93
Research Director	18.75
Research Scientist	18.49
Sales Executive	21.17
Sales Representative	14.46

EMPLOYMENT SATISFACTION CONCLUSION

- Based on the result, we can conclude that the role has highest satisfaction level is Sales Executive (34.36 %), and followed by Healthcare Representative (32.83 %) and Research Scientist (32.53 %). The role has the lowest satisfaction level is Laboratory Technician (21.62 %), and followed by Sales Executive (21.17%) and Manager (20.59 %). Overall, the total number of employees who have the highest satisfaction is more than the employees who have the lowest satisfaction.

WORK LIFE BALANCE

- Which role has the highest level of WorkLifeBalance?

```
            JobRole count(WorkLifeBalance)
1 Healthcare Representative                 11
2            Human Resources                10
3       Laboratory Technician               25
4                    Manager                12
5      Manufacturing Director               14
6          Research Director                12
7          Research Scientist               24
8            Sales Executive                36
9       Sales Representative                 9
```

- Which role has the lowest level of WorkLifeBalance?

```
            JobRole count(WorkLifeBalance)
1 Healthcare Representative                 10
2            Human Resources                 4
3       Laboratory Technician               20
4                    Manager                 6
5      Manufacturing Director                7
6          Research Director                 4
7          Research Scientist               16
8            Sales Executive                12
9       Sales Representative                 1
```

WORK LIFE BALANCE

- Total Employee Number by Role (RECALL)

```
                         JobRole sum (EmployeeCount)
1 Healthcare Representative                    131
2            Human Resources                    52
3      Laboratory Technician                   259
4                    Manager                   102
5      Manufacturing Director                   145
6          Research Director                    80
7         Research Scientist                   292
8            Sales Executive                   326
9       Sales Representative                    83
```

- Code:

- 1.select JobRole, count(WorkLifeBalance) from HR where WorkLifeBalance=4 group by JobRole.

- 2. select JobRole, count(WorkLifeBalance) from HR where WorkLifeBalance=1 group by JobRole

WORK LIFE BALANCE

Highest level of WorkLifeBalance 4

Job Role	Percent (%) [count/total employees]
Healthcare Representative	8.40
Human Resources	19.23
Laboratory Technician	9.65
Manager	11.76
Manufacturing Director	9.66
Research Director	15.00
Research Scientist	8.22
Sales Executive	11.04
Sales Representative	10.84

Lowest level of WorkLifeBalance 1

Job Role	Percent (%) [count/total employees]
Healthcare Representative	7.63
Human Resources	7.69
Laboratory Technician	7.72
Manager	5.88
Manufacturing Director	4.83
Research Director	5.00
Research Scientist	5.48
Sales Executive	3.68
Sales Representative	1.20

WORK LIFE BALANCE CONCLUSION

- The role of HR has the highest work-life balance level (19.23%), followed by research director (15%) and manager (11.76%). However, the role of HR also has the lowest work-life balance level (7.69%), compared with others. Based on the percentage, there is a big difference between 19.23% and 7.69%, so we can say most of the HR employees feel a work-life balance. Others roles, such as healthcare representative, there is a small difference between the percentage of highest work-life balance level and lowest work-life balance level. So, it indicates that we can communicate with these employees and make an improvement.

EMPLOYEE PROMOTION

- select YearsAtCompany, YearsInCurrentRole from HR where YearsSinceLastPromotion > 10 and YearsInCurrentRole>10

- We got 26 observations where years since last promotion is more than 10 years ago and work in the current role also more than 10 years. Also, most of the employees work for company 20 years.

	YearsAtCompany	YearsInCurrentRole
1	22	15
2	17	14
3	13	12
4	20	18
5	24	13
6	33	18
7	22	17
8	20	16
9	13	12
10	18	13
11	18	16
12	17	11
13	31	15
14	17	13
15	12	11
16	17	13
17	26	15
18	20	11
19	19	16
20	12	11
21	18	15
22	23	15
23	22	14
24	22	11
25	17	14
26	15	12

GENERAL INFORMATION OF COMPANY'S SALARY

The Hourly Rate Distribution

Hourly Rate overview
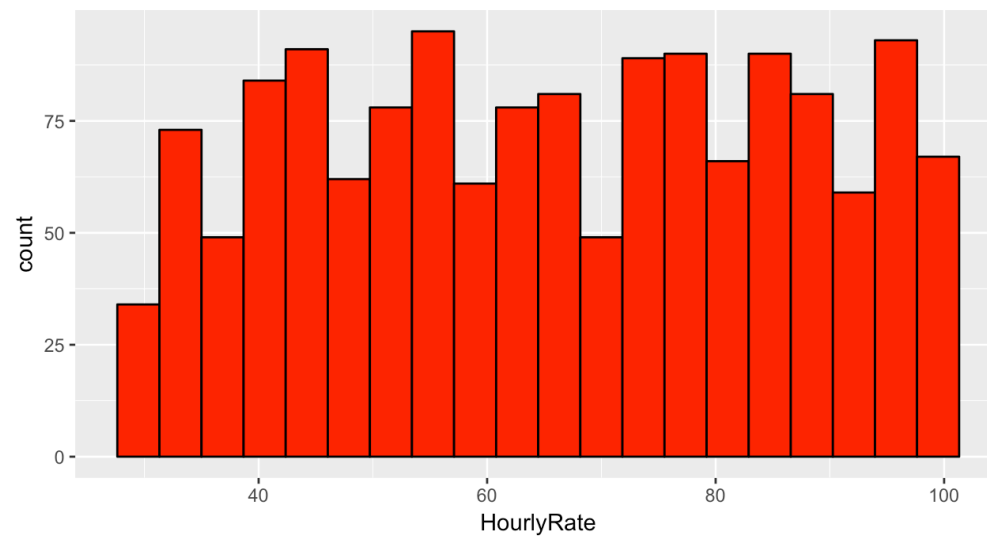
AVERAGE HOURLY RATE

- **The average hourly Rate by Department**

```
##                   Department Avg_HourlyRate
## 1       Human Resources        64.30159
## 2 Research & Development        66.16753
## 3                 Sales        65.52018
```

- **Insight**: Hourly rate are closed, since there is no location reference, it is not appropriate compare the results with industry level

- According to Glassdoor, all industry level average of HR is around $39.59/hr and $46.875/hr for product team. (estimated by average annual salary)

POTENTIAL RISK DETECTION FOR HR DEPARTMENT

- **What risk that HR department has?**

- Violation of Youth Rules & Violation of Equal Job Opportunity.

- Check if potential risk of violation of Youth Rules:

```
sqldf('SELECT EmployeeNumber
       FROM hr
       WHERE Age < 18')
```

```
## [1] EmployeeNumber
## <0 rows> (or 0-length row.names)
```

- Conclusion: NO employee under 18; no need worried about violation of youth rules.

POTENTIAL RISK DETECTION FOR HR DEPARTMENT

- Check if potential risk of violation of Equal Job

```
sqldf('SELECT Gender, COUNT(EmployeeNumber) AS Employee_Count
      FROM hr
      GROUP BY Gender')
```

```
##    Gender Employee_Count
## 1 Female            588
## 2   Male            882
```

- **Insight**: male-female ratio is imbalanced; according to US equal employment opportunity commission, the ratio of male-female employee itself could not be used as evidence against the company. however, for individual case, it can be used as supporting evidence.

- **Suggestion**: need pay attention on this area to avoid disputes in the future.

EMPLOYEES' BEHAVIOR TRACKING

- Two major metric: **Attrition** and **Performance**.

- **Attrition** Definition: "Employees who move, retire, pass away or leave the company to raise a family or attend school represent the usual ebb and flow of staffers through a business.(McQuerrey)"

- Also, when you talk about attrition in a company, there is a similar term called **employee turnover rate**, which refers employees who leave the company due to termination, taking a better job, or because they felt there was no room for growth, or worse, that they were dealing with a hostile or discriminatory work environment (McQuerrey).

ATTRITION & EMPLOYEE TURNOVER RATE

- "A **high turnover rate** typically means working conditions are not optimal, pay is below market average, or staffers are not well trained. Concurrently, a **low turnover rate** is indicative of a work environment where staffers feel appreciated, work as a team, have room to move up the corporate ladder, and are satisfied with their jobs (McQuerrey)."

ATTRITION & EMPLOYEE TURNOVER RATE

```
sqldf('SELECT COUNT(EmployeeNumber) AS EmployeeCount
    From hr
    GROUP BY Attrition')
```

```
##    EmployeeCount
## 1          1233
## 2           237
```

- **Insight**: Turnover rate in IBM company based on the data that gives us is

$$turnover = \frac{Attrition}{TotalEmployee} : 16.12\%;$$

- Based on the AON (2017), as comparison, the statistics shows that the average turnover in Software company is 22.4%, which higher than IBM.

- So, it indicates that employees in the company feel appreciated about the work environment and satisfied with their job.

PERFORMANCE RATING

- Performance rating: internal performance tracking, expected to be high for individual, it reflects the HR's management working quality.

- Code: select PerformanceRating, count(EmployeeNumber) from HR group by PerformanceRating

```
  PerformanceRating count(EmployeeNumber)
1                 3                  1244
2                 4                   226
```

- Conclusion: The performance rating only includes level 3 and 4, which indicate excellent and outstanding performance.

REFERENCES

- IBM HR data. Kaggle.com. *IBM HR Analytics Employee Attrition & Performance.* Retrieved from: https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

- Santiago C.D, Swati L., and Gerard S. (2016). *An operating model for company-wide agile development.* Retrieved from: https://www.mckinsey.com/business-functions

- Glassdoor. Glassdoor.com. *Salary.* Retrieved from: https://www.glassdoor.com/Salaries

- Youth Rules. Government website. Retrieved from: https://www.youthrules.gov/

REFERENCES

- Turn Over Rate. AON. Retrieved from: https://radford.aon.com/insights/infographics/2017/Technology/Q1-2017-Turnover-Rates-Hiring-Sentiment-by-Industry-at-US-Technology-Companies

- Equal Job Opportunity. U.S Equal Employment Opportunity Commission. Retrieved from: https://www.eeoc.gov/eeoc/newsroom/wysk/enforcement_protections_lgbt_workers.cfm

- McQuerrey, L. (2018, June 29). *Employee Turnover Vs. Attrition*. Retrieved from https://smallbusiness.chron.com/employee-turnover-vs-attrition-15846.html

- THANK YOU!