ATTRITION RATE ANALYSIS

GROUP – 7

OBJECTIVE:

To find the correlation and effect of different factors on the attrition rate of employees working in the healthcare industry.

Dependent variable: Attrition Rate

More about the dataset: The data is based on the IBM Watson employees for attrition and was captured in the year of 2021.

CHOOSING THE INDEPENDENT VARIABLES - I

```
> chisq.test(d$Attrition ,d$WorkLifeBalance)

        Pearson's Chi-squared test

data:  d$Attrition and d$WorkLifeBalance
X-squared = 25.063, df = 3, p-value = 1.498e-05

> chisq.test(d$Attrition,)
Error in sum(x) : invalid 'type' (character) of argument
> chisq.test(d$Attrition,d$EnvironmentSatisfaction)

        Pearson's Chi-squared test

data:  d$Attrition and d$EnvironmentSatisfaction
X-squared = 23.315, df = 3, p-value = 3.471e-05

> chisq.test(d$Attrition,d$Education)

        Pearson's Chi-squared test

data:  d$Attrition and d$Education
X-squared = 9.0625, df = 4, p-value = 0.05956

> chisq.test(d$Attrition,d$Gender)

        Pearson's Chi-squared test with Yates' continuity correc

data:  d$Attrition and d$Gender
X-squared = 0.59123, df = 1, p-value = 0.4419

> chisq.test(d$Attrition,d$MaritalStatus)

        Pearson's Chi-squared test

data:  d$Attrition and d$MaritalStatus
X-squared = 72.489, df = 2, p-value < 2.2e-16
```

We have a long list of variables (all categorical) to analyze but on conducting the chi square test of independence, we eliminated Gender, & Education Level because their p-value was greater than 0.05.

CHOOSING THE INDEPENDENT VARIABLES - II

```
Deviance Residuals:
    Min       1Q    Median       3Q       Max
-1.8145   -0.5041  -0.3149  -0.1546    3.6433

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                    5.62781    0.67202   8.375  < 2e-16 ***
data$JobLevel                 -0.62681    0.12705  -4.934 8.07e-07 ***
data$JobInvolvement           -0.73893    0.11093  -6.661 2.72e-11 ***
data$JobSatisfaction          -0.31389    0.07427  -4.226 2.38e-05 ***
data$Age                      -0.07582    0.01254  -6.046 1.49e-09 ***
data$BusinessTravel            0.58889    0.15606   3.773 0.000161 ***
data$WorkLifeBalance          -0.39800    0.11209  -3.551 0.000384 ***
data$EnvironmentSatisfaction  -0.32053    0.07474  -4.288 1.80e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.46  on 1675  degrees of freedom
Residual deviance:  980.51  on 1668  degrees of freedom
AIC: 996.51

Number of Fisher Scoring iterations: 6
```

- Finally, we build a logistic regression model for finalizing our set of IVs.

- ***P-value*** is less than 0.05 for all tested variables.

- So, Independent Variables - JobLevel, JobInvolvement, JobSatisfaction, Age, BusinessTravel, WorkLifeBalance, EnvironmentSatisfaction

WORK LIFE
BALANCE

Does poor work life balance result in higher attrition rate?

ENVIRONMENT
SATISFACTION

Does greater environment satisfaction result in lower attrition rate?

JOB LEVEL

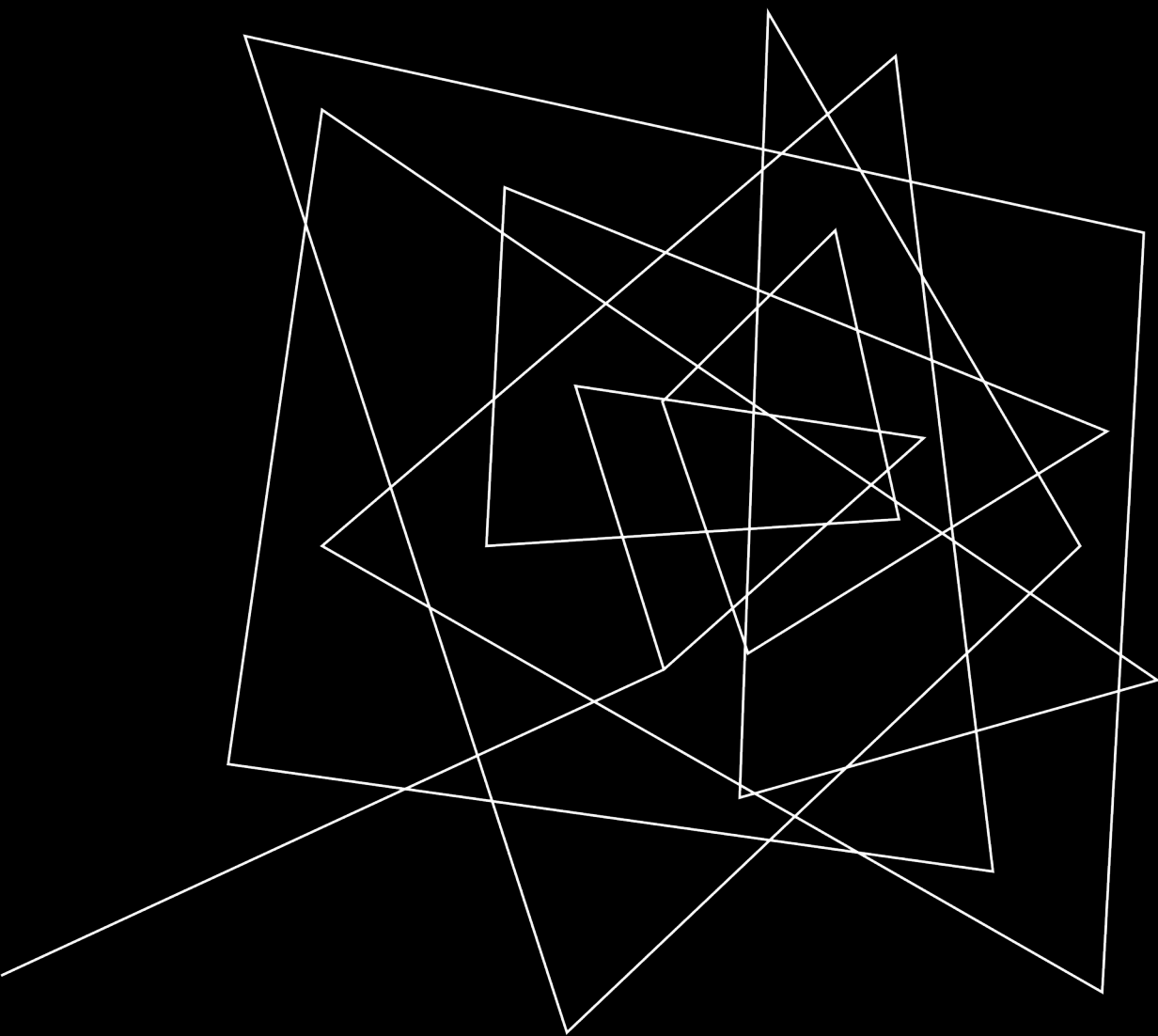Is there a lower attrition rate for a higher job level?

AGE + BUSINESS TRAVEL

Does age impact the attrition rate of employees depending upon their past travel records?

JOB INVOLVEMENT +
SATISFACTION

Do employees who are involved and satisfied in their job tend to switch companies faster?

QUESTIONS

WE PERFORM LOGISTIC REGRESSION FOR EVERY PREDICTOR VARIABLE(S) MENTIONED IN THE PREVIOUS SLIDE

1. WORK LIFE BALANCE

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7876  -0.5088  -0.4524  -0.4524   2.1588

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -1.0116     0.2384  -4.244 2.20e-05 ***
as.factor(WorkLifeBalance)2  -0.8015     0.2799  -2.863  0.00419 **
as.factor(WorkLifeBalance)3  -1.2163     0.2606  -4.668 3.04e-06 ***
as.factor(WorkLifeBalance)4  -0.9678     0.3332  -2.905  0.00368 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.5  on 1675  degrees of freedom
Residual deviance: 1200.5  on 1672  degrees of freedom
AIC: 1208.5

Number of Fisher Scoring iterations: 4
```

- Moving from a rating of 1 to 2 in Work Life Balance, the log of odds in attrition rate decreases by 0.801.

- Moving from a rating of 2 to 3 in Work Life Balance, the log of odds in attrition rate decreases by 1.216.

- Moving from a rating of 3 to 4 in Work Life Balance, the log of odds in attrition rate decreases by 0.967.

PROBABILITY OUTCOME OF ATTRITION RATE FOR A RATING VALUE OF WORK LIFE BALANCE:

Rating of 1 to 4 where:

1 = Unhealthy and 4 = Healthy

RATING	PROBABILITY
1	0.267
2	0.14
3	0.097
4	0.121

PREDICTION: With an increase in work-life balance for employees, there is a less chance of them quitting their firms.

2. ENVIRONMENT SATISFACTION

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6566  -0.4979  -0.4504  -0.4437   2.1758

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                         -1.4246     0.1392 -10.232  < 2e-16 ***
as.factor(EnvironmentSatisfaction)2 -0.6008     0.2217  -2.709  0.00674 **
as.factor(EnvironmentSatisfaction)3 -0.8441     0.2058  -4.101 4.12e-05 ***
as.factor(EnvironmentSatisfaction)4 -0.8126     0.2049  -3.967 7.29e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.5  on 1675  degrees of freedom
Residual deviance: 1200.3  on 1672  degrees of freedom
AIC: 1208.3

Number of Fisher Scoring iterations: 4
```

- Moving from a rating of 1 to 2 in Environment Satisfaction, the log of odds in attrition rate decreases by 0.6.

- Moving from a rating of 2 to 3 in Environment Satisfaction, the log of odds in attrition rate decreases by 0.844.

- Moving from a rating of 3 to 4 in Environment Satisfaction, the log of odds in attrition rate decreases by 0.812.

PROBABILITY OUTCOME OF ATTRITION RATE FOR A RATING VALUE OF ENVIRONMENT SATISFACTION:

Rating of 1 to 4 where:

1 = Poor and 4 = Excellent

RATING	PROBABILITY
1	0.193
2	0.116
3	0.093
4	0.096

PREDICTION: With a better work environment for employees, there is a less chance of them quitting their firms.

3. JOB LEVEL

```
Call:
glm(formula = as.factor(Attrition) ~ as.factor(d$JobLevel), family = binomial())

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-0.7206  -0.7206  -0.3398  -0.3398   2.7223

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)             -1.21587    0.09555 -12.725  < 2e-16 ***
as.factor(d$JobLevel)2  -1.60690    0.20072  -8.006 1.19e-15 ***
as.factor(d$JobLevel)3  -1.32310    0.26281  -5.034 4.79e-07 ***
as.factor(d$JobLevel)4  -2.46464    0.59233  -4.161 3.17e-05 ***
as.factor(d$JobLevel)5  -2.46043    0.72234  -3.406 0.000659 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.5  on 1675  degrees of freedom
Residual deviance: 1105.4  on 1671  degrees of freedom
AIC: 1115.4

Number of Fisher Scoring iterations: 6
```

- Moving from a rating of 1 to 2 in Job Level, the log of odds in attrition rate decreases by 1.606.

- Moving from a rating of 2 to 3 in Job Level, the log of odds in attrition rate decreases by 1.323.

- Moving from a rating of 3 to 4 in Job Level, the log of odds in attrition rate decreases by 2.464.

- Moving from a rating of 4 to 5 in Job Level, the log of odds in attrition rate decreases by 2.46.

PROBABILITY OUTCOME OF ATTRITION RATE FOR A RATING VALUE OF JOB LEVEL:

Rating of 1 to 5 where:

1 = Lower and 5 = Higher

RATING	PROBABILITY
1	0.228
2	0.056
3	0.073
4	0.024
5	0.024

PREDICTION: As depicted by the result, the highest chance of attrition is when the employee is at an entry-level position. And then it goes to a gradual decline with a slight increase in the medium level job roles.

4. AGE + BUSINESS TRAVEL

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6582  -0.5139  -0.5139  -0.2833   2.5438

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.95793    0.09369 -20.898  < 2e-16 ***
as.factor(ABT)2    0.53853    0.18072   2.980  0.00288 **
as.factor(ABT)3   -1.23734    0.30899  -4.004 6.21e-05 ***
as.factor(ABT)4   -0.30384    0.47918  -0.634  0.52603
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.5  on 1675  degrees of freedom
Residual deviance: 1184.2  on 1672  degrees of freedom
AIC: 1192.2

Number of Fisher Scoring iterations: 5
```

- Moving from a rating of 1 to 2 in Age+Travel, the log of odds in attrition rate increases by 0.53.

- Moving from a rating of 2 to 3 in Age+Travel, the log of odds in attrition rate decreases by 1.23.

- Moving from a rating of 3 to 4 in Age+Travel, the log of odds in attrition rate decreases by 0.303.

PROBABILITY OUTCOME OF ATTRITION RATE FOR A CONJOINED VALUE OF BOTH AGE AND TRAVEL FREQUENCY:

Rating of 1 to 4 where:

1 = <45 travelling rarely or not at all

2 = <45 travelling frequently

3 = 45+ travelling rarely or not at all

4 = 45+ travelling frequently

RATING	PROBABILITY
1	0.123
2	0.194
3	0.039
4	0.094

PREDICTION: After checking for preference to travel for work across different age groups, we arrived on the conclusion that employees who are made to travel have a higher attrition rate.

5. JOB INVOLVEMENT + SATISFACTION

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7100  -0.6069  -0.3822  -0.3822   2.3036

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -2.5802     0.1467 -17.591  < 2e-16 ***
as.factor(JSI)2    0.5110     0.2110   2.423   0.0154 *
as.factor(JSI)3    0.9817     0.2083   4.714 2.43e-06 ***
as.factor(JSI)4    1.3306     0.2238   5.944 2.77e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1221.5  on 1675  degrees of freedom
Residual deviance: 1179.0  on 1672  degrees of freedom
AIC: 1187

Number of Fisher Scoring iterations: 5
```

- Moving from a rating of 1 to 2 in Satisfaction+Involvement, the log of odds in attrition rate increases by 0.511.

- Moving from a rating of 2 to 3 in Satisfaction+Involvement, the log of odds in attrition rate increases by 0.981.

- Moving from a rating of 3 to 4 in Satisfaction+Involvement, the log of odds in attrition rate increases by 1.33.

PROBABILITY OUTCOME OF ATTRITION RATE FOR A CONJOINED VALUE OF BOTH AGE AND TRAVEL FREQUENCY:

Rating of 1 to 4 where:

1 = High job involvement & high job satisfaction

2 = High job involvement & low job satisfaction

3 = Low job involvement & high job satisfaction

4 = Low job involvement & low job satisfaction

RATING	PROBABILITY
1	0.070
2	0.112
3	0.168
4	0.222

PREDICTION: High levels of Job involvement and satisfaction when put together can exponentially reduce the attrition rate of an employee according to the probability analysis done above.

THANK YOU

Adhira Ranjithkumar

Asmita Samanta

Mingxi Xu

Abdul Shaik