Employee Attrition and Performance

Group 6

Angela Tseng (Technical Analyst)
Harshitha Ramachandra (Technical Analyst)
Sakshi Patil (Subject Matter Analyst)
Srikanth Parvathala (Project Manager)

Research Question

Unit of Analysis

Individuals: employee

Independent Variable

Monthly Income

Does monthly income have an impact on employee job attrition?

Tools

- Logistic Regression
- Descriptive statistics
- Histogram

Dependent Variable

Attrition: Yes/No

IBM HR Analysis Employee Attrition and Performance

IBM HR Employee Attrition Database

35 attributes
1470 records

01

Attrition

Categorical ordinal

02

Monthly Income

Numerical continuous

03

Job Satisfaction

Categorical ordinal

04

Work Life Balance

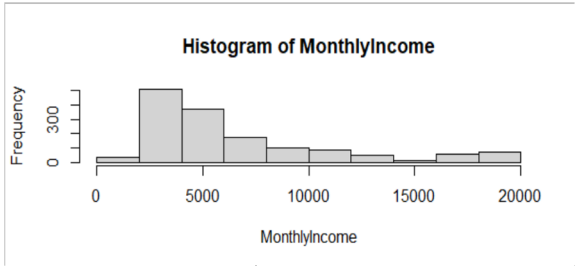
Categorical ordinal

Analysis

```
> levels(df$Attrition)
[1] "Yes" "No"
```

Monthly income:

Min.	1st Qu.	Median	**Mean**	3rd Qu.	Max.
1009	2911	8379	4919	19999	**6503**



Dependent variable is categorical with two levels and independent variable is numeric(continuous).

Descriptive Statistics

Histogram

Logistic Regression

Attrition

Statistical analysis

01

Null Hypothesis

Ho: b1=0
There is no significant relationship between Monthly Income and Attrition.

02

Alternative Hypothesis

Ha : b1 != 0
There is a significant relationship between Monthly Income and Attrition.

Monthly income has a significant effect on Job Attrition

03

Logistic Regression

```
fit <- glm(df$Attrition ~ df$MonthlyIncome, family=binomial())
```

Coefficients

	Estimate	Std.Error	z value	Pr(>|z|)
(Intercept)	-9.291e-01	1.292e-01	-7.191	6.43e-13
df$MonthlyIncome	-1.271e-04	2.162e-05	-5.879	4.12e-09

Interpretation of Result

$$ln(\frac{prob(event)}{1-prob(event)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$

where
bo= -0.929
b1= 0.000127

01

The p value associated with coefficient obtained is 4.12e-09 which is less than the significance level(0.05), hence based on our p value we reject the null hypothesis.

02

This implies that monthly income has a significant effect on Job Attrition.

03

Each one unit change in monthly income will decrease the log odds of determining attrition by 0.000127.

Impact of other Independent Variables

Chi Square Test of Independence for "JobSatisfaction" and "Attrition"

H0
There is no relationship between Job Satisfaction and Attrition

HA
There is a relationship between Job Satisfaction and Attrition

Output
X-squared = 17.505, df = 3, p-value = 0.0005563

Interpretation
Reject our null hypothesis: we conclude that there is a significant relationship between Job Satisfaction and Attrition.

01

02

03

04

Chi Square Test of Independence for "WorkLifeBalance" and "Attrition"

H0
There is no relationship between WorkLifeBalance and Attrition

HA
There is a relationship between WorkLifeBalance and Attrition

Output
X-squared = 16.325, df = 3, p-value = 0.0009726

Interpretation
Reject our null hypothesis: we conclude that there is a significant relationship between WorkLifeBalance and Attrition.

Further Analysis – Logistical Regression

	Estimate	Std. Error	z value	Error Pr(>|z|)
(Intercept)	3.769e-01	3.026e-01	1.245	0.21300
df$MonthlyIncome	-1.280e-04	2.176e-05	-5.884	4.01e-09 ***
df$WorkLifeBalance2	-7.615e-01	2.896e-01	-2.630	0.00855 **
df$WorkLifeBalance3	-9.994e-01	2.678e-01	-3.732	0.00019 ***
df$WorkLifeBalance4	-6.967e-01	3.307e-01	-2.107	0.03511*
df$JobSatisfaction2	-4.535e-01	2.191e-01	-2.070	0.03844*
df$JobSatisfaction3	-4.283e-01	1.946e-01	-2.201	0.02771 *
df$JobSatisfaction4	-8.910e-01	2.080e-01	-4.284	1.83e-05 ***

glm(formula = df$Attrition ~ df$MonthlyIncome + df$WorkLifeBalance + df$JobSatisfaction, family = binomial())

From the below result, it is observed that all the variables have a p value less than .05. Hence we reject the null hypothesis that the coefficient is equal to zero. **Monthly Income, Work life balance and Job satisfaction has a significant impact on the attrition.**

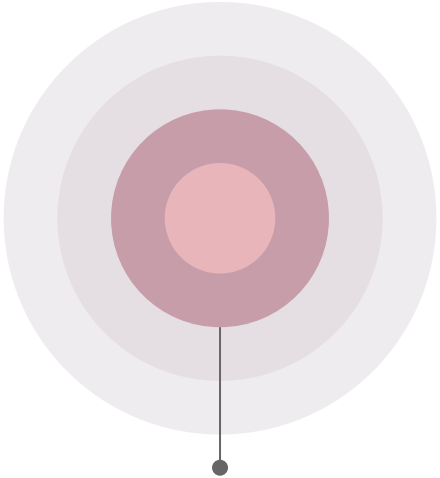
Further Analysis

	Estimate	Std. Error	z value	Error Pr(>|z|)
(Intercept)	3.769e-01	3.026e-01	1.245	0.21300
df$MonthlyIncome	-1.280e-04	2.176e-05	-5.884	4.01e-09 ***
df$WorkLifeBalance2	-7.615e-01	2.896e-01	-2.630	0.00855 **
df$WorkLifeBalance3	-9.994e-01	2.678e-01	-3.732	0.00019 ***
df$WorkLifeBalance4	-6.967e-01	3.307e-01	-2.107	0.03511*
df$JobSatisfaction2	-4.535e-01	2.191e-01	-2.070	0.03844*
df$JobSatisfaction3	-4.283e-01	1.946e-01	-2.201	0.02771 *
df$JobSatisfaction4	-8.910e-01	2.080e-01	-4.284	1.83e-05 ***

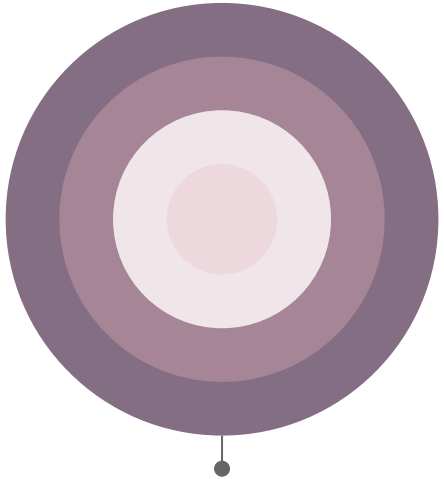
glm(formula = df$Attrition ~ df$MonthlyIncome + df$WorkLifeBalance + df$JobSatisfaction, family = binomial())

From the below result, it is observed that all the variables have a p value less than .05. Hence we reject the null hypothesis that the coefficient is equal to zero. **Monthly Income, Work life balance and Job satisfaction has a significant impact on the attrition.**

Limitations



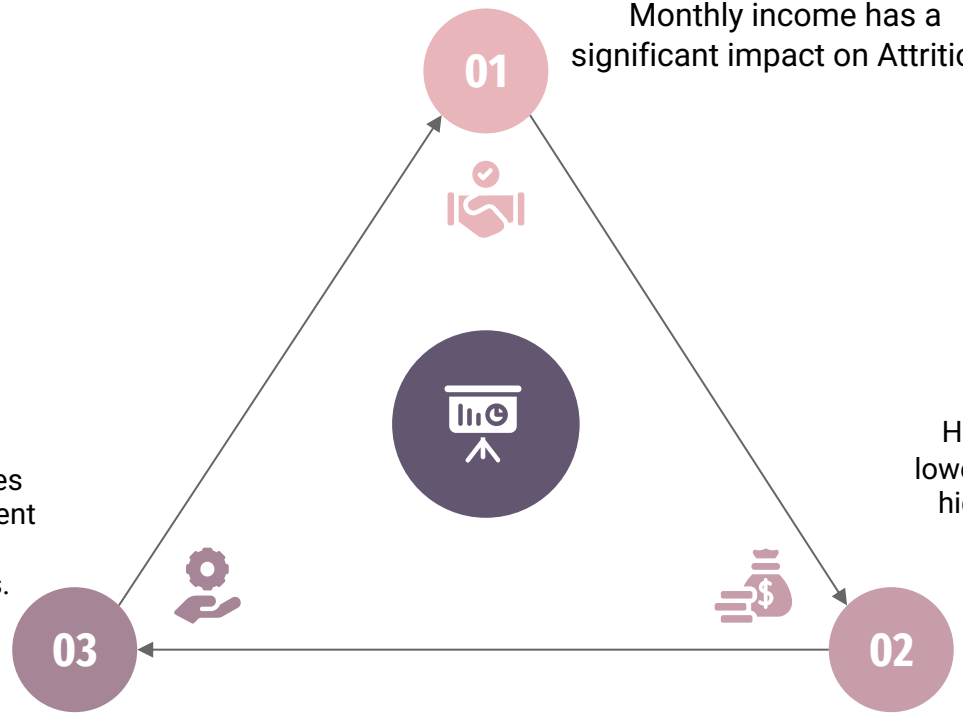
Fictional data set created by IBM data scientists



Dataset doesn’t consist of details of the company that employees joined after leaving current job.

This data will help us to analysis further regarding outsourcing or poaching which is also one of the main factor for attrition in IT companies.

Conclusion



Monthly income has a significant impact on Attrition.

Higher the Job Satisfaction lower the attrition rate, similarly higher the Work life balance lower the attrition rate.

For future analysis one can consider the bottom up approach, with other attributes from the dataset as independent variables to determine how probability of attrition varies.

THANK YOU!