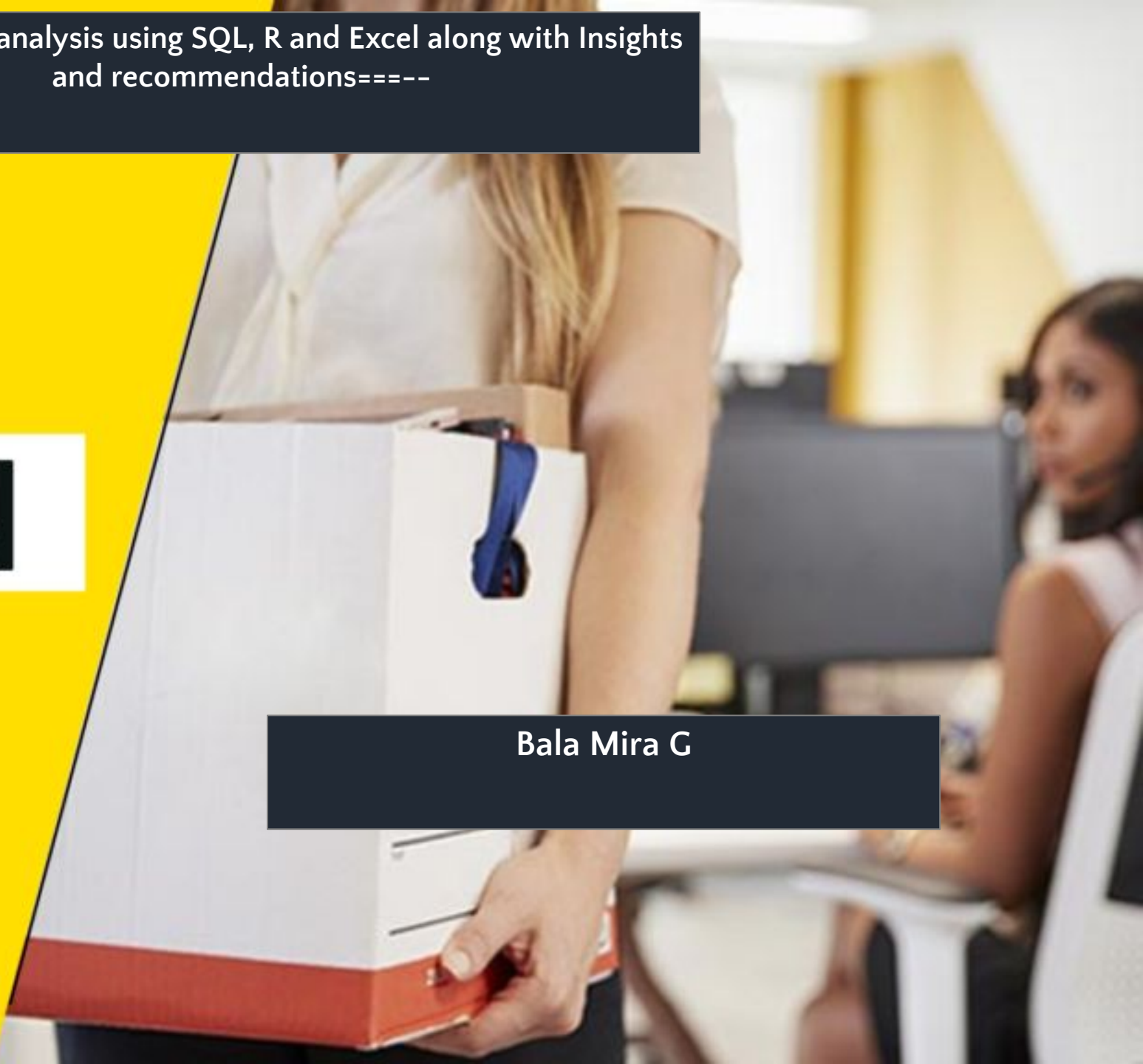


-----Attrition analysis using SQL, R and Excel along with Insights
and recommendations-----

ATTRITION

Bala Mira C



-----Insights-----

-----Gender attrition % ----

```
with cte as (select * from hr where attrition="Yes")
SELECT
  gender,
  round((COUNT(*) * 100 / (SELECT COUNT(*) FROM cte)),0)AS
gender_percentage
FROM cte
GROUP BY
gender order by gender_percentage desc;
```

	Gender	gender_percenta...
	Male	63
	Female	37

Insights and Recommendations:

Males have a higher attrition % than females.

-----Attrition by Gender and Marital status-----

```
SELECT `Marital status`,gender,count(*) as totalattrition
FROM
  hr
WHERE
  attrition = 'Yes' group by `Marital status`,gender order by totalattrition desc;
```

	Marital status	gender	totalattrition
	Single	Male	147
	Married	Male	109
	Single	Female	95
	Married	Female	65
	Divorced	Male	55
	Divorced	Female	20

Insights and Recommendations:

Single Men and Married men are the highest and second highest no of people under attrition.

----- Attrition % and Avg salary by Department and Education level -----

	Department	attrition_percent...
	R&D	57
	Sales	38
	HR	5

	Department	avgincome
	Sales	5920
	R&D	4147
	HR	3716

Education	attrition_percent...
Bachelor's Degree	42
Master's Degree	24
Associates Degree	19
High School	13
Doctoral Degree	2

Education	avgincome
Doctoral Degree	5850
Master's Degree	5493
Bachelor's Degree	4759
High School	4327
Associates Degree	4195

Insights and Recommendations:

- The attrition rate in HR department is very low (even though the avg income is the lowest) when compared to R&D and Sales departments
- Steps which are followed in the HR dept can be implemented to improve the attrition% in the R&D department since their attrition % is more than 50%.
- The bachelor's degree holders form the highest attrition % since young undergraduate professionals always look for further studies and career opportunities to enrich their career.
- The doctoral degree holders have a lower attrition % thus attributing to their high income level.
- The High schoolers might not have enough financial resources to pursue higher studies or are unable to get opportunities from other companies and hence the low attrition.

----- Attrition % and Avg salary by Extended working hours -----

	Over Time	attrition_percent...
	Yes	53
	No	47

```
with cte as (select * from hr where
attrition="Yes")
SELECT
  `Over time`,
  round(avg(`Monthly Income`),0) AS
avgincome
FROM cte
GROUP BY
  `Over time` order by avgincome desc;
```

	Over Time	avgincome
	No	4946
	Yes	4663

```
with cte as (select * from hr where
attrition="Yes")
SELECT
  `Over time`,
  round(((COUNT(*) * 100 / (SELECT COUNT(*)
FROM cte )) ,0) AS attrition_percentage
FROM cte
GROUP BY
  `Over time` order by attrition_percentage desc;
```

Insights and Recommendations:

- The attrition rate % for people working overtime are more than the people who do not work overtime.
- Further to this, it may be due to the fact that they are being paid less when compared to people who do not work over time.
- Steps can be taken to decrease the workload for people working extended time and also increase their average income.

----- Attrition % and Avg salary by Extended working hours -----

	Job Satisfaction	attrition_percentage
	3	31
	1	27
	4	23
	2	19

```
with cte as (select * from hr where attrition="Yes")
SELECT
  `Job Satisfaction`,
  round((COUNT(*) * 100 / (SELECT COUNT(*) FROM cte)),0) AS attrition_percentage
FROM cte
GROUP BY
  `Job Satisfaction` order by attrition_percentage desc;
```

Insights and Recommendations:

- The attrition rate % for people having a lower job satisfaction is around 27% which is second highest.
- Steps can be taken to increase the satisfaction score for the employees.

----- Correlation between Age and Attrition % -----

	Age	Attrition_percentage
Age	1	
Attrition_percentage	-0.473783974	1

Insights and Recommendations:

- To perform the correlation, I used Excel's Correl function.
- A negative correlation coefficient indicates that as one value increases the other one decreases.
- As age increases, the attrition rate decreases and vice versa.
- Steps can be taken to reduce the attrition level at lower age groups by providing ample career and growth opportunities within the company.

----- Correlation between Performance Rating and Attrition % -----

	Performance Rati...	attrition_percentage
	3	84
	4	16

	Performance Rating	Attrition_percentage
Performance Rating	1	
Attrition_percentage	-1	1

Insights and Recommendations:

- The data suggests a strong negative correlation between Performance Rating and Attrition Percentage.
- High Performance Ratings are associated with lower Attrition Percentages, and low Performance Ratings are associated with higher Attrition Percentages.
- Company should focus on maintaining a performance-driven culture, as higher performance seems to be linked with lower attrition. Consider strategies and initiatives to improve and recognize employee performance.

----- Correlation between Percent Salary Hike and Attrition % -----

	Percent Salary Hike	attrition_percentage
	11	18
	13	15
	12	14
	14	10
	15	7
	17	6
	16	6
	22	5
	18	5
	19	4
	20	3
	21	3
	23	2
	24	2

	Percent Salary Hike	Attrition_percentage
Percent Salary Hike	1	
Attrition_percentage	-0.9153563	1

Insights and Recommendations:

- The correlation coefficient of -0.915 indicates a very strong relationship between these two variables.
- As Percent Salary Hike increases, Attrition Percentage tends to decrease, and vice versa.
- Attrition can be reduced by offering salary hikes to employees and also by performing Competitor Benchmarking. This can help in retaining top talent.
- Implement more employee engagement and retention programs to enhance growth within the company.

----- Correlation between Years at Company and Attrition using R %

=====

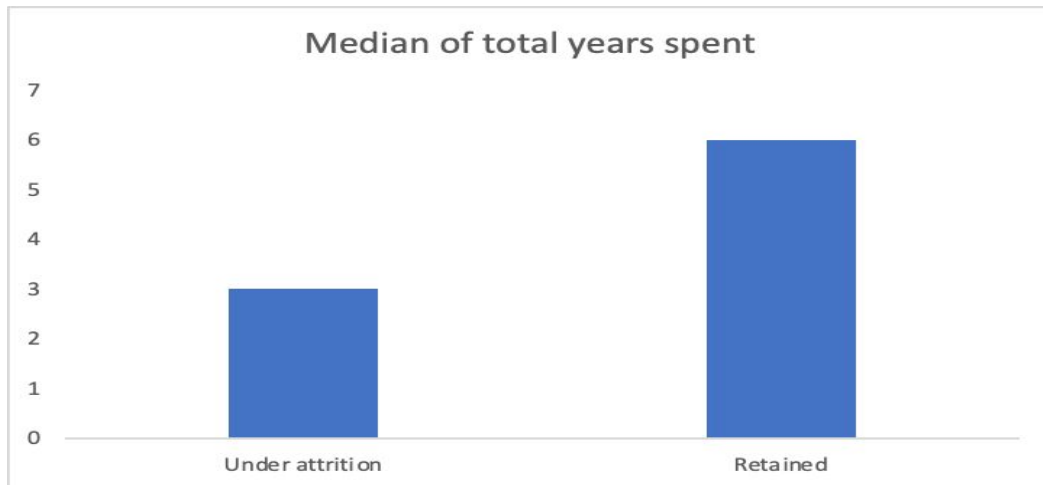
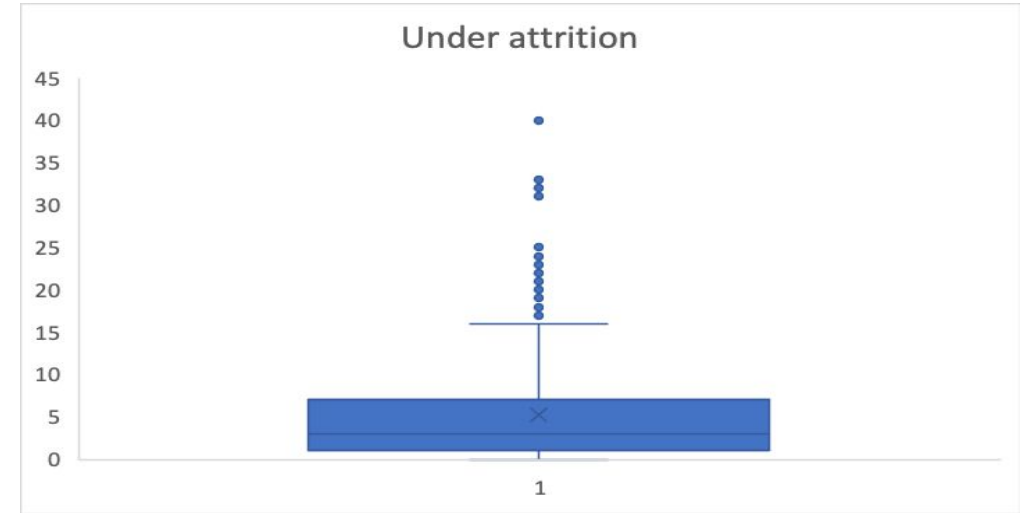
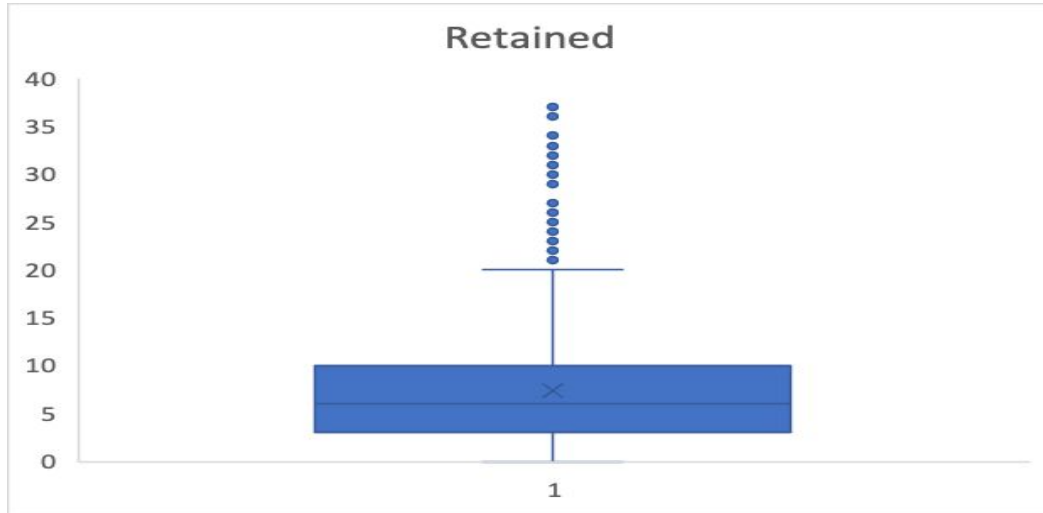
```
> # Create the data
> years_at_company <- c(40, 33, 32, 31, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10,
9, 8, 7, 6, 5, 4, 3, 2, 1, 0)
> attrition_percentage <- c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 8, 4, 4, 5, 4, 9, 8,
9, 11, 24, 7)
>
> # Calculate correlation coefficient
> correlation_result <- cor(years_at_company, attrition_percentage)
>
> # Print the result
> print(correlation_result)
[1] -0.6822342
```

	Years At Company	attrition_percentage
	40	0
	33	0
	32	0
	31	0
	25	0
	24	0
	23	0
	22	0
	21	0
	20	0
	19	0
	18	0
	17	0
	16	0
	15	0
	14	0
	13	0
	12	0
	11	0
	10	0
	9	0
	8	0
	7	0
	6	0
	5	0
	4	0
	3	0
	2	0
	1	0
	0	0

Insights and Recommendations:

- The R code creates two vectors representing the tenure of employees and attrition_percentage. The result is stored in a variable called Correlation_result
- The correlation between the number of years an employee has been in their current role and the attrition percentage is -0.6822342.
- A negative correlation of suggests that as the number of years in the company increases, the attrition percentage tends to decrease.
- However correlation does not imply causation. This can be further analyzed by doing statistical tests to examine the relationship between these variables.

-----Median of Total years spent in the company-----



- We can further our analysis by replacing age with total years at company.
- The median tenure of people under attrition and not under attrition can be plotted in a Box and a Whisker plot which shows the median tenure of non - attrited people is greater than that of attrited ones.
- This statement can be statistically proved using Wilcoxon Rank sum test which would be followed in the next slide.

-----Wilcoxon rank sum test using R for Tenure-----

- The R code creates two vectors representing the tenure of employees under attrition and those not under attrition.
- It then uses the `wilcox.test` function in R to perform the Wilcoxon Rank-Sum test.
- The result is stored in the variable `wilcox_test_result`, and you can print it to see the test statistics.
- Since the p-value is very small, you have evidence to reject the null hypothesis. This indicates that there is a statistically significant difference between the paired samples. In the context of your data, it suggests that there is a significant location shift between the "Yes" and "No" groups.
- Given the rejection of the null hypothesis, we have enough statistical evidence suggesting a significant difference in median tenure between those who were attrited and those who were not. This implies that tenure may have played a role in the selection for attrition.

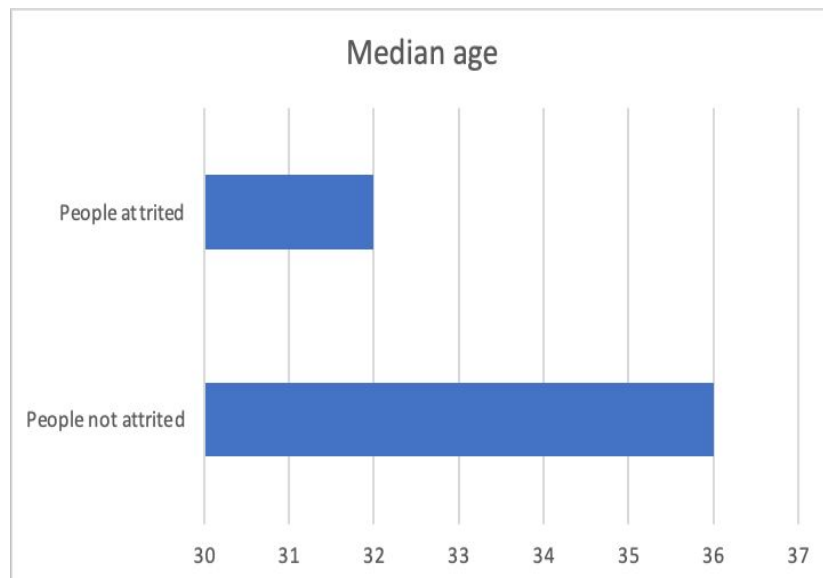
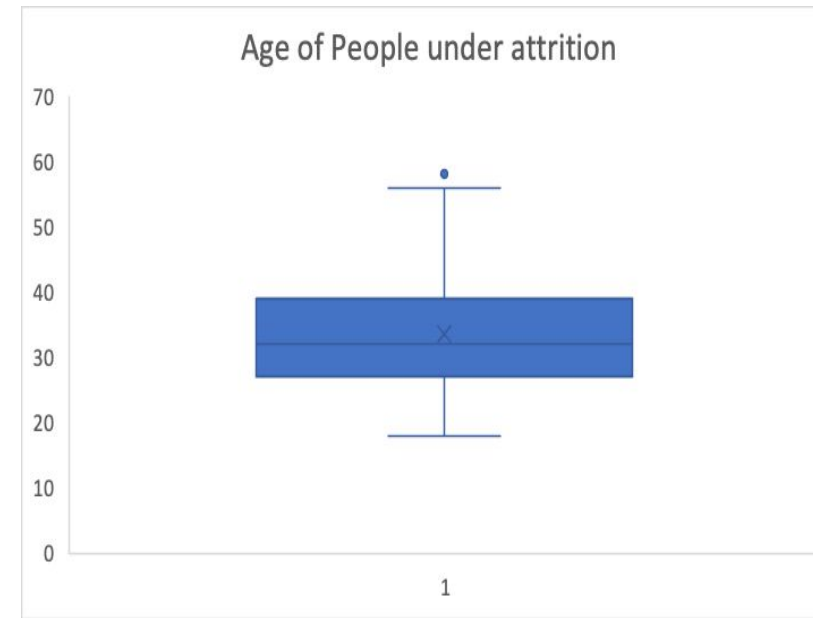
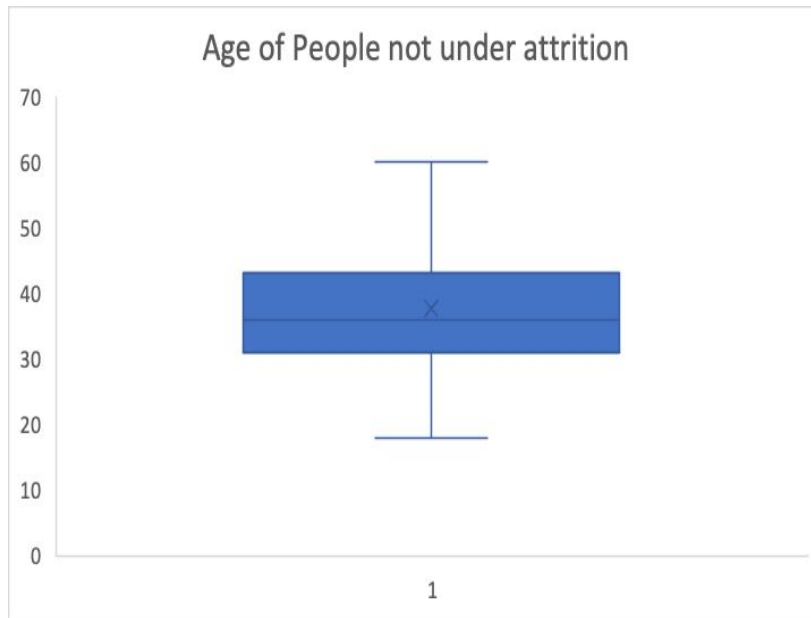
data: Yes and No

$W = 99363$, $p\text{-value} = 1.365e-07$

alternative hypothesis: true location shift is not equal to 0



-----Median Age analysis-----



- The problem statement is that ages of people under attrition is less than that of people not under attrition. This is due to the fact that the company hired many young employees and fresh graduates to manage the high workload during the hiring surge.
- The median age of people under attrition and not under attrition can be plotted in a Box and a Whisker plot which shows the median age of non - attrited people is greater than that of attrited ones.
- This statement can be statistically proved using Wilcoxon Rank sum test which would be followed in the next slide.

-----Wilcoxon rank sum test for age analysis using R-----

- The R code creates two vectors representing the ages of employees under attrition and those not under attrition.
- It then uses the `wilcox.test` function in R to perform the Wilcoxon Rank-Sum test.
- The result is stored in the variable `wilcox_test_result`, and you can print it to see the test statistics.
- The result of the Wilcoxon Rank-Sum test indicates a p-value of 0.006004, which is less than the commonly used significance level of 0.05. With such a small p-value, you would typically reject the null hypothesis, suggesting that there is a significant difference in the median ages between the two groups. The null hypothesis stated that there is no difference in the median ages between the two groups.
- Given the rejection of the null hypothesis, we have enough statistical evidence suggesting a significant difference in median ages between those who were attrited and those who were not. This implies that age may have played a role in the selection for attrition.

```
> wilcox_test_result <- wilcox.test(attritionyes, attritionno)
> print(wilcox_test_result)
```

Wilcoxon rank sum test with continuity correction

data: attritionyes and attritionno

W = 7892.5, p-value = 0.006004

alternative hypothesis: true location shift is not equal to 0

Key Findings:

- This might be due to the fact that the company due to its cost cutting exercise had to let go of many young employees they hired recently due to the hiring boom.
- Moreover the company wanted to retain only older employees who hold key positions within the company.

Recommendations:

Strategic Workforce Planning:

- Develop a strategic workforce plan that considers the optimal composition of the workforce based on business requirements.
- Align hiring and attrition strategies with long-term organizational goals.
- Company can also ensure that it makes optimal use of its current workforce and engage in a minimal need based hiring .