

Exploratory Data Analysis (EDA) on Attrition Dataset

Rounak Kumar
Aswin R
Andrew George Cherian

The attrition dataset gives us an insight into the different factors affecting the attrition rates of employees in the given company. The dataset is given in the form of a csv file which is analyzed using python.

A few rough insights that have been observed from the dataset are:

The attrition rate at the company stands at 16.12% and a total of 237 out of 1470 employees have left the company. The attrition variable forms the target variable for our analysis.

The gender ratio of the company stands at 3:2 with 60% males and 40% females.

The mean age of the company stands at 36 which shows that the company has a relatively young workforce.

The list of predictors affecting attrition in the dataset that will be used by us in our analysis are:

- Age
- Attrition
- Environment Satisfaction
- Job Involvement
- Job Level
- Job Satisfaction
- Monthly Income
- Total Working Years
- Work Life Balance
- Years At Company
- Years In Current Role
- Years Since Last Promotion
- Years With Current Manager
- Distance From Home

Univariate Analysis

Analysis of Categorical Variables

1.Business Travel

Business travel is defined as the travel commuted by employees for work related purposes.

1043 out of 1470 employees or 70.95% of the total employees travel rarely.

We can observe from this that under our analysis in our observations about the data, a majority of the employees work from the office setup.

2.Department

Department is defined as the different sectors of focus within a company working on one or more key areas.

961 out of 1470 employees or 65.37% of the total employees work in the R&D department.

We can observe from this that under our analysis in our observations about the data that the company has its maximum workforce in R&D.

3.Gender

Gender diversity is defined as the no. Of employees of different sexes working in an organization.

882 of the total 1470 employees or 60% of the employees are males and the gender ratio stands at 3:2.

We observe from this that under our analysis in our observations about the data that the workforce of the company is dominated by males.

4.Marital Status

Marital status is defined as the status of person divided in the categories of married ,unmarried and divorced.

673 of the total 1470 or 45.7% of the employees are married .

We observe from this that under our analysis in our observations about the data that majority of employees the company are married.

Analysis on discrete variables

Environment satisfaction

It is defined as a variable that measures the depth of content feeling in the company among employees.

61.16% of the total employees or 899 of the total 1470 employees gave a rating of 3 and above on a scale of 4 giving a good outlook about Environment satisfaction in the company.

Work Life Balance

Work Life Balance is defined as the division of one's time and focus between working and family or leisure activities.

In our observations under our analysis, we can say that 60.75% employees have work life balance of 3 on a scale of 1-4 forming a majority of the employees that have positive outlook about Worklife balance.

Job Involvement

Job Involvement refers to the psychological and emotional extent to which someone participates in his/her work, profession, and company.

68.85% of the total employees were rated 3 and above on a scale of 4 therefore giving a fairly positive impression about the employees of the company

Job level

Job levels, also known as “job grades” or “classifications,” are categories with different titles and salary ranges within a workplace.

36.4% of the employees were on the job level of 1 forming the largest portion of employees at a single level. 12 % of the employees were on the level of 4 and above. This shows that majority of the employees are working at junior level in the organization.

Job Satisfaction

Job satisfaction is defined as the level of contentment employees feel with their job.

31.22% employees have job satisfaction of scale 4 on a scale of 1 to 4. 30.07% employees have job satisfaction of scale 3 on a scale of 1 to 4. giving a positive outlook about the Job satisfaction about Job satisfaction among employees in the company

How does Attrition change with Age

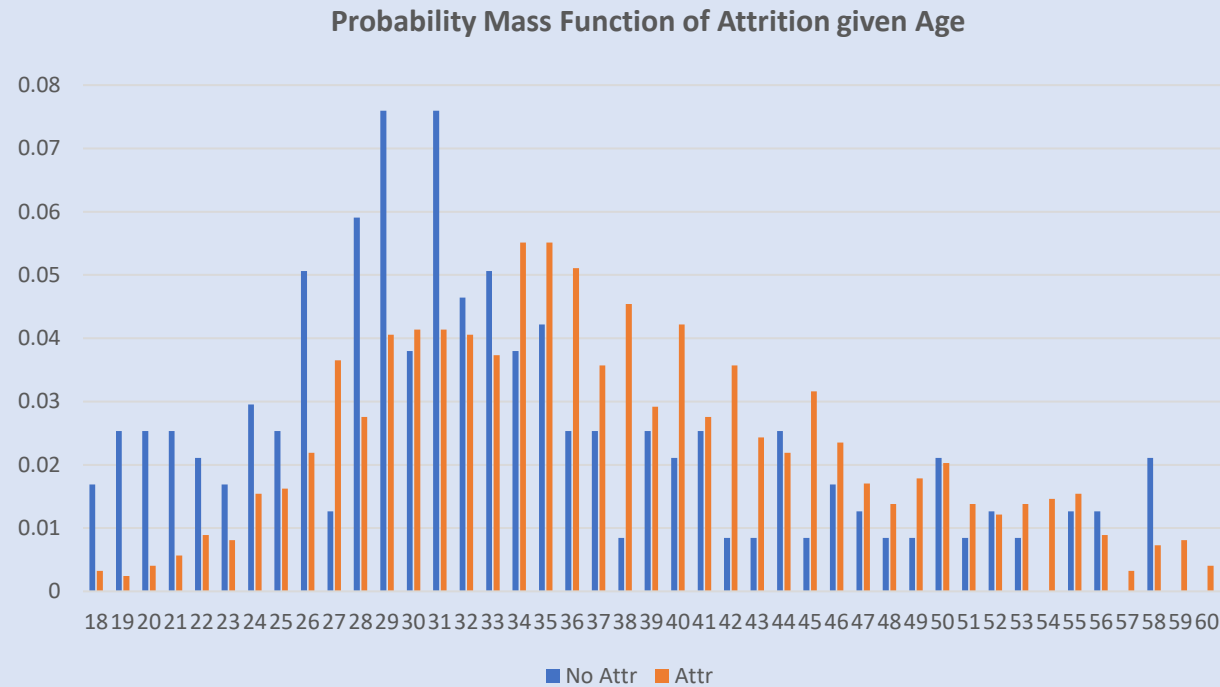


Fig 1

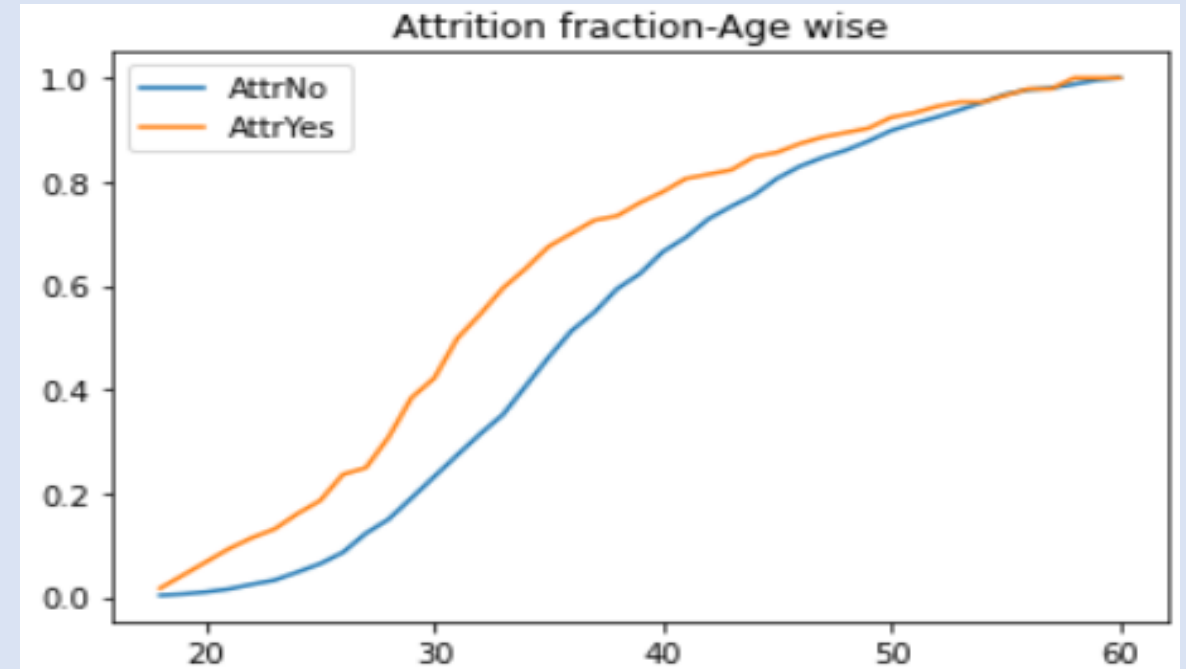


Fig 2

- On a closer observation of age of employees with attrition it was observed that the probability of attrition roughly increases till the age of ~35 after which it is observed to drop. (Fig 1)
- While observing the cumulative attrition fraction below an age group, it is observed that fraction of employees leaving the company is greater than fraction of existing employees. (Fig 2)

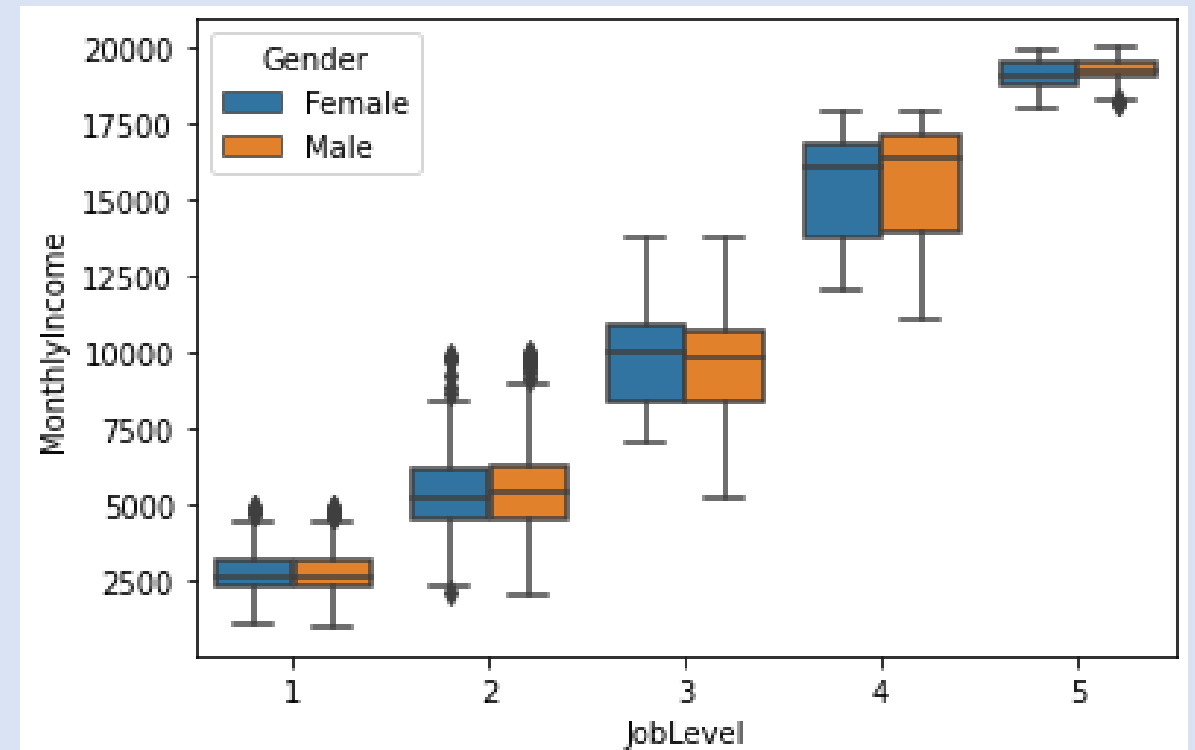
The difference in height between the lines increase till mid-30s and the gap reduces, which confirms the insight we obtained from previous PMF distribution.

Is Attrition more in Males?

Is there a difference in pay
between male employees and
Female Employees?

- By carrying out a preliminary analysis, we can see that the probability of Male given attrition is roughly 63%.
- To check whether there is any disparity in pay among Male employees and female employees, we can see from the box plot given below that the median income for each job level is roughly the same for both male and female employees. This gives us an indication that there is no disparity in pay due to gender.

Attrition	Yes (%)	No (%)
Gender		
Female	37	41
Male	63	59



How does Attrition change with Income Level and Over Time?

- From the above graphs, it can be observed that out of the 237 people who left the company under our analysis, the proportion of employees who worked overtime was roughly equal.
- The employees under our analysis did not have get paid for the overtime they were putting into their work.
- Given Attrition the income level of the employees under our analysis was lower.
- Thus, it is intuitive to expect the probability of attrition given lower salary range and over time to rise as seen in Fig 4.

Fig 1

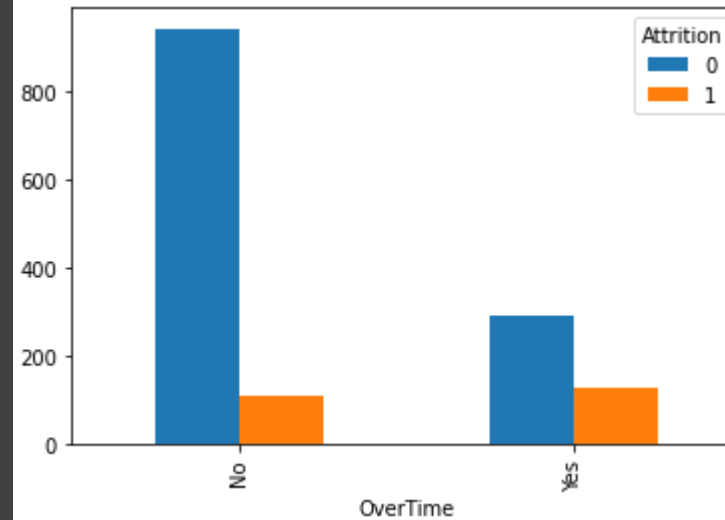


Fig 2

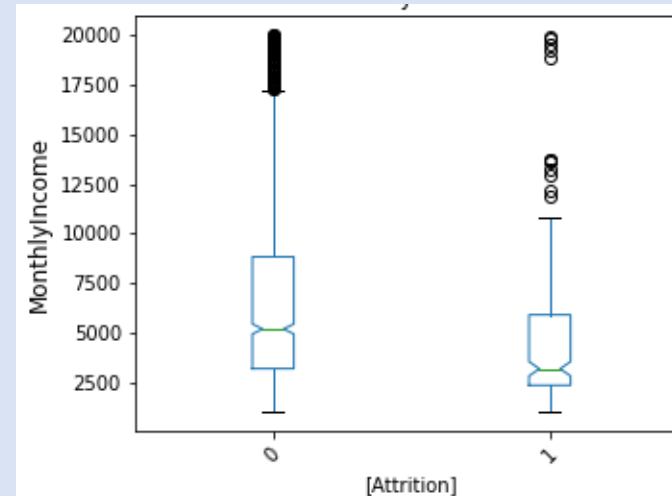
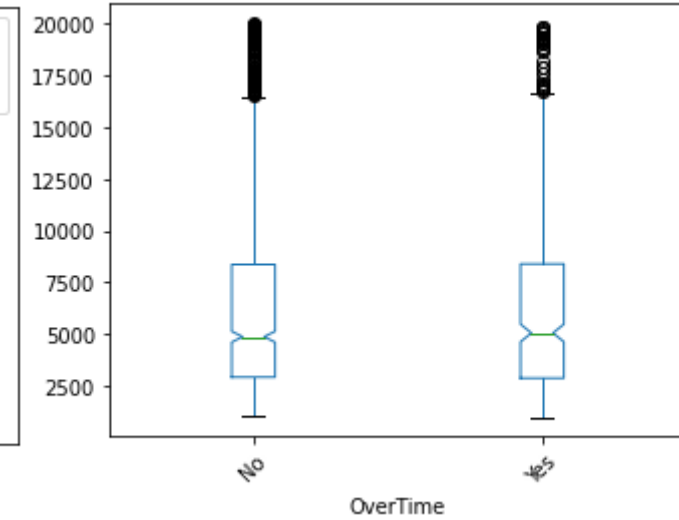


Fig 3

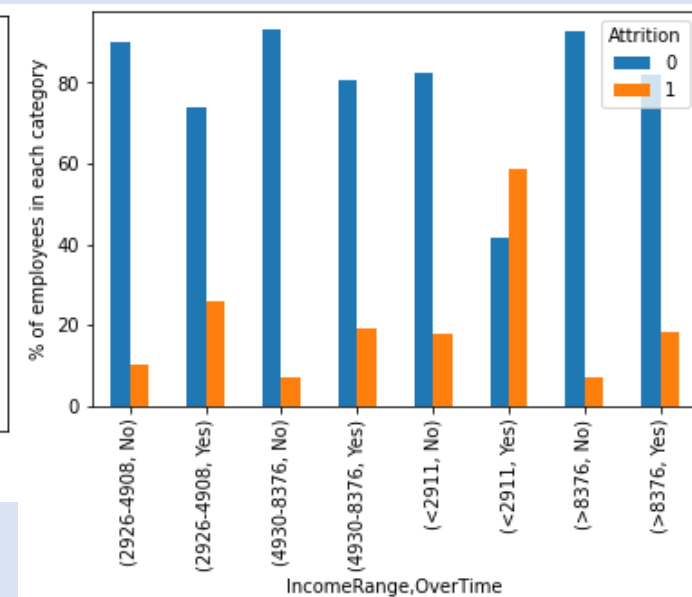
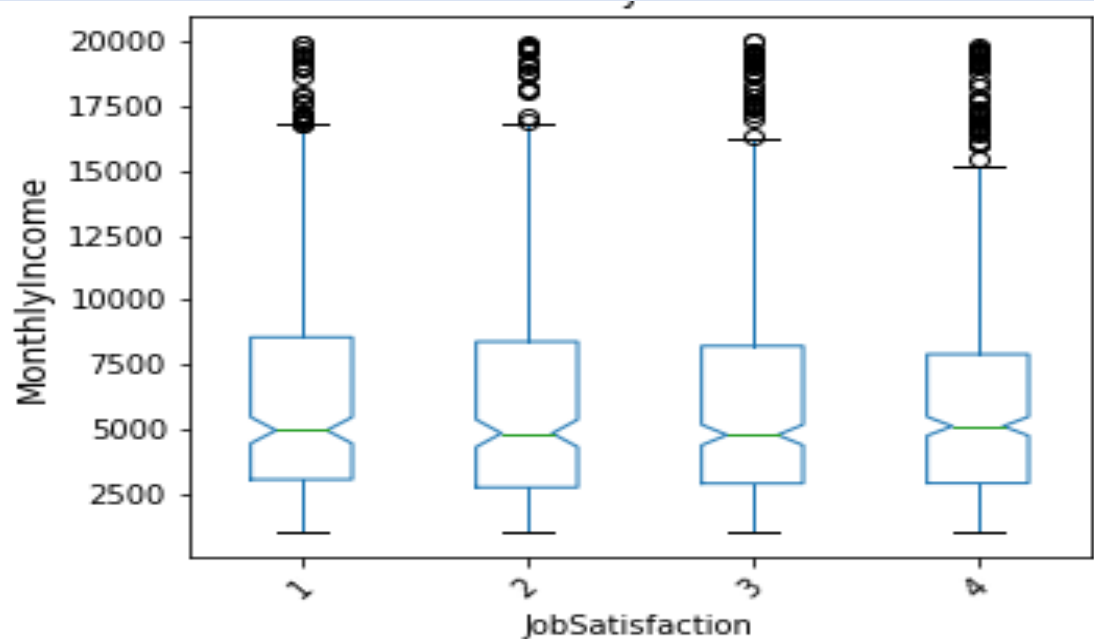


Fig 4

Attrition	No(%)	Yes (%)
Job Satisfaction		
1	77	23
2	83	17
3	84	16
4	89	11

Attrition	No(%)	Yes (%)
Monthly Income		
Less than Rs.2911	71	29
2926-4908	86	14
4930-8376	89	11
Greater Than Rs 8376	90	10



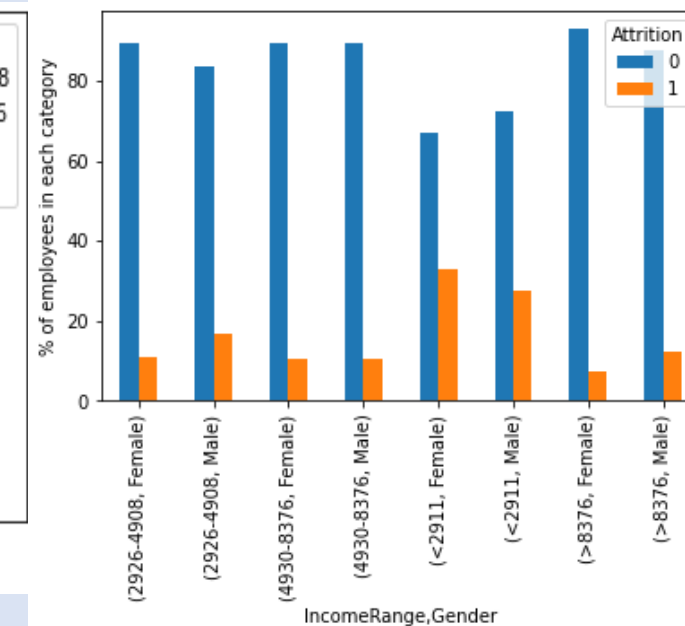
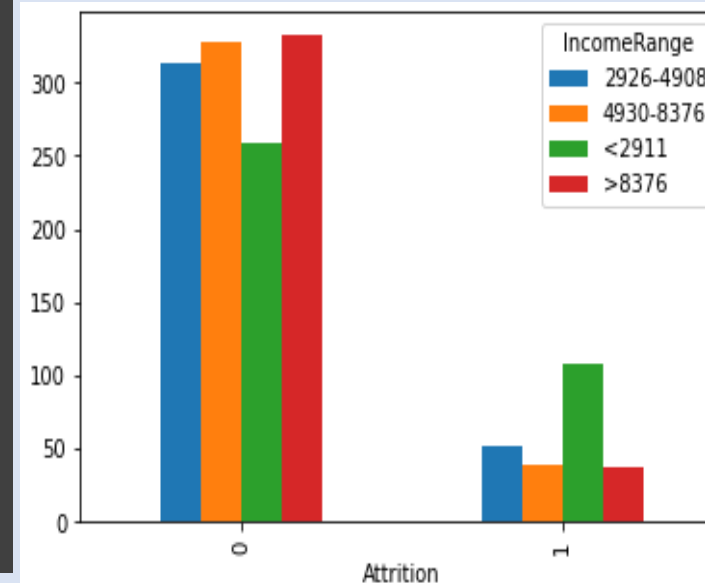
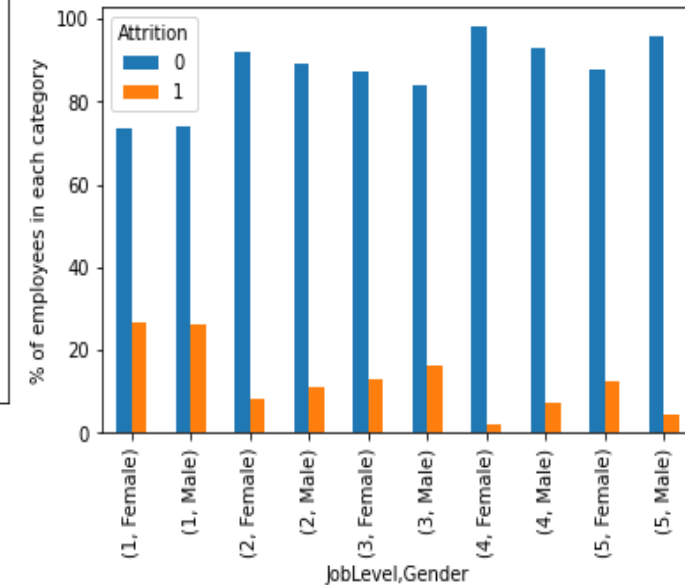
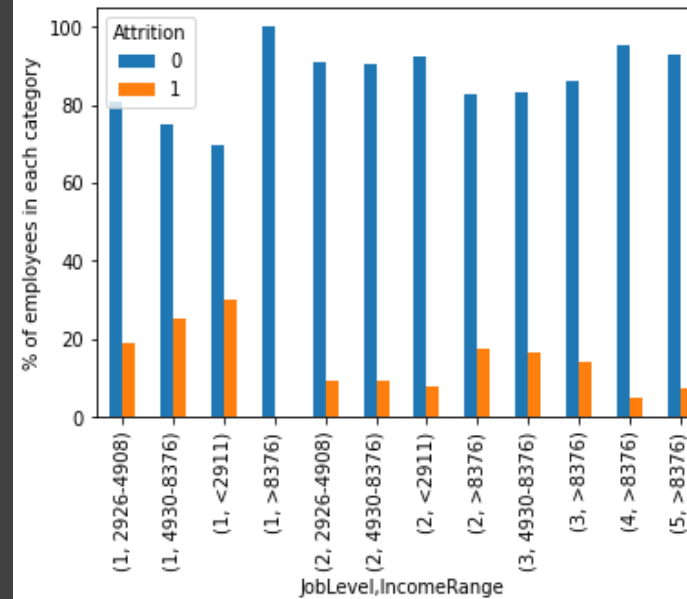
Is monthly Income affecting Job Satisfaction?

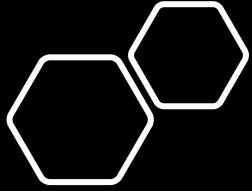
- From our analysis we observed that Attrition percentage among people who has rated 1 for their job satisfaction on a scale of 4 is 23% and highest among all other levels of job satisfaction.
- Similarly we segregated the employees into 4 income ranges (4 quartiles of the income), we observed roughly 29% attrition among employees who had monthly income below first quartile.

But when observing closely from the boxplot, it showed that median and IQR of monthly income is roughly the same for all levels of Job satisfaction. **Thus there is no association between Income and Job satisfaction**

How does Attrition change with Income Level, Job Level and Gender?

- From the above graphs, it can be observed that out of the 237 people who left the company under our analysis, roughly 100 employees had income less than INR 2911.
- Among the employees attrition is highest given job level is 1 and income less than INR 2911
- While examining the effect of Gender on Attrition with Job Level and Income Range, it is observed that Gender does not have a large difference in the attrition.





How attrition affects LATERAL and OLD Employees with Age and Job level

- We have categorized the employees as OLD employees and Lateral Employees based on the fact (if Total Experience is not equal Years at company, the employee is categorized as a Lateral Employee)
- From Fig 3 and 4 among the employees under our analysis, the fraction of Lateral employees is increasing with age among the existing employees as well as among the employees who have resigned.
- The variation of job level is as expected in an organization(from Fig 5), Median age in each job level is increasing as indicated by the correlation coefficient.
- From Fig 1 it is seen that the proportion of lateral employees in the company is increasing with job level and more than 80% of the employees in the top management is comprised of Lateral employees.
- From Fig 2 it is seen that out of all the employees who have resigned in our analysis having job level higher than 3, 100% of the employees are Lateral employees.
- This indicates that overall loyalty of employees towards the company increases after crossing a job level 3, and the loyalty of OLD employees of the company is 100% in such situation.

Job Level Wise - Lateral Employees
Among Existing Employees

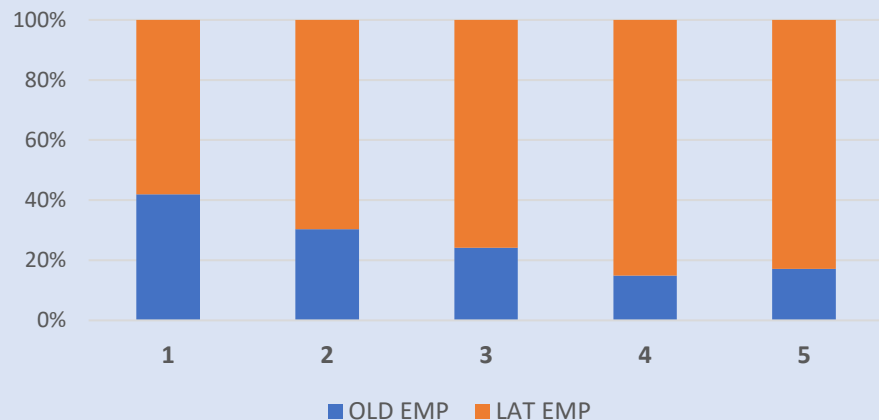


Fig 1

Job Level Wise - Lateral Employees
Among Resigned Employees



Fig 2

Fraction of Lateral Joinees Among Resigned
employees

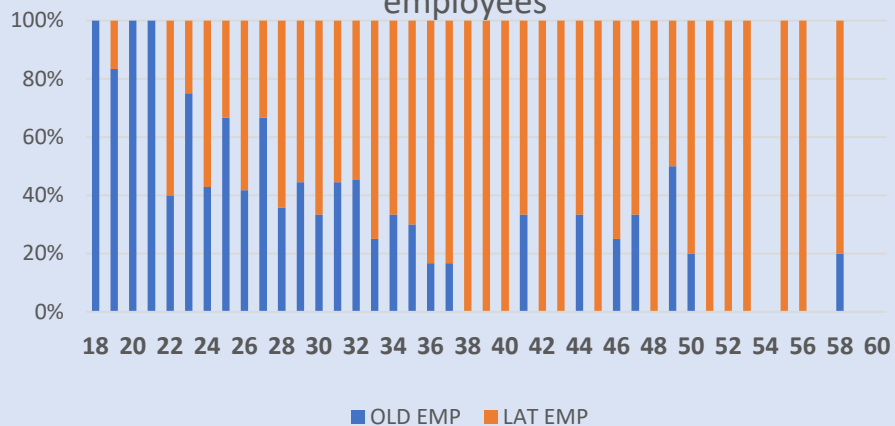


Fig 3

Fraction of Lateral Joinees Among Existing
employees

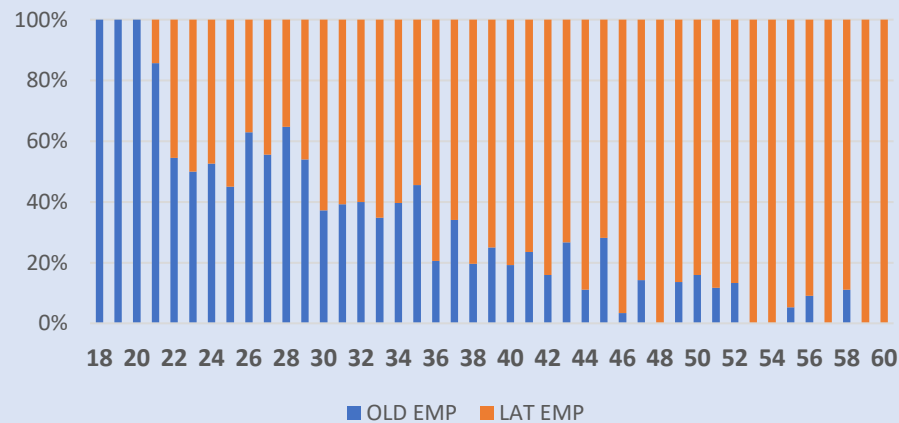


Fig 4

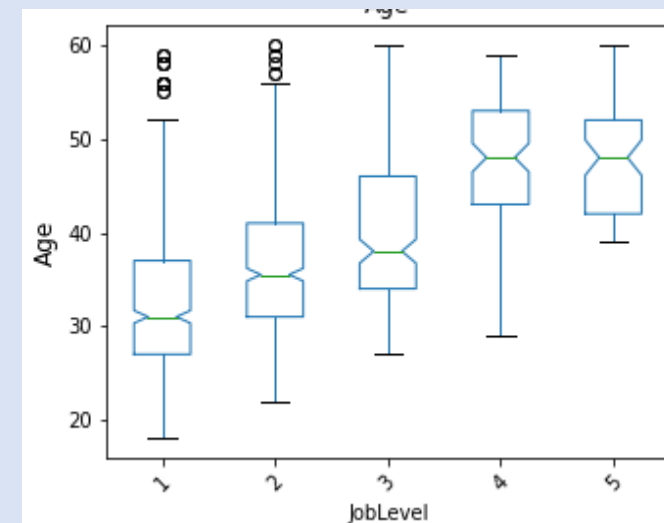


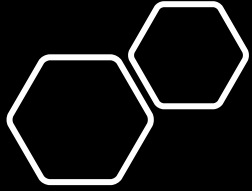
Fig 5

	Business Travel	Department	Education Field	Environment Satisfaction	Job Involvement	Job Level	Job Satisfaction	Marital Status	Income Range	Over Time	First Letter Represents
BusinessTravel			FT			F1		FS	F<2911	FY	Travel Frequently
Education Field				M1		M1		TS	T<2911	MY	M - Marketing T - Technical
JobInvolvement		1S		11					1<2911	1Y	Job Involvement 1
JobLevel	1F	1S		11			11	1S		1Y	Job Level 1
JobSatisfaction				11		11		1S	1<2911	2Y	Job Satisfaction 1
MaritalStatus	SF	SS				S1	S1		S<2911	SY	Single
IncomeRange	<2911 F		<2911 M	<2911 1			<2911 1	<2911 S		<2911 Y	Income < 2911
OverTime				Y1	Y1	Y1		YS		Y<2911	Yes
WorkLifeBalance				11	11	11	11	1S	1<2911	1Y	Work Life Balance 1
Second Letter Represents	Travel Frequently	Sales	Marketing	Environment Satisfaction 1	Job Involvement 1	Job Level 1	Job Satisfaction 1	Single	Income < 2911	Yes	

How to read the table?

An XY in the table represents: Among all the people in group sub group X, people belonging to Y sub group have much higher attrition percentage (>40) when comparing with other sub groups along Y

For example, FT represents: Among all the people who travel frequently, people with Education ‘Technical’ has much higher attrition percentage when comparing with other education fields



Conclusions

- **Age:** It is highly associated with attrition; younger employees tend to leave the company more. It may be for better growth opportunity and career growth. Also once assigned with higher and better responsibilities, employees tend to avoid leaving the company.
- **Job Satisfaction :** People with low job satisfaction tend to leave the company more than employees with higher job satisfaction
- **Monthly Income:** Employees with lower monthly income was observed to leave more frequently. It may be argued that employees with lower income may be the same employees with low job level or less age.
- **Lateral Employees** tend to join more in the age range of 20~35. They tend to leave more in the age range above ~36.
- **Overtime:** Employees who tend to work overtime without any additional pay tend to leave the company more.
- Thus the company may want to act in increasing job satisfaction of Lateral Employees in order to reduce attrition among Lateral employees
- Similarly act to increase job satisfaction and monthly salary of employees in lower job levels to curb the attrition rate.