Faculty of Engineering
Computer Engineering Department

WHY EMPLOYEES QUIT?

A Research That could change the way you see the Employment Statistics

By:

Abnormal distribution Team

Supervisor:

DR/ Maha Amin Hassanein

Team Members:

Name:	Sec:	Code:
Kareem Mohamed Rashed	2	9220600
Kareem Mahmoud Ahmed	2	9220601
Kareem Wael Mohamed	2	9220602
Mohamed Abd El Rahman El Khiat	2	9211015
Mostafa Mohamed Rabie	2	9211196
Nour Khaled Ahmed	2	9220921

Contents

Introduction

In today's dynamic workplace landscape, understanding and anticipating factors influencing employee retention are crucial for organizational success. Through advanced analytics and machine learning, our project aims to unveil patterns and indicators contributing to turnover, providing valuable insights for proactive talent management and fostering a more resilient and engaged workforce. Join us on this journey to harness the power of data to enhance employee retention strategies and optimize organizational performance.

The act of an employee quitting their job can have a negative impact on the workplace, reducing efficiency and productivity. This study aims to investigate the various factors that contribute to employees leaving their jobs. These factors can range from external reasons, such as a negative workplace environment that can lead to an employee feeling undervalued and unappreciated, to personal reasons, such as work-life imbalances or the desire to seek a different career path. Understanding these reasons is crucial for employers so that they can take appropriate measures to create a better workplace environment for their employees. To achieve this, we have gathered relevant data sets from Kaggle that contain information about employees. By studying these data sets, we aim to form a model that accurately predicts whether an employee is likely to quit or not

Methodology

1. Take an overview of our data (check for null, duplicates... and discover the central tendency and variability of features).
2. Analyze each feature independently and with other features:
   - Divide features into technical features and personal ones.
   - Formulate our hypothesis that (Null (H_0): Any feature has no effect on attrition and each group of data is independent of others in whole features, Alternative (H_a): those features are dependent and affect each other).
   - At each feature we use the appropriate statistical test to check our hypothesis:
   - **T_Test** for two groups of numerical groups.
     ANOVA if more than two groups.
     Post Hoc to know exactly any pair of groups are significantly independent (if needed).
     Chi square $\chi 2$ test for two categorical variables.
   - Utilize visual tools like histograms, box plots, scatter plots, and pie charts to uncover patterns or connections between this features and attrition or other features also according to our testing needs.
   - Then we incorporate our observations based on these tests and visualizations.
3. Build machine learning model that predict likelihood of turnover.

Technical Reasons

Department Analysis:

Is Department has an effect on the decision of quit?

To answer this question we will do Chi square test.

Null Hypothesis H_0: Department has no effect on the decision of quit.
Alternative Hypothesis H_a: Department has a significant effect on the decision of quit.

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

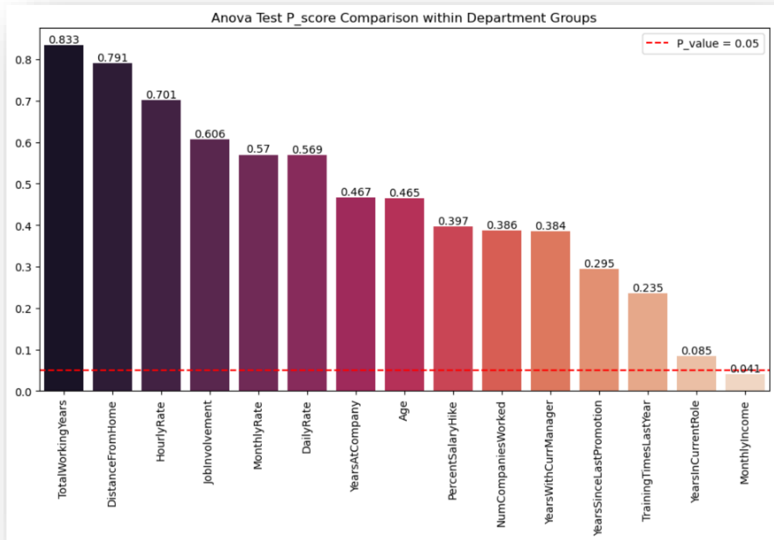
$$\chi^2 = 10.796\,,\ \ \chi_c^2 = 5.991, \qquad \text{Pvalue} = 0.0045$$

Since $\chi^2 > \chi_c^2$ or $\text{Pvalue} < 0.05$ we will reject our Null Hypothesis and accept the alternative that says **"Department has a significant effect on the decision of quit ".**

Is there a significant difference among groups of different departments?

To answer this question we will do Analysis of Variance **ANOVA** test.

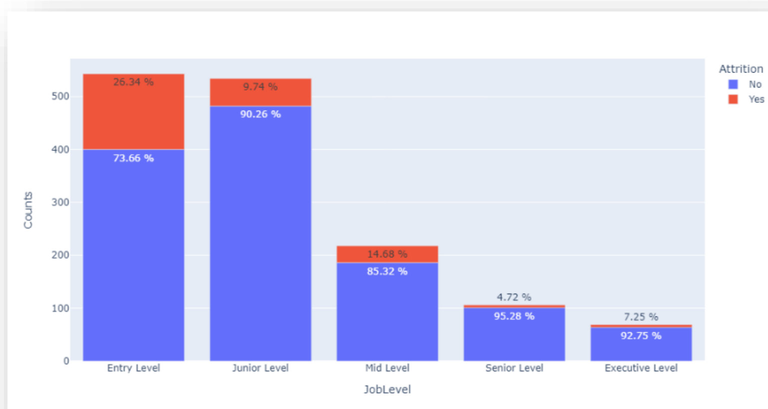
We extract p value from the test which shows that there is a significant difference in groups from each department in their Monthly income, we will know exactly any pair of groups between all the groups has a significant difference in Monthly income analysis.



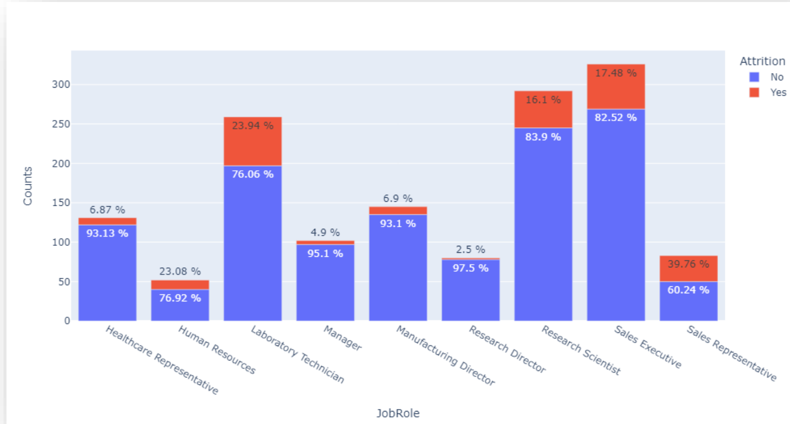
Job Role/Level Analysis:

Which job role/level has the highest attrition rate and which has the lowest?

- Most employees is working as Sales executive, Research Scientist or Laboratory Technician in this organization.
- Highest attrition rates are in sector of Research Director, Sales Executive, and Research Scientist.
- Laboratory Technician and HR job roles also has a high attrition rate



- Highest attrition rates are in Entry and mid Job level.
- Employees that have High Job level such as seniors and Executive level tend to stay in company more than the others.



Is there a significant difference in the attrition rates between different job roles/levels?

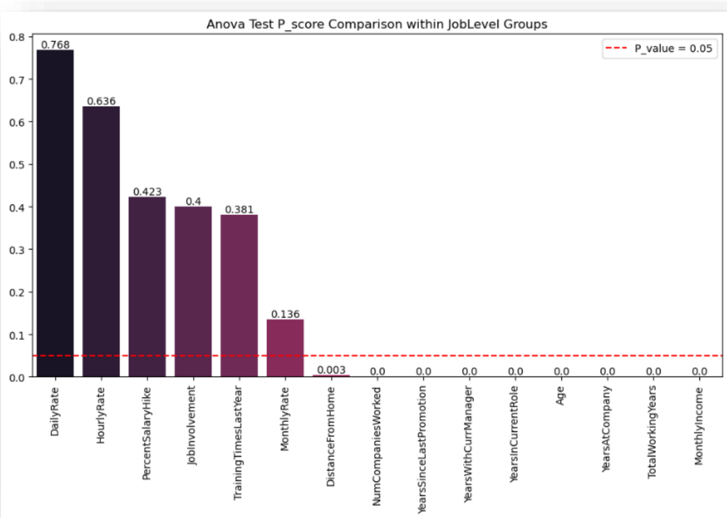
Let’s execute chi square test as above to check this

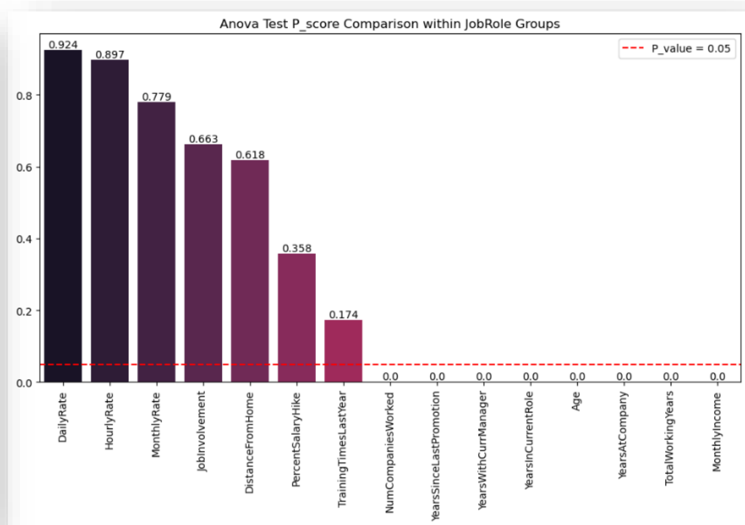
Category	χ^2	χ_c^2	P value
Job Role	86.1903	15.5073	$2.7525*10^{-15}$
Job Level	72.5290	9.4877	$6.6347*10^{-15}$

We can conclude that **"Both Job Role/Level have a significant effect on the decision of quit ".**

Is there a significant difference among groups of different Job roles/Levels?

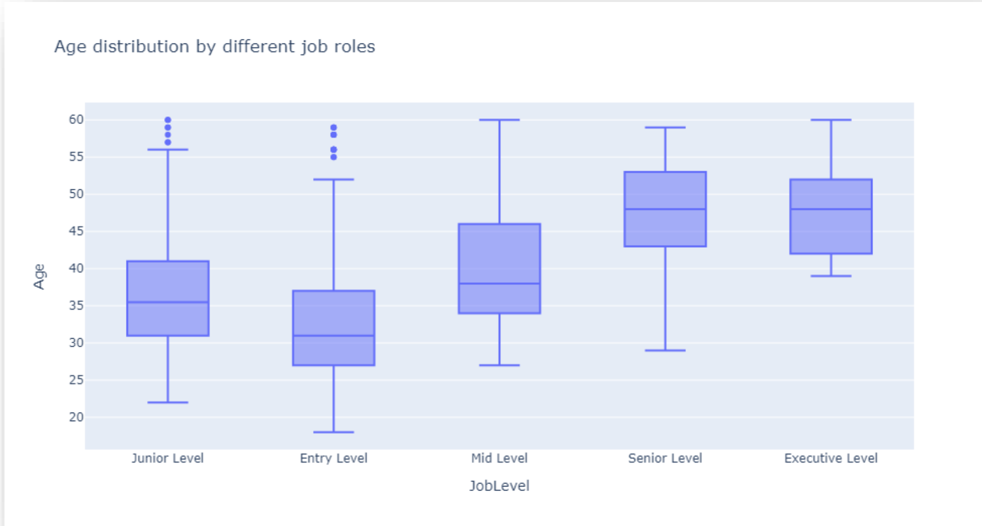
Let's do ANOVA test to know:



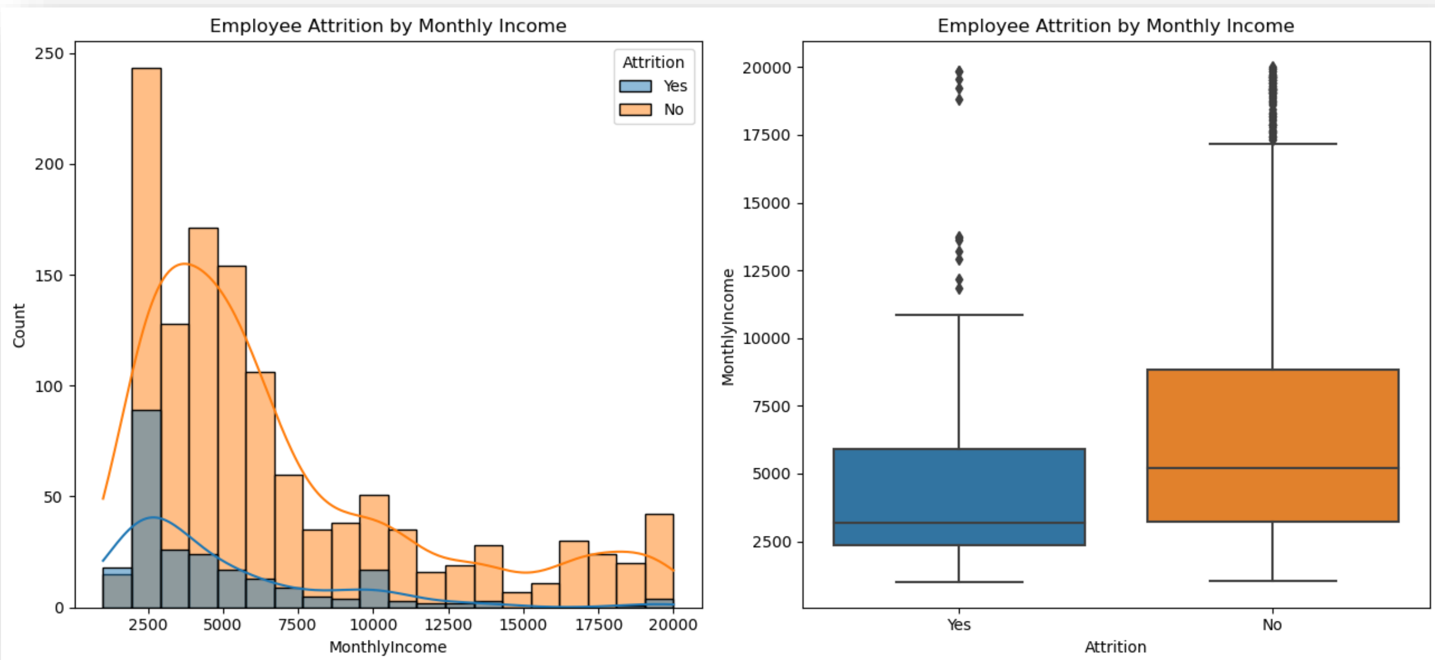


The test shows that there is a significant difference between groups from different job roles/levels in values such as **Age , Monthly Income ,Number of Companies Worked ,Total Working Years ,Years At Company , Years In Current Role, Years Since Last Promotion, Years With Current Manager**.

For example this graph shows that the Age distribution from some of job levels is significantly different which agree with ANOVA results



Monthly Income Analysis:



Most of the employees are getting paid less than 10000 in the organization

Less paid employees tend to quit more than other which is logically.

But to know if the Monthly income plays significant role in attrition,

We can apply T_Test between employees who quit and the others who stayed.

Null Hypothesis H_0: Monthly income does not affect the attrition.

Alternative Hypothesis H_a: Monthly income has a significant effect on attrition.

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = -6.204, \qquad t_c = -1.96, \qquad \text{Pvalue} = 7.1474 * 10^{-15}$$

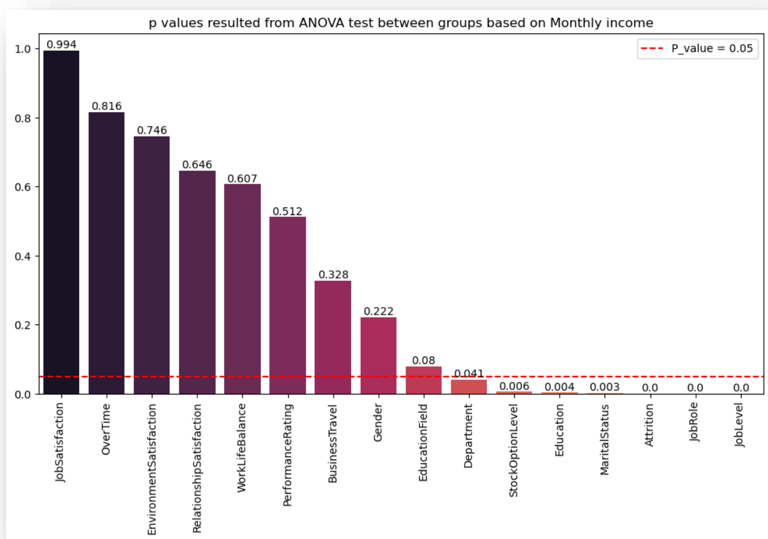
Since $t < | t_c |$ or $p \ll 0.05$we will can reject Null Hypothesis and accept alternative that says **"Monthly income has a significant effect on attrition".**

Are there any groups that differ significantly in Monthly income?

Let's execute ANOVA test to know which categorical variables vary in monthly income:

ANOVA test results shows that there is significant difference in monthly income between different groups of:

- Department
- Education
- Job levels
- Job roles.



So let's analyze some of them independently

Monthly income by department:

By applying ANOVA test we found:

$$f = 3.2017, \; f_c = 3.0018, \qquad \text{Pvalue} = 0.04097$$

Since $\mathrm{f} < \mathrm{f_c}$ or $\mathrm{p} < 0.05$ we can say that *the mean of monthly income is different from department to another*.

But **which pair of groups between all the groups that has a significant difference in Monthly income?**

To know that let's do post hoc test by help of T_Test

This table show p value for each test between each group and others, the results show that the groups that has a significant difference in Monthly income are Research & Development and Sales departments.

	Sales	Research & Development	Human Resources
Sales	1.00000	0.01100	0.59948
Research & Development	0.01100	1.00000	0.56254
Human Resources	0.59948	0.56254	1.00000

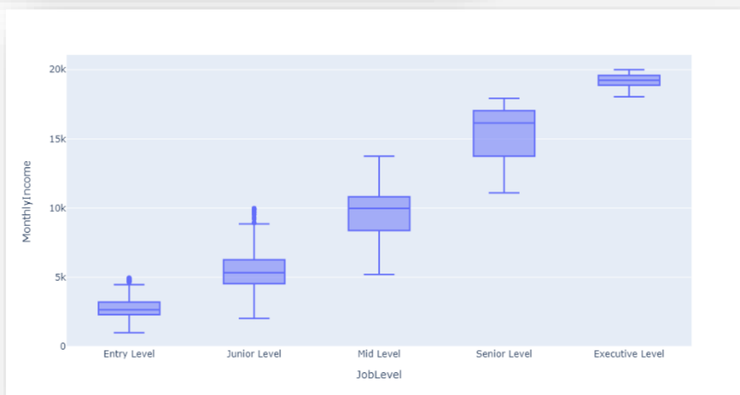
We can calculate maximum error of estimation as

$$E = \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \,, \qquad t_{\frac{\alpha}{2}} = 1.96 \,, \qquad CI = 95\%$$

Job Level with monthly income:

JobLevel	Total number	Standard Deviation	Mean	Max Error
Junior Level	513	1161.309722	5334.621832	+/- 100.7315
Entry Level	526	665.460059	2721.863118	+/- 57.0006
Mid Level	218	1805.999233	9817.252294	+/- 241.0828
Senior Level	106	1816.239003	15503.783019	+/- 349.7859
Executive Level	69	512.383127	19191.826087	+/- 123.0879

It is obvious that groups of different job levels vary a lot in their Monthly income as shown in this plot.



Let’s move to Job Role feature

JobRole	Total number	Standard Deviation	Mean	Max Error
Sales Executive	325	2338.860456	6902.901538	+/- 255.2325
Research Scientist	285	1036.962913	3146.663158	+/- 120.9048
Laboratory Technician	254	1049.060730	3168.397638	+/- 129.6326
Manufacturing Director	145	2676.745753	7295.137931	+/- 439.3761
Healthcare Representative	131	2542.550170	7528.763359	+/- 439.4846
Manager	94	1751.984116	17644.212766	+/- 358.8411
Sales Representative	72	447.921852	2550.986111	+/- 105.2565
Research Director	80	2827.621369	16033.550000	+/- 629.2563
Human Resources	52	2438.849744	4235.750000	+/- 678.9801

This graph show the mean of groups from different job roles with MEE but also separated by Attrition

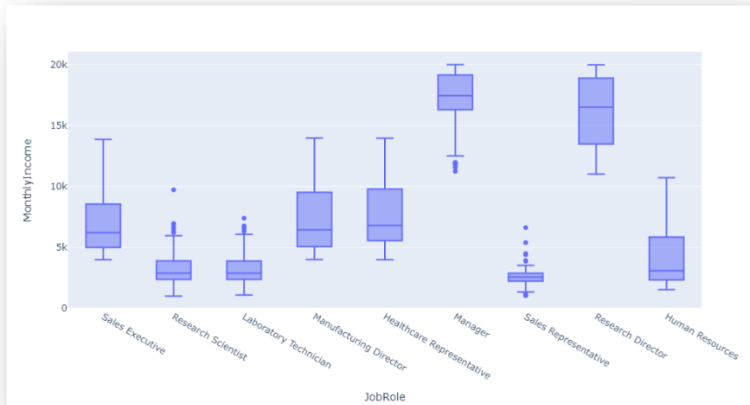
We can see that employees who quit as a manager, manufacturing director or Research director vary greatly in monthly income.

All above calculations and plots strongly agree with what we conclude from ANOVA test that groups of different job levels/roles vary in their Monthly income.

Which pair of groups between all the groups that has a significant difference in Monthly income?

	Sales Executive	Research Scientist	Laboratory Technician	Manufacturing Director	Healthcare Representative	Manager	Sales Representative	Research Director	Human Resources
Sales Executive	1.00000	0.00000	0.00000	0.13262	0.01607	0.00000	0.00000	0.00000	0.00000
Research Scientist	0.00000	1.00000	0.97773	0.00000	0.00000	0.00000	0.00002	0.00000	0.00001
Laboratory Technician	0.00000	0.97773	1.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00001
Manufacturing Director	0.13262	0.00000	0.00000	1.00000	0.45905	0.00000	0.00000	0.00000	0.00000
Healthcare Representative	0.01607	0.00000	0.00000	0.45905	1.00000	0.00000	0.00000	0.00000	0.00000
Manager	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	0.00298	0.00000
Sales Representative	0.00000	0.00002	0.00001	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
Research Director	0.00000	0.00000	0.00000	0.00000	0.00000	0.00298	0.00000	1.00000	0.00000
Human Resources	0.00000	0.00001	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

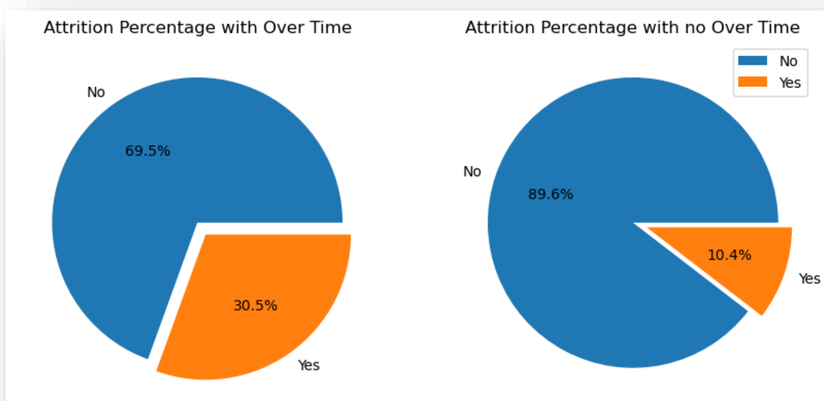
Any value of p below 0.05 in post hoc results table indicates that those pairs are varying in their Monthly income, that clearly obvious in this plot



Over Time Analysis:

Over 30% of employees who work over time quit the organization with percentage higher than who do not work over time.

This should motivate us to ask ourselves some questions to know if the option of over time affect the decision of quit.



Are employees who work over time tends to quit more than another?

We can ask this question by execute chi square test and show results and test our hypothesis, we found that:

$$\chi^2 = 87.5643,\ \ \chi_c^2 = 3.8414, \qquad \text{Pvalue} = 8.1584 * 10^{-21}$$

Since $\chi^2 \gg \chi_c^2\ or\ p \ll 0.05$ We will reject our Null Hypothesis and accept the alternative that says **"Overtime has a significant effect on the decision of quit "**.

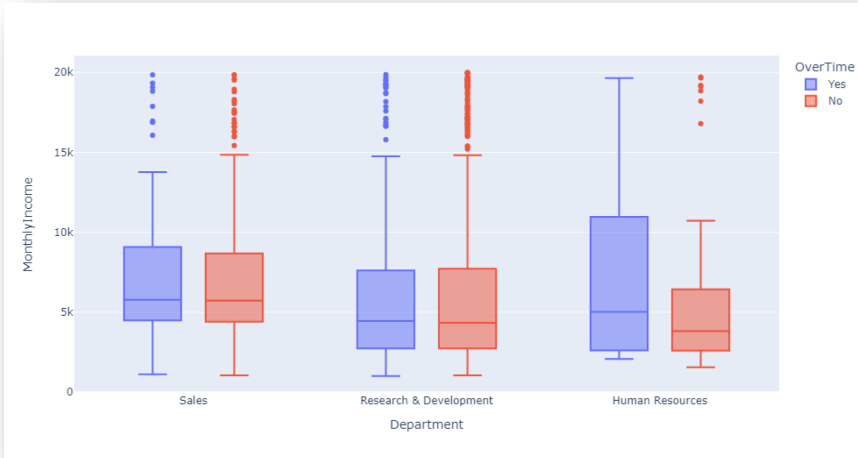
Is overtime option dependent on specific departments?

By applying chi square test we obtained that:

$$\chi^2 = 0.0936,\ \ \chi_c^2 = 5.9914$$

$$\text{Pvalue} = 0.9543$$

We can conclude that Overtime is **independent of departments** which is also noticed from this plot.



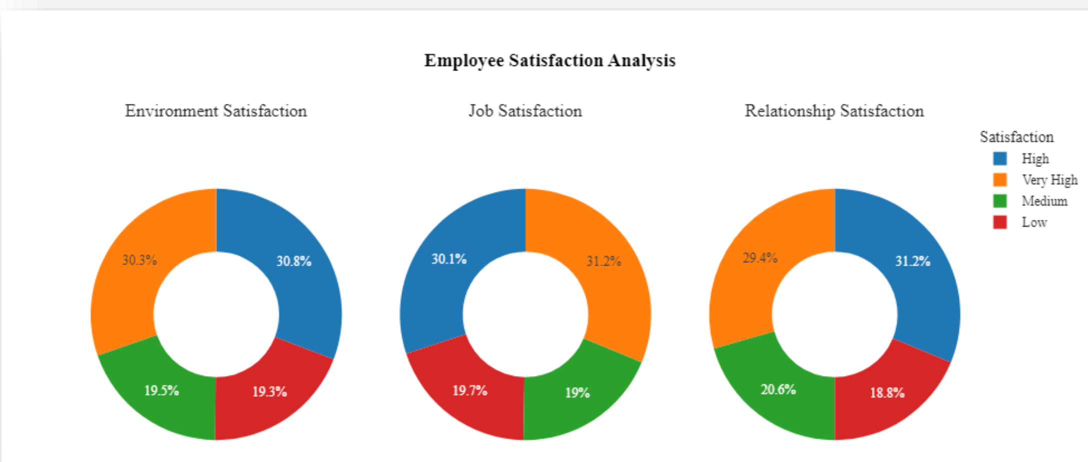
Are employees who work overtime get paid more than others who don't?

We can apply T_Test on the two groups in monthly income, we found:

$$t = 0.2333, \qquad t_c = 1.96\,, \qquad Pvalue = 0.8155$$

We can say that **"There is no significant difference in monthly income between employees who work over time and others who don't,** which may lead employees who work over time to quit.

Satisfaction rates:



Has level of satisfaction a relation with Attrition?

Let's do ANOVA test to know.

Null hypothesis: level of satisfaction does not affect attrition.

Alternative hypothesis: level of satisfaction has an effect on attrition.

Satisfaction Type	χ^2	χ_c^2	**P value**
Environment	22.5038	7.8147	$5.123 * 10^{-5}$
Job	17.5051		0.0005
Relationship	5.2411		0.1549

Results show that Environment and job satisfaction have an effect on attrition but relationship satisfaction don't have

Personal Reasons

Age Analysis:

Is there a significant association between age and attrition?

$$t = -6.1786, \qquad t_c = 1.96, \qquad Pvalue = 8.356 * 10^{-10}$$

Since $t < |t_c|$ or $p \ll 0.05$ we can say that **"There is a significant difference in ages between employees who quit and others who stayed which indicates that age has an effect on attrition".**

Are elder get more monthly income?

We can calculate the correlation between age and monthly income to and check for p value to know that

$$Pearsonr = 0.4978, \qquad Pvalue \approx 0$$

So we can conclude that there is a moderate positive correlation between Monthly income and age which means that Monthly income increase as the employee became older.

Prediction Interval:

This is a plot for prediction interval which means that I am confident with 95% that the future observations will be placed in this red shaded area.

Formula of prediction interval

$$\widehat{y_h} \pm t_{\alpha\backslash 2} \times \sqrt{MSE \times (1 + \frac{1}{n} + \frac{(x_h - \bar{x})}{\sum(x_i - \bar{x})}}$$

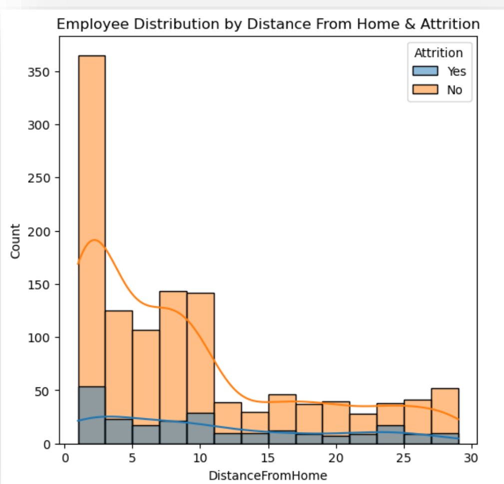
Distance from Home Analysis:

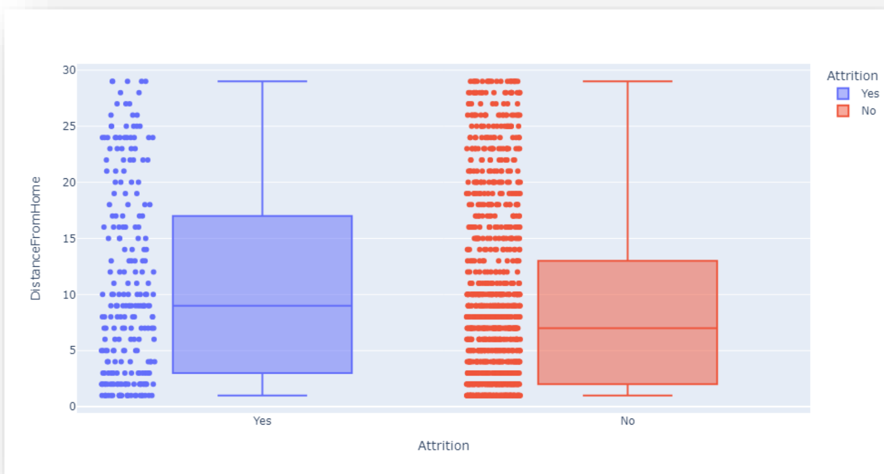
Does the distance from home play a role in turnover?

Let's discover that by applying T_Test on groups of employees who quit and others who don't.

$$t = 2.9947, \qquad t_c = 1.96, \qquad Pvalue = 0.0028$$

Since $t < |t_c|$ or $p < 0.05$ we can conclude that **the distance from home play a role in turnover**.



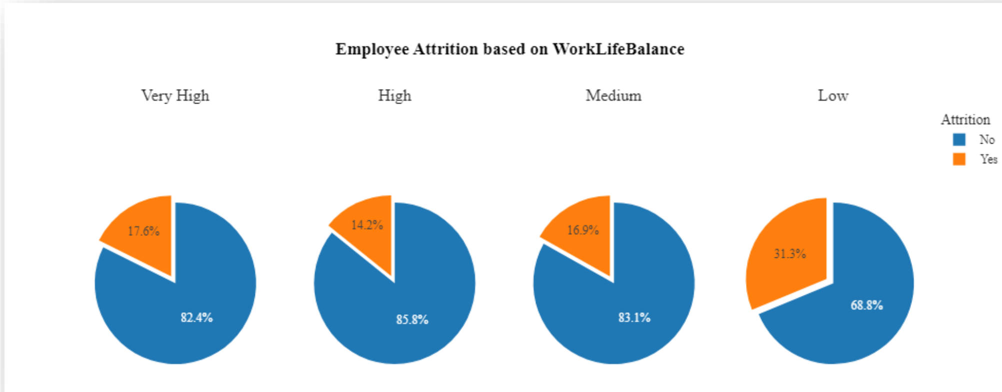


From plots we can see that most employees Located at distance from 0 to 10 to the company and the attrition rate is high in this region.

Work Life Balance Analysis:

Has Work life Balance an effect on employees attrition?

We can see that employees who have low WorkLife Balance are more likely to quit. So let's check if work life balnance play a role in the decision of quit by applying chi square tes.



We found that:

$$\chi^2 = 16.3251, \ \chi_c^2 = 7.8147, \qquad \text{Pvalue} = 0.0009$$

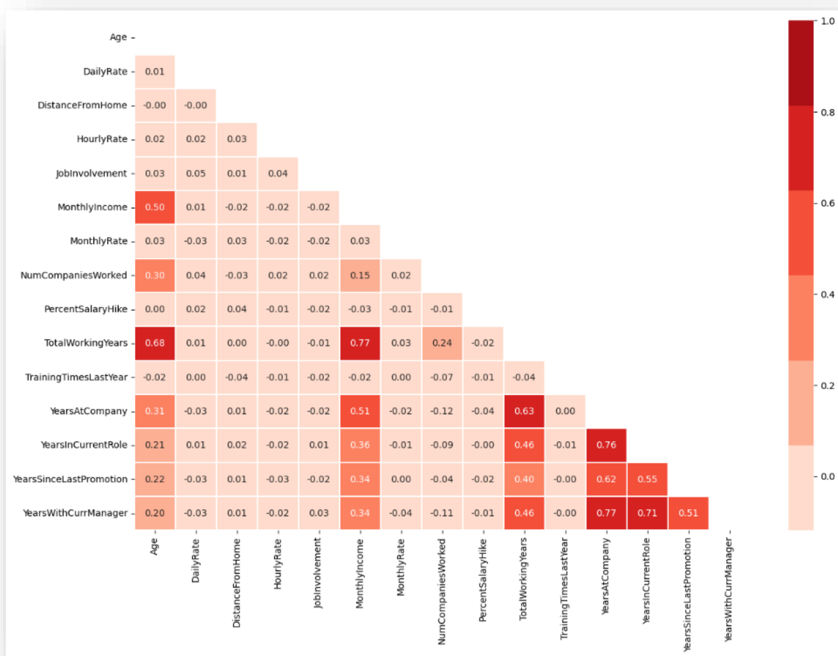
We can conclude that **"Work life Balance has an effect on employees attrition".**

Correlation Analysis:

Correlation Matrix:

We can see that:

- Features related with years of service are high correlated with each other.
- Age is moderate correlated with monthly income but highly correlated with total working years.
- The higher the total working years higher the monthly income.



We will use this analysis when training a model to drop some features which are highly correlated with each other but select some of them to be used in model training.

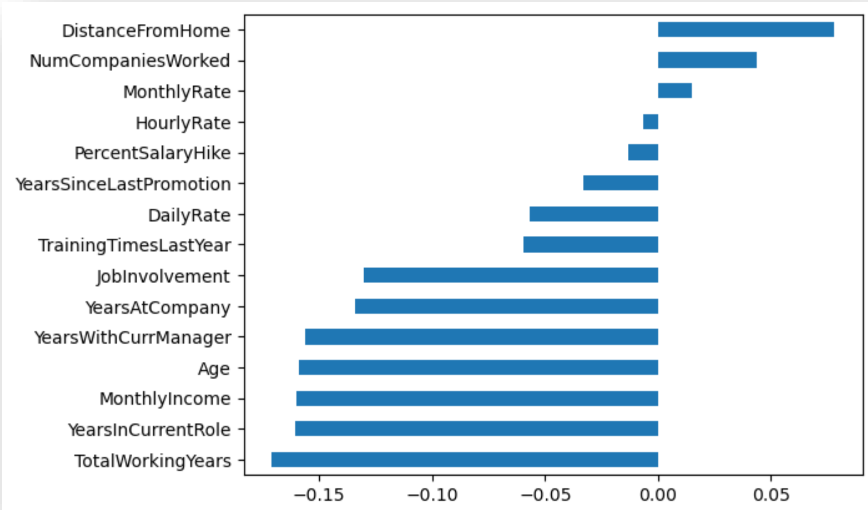
Point bi-serial correlation:

We can use this type of correlation to see which features are highly correlated with our target value, we use this type as our target is a binary categorical variable.

We will formulate our hypothesis as:

Null hypothesis (H_0): There is no association between considering numerical feature and Employee Attrition

Alternative hypothesis (H_a): There is a significant different and this numerical feature affect Attrition.



Features	Corr	P_value	Result
Age	-0.159205	8.356308e-10	Reject
DailyRate	-0.056652	2.985816e-02	Reject
DistanceFromHome	0.077924	2.793060e-03	Reject
HourlyRate	-0.006846	7.931348e-01	Accept
JobInvolvement	-0.130016	5.677065e-07	Reject
MonthlyIncome	-0.159840	7.147364e-10	Reject
MonthlyRate	0.015170	5.611236e-01	Accept
NumCompaniesWorked	0.043494	9.552526e-02	Accept
PercentSalaryHike	-0.013478	6.056128e-01	Accept
TotalWorkingYears	-0.171063	4.061878e-11	Reject
TrainingTimesLastYear	-0.059478	2.257850e-02	Reject
YearsAtCompany	-0.134392	2.318872e-07	Reject
YearsInCurrentRole	-0.160545	6.003186e-10	Reject
YearsSinceLastPromotion	-0.033019	2.057900e-01	Accept
YearsWithCurrManager	-0.156199	1.736987e-09	Reject

The results show that:

- Business Travel
- Department
- Education Field
- Environment Satisfaction
- Job Level
- Job Role
- Job Satisfaction
- Marital Status
- Over Time
- Stock Option Level
- Work Life Balance

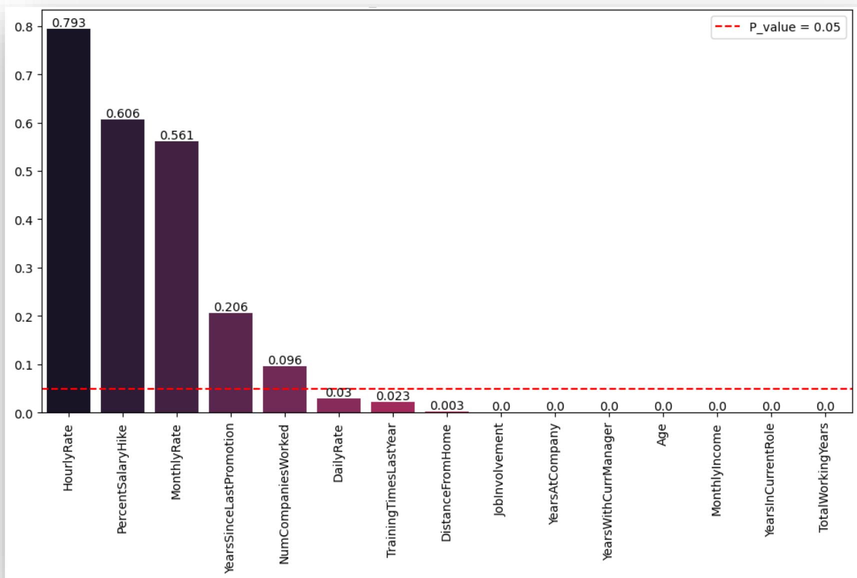
Each has a good correlation score that mean that there is a relation between those features and the attitude of quit, those results agree with what we conclude over our analysis.

Summary of t and Chi square tests:

We have already execute t test and chi test on most of those features along our previous analysis so let's do a summary of overall results with all features with Attrition.

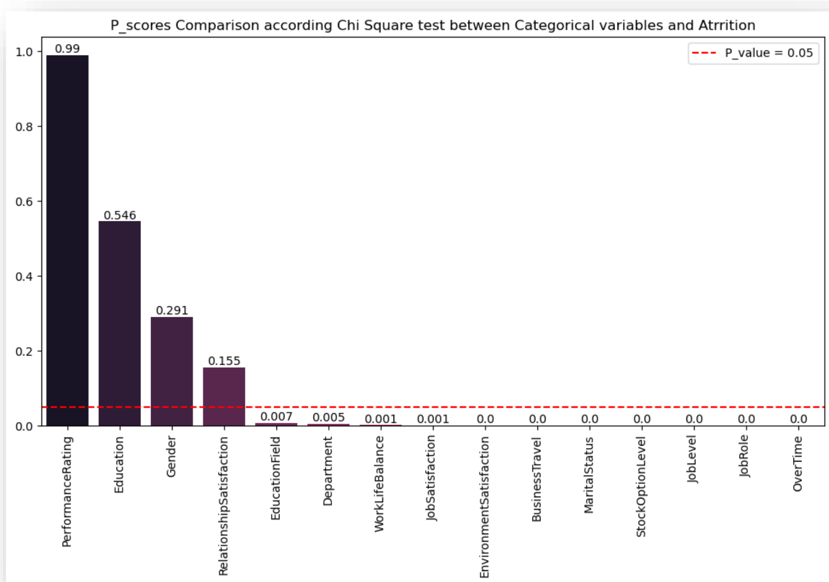
T-Test Summary

Features	T Value	Critical Value	P_value	Result
Age	-6.178664	1.961581	8.356308e-10	Reject
DailyRate	-2.174084	1.961581	2.985816e-02	Reject
DistanceFromHome	2.994708	1.961581	2.793060e-03	Reject
HourlyRate	-0.262290	1.961581	7.931348e-01	Accept
JobInvolvement	-5.024140	1.961581	5.677065e-07	Reject
MonthlyIncome	-6.203936	1.961581	7.147364e-10	Reject
MonthlyRate	0.581306	1.961581	5.611236e-01	Accept
NumCompaniesWorked	1.668019	1.961581	9.552526e-02	Accept
PercentSalaryHike	-0.516457	1.961581	6.056128e-01	Accept
TotalWorkingYears	-6.652255	1.961581	4.061878e-11	Reject
TrainingTimesLastYear	-2.282903	1.961581	2.257850e-02	Reject
YearsAtCompany	-5.196309	1.961581	2.318872e-07	Reject
YearsInCurrentRole	-6.232038	1.961581	6.003186e-10	Reject
YearsSinceLastPromotion	-1.265788	1.961581	2.057900e-01	Accept
YearsWithCurrManager	-6.059069	1.961581	1.736987e-09	Reject



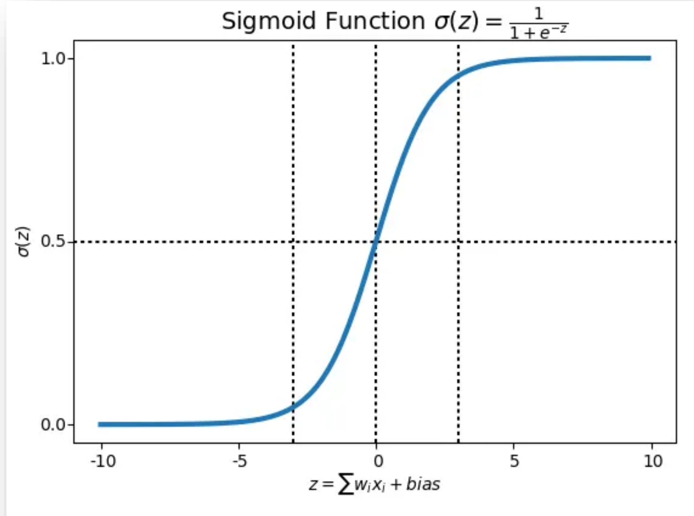
Chi square test Summary

Features	Chi_2 Statistic	Critical Value	P_value	Result
BusinessTravel	24.182414	5.991465	5.608614e-06	Reject
Department	10.796007	5.991465	4.525607e-03	Reject
Education	3.073961	9.487729	5.455253e-01	Accept
EducationField	16.024674	11.070498	6.773980e-03	Reject
EnvironmentSatisfaction	22.503881	7.814728	5.123469e-05	Reject
Gender	1.116967	3.841459	2.905724e-01	Accept
JobLevel	72.529013	9.487729	6.634680e-15	Reject
JobRole	86.190254	15.507313	2.752480e-15	Reject
JobSatisfaction	17.505077	7.814728	5.563005e-04	Reject
MaritalStatus	46.163677	5.991465	9.455511e-11	Reject
OverTime	87.564294	3.841459	1.000000e-20	Reject
PerformanceRating	0.000155	3.841459	9.900745e-01	Accept
RelationshipSatisfaction	5.241068	7.814728	1.549724e-01	Accept
StockOptionLevel	60.598301	7.814728	4.379390e-13	Reject
WorkLifeBalance	16.325097	7.814728	9.725699e-04	Reject



Predictive Model (Logistic Regression)

Logistic regression is a Supervised Machine Learning algorithm mainly used for binary Classification where we use a logistic function, also known as a sigmoid function that takes input as independent variables and produces a probability value between 0 and 1, this model is commonly estimated via maximum likelihood estimation (MLE)



Sigmoid Function Graph

Given X (which means feature matrix also) what is the probability the y (which means target value) = 1

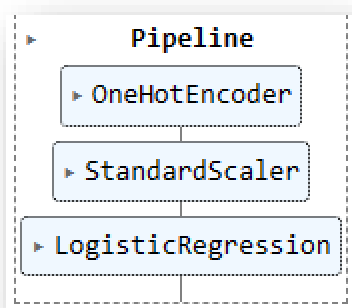
$$\hat{y} = P(y = 1\,|X)$$

That's what our model try to get.

We make a pipeline that take the feature matrix and encode the categorical variables because the model can handle only numerical values, then standard scalar to to ensure that features are on a similar scale because having features on different scales can affect the algorithm's performance, as logistic regression coefficients represent the change in the log-odds of the response variable per one-unit change in the predictor variable. If one predictor has a larger scale than another, it might dominate the learning process, leading to biased coefficients.

Standard Scalar use z-score normalization

$$z = \frac{x - \mu}{\sigma}, \qquad z \sim N(0,1)$$



Where:

- (z) is the standardized value,

- (x) is the original value of the feature,

- (μ) is the mean of the feature's values,

- (σ) is the standard deviation of the feature's values.

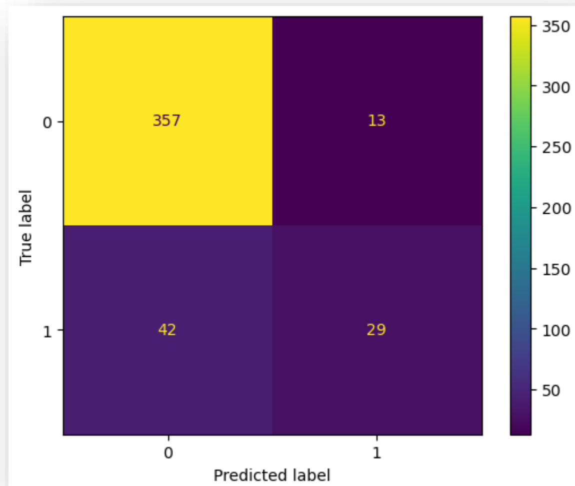
We use a 70% of the data to be trained and 30% for testing.

WE get that:

```
ACCURACY SCORE: 0.8753
Precision score: 0.69
Recall Accuracy: 0.41
Area under Curve AUC: 0.69
```

The model give a bad scores because the data is imbalanced as there are only 15% of employees who quit and the 85% stayed.

Confusion Matrix:

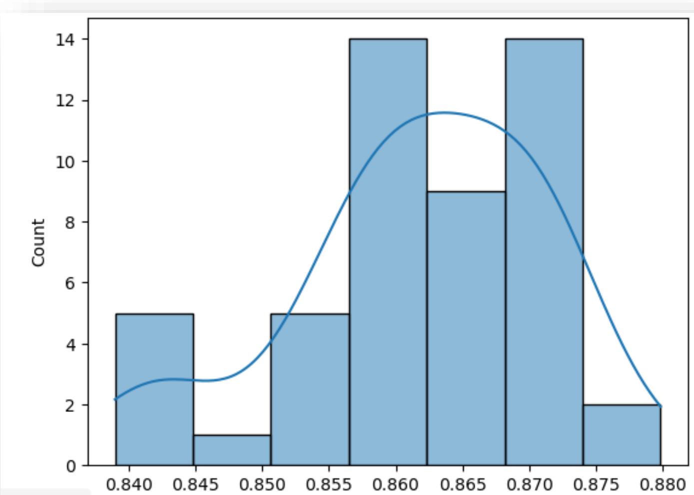


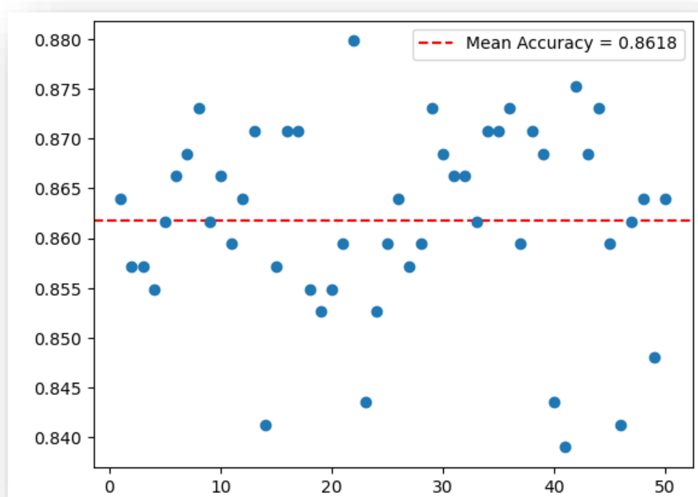
That means

- 375 of true labels have been predicted successfully, 29 of the false ones also.
- 42 of true labels have been misclassified and predicted as false.
- 13 of false labels have been misclassified and predicted as true.

Bootstrap confidence interval

By training our model on 50 bootstrapped samples we get that the mean of accuracies ≈ 86.2%

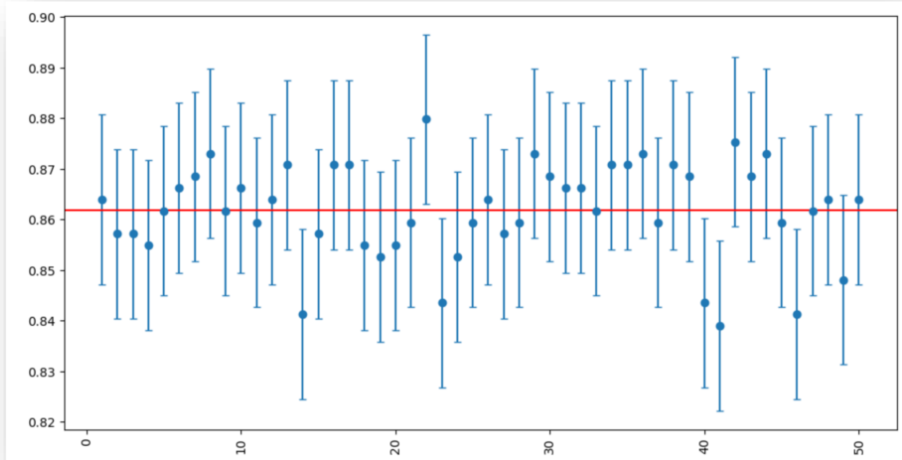




We plot error bars plot with confidence level 95%, which means that I am confident with 95% that the model accuracy on different samples

$= (86.2 \pm 1.7)\%$

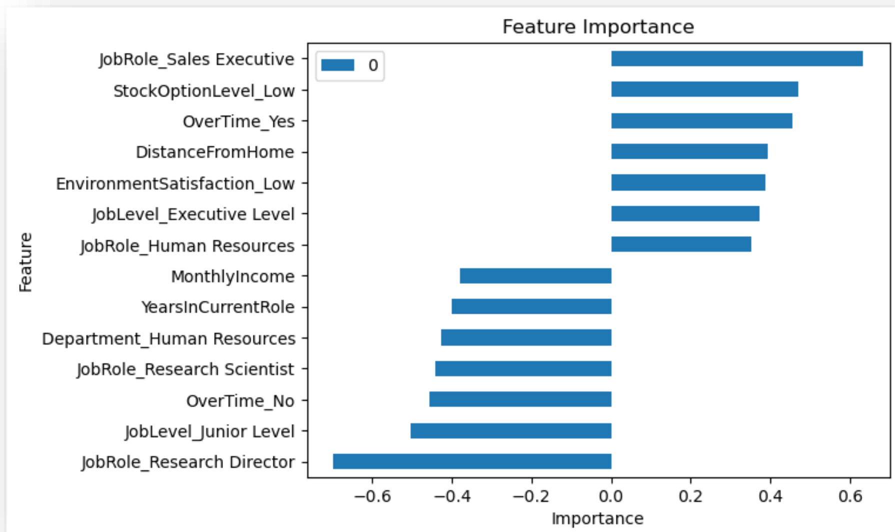
We can see that some models are out of interval as the 95% CI does not capture all real accuracies.



Feature importance:

The larger the absolute value of a coefficient, the more influential that particular feature is in predicting the outcome. The sign of the coefficient indicates the direction of the impact (positive or negative).

For example Sales Executive and Research Director Job roles each has a high importance in predicting the target, however the first one is in positive way and the second in negative.



Risk Level:

Finally we use the probability from the model to build an Early Warning System that try to provide alerts when an employee is about to quit, for example this is a subset of the new Data Frame.

	Attrition	AttrionLikelihood	RiskLevel
0	Yes	0.876117	Strong
1	No	0.001710	Weak
2	Yes	0.173188	Weak
3	No	0.155206	Weak
4	No	0.084546	Weak
5	No	0.254310	Weak
6	No	0.678870	Strong

For new employees we build a dashboard that predict the likelihood to quit with a risk level, you can find it on the full project on the Jupyter notebook in Appendix Section.

Discussion

<u>And now we will Discuss how our results could be utilized in further studies or practical applications.</u>

1. Identifying Predictors of Attrition:

- Explore which features or factors are strong indicators of attrition in the company. Understanding these predictors can guide future studies to delve deeper into the root causes of employee turnover.

2. Targeted Intervention Strategies:

-For instance, if certain departments or roles exhibit higher attrition rates, HR teams could implement retention initiatives tailored to those specific areas.

3. Early Warning System:

- Our Implemented Program could be used as an early warning system. This system could provide alerts when employees show signs of being at a higher risk of attrition, allowing proactive measures to be taken to retain valuable talent.

4. Benchmarking and Comparison:

- our study could be extended by benchmarking the attrition rates against industry standards or similar companies. This comparative analysis could provide valuable insights into whether the attrition rates are within expected ranges or if there are specific areas where the company stands out.

5. Collaboration with HR Professionals:

- Encourage collaboration between data scientists and HR professionals to implement and fine-tune strategies based on the model's recommendations. The combination of data-driven insights and HR expertise can lead to more effective solutions.

Some additional questions and hypotheses that arose from Our analysis which could be the focus of subsequent investigations.

Impact of Remote Work: Does the shift towards remote work influence attrition rates? Hypothesis: Increased remote work might either reduce attrition due to improved work-life balance or increase it due to decreased personal connections and engagement.

Managerial Styles and Attrition: How do different managerial approaches affect attrition? Hypothesis: Supportive and communicative managers might have lower attrition rates compared to those who are more authoritative or hands-off.

Career Development and Attrition: Does a lack of career development opportunities contribute to higher attrition? Hypothesis: Employees with limited growth prospects within an organization might be more likely to leave for better opportunities elsewhere.

Workplace Culture and Attrition: How does organizational culture impact attrition rates? Hypothesis: Companies fostering a positive, inclusive, and engaging culture are likely to experience lower attrition rates.

Diversity and Inclusion Effects: Does diversity and inclusion within a company influence attrition rates? Hypothesis: Companies with inclusive policies may experience lower attrition rates due to increased employee satisfaction and a sense of belonging.

Exit Interview Insights: What patterns emerge from exit interviews? Hypothesis: Common reasons for attrition revealed in exit interviews might point to systemic issues (e.g., poor management, lack of growth opportunities) contributing to high turnover.

Industry Comparison: Do certain industries have consistently higher attrition rates than others? Hypothesis: High-stress industries or those with limited advancement opportunities might have higher attrition rates compared to industries with better work-

Conclusions

High employee turnover has high costs to a company, both financially as well as culturally. Job dissatisfaction is a critical factor that leads employees to leave their jobs as well as poor benefit packages, poor culture and poor career development opportunities. Employers can use digital coaching to address these issues and accelerate progress in areas such as personal and professional growth, improving compensation and benefits packages, and creating a positive workplace culture and management culture. It is vital to address these issues in order to improve employee retention as well as company culture.

One of the best ways to prevent a worker from leaving an organization is to get regular feedback from them. It is essential not only to provide the working staff with feedback on their work but also to ask about their opinion. This can help to establish effective communication and identify a problem at an early stage. Moreover, a friendly atmosphere at work and a good relationship with the boss play a crucial role in retaining professionals in a company.

<u>Understanding and managing attrition is crucial for various reasons, and its implications have a significant impact on the business landscape. Here are some key points to consider:</u>

Cost Implications:

High attrition rates can result in substantial financial losses for a company. The cost of recruiting, hiring, and training new employees is often higher than retaining existing ones. Additionally, the loss of institutional knowledge and productivity during the transition period can further contribute to financial setbacks.

Organizational Stability:

Attrition can affect the overall stability and continuity of an organization. Frequent turnover can disrupt team dynamics, project timelines, and the overall work environment. Maintaining a stable workforce is essential for ensuring consistent performance and achieving long-term business objectives.

Employee Morale and Productivity:

High attrition rates can negatively impact the morale of remaining employees. When colleagues leave frequently, it can create a sense of instability and insecurity among those who remain, potentially leading to decreased productivity and engagement.

Talent Management and Development:

Understanding the reasons behind attrition can provide valuable insights into talent management and development strategies. Identifying common patterns or issues that lead employees to leave allows companies to address those concerns and implement measures to retain valuable talent..

Strategic Planning:

Attrition data can inform strategic planning by highlighting areas that require attention. For instance, if a particular department consistently experiences high turnover, it may signal issues related to leadership, work culture, or job satisfaction that need to be addressed strategically.

Resources

- Logistic Regression in Machine Learning - GeeksforGeeks

- https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148

- Prediction Interval for a New Response | STAT 501 (psu.edu)

- Calculating Confidence Intervals with Bootstrapping | by Barış Hasdemir | Towards Data Science

Appendix

Here is a copy of the notebook if you want to take a look for more details:

Elkhiat15/Probability-Project (github.com)