MINI PROJECT

Initial Report

Topic: HR Analytics – Analysing Employee Attrition Rate

By **Ajmal M S**

CONTENTS

PROBLEM STATEMENT

ABSTRACT

WHAT ARE HR ANALYTICs

PROCESS INVOLVED

APPLICATIONS

DATA SET AND TECHNIQUES USED

METHODOLOGY DIAGRAM

CONCLUSION

PROBLEM STATEMENT

Human Resources are pillar of any organization. Involving from hiring to firing, HR department take care of all employee related things. In an organization, the time consumed is for hiring or retention process of work force. This time consuming yet necessary job can be done efficiently using HR Analytics. HR Analytics is a data set with over 30 characteristics, categorized by numeric and text data, and discrete. With the advent of storing data in digital form and the realization of its value, a race has begun to automate many older systems in order to improve speed accuracy! Automate your organization's system of hiring and firing. This is made possible with the help of data science and machine learning techniques.

ABSTRACT

Talent management in the 21st century is becoming more efficient by taking full advantage of technological advances. From finding the right talent to retaining top talent, organizations strive to make many smart decisions. Decision-making in HR is primarily based on trust and relationships, unlike other administrative functions. From our perspective, HR is a highly neglected area compared to other functional areas, but each business operation needs the right people for better results. However, after the Great Recession of 2008, most organizations recognized the need for more accurate and evidence-based talent management. Fortunately, HR big data has brought HR analytics to the concept of evidence-based HRM. Data-driven and evidence-based HR managers must practice analysis, decision-making and problem-solving to make accurate HR decisions. Therefore, the concept of evidence-based HRM with effective HR analysis tools enhances the accurate decision-making power of HRM. This white paper highlights the importance of HR analytics practice and its applicability in various sectors. We also attempt to follow the regular development of HR analytics as an effective, evidence-based HR management tool.

My aim is;

- To classify / predict whether an employee continues with the company or not!
- To draw insights about employee performance as well as employee retention / departure from the employee data!

HR ANALYTICS

HR analytics is the process of collecting and analysing Human Resources (HR) data to improve employee performance in an organization. This process is sometimes called Talent Analytics, People Analytics, or Workforce Analytics.

This data analysis method takes data collected by HR on a regular basis and relates it to HR and organizational goals. This provides measurable evidence of how HR initiatives are contributing to organizational goals and strategies. For example, if a software engineering firm has high employee turnover, the company is not operating at a fully productive level.

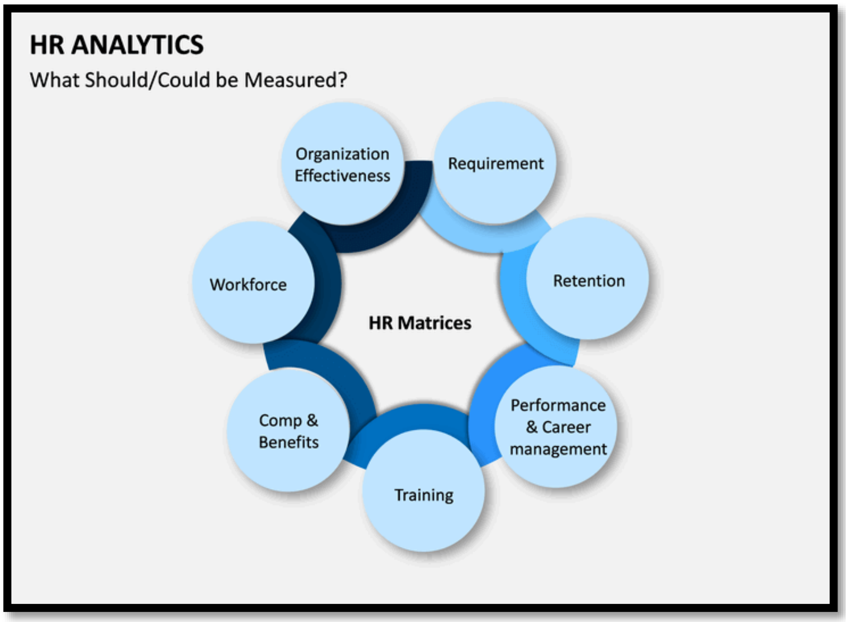

HR ANALYTICS
What Should/Could be Measured?
Organization Effectiveness
Requirement
Retention
Performance & Career management
Training
Comp & Benefits
Workforce
HR Matrices

Figure 1 - HR Analytics

It takes time and investment to get your workforce up to a fully productive level. HR analytics provides data-driven insights into what is working and what isn't, so organizations can make improvements and plan more effectively for the future. As in the example above, knowing what causes a company's high turnover rate can provide valuable insight into how to reduce it. By reducing turnover, the company can increase turnover and productivity. Most companies already have data collected on a regular basis. So why a special form of analytics? Can't HR just see the data they already have? Unfortunately, raw data alone does not provide useful insights. It's like looking at a big table full of numbers and words. Without organization and direction, data seems meaningless. Organizing, comparing, and analysing this raw data can yield useful insights. They can help answer questions like:

- What patterns can be revealed in employee turnover?
- How long does it take to hire employees?
- What amount of investment is needed to get employees up to a fully productive speed?
- Which of our employees are most likely to leave within the year?

- Are learning and development initiatives having an impact on employee performance?

Organizations can concentrate on making the required improvements and planning for future initiatives when they have data-supported evidence. It is not surprising that many companies using HR analytics are attributing performance improvement to HR initiatives given the capability to provide accurate answers to crucial organizational questions.

PROCESS INVOLVED

HR Analytics is made up of several components that feed into each other.

1. To gain the problem-solving insights that HR Analytics promises, data must first be **collected**.
2. The data then needs to be monitored and **measured** against other data, such as historical information, norms or averages.
3. This helps identify trends or patterns. It is at this point that the results can be **analysed** at the analytical stage.
4. The final step is to **apply** insight to organizational decisions.

Let’s take a closer look at how the process works:

1. Collecting data

Figure 2 - Collecting Data

Big data refers to the large quantity of information that is collected and aggregated by HR for the purpose of analysing and evaluating key HR practices, including recruitment, talent management, training, and performance

Collecting and tracking high-quality data is the first vital component of HR analytics.

The data needs to be easily obtainable and capable of being integrated into a reporting system. The data can come from HR systems already in place, learning & development systems, or from new data-collecting methods like cloud-based systems, mobile devices and even wearable technology.

The system that collects the data also needs to be able to aggregate it, meaning that it should offer the ability to sort and organize the data for future analysis.

What kind of data is collected?

- employee profiles
- performance
- data on high-performers
- data on low-performers
- salary and promotion history
- demographic data
- on-boarding
- training
- engagement
- retention
- turnover
- absenteeism

2. Measurement

At the measurement stage, the data begins a process of continuous measurement and comparison, also known as HR metrics.

HR analytics compares collected data against historical norms and organizational standards. The process cannot rely on a single snapshot of data, but instead requires a continuous feed of data over time.

The data also needs a comparison baseline. For example, how does an organization know what is an acceptable absentee range if it is not first defined.

In HR analytics, key metrics that are monitored are:

Organizational performance
Data is collected and compared to better understand turnover, absenteeism, and recruitment outcomes.

Operations
Data is monitored to determine the effectiveness and efficiency of HR day-to-day procedures and initiatives.

Process optimization
This area combines data from both organizational performance and operations metrics in order to identify where improvements in process can be made.

Examples of HR analytics Metrics

Here are some examples of specific metrics that can be measured by HR:

- Time to hire – The number of days that it takes to post jobs and finalize the hiring of candidates. This metric is monitored over time and is compared to the desired organizational rate.
- Recruitment cost to hire – The total cost involved with recruiting and hiring candidates. This metric is monitored over time to track the typical costs involved with recruiting specific types of candidates.
- Turnover – The rate at which employees quit their jobs after a given year of employment within the organization. This metric is monitored over time and is compared to the organization’s acceptable rate or goal.
- Absenteeism – The number of days and frequency that employees are away from their jobs. This metric is monitored over time and is compared to the organization’s acceptable rate or goal.
- Engagement rating – The measurement of employee productivity and employee satisfaction to gauge the level of engagement employees have in their job. This can be measured through surveys, performance assessments or productivity measures.

3. Analysis

The analytical stage reviews the results from metric reporting to identify trends and patterns that may have an organizational impact.

Figure 3 - Data analysis

There are different analytical methods used, depending on the outcome desired. These include: descriptive analytics, prescriptive analytics, and predictive analytics.

Descriptive Analytics is focused solely on understanding historical data and what can be improved.

Predictive Analytics uses statistical models to analyse historical data in order to forecast future risks or opportunities.

Prescriptive Analytics takes Predictive Analytics a step further and predicts consequences for forecasted outcomes.

Examples of Analytics:

Here are some examples of metrics at the analytics stage:

- Time to hire – The amount of time between a job posting and the actual hire is a metric that enables HR to gain insight into the efficiency of the hiring process; it prompts investigation into what is working and what is not working. Does it take too long to find the right candidate? What factors could be impacting the result?
- Turnover – Turnover metrics that indicate the rate at which employees leave the organization after hire can be analysed to determine what specific departments within the organization are struggling with retention and the possible factors involved, such as work environment dissatisfaction or lack of training support.

- Absenteeism – The metric indicating how often and how long employees are away from their jobs as compared to the organization's established norm could be an indicator of employee engagement. As absenteeism can be costly to the productivity of an organization, the metric enables HR to investigate the possible reasons for high absence rates.

APPLICATIONS

How can HR Analytics be used by organizations?

Let's take a look at a few examples using common organizational issues:

1. Turnover

Figure 4 - Turnover

When employees quit, there is often no real understanding of why. There may be collected reports or data on individual situations, but no way of knowing whether there is an overarching reason or trend for the turnover.

With turnover being costly in terms of lost time and profit, organizations need this insight to prevent turnover from becoming an on-going problem.

HR Analytics can:

- Collect and analyse past data on turnover to identify trends and patterns indicating why employees quit.
- Collect data on employee behaviour, such as productivity and engagement, to better understand the status of current employees.
- Correlate both types of data to understand the factors that lead to turnover.
- Help create a predictive model to better track and flag employees who may fall into the identified pattern associated with employees that have quit.
- Develop strategies and make decisions that will improve the work environment and engagement levels.
- Identify patterns of employee engagement, employee satisfaction and performance.

2. Recruitment

Figure 5 - Recruitment

Organizations are seeking candidates that not only have the right skills, but also the right attributes that match with the organization’s work culture and performance needs.

Sifting through hundreds or thousands of resumes and basing a recruitment decision on basic information is limiting, more so when potential candidates can be overlooked. For example,

one company may discover that creativity is a better indicator of success than related work experience.

HR Analytics can:

- Enable fast, automated collection of candidate data from multiple sources.
- Gain deep insight into candidates by considering extensive variables, like developmental opportunities and cultural fit.
- Identify candidates with attributes that are comparable to the top-performing employees in the organization.
- Avoid habitual bias and ensure equal opportunity for all candidates; with a data-driven approach to recruiting, the viewpoint and opinion of one person can no longer impact the consideration of applicants.
- Provide metrics on how long it takes to hire for specific roles within the organization, enabling departments to be more prepared and informed when the need to hire arises.
- Provide historical data pertaining to periods of over-hiring and under-hiring, enabling organizations to develop better long-term hiring plans

Here are some examples of how to apply the analysis gained from HR analytics to decision-making:

- Time to hire – If findings determine that the time to hire is taking too long and the job application itself is discovered to be the barrier, organizations can make an informed decision about how to improve the effectiveness and accessibility of the job application procedure.
- Turnover – Understanding why employees leave the organization means that decisions can be made to prevent or reduce turnover from happening in the first place. If lack of training support was identified as a contributing factor, then initiatives to improve on-going training can be put together
- Absenteeism – Understanding the reasons for employee long-term absence enables organizations to develop strategies to improve the factors in the work environment impacting employee engagement.

Examples of how to apply HR analytics insights:

Here are some examples of how to apply the analysis gained from HR analytics to decision-making:

- Time to hire – If findings determine that the time to hire is taking too long and the job application itself is discovered to be the barrier, organizations can make

an informed decision about how to improve the effectiveness and accessibility of the job application procedure.

- Turnover – Understanding why employees leave the organization means that decisions can be made to prevent or reduce turnover from happening in the first place. If lack of training support was identified as a contributing factor, then initiatives to improve on-going training can be put together.

- Absenteeism – Understanding the reasons for employee long-term absence enables organizations to develop strategies to improve the factors in the work environment impacting employee engagement.

DATA SET AND TECHNIQUES USED

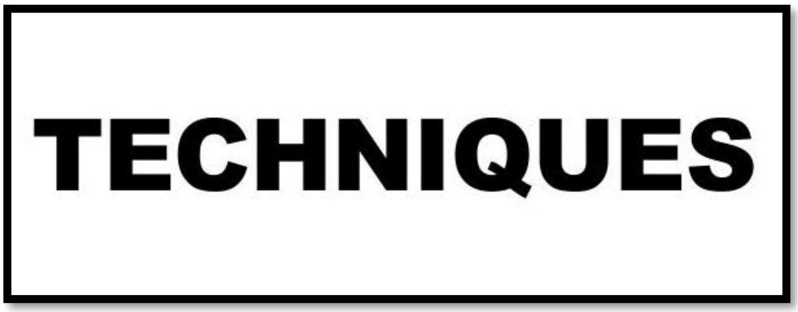


Figure 6 - Techniques

I utilized the dataset of IBM, about their employee data with the attributes like;

- Age: Numerical Discrete Data
- Attrition: Text Categorical Data
- Business Travel: Text Categorical Data
- Daily Rate: Numerical Discrete Dat
- Department: Text Categorical Data
- Distance From Home: Numerical Discrete Data
- Education: Numerical, Categorical Data
    - 1: Below College
    - 2: College
    - 3: Bachelor
    - 4: Master
    - 5: Doctor
- Education Field: Text Categorical Data

- Employee Count: Numerical Categorical Data
- Employee Number: Numerical Categorical Data
- Environment Satisfaction: Numerical Categorical Data
    - 1: Low
    - 2: Medium
    - 3: High
    - 4: Very High
- Gender: Text Categorical Data
- Hourly Rate: Numerical Discrete Data
- Job Involvement: Numerical Categorical Data
    - 1: Low
    - 2: Medium
    - 3: High
    - 4: Very High
- Job Level: Numerical Categorical Data
- Job Role: Text Categorical Data
- Job Satisfaction: Numerical Categorical Data
    - 1: Low
    - 2: Medium
    - 3: High
    - 4: Very High
- Marital Status: Text Categorical Data
- Monthly Income: Numerical Discrete Data
- Monthly Rate: Numerical Discrete Data
- Number Companies Worked: Numerical Discrete Data
- Over 18: Text Categorical Data
- Over Time: Text Categorical Data
- Percent Salary Hike: Numerical Discrete Data
- Performance Rating: Numerical Categorical Data
    - 1: Low
    - 2: Good
    - 3: Excellent
    - 4: Outstanding
- Relationship Satisfaction: Numerical Categorical Data
    - 1: Low
    - 2: Medium
    - 3: High
    - 4: Very High
- Standard Hours: Numerical Discrete Data
- Stock Option Level: Numerical Categorical Data
- Total Working Years: Numerical Discrete Data

- Training Times Last Year: Numerical Discrete Data
- Work Life Balance: Numerical Categorical Data
    - 1: Bad
    - 2: Good
    - 3: Better
    - 4: Best
- Years At Company: Numerical Discrete Data
- Years In Current Role: Numerical Discrete Data
- Years Since Last Promotion: Numerical Discrete Data
- Years as Manager: Numerical Discrete Data

All the techniques I used which are important for data analytics are:

Exploratory Data Analytics (EDA), Summary of EDA, Feature Engineering (Data Balancing using SMOTE and Data leakage), and Modelling.

Exploratory Data Analytics

It's a technique also referred as EDA, which is utilized to represent the data and its relation in an illustrational manner. These are done using libraries of python language like pandas, Seaborn, Matplotlib and more.

Pandas

An open source widely used library used for data analysis. Used for writing and reading data from file sources and data handling.

Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics

Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes representations easy and reader friendly.

Summary of EDA

Tried to point out the summarized texts from what I get from explorations of the data set.

Feature Engineering

Figure 7 – Feature Engineering

I used feature engineering to make the necessary attributes to a desired data format. Feature engineering or feature extraction or feature discovery is the process of using domain knowledge to extract features (characteristics, properties, attributes) from raw data.

Data Balancing using SMOTE

Imbalanced data is a common problem in machine learning, which brings challenges to feature correlation, class separation and evaluation, and results in poor model performance, this may create oversampling problem in our data.

SMOTE (synthetic minority oversampling technique) is one of the most commonly used oversampling methods to solve the imbalance problem.

Data Leakage

Data leakage is one of the major problems in machine learning which occurs when the data that we are using to train an ML algorithm has the information the model is trying to predict. It is a situation that causes unpredictable and bad prediction outcomes after model deployment.

Modelling

The necessary models for computing the machine learning process are;

train_test_split

The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets.

feature_selection

In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction

Data scaling(pre-processing)

Data Scaling is a method of standardization that’s most useful when working with a dataset that contains continuous features that are on different scales

XGBoost

XGBoost is an extension to gradient boosted decision trees (GBM) and specially designed to improve speed and performance.

LGBM classifier (and regression)

Machine Learning LightGBM is a gradient boosting classifier in machine learning that uses tree-based learning algorithms. It is designed to be distributed and efficient with faster drive speed and higher efficiency, lower memory usage and better accuracy. LightGBM can be used for regression, classification, ranking and other machine learning tasks

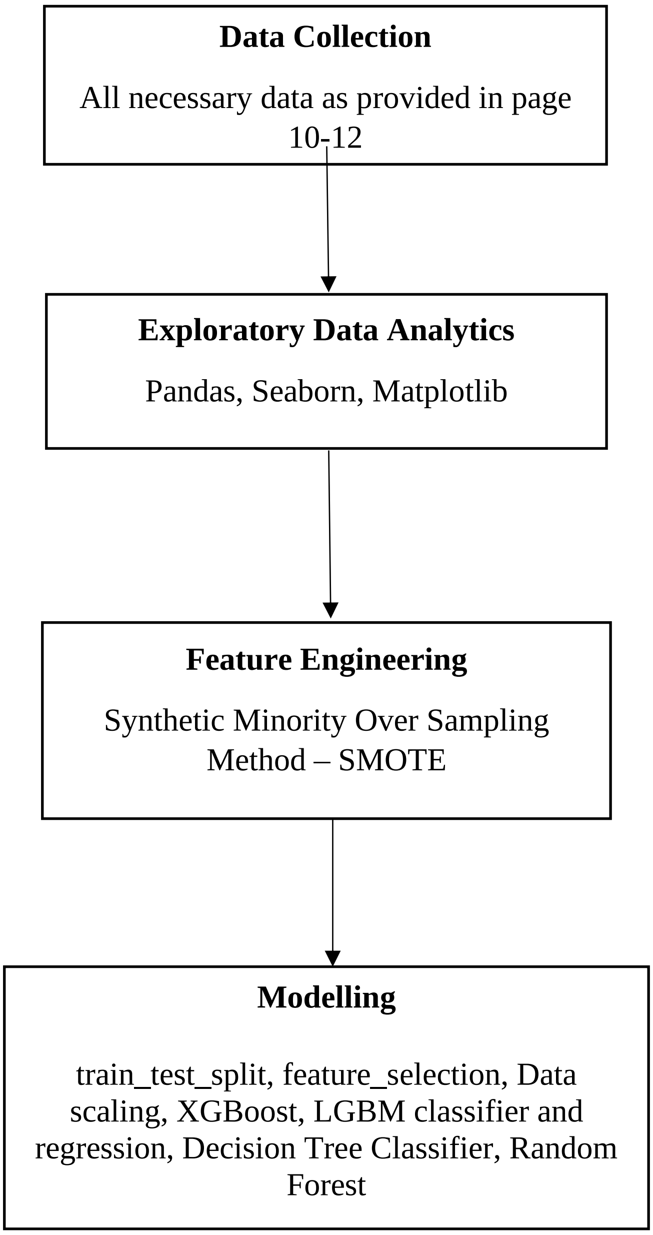
Decision Tree Classifier

Decision Tree is a Supervised Machine Learning Algorithm that uses a set of rules to make decisions, similarly to how humans make decisions.

Random Forest

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

METHODOLOGY DIAGRAM



Data Collection

All necessary data as provided in page 10-12

Exploratory Data Analytics

Pandas, Seaborn, Matplotlib

Feature Engineering

Synthetic Minority Over Sampling Method – SMOTE

Modelling

train_test_split, feature_selection, Data scaling, XGBoost, LGBM classifier and regression, Decision Tree Classifier, Random Forest

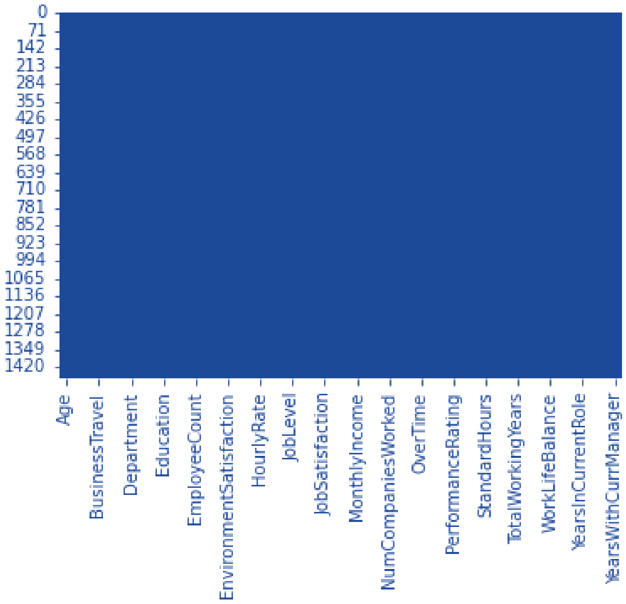
Figure 8 – Methodology diagram

EXPERIMENTAL STUDY

All the data features may not perform equally in all study cases and analysis. Each feature may exhibit each behavioural pattern. So not all the features are extremely good together and some of the feature together would be good. For example, employee age and job engagement are positively correlated. So, this together data might be very helpful for analysing.

The heat map of the data set is found to be optimal without any null values after the pre-processing stage as shown in figure 2.

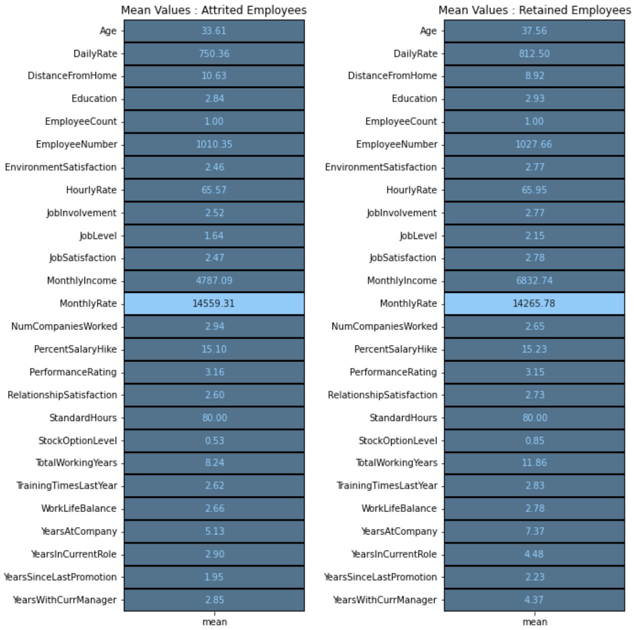
Figure 9 - Heatmap



MEAN VALUES

The mean values of all the features are analysed. It is then divided to cases of Retention and Attrition faced Employees as shown in figure 3.

Figure 10- Mean values



When considering the feature age, mean values of staying employees are 37.56. Meaning, more than the leaving employees, that is 33.61. Similarly, Job level and Daily rate is higher for staying employees than leaving employees. It is also noticeable that staying employees have higher values for features like – Years at company, Years with current manager, Total working years and Years in current role.

DISTRIBUTION OF CATEGORICAL FEATURES

Distribution of features are analysed to check the value distribution and skewness.

Figure 11 Distribution of Attrition

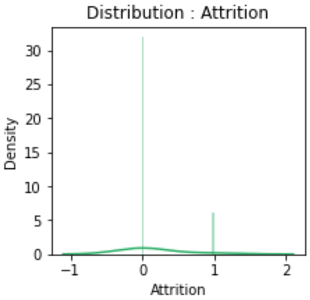


Figure 12 Distribution of Business Travel

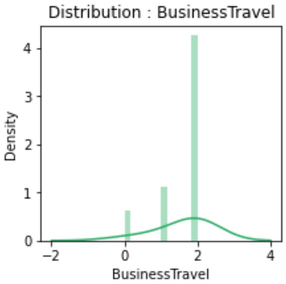


Figure 13 Distribution of Department

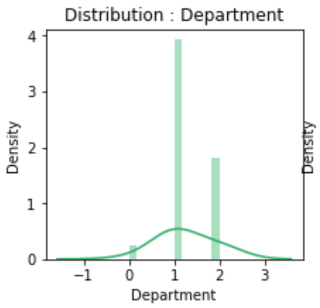


Figure 14 Distribution of Education

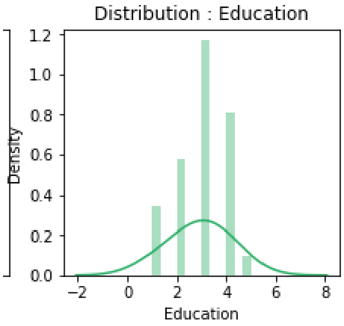


Figure 15 Distribution of Employee count

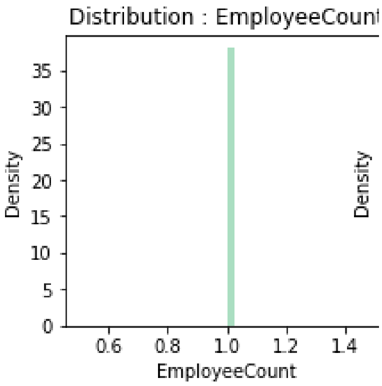


Figure 16 Distribution of Employee Number

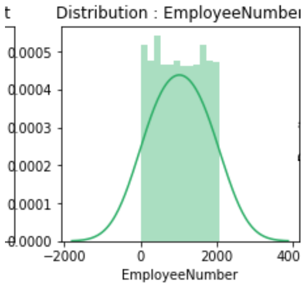
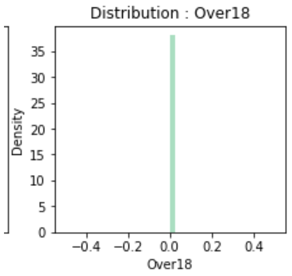


Figure 17 Distribution of over 18



From the above figures, it is clear that Employee Number is just a unique identifying number with no repetitive elements. Hence, it should be dropped. A bimodal distribution can be observed for the Job role. Over 18 and Employee count are single value features. Attrition is the target variable, so it could be considered later after dropping.

DISTRIBUTION OF DISCRETE FEATURES

Figure 18 Distribution of Hourly Rate

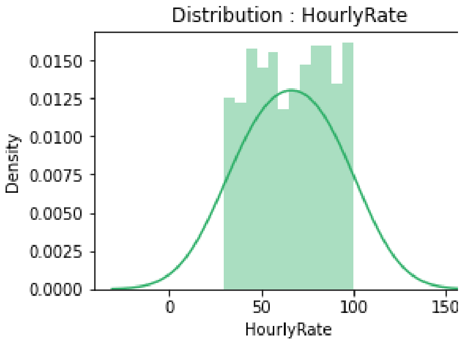


Figure 19 Distribution of Monthly Income

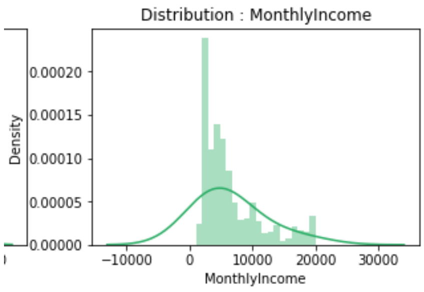


Figure 20 Distribution of Monthly Rate

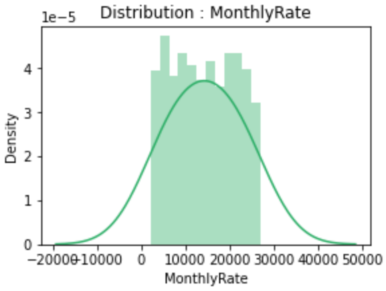


Figure 21 Distribution of Distance from home

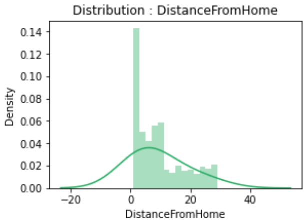


Figure 22 Distribution of Number of companies worked

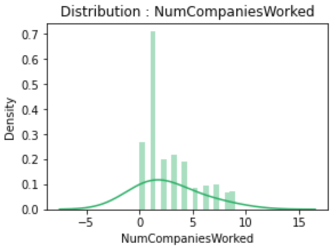


Figure 23 Distribution of Salary hike

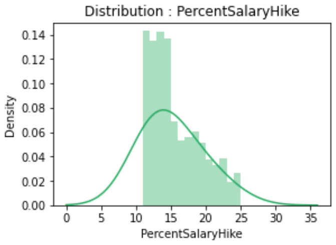


Figure 24 Distribution of Total working years

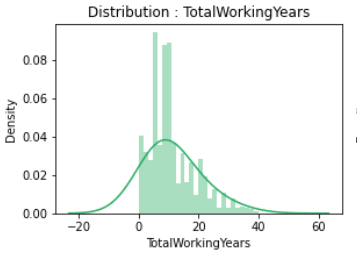


Figure 25 Distribution of Years at company

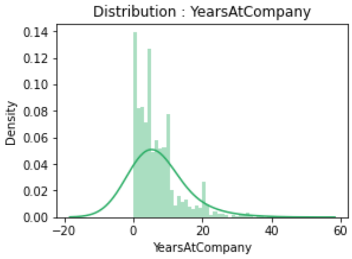


Figure 26 Distribution of Years since last promotion

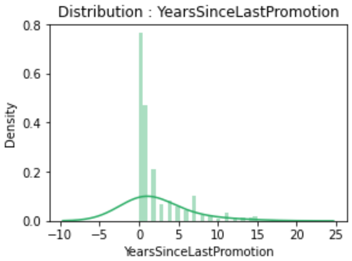


Figure 27 Distribution of Years in current role

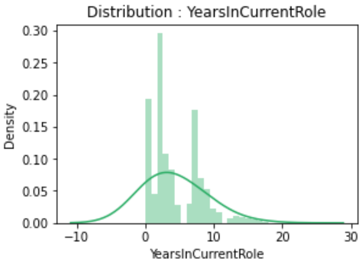
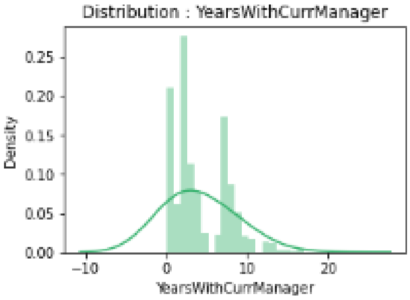


Figure 28 Distribution of Years with current manager



The hourly, Monthly and Daily Rate display graphs that are usually found in time series. These graphs change with respect to time. Distance from home, Monthly income, Number of companies worked, Percentage salary hike, Total working years, Years at company and Year since last promotion shows a rightly skewed data distribution. Years in current role and Years with current manager have a bimodal data distribution. Standard hours is a single value feature.

TARGET VARIABLE VISUALIZATION (ATTRITION)

Figure 29 Plot of Attrition

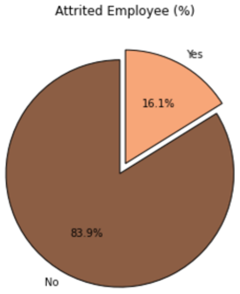
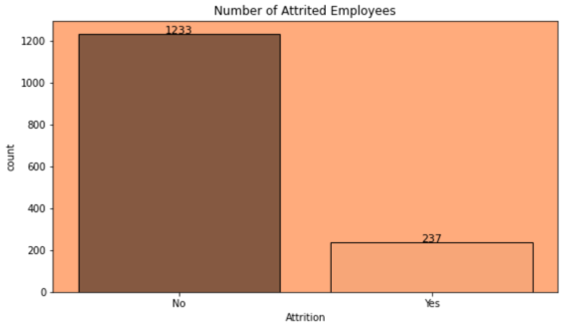
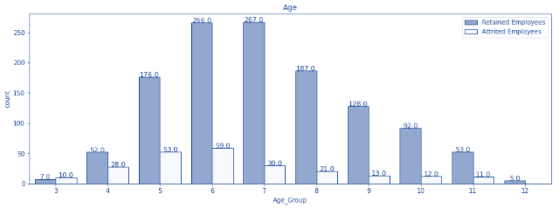


Figure 30 Histogram of Attrition



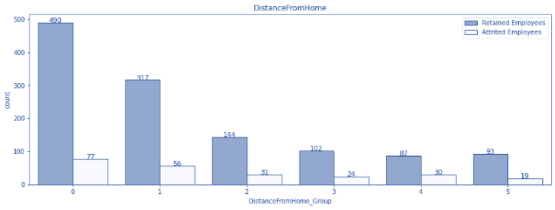
The data set is unbalanced highly, the ratio is found 5.2:1 ratio for Retained: Attrited Employees. This makes predictions biases towards the retention cases. Removing target feature and grouping other features as shown in below figures.

Figure 31 Attrition – Age Relational plot



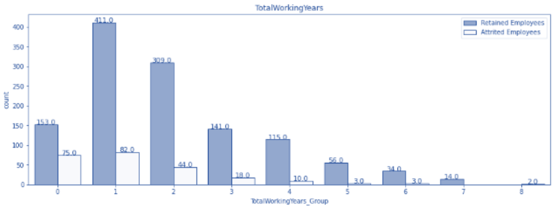
Attrition is found in all the age groups. Between group 30 to 34, highest number of employees left the organization.

Figure 32 Attrition – Distance from home Relational plot



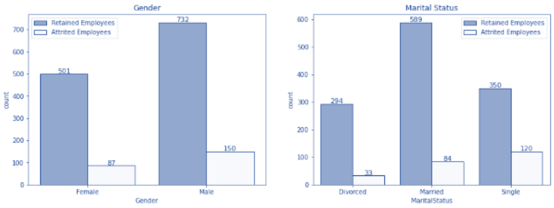
The employees living away from organization have been attrited the most and employees living near, attrited the least.

Figure 33 Attrition – Retained employee Relational plot



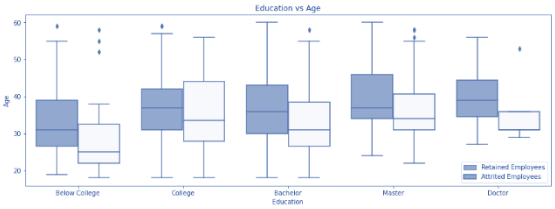
From figure 25, it is clear that employees with in first ten years of experience are highly prone to being removed. As the experience increases, chances of attrition reduce.

Figure 34 Attrition – Retained marriage Relational plot



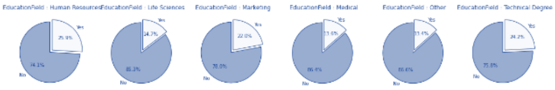
More male employees have been removed than the female employees. Single employees have been faced attrition the most.

Figure 35 Education v/s Age boxplot



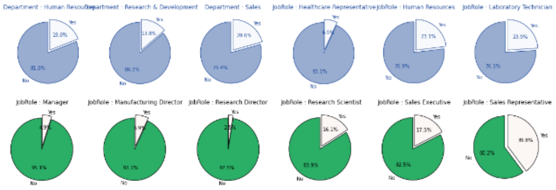
When it comes to education, employee’s college degree has a high range of age values, making them more prone to removal from the organization. The higher education degrees like Masters, Doctorate employees are the least prone to the departure from the company.

Figure 36 Attrition to Education field ratio



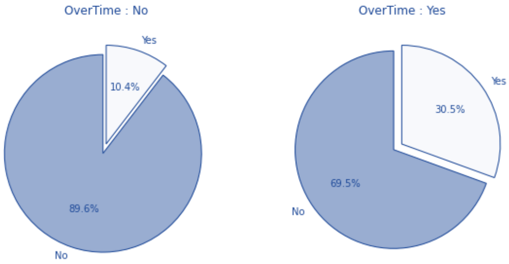
Employees with Educational background in the field of degrees like Technical, Human resources and marketing have a high chance of being removed.

Figure 37 Attrition to Education field ratio



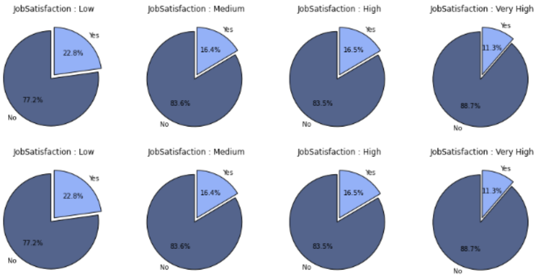
According to the pie charts above, personnel in the sales and human resources departments are more likely to leave the company than those in the research and development department.
In terms of Job Role, 4 of the 9 jobs show an attrition rate of less than 7%, while the remaining 5 positions show an attrition rate of more than 15%.

Figure 38 Attrition to over time ratio



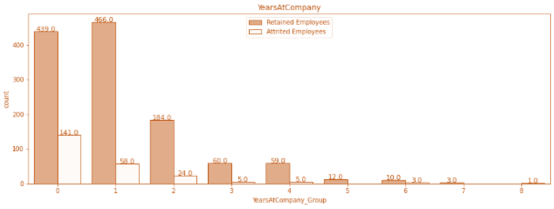
We can observe that those who put in overtime are more likely to lose their jobs. It has a 30% attrition rate, which is significantly lower than that of employees who don't put in overtime.

Figure 39 Attrition to Involvement ratio



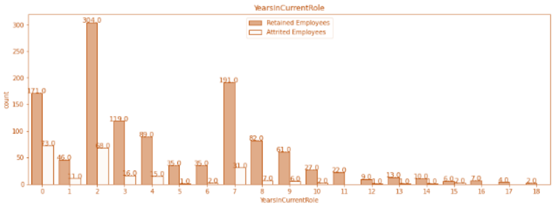
We can see that the attrition rate decreases as job involvement increases.
It is possible to see a similar pattern for job satisfaction.

Figure 40 Attrition to years at company relation



It is obvious that workers who have worked with the company for 0 to 4 years have been let go the most frequently.
Attrition declines when employees gain experience at the organisation.

Figure 41 Attrition to years at current role

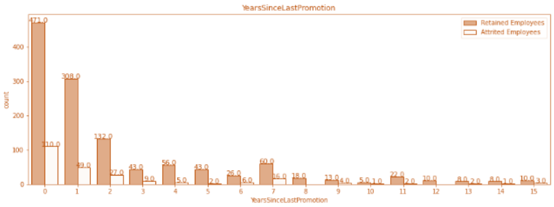


Employees in their first role are quite volatile and seek an early leave, as is now to be expected. Employee attrition also increases when they have been in their current position for two years. Either individuals are trying to get better at their jobs, or employers have evaluated their workforce and made a decision.

After that, there is attrition in years three and four. This is most likely an extension of the attrition from year 2.

In year seven of their current position, a further big increase can be seen as the personnel may seek advancement or the corporation may opt to make changes.

Figure 42 Attrition to years since last promotion



We can observe that there is a tonne of attrition situations for value 0. The majority values, I suppose, are those of the company's new hires.

A considerable number of employee dismissal cases have been reported in the 1 and 2 years since the last promotion.

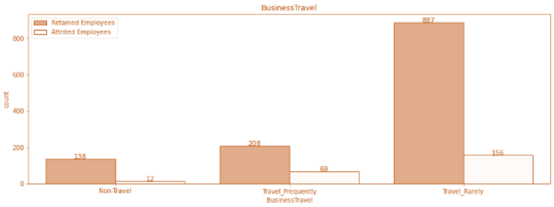
A respectable number of employee dismissal cases have occurred in the seven years since the last promotion. Years in Current Role & Years with Current Manager, the preceding 2 graphs, and this value appear to be somewhat correlated.

Similarly, a bad work life balance too results in a huge attrition percentage of 31.2%.

For all of the Work Life Balance values, laboratory technicians, research scientists, and sales executives have recorded high numbers.

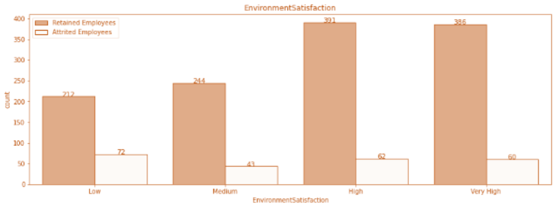
The percentage of attrition employees declines as the salary increase percentage rises.

Figure 43 Attrition to Travel



Employees at Travel Frequently have a 25% chance of leaving the organisation in terms of attrition rate.

Figure 44 Attrition to Environment satisfaction



The most frequently noted Environment Satisfaction values are High & Very High.

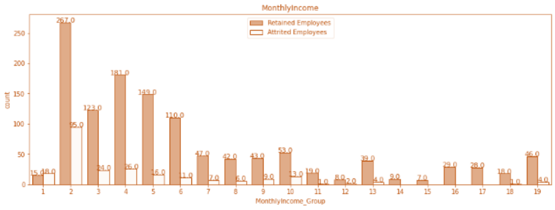
They have a lower attrition rate than Low & Medium Environment Satisfaction, as was to be expected.

As the environment is more favourable, the attrition rate rises.

Environment Satisfaction and Relationship Satisfaction are extremely similar.

Attrition rate decreases as Relationship Satisfaction values increase.

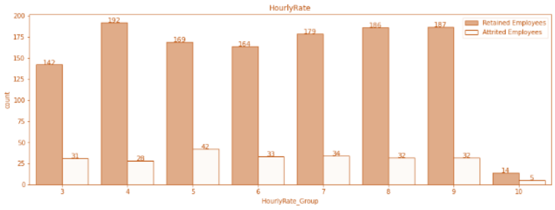
Figure 45 Attrition to Monthly income



The graph of figure 37 shows that the count of the values has generally decreased.

There are a lot of monthly income amounts between 1000 and 2000. Values in the range of 3000 to 4000 come in second, with more than 200 values being present.

Figure 46 Attrition to Hourly Rate



Values between 30 and 100 are present for Hourly Rate, with a count of greater than 175 for each.

These values have a low attrition rate and are quite similar to one another.

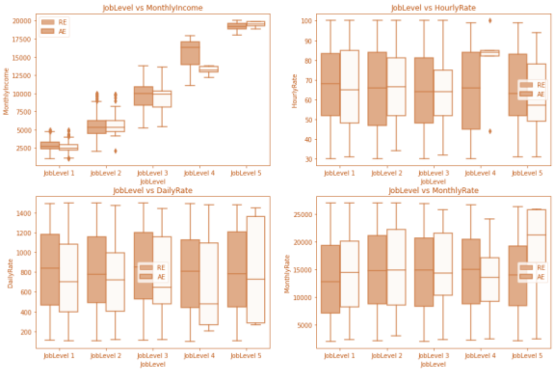
Because there are so few results for Hourly Rate values greater than 100, attrition is also high.

There are too many outlier numbers in Monthly Income. These anomalies most likely come from Job Level 5, which has a low attrition rate and low numbers.

Research & Development and Sales departments for Hourly Rate essentially fall within the same range of values for attrition and non-attrition. There is a fairly narrow range of attrition values for human resources.

Similar to Hourly Rate, the Research, Development, and Sales departments' Daily Rate and Monthly Rate patterns can be seen.

Figure 47- Job level v/s Monthly income, Hourly rate, Daily rate, Monthly rate



As might be expected, monthly income rises with job level! The preceding Job Level value's upper limit value is less than the subsequent Job Level value's lower limit value.

The upper and lower limits of the hourly rate by job level are extremely close to one another. There isn't enough of a difference to distinguish it as monthly income.

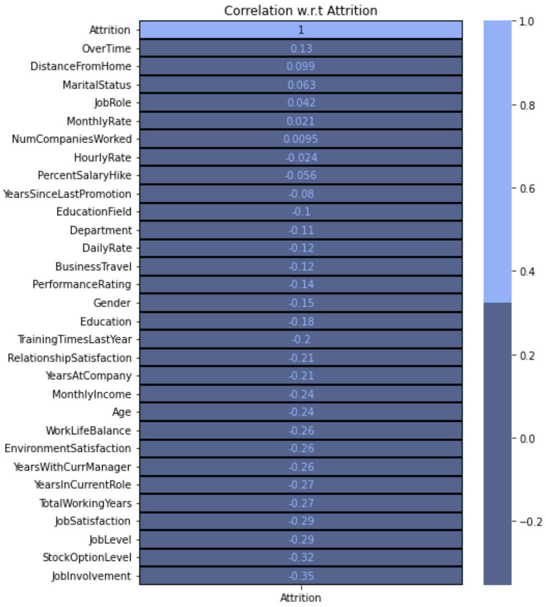
The daily rate and monthly rate are essentially the same. The maximum Daily Rate and Monthly Rate for Job Level 5 are distinct from one another.

CORRELATION MATRIX

We establish a new data frame in order to display the correlation matrix.

To prevent data leakage, we therefore discard anything not contained in the training data.

Figure 48 Correlation Matrix

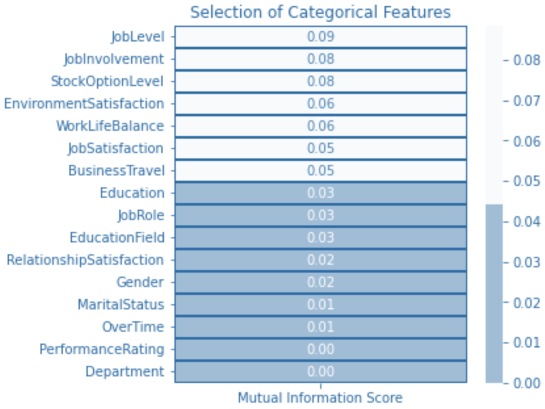


FEATURE SELECTION – CATEGGORICAL VALUES

From figure 40, There is no discernible strong positive or negative link between any of the variables and attrition. Values for the majority of the characteristics range from [-0.3 to 0.14].

MUTUAL INFORMATION TEST

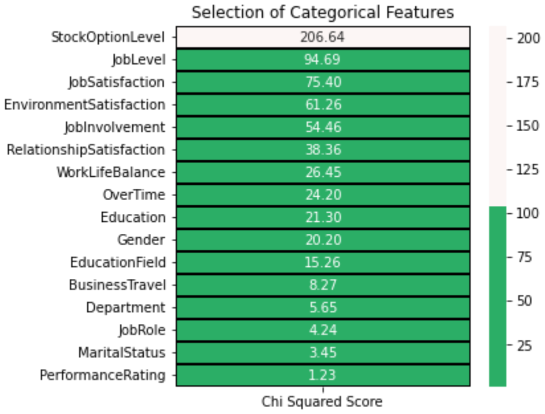
Figure 49 Mutual information test



With categorical features, the Mutual Information Score of Attrition shows extremely low results. The aforementioned scores indicate that none of the traits should be used for modelling.

CHI SQUARED TEST

Figure 50 Chi squared test



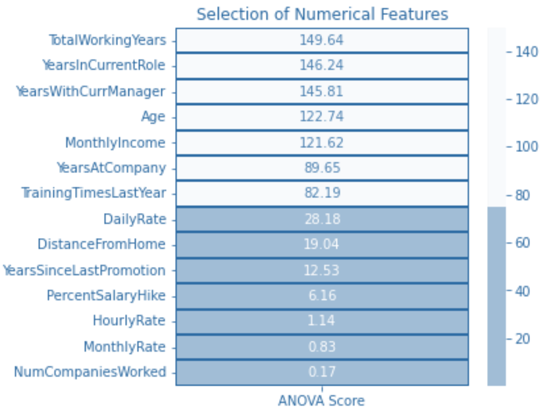
We shall remove the following characteristics from the aforementioned Chi Squared Score Test: Performance Rating, Department, Job Role, Education Field, Business Travel, Marital Status, and Gender.

FEATURE SELECTION – NUMERICAL VALUES

Similarly, as categorical value test is undergone, the test for numerical values should be done. The tests used here ANOVA test.

ANOVA TEST

Figure 51 ANOVA Test



We will eliminate the following features from the aforementioned ANOVA Score Test: Monthly Rate, Hourly Rate, Number of Companies Worked, Percent Salary Increase, Years Since Last Promotion, Distance from Home, and Daily Rate.

By removing the characteristics based on the aforementioned statistical tests, we prepare the datasets for data scaling.

CONCLUSION

HR analytics is fast becoming a desired addition to HR practices.

Data that is routinely collected across the organization offers no value without aggregation and analysis, making HR analytics a valuable tool for measured insight that previously did not exist.

But while HR analytics offers to move HR practice from the operational level to the strategic level, it is not without its challenges.

Using HR analytics, we can make;

More accurate decision-making

Strategies to improve retention

Employee engagement can be improved by analysing data about employee behaviour, such as how they work with co-workers and customers, and determining how processes and environment can be fine-tuned.

Recruitment and hiring can be better tailored to the organization's actual skillset needs by analysing and comparing the data of current employees and potential candidates.

What so ever, many HR departments lack the statistical and analytical skillset to work with large datasets. Also, different management and reporting systems within the organization can make it difficult to aggregate and compare data. HR Analytics in one case can make all the difficulties overcome, this will be again fine-tuned by implementing new age technologies like cloud, big data and more.

FUTURE ENHANCEMENT

Every project needs a future enhancement. The future enhancement of this project is adding a new feature. The features are always good in the data set. But, following the recession in 2022, employees mainly form IT field are getting lay-offs. Lay – offs are usually very misunderstood these days. This misunderstanding would lead to panic in young working community. After detailed research, I'm concluding that layoffs are done only for;

Bench people, Employee without skill upgrade, Employees working on same technology for five plus years, Employees doing parallel jobs, Employees working from home. Including these much of features would help this project far more.