INTERNSHIP: PROJECT REPORT

Internship Project Title	RIO 45 - Employee Attrition & Performance Analysis
Name of the Company	TCS iON
Name of the Industry Mentor	Debashis Roy
Name of the Institute	ICT Academy, Kerala

Start Date	End Date	Total Effort (hrs.)	Project Environment	Tools used
23-07-2023	20-08-2023	45	Windows	Microsoft Power BI, MS Word

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank **TCSiON** for this wonderful virtual internship experience. This is a great chance for me to explore how an industry project works and understand its different stages. The learning platform provided by **TCSiON** greatly guides the students and provides a detailed plan.

I would also extend my gratitude to ICT Academy, Thiruvananthapurm in making way for the internship opportunity.

OBJECTIVE

The objective of the project is to develop analytics and reports using Data Science tools to provide detailed insights on HR Analytics focusing on **employee attrition** and **performances**.

a) To understand and analyse factors that lead to employee attrition.

b) To analyse factors affecting employee performance and means to optimize the same.

INTRODUCTION TO INTERNSHIP

The project is titled "**RIO-45: Employee Attrition and Performance Analysis**". It is a virtual learning experience on industry projects for 45 hours. This project aims to develop analytics and reports using Data Science tools to provide comprehensive insights on HR Analytics focusing on employee attrition and performances. These insights will help to optimize employee performance and control attrition.

Low attrition rate is preferred by many companies because it implies that employees are satisfied with their workplace and the company does not have to hire any new resources. The attrition may be caused by several reasons including job satisfaction, salary, nature of work, job level, environment, etc to name a few. Attrition can reduce the overall performance of an organization and incur huge costs. The teams may lack experienced people and find it difficult to manage various tasks. Thus, it is important that organizations monitor employee performance, track attrition and takes steps to control it, if required.

The project comprises of all the steps needed to work on the HR dataset to provide insights on Employee Attrition and performance management using tools like Tableau/Power BI. Here we use Microsoft Power BI as the data visualization and analysis tool.

INTERNSHIP ACTIVITIES

The internship consists of working on industry projects for 45 hours. The different learning stages and activities are as follows:

- Pre-Project Test – a simple test to check the learner's knowledge level.
- Project References – references to various learning resources for self-learning.
- Knowledge sharing – digital discussion rooms to make live interactions between learners and between the learner and the mentor.
- Industry Project – the industry project on some trending topic, where you will be working on.
- Activity Reports – The daily activity reports for tracking the day-wise progress of learner, which comprises of learning, project activities and interactions.
- Project Report – This report is submitted towards the end of the project and consists of report findings, outcomes, activities done etc.
- Project Test and Viva – these are conducted towards the end of the project, to test the technical knowledge gained from the internship journey.

SUMMARY OF DATA

The dataset consists of consists of 1470 observations (rows) of employee information and 35 features (variables). This fictional data is created by IBM scientists. There are no cases of any missing data. The dataset dealt with different datatypes. Attrition is the label in the dataset and we would like to find out why employees are leaving the organization. 1237 (84% of cases) employees did not leave the organization while 237 (16% of cases) did leave the organization making the dataset to be considered imbalanced since more people stay in the organization than they leave.

The various attributes and their types are listed below:

Feature	**Data Type**
Age	Numerical Discrete Data
Attrition	Text Categorical Data
Business Travel	Text Categorical Data
Daily Rate	Numerical Discrete Data
Department	Text Categorical Data
Distance From Home	Numerical Discrete Data
Education	Numerical, Categorical Data
Education Field	Text Categorical Data
Employee Count	Numerical, Categorical Data
Employee Number	Numerical, Categorical Data
Environment Satisfaction	Numerical, Categorical Data
Gender	Text Categorical Data
Hourly Rate	Numerical Discrete Data
Job Involvement	Numerical, Categorical Data
Job Role	Text Categorical Data
Job Satisfaction	Numerical Categorical Data
Marital Status	Text Categorical Data
Monthly Income	Numerical Discrete Data
Monthly Rate	Numerical Discrete Data

APPROACH

The first step in designing the project consisted of gaining background information about the dataset to be worked on. The attributes were studied and their types were classified. The next step was analyzing those that contributed to the decision-making process. The key pointers were noted and an idea of how to design the reports were made. The logic flow was designed. The visualizations that could best depict each pointer was decided. The number of pages in the report and visuals on each page were finalized.

To start on building visualizations, first the data set was loaded into Microsoft Power BI Desktop and the column quality was checked. This was followed by pre-processing of the data. Data encoding was performed. The key columns that affect the decision-making were analyzed again.

Initially the KPIs were created in the reports. For this, the required measures and calculated columns were created. Next the rest of visuals were designed. This consisted of clustered bar charts, stacked bar charts, line charts, funnel charts, tree maps, pie-charts, doughnut charts, and area charts. Once the visuals were completed, the relevant insights were summarized in the last page of the report. Finally, the various steps in the project work were documented in the project report.

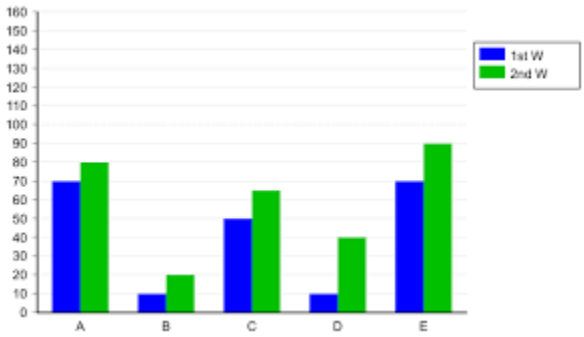
EXCLUSIONS

While analyzing the dataset, two attributes were found that did not contribute much to the analysis and contained the same values for all employees. These attributes were StandardHours, and Over18, which contains numerical discrete data and text categorical data respectively. Hence, they were not used in the visualizations.

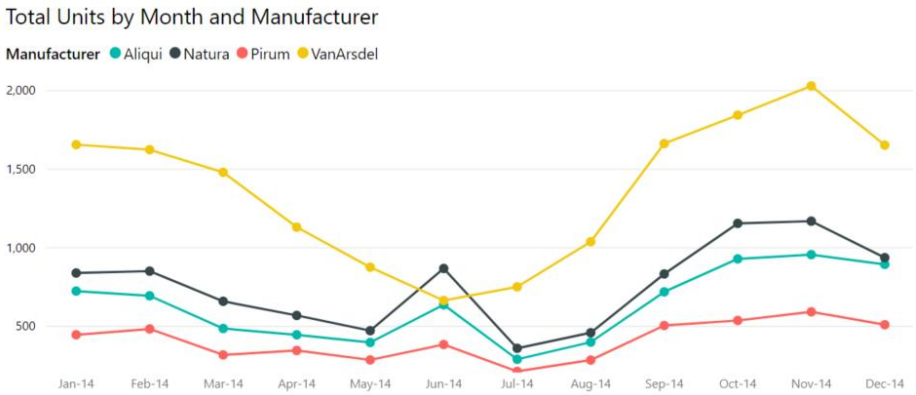
CHARTS AND DIAGRAMS

A variety of charts and diagrams were used in building visualizations of the project. They are listed below:

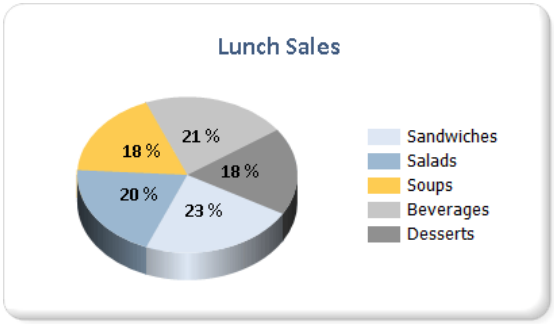
- Bar Charts - Bar charts organize data into rectangular bars that make it easy to compare related data sets. Mostly when you want to compare two or more values in the same category, or you want to compare parts of a whole.



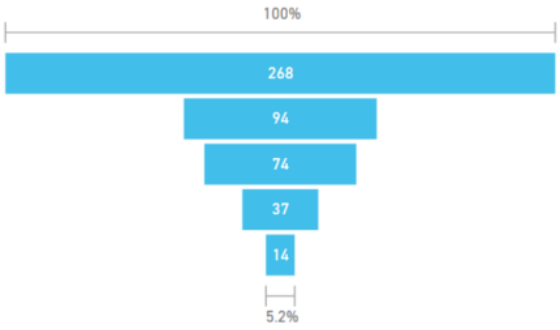
- Line charts – Line charts are used to show resulting data relative to a continuous variable - most commonly time or money. Line charts are helpful to understand trends, patterns, and fluctuations in your data. They are used when you want to compare different yet related data sets with multiple series. Line charts make projections beyond your data and thus are used as trend graphs for forecasting purposes. However, they do not demonstrate an in-depth view of your data.



- Pie-charts – Pie-charts allow you to compare relative values. You use them when you want to compare parts of a whole.

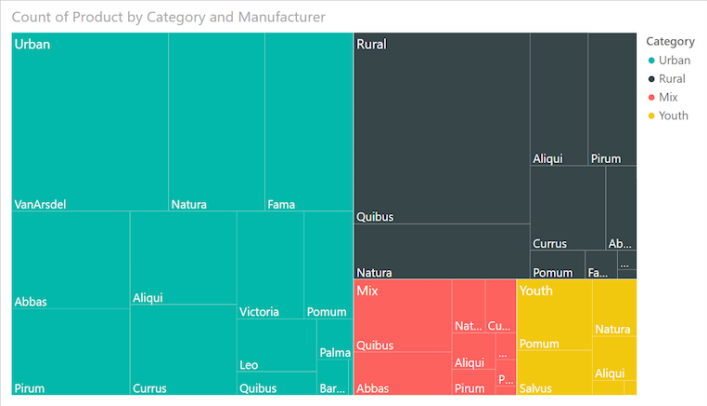


- Funnel Chart - A funnel chart is your data visualization of choice if you want to display a series of steps and the completion rate for each step. Funnel charts are most often used to represent how something moves through different stages in a process. A funnel chart displays values as progressively decreasing proportions amounting to 100 percent in total.



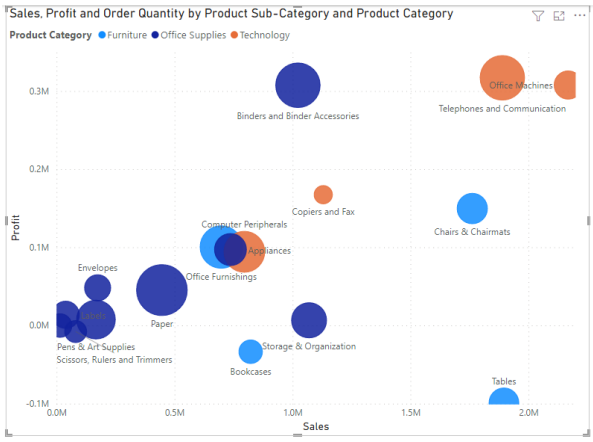
- Slicers - Slicers are another way of filtering. They are displayed on the report page, and narrow the portion of the dataset that's shown in the other report visualizations. Slicers are a great choice when you want to display commonly used or important filters on the report canvas for easier access. They make it easier to see the current filtered state without having to open a drop-down list. By putting slicers next to important visuals, we can create more focused reports.

- TreeMaps - A treemap is a visual tool that can be used to break down the relationships between multiple variables in your data. They can be used strictly as a presentation vehicle to show how your products roll up into different categories.

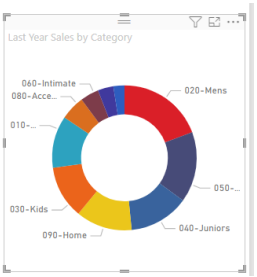


Treemap

- Scatter Plot - Scatterplots are the right data visualizations to use when there are many different data points, and you want to highlight similarities in the data set. This is useful when looking for outliers or for understanding the distribution of your data.



- Doughnut Chart - A doughnut chart is like a pie chart in that it shows the relationship of parts to a whole. The only difference is that the center is blank and allows space for a label or icon.



MEASURES

Various measures and calculated columns have been created to aid in the analysis. They are listed below:

- Age Group – calculated column, for categorizing employees according to their age.
- Salary Slab - calculated column, for categorizing employees according to their monthly income.
- Attrition Rate – measure for finding the attrition rate.
- Attrition count – calculated column, for encoding attrition labels.
- Average Age – measure for finding the average age of employees leaving the organization.
- Avg Years at Company – measure for calculating the average of years spent by resigned employees in the organization.
- AvgMonthlyIncomeAttrition – measure for calculating the average monthly income of resigned employees.
- AvgMonthlyIncomeNoAttrition – measure for calculating the average monthly income of current employees.
- CurrentEmployees – measure for calculating the number of current employees in the company.
- Ex-Employees - measure for calculating the number of exited employees in the company.
- MinimumHrlyRate - measure for calculating the minimum hourly rate of exited employees in the company.
- Oldest Age - measure for finding the oldest among exited employees in the company.
- Youngest Age - measure for finding the youngest among exited employees in the company.
- TotalEmployees – measure for calculating the total number of employees in the company.

VISUALIZATIONS AND INSIGHTS

Visualizations :-

The report consists of three pages. The last page summarizes the relevant insights from the visualizations. The various visualizations used in the report are listed as follows:

- Cards – The card visuals are used to depict the various KPIs. These include:
  - ✓ Attrition rate
  - ✓ Employee count
  - ✓ Current employees
  - ✓ Ex-employees
  - ✓ Average job satisfaction
  - ✓ Average years at company
  - ✓ Youngest age
  - ✓ Average age of attrition
  - ✓ Average monthly income of current employees
  - ✓ Minimum hourly rate
- Pie-chart – Pie-chart represents the attrition based on gender.
- Stacked column chart – This represents the attrition based on the different age groups and gender.
- Area chart – This shows the distribution of attrition rate across different job roles and by average of salary.
- Line chart – This shows the trend of attrition based on distance from home and gender.
- Stacked column chart – This shows the distribution of attrition based on percent salary hike and education.
- Stacked line chart – This represents the attrition by job satisfaction.
- Area chart – This shows the attrition according to job role.
- Stacked bar chart – This represents the pattern of attrition according to monthly income and gender.
- Funnel chart – This shows the attrition distribution based on education field.
- Doughnut charts – A set of donought charts were used to represent the effect of the following factors on attrition: environment satisfaction, business travel, performance rating and overtime.

- Slicers – Various slicers are used in the reports for obtaining useful insights. They are based on gender, departments, and age groups.
- Page navigators – Page navigators are used to navigate between the different pages in the report. The first two pages of the report covers the various visualizations. The last page provides a summary of useful insights from the visualizations.

Insights :-

Various insights could be analyzed from the reports. Few relevant ones are summarized below:

- The attrition rate of the organization is 16.12%. Out of the total 1470 employees in the organization, 237 of them have quitted their jobs and 63.29 % of them are males.
- Among all the departments, the Sales has the highest attrition rate, followed by HR and then R&D.
- Employees with attrition have a lower monthly income. Moreover, the employees getting higher salary hikes, such as sales representatives, tend to quit their jobs.
- Employees with lower job level are more likely to leave the job.
- As job level increases, the number of employees who want to leave the company decreases.
- Job roles such as Research Director, Manager, Manufacturing Director etc work in company for many long years and have very low attrition rates. They leave the company at their retirement ages.
- Job roles such as Sales representative, Laboratory Technician etc. with lower salaries have very high attrition rates.
- Attrition is highest among the age group of 25-34 years and decreases as people gets aged and experienced. This is because the younger people tend to quit their jobs either for better opportunities or higher studies.
- Most employees with attrition have high job satisfaction and come from Life Sciences field.
- The average job satisfaction score is 2.73, on a scale of 5.
- Most employees spent an average of 7 years at the company.
- Employees staying near to the company and having no overtime are more likely to quit.
- An alarming fact is that most employees with excellent ratings are likely to quit their jobs.
- Employees with attrition tend to have low environment satisfaction and travels rarely.
- With lower education level, employees have higher attrition rates.

CHALLENGES & OPPORTUNITIES

Challenges:

The major task was to understand the background of work and the nature of different attributes present in the dataset. Besides, pre-processing of data and making it suitable for report making was another task that required quite efforts.

Opportunities:

The project presents excellent opportunity to work on industry projects and understand their working. It helps us analyze the various factors that optimize employee performance and control attrition in an organization. Thus, this will help the organization take better steps that could retain employees and enhance their performance.

REFLECTIONS ON THE INTERNSHIP

The internship provides an enriching experience in industry projects for students. The project makes them delve into the various elements that affect an organization and its employees. Those factors that affect employee attrition and optimize employee performance in an organization are studied. This will act as a pointer to the organization to take measures that could retain employees and enhance their performance. The project is an excellent opportunity to apply data science tools to practical problems and gain insights out of it.

CONCLUSIONS

The project was successful in developing analytics and reports using Data Science tools and presented a comprehensive study on HR Analytics focusing on employee attrition and performance. It made us understand and analyze factors that lead to employee attrition as well as affecting employee performance and means to optimize the same.

Microsoft Power BI was used as data visualization tool. For building reports, the following visuals were used – different types of bar charts, pie-charts, doughnut charts, scatter charts, line charts, tree maps, slicers, and funnel charts. The reports could accomplish the trend and the various elements that contributed to employee attrition. This was further explained by the measures created.

Some of the top factors that affect the employee attrition are found to be overtime, age, monthly income, distance from home, job role, salary hike among others. Employees who do not have overtime are most likely to leave the organization. This could be that they would like to have a higher amount of income or they could feel that they are not utilized. Income is a huge factor affecting attrition, since employees leave the organization in search for a better salary. Moreover, the employees getting higher salary hikes, such as sales representatives, tend to quit their jobs. Employees who have very few business travels and dissatisfied with their environments are most likely to leave the organization. Employees with lower job level are more likely to leave the job. As the job level increases, the number of employees who want to leave the company decreases. Moreover, with lower education level, employees have higher attrition rates. These insights could pave way for fruitful changes in employee management and thus, help the organization to reduce the level of attrition in the organization.

To cap it all, the project provided an excellent practical learning experience and demonstrated how data science tools could be used to solve real-life problems.

ENHANCEMENT SCOPE

The project provides us with scope for several enhancements that will fuel our decisions regarding why employees leave the organization. Having good knowledge of these facts will help the organization to tackle attrition and take steps to optimize employee performance.

The various key areas which can be further analyzed for attrition are:

- salary distribution by gender
- total working years
- education wise attrition based on monthly income
- education wise attrition based on percent salary hike
- job wise attrition based on distance from home
- daily rate
- number of companies worked
- stock option level
- marital status
- years with current manager
- attrition based on distance from home and business travel
- job satisfaction and work life balance
- Daily commute
- Job related features
- Educational factors (higher studies/credentials)

In finding the measures for optimizing employee performance, the following elements can be analyzed:

- Working environment by job role
- Salary by job role
- Level of income by job satisfaction
- Differences in daily rate
- Job satisfaction by gender

LINK TO CODE AND EXECUTABLE FILE

The code/executable has been shared in the following drive:

https://drive.google.com/file/d/1JBhswnwnGrXOwwLyuU3AC_shZl9n2nWf/view?usp=sharing

RESEARCH QUESTIONS AND RESPONSES

Few research questions and their answers are listed as follows:

1. What is the average job satisfaction by attrition status? Is any type of gender more dissatisfied than the other?

Ans: Job Satisfaction by Gender**:** Average job satisfaction score is 2.72. For people who left the organization, females had a lower satisfaction level as opposed to males.

2. What is the effect of education on attrition?

Ans: Attrition by Level of Education: Bachelors are the ones showing the highest level of attrition and they mostly come from life sciences.

3. Do employees who quit the organization have a much lower income than people who didn't quit the organization?

Ans: Yes, employees with lower income levels look for better job opportunities outside the company.

4. Does job satisfaction have an upper hand in causing attrition?

Ans: Income by Job Satisfaction**:** It seems the lower the job satisfaction the wider the gap by attrition status in the levels of income.

5. Will overtime contribute to attrition?

Ans: Yes, exhaustion at work could be a reason for attrition. Around 54% of workers who left the organization worked overtime. This, combined with various other factors like income, job role, years since last promotion etc could have fuelled attrition.

6. Which job role has the highest attrition rate? Which has the lowest?

Ans: Attrition by Job Role: Sales Representatives, HealthCare Representatives and Managers have the highest attrition rates. Research Director has the lowest attrition rate.

REFERENCES

- https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset/code
- https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview
- https://www.spec-india.com/blog/power-bi-dashboard-examples
- https://learn.microsoft.com/en-us/power-bi/consumer/end-user-dashboards
- https://data-flair.training/blogs/power-bi-dashboard/
- https://medium.com/swlh/exploratory-data-analysis-employee-attrition-rate-591ce8e7518f
- https://powerbidocs.com/2020/02/16/power-bi-clustered-bar-chart/
- https://www.youtube.com/watch?v=cfogi9vGQLc&list=PL1N57mwBHtN2NhxTvyO6TquHuZOMh4BMq&index=2
- https://www.udemy.com/course/microsoft-power-bi-up-running-with-power-bi-desktop/