Project Report

Product Name	Implement Data Visualisation techniques using MS Power BI
Qualification Name (NICF)	NICF Diploma in Infocomm Technology(Data)
Product Name	NICF-Data Queries and Visualization Basics(SF)
Module Name (NICF)	**NICF-**Data Queries and Visualization Basics**(SF)**

Student name	**Assessor name**	
Ahmad Faiz Bin Jaafar		
Date issued	**Completion date**	**Submitted on**
31/03/2022	13/04/2022	13/04/2022

Project title	**HR Attrition Project**

Learner declaration
I certify that the work submitted for this assignment is my own and research sources are fully acknowledged. Student signature: Date: 13/04/2022

Content

1. Project Background

The aim of this project is to understand the HR attrition data of a company by means of several different visualizations on Power BI services.

Employee attrition is defined as workers' departures from company, leaving roles unfilled. This concept is similar to employee turnover. By understanding attrition, leaders can design smarter retention strategies and revamp the interview or hiring process.

There are many factors that affect employee attrition. Based on the HR attrition data which we are working with, the factors which were taken into consideration are as follows:

Factors	Data Indicators (From datasheet)
Life Circumstances	• Age • Age group • Gender • Marital Status
Educational background	• Education • Educational field • Job level • Job role
Working environment	• Department • Environment satisfaction • Work life balance • Years at company
Job Satisfaction	• Department • Job satisfaction • Training times last year • Relationship satisfaction • Years with current manager
Appraisal & Remuneration	• Monthly income • Overtime • Percent salary hike • Performance rating • Average years since last promotion

Table 1. *Assumption of factors contributing to the employee attrition based on the given dataset*

Based on these data indicators, we are able to understand the factors which are affecting the attrition of the company.

2. Project Objectives

The objectives of this project are:

a. to analyze cause(s) of employee attrition
b. to offer recommendation(s) to improve current employee attrition rate from 16% to 10% or lesser
c. to build reports incorporating the use of visualizations which are appropriate to the data used and easy to understand

3. Project Specifications

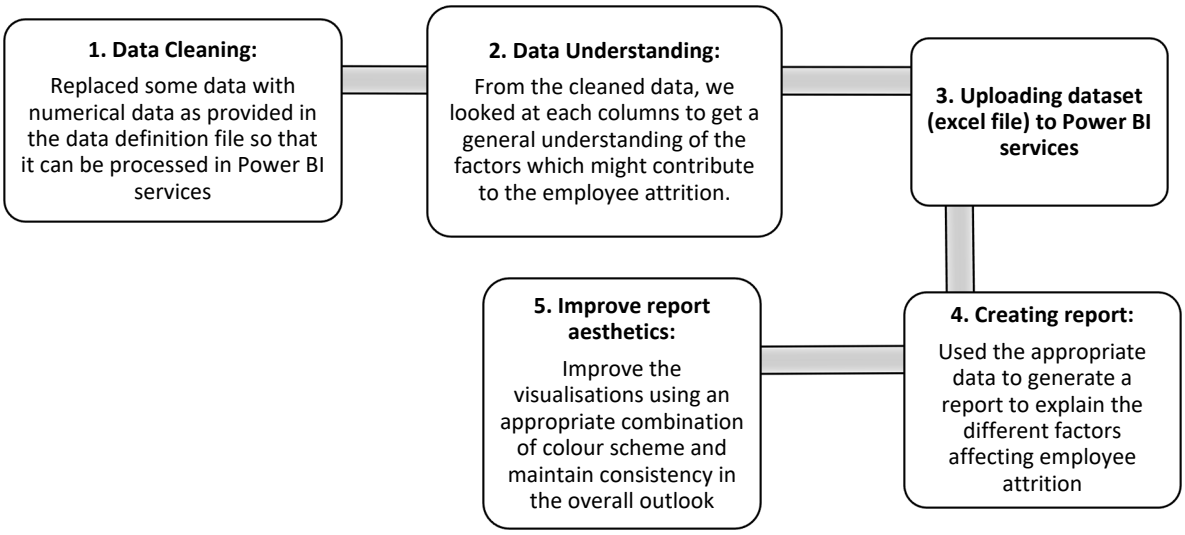
Technical tools used:

a. MS Excel (data attrition dataset, data definition file)

b. MS Word

c. Power BI Services

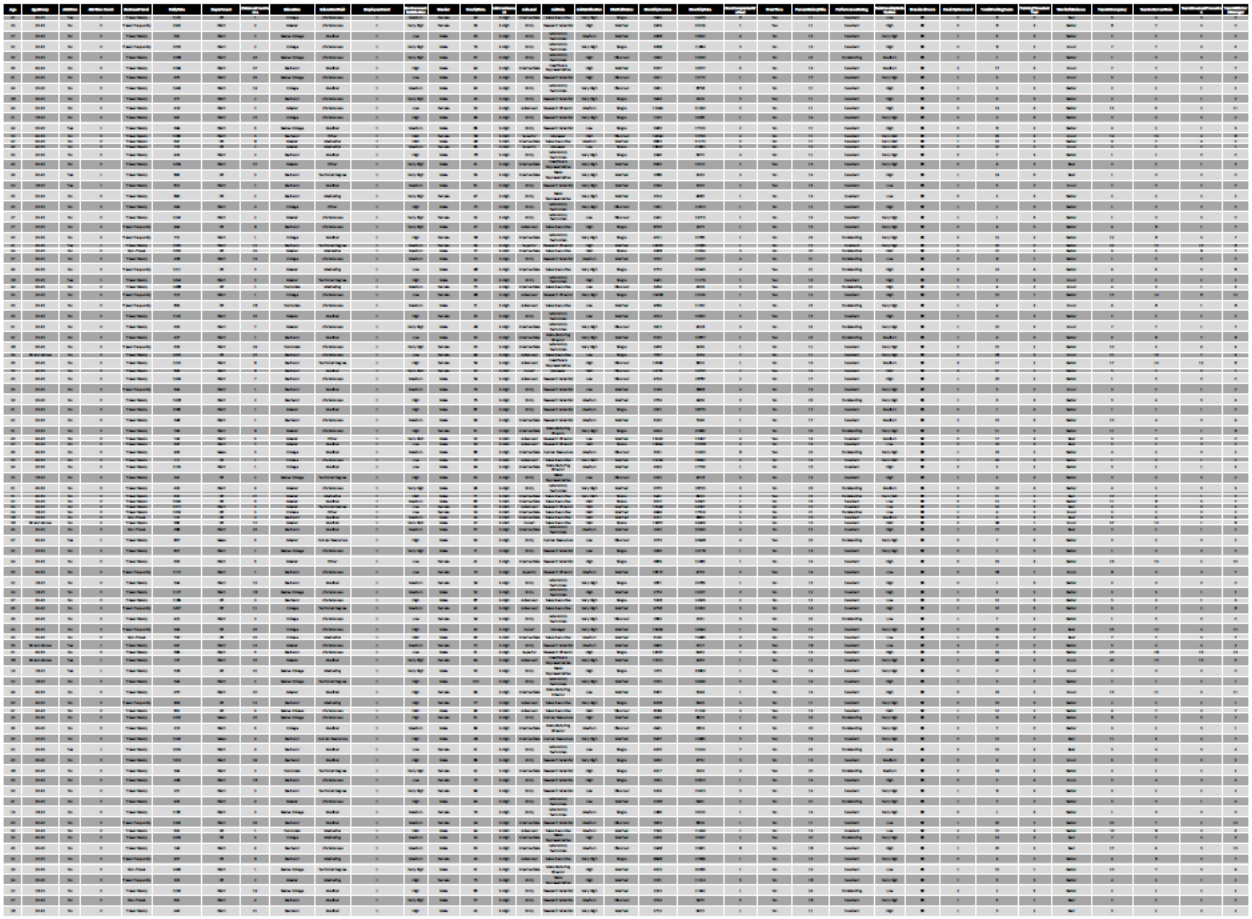
Project workflow:



4. Activity 1

4.1. Extracting the data from the HR Attrition Excel

The dataset consists of 1471 rows by 34 columns:

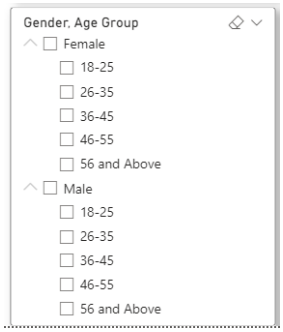
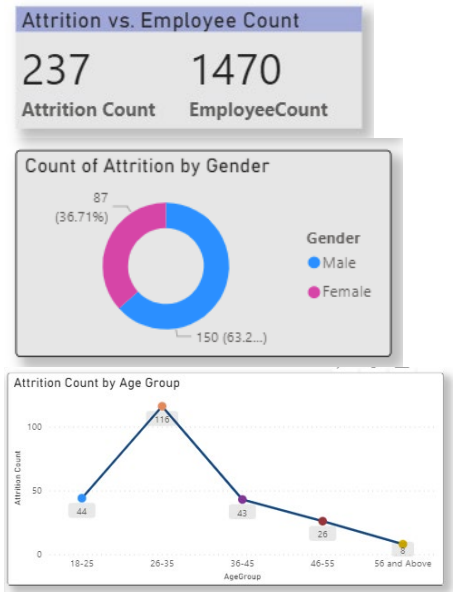


The 34 attributes of the data are shown in the table below:

Age	Distance from home	Hourly rate	Monthly income	Relations satisfaction	Years at company
Age group	Education	Job involvement	Monthly rate	Standard hours	Year in current role
Attrition	Education field	Job level	No. of companies worked	Stock options level	Years since last promotion
Business travel	Employee count	Job role	Overtime	Total working years	Years with current manager
Daily rate	Environment satisfaction	Job satisfaction	Percent hike salary	Training times last year	
Department	Gender	Marital status	Performance rating	Work life balance	

4.2. Visualizations required

Both static and interactive visuals were used in the project.

Static	Visuals which are not changed with respect to the data. Examples of visuals used: filter.	
Interactive	Visuals which get changed with respect to the data. Examples of visuals used: bar chart, donut chart, data card.	

4.3. Audience of data visualization and their size

The audience of this data visualization presentation are the HR department and the heads of department. The audience comprises of 354 people.

4.4. Level of details required for data visualization

When comparing two categories of data (e.g. Attrition vs. Age group), a **bar/column chart** is normally usually used as it allows the audience to compare, recognize patterns or trends far more easily than a table of numerical data.

When comparing more than two categories of data (e.g. Attrition vs. Education and Education field), a **tree map**, a **matrix table,** a **clustered column/bar chart** or a **decomposition tree** is used as it is able to display large amounts of hierarchical data.

4.4. Layout and colors required for the visualization

For a visually enticing display, the colors of the visuals were kept to a minimalist theme. After analyzing the dataset, it is found that the company's attrition can be analyzed based on the factors listed in the table below using the appropriate data indicators.

Factors	**Data Indicators (From datasheet)**
Life Circumstances	• Age • Age group • Gender • Marital Status
Educational background	• Education • Educational field • Job level • Job role
Working environment	• Department • Environment satisfaction • Work life balance • Years at company
Job Satisfaction	• Department • Job satisfaction • Training times last year • Relationship satisfaction • Years with current manager
Appraisal & Remuneration	• Monthly income • Overtime • Percent salary hike • Performance rating • Average years since last promotion

5. Activity 2

The dataset deals with 1470 employees of a company and each employee is analyzed based on 34 different attributes as listed below:

Age	Distance from home	Hourly rate	Monthly income	Relations satisfaction	Years at company
Age group	Education	Job involvement	Monthly rate	Standard hours	Year in current role
Attrition	Education field	Job level	No. of companies worked	Stock options level	Years since last promotion
Business travel	Employee count	Job role	Overtime	Total working years	Years with current manager
Daily rate	Environment satisfaction	Job satisfaction	Percent hike salary	Training times last year	
Department	Gender	Marital status	Performance rating	Work life balance	

6. Activity 3

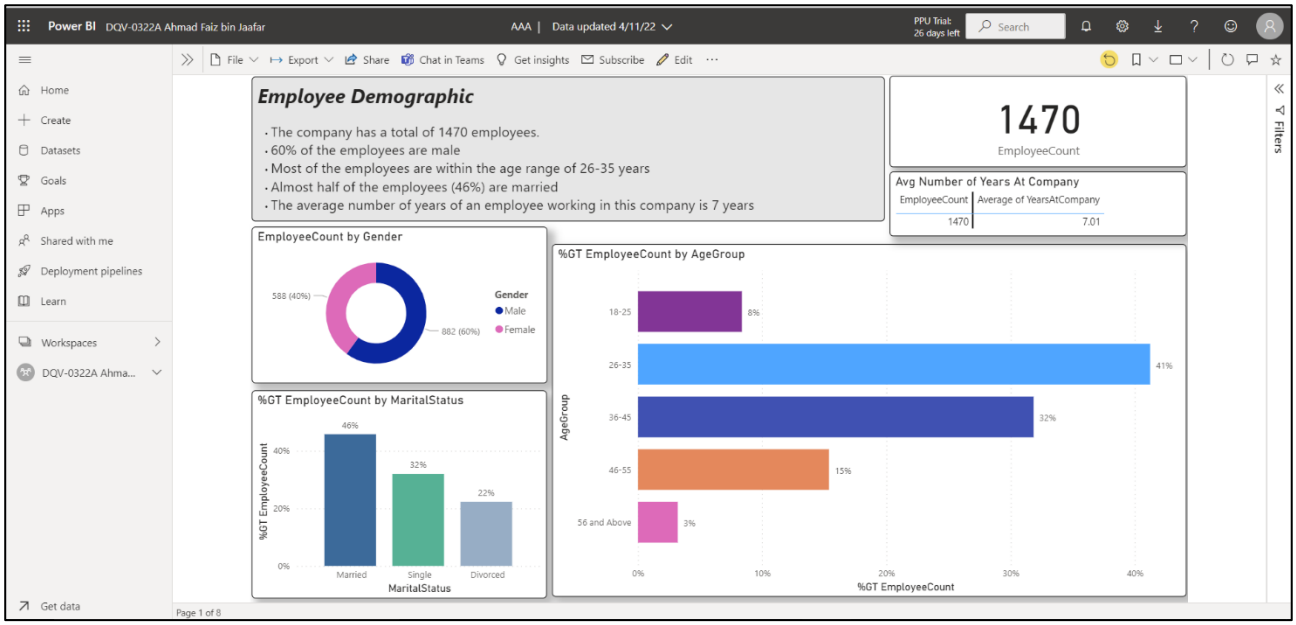
Since one of our objectives is to build reports incorporating the use of visualizations which are appropriate to the data used and easy to understand, below are the list of visualizations that were used throughout this project:

Chart	Justification
Card	Card is a useful visual to display a single number. It can also be used to show comparisons between categories of data.
Pie chart	A pie chart is a circular graph that's used to show or compare different segments of data. Each slice represents a proportion that relates to the whole. Pie charts are best used for showcasing part-to-whole relationships.
Donut/ Pie chart	A donut chart is almost identical to a pie chart, but the centre is hollow. Donut charts are also used to show part-to-whole relationships, but the centre can also be used to display data. However, with donut charts, it can accommodate fewer categories than pie charts — anywhere from 2 to 5. That's because you want your audience to be able to quickly tell the difference between arc lengths.
Stacked bar chart	Stacked bar chart allows the direct comparison of multiple data series stacked in horizontal bars. Each bar displays a total section of data, broken down into sub-sections. Equivalent subsections are the same colour in each bar. Each horizontal bar is further divided into subsections which represent another category of data. This formatting makes it easy to compare both the whole picture and the components of each bar.
Clustered column/bar chart	Clustered column/bar chart allows the direct comparison of multiple data series which is displayed in clustered vertical or horizontal columns.
Line and clustered column chart	A line and clustered column chart is used when there are more than one data column which all add up to the total trend. The clustered column chart will represent one category of data, while the line chart will represent another. This allows one to see the relationship trend between these two categories of data across an axis.
Line chart	A line chart is easy to understand and simple in form, typically only depicting changes in a data attribute over time. Usually when time factor is involved in analyzing the data, this chart is used in particular.
Waterfall chart	The point of the waterfall chart is to show both positive and negative values over a period of time, while pointing out the initial and end values as well.
Matrix table	Matrix is a table which can automatically aggregate the data and enables drill down. It allows one to compare multiple categories of hierarchical data all in one table.
Slicer	Slicers provide buttons that you can click to filter tables or charts. In addition to quick filtering, slicers also indicate the current filtering state, which makes it easy to understand what exactly is currently displayed.
Decomposition tree	The decomposition tree displays the data in multiple dimensions. It automatically aggregates the data and enables drilling down into dimensions in any order.
Treemap	Treemap is used to display large amounts of hierarchical data using nested rectangles. It shows the proportions between each part and the whole in a tree-like structure.

7. Activity 4

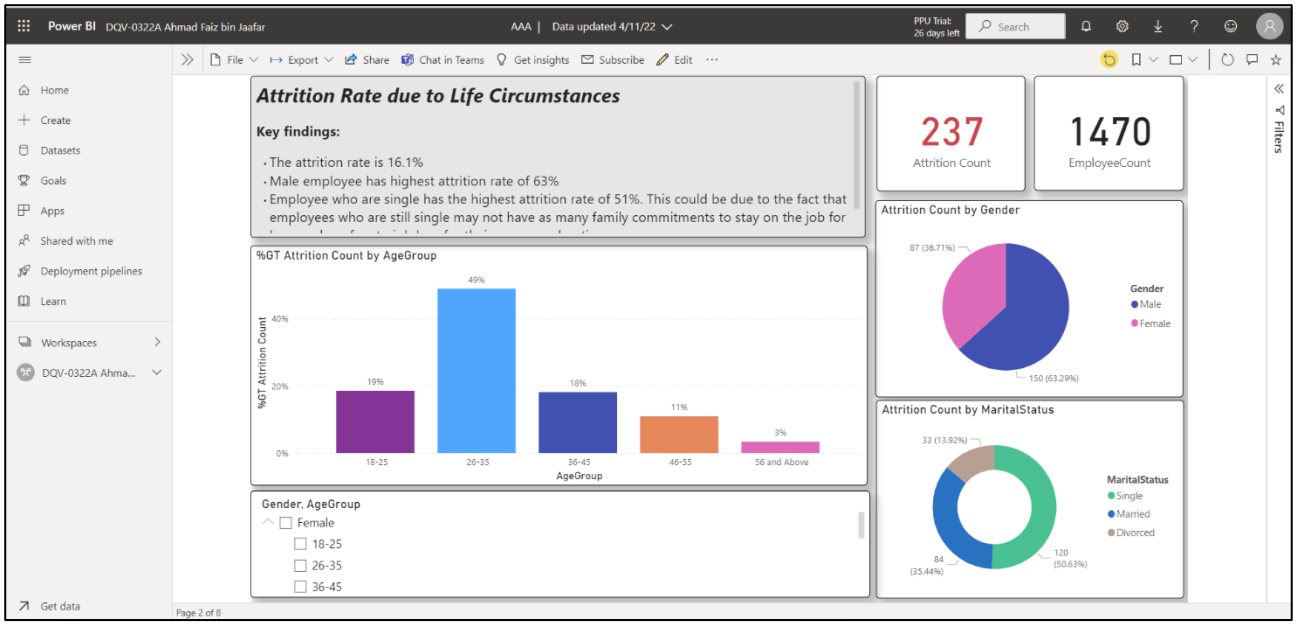
7.1. Page 1 – Employee demographics

Before we try to explain about the factors affecting the attrition rate, it is first important to establish the overall background of the employee demographics to the audience.



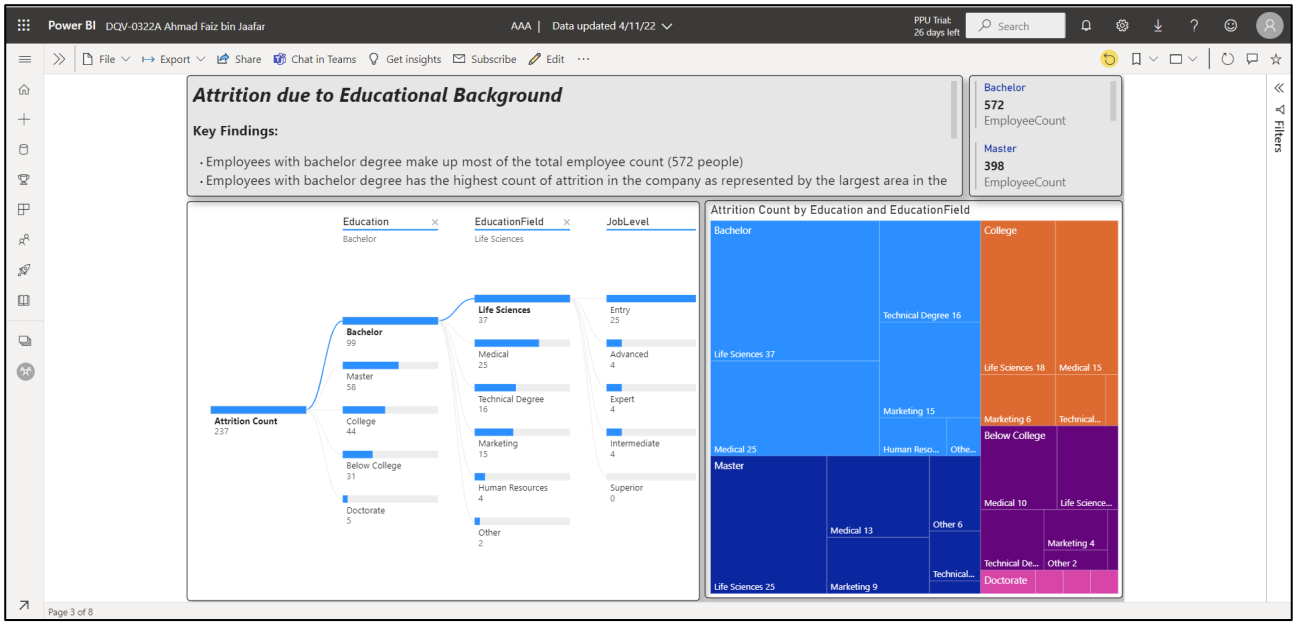
Visual	Outcome	Justification
<u>Card</u> - Employee count	• Total no. of employee is 1470	Card is a useful visual to display a single number. In this case, it is used to track the total count of employee based on the selected demographic attribute.
<u>Donut Chart</u> - employee count by gender	• shows a breakdown of the employee count based on the respective gender (male & female) • There are 60% of employees who are male and 40% who are female	Donut chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show value of each gender to the total number of employee count.
<u>Clustered Column Chart</u> - percentage of employee by marital status	• shows percentage of employee by marital status • 46% of employees are married, 32% of employees are single while the other 22% are divorced	Clustered column chart allows the direct comparison of multiple data series which is displayed in clustered vertical columns. In this case, it is used to show the percentage of employee count by marital status.
<u>Clustered Bar Chart</u> - employee count by age group	• Shows number of employees belonging to different age group • Most of the employees are within the age range of 26-35 years	Clustered bar chart allows the direct comparison of multiple data series which is displayed in clustered horizontal columns. In this case, it is used to show the employee count by age group.

7.2. Page 2 – Attrition due to life circumstances



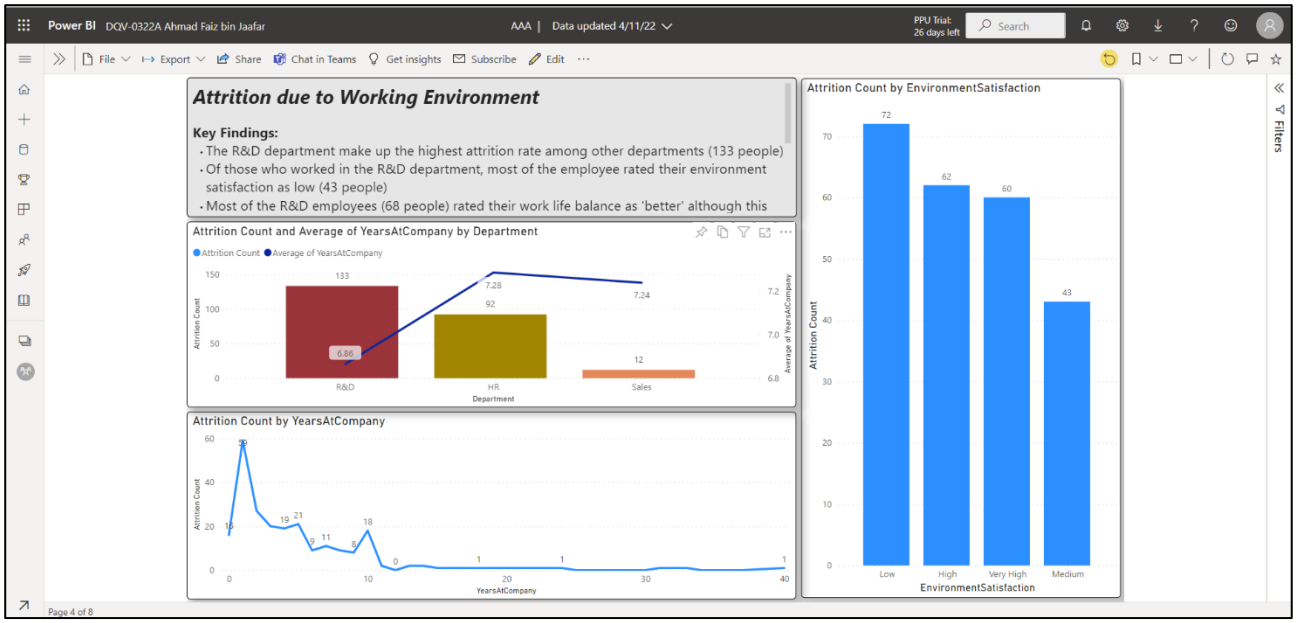
Visual	Outcome	Justification
<u>Card</u> - Employee count - Attrition count	• Total no. of employee is 1470 • Attrition count is 237	Card is a useful visual to display a single number. In this case, it is used display the total count of employee and attrition count based on the selected demographic attribute.
<u>Pie Chart</u> - attrition count by gender	• shows a breakdown of the attrition count based on the respective gender (male & female) • Of those who left the company, 63% were male and 37% were female	Pie chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show the number and percentage of attrition based on gender.
<u>Donut Chart</u> - attrition count by marital status	• shows number and percentage of employee attrition by marital status • Of those who left the company, 51% were single, 35% were married and 14% were divorced	Donut chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show the percentage of employee attrition by marital status.
<u>Clustered Bar Chart</u> - attrition count by age group	• shows number of employees leaving the company based on different age group • employees within the age range of 26-35 years account for the highest attrition in the company	Clustered bar chart allows the direct comparison of multiple data series which is displayed in clustered horizontal columns. In this case, it is used to show the attrition count by age group.
<u>Slicer</u> - filter attrition count by gender and/or age group	• the slicer shows a list of age group by gender so that one can filter according to the data that is needed	Slicers provide buttons that you can click to filter tables or charts. In addition to quick filtering, slicers also indicate the current filtering state, which makes it easy to understand what exactly is currently displayed.

7.3. Page 3 – Attrition due to educational background



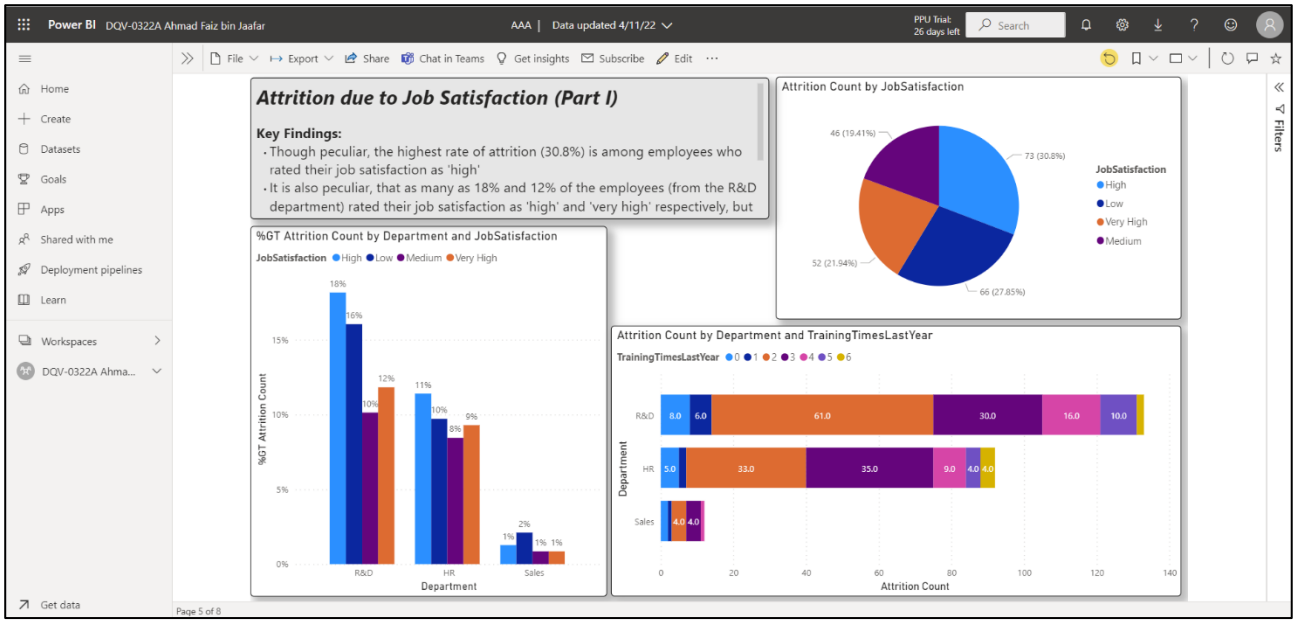
Visual	Outcome	Justification
<u>Multi-row card</u> - Employee count based on education level	• out of the 1470 employees in the company, highest education attained at bachelor is 572 employees, highest education attained at masters is 398 employees, highest education attained at college is 282 employees, highest education attained below college is 170 employees and highest education attained at doctorate is 48 employees	Multi-Row cards allows values to be grouped and displayed in a summarized form. These types of cards are used to present amounts or a set of figures independent of other references. In this case, the employee count based on education level is just as a form of reference independent of the other data in this page report.
<u>Treemap</u> - Attrition count by Education & Education Field	• shows the attrition count based on the education level and education Field • employees with bachelor degree has the highest count of attrition in the company as represented by the largest area in the treemap (99 people) • most of the employees with bachelor degree who left the company worked in the life sciences field (37 people)	Treemap is used to display large amounts of hierarchical data using nested rectangles. It shows the proportions between each part and the whole in a tree-like structure.
<u>Decomposition Tree</u> - Attrition by: • Education/ • Education field/ • Job Role/ • Job level	• shows the attrition count based on 4 different attributes • of those employees in the life sciences, most of them work as laboratory technicians (33 people) and they make up the highest attrition count as compared to other job roles • employees who are at the entry level make up most of the attrition rate (143 people) • the tendency of employees doing entry level job roles leaving the company holds true for every education level and education field.	The decomposition tree displays the data in multiple dimensions. It automatically aggregates the data and enables drilling down into dimensions <u>in any order</u>.

7.4. Page 4 – Attrition due to working environment



Visual	Outcome	Justification
<u>Clustered Column Chart</u> - attrition count by department and environment satisfaction and work life balance (drill-down)	• there are 3 different departments in the company namely, R&D, HR and sales. • the R&D department make up the highest attrition rate among other departments (133 people) • of those who worked in the R&D department, most of the employee rated their environment satisfaction as low (43 people) • most of the R&D employees rated their work life balance as 'better'	Clustered column chart allows the direct comparison of multiple data series which is displayed in clustered vertical columns. Using the drill-down method, we can gain more insight into the root causes of a problem within a department. After the root causes are known, a larger plan can be devised to address the problem.
<u>Line and clustered column chart</u> - attrition count and average number of years at company by department	• employees who worked in the R&D department has the lowest average number of years working at the company which is below the average of 7 years • this could be a possible explanation as to why this department has the highest rate of attrition as compared to other departments	A line and clustered column chart are used when there are more than one data column which all add up to the total trend. In this case, we would like to establish a relationship trend between the attrition count and the average number of years at company by department.
<u>Line Chart</u> - attrition count by years at company	• it seems that there is a high number of attrition for employees who worked in the company for 10 years or lesser • however, the attrition rate is highest for employees who work in the company between 0-2 years	A line chart is easy to understand and simple in form, typically only depicting changes in a data attribute over time.

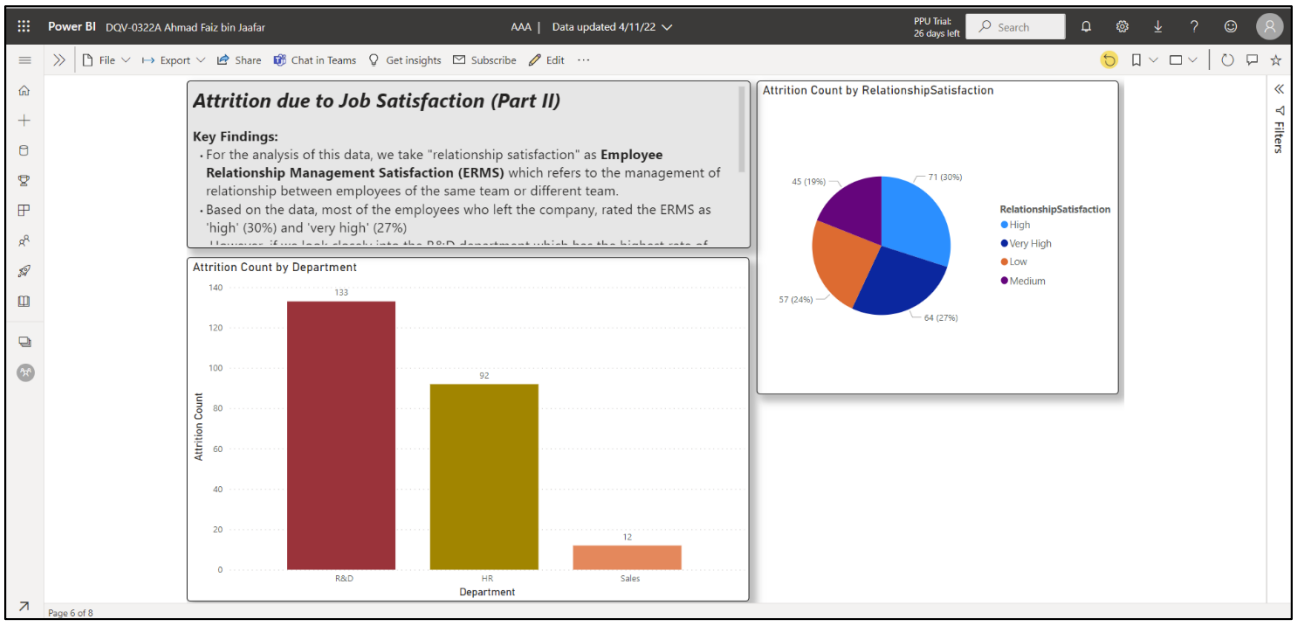
7.5. Page 5 – Attrition due to job satisfaction (part I)



Visual	Outcome	Justification
<u>Pie Chart</u> - attrition count by job satisfaction	• shows a breakdown of the attrition count based on the job satisfaction • the highest rate of attrition (30.8%) is among employees who rated their job satisfaction as 'high'	Pie chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show the number and percentage of attrition based on job satisfaction.
<u>Clustered Column Chart</u> - attrition count by department and job satisfaction	• as many as 18% and 12% of the employees (from the R&D department) rated their job satisfaction as 'high' and 'very high' respectively, but the R&D department still suffers from the highest rate of attrition	Clustered column chart allows the direct comparison of multiple data series which is displayed in clustered vertical columns. In this case, the chart is used as we would like to analyse the attrition count based on the 3 different departments and how the employees who left rated their job satisfaction based on the department where they came from.
<u>Stacked Bar Chart</u> - attrition count by department and training times last year	• we can see from each department, the number of employees who left the company based on the number of training hours they had last year • it can be seen from R&D and HR department that employees are more likely to leave the company when the hours spent on training is only between 2 to 3 hours. This is probably the likely cause of employees in the R&D department leaving the company	A stacked bar chart allows the direct comparison of multiple data series stacked in horizontal bars. Each bar displays a total section of data, broken down into sub-sections. Equivalent subsections are the same colour in each bar. In this case, each horizontal bar represents the attrition count for each department. Each horizontal bar is further divided into subsections which represent the number of employees based on the number of training hours they had last year. This formatting makes it easy to compare both the whole picture and the components of each bar.

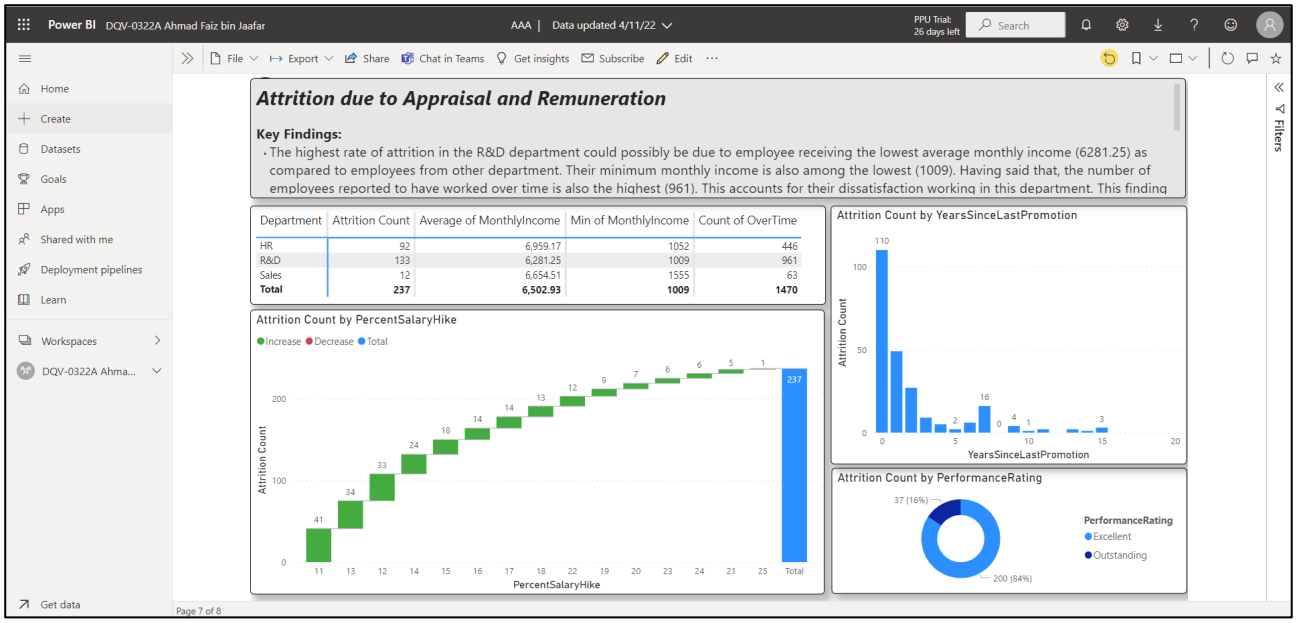
7.6. Page 6 – Attrition due to job satisfaction (part II)

For the analysis of this data, we take "relationship satisfaction" as **Employee Relationship Management Satisfaction (ERMS)** which refers to the management of relationship between employees of the same team or different team.



Visual	Outcome	Justification
<u>Pie Chart</u> - attrition count by relationship satisfaction	• shows a breakdown of the attrition count based on the relationship satisfaction • Based on the data, most of the employees who left the company, rated the ERMS as 'high' (30%) and 'very high' (27%)	Pie chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show the number and percentage of attrition based on relationship satisfaction.
<u>Clustered Column Chart</u> - attrition count by department and years with current manager (drill down)	• of those who left the company from the R&D department (133 people), most of them rated the ERMS as 'low' (37 people). This accounts for why R&D department has got the highest rate of attrition • when we drill down, we also see a trend that where employees spend more years with their current manager, the attrition number decreases. This trend is maintained across all 3 departments	Clustered column chart allows the direct comparison of multiple data series which is displayed in clustered vertical columns. In this case, the chart is used as we would like to analyse the attrition count based on the 3 different departments and whether the number of years with their current managers affects the attrition count across all 3 departments

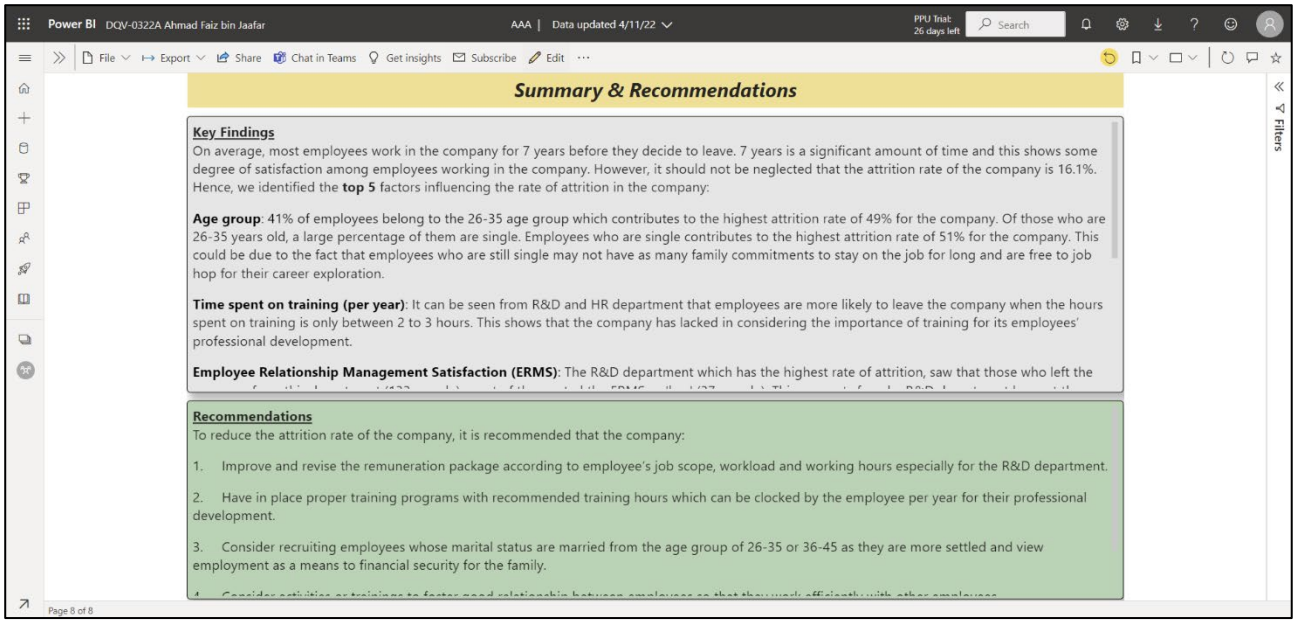
7.7. Page 7 – Attrition due to appraisal and remuneration



Visual	Outcome	Justification
<u>Donut Chart</u> - attrition count by performance rating	• there are only 2 performance rating namely, 'excellent' and 'outstanding' • all of the employees who left the company received either an 'excellent' (84%) or 'outstanding' (16%) performance appraisal	Donut chart is used to illustrate the distribution of different value from each category of an attribute to a total. In this case, it is used to show the percentage of employee attrition by performance rating.
<u>Matrix Table</u> - departmental Attrition by average monthly income, minimum monthly income and count of overtime	• shows value of attrition based on these attributes: average monthly income, minimum monthly income and count of overtime by department • the highest rate of attrition in the R&D department could possibly be due to employee receiving the lowest average monthly income (6281.25) as compared to employees from other departments. Their minimum monthly income is also among the lowest (1009). Having said that, the number of employees reported to have worked over time is also the highest (961).	Matrix is a table which can automatically aggregate the data and enables drill down. In this case, it is used to show departmental attrition based on average monthly income, minimum monthly income and count of overtime.
<u>Waterfall chart</u> - attrition count by percentage salary hike	• of those who left the company, no employees reported a decrease in their salary • the lowest percentage salary hike is 11% while the highest is 25%. • the attrition count decreases with the increase in the percentage of salary hike • the attrition count is highest at 41 when the percentage salary hike is the lowest. The attrition count is lowest at 1 when the percentage salary hike is the highest	The point of the waterfall chart is to show both positive and negative values over a period of time, while pointing out the initial and end values as well. In this case, we can see the green bars which represent the count of attrition gets smaller with increasing percentage salary hike.

Clustered column chart - attrition count by number of years since last promotion	• The attrition count seems to decrease with the increase in the number of years since the last promotion. • except that we see an outlier where the attrition count suddenly increases significantly to 16 at 7 years since last promotion	Clustered column chart allows the direct comparison of multiple data series which is displayed in clustered vertical columns. In this case, the chart is used as we would like to analyse the attrition count based on the number of years since last promotion.

7.8. Page 8 – Summary



Key Findings	Recommendations
On average, most employees work in the company for 7 years before they decide to leave. 7 years is a significant amount of time and this shows some degree of satisfaction among employees working in the company. However, it should not be neglected that the attrition rate of the company is 16.1%. Hence, we identified the top **5** factors influencing the rate of attrition in the company: 1. **Age group & Marital status**: 41% of employees belong to the 26-35 age group which contributes to the highest attrition rate of 49% for the company. Of those who are 26-35 years old, a large percentage of them are single. Employees who are single contributes to the highest attrition rate of 51% for the company. This could be due to the fact that employees who are still single may not have as many family commitments to stay on the job for long and are free to job hop for their career exploration. 2. **Time spent on training (per year)**: It can be seen from R&D and HR department that employees are more likely to leave the company when the hours spent on training is only between 2 to 3 hours. This shows that the company has lacked in considering the importance of training for its employees' professional development. 3. **Employee Relationship Management Satisfaction (ERMS)**: The R&D department which has the highest rate of attrition, saw that those who left the company from this department (133 people), most of them rated the ERMS as 'low' (37 people). This accounts for why R&D department has got the highest rate of attrition. This possibly means that the company has lacked in its	To reduce the attrition rate of the company, it is recommended that the company: 1. Improve and revise the renumeration package according to employee's job scope, workload and working hours especially for the R&D department. 2. Have in place proper training programmes with recommended training hours which can be clocked by the employee per year for their professional development. 3. Consider recruiting employees whose marital status are married from the age group of 26-35 or 36-45 as they are more settled and view employment as a means to financial security for the family. 4. Consider activities or trainings to foster good relationship between employees so that they work efficiently with other employees. 5. Consider revising the appraisal methods for employees so that the employees can be appraised at least 2 times a year and review their job scope and professional development.

management of relationship between employees as well as employees between employers which led to some dissatisfaction among many. 4. **Salary**: The highest rate of attrition in the R&D department could possibly be due to employee receiving the lowest average monthly income (6281.25) as compared to employees from other departments. Their minimum monthly income is also among the lowest (1009). This accounts for their dissatisfaction working in this department. This finding is also congruent with the finding where the attrition count decreases with the increase in the percentage of salary hike. The attrition count is highest at 41 when the percentage salary hike is the lowest (11%). The attrition count is lowest at 1 when the percentage salary hike is the highest (25%). This means that employees highly regard the importance of being paid fairly as one of the key factors to remain in the company. 5. **Overtime**: The number of employees reported to have worked over time is also the highest (961) for the R&D department. This accounts for their dissatisfaction working in this department which has the hights rate of attrition. Although many employees from this department reported having the highest overtime, the month average income is still the lowest. Hence, they could have been underpaid, or unpaid for their overtime hours in the company.	

8. Annexure (Excel Workbook)

1. Project Report:

https://app.powerbi.com/links/aQ2f7sHo5i?ctid=cbdb10cb-b644-4887-995c-8ea0e1d40752&pbi_source=linkShare&bookmarkGuid=ef181780-c29e-4f10-b0af-19474ed3d85d

2. HR Attrition dataset (edited):

https://app.powerbi.com/groups/6c7ac6e2-b1cf-4733-9bab-4275e3e9b881/workbooks/1_38672_840943/new

3. HR Attrition data definition:

https://app.powerbi.com/groups/6c7ac6e2-b1cf-4733-9bab-4275e3e9b881/workbooks/1_38673_840945/new