Project Report

|  |  |
| --- | --- |
| Product Name | Implement Data Visualisation techniques using MS PowerBI |
| Qualification Name (NICF) | NICF Diploma in Business Analytics |
| Product Name | Data Queries and Visualization Basics |
| Module Name (NICF) | NICF Data Queries and Visualization Basics |

|  |  |  |  |
| --- | --- | --- | --- |
| Student name | | Assessor name | |
| Neo Wei Siang | | Suryanto | |
| Date issued | Completion date | | Submitted on |
| 16 April 2020 | 24 April 2020 | | 25 April 2020 |
|  | |  | |
| Project title | Data Queries and Visualization Basics | | |

|  |
| --- |
| Learner declaration |
| I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.    Student signature: Date: 25 April 2020 |

Content

|  |  |
| --- | --- |
| Project sections | Page Number |
| Project Overview | 2 |
|  | |
| Factors that influence Data Visualization | 3 to 4 |
|  | |
| Visualization Techniques | 5 |
|  | |
| Data Set Analysis | 6 |
|  | |
| Data visualisation | 7 to 8 |

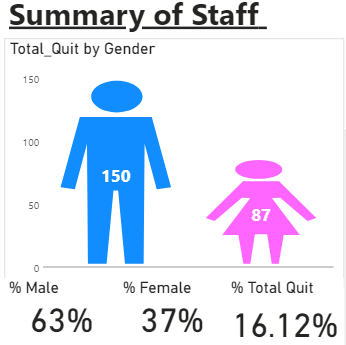
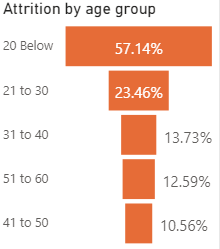
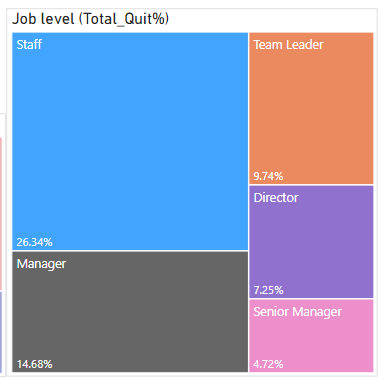
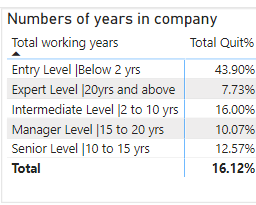
1. Project Overview: Describe the Project along with Project Outcomes and tools used (Explain the Project in your own words in 15 – 20 lines)

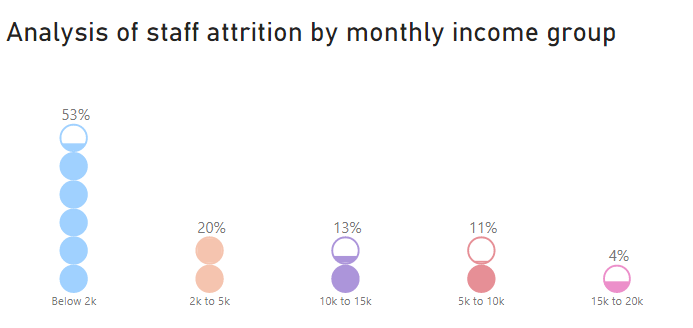
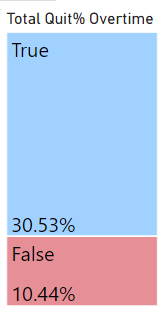
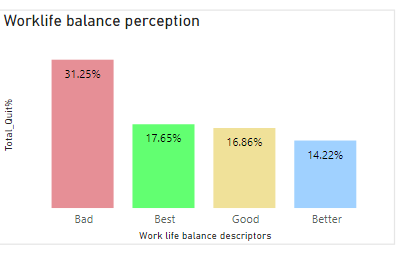
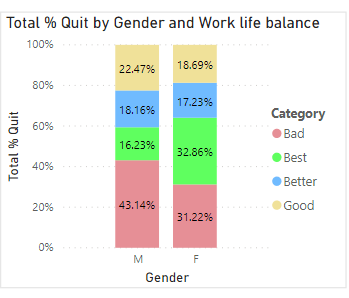
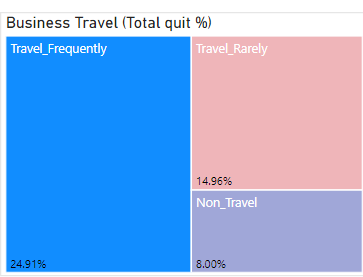
In this project, we are given data from the SQL service with information such as employees information, attrition status, academic etc. However, some variables are not in the correct data type and needed to be corrected. This is shown in the table below:

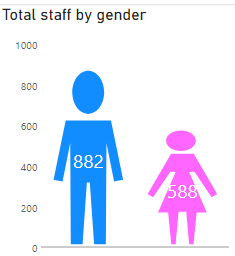
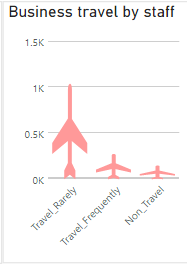
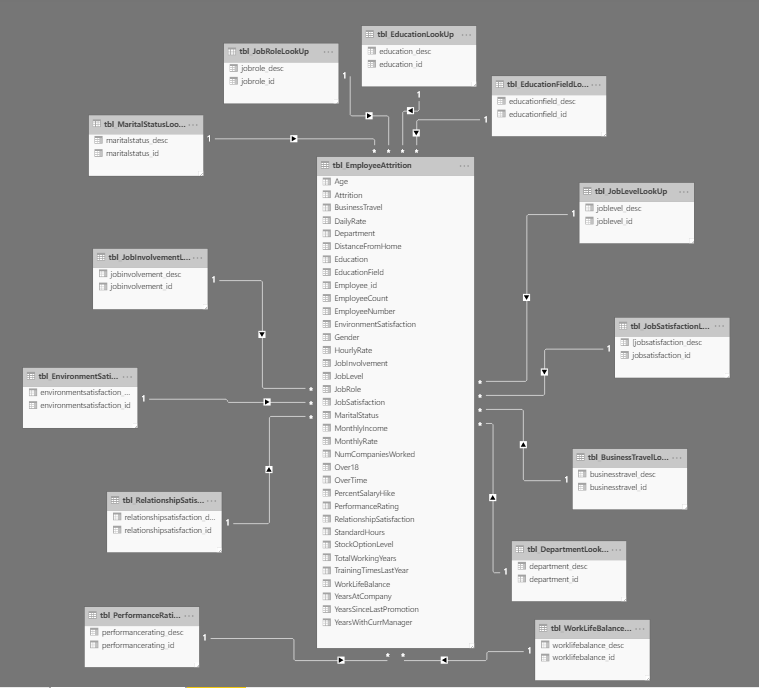
|  |  |  |  |
| --- | --- | --- | --- |
| * No | * Column name | * From data type | * To data type |
| * 1 | * PercentSalaryHike | * Currency | * Whole number |

After the data has been cleansed/transformed, we also start to create new columns (i.e. to perform age groups and income groups of current staff etc) and Data Analysis Expressions aka DAX measures (i.e. calculating number of total employees, sub groups of details) to perform data visualisation.   
  
The visualisation software to use for the project is Microsoft Power BI (“Power BI”). We can then proceed to use suitable visualisation techniques to analyse the variables that might be the contributing factors of employee attrition.  
  
When the factors are identified, we will then further drill down on underlying factors that have relationships with the variables using Power BI visualization techniques.

Project outcomes

1. Using visualisation tools to analyse the main factors resulting in staff attrition;
2. After using Power BI desktop, publish the results in Power BI services online and pin a few important visuals as a dashboard; and
3. Prepare the project report.
4. Factors that influence Data Visualization: (List and Explain the factors such as static or interactive data, audience of data visualization, frequency of data visualization, level of details etc.)   
     
   The factors influencing Data visualization are listed as below:  
   a) Age Group and Gender (Attrition)  
     
   ai) From the infographic and cards visualization tools used below, there are 223 (150+87) resigned employees out of a total of 1470 employees in the company.  
   aii) Attrition rate for male and females are 63% and 37% respectively.  
   aiii) Attrition by age group using a funnel visualization tool showed that resigned staff below 20 formed the majority (57.14%).   
     
     
     
     
     
     
     
     
     
     
     
     
     
     
   Observation: The profile of a resigned employee tend to be male and below age 20 from the above preliminary data analysed.  
     
     
   b) Total working years with the company, job position and job level (attrition)  
   bi) We have used a matrix and treemap in Power BI to generate the below information:  
     
     
   Observation: The resignees whom left the company worked for less than two years (43.90%), working at an entry level (i.e. staff level at 26.34%).

Factors that influence Data Visualization (Cont’d):   
  
c) Staff attrition by monthly income group and total overtime  
   
Observation: Using a bubble stack visualization tool, the resigned staff earning below $2k (53%) tend to leave the company due to possibly lower pay.In addition, working overtime (i.e. “True” at 30.53%) was also a factor in resigned staff leaving the company.  
  
  
d) Total staff attrition by gender and their perception of work life balance  
  
Observation: Using a stacked column chart (first image), we analysed that for both male (43.14%) and female (31.22%) resignees rated the worklife balance as “bad” would leave the company. The second image confirmed the first visual’s represented data that “bad” worklife balance would cause a staffer to leave the company.  
  
  
e) Attrition based on the frequency of business travel   
  
Observation: Using the treemap visualization,previous employees traveling frequently (24.91%) for business formed the bulk of the resigned staff.

1. Visualization Techniques: (List and describe the visualization techniques considered along with justification why they were used or not used)  
     
   Power BI Visualization tools used for this project  
   1. Card  
   As the project’s focus is on employee attrition, the attrition rate of 16.12% (237 resignees out of 1470 employees) is the basis for data analysis. This figure is easily generated and can be presented in the report or dashboard.   
     
   2. Table and matrix  
   Table and matrix help to organize the information based on the chosen factors by arranging them in columns and we have an option to arrange in ascending or descending order. The headers would describe the variables and relationships of the figures can be interpreted.   
     
   3. Column and Bar Chart  
   The charts help to display column of varying length representing data value. The charts are also to compare numeric quantities or counts across categories.   
     
   4. Combination Chart  
   Such charts display dataset with measures having different scales, eg primary scale is in quantity and the secondary scale is in percentage with share axis, allowing the data presented to be more useful for decision-making.  
     
   5. Slicer  
   Enable user to examine part of a data visualization through filtering (i.e. selecting one department or all departments showing data with different figures)  
     
   6. Pie Charts and Doughnut Charts  
   Display data categories as proportion of a whole  
     
   7. Sunburst (installed visualization tool)  
   To display data sets in hierarchy format (i.e. analysis by department and job role)  
     
   8. Enlighten Bubble stack (installed visualization tool)  
   Using an alternative method instead of column bar charts to present columnal data.   
     
   9. Infographic Designer 1.8.7 (installed visualization tool)  
   Using an visual images such as male, female to represent gender and aeroplane icon to demonstrate the concept of business travel. Please see below for samples.  
     
     
     
   Visualization tools not used in this report  
   1. Map  
   To display information in geographic format (there is no country information given)   
     
   2. Scatter Chart  
   To display outliers and correlations in sets of data (for this case, we are analysing possible factors in employee attrition. Outliers may not be the top concern).
2. Data Set Analysis: (Analyse and list down your observations on total volume of data, growth rate, diversity of data and relationships between various columns of data): 

Please refer to the screenshot above for the relationships among the tables with the various of data. There are 1470 data records with 35 variables in the table EmployeeAttrition. This is before the new columns and new measures are created for the data visualization process.

5. Data Visualization: (Explain the process in which data visualization has been accomplished   
 using the spreadsheet which you have created with data visualization)  
  
a) Before data visualization is done, five steps are needed to convert the raw data imported from the SQL database into a format for data visualization. Please see the details below:  
  
Step 1: Data preparation  
To check that the variables are in the correct data type/format. We have to change the data type to the correct type so that data visualisaiton can be done properly.  
  
  
Step 2: Data Cleaning  
To remove any duplicated records, eliminate errors in data type before doing the visualisation  
  
  
Step 3: Create New Columns  
To create new column for grouping column variables and for variable with “True/False” status.  
  
New grouping column created in this project are:  
a) Age Group  
b) Length of Service Group  
c) Monthly Income Group  
d) Total working years Group  
e) Years with Current Manager Group  
  
New Columns with text status for below “True/False” data type  
a) Attrition 🡺 Convert to “Stayed” or “Left”  
  
  
Step 4: Creating New Measures  
Some of the new measures (DAX) are created for data visualization:  
a) “Total Left” 🡺 To calculate the total number of Employee Left  
b) “Total Stayed” 🡺 To calculate the total number of Employee Stayed  
c) “Total Employee” 🡺 To calculate the total number Employee  
d) “%Attrition By Category” 🡺 To find the attrition rate

Data Visualization (Cont’d):  
  
Step 5: Proceed with Data Visualization and findings

Please refer to the embedded file to assess the power BI with the visuals.   
  
The power BI desktop file has also been published in Power BI online services in a dashboard. Please see the below two screenshots below for the five visuals tagged:  
