Project Report

|  |  |
| --- | --- |
| Product Name | Implement Data Visualisation techniques using MS Excel |
| 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 | |
| Lim Qiang Sheng Michael | | SURYANTO (Ryan) | |
| Date issued | Completion date | | Submitted on |
|  | 21/02/2019 | | 21/02/2019 |
|  | |  | |
| Project title | Student Registration Form Development | | |

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

Content

1. Project Overview 2
2. Steps for Preparation of Data 3
3. Data Visualization Techniques 4
4. Data Set Analysis 5
5. Data Visualization 6

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

This project takes a look at survey data taken from within an unnamed company with only 3 departments of HR, Sales and R&D. The survey data looks at factors which affect company attrition or turnover. Examples of these factors include, Education, Age, Gender, Distance from home etc.

This project was done solely with Excel, taking data from existing workbooks. Although it was not needed for this project, as part of my skillset I have learnt how to select the relevant data from large databases using SQL or even from multiple files or from text or pdf.

For the purpose and scope of this project, only 5 factors have been found to be useful. Namely, whether or not the employee does OT (overtime), total working years, age, marital status and job level has been found to have the highest correlation.

Unfortunately, this data only shows obvious and logical things like employees having overtime and from lower job levels and younger employees or those on their first job are likely to leave. While this information is useful to management and steps can be taken in terms of welfare and hiring practices, overtime cannot be eliminated totally. Nor can hiring of young people on their first job be eliminated.

We will then again look at this subset of young employees and analyze again this subset to see what factors affect them.

Steps for preparation of data:

Steps used for preparation of data or cleaning of data include.

* Hiding or removing duplicate data, getting rid of unwanted columns,
* Normalization - Using find and replace to change text to numbers to prepare data for Data Analysis.
* Using Data Analysis tool to produce a correlation matrix out of all the columns from the attrition data sheet in order to find which factors influence attrition the most.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Age\_Norm* | *Attrition* | *BizTrav\_Norm* | *Dept\_Norm* | *Dist\_Norm* |
| Age\_Norm | 1 |  |  |  |  |
| Attrition | -0.159205007 | 1 |  |  |  |
| BizTrav\_Norm | -0.011807332 | 0.127006483 | 1 |  |  |
| Dept\_Norm | -0.031882283 | 0.063990596 | -0.002639604 | 1 |  |
| Dist\_Norm | -0.00168612 | 0.077923583 | -0.009696041 | 0.017224804 | 1 |
| EduLevel\_Norm | 0.208033731 | -0.03137282 | -0.00866984 | 0.007996422 | 0.021041826 |
| Environment\_Norm | 0.010146428 | -0.103368978 | -0.011309935 | -0.019395271 | -0.016075327 |
| Gender | -0.03631055 | 0.029453253 | -0.044895513 | -0.04158329 | -0.001850528 |
| JobInvolvement\_N | 0.029819959 | -0.130015957 | 0.029299959 | -0.024586062 | 0.00878328 |
| JobLvl\_Norm | 0.509604228 | -0.169104751 | -0.01169583 | 0.101963106 | 0.005302731 |
| JobSatisfaction\_N | -0.004891877 | -0.103481126 | 0.008665959 | 0.021000879 | -0.003668839 |
| Marital\_Norm | 0.09502891 | -0.162070235 | -0.030914988 | -0.056073435 | 0.014437031 |
| MTH\_INCOME\_N | 0.497854567 | -0.159839582 | -0.013449947 | 0.053129698 | -0.017014445 |
| CompWorked\_Norm | 0.299634758 | 0.043493739 | -0.030742984 | -0.035881612 | -0.029250804 |
| OverTime | 0.028062357 | 0.246117994 | 0.042751516 | 0.007480968 | 0.025513635 |
| %SalaryHike\_Norm | 0.003633585 | -0.013478202 | -0.025726943 | -0.007840161 | 0.040235377 |
| Performance\_Norm | 0.001903896 | 0.002888752 | 0.001683337 | -0.024603543 | 0.027109618 |
| RSSatisfaction\_Norm | 0.05353472 | -0.045872279 | 0.008926303 | -0.022414425 | 0.006557475 |
| StockOption\_Norm | 0.037509712 | -0.137144919 | -0.028256953 | -0.012192914 | 0.044871999 |
| TotWorkingYrs\_N | 0.680380536 | -0.171063246 | 0.00797211 | -0.015761512 | 0.004628426 |
| TimesTrainingLastYr\_N | -0.019620819 | -0.059477799 | 0.016357219 | 0.036875066 | -0.036942234 |
| WorkLifeBal\_Norm | -0.021490028 | -0.063939047 | 0.004208799 | 0.026382525 | -0.026556004 |
| YearsAtCo\_Norm | 0.31130877 | -0.134392214 | 0.005212128 | 0.022920442 | 0.00950772 |
| YrsSincePromo\_Norm | 0.216513368 | -0.033018775 | 0.00522202 | 0.040060967 | 0.010028836 |
| YearsWithMgr\_Norm | 0.202088602 | -0.156199316 | -0.000228512 | 0.034282473 | 0.014406048 |

*Correlation matrix only shown up to column F*

After producing the correlation matrix and using data to form charts in pivot fields, it was found that 2 particular columns, age and monthly income, had too many data points to use to form neat charts. Steps taken to make this data more useful, and easy to interpret include:

* Using IF function to sort age into age groups
* Converting monthly income to income per annum
* Grouping the annual incomes into ranges

This data was then used to create *attrition by age groups* and *attrition by income categories* shown in this report below.

Project Visualization Techniques: (List and describe the visualization techniques considered along with justification why they were used or not used)

Pivot charts and pie charts are mainly used to ensure clarity in visualizing the data. While the data shown is simple, generous use of slicers make the dashboard more interactive and allow greater detail to be seen by anyone who sees this as a presentation or to use as a tool for further analysis.

Data Set Analysis: (Analyse and list down your observations of the 5-6 strongest factors related to attrition in charts.)

Out of those who Do Overtime, 31% left the company.

Out of those who left the company 53.9% do overtime.

This chart shows age groupings and which group has the highest total populations, and also which group has the highest attrition rate. It has also been filtered (by slicer) to show only those who do overtime. In the legend Orange ‘Yes’ shows the count of people who have left the company. Blue ‘No’ shows the count of people who have stayed.

This chart shows Income Categories, once again filtered to show only those who do overtime.

This chart shows that *many leave in their first year of work*, especially after filtering to show only those who work at *job level 1*, and only showing those who do *overtime*. From playing with the filters, we also discover that most of the attrition comes from Research and Sales. Strangely if you toggle the slicers to show broader data that shows all the job levels, many people leave in year 10 as well.

This chart shows that *attrition* is inversely related to *stock option level*. This relationship is especially true for option levels 0 and 1.

Data Visualization: (Explain the process in which data visualization has been accomplished using the spreadsheet which you have created with data visualization)

