**Group Name: Teamwork Pays**

**Group 9 Project Assignment report**

**1. Iwuji Emmanuel Chukwuebuka (M)**

**2. Ilemona Muhammed Ebiloma (M)**

**3. Anas Muhammad Abdulmumini (M)**

**4. Igbafe Sylvia (F)**

**5. Nonso N (F)**

**INTRODUCTION**

In the context of today's data-driven landscape, organizations increasingly depend on data analysis to guide decision-making processes. This project focuses on the application of data analysis techniques utilizing Microsoft Excel, leveraging a real-world dataset titled "Employee.csv" for analysis.

**Aim**

The aim of this report is to summarize the data, analyze the trends, evaluate performance and make business decisions based on the dataset provided which can be examined by management.

**Objectives**

* Apply descriptive statistics to summarize the dataset and identify key trends.
* Perform hypothesis testing to validate assumptions and explore relationships between different variables.
* Use inferential statistics to draw conclusions about the employee population from the sample data.

**METHODOLOGY AND RESULTS**

**Descriptive Statistics**

The total number of employees equate to **1470** and they are distributed in 3 separate departments (Research and Development, Sales and Human Resources). Using basic functions of Excel to calculate the mean, median and mode, the ages are first rearranged in ascending order from 18 – 59 to enable accurate results. The mean is the average age of the employees, the median is the middle value of ages when sorted and the mode is the most frequently occurring age. Respective results were given as Mean = 37, Median = 36 and Mode = 35

The mean, median and range of the salaries (monthly income) was also generated and respective results are as follows: Mean: 6503, Median: 4919 and Range which is the maximum value – minimum value: 19999 – 1009 = 18990

The standard deviation identifies the how much salaries deviate from the mean salary. A high standard deviation suggests a wide range of salaries, while a low standard deviation indicates that salaries are closely clustered. The standard deviation was calculated as 4706.355165. which implies that the monthly income are closely clustered

The frequency table for employees in each department shows the number of employees that work in each department as seen in Sheet 4 on the report. Results show that the Research and Development department have more employees than the Sales and Human Resources departments.

Gender distribution of employees which separates the number of males and females, results are as follows: Males: 882 and Female: 588

**Hypothesis Testing**

To solve for the average salary of male vs female employees, a t-test is implemented to determine if there is a significant difference between the average salaries and of male and female employees. The data is organized into groups of male and female, then loaded on Analysis Toolpak which is enabled and the data is keyed in, Excel will output a table which will include the p-value which is compared. If p > 0.05, reject the null hypothesis (meaning significant difference exists between groups). If p < 0.05, fail to reject the null hypothesis (no significant difference)

Results show that P value = 0.221836878 which means that there IS significant difference between the salaries of male and female employees.

The proportion of male employees in Sales department is determined by a one sample proportion test to check if the proportion of male employees in the sales departments exceeds 50% which is indicated in Table 1.1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Department** | **Male** | **Female** | **Total** |
| Human Resources | 43 | 20 | 63 |
| Research and Development | 582 | 379 | 961 |
| Sales | 257 | 189 | 446 |
| Total | 882 | 588 | 1470 |

Table 1.1 Gender distribution among departments

**ANOVA for Salary differences**: among departments was performed to compare and check for any significant differences across the different departments. The result amounted to p = 0.04, therefore there IS significant difference in the salaries across different departments

**Independent T-Test**: A statistical test used to compare the means of two independent groups to determine if there is a statistically significant difference between them. For one tail upper t test 0.05 at 9 degree of freedom = 1.83 T > t 0.05 reject hypothesis recognition has enhanced product. In this case where **the average salary of male vs. female, average monthly income based on years at the company, average salary based on education and average performance rating based on marital status** and the results yielded values greater than 0.05, therefore the null hypothesis was NOT rejected.

**Chi-Square Test of Independence**: Assesses whether two categorical variables are independent of each other, based on the frequencies observed in a contingency table.

If the p-value from the test is less than a chosen significance level (0.05), the null hypothesis is rejected, indicating that a significant relationship exists between the variables. The results yields no significant difference.

**Inferential Statistics**

The confidence interval of 95% in two statistical cases brought about certain conclusions in two distinct cases.

Firstly, we were able to confidently determine the average employee salaries through a range of 40 which was derived from a margin of error of 240.87.

Secondly, it was established that the true employee tenure falls within a range determined by a margin of error of 0.313445196