Project Title: The Impact of Work Location on Employees: A Comparative Analysis of Remote and On-Site Work.

Introduction: The motive of this project is to identify and analyze the mental health of an employee in his workplace. The outcome will suggest employee survivability in the workplace. By using data analysis, correlation, classifications the outcome will be predicted.

The goals and objectives:

1. Analyzing the data
2. Data cleaning and preprocessing.
3. Data visualization.
4. Predictive modelling.
5. Descriptive analysis
6. Hypothesis testing.

Task Definition

Problem Statement:

The Concept of the dataset is to categorize the employees with respect to their level of work stress in their professional lives. After the pandemic the employees are asked to come back to the office, which increases their stress levels, while few are happy to go back to their regular routine, but few tend to drop due to the pressure level and lack of comfort from the workstation. This will be a huge change in the work life of an employee by bring this drastic change and will increase the amount of frazzle an employee will go through.

Input/Output Examples:

The data set comprises of 9 columns in which 3 major categories will act as the input, while the decision which will be in the format of descriptive will be the output of the project. An employee with WFH resources and less stress and less burnout are more likely to continue his working from home, while a person with increased stress levels and no resources will tend to resign from the job, and the guy with medium stress and no resource allocation will tend to go back to the office.

Approach

The output of the task will be done using the predictive analysis and comparative analysis.

Proposed Method

Analyzing the data, Data cleaning and preprocessing, Data visualization, Predictive modelling Descriptive analysis, Hypothesis testing, Linear Regression Models
In this project we are going to add a t-test which will identify whether the employee will continue working for the office or will he/she quit by comparing it with the null-hypothesis.

Evaluation

Evaluation Metric

T-test (hypothesis testing), one-way anova, linear regression, logistic regression, random classifier, average, mean, median, variance, standard deviation, f-1 score, accuracy, Precision, Recall, ROC curve, AUC score.

Data

Dataset Description:

1. **Data collection**
   - The dataset is acquired from Kaggle and has the size of 204750. The attributes like employee id, date of joining, gender, company type, WFH setup availability, designation, resource allocation, mental fatigues, burn rate. (22750,9)
   - The dataset contains integers, strings, Boolean, date, float, and a few NULL values.
   - Due to the large size of the data the preprocessing is done.
2. **Pre-Processing**
   - Handling the missing / NULL values.
   - Updating the new dataset
   - Shape and size reduced to 169726 and (18254,9)
   - Making it easy to analyze

3. **Analyzing the data**
    - Various techniques of analysis are used to understand the data like head, counting the data and finding the shape and size of the dataset.
    - Highlighting the main attributes used to perform further tests.
    - The data is found to have dependent and independent attributes that will be used in further analysis.

Motivation and Significance:

The motivation of this project is to consider the stress levels of employees in the organization who work from the office and whether they prefer to work from home. High stress levels can lead to increased risk of health problems in the employees of the organization.

The significance of this project lies in the decrease in productivity of employees due to the higher stress levels and if at all there is a decrease if they work from home.

Literature Survey:

In this project, we have gone through extensive research covering factors such as stress levels of employees, their designations in the company, also the resources allocated to them each day and many more. Based on these factors, we can develop strategies to decrease the stress levels of employees in the organization and map an approach for a hybrid work environment.

Recommendations

- Most of the employees who have their stress levels high are those who work on site. So, we recommend a hybrid work environment approach to deal with these stress levels of employees.
- Employees who have their resources very high resources allocated to them have been observed with higher stress levels than those employees who have very few resources allocated to them. So, it's recommended to share the resources between the employees.

- The burnout factor of employees is high who work from the office and have a greater number of resources allocated to them. So, it's recommended for them to work from home.

Conclusion:

This project aims to provide powerful insights and predictive models on the employees' choice of whether to work from home or to work on site based on their stress levels. By analyzing the underlying patterns and utilizing appropriate classifiers, organizations can make crucial decisions regarding the deployment of a hybrid work environment.

This is the initial proposal of our project indicating the expected outcomes and highlighting its objectives and methodologies.

Plan and Team Roles

Team Members:

1. Mohammad Faiyaz Pasha – 11629829 **(Team Leader)**
2. Shyam Rahul Chennupati - 11639795
3. Sai Adhinatha Reddy Gona - 11724867
4. Sujan Lanka -11702061

Division of Work:

Mohammad Faiyaz Pasha **–** Assigned work to team members. (motivation and significance)

Shyam Rahul Chennupati – Project Proposal **(**Literature Survey, Recommendations, conclusions)

Sai Adhinatha Reddy Gona – Project Proposal (Dataset Description, Approach, and evaluation metrics)

Sujan Lanka – Project Proposal (Introduction, Task definition, Goals, and objectives)