Project Description

Name: Aws Alyasjeen

Project Name: “From Office to Home: Remote Work’s Impact on Productivity and Resignation”

## Problem Statement:

Remote work has become a cornerstone of modern workplace dynamics, particularly in the wake of the COVID-19 pandemic. Businesses across industries are increasingly adopting hybrid and fully remote work models, leading to significant changes in employee productivity, satisfaction, and turnover rates. However, the long-term effects of remote work on these key employee outcomes remain unclear, leaving companies uncertain about how to optimize their remote work policies.

Using this dataset, you will investigate how remote work frequency correlates with employee performance, job satisfaction, and resignation rates. The objective is to uncover actionable insights into how different levels of remote work impact employees’ productivity and likelihood of resigning. With this data, the project aims to help organizations design evidence-based policies that balance productivity and employee retention while supporting a flexible work culture.

## Possible Impact of Your Analysis:

This problem matters because remote work has fundamentally reshaped workplace dynamics, and its long-term effects on employee productivity, satisfaction, and retention are still not fully understood. By investigating how remote work frequency correlates with these outcomes, this analysis will provide actionable insights that organizations can use to:

1. **Enhance Employee Retention:** Understand how remote work policies influence resignation rates and take proactive steps to retain talent.
2. **Optimize Productivity:** Identify the optimal balance of remote work frequency that maximizes employee performance while minimizing burnout.
3. **Improve Job Satisfaction:** Address key pain points associated with remote work, such as isolation and overwork, to foster a positive work environment.
4. **Support Evidence-Based Policy Making:** Equip HR leaders and decision-makers with data-driven insights to design flexible work arrangements that benefit both employees and the organization.
5. **Change in employment location over time(time series analysis):** In this dataset, I have a hire date column, so I plan to track the number of employees working from home at different time intervals. By grouping the data based on hire dates, I can analyze trends in hiring for remote roles over time. For example, I could aggregate the data monthly or yearly to observe the cumulative growth of remote workers or identify spikes and shifts in hiring practices. This approach will also allow me to explore correlations with external events, such as changes in company policies or global trends. While this is just a brief overview, the full analysis will include visualizations to clearly communicate these trends.

Ultimately, this analysis has the potential to help businesses adapt to evolving workplace trends while maintaining a competitive edge in attracting and retaining top talent.

## Dataset(s):

(Include information on where your dataset is from, the license for using the dataset, and a brief description of the variables included in your dataset.)

[Source:https://www.kaggle.com/datasets/mexwell/employee-performance-and-productivity-data/data](https://www.kaggle.com/datasets/mexwell/employee-performance-and-productivity-data/data)

Dataset size : 100,000 rows & 20 Columns .

This dataset contains 100,000 rows of data capturing key aspects of employee performance, productivity, and demographics in a corporate environment. It includes details related to the employee's job, work habits, education, performance, and satisfaction. The dataset is designed for various purposes such as HR analytics, employee churn prediction, productivity analysis, and performance evaluation.

* Employee\_ID: Unique identifier for each employee.
* Department: The department in which the employee works (e.g., Sales, HR, IT).
* Gender: Gender of the employee (Male, Female, Other).
* Age: Employee's age (between 22 and 60).
* Job\_Title: The role held by the employee (e.g., Manager, Analyst, Developer).
* Hire\_Date: The date the employee was hired.
* Years\_At\_Company: The number of years the employee has been working for the company.
* Education\_Level: Highest educational qualification (High School, Bachelor, Master, PhD).
* Performance\_Score: Employee's performance rating (1 to 5 scale).
* Monthly\_Salary: The employee's monthly salary in USD, correlated with job title and performance score.
* Work\_Hours\_Per\_Week: Number of hours worked per week.
* Projects\_Handled: Total number of projects handled by the employee.
* Overtime\_Hours: Total overtime hours worked in the last year.
* Sick\_Days: Number of sick days taken by the employee.
* Remote\_Work\_Frequency: Percentage of time worked remotely (0%, 25%, 50%, 75%, 100%).
* Team\_Size: Number of people in the employee's team.
* Training\_Hours: Number of hours spent in training.
* Promotions: Number of promotions received during their tenure.
* Employee\_Satisfaction\_Score: Employee satisfaction rating (1.0 to 5.0 scale).
* Resigned: Boolean value indicating if the employee has resigned.

[License - CC0: Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)

Project Scoping Document

## Business Problem

Remote work has become an integral part of modern employment, with employees increasingly requesting it as part of their contracts and companies recognizing its potential benefits. This shift has significant implications for productivity, employee satisfaction, and retention. By studying the relationships between work location, productivity, comfort, and other key factors, this analysis aims to uncover actionable insights.

Discussing this topic will empower employers to make informed decisions about workplace policies, reduce uncertainties surrounding remote work, and provide a clearer understanding of its broader impacts on business performance and employee well-being.

## Business Impact

The analysis will contribute to the growing body of knowledge on remote work by providing actionable insights that benefit both employees and organizations. It will demonstrate how remote work influences productivity, job stability, and overall workplace sustainability. Additionally, it will highlight the significant impact of workplace policies on employees’ real-life experiences.

Since remote work is still a relatively new phenomenon, this analysis is particularly timely. It will help break down existing fears and misconceptions about remote work and show that, when managed effectively, it can yield positive outcomes for all parties involved.

## Dataset(s)

# **Employee Performance and Productivity Data**

### Strengths:

This dataset includes 100,000 rows of data, capturing critical aspects of employee performance, productivity, and demographics within a corporate environment. Its comprehensive nature provides a robust foundation for analyzing the relationships between remote work, employee satisfaction, and resignation trends.

### Weaknesses:

While the dataset offers valuable insights, it has some limitations:

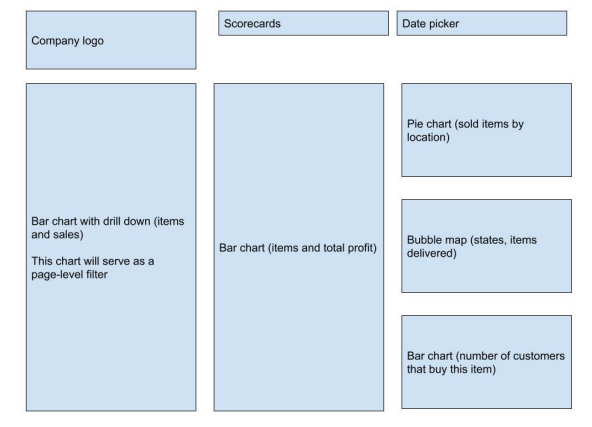
1. **Irrelevant Data:** Certain fields may not contribute directly to this analysis and will need to be filtered out.
2. **Missing Data:** There may be gaps in key variables that require cleaning before analysis.
3. **Lack of Additional Context:** The dataset does not include important factors such as the cost of commuting or details about employees’ conditions before and after transitioning to remote work, which could enhance the depth of the analysis.

## Methods

* Inspecting general insights about the dataset (Excel).
* Checking for missing or irrelevant data and cleaning it (Excel).
* Calculating descriptive statistics like averages, distributions, and correlations (Excel).
* Visualizing relationships between key variables such as Remote\_Work\_Frequency, Resigned, and Employee\_Satisfaction\_Score (Excel).
* Creating dashboards to illustrate trends and insights (Tableau).
* Compiling findings into a Datafolio for clear communication.
* Developing the Final Report with actionable insights and recommendations.
* Reviewing and proofreading all project documents.

## Dashboard

The dashboard will display key metrics, highlight resignation trends, and show the impact of remote work on employee productivity and satisfaction.



## Milestones

* Identify characteristics of variables.
* Clean and preprocess data.
* Perform exploratory data analysis (EDA) to uncover initial trends.
* Create visualizations to illustrate relationships between variables.
* Analyze correlations between remote work frequency, satisfaction, and resignation rates.
* Develop an interactive dashboard.
* Write and finalize the project report with key insights and recommendations.

## Timeline

|  |  |
| --- | --- |
| Week | Tasks |
| Week 1 | Project Description & Project Scoping |
| Week 2 | Data Curation |
| Week 3 | Exploratory Data Analysis |
| Week 4 | Datafolio & Dashboard |
| Week 5 | Final Report |

**From Office to Home: Remote Work’s Impact on Productivity and Resignation**

**Data Curation** **By Aws Alyasjeen**  
  
**General Dataset Information:**

File name:Extended\_Employee\_Performance\_and\_Productivity\_Data.csv

Description: This dataset contains information for 100,000 employees hired between 2014 and 2024 in a corporate environment. It provides comprehensive details related to their workplace, performance, productivity, and demographics.

Dataset Details: 100,000 rows & 20 Columns.

Size: 10,969.6 KB (10.7 MB)

Source: Kaggle - [Dataset\_link](https://www.kaggle.com/datasets/mexwell/employee-performance-and-productivity-data/data)

**Data Profile:**

* Download the dataset and open it in Excel.
* applied a filter in Excel to extract unique values from the specified columns, YEAR (), UNIQUE(), Conditional Formatting,MIN(),MAX(),pivot table,CountIf/Ifs.

**Notable Features:**

* 9 Unique Departments: Customer Support, Engineering, Finance, HR, IT, Legal, Marketing, Operations, Sales.
* 7 Unique Job Title: Analyst, Consultant, Developer, Engineer, Manager, Specialist, Technician.
* 4 Unique Education Level: High School, Bachelor, Master, PhD.

**Recorded Inconsistency:**

* Hire\_Date: It contained the appointment date, year, day, and hour, and we only needed the year number make it better.
* The dataset was thoroughly checked, and no null values or inconsistencies were identified, ensuring data completeness and integrity.
* Some columns initially contained numeric values, which were converted to text for improved clarity and better interpretability.
* The Gender column initially contained three values: Male, Female, and Other. Rows with the "Other" category were removed entirely to improve analysis and visualization clarity.
* Years\_At\_Company column contained some values of 0, representing employees with no tenure. These values were categorized into Junior (0-3 years) to provide more meaningful insights for analysis.

**Data Wrangling:**

* The **Performance\_Score** (1 to 5 scale) was categorized into three groups: **Low Performer (1-2)**, **Average Performer (3)**, and **High Performer (4-5)** to ensure a balanced distribution for analysis.
* The **Monthly\_Salary** (ranging from 3,850 to 9,000) was categorized into three groups: **Low Salary (3,850-5,500)**, **Medium Salary (5,501-7,000)**, and **High Salary (7,001-9,000)** for better analysis and comparison.
* The **Remote\_Work\_Frequency** was categorized into three groups: **Onsite (0%)**, **Hybrid (25%-75%)**, and **Remote (100%)** to better represent work arrangements.
* The **Employee\_Satisfaction\_Score** (1.0 to 5.0 scale) was categorized into three groups: **Dissatisfied (1.0-2.5)**, **Neutral (2.6-3.5)**, and **Satisfied (3.6-5.0)** to facilitate analysis of employee sentiment.
* The **Years\_At\_Company** (0-10 years) was categorized into three groups: **Junior (0-3 years)**, **Senior (4-7 years)**, and **Expert (8-10 years)** to reflect career progression and tenure.
* The **Hire\_Date** column, initially in the format 1/19/2022 8:03:06 AM, was simplified using the YEAR function in Excel to display only the year, resulting in 2022 for easier analysis of annual trends.
* The cleaned data was isolated in a new worksheet, while the deleted rows were moved to a separate worksheet for reference. The original dataset contained 100,000 rows, and after removing 3,968 rows (due to the exclusion of "Other" from the Gender column), the cleaned dataset now contains **96,032 rows**.

**Data Table Schema**:

|  |  |  |  |
| --- | --- | --- | --- |
| **Field** | **Type before Wrangling** | **Description** | **Type After Wrangling** |
| employee\_id | INTEGER | Unique identifier for each employee (e.g., 1, 2, 3). | INTEGER |
| department | STRING | The department in which the employee works (e.g., Sales, HR, IT). | STRING |
| gender | STRING | Gender of the employee (Male, Female, Other). Rows with "Other" were removed. | STRING |
| age | INTEGER | Employee's age, ranging between 22 and 60. | INTEGER |
| job\_title | STRING | The role held by the employee (e.g., Manager, Analyst, Developer). | STRING |
| hire\_date | DATE (e.g., MM/DD/YYYY HH:MM:SS AM/PM) | The date the employee was hired (e.g., 1/19/2022 8:03:06 AM). Transformed to display only the year (e.g., 2022). | INTEGER (Year) |
| years\_at\_company | INTEGER | The number of years the employee has been working for the company. Categorized into **Junior (0-3 years)**, **Senior (4-7 years)**, and **Expert (8-10 years)**. | STRING |
| education\_level | STRING | Highest educational qualification (e.g., High School, Bachelor, Master, PhD). | STRING |
| performance\_score | INTEGER | Employee's performance rating on a 1 to 5 scale. Categorized into **Low (1-2)**, **Average (3)**, and **High (4-5)**. | STRING |
| monthly\_salary | FLOAT | The employee's monthly salary in USD, correlated with job title and performance score. Categorized into **Low (3,850-5,500)**, **Medium (5,501-7,000)**, and **High (7,001-9,000)**. | STRING |
| work\_hours\_per\_week | INTEGER | Number of hours worked per week. | INTEGER |
| projects\_handled | INTEGER | Total number of projects handled by the employee. | INTEGER |
| overtime\_hours | INTEGER | Total overtime hours worked in the last year. | INTEGER |
| sick\_days | INTEGER | Number of sick days taken by the employee. | INTEGER |
| remote\_work\_frequency | STRING | Percentage of time worked remotely (0%, 25%, 50%, 75%, 100%). Categorized into **Onsite (0%)**, **Hybrid (25%-75%)**, and **Remote (100%)**. | STRING |
| team\_size | INTEGER | Number of people in the employee's team. | INTEGER |
| training\_hours | INTEGER | Number of hours spent in training. | INTEGER |
| promotions | INTEGER | Number of promotions received during their tenure. | INTEGER |
| employee\_satisfaction\_score | FLOAT | Employee satisfaction rating on a 1.0 to 5.0 scale. Categorized into **Dissatisfied (1.0-2.5)**, **Neutral (2.6-3.5)**, and **Satisfied (3.6-5.0)**. | STRING |
| resigned | BOOLEAN | Boolean value indicating if the employee has resigned (e.g., TRUE for resigned, FALSE for active). | BOOLEAN |
|  |  |  |  |