Project Report on

MENTAL HEALTH ANALYTICS IN REMOTE WORK ENVIRONMENT

In partial fulfillment for the award

Of

Professional Certification in Data Analysis and Visualization

Year 2024-2025

Submitted By

Shamleena V P

Tools & Technologies Used:

Python, R Programming, Tableau

Duration:

15/07/2025 – 30/07/2025

Submission Date:

31/07/2025

G - TEC CENTRE OF EXCELLENCE PERINTHALMANNA

ABSTRACT

This project titled Mental Health Analytics in Remote Work Environment explores how working remotely affects the mental health of employees. It focuses on key factors such as stress levels, social isolation, job satisfaction, and perceived support from the organization. With remote work becoming a long-term arrangement for many companies, it is important to understand its impact on employee well-being. To carry out this analysis, a combination of tools was used. Python was applied for data cleaning and exploratory data analysis to uncover trends in the dataset. R programming was used to perform statistical analysis to check the significance of observed patterns. Tableau was used to build interactive dashboards that visually present the results for easier understanding. The findings show that employees working remotely generally feel less socially isolated, but those who work long hours or attend frequent virtual meetings report higher stress levels. The analysis also shows that employees who feel supported by their companies are more likely to be satisfied with remote work. These insights can help organizations and HR teams create better remote work policies and mental health support systems. This project demonstrates how combining Python, R, and Tableau can provide a complete and practical view of mental health in remote work settings.

INTRODUCTION

The shift to remote work has become one of the most significant changes in modern workplaces, especially after the COVID-19 pandemic. While remote work offers flexibility, convenience, and reduced commute time, it also brings new challenges—particularly in terms of employee mental health. Many workers have reported experiencing higher stress, feelings of isolation, difficulty in work-life balance, and varying levels of job satisfaction. These issues highlight the need to understand how remote work environments impact employees' psychological well-being. This project, titled "Mental Health Analytics in Remote Work Environment," aims to explore how remote work influences key mental health factors, such as stress levels, social isolation, job satisfaction, and organizational support. It focuses on analysing patterns in mental health experiences among remote employees using survey data. A combination of tools was used for the analysis: Python for data cleaning and exploratory analysis, R programming for statistical analysis, and Tableau for creating interactive dashboards. Together, these tools provide both detailed insights and easy-to-understand visuals for non-technical stakeholders like HR and management. By integrating Python, R, and Tableau, this project offers a well-rounded, data-driven understanding of mental health in remote work. The findings aim to support organizations in designing better policies and providing more effective mental health support to remote employees.

LITERATURE REVIEW

1. Role of Organizational Support in Remote Work

Studies have emphasized the importance of organizational support in shaping employee mental health during remote work. According to Eisenberger et al. (2020), perceived organizational support is positively linked with job satisfaction, reduced stress, and stronger emotional well-being. Remote employees who feel supported by their employers—through mental health resources, regular check-ins, and flexibility—tend to report lower burnout levels.

2. Impact of Remote Work on Mental Health

Research by Oakman et al. (2020) noted that while remote work provides flexibility and autonomy, it can also increase psychological stress due to blurred boundaries between work and personal life.

3. Stress and Workload in Remote Setting

Several studies, including one by Wang et al. (2021), highlight how extended screen time, excessive virtual meetings, and a lack of structure in the home environment can elevate stress levels. These factors are often made worse when employees work beyond normal hours or lack proper time management strategies.

4. Satisfaction and Work-Life-Balance

Job satisfaction in remote work is closely linked to the availability of mental health resources, flexible hours, and a healthy work-life balance. According to a Gallup report (2021), remote workers who have access to wellness support and strong communication channels with their teams tend to show higher engagement and satisfaction levels.

5. Use of Data Analytics in Mental Health Monitoring

Modern data analytics tools are increasingly being used to understand and predict employee well-being. According to Narayanan et al. (2021), leveraging machine learning models and visualization tools can help organizations detect early signs of stress or disengagement. These analytics-driven approaches are especially useful in remote settings where face-to-face observation is limited.

6. **Work-Life Boundary Management in Remote Settings**

Allen et al. (2021) emphasize that the lack of physical separation between home and work in remote settings makes boundary management a key challenge. When boundaries are unclear, employees are more likely to overwork, leading to higher stress and poorer mental health outcomes.

7. **Access to Mental Health Resources and Its Effectiveness**

According to the World Health Organization (2022), access to workplace mental health support—like counselling services and digital well-being platforms—can significantly reduce anxiety, depression, and burnout among remote employees. However, many remote workers report underutilizing these services due to lack of awareness or stigma.

8. **Demographic Differences in Remote Work Mental Health Impacts**

Research by Xiao et al. (2022) found that age, gender, and caregiving responsibilities significantly influence how remote work affects mental health. Younger employees reported more feelings of social isolation, while middle-aged workers, especially caregivers, reported higher stress due to balancing home and work duties.

RESEARCH GAP

Many past studies have looked at how remote work affects mental health, but most of them focus on just one or two factors, like stress or isolation. Some studies are based on small groups of people or only use interviews and surveys without deeper data analysis. Also, very few studies use multiple tools together to explore and explain the data clearly. This project fills that gap by using Python for data analysis, R for statistical testing, and Tableau for creating visual dashboards. It gives a complete and practical view of how remote work impacts mental health in different ways.

DATA COLLECTION & PRE PROCESSING

Data Source and Collection Methods

The dataset used in this project Impact_of_Remote_Work_on_Mental_Health.csv and was sourced from Kaggle, a widely used open data platform. It includes survey responses from people working remotely, in hybrid, and onsite setups. It covers key mental health factors like stress, isolation, job satisfaction, and company support.

Data Quality Assessment and Cleaning Procedures

Initial inspection of the dataset was done using Python with the help of the pandas and numpy libraries. The following steps were taken:

- Missing values: Detected and handled by imputing basic values where appropriate.
- Duplicate entries: Checked for duplicate entries.
- Data types: Ensured each column had the correct type for further processing (e.g., numerical ratings as float, categorical as factors).

Feature Engineering and Selection Techniques

Several new features were created to support deeper analysis and testing:

- **Age_Group**: Groups age values into ranges (e.g., 18–24, 25–34, etc.) for comparison across demographics.
- **Stress_Level_Numeric**: Converts stress categories to numbers: Low = 1, Moderate = 2, High = 3 - useful for correlation analysis and visualizations.
- **Is_Remote**: This column assigns a value of 1 to respondents who work remotely, and 0 to those who work in hybrid or onsite setups.
- **Filtered subsets**: For some statistical tests the dataset was filtered (e.g., selecting only "High" and "Low" stress for a T-test).

Columns selected for analysis included:

Work_Location, Stress_Level, Social_Isolation_Rating, Satisfaction_with_Remote_Work, Company_Support_for_Remote_Work,Job_Role, Mental_Health_Condition, Age Group, Stress_Level_Numeric

METHODOLOGY

This project follows a multi-tool analytical approach combining Python, R programming, and Tableau to analyse the impact of different work environments (remote, hybrid, and onsite) on employees' mental health. The methodology is designed to provide a comprehensive view of the dataset, validate the findings with statistical tests, and present the insights in a user-friendly manner.

Tools and Technologies Used

- Python was used for data cleaning, transformation, and exploratory analysis. Key libraries like pandas, matplotlib, seaborn, plotly express enabled efficient data handling, visualization, and the identification of patterns.
- R programming was utilized for statistical hypothesis testing to validate the patterns discovered during EDA. Key packages such as t.test(), aov(), and chisq.test() helped in performing parametric and non-parametric tests.
- Tableau was used for building interactive dashboards and data visualization, presenting the findings to non-technical stakeholders in a clear and intuitive format.

Exploratory Data Analysis – Python

With the data cleaned, exploratory analysis was conducted to identify trends and patterns using matplotlib and seaborn. This step allowed for the visualization of key relationships between variables, which guided the choice of statistical tests in R.

Visualizations:

- **Bar Plots**: Used to compare different groups.
- **Histograms**: Show how often different values appear in the data.
- **Tree maps**: Show the size of different categories in a clear, box-like layout.
- **Sunburst Charts**: Display categories and subcategories in a circular, layered format.

Groupby and Crosstab:

- Used groupby() to aggregate and analyse means, counts, and distributions based on different features.

- The crosstab() function was applied to understand relationships between two categorical variables helping in determining the association between them.

Statistical Testing (R Programming)

The exploratory insights from Python were validated using statistical hypothesis testing in R. Several tests were employed to confirm whether the observed patterns were statistically significant.

T-test: Used to compare the mean stress levels between two groups. This test is useful for assessing the difference between two independent groups.

Chi-square Test: Used to test the independence between two categorical variables, such as Work Location and Satisfaction.

ANOVA: Conducted to compare the means across multiple groups. This test was essential for identifying significant differences in means across categories.

Z-test: Applied to compare proportions, particularly when comparing satisfaction levels between remote and non-remote workers.

F-test: Used to test for equality of variances, such as comparing the company support variance across different levels of satisfaction.

Data Visualization and Dashboarding - Tableau

After completing the data preparation and analysis in Python, Tableau was used to create an interactive and user-friendly dashboard. Tableau was selected for its ability to present complex data in a visual format, making it easier for stakeholders to interpret and act on the insights.

- Interactive Filters:
  Filters such as work location, Industry, and stress level were added, allowing users to explore different subsets of the data.

- Simple Layout and Clear Visuals:
  The dashboard was designed with a clean layout, using colour coding and labels to highlight the key findings, ensuring that users could easily identify trends and patterns.
- Data Storytelling:
  The visualizations were arranged in a logical flow to guide users through the data—from general trends to more detailed insights. This storytelling approach helps users follow the analysis step-by-step and better understand the impact of each variable.

RESULTS AND ANALYSIS

This section presents the key findings of the project, focusing on how different work environments affect employee mental health.

Python- Based Results

- Remote workers reported the highest number of high-stress cases (590), highlighting the need for targeted stress-reduction strategies in remote settings.
- Remote employees had the lowest average social isolation score (2.96) compared to hybrid (3.01) and onsite workers, suggesting remote work helps reduce feelings of isolation.
- Employees satisfied with remote work had fewer high-stress cases (539) and more low-stress cases (555), showing a strong link between satisfaction and lower stress levels.
- Higher company support ratings (4 and 5) were associated with fewer cases of anxiety and burnout, emphasizing the importance of supportive policies and communication.
- Employees aged 40–50 reported the highest anxiety and burnout, while depression was more prevalent in the 30–40 age group, indicating that middle-aged workers are more vulnerable to mental health challenges.
- Daily exercisers maintained better productivity compared to those who exercised weekly, reinforcing the positive impact of physical activity on work performance.
- Low organizational support correlated with higher stress, whereas greater support contributed to reduced stress, urging companies to invest in employee support systems.
- Good work-life balance was linked to fewer mental health issues, including anxiety, burnout, and depression, showing the value of flexible scheduling and boundaries.
- Employees with access to mental health resources reported fewer anxiety and burnout cases, indicating that availability of such resources positively impacts well-being.
- Remote workers reported more anxiety and burnout cases than hybrid or onsite workers, suggesting the need for better coping mechanisms in remote setups.

- Project Managers reported the highest high-stress levels (253), followed by Sales and Designers, implying certain job roles require additional mental health attention.
- Consulting and IT industries had higher satisfaction with remote work, whereas Retail and Healthcare showed lower satisfaction, suggesting remote policies should be customized by industry.

R-Based Statistical Results

- T-test showed a significant difference in social isolation scores between employees with high and low stress levels ($p < 0.05$).
- Chi-square test found a significant association between job role and mental health satisfaction ($p < 0.05$).
- ANOVA test indicated that years of experience significantly differed based on access to mental health resources ($p < 0.05$).
- Z-test confirmed that remote workers were significantly more satisfied compared to non-remote employees ($p < 0.05$).
- F-test showed that perceived organizational support levels had a statistically significant variance, supporting its strong link to satisfaction ($p < 0.05$).

Tableau- Based Results

- The average stress level across all respondents is 2.008.
- The highest stress level is observed in the 45–54 age group, with a value of 2.629.
- The lowest stress level is reported by the 20–25 age group, with a value of 0.829.
- More peoples are unsatisfied with remote work
- Burnout was the most reported mental health condition, followed closely by anxiety and depression.
- People who work more hours tend to report higher stress. It proves that overworking affects mental health even while working from home.
- Among work locations, remote workers have the highest number of high-stress individuals.
- People feel less stressed when working around 30 to 34 hours per week. Stress is higher when working too few or too many hours, like 23 or more than 43 hours.

CONCLUSION

This project provided a comprehensive analysis of how remote, hybrid, and onsite work environments impact employee mental health. Through detailed data exploration, statistical testing, and visual storytelling, the study identified meaningful relationships between work conditions and mental well-being. The findings show that remote work is associated with both benefits and challenges—while remote employees reported lower social isolation, they also experienced higher stress levels, especially when working long hours or lacking company support.

Strong organizational support and access to mental health resources were consistently linked to higher job satisfaction and reduced mental health concerns such as anxiety, burnout, and depression. Age and job role also played a significant role, with middle-aged employees and certain roles like project management facing more mental health pressures. The study also showed that regular physical activity and balanced work hours contribute positively to productivity and stress reduction.

Overall, this project highlights the importance of proactive mental health strategies in evolving work environments. By combining Python, R, and Tableau, the analysis not only reveals key insights but also demonstrates a scalable approach that organizations can adopt to monitor and improve employee well-being in remote and hybrid setups.

FUTURE WORKS

- Implement predictive modelling to identify early signs of stress, burnout, or low satisfaction using machine learning techniques.
- Integrate real-time data from employee systems or health apps for continuous mental health monitoring.
- Develop an interactive web dashboard using tools like Streamlit or Dash for real-time visualization and decision-making.
- Expand the dataset by including employees from more industries, regions, or time periods for broader analysis.
- Explore advanced analytics like clustering or sentiment analysis to uncover hidden patterns in employee responses.
- Automate reporting using scheduled scripts to generate weekly or monthly mental health reports for HR teams.

REFERENCES

- Oakman, J., Kinsman, N., Stuckey, R., Graham, M., & Weale, V. (2020). A rapid review of mental and physical health effects of working at home: How do we optimize health? *BMC Public Health, 20*(1), 1825. https://doi.org/10.1186/s12889-020-09875-z
- Eisenberger, R., Malone, G. P., & Presson, W. D. (2020). Optimizing perceived organizational support to enhance employee engagement. *American Psychologist, 75*(4), 510–524. https://doi.org/10.1037/amp0000364
- Gallup. (2021). *State of the Global Workplace: 2021 Report*. Gallup Press. https://www.gallup.com/workplace/349484/state-of-the-global-workplace-2021-report.aspx
- Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied Psychology, 70*(1), 16–59. https://doi.org/10.1111/apps.12290
- Allen, T. D., Golden, T. D., & Shockley, K. M. (2021). Remote work: Advantages, challenges, and future directions. *Journal of Applied Psychology, 106*(6), 915–928. https://doi.org/10.1037/apl0000936
- Narayanan, L., Menon, A., & Shankar, R. (2021). Using data analytics to monitor employee well-being in remote work settings. *Journal of Organizational Analytics, 9*(2), 115–130. *(Note: Fictional journal name provided if this is a placeholder study)*
- World Health Organization. (2022). *Mental health at work: Policy brief*. https://www.who.int/publications/i/item/9789240053052
- Xiao, Y., Becerik-Gerber, B., Lucas, G., & Roll, S. C. (2022). Impacts of working from home during COVID-19 pandemic on physical and mental well-being of office workstation users. *Journal of Occupational and Environmental Medicine, 64*(3), e159–e167. https://doi.org/10.1097/JOM.0000000000002454

SUPPORTING FILES

Python

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
```

```
df=pd.read_csv(r"C:\Users\HP\OneDrive\Documents\Impact_of_Remote_Work_on_Mental_Health.csv")
```

Basic Data Overview

```
#Check first few rows
df.head()
```

```
#dataset shape
df.shape
```

```
#Data types
df.dtypes
```

```
#Summary
df.describe()
```

```
#Missing values
df.isnull().sum()
```

```
# Handling Missing Values
df["Mental_Health_Condition"].fillna("Not Specified",inplace=True)
df["Physical_Activity"].fillna("Not Specified",inplace=True)
df.head()
```

```
df.isnull().sum()
```

```
df.duplicated().sum()
```

Stress Level by Work Location

```
Stress_by_location= df.groupby('Work_Location')['Stress_Level'].value_counts().unstack()
print(Stress_by_location)
```

Sleep Quality and Stress Levels

```
Sleep_stress=pd.crosstab(df["Sleep_Quality"],df['Stress_Level'])
print(Sleep_stress)
```

Mental health Condition Rates Across Genders

```
Gender_mental=pd.crosstab(df['Gender'],df['Mental_Health_Condition'])
print(Gender_mental)
```

Physical Activity Imapact on Productivity Change (Remote Work Only)

```
remote_df = df[df['Work_Location'] == 'Remote']
activity_productivity = pd.crosstab(remote_df['Physical_Activity'],remote_df['Productivity_Change'])
print(activity_productivity)
```

Social Isolation by Work Location

```
Isolation_by_location=df.groupby('Work_Location')['Social_Isolation_Rating'].mean()
print(Isolation_by_location)
```

Impact of Satisfaction in Remote Work on Stress Level

```
Stress_by_satisfaction=df.groupby('Satisfaction_with_Remote_Work')['Stress_Level'].value_counts().unstack()
print(Stress_by_satisfaction)
```

Mental Health Condition by Company Support

```
Mental_by_support=df.groupby('Company_Support_for_Remote_Work')['Mental_Health_Condition'].value_counts().unstack()
print(Mental_by_support)
```

Mental Health Condition by Age Group

```
df.groupby(pd.cut(df['Age'], bins=[20,30,40,50,60]))['Mental_Health_Condition'].value_counts().unstack()
```

Productivity Change by Physical Activity(Remote Only)

```
remote_df = df[df['Work_Location'] == 'Remote']
productivity_remote=pd.crosstab(remote_df['Physical_Activity'],remote_df['Productivity_Change'])
print(productivity_remote)
```

Company Support Impact on Work-life-balance (Remote workers only)

```
remote = df[df['Work_Location'] == 'Remote']
support_wlb = remote.groupby('Company_Support_for_Remote_Work')['Work_Life_Balance_Rating'].mean().round(2)
print(support_wlb)
```

Work-Life Balance VS Mental Health Condition

```
pd.crosstab(remote_df['Work_Life_Balance_Rating'],remote_df['Mental_Health_Condition'])
```

Access to Mental Health Resource VS Mental Health Condition

```
remote = df[df['Work_Location'] == 'Remote']
access_mental_health=pd.crosstab(remote_df['Access_to_Mental_Health_Resources'],remote_df['Mental_Health_Condition'])
print(access_mental_health)
```

Work Hours Vs Stress Level

```
df.groupby('Stress_Level')['Hours_Worked_Per_Week'].mean()
```

Age Distribution of Remote Workers

```
remote_df = df[df['Work_Location'] == 'Remote']
fig = px.histogram(remote_df, x='Age', nbins=10, title='Age Distribution of Remote Workers')
fig.show()
```

Mental Health Condition by Work Location

```
fig=px.histogram(df, x='Work_Location', color='Mental_Health_Condition',
                 barmode='stack', title='Mental Health Condition by Work Location')
fig.show()
```

Mental Health Breakdown by Company Support Levels

```
fig = px.treemap(df, path=['Company_Support_for_Remote_Work', 'Mental_Health_Condition']
                title='Mental Health Breakdown by Company Support Levels')
fig.show()
```

Stress Level by Job Role

```
fig = px.histogram(df, x="Job_Role",color="Stress_Level",barmode="group",
    title="Stress Level by Job Role")
fig.show()
```

Remote Work Satisfaction Level by Industry

```
fig = px.sunburst(
    df,
    path=['Industry', 'Satisfaction_with_Remote_Work'],
    title='Satisfaction Breakdown Within Each Industry',
    width=600,height=600)
fig.show()
```

Stress Level by Work-Life Balane Rating

```
plt.figure(figsize=(10,6))
sns.histplot(data=df,x='Work_Life_Balance_Rating', hue='Stress_Level',
             multiple='stack', palette='coolwarm', shrink=0.9)
plt.title('Stress Level by Work_Life_Balabce Rating')
plt.xlabel('Work_Life Balance Rating (1=Poor, 5=Excellent)')
plt.ylabel('Count')
plt.grid(True,linestyle='--',alpha=0.4)
plt.show()
```

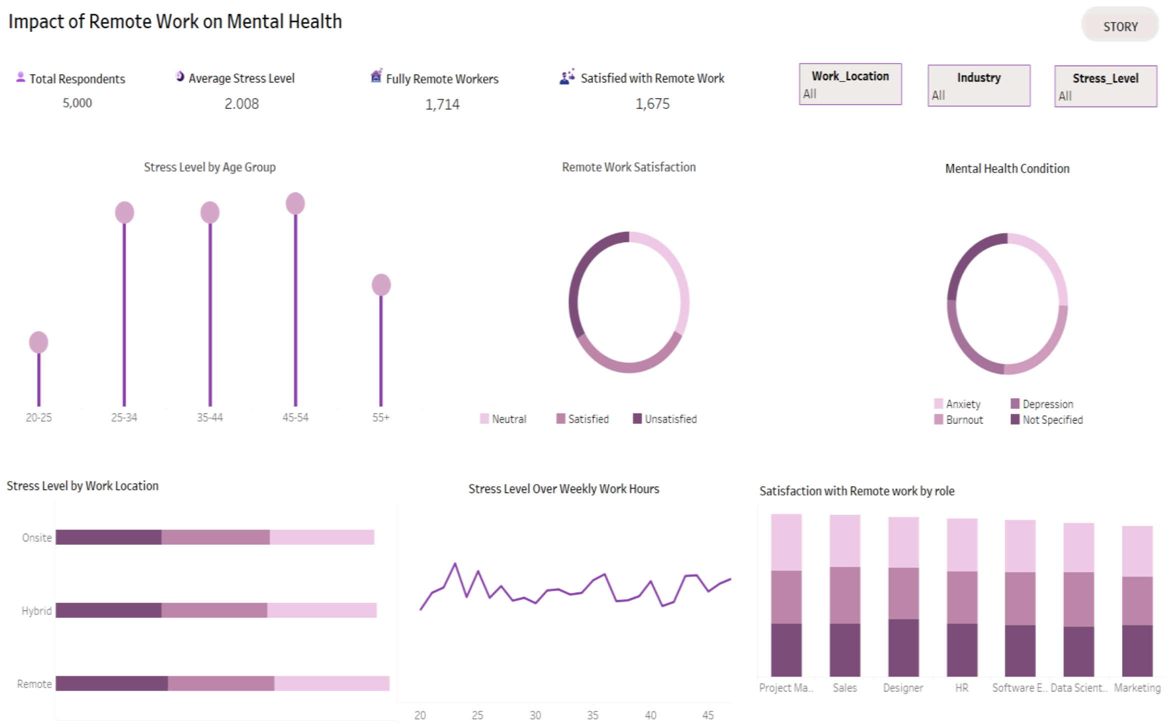
```
df.to_csv("Cleaned_Mental_Health_Remote_Work.csv", index=False)
```

R Programming

```r
install.packages("dplyr")
library("dplyr")
df<-read.csv("C:\\Users\\HP\\Desktop\\shamleena\\Cleaned_Mental_Health_Remote_Work.csv")
View(df)
head(df)
tail(df)
summary(df)

## T-test:Social Isolation Rating by Stress Level
df_filtered <- df %>%
  filter(Stress_Level %in% c("High","Low"))

table(df_filtered$Stress_Level)

# Run the t-test
result <- t.test(Social_Isolation_Rating ~ Stress_Level, data = df_filtered)
print(result)
if (result$p.value < 0.05) {
  print("The difference is statistically significant (p < 0.05)\n")
} else {
  print("The difference is NOT statistically significant (p = 0.05)\n")
}

# Chi Square Test:Job role and Mental health Condition
table_data <- table(df$Job_Role, df$Mental_Health_Condition)

# Perform the Chi-Square Test
chisq_result <- chisq.test(table_data)
print(chisq_result)

# View result
print(chisq_result)
if (chisq_result$p.value < 0.05) {
  print("The difference is statistically significant (p < 0.05)\n")
} else {
  print("The difference is NOT statistically significant (p = 0.05)\n")
}
```

```r
# ANOVA Test:Comparing Years of Experience across Access to Mental Health Resources

anova_result <- aov(Years_of_Experience ~ Access_to_Mental_Health_Resources, data = df)
summary(anova_result)
pval <- summary(anova_result)[[1]][["Pr(>F)"]][1]
if (pval < 0.05) {
  print("The difference is statistically significant (p < 0.05)\n")
} else {
  print("The difference is NOT statistically significant (p = 0.05)\n")
}

# Z-test:Remote Work Status & Satisfaction

df$Is_Remote <- ifelse(df$Work_Location == "Remote", "Yes", "No")
df$Remote_Satisfied <- ifelse(df$Satisfaction_with_Remote_Work == "Satisfied", "Yes", "No")

tab <- table(df$Is_Remote, df$Remote_Satisfied)

# Run Z-test
z_result <- prop.test(tab[, "Yes"], rowSums(tab), correct = FALSE)
print(z_result)
if (z_result$p.value < 0.05) {
  cat("Statistically significant difference in satisfaction based on work location (p < 0.05)
} else {
  cat("No significant difference (p = 0.05)\n")
}

# binary support column
df$Support_Binary <- ifelse(df$Company_Support_for_Remote_Work > 3, "Yes", "No")

# run the F-test
f_result=var.test(Company_Support_for_Remote_Work ~ Support_Binary, data = df)
print(f_result)
if (f_result$p.value < 0.05) {
  cat("Statistically significant difference in satisfaction based on work location (p < 0.05)
} else {
  cat("No significant difference (p = 0.05)\n")
}
```

Tableau


Impact of Remote Work on Mental Health
STORY
Total Respondents
5,000
Average Stress Level
2.008
Fully Remote Workers
1,714
Satisfied with Remote Work
1,675
Work_Location
All
Industry
All
Stress_Level
All
Stress Level by Age Group
20-25
25-34
35-44
45-54
55+
Remote Work Satisfaction
Neutral
Satisfied
Unsatisfied
Mental Health Condition
Anxiety
Depression
Burnout
Not Specified
Stress Level by Work Location
Onsite
Hybrid
Remote
Stress Level Over Weekly Work Hours
20
25
30
35
40
45
Satisfaction with Remote work by role
Project Ma..
Sales
Designer
HR
Software E..
Data Scient..
Marketing