1. Data Import & Initial Inspection

The dataset was loaded successfully and examined for structure and completeness.

No null values were found in the dataset.

Basic statistical information was reviewed to understand variable distribution.

2. Data Cleaning & Preprocessing
a) Outlier Detection and Removal

Outliers were identified in columns such as:

PhysicalActivity

SleepHours

Workload

Stress

IQR (Interquartile Range) method was used to clean the dataset:

Lower Limit = Q1 − 1.5 × IQR

Upper Limit = Q3 + 1.5 × IQR

b) Fixing Spelling & Label Issues

Categorical columns were checked for:

Duplicate labels

Spelling mistakes

Inconsistent categories

Columns reviewed:

Gender

Marital Status

Job Level

Employee Type

Commute Mode

Education Level

OT Status

Gender column initially contained mislabeled categories but was cleaned and standardized.

3. Exploratory Data Analysis (EDA)
Correlation Findings

Workload and Stress show a strong positive relationship.

Workload and Job Satisfaction show a negative correlation.

Experience correlates positively with the number of previous companies.

Univariate Analysis

Distribution of key variables analyzed:

Job Satisfaction

Work Environment Rating

Sleep Hours

Workload

Stress Levels

Bivariate Analysis

Work Environment vs Job Satisfaction:
Better work environment → Higher satisfaction.

Commute Distance vs Sleep Hours:
Longer commute → Reduction in sleep hours.

Physical Activity vs Stress:
Higher physical activity → Lower stress levels.

4. Department & Job-Level Insights
Pivot Table Analysis

Multiple pivot tables were generated to compare metrics across departments, job levels, gender, and marital status:

Average Stress by Job Level & Department

Average Experience comparison

Workload Variations

Work–Life Balance (WLB) patterns

Overtime Percentage (OT %) distribution

Key Observations

Senior job levels generally show lower stress and better WLB.

Overtime percentage varies significantly by department.

Gender-wise comparisons highlighted differences in:

Workload

Stress

Work–Life Balance

5. Visualization & Insights

I have generated multiple visualizations, including:

Boxplots for outlier analysis

Count plots for categorical distribution

Line plots, bar plots, and lmplots for relationship analysis

Detailed department-level comparisons with heatmaps and pivot charts

These visualizations highlighted the employees' mental health, work patterns, and satisfaction metrics.

Conclusion

The analysis provides a strong understanding of:

Employee workload patterns

Stress contributors

Satisfaction factors

Department and job-level differences

Gender and marital-status-based disparities

This final cleaned dataset is now ready for:

Dashboard creation (Power BI / Tableau)

Predictive modeling

HR insights reporting