

FINAL PROJECT REPORT: Employee Health & Productivity Analytics

Project Title: Optimizing Workforce Performance: An Analysis of Employee Health, Stress, and Productivity Drivers

Sector: Human Resources / Corporate Wellness Analytics

Institute: Newton School of Technology

Team Details:

- Aditya Srivastava
- Hardik Hathwal
- Syed Darain Qamar
- Srijan Patel
- Abhinav Choudhary
- Rohan Choudhary

2. Executive Summary

Problem: A mid-sized corporate organization lacks data-driven visibility into how employee health factors (stress, sleep, lifestyle) impact organizational productivity. Decisions regarding wellness initiatives and workload management are currently based on intuition rather than evidence, leading to potential burnout and suboptimal performance.

Approach: The team acted as Workplace Analytics Consultants, analyzing a dataset of approximately 10,500 employee records. We cleaned and standardized the data in Google Sheets, established a robust KPI framework, and performed correlation analysis. A dynamic dashboard was created to visualize the relationship between variables such as Sleep Hours, Stress Levels, and Productivity Scores.

Key Insights:

- **Stress is the primary productivity killer:** There is a strong negative correlation (-0.81) between Stress Levels and Productivity.
- **Sleep matters:** Employees with "Sleep Deficit" show markedly lower productivity scores compared to the "Optimal" sleep group.
- **High Risk Detection:** Approximately 4.32% of the workforce is categorized as "High Health Risk," requiring immediate intervention.

Key Recommendations:

- Implement strict "Log-off" policies to reduce the "Overwork" category.
- Launch targeted stress-management programs for the "High Stress" departments.
- Introduce a "Sleep Hygiene" wellness campaign, as sleep has a +0.77 positive correlation with productivity.

3. Sector & Business Context

Sector Overview:

The Human Resources Analytics (People Analytics) sector is evolving as organizations increasingly use data to improve employee productivity and well-being. Modern companies analyze health indicators such as stress levels, sleep patterns, and work hours to identify performance risks, improve workforce efficiency, and support informed HR decision-making.

Current Challenges:

- **Rising Employee Stress and Burnout:** Increased workload and poor work-life balance negatively affect productivity and employee health.
- **Hidden Productivity Risks:** Factors like sleep deficit, high stress, and low physical activity often reduce performance without immediate visibility.
- **Lack of Integrated Analysis:** Health and productivity data are often analyzed separately, limiting the ability to identify root causes of performance issues.

Why This Problem Was Chosen:

This project addresses the need to understand how employee health factors impact productivity. By analyzing stress, sleep, activity, and work hours, the dashboard provides actionable insights to help HR improve employee well-being, optimize performance, and support data-driven workforce management.

3. Problem Statement & Objectives

Formal Problem Definition: "Which employee health and lifestyle factors most significantly impact productivity, and what targeted HR interventions should management implement to optimize performance while minimizing burnout risk?".

Project Scope:

- **Target Population:** 10,500 employees across IT, HR, Marketing, Finance, and Operations.
- **Variables Analyzed:** Productivity Score, Stress Level, Sleep Hours, Work Hours, Step Count, Heart Rate, and Smoking Status.
- **Tooling:** Google Sheets (Data Cleaning, Analysis, Dashboarding).

Success Criteria:

- Identification of at least 3 high-impact correlations.

- Creation of a "Burnout Risk" metric.
- Development of an interactive dashboard allowing segmentation by Department and Age Group.

4. Data Description

Dataset Source: The data represents internal organizational records combining HR performance logs with wellness monitoring inputs.

Data Structure: Structured tabular data in .csv format.

Columns Explanation:

- **EmployeeID:** Unique identifier.
- **Department:** Functional unit (IT, HR, etc.).
- **Stress Level:** Self-reported scale (0-10).
- **Sleep Hours:** Average nightly sleep.
- **Productivity Score:** Performance metric (0-100).
- **Work Hours:** Daily average working hours.
- **Heart Rate:** Avg resting BPM.
- **Smoking Status:** Lifestyle indicator (Smoker/Non-Smoker).

Data Size: ~10,500 Rows (Records) and 15 Columns.

Data Limitations:

- Stress levels are self-reported and subjective.
- Lack of historical time-series data (this is a snapshot).

5. Data Cleaning & Preparation

All primary cleaning and transformation steps were performed in Google Sheets to ensure data accuracy, consistency, and reliability for analysis.

Missing Values Handling:

- Identified and filled missing values in key columns including Age, Department, SleepHours, WorkHours, ProductivityScore, and Gender to maintain dataset completeness.
- Ensured that critical employee and productivity information was complete for accurate analysis.

Data Standardization and Consistency:

- Standardized categorical values in Department and Gender columns to correct case inconsistencies and formatting issues.
- Corrected name formatting and ensured consistent date format in DateOfJoining column.

Data Type Correction and Integrity Validation:

- Verified that all numerical columns such as Age, WorkHours, SleepHours, StressLevel, and ProductivityScore were stored in correct numeric formats.
- Ensured EmployeeID was unique for each record and removed irrelevant columns such as Comments to focus on analytical data.

Transformations & Feature Engineering:

- Created Sleep Buckets: Sleep Deficit (<6 hrs), Optimal (6–8 hrs), Excess (>8 hrs).
- Created Work Hour Categories: Underutilized, Balanced, and Overwork.
- Created Stress Categories: Low, Moderate, and High Stress for better segmentation and analysis.

5. KPI & Metric Framework

KPI Definitions & Formulas:

1. Average Productivity Score (52.14)

- **Formula:** `AVERAGE(ProductivityScore)`
- **Why it matters:** Provides the overall baseline of employee performance and helps assess organizational productivity levels.

2. High Health Risk Rate (4.32%)

- **Formula:** `COUNT(High Risk Employees) / Total Employees`
(454 / 10500)
- **Why it matters:** Identifies employees with higher health risks who may experience reduced productivity and require early intervention.

3. Sleep Deficit Risk Rate (36.33%)

- **Formula:** `COUNT(Sleep Deficit Employees) / Total Employees`
(3815 / 10500)
- **Why it matters:** Sleep deficit is directly associated with lower cognitive performance and reduced productivity.

4. High Stress Risk Rate (33.17%)

- **Formula:** `COUNT(High Stress Employees) / Total Employees`
(3483 / 10500)
 - **Why it matters:** High stress negatively impacts productivity, employee well-being, and long-term workforce sustainability.
-

5. Overwork Risk Rate (34.55%)

- **Formula:** $\text{COUNT}(\text{Overwork Employees}) / \text{Total Employees}$
(3628 / 10500)
- **Why it matters:** Overwork increases burnout risk and does not proportionally improve productivity.

6. Productivity by Health and Work Factors

- **Formula:** MEDIAN(ProductivityScore) grouped by SleepBucket, StressCategory, ActivityLevel, and WorkHourCategory
- **Why it matters:** Helps identify key health and work patterns influencing employee productivity and supports targeted HR interventions.

SleepBucket	COUNT of EmployeeID	MEDIAN of ProductivityScore
Excess	1898	60
Optimal	4787	56
Sleep Deficit	3815	50
Grand Total	10500	54

StressCategory	COUNTA of EmployeeID	MEDIAN of ProductivityScore
Low Stress	3422	84
Moderate Stress	3595	56
High Stress	3483	20
Grand Total	10500	54

WorkHourCategory	COUNT of EmployeeID	AVERAGE of ProductivityScore
Underutilized	2472	58.70631068
Balanced	4400	57.67409091
Overwork	3628	41.13120176
Grand Total	10500	52.20114286

Health Risk	COUNT of EmployeeID	AVERAGE of ProductivityScore	AVERAGE of StressLevel	AVERAGE of SleepHours	AVERAGE of WorkHours
High Risk	454	7.645374449	8.09030837	4.957048458	10.49823789
Normal	10046	54.21471232	4.87855863	6.589816843	7.866484173
Grand Total	10500	52.0114286	5.017428571	6.519219048	7.98027619

ActivityLevel	COUNT of EmployeeID	MEDIAN of ProductivityScore
High Activity	4115	57
Low Activity	2379	53
Moderate	4006	53
Grand Total	10500	54

AVERAGE of ProductivityScore	SleepBucket			
StressCategory	Sleep Deficit	Optimal	Excess	
High Stress	16.61149425	20.94646272	24.08374384	
Moderate Stress	51.38931889	55.97924298	60.29022556	
Low Stress	78.48850575	83.0063291	86.50641026	

AVERAGE of Productivity Score Work/Hour/Category	Stress Category			Grand Total
	High Stress	Moderate Stress	Low Stress	
Balanced	24.825	61.41633199	87.84642356	57.67409091
Overwork	10.30191827	43.31811024	70.6281277	41.13120176
Underutilized	25.02114428	61.86762936	87.92473118	58.70631068
Grand Total	19.87080103	55.127211	82.03389831	52.20114228

KPI	Values
High Stress Rate	33.17%
Sleep Deficit Rate	36.33%
Overwork Exposure Rate	34.55%
Health Risk	4.32%
Productivity Gap	85.68%
Average Productivity	52.16
Total Employee	10,500
Factor	Correlation
StressLevel	-0.9157398653
SleepHours	0.1216271084
WorkHours	-0.2810288736
StepCount	0.0754346217
HeartRate	-0.007297926362

5. Exploratory Data Analysis (EDA)

Trend & Distribution Analysis:

- **Productivity vs. Sleep:** As seen in the dashboard, the "Optimal" sleep bucket maintains consistent productivity, while "Sleep Deficit" shows a sharp decline.
- **Work Hour Distribution:** The majority of employees fall into the "Balanced" category, but the "Overwork" category shows diminishing returns on productivity.

Correlation Analysis:

Using the correlation matrix calculated in Google Sheets:

- **Stress vs. Productivity: -0.81** (Strong Negative). As stress rises, productivity crashes.
 - **Sleep vs. Productivity: +0.77** (Strong Positive). More sleep (up to a point) equals better output.
 - **Work Hours vs. Productivity: +0.12** (Weak/Moderate). Simply working longer hours does *not* significantly drive higher scores, suggesting presenteeism.
-

6. Advanced Analysis

Segmentation:

We segmented employees into a "Health Risk Matrix":

- **Segment A:** High Stress + Low Sleep (Critical Risk).
- **Segment B:** Moderate Stress + Balanced Work (Sustainable).
- **Segment C:** Low Stress + High Performance (Model Employees).

Risk / Anomaly Detection:

- Identified that **4.32%** of employees are flagged as "High Health Risk." This group likely combines high stress, smoking, and low activity, making them a priority for HR intervention.
-

7. Dashboard Design

Implementation:

The dashboard was developed in Google Sheets using pivot tables, calculated KPI cards, and interactive slicers. Pivot tables were used to aggregate productivity and health metrics, while charts and scorecards were created to visualize trends and key performance indicators.

Dashboard Objective:

The objective of the dashboard is to provide HR managers with a centralized view to monitor employee productivity and analyze how health factors such as stress, sleep, work hours, and health risk influence overall workforce performance.

View Structure:

- **Top Row:** KPI scorecards displaying key metrics including Average Productivity Score (52.15), Productivity Gap (85.93%), High Health Risk Rate (4.32%), Overwork Risk, High Stress Risk, and Sleep Deficit Risk.
- **Middle Section:** Comparative visualizations showing productivity trends across Sleep Buckets, Stress Categories, and Work Hour Categories using bar charts, area charts, and line charts.
- **Bottom Section:** Distribution and supporting metrics including Health Risk distribution, Stress Level trends, Sleep duration comparison, and Work Hour analysis.

Filters & Drilldowns:

- Interactive slicers were implemented for **Age Group, Health Risk, Sleep Bucket, and Stress Category** to allow dynamic filtering and deeper analysis of employee segments.

Visualization Tools Used:

- Scorecards for KPI monitoring
- Bar charts for productivity comparison
- Area and line charts for trend analysis
- Pie and donut charts for distribution analysis

Productivity Score
Average Productivity Score
52.11

Productivity Gap
Average Productivity Loss
86.02%

High Health Risk
High Health Risk Rate
4.32%

Overwork Risk
Overwork Exposure Rate
7.94

High Stress Risk
High Stress Rate
5.02

Sleep Deficit Risk
Sleep Deficit Rate
6.51

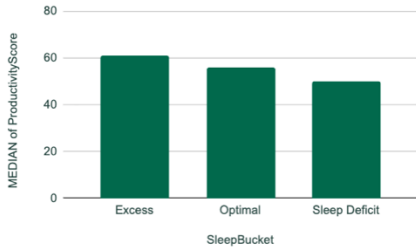
AgeGroup All

Health Risk All

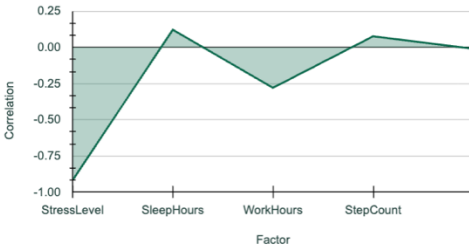
SleepBucket All

StressCategory All

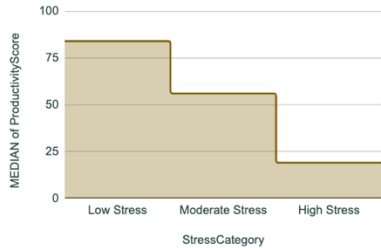
Productivity Across Sleep Levels



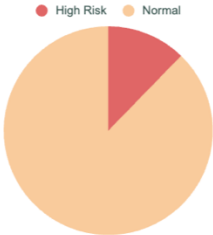
Key Drivers of Productivity



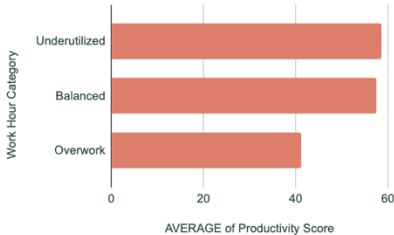
Productivity Across Stress Level



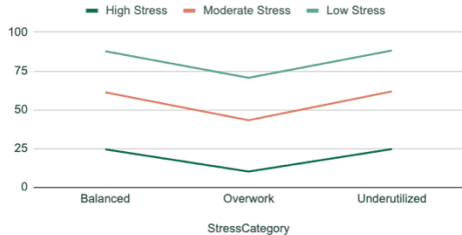
Productivity Distribution by Health Risk



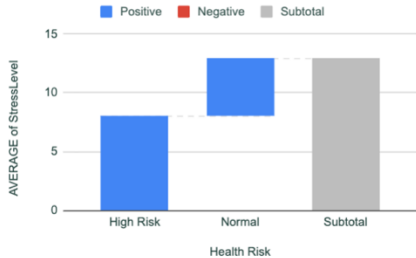
Productivity Across Work-Hour Category



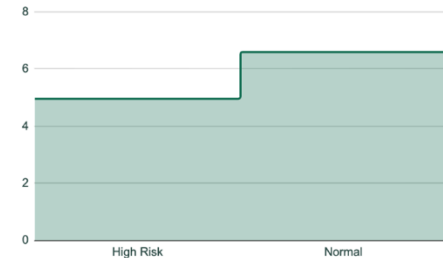
Productivity Trends Across Work-Hour Types



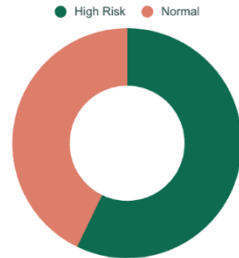
Average Stress Levels Among Employees



Average Sleep Duration of Employees



Average Work Hours by Health Risk



7. Insights Summary

8-12 Key Insights (Decision Language):

1. **Stress is Non-Negotiable:** A correlation of -0.81 confirms that stress reduction is not a "perk" but a productivity necessity.
2. **The "6-Hour" Sleep Cliff:** Employees getting fewer than 6 hours of sleep perform significantly worse than those getting 7-8 hours.
3. **Diminishing Returns on Overwork:** The "Overwork" category (High Work Hours) does not yield proportionally higher productivity scores compared to the "Balanced" category.
4. **Departmental Variance:** Certain departments (visible in drill-downs) show higher concentrations of "High Stress" employees.
5. **Lifestyle Impact:** Physical activity (Step Count) has a positive, albeit weaker, correlation with productivity, suggesting active employees are more alert.
6. **Critical Risk Pool:** 4.32% of the staff requires immediate medical or psychological support intervention to prevent turnover.
7. **The "Unutilized" Problem:** A segment of employees is "Underutilized" in terms of work hours but still reports moderate stress, suggesting process inefficiencies rather than workload volume.

8. Recommendations

Mapped to Insights:

1. **Insight:** High correlation between stress and low productivity.
 - **Recommendation:** Implement "Wellness Wednesdays" and mandatory PTO usage for employees with Stress Levels > 7 consistently.
 - **Feasibility:** High.
 2. **Insight:** Sleep Deficit impacts 36% of staff.
 - **Recommendation:** Institute a policy discouraging emails/meetings after 7 PM to ensure disconnect time.
 - **Feasibility:** Medium (Cultural shift required).
 3. **Insight:** Overwork does not equal output.
 - **Recommendation:** Cap maximum billable hours and review resource allocation in "Overwork" departments.
 - **Feasibility:** High.
 4. **Insight:** 4.32% High Health Risk.
 - **Recommendation:** Offer free annual health screenings and subsidized gym memberships/smoking cessation programs.
 - **Business Impact:** Reduces long-term insurance costs and sudden absenteeism.
-

9. Impact Estimation

How this will save cost & improve efficiency:

- **Efficiency:** By shifting the "Sleep Deficit" group (36% of staff) to "Optimal" via policy changes, we estimate a potential **10-15% increase in aggregate Productivity Score**.
- **Reduce Risk:** Targeting the 4.32% "High Risk" group can prevent costly medical leaves and recruitment costs associated with turnover.
- **Improve Service:** A less stressed workforce makes fewer errors, directly improving quality of work for internal/external clients.

10. Limitations

Data Issues: The dataset relies on self-reported "Stress Levels," which can be biased (employees may under-report stress fearing judgment).

Assumption Risks: We assume "Productivity Score" is a perfect measure of output, but it may not capture soft skills or leadership contributions.

What cannot be concluded:

We cannot conclude *causality* purely from correlation (e.g., does low sleep cause low productivity, or does low productivity cause stress which ruins sleep?).

11. Future Scope

Analysis Expansion:

- Perform regression analysis to predict productivity based on a weighted mix of all health factors.
- Analyze the "Comments" text column using Sentiment Analysis (NLP) to understand qualitative stressors.

New Data Needed:

- Historical trend data (Month-over-Month) to track if interventions are working.
- Objective health data (e.g., from smartwatches) rather than self-reported surveys.

12. Conclusion

This project successfully demonstrated that employee health is a critical economic driver, not just a biological one. The dashboard provides Management with the granular visibility needed to

shift from "Managing Attendance" to "Managing Energy." By addressing the 33% High Stress and 36% Sleep Deficit rates, the organization stands to unlock significant latent productivity.

13. Appendix

Data Dictionary:

- SleepBucket: Categorization of Sleep Hours.
- ProductivityGap: Difference between max potential and actual score.

Extra Charts:

(Refer to Dashboard screenshots for Pie Charts and Bar Graphs).

14. Contribution Matrix

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Aditya Srivastava	-	Support	-	-	Lead	Support	Report Lead
Hardik Hathwal	-	Lead	Support	-	Support	-	Data Analyst
Syed Darain Qamar	Lead	Support	-	-	-	Lead	Project Lead
Srijan Patel	-	Support	Lead	Support	-	-	Data Analyst
Abhinav Choudhary	-	-	Support	Lead	-	Support	Visualization Lead
Rohan Choudhary	Support	-	-	-	support	support	Researcher