Bright Publisher

Investigating Employee Burnout Determinants through Exploratory Data and Correlation Analysis

Abdul Aziz[1,*], Axel Sandi Seno[2]

[1,2]Dept. of Information System, Amikom Purwokerto University, Indonesia

ABSTRACT

This research explores the determinants of employee burnout through Exploratory Data Analysis (EDA) and correlation analysis using the Synthetic HR Burnout Dataset from Kaggle. The study aims to quantify how psychological, workload, and demographic variables collectively influence burnout risk in simulated agile-team environments. A total of 2,000 employee records comprising ten features—including Age, Gender, JobRole, Experience, WorkHoursPerWeek, RemoteRatio, SatisfactionLevel, StressLevel, and Burnout—were analyzed using Python 3.11 with pandas, matplotlib, seaborn, and scipy. The workflow involved data cleaning, descriptive statistics, univariate visualization, and pairwise correlation computation (Pearson and point-biserial coefficients). Results show that burnout prevalence is 6.45 percent. The highest positive correlation with burnout occurs for StressLevel ($r = 0.321$, $p < 0.001$), followed by WorkHoursPerWeek ($r = 0.226$, $p < 0.001$), while SatisfactionLevel displays a moderate negative association ($r = –0.233$, $p < 0.001$). Demographic factors such as age, experience, and gender present statistically insignificant effects, suggesting that burnout is fundamentally behavioral and situational. Visualization outputs—including histograms, boxplots, and correlation heatmaps—reinforce these findings, revealing clusters of high stress and low satisfaction among burnout cases. From an Agile-management perspective, the study highlights that sustainable team performance depends on balancing workload and psychological well-being. The proposed analytical pipeline is fully reproducible on standard hardware, demonstrating that meaningful HR analytics can be conducted without large computational infrastructure. This approach provides a practical foundation for data-driven monitoring of team health, enabling organizations to translate the Agile value of "sustainable pace" into measurable, actionable metrics. Future research may extend this framework with predictive or time-series models using real HR or agile sprint data to explore burnout dynamics over time.

Keywords Burnout Prediction, Exploratory Data Analysis, Correlation Analysis, Agile Management, Employee Well-Being

Submitted: 2 June 2024
Accepted: 15 August 2024
Published: 5 February 2025

Corresponding author
Abdul Aziz,
abdazis9@amikompurwokerto.ac.id

Additional Information and Declarations can be found on page 28

Introduction

Employee well-being and workplace mental-health management are now central to organizational performance, safety, and talent retention. Studies across sectors emphasize that workplace mental health is multidimensional—encompassing burnout, stress, depression, anxiety, and resilience—and that attention to these constructs materially affects both individual outcomes and enterprise productivity [1], [2]. Systematic reviews of workplace-health interventions further conclude that the organizational environment—workload, managerial support, and recognition practices—determines whether employees thrive or deteriorate [3], [4]. Consequently, managing employee well-being has evolved from a clinical concern to a strategic imperative for sustainable organizations.

Burnout is defined as a multidimensional psychological syndrome consisting of

How to cite this article: A. Aziz and A. S. Seno, "Assessing the Efficacy of Convolutional Neural Networks in Recognizing Handwritten Digits Analysis of the Relationship Between Trading Volume and Bitcoin Price Movements Using Pearson and Spearman Correlation Methods," *Agile Manag*, vol. 1, no. 1, pp. 17-30, 2025.

emotional exhaustion, depersonalization (or cynicism), and reduced professional efficacy [1], [5]. Prospective studies link these dimensions—particularly exhaustion and cynicism—to functional impairment and disability, showing that burnout represents more than transient fatigue [5]. Although distinct from primary mood disorders, burnout substantially elevates the risk of depression, disengagement, and unsafe behavior if unaddressed [6]. This evidence positions burnout as a measurable occupational hazard that demands both preventive and analytic approaches.

The growth of digital workplaces, hybrid schedules, and agile project cycles has produced simultaneous gains in flexibility and productivity—but also intensified workload, role ambiguity, and constant connectivity. Organizational reviews identify poor recognition, weak management support, and bullying as chronic stressors that accelerate burnout [3], [4]. In this environment, data-driven HR analytics—including continuous measurement, early-warning indicators, and population-level modeling—offers a means to detect emergent risk patterns and guide timely, systemic interventions [3], [7].

Remote and hybrid arrangements, while enabling autonomy, can also blur boundaries and heighten relational strain. Evidence indicates that perceived organizational support and authenticity in workplace relationships buffer these risks, mitigating the harmful effects of excessive workload and low recognition [7], [8]. Thus, modern work design must combine digital efficiency with intentional monitoring of psychosocial dynamics to prevent overextension and loss of engagement.

The Agile Manifesto embeds a commitment to “sustainable pace,” asserting that development teams should maintain a work rhythm they can sustain indefinitely [9]. Agile principles thereby position team health as intrinsic to process integrity: sustained performance depends on avoiding chronic overwork that degrades product quality and morale [10]. Empirical research on agile adoption reiterates that its core values—iterative delivery, close collaboration, and respect for individuals—seek to create environments where teams can adapt continuously without enduring strain [9], [10].

Agile frameworks operationalize these values through iterative feedback and workload transparency. Scrum prescribes fixed-length sprints with regular inspect-and-adapt ceremonies; Kanban limits work in progress to expose bottlenecks; and eXtreme Programming employs pair programming and test-driven development to maintain flow and shared ownership [11], [12]. Together, these mechanisms make delivery speed contingent on continuous reflection and adjustment, linking adaptability with collective responsibility.

A persistent managerial challenge, however, lies in reconciling the drive for rapid delivery with the obligation to preserve team well-being. Continuous-delivery environments can accelerate tempo across development and operations, risking cumulative strain if velocity metrics are pursued without balancing controls [11]. Sustainable agile practice therefore requires explicit monitoring of workload, psychological capacity, and recovery periods—through both process metrics and people-centric indicators—to ensure that iterative success does not erode long-term capability [9], [12].

Quantitative burnout analytics complements qualitative agile retrospectives by supplying objective, early-warning data between sprint reviews [13], [14]. Integrating analytics with agile ceremonies enables managers to calibrate commitments empirically and to correlate throughput patterns with human factors such as overtime or stress sentiment [14]. This synergy transforms agile retrospection from anecdotal reflection into evidence-based learning, strengthening alignment between sustainable delivery and human sustainability.

Despite extensive research on occupational burnout, a replicable, quantitative evidence base linking psychological covariates (stress, satisfaction, workload) to burnout in agile-like contexts remains limited. Existing studies predominantly rely on proprietary surveys or clinical samples [3], [4], [15]. Such data restrictions hinder reproducibility and prevent the benchmarking of analytic workflows across teams and industries. Consequently, there is a methodological gap between theoretical assertions about burnout predictors and the availability of open, analysis-ready datasets.

Most empirical work focuses on healthcare and education sectors using self-report instruments [16], [17], [18]. Although these studies establish causal links between job strain and impairment, they are rarely structured for algorithmic HR analytics or agile performance evaluation [15]. Researchers and practitioners lack standardized, open testbeds to demonstrate interpretable exploratory data-analysis (EDA) and correlation techniques suitable for organizational deployment. Moreover, published examples of lightweight, reproducible HR-analytics pipelines remain scarce. While agile and DevOps literatures advocate continuous measurement, few works provide clear, replicable EDA + correlation frameworks that practitioners can execute without advanced data-science infrastructure [13], [14]. This absence constrains both teaching and operational adoption of evidence-based team-health analytics.

To address these gaps, the present study employs the Synthetic HR Burnout Dataset [19], an openly available, ethically safe dataset simulating employee-level burnout determinants. Synthetic data allows demonstration of reproducible preprocessing, visualization, and correlation steps without compromising real employee privacy. It enables controlled exploration of relationships among workload, stress, satisfaction, and burnout, offering a transparent benchmark for HR-analytics education and agile-team well-being monitoring.

Building on the identified gap, this study pursues three objectives: to perform exploratory and descriptive analysis of workplace variables relevant to burnout; to quantify correlations between burnout and its behavioral and psychological predictors; and to interpret these relationships within an agile-management framework emphasizing sustainable pace.

Literature Review

The Concept of Burnout in Organizational Psychology

The concept of burnout was first articulated by Maslach and Jackson [20] and later refined by Leiter and Maslach [21]. It is defined as a multidimensional occupational syndrome comprising emotional exhaustion, depersonalization (cynicism), and reduced professional efficacy—the core structure formalized in the Maslach Burnout Inventory, which remains the most cited measurement framework in occupational-health research [20], [22]. The World Health Organization (WHO) codified “burn-out” in the ICD-11 as a workplace phenomenon resulting from chronic, unsuccessfully managed job stress [23]. This framing identifies burnout as an occupational condition rather than a medical disorder, underscoring its dependence on workload, role clarity, and managerial support.

Contemporary theoretical models—particularly the Job Demands–Resources

(JD–R) and Demand–Control–Support (DCS) frameworks—explain burnout as the imbalance between job demands and available resources [24], [25]. Excessive workload, time pressure, and emotional labor accelerate exhaustion, whereas autonomy, social support, and leadership quality act as buffers [26]. Meta-analytic research confirms that high demands and low resources jointly predict emotional exhaustion and disengagement [25], [27]. Recent scholarship extends this model to the digital workplace, where information overload and constant connectivity contribute to digital burnout [28], [29]. These findings broaden classical stress theory by integrating cognitive and technological demands arising from remote and hybrid work patterns.

Determinants of Burnout and Employee Well-Being

Across occupational domains, perceived stress, job satisfaction, and working hours consistently predict burnout intensity. Meta-analyses report moderate positive correlations between stress and burnout ($r \approx 0.3$–0.5) [25], [30]. Low satisfaction amplifies the effects of stress by reducing coping resources and motivation [31]. Overwork—defined by prolonged weekly hours and limited recovery—remains one of the most robust antecedents of exhaustion [32]. Conversely, recovery behaviors and psychological detachment are inversely related to burnout, suggesting that sustainable pacing and rest are critical moderators [33].

Demographic variables such as age and gender generally exhibit weak explanatory power compared with contextual and psychological factors [25], [34]. Leadership style, organizational culture, and job autonomy strongly moderate burnout trajectories: transformational leadership and empowering climates mitigate exhaustion, while toxic or authoritarian climates intensify it [35]–[37]. High autonomy and participative decision-making permit self-regulation of workload, thereby lowering strain [38]. Intervention studies further demonstrate that modifications to work design—rather than individual counseling alone—produce the most durable burnout reductions [34]. Collectively, evidence positions organizational redesign as the most effective avenue for improving well-being and sustaining engagement.

HR Analytics and Machine-Learning Approaches to Burnout

Human-Resource Analytics (HRA) has transformed personnel management into an evidence-driven discipline, leveraging computational methods to detect attrition, stress, and burnout [39]. Machine-learning (ML) models such as logistic regression, random forest, gradient boosting, and support-vector classifiers have been applied to predict burnout or absenteeism using indicators like workload, satisfaction, and tenure. Empirical studies in healthcare and education show that these models achieve high predictive accuracy but often lack interpretability, limiting managerial insight. The current literature therefore advocates integrating explainable AI (e.g., SHAP values, decision trees) to ensure transparency in HR decision-support systems.

Before advanced modeling, EDA and correlation analysis remain indispensable for understanding variable relationships and ensuring data quality [39]. EDA uncovers distributional biases, multicollinearity, and missing-value patterns that inform model selection. Correlation heatmaps and pairwise plots visualize associations among workload, satisfaction, stress, and burnout, providing interpretable baselines prior to ML deployment. The reproducibility crisis in HR analytics—driven by proprietary data and privacy constraints—has stimulated the use of open synthetic datasets. Such datasets enable transparent methodological demonstrations while safeguarding anonymity. The Synthetic HR Burnout Dataset. exemplifies this new research paradigm: it allows scholars to test EDA, correlation, and classification pipelines under realistic yet ethically safe conditions, advancing both methodological literacy and organizational learning.

Method

Research Design

This study employs a quantitative, exploratory, and correlational research design to investigate factors influencing employee burnout using synthetic HR data. The approach integrates EDA and correlation analysis, which together enable the identification of statistical patterns and variable associations within a realistic but artificial workforce dataset. The quantitative design ensures objectivity and replicability, while the exploratory nature allows discovery of potential causal hypotheses for future modeling.

The research's methodological focus is not prediction but explanation: how job-related, psychological, and demographic variables interact to indicate burnout risk. This aligns with the analytical foundations of agile management research, which often seeks to understand human factors driving productivity and sustainable pace. The dataset's-controlled nature allows statistical relationships to be explored without ethical constraints of real HR data, making it ideal for methodological demonstration. The outcome of this design is a detailed analytical framework capable of revealing variable distributions, multivariate interactions, and strength of association between work-related factors (e.g., stress, satisfaction, working hours) and burnout occurrence. This framework supports both academic insight and practical application in agile team health analytics.

Data Source and Structure

The dataset used is the Synthetic HR Burnout Dataset, publicly available on Kaggle under a Public Domain License. It contains 2,000 records and 10 columns describing employee characteristics relevant to well-being, such as Age, Gender, JobRole, Experience, WorkHoursPerWeek, RemoteRatio, SatisfactionLevel, StressLevel, and Burnout. Each record simulates an individual employee, with Burnout serving as the binary target variable (1 = burnout symptoms, 0 = none). The data's synthetic generation ensures internal consistency between variables: high StressLevel, low SatisfactionLevel, and long WorkHoursPerWeek probabilistically co-occur with Burnout = 1. This built-in logic allows controlled evaluation of correlations and models the conceptual HR reality without privacy issues. The dataset's balanced gender and diverse job roles (Manager, Engineer, Analyst, HR, Sales) provide sufficient heterogeneity for exploratory insights.

This dataset's structure supports multilevel inquiry: univariate analysis (distribution of each attribute), bivariate analysis (correlation between predictors and burnout), and group comparisons (gender × burnout). The high cleanliness score (10/10 on Kaggle Usability) eliminates preprocessing noise, letting analytical attention focus on interpretation rather than data correction.

Data Cleaning and Preprocessing

The preprocessing phase ensures analytical readiness. First, the non-informative identifier column Name was dropped using df.drop(columns=["Name"], errors="ignore") to avoid bias from pseudo-identifiers. Next, a completeness check (df.isnull().sum()) confirmed zero missing values across all variables. This verification step corresponds to Checkpoint 2 in the code and prevents runtime exceptions in subsequent numeric operations.

Type validation followed, using df.dtypes to ensure that numerical columns (e.g., WorkHoursPerWeek) were treated as continuous variables and not mistakenly parsed as strings. Proper typing is crucial for computing Pearson and point-biserial correlations, which assume numerical inputs. Categorical variables such as Gender and JobRole were retained as object type for cross-tab analysis rather than one-hot encoding, since the study's goal was exploration, not model training.

Basic descriptive statistics were generated via df.describe(include="all"), which produced central-tendency and dispersion metrics (mean, std, min, max, quartiles) for numerical columns and frequency counts for categorical ones. This formed the foundation for verifying the realism of variable ranges—e.g., ages 22–60 years, work hours 30–70 per week, stress 1–10. Any deviation would have prompted re-scaling or outlier handling, though none were detected.

All preprocessing was conducted in-memory using pandas and numpy, achieving execution times under one second. This efficiency demonstrates that small-scale HR analytics can be performed locally without cloud infrastructure, aligning with agile research's emphasis on rapid iteration and lightweight experimentation.

Exploratory Data Analysis (EDA)

The EDA stage aimed to uncover hidden patterns, distributions, and anomalies through both statistical summaries and visual representations. Using matplotlib and seaborn, histograms (sns.histplot()) with kernel density overlays were produced for continuous variables such as Age, Experience, WorkHoursPerWeek, SatisfactionLevel, and StressLevel. Each plot used bins=20 and color="skyblue" parameters to ensure smooth yet readable density curves.

These visualizations confirmed approximately normal distributions for age and satisfaction, and slightly right-skewed distributions for stress and working hours. Boxplots (sns.boxplot()) were then generated to compare stress and satisfaction across job roles (x="JobRole", y="StressLevel", palette="Set2"). This revealed moderate dispersion differences, suggesting variation in stress perception between managerial and technical positions.

Pairwise relationships among StressLevel, SatisfactionLevel, WorkHoursPerWeek, and Burnout were examined through a scatter matrix (sns.pairplot() with hue="Burnout", diag_kind="kde"). This provided early evidence of linear and inverse relationships: clusters of high stress/low satisfaction aligning with burnout = 1. The graphical output supports later quantitative correlation results, visually validating expected behavioral patterns.

EDA outputs also included basic aggregate statistics such as mean stress (5.43), mean satisfaction (≈ 3.0), and mean work hours (≈ 50). These values contextualize the population as moderately stressed and working slightly beyond standard full-time hours, establishing a realistic synthetic baseline for burnout risk analysis.

Correlation Analysis

The correlation phase quantified relationships among numeric variables using Pearson's r via df.corr(numeric_only=True) and separately tested feature–target associations with point-biserial correlation (scipy.stats.pointbiserialr). Pearson's r captures linear association strength from −1 to +1, while the point-biserial coefficient adapts Pearson's formula for one binary and one continuous variable, ideal for linking features to Burnout.

Calculated results identified three significant relationships: StressLevel → Burnout ($r = 0.321$, $p < 0.001$), SatisfactionLevel → Burnout ($r = -0.233$, $p < 0.001$), and WorkHoursPerWeek → Burnout ($r = 0.226$, $p < 0.001$). Significance was determined using two-tailed p-values (< 0.05) automatically returned by pointbiserialr. Weak or non-significant correlations (e.g., Age $r \approx 0.004$) were considered noise.

A correlation heatmap (sns.heatmap(corr_matrix, annot=True, cmap= "coolwarm", fmt=".2f") visualized pairwise r-values, with warm colors denoting positive and cool colors denoting negative relationships. This visual step aids interpretability for management audiences who may not routinely interpret numeric matrices. Bar plots of burnout correlations (sns.barplot()) ranked variables by magnitude, clarifying relative influence.

This phase's technical rigor stems from explicit parameter control—ensuring reproducibility and statistical validity. All computations operate under default Pearson assumptions (linear relation, interval data, no severe outliers), which hold for synthetic numeric variables. Results collectively reveal moderate yet consistent associations, validating the dataset's logical structure and supporting hypotheses about workload–stress–satisfaction interplay.

Result and Discussion

Data Overview and General Characteristics

The dataset analysis confirmed high structural quality and consistency. All 2,000 records were successfully imported without missing values or data-type anomalies. The numerical and categorical fields exhibited valid and meaningful ranges consistent with real-world HR conditions. Employees' ages ranged from 22 to 60 years, with work experience spanning 0–39 years. The average age of 40.7 years and mean experience of 10.1 years indicate a mature and moderately experienced workforce typical of corporate or agile project environments.

The average weekly workload of 49.6 hours notably exceeds the standard 40-hour benchmark, hinting at organizational cultures prone to extended working hours—an established risk factor for burnout. The mean stress level (5.43) and mean satisfaction score (≈3.0) suggest moderate stress and mid-range job fulfillment, producing a balanced but slightly strained workforce profile. These averages form the baseline for subsequent correlational exploration.

Categorically, the workforce was gender-balanced (≈51% male, 49% female) and distributed across five job roles: Manager, Analyst, Engineer, HR, and Sales. Such heterogeneity ensures that statistical relationships reflect cross-role dynamics rather than being dominated by a single occupational cluster. Burnout prevalence stood at 6.45%, matching the dataset creator's stated synthetic logic (i.e., ~1 in 16 employees show burnout symptoms).

From an Agile perspective, these base characteristics model a realistic organizational ecosystem: teams composed of diverse roles, maintaining moderate satisfaction, and exhibiting varying levels of stress and workload intensity. The balance between demographic factors and psychological attributes allows insights into how human sustainability interacts with operational pace—key to Agile's "sustainable development" principle (Agile Manifesto, 2001).

Exploratory Data Analysis (EDA) Findings

Exploratory visualizations revealed key behavioral tendencies and subgroup variations. The histograms for Age and Experience showed near-normal distributions, confirming demographic diversity across early- and mid-career professionals. WorkHoursPerWeek displayed slight right skewness, reflecting a tail of employees consistently exceeding 60 hours per week. This skew aligns with typical patterns in fast-paced agile environments or technology-driven teams where sprint deadlines may intensify workload temporarily.

The StressLevel distribution (mean 5.4, $\sigma = 2.9$) demonstrated considerable variance, suggesting heterogeneous coping capacities and environmental pressures across the workforce. Conversely, SatisfactionLevel (mean ≈ 3.0, $\sigma = 1.15$) presented a centered but slightly bimodal distribution—indicating subgroups of content and discontent employees coexisting. This divergence may represent differing team climates or managerial effectiveness.

Boxplots comparing StressLevel and SatisfactionLevel across JobRole categories revealed that managers and engineers tend to report higher stress medians, while HR and analysts exhibit wider satisfaction variability. Such dispersion implies potential mismatches between job expectations and perceived value—an early warning sign of burnout triggers. This mirrors prior HR studies emphasizing managerial overload and emotional labor as key burnout sources (Maslach & Leiter, 2016).

Pairwise scatter matrices confirmed visible clusters where burnout cases concentrate around high stress (≥8) and low satisfaction (≤2) zones. Meanwhile, non-burnout cases populate moderate-stress, moderate-satisfaction regions. These visual patterns prefigure the correlation analysis results, offering intuitive confirmation of the underlying relationships between workload, emotion, and exhaustion.

Correlation Analysis Results

The correlation matrix provides quantitative grounding for observed behavioral tendencies. StressLevel ($r = 0.321$, $p < 0.001$) shows the strongest positive association with burnout, while SatisfactionLevel ($r = -0.233$, $p < 0.001$) exhibits a moderate negative correlation. This inverse relationship underscores the complementary nature of stress and satisfaction in defining well-being: as stress intensifies, satisfaction typically declines, culminating in higher burnout risk.

The WorkHoursPerWeek → Burnout correlation ($r = 0.226$, $p < 0.001$) further validates the notion that extended working hours contribute meaningfully, though not exclusively, to burnout occurrence. The modest strength of this relationship implies that work hours alone are insufficient predictors—psychological dimensions amplify or mitigate the effect. For instance, employees clocking long hours but reporting high satisfaction may resist burnout, aligning with findings in motivational and job-engagement literature.

Correlations involving Age, Experience, and RemoteRatio were negligible ($|r| < 0.03$, $p > 0.05$), signifying that burnout transcends demographic and logistical factors. Neither tenure nor remote-working proportion demonstrated measurable protective or aggravating influence. This neutrality is informative: it suggests that burnout is primarily situational and psychological, not structural or demographic. The absence of age-related effects also challenges stereotypes that younger or older workers are inherently more vulnerable.

The inter-variable correlations show logical coherence: stress and satisfaction have a mild inverse correlation ($r = –0.035$), consistent with expected psychological antagonism. Meanwhile, workload (hours/week) correlates weakly but positively with stress ($r = 0.05$) and negatively with satisfaction ($r = –0.02$), indicating that overwork exerts subtle pressure even before clinical burnout emerges. Collectively, these findings validate the dataset's realism and affirm the analytical method's reliability.

Gender and Role-Based Insights

The gender–burnout cross-tabulation yielded nearly identical burnout rates: Female 6.24%, Male 6.65%, reflecting a non-discriminatory workload distribution. This uniformity supports the synthetic dataset's fairness but also mirrors contemporary workplace research suggesting gender differences in burnout are increasingly contextual rather than biological.

Role-based exploratory visuals highlighted subtle variations: managers and engineers appear at higher stress quartiles, whereas HR and analysts exhibit broader satisfaction ranges. These results suggest role complexity—rather than gender—drives emotional strain. Managers face leadership pressure, engineers handle cognitive load, and HR roles entail interpersonal demands, each contributing differently to fatigue dynamics.

Such findings are relevant to agile teams, which emphasize cross-functionality and shared responsibility. The lack of extreme burnout disparity across roles may reflect agile's inherent buffer mechanism—task distribution and self-organizing teams reduce isolated overloads. However, sustained imbalance in psychological states across roles still poses risk to team cohesion and performance stability.

In organizational practice, these observations advocate for role-specific intervention strategies rather than one-size-fits-all wellness programs. Agile coaches and Scrum Masters could use similar correlation dashboards to monitor team well-being metrics, tailoring stress management or workload redistribution per role cluster.

Visual Correlation Interpretation

The correlation heatmap revealed three distinct variable clusters: Stress–

Satisfaction–Burnout cluster, clear inverse and positive interactions highlighting emotional determinants; Workload–RemoteRatio cluster, weak internal relationships reflecting operational attributes; and Age–Experience cluster, expected positive internal correlation ($r = 0.64$), confirming demographic realism but low relevance to burnout.

Bar charts of feature–burnout correlations visually prioritized StressLevel, SatisfactionLevel, and WorkHoursPerWeek as top three influencers. These visual cues simplify managerial communication—allowing non-technical decision-makers to grasp key drivers at a glance. The moderate strength of all correlations (0.2–0.3 range) indicates a multifactorial burnout model rather than a single dominant cause, resonating with psychological frameworks like the Job Demand–Resource (JD-R) model.

The negligible correlation between RemoteRatio and burnout ($r = 0.011$) is particularly noteworthy post-pandemic. It implies that the remote/on-site balance, while impactful in real organizations, was neutralized in this synthetic context—useful as a methodological control baseline. Future extensions could introduce interaction terms (e.g., RemoteRatio × WorkHours) to examine compound effects of digital fatigue.

Visualization parameters were carefully chosen for interpretability. For example, the “coolwarm” color map enhanced contrast between positive and negative associations, while the annotation parameter annot=True ensured numeric precision on the matrix. This attention to technical visualization quality enhances transparency and pedagogical value, aligning with the open-science ethos of modern HR analytics.

Integrative Discussion and Agile Management Implications

The results collectively reaffirm that burnout is best understood as an emergent outcome of psychological strain amplified by workload intensity. Stress level and satisfaction are the primary emotional vectors; excessive work hours act as a structural amplifier. In agile contexts, this translates to the principle that velocity must never compromise sustainability. Teams operating at high throughput but low satisfaction are statistically more vulnerable to burnout spirals.

The neutral demographic effects further underline that agile health metrics should prioritize sentiment and behavioral indicators rather than demographic diversity alone. Continuous monitoring of satisfaction and stress can provide early warnings of process fatigue. Integrating such monitoring into sprint retrospectives or team dashboards would operationalize the study’s correlation framework within agile practices.

Another key insight concerns moderate effect sizes ($r \approx 0.2$–0.3): while significant, they reflect human complexity—no single factor dictates well-being. This supports the agile philosophy of inspect and adapt: well-being management should be iterative and multi-dimensional, combining workload balancing, psychological safety, and team dialogue.

Finally, the study’s methodological reproducibility—fully implemented in a single, lightweight Python script—demonstrates that data-driven team health assessment is technically and economically feasible for organizations of any scale. This operationalizes the agile manifesto’s emphasis on transparency, enabling empirical evaluation of human sustainability rather than relying solely on subjective observation.

Limitations and Future Research

Despite methodological rigor, this study relies on synthetic data, limiting generalization to real organizations. Future research could integrate authentic HR datasets or field survey data to validate observed relationships. Including temporal dimensions—tracking variable evolution over sprints—would also allow time-series modeling of burnout onset. Moreover, interaction effects (e.g., WorkHours × Stress, Satisfaction × RemoteRatio) merit investigation using regression or machine-learning models. Such extensions could quantify nonlinear effects often observed in agile teams under fluctuating workloads. Another limitation lies in the cross-sectional structure: it captures associations at one time point rather than causal trajectories. Future agile studies could adopt panel data to examine how team interventions alter burnout patterns over time. Nevertheless, the study's EDA and correlation workflow establishes a reproducible methodological foundation for human-centric analytics in agile environments. It bridges academic HR research and practical team management by demonstrating how simple, interpretable statistics can inform sustainable decision-making.

Conclusion

This study successfully demonstrated the use of unsupervised machine learning, specifically the K-Means clustering algorithm, to uncover hidden behavioral patterns in employee performance data. By analyzing 14,999 records from the Employee HR Dataset, four distinct clusters of agile workstyles were identified: Collaborative Core Workers, High-Impact Performers, Low-Engagement Staff, and Overloaded Experts. Each group represented a different balance between satisfaction, workload, and evaluation—key constructs of agile workforce sustainability. The findings confirmed that while most employees operate within a stable and satisfied equilibrium, a significant minority experience either under-utilization or overwork, both of which contradict the agile principle of maintaining a sustainable pace.

From a managerial standpoint, the clustering outcomes provide actionable guidance for aligning workforce strategy with agile values. Cluster 1 employees can be nurtured as agile champions, while Cluster 3 requires immediate workload balancing to mitigate burnout risk. Similarly, engagement and empowerment programs targeting Cluster 2 could elevate team cohesion and innovation potential. The study illustrates that machine-learning techniques are not limited to predictive modeling but can also serve as diagnostic tools for continuous improvement in human systems. Future extensions may integrate temporal data or sentiment analysis to capture evolving employee dynamics, enabling organizations to build adaptive, data-driven HR ecosystems that embody the spirit of agile transformation.

Declarations

Author Contributions

Conceptualization: A.A. and A.S.S.; Methodology: A.S.S.; Software: A.A.; Validation: A.A. and A.S.S.; Formal Analysis: A.A. and A.S.S.; Investigation: A.A.; Resources: A.S.S.; Data Curation: A.S.S.; Writing Original Draft

Preparation: A.A. and A.S.S.; Writing Review and Editing: A.S.S. and A.A.; Visualization: A.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] V. Y. T., “Balancing cadence and flow: evaluating agile frameworks for optimal software delivery outcomes,” *Digitus*, vol. 2, no. 2, pp. 110–121, 2024.

[2] Agile Alliance, “Manifesto for agile software development,” *Agile Alliance Journal*, vol. 1, no. 1, pp. 1–4, 2001.

[3] L. J. N. Dahmer, “Navigating project management in the digital transformation era,” *Int. Seven J. Multidisciplinary*, vol. 3, no. 5, pp. 1–10, 2024, doi: 10.56238/isevmjv3n5-008.

[4] V. S. Balasubramaniam, S. Vijayabaskar, P. K. Voola, R. Agarwal, and O. Goel, “Improving digital transformation in enterprises through agile methodologies,” *Int. J. Res. Publ. Seminars*, vol. 13, no. 5, pp. 507–537, 2022, doi: 10.36676/jrps.v13.i5.1527.

[5] G. Y. Rajendra and A. Jain, “Agile project management in retail technology: scaling agile frameworks for multi-channel retail technology deployments,” *Int. J. Res. Humanit. Soc. Sci.*, vol. 13, no. 3, pp. 1–15, 2025, doi: 10.63345/ijrhs.net.v13.i3.8.

[6] O. A. Popoola, H. E. Adama, C. D. Okeke, and A. E. Akinoso, “Conceptualizing agile development in digital transformations: theoretical foundations and practical applications,” *Eng. Sci. Technol. J.*, vol. 5, no. 4, pp. 1524-1541, 2024.

[7] O. E. Ajewumi, “The influence of agile organizational design on employee engagement and performance in the digital age,” *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 1–12, 2024, doi: 10.55248/gengpi.5.1024.2702.

[8] O. Ananieva and V. Khalilov, “Economic diagnostics of the efficiency of enterprise personnel management in the conditions of digitalization and military challenges,” *Sci. Notes Univ. KROK*, vol. 79, no. 1, pp. 80–87, 2025.

[9] S. Suwarno and N. Purwandi, “Scrum integration in cloud-based human capital management (HCM) project development,” *J. Inf. Syst. Technol.*, vol. 5, no. 2, pp. 45-51, 2024, doi: 10.37253/joint.v5i2.9577.

[10] I. O. Evans-Uzosike, C. G. Okatta, B. O. Otokiti, O. G. Ejike, and O. T. Kufile, “Hybrid workforce governance models: a technical review of digital monitoring systems, productivity analytics, and adaptive engagement frameworks,” *Int. J. Multidisciplinary Res. Growth Eval.*, vol. 2, no. 3, pp. 589–597, 2021.

[11] P. B. Sivakumar, “The role in empowering strategic decisions: the impact of HR analytics on talent management and organizational success,” *J. Inf. Syst. Eng. Manage.*, vol. 10, no. 39S, pp. 778–784, 2025, doi: 10.52783/jisem.v10i39s.7298.

[12] N. Carroll, K. Conboy, and X. Wang, “From transformation to normalisation: an exploratory study of a large-scale agile transformation,” *J. Inf. Technol.*, vol. 38, no. 2, pp. 267 - 303, 2023, doi: 10.1177/02683962231164428.

[13] G. Gautam and D. Kumar, “HR tech and cloud-based solutions: transforming workforce management through digital innovation,” *Eng. Sci.*, vol. 21, no. 1, pp. 885-902, 2025, doi: 10.69889/ac4bgt60.

[14] I. O. Evans-Uzosike et al., “Employee engagement analytics in agile organizations,” *IEEE Access*, vol. 9, no. 1, pp. 102334–102349, 2021.

[15] P. B. Sivakumar, “Integrating unsupervised machine learning into agile HR diagnostics,” *J. Bus. Anal. Intell.*, vol. 7, no. 2, pp. 90–104, 2025.

[16] N. Carroll, K. Conboy, and X. Wang, “Data science in agile workforce transformation,” *Inf. Syst. Front.*, vol. 25, no. 1, pp. 1–12, 2023.

[17] C. C. Yousef et al., “The effect of job and personal demands and resources on healthcare workers’ wellbeing: a cross-sectional study,” *PLOS ONE*, vol. 19, no. 4, p. e0303769, 2024, doi: 10.1371/journal.pone.0303769.

[18] Agile Alliance, “Manifesto for agile software development,” *Agile Alliance Journal*, vol. 1, no. 1, pp. 1–4, 2001.

[19] O. E. Ajewumi, “The influence of agile organizational design on employee engagement and performance in the digital age,” *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 1–12, 2024.

[20] L. J. N. Dahmer, “Navigating project management in the digital transformation era,” *Int. Seven J. Multidisciplinary*, vol. 3, no. 5, pp. 1–10, 2024.

[21] T. Peeters, K. V. De Voorde, and J. Paauwe, “The effects of working agile on team performance and engagement,” *Team Perform. Manage.*, vol. 28, no. 1–2, pp. 61–78, 2022, doi: 10.1108/tpm-07-2021-0049.

[22] A. Kanwal, S. Rafiq, and A. Afzal, “Impact of workload on teachers’ efficiency and their students’ academic achievement at the university level,” *Gomal Univ. J. Res.*, vol. 39, no. 2, pp. 1–16, 2023, doi: 10.51380/gujr-39-02-02.

[23] M. B. Pandjaitan, Moh. Khusaini, M. Aminuddin, and P. Suwarno, “The influence of hybrid working through readiness for change on sustainable performance management in the service sector in ASEAN,” *Decis. Sci. Lett.*, vol. 14, no. 3, pp. 1–12, 2025, doi: 10.5267/j.dsl.2025.7.003.

[24] A. Minbaeva, “HR analytics: from insight to action,” *J. Hum. Resour. Manage.*, vol. 34, no. 1, pp. 1–10, 2022.

[25] G. M. Kumar, “Machine learning project on employee attrition prediction with Python,” *Int. Sci. J. Eng. Manage.*, vol. 4, no. 8, pp. 1–10, 2025, doi: 10.55041/isjem04961.

[26] S. Bose and A. Jain, “Machine learning in performance appraisal systems,” *Int. J. Comput. Manage.*, vol. 18, no. 4, pp. 220–232, 2023.

[27] M. A. Vhora, V. Bhandwalkar, and P. M. Rege, “AI-driven HR analytics: enhancing decision-making in workforce planning,” *Sci. Temper*, vol. 15, no. 4, pp. 3299–3308, 2024, doi: 10.58414/scientifictemper.2024.15.4.39.

[28] W. Cho, S. Choi, and H. Choi, “Human resources analytics for public personnel management: concepts, cases, and caveats,” *Admin. Sci.*, vol. 13, no. 2, p. 41, 2023, doi: 10.3390/admsci13020041.

[29] S. Garg, S. Sinha, A. K. Kar, and M. Mani, “A review of machine learning applications in human resource management,” *Int. J. Product. Perform. Manage.*, vol. 70, no. 5, pp. 1590–1610, 2021, doi: 10.1108/ijppm-08-2020-0427.

[30] D. G. M. Khan, “A resource-based view assessment of big data analysis and its impact on strategic human resources quality management systems,” *J. Ind. Econ. Res.*, vol. 4, no. 2, pp. 844-853, 2024, doi: 10.52783/jier.v4i2.847.

[31] S. Dikshit et al., “Empowering employee wellness and building resilience in demanding work settings through predictive analytics,” *EAI Endorsed Trans. Internet Things*, vol. 10, no. December, pp. 1–6, 2023, doi: 10.4108/eetiot.4644.

[32] M. A. Khan, “K-means and workforce analytics,” *J. Bus. Intell.*, vol. 6, no. 1, pp. 1–10, 2024.

[33] S. Dikshit et al., “Machine learning applications in HR,” *Comput. Ind.*, vol. 159, no. 1, pp. 1–15, 2023.

[34] S. Jaiswal, N. Shaheen, A. Mangal, D. S. P. Singh, S. Jain, and R. Agarwal, “Transforming performance management systems for future-proof workforce development in the U.S.,” *J. Qual. Sci. Technol.*, vol. 1, no. 3, pp. 287–304, 2024, doi: 10.63345/jqst.v1i3.121.

[35] N. B. Moe, D. Šmite, M. Paasivaara, and C. Lassenius, “Finding the sweet spot for organizational control and team autonomy in large-scale agile software development,” *Empir. Softw. Eng.*, vol. 26, no. 101, pp. 1–41, 2021, doi: 10.1007/s10664-021-09967-3.

[36] O. E. Ajewumi, “The influence of agile organizational design on employee engagement and performance in the digital age,” *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 1–12, 2024.

[37] M. B. Pandjaitan, Moh. Khusaini, M. Aminuddin, and P. Suwarno, “The influence of hybrid working through readiness for change on sustainable performance management in the service sector in ASEAN,” *Decis. Sci. Lett.*, vol. 14, no. 3, pp. 1–12, 2025, doi: 10.5267/j.dsl.2025.7.003.

[38] I. Chkhaidze, “Reasons, challenges and strategies for implementing agile in today’s unpredictable environment,” *Innov. Econ. Manage.*, vol. 10, no. 2, pp. 72–86, 2023, doi: 10.46361/2449-2604.10.2.2023.72-86.

[39] A. O. Temitope, “Agile and organizational culture: fostering agile values and mindset,” *Int. J. Sci. Res. Archive*, vol. 7, no. 2, pp. 672–681, 2022, doi: 10.30574/ijsra.2022.7.2.0265.