Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

A STUDY ON HEALTHCARE'S PROJECT OXYGEN: A DATA-INFORMED APPROACH TO CULTIVATING HIGH-IMPACT MANAGERS

Smt. Garima Agrawal
Assistant Professor, Indian Institute of Management and Commerce
Khairtabad. Hyderabad, Telangana

Abstract

By utilizing the power of People Analytics, this study seeks to replicate the fundamental ideas of Google's "PROJECT OXYGEN" in a different industry context, namely the healthcare sector. This study starts by examining employee attrition trends to comprehend organizational dynamics and the possible managerial factors influencing employee retention, whereas Project Oxygen concentrated on determining the qualities of successful managers within Google.

In order to identify the main causes of attrition, such as overtime, monthly income, and tenure with the current manager, the study uses a combination of exploratory data analysis (EDA), logistic regression, and random forest modeling on employee data from a corporate HR environment. The results emphasize the value of supportive leadership and employee engagement by showing that longer relationships with managers typically result in lower attrition.

The study presents a transferable analytical framework that can be used to inform leadership development and HR decisions in the healthcare sector or other people-intensive industries. It highlights how data-driven insights can assist organizations in determining the managerial characteristics most compatible with long-lasting, high-performing teams, in addition to retention risks.

Indexed Words: Project Oxygen, People Analytics, Healthcare, HR Environment, logistic regression, Random Forest Modeling.

I. Introduction

In today's data-driven world, Human Resource (HR) Analytics has emerged as a transformation tool, empowering organizations to make evidence-based decisions that directly enhance productivity, engagement, and organizational culture. One of the most influential examples of applying HR analytics is Google's renowned "Project Oxygen", which revolutionized the understanding of managerial effectiveness.

Launched in the early 2000s, Project Oxygen was driven by a simple but critical question: "Do managers matter?" Google's People Analytics team collected vast amounts of data from employee surveys, performance reviews to identify behaviors that consistently led to high-performing teams. The analysis revealed eight core behaviors shared by effective managers, including coaching ability, empowering employees, effective communication, and technical expertise. These insights led to

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

targeted manager development programs and a measurable improvement in employee satisfaction and performance across Google.

Inspired by its success, organizations across sectors have shown interest in replicating Project Oxygen within their own unique contexts. One sector that stands to benefit immensely is **healthcare.** While sectors like banking and finance also rely heavily on leadership and analytics, healthcare presents a particularly compelling case. It is a labor-intensive, emotionally demanding industry where effective leadership can directly influence not only employee morale but also **patient outcomes** – making it a high-impact environment for studying managerial effectiveness.

Consider the following contrast: In banking, a manger's primary role may revolve around achieving sales targets, maintaining regulatory compliance, or managing risk – functions that are relatively structured and metric-drive. However, in healthcare, a manager may need to juggle complex schedules, address staff burnout, resolve interpersonal conflicts, ensure compliance with constantly evolving safety standards, and above all, uphold patient care quality. These diverse responsibilities demand a leadership style that is empathetic, agile, and communicative.

From emergency room supervisors managing high-pressure situations to department heads implementing digital health solutions, managerial behavior in healthcare can make or break team performance. Given these stakes, adapting Project Oxygen to this sector could uncover new dimensions of leadership that are vital yet underexplored. This paper proposes a framework for such a replication, focusing on identifying relevant behavior indicators, data points to collect, and methods of analysis.

The ultimate goal is to explore how HR analytics can support healthcare institutions in developing and sustaining effective managers, thereby enhancing both employee well-being and patient care outcomes. This research will provide practical insights for HR professionals and hospital administrators aiming to build a resilient, people-centric workplace culture through data-driven leadership strategies.

A. Rationale &Relevance in Relation to HR Analytics

Project Oxygen illustrates the power of HR analytics in converting qualitative aspects of human behavior into measurable, actionable insights. Replicating it in the healthcare industry is not only relevant but essential for several reasons:

- **Direct impact on patient outcomes:** Unlike many sectors, the consequences of poor management in healthcare can include compromised patient safety and well-being.
- **Burnout and mental health crises:** Healthcare professionals face high emotional stress, irregular hours, and critical decision-making – making empathetic and supportive leadership indispensable.
- **Team-based, multidisciplinary environment:** Healthcare delivery depends on effective collaboration among doctors, nurses, technicians, and administrative staff.
- **Retention and talent shortages:** Understanding what makes managers effective can help reduce attrition, especially among nurses and frontline workers.

HR analytics enables the identification of patterns in staff performance, satisfaction, and turnover. With the right data, healthcare institutions can design leadership development programs tailored to the needs of their workforce. Metrics such as employee engagement, absenteeism, patient care ratings, and incident reports can provide rich insights into how leadership affects organizational health.

Thus, by adapting Project Oxygen to the healthcare setting, organizations can bridge the gap between operational efficiency and compassionate caregiving – making HR analytics an indispensable tool for building the future of healthcare leadership.

B. Key Research Objectives

- To understand the core components and outcomes of Google's Project Oxygen.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- To identify how the principles of Project Oxygen can be adapted to the healthcare industry.
- To determine what data points should be collected to evaluate managerial effectiveness in healthcare.
- To analyze the potential HR and organizational outcomes of implementing such an initiative in the healthcare sector.

II. Literature Review/ Background

A. Concepts Relevant to the topic

1. Project Oxygen – A Case Study of Evidence-Based HR

Google's Project Oxygen (2010) emerged from an internal research initiative to determine the characteristics of effective managers. By analyzing performance reviews, employee feedback, and manager evaluations, Google identified eight key behaviors that effective managers share:

- Be a good coach
- Empower the team and do not micromanage
- Express interest in team members' success and well-being
- Be productive and results-oriented
- Be a good communicator and listen to the team
- Help with career development
- Have a clear vision and strategy for the team
- Possess key technical skills

These insights were derived using structured interviews, natural language processing, and performance analysis, reflecting a data-driven HR strategy that emphasizes behavioral patterns over subjective opinions *(Garvin, 2013).*Project Oxygen demonstrates that analytics can provide a framework for leadership development and performance management by replacing gut-feel assessments with actionable data.

2. HR Analytics – Turning Data into Decisions

HR analytics involves the systematic collection, analysis, and interpretation of workforce data to improve human resource decisions. According to Bassi (2011), effective HR analytics aligns with business outcomes and focuses on metrics that matter – such as engagement, productivity, and turnover. In healthcare, such analytics could evaluate indicators like:

- Nurse turnover rates
- Employee satisfaction surveys
- Incident reports and patient safety errors
- Overtime hours and absenteeism
- Training effectiveness and performance ratings

These data sources can help HR professionals uncover which managerial traits correlate with better outcomes. For instance, is a ward with fewer incidents led by a manager who communicates regularly and shows empathy? Analytics can help identify such links.

3. Leadership Theories and Frameworks

To replicate Project Oxygen in healthcare, it is crucial to contextualize the identified managerial behaviors within established leadership frameworks:

- **Transformational Leadership (Bass & Avolio, 1993):** Focuses on inspiring team members, setting clear goals, and aligning them with organizational vision. In healthcare, transformational leaders and linked to higher staff morale and innovation in care delivery.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- **Servant Leadership (Greenleaf, 1977):** Prioritizes the growth and well-being of employees. It emphasizes listening, empathy, and healing – essential traits in emotionally challenging environments like hospitals.
- **Emotional Intelligence (Goleman, 1995) :** Includes self-awareness, empathy, and emotional regulation. Healthcare leaders with high emotional intelligence are better equipped to manage conflict, support staff, and enhance patient interaction quality.

These frameworks provide theoretical backing to the behaviors uncovered by Project Oxygen and support their application in the health care industry.

4. HR Challenges in Healthcare:

The healthcare sector faces several industry-specific HR issues that make the implementation of HR analytics especially valuable:

- **High levels of burnout and stress** (Shanafelt et al., 2015): Managerial support and workload management can reduce burnout rates.
- **Skill shortages and high turnover rates** (WHO, 2020): Identifying traits of managers who retain staff can inform recruitment and training strategies.
- **Regulatory and compliance pressures:** HR analytics can flag trends in non-compliance that may stem for poor leadership or communication gaps.
- **Emotional nature of caregiving:** Managers who can offer emotional support and flexibility enhance team resilience.
- **Resistance to change and innovation:** Analytics can track how leadership behavior influences adoption of new technologies and protocols.

By addressing these challenges through a data-driven lens, HR departments can create more targeted leadership interventions that align with the unique needs of the healthcare sector.

III. Methodology/Approach

A. Research Design and Strategy

This study follows a **mixed-methods exploratory research design,** combining qualitative and quantitative data collection and analysis. The purpose of this approach is to first explore and understand the context-specific behaviors that define effective management in healthcare, and then quantify those behaviors through HR analytics to assess their impact on organizational and patient outcomes.

Inspired by Google's Project Oxygen, the study seeks to **replicate its core methodology** while adapting it to the unique demands of the healthcare sector. Project Oxygen began with open-ended data collection and thematic analysis, and then moved to quantifiable, pattern-based analytics. Our research similarly starts with **qualitative exploration,** followed by the development of **quantitative indicators** to evaluate and model managerial effectiveness.

This approach is particularly suitable for healthcare, where leadership is shaped not just by technical skills or outcomes, but also by emotional labor, team dynamics, and adaptability to frequent change. By blending qualitative insights with data-driven validation, we aim to develop a robust, context-sensitive framework for identifying and cultivating effective leaders in healthcare organizations.

B. Phase 1: Data Collection

1. Qualitative Methods – Understanding Managerial Behaviors

To replicate the foundation of Project Oxygen, we begin with in-depth qualitative data collection focused on identifying key managerial behaviors in healthcare settings.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- **Structured Interviews and Focus Groups**

We will conduct interviews and focus groups with a diverse group of hospital staff including doctors, nurses, technicians, and administrators. Questions will explore:

  - What traits do they value in their managers?
  - Which leadership behaviors positively or negatively affect their performance?
  - How do managers influence team morale and patient care?

- **Document Review and Observational Notes**

Review internal HR documents (e.g., performance reviews, employee feedback reports) and gathering observational data where feasible (e.g., shadowing managers during ward rounds or staff meetings) will help enrich the behavioral profile.

- **Thematic Coding and Synthesis**

Using grounded theory or thematic analysis, common themes and behaviors will be extracted. This mirrors Google's methodology, which involved coding open-ended responses to identify recurring leadership traits.

The outcome of Phase 1 will be a **preliminary list of effective managerial behaviors** specific to healthcare – both universal traits (e.g., communication, empathy) and healthcare-specific ones (e.g., patient advocacy, crisis management).

C. Phase 2: Developing Key Performance Indicators (KPIs)

Once themes have been identified, the next step is to translate them into **measurable variables** that can be tracked through HR analytics.

For example:

- **Empathy and Support:** Scores from employee engagement surveys (e.g., 'My manager listens to my concerns")
- **Communication Clarity:** Incident reports or staff satisfaction on information dissemination
- **Burnout Prevention:** Staff turnover rates, absenteeism, and overtime hours
- **Team Effectiveness:** Department-level patient satisfaction scores, peer review ratings.

These KPIs will be used to evaluate the presence and impact of the identified managerial behaviors.

D. Phase 3: Quantitative Data Collection

We will collect quantitative data from internal HR systems and hospital databases across selected healthcare institutions. The dataset may include:

- **Employee Metrics:**
  - Performance appraisal scores
  - Retention and attrition rates
  - Sick leaves and absenteeism
  - Employee engagement survey responses
  - Promotion and career progression data
- **Team and Unit-Level Metrics:**
  - Patient satisfaction and care quality ratings
  - Frequency of clinical errors or safety incidents
  - Team collaboration or conflict metrics (via pulse surveys)
  - Compliance with hospital protocols or training completion rates

- **Demographic and Contextual Variables:**
    - Department type (e.g., Emergency, Surgery, Admin)
    - Experience level of manager
    - Workload (e.g., staff-to-patient ratio)
    - Shift types (day/night/rotational)

Where possible, **data will be anonymized and aggregated** to maintain privacy and comply with healthcare data regulations (e.g., HIPAA and GDPR equivalents).

E. Phase 4: Data Analysis

1. Descriptive Analysis

To identify general trends, we will begin with descriptive statistics to understand the distribution of managerial behaviors and corresponding outcomes across the sample.

2. Correlation and Regression Analysis

We will run statistical models (e.g., multiple regression, logistic regression) to examine the strength and nature of relationships between managerial behaviors (independent variables) and employee/patient outcomes (dependent variables). For example:

- Does manager's communication score predict lower staff turnover?
- Does empathetic leadership correlate with higher patient satisfaction?

3. Cluster and Factor Analysis

Using techniques like cluster analysis or factor analysis, we may group managers into typologies (e.g. "supportive but disorganized" vs. "task-focused and efficient") and identify latent traits contributing to effectiveness.

4. Predictive Modeling

If the dataset is large enough, machine learning models (e.g. decision trees or random forests) can be applied to predict outcomes such as employee engagement or patient safety based on managerial behavior patterns.

F. Phase 5: Feedback and Refinement

The findings will be shared with participating institutions through workshops or debriefs. This allows:

- Validation of insights through stakeholder feedback
- Refinement of behavioral definitions or KPIs based on practitioner experience
- Identification of implementation barriers and contextual constraints

This feedback loop ensures the model is not just data-driven but also **practically relevant** to real-world healthcare environments.

G. Ethical Considerations

- **Informed Consent:** All participants in interviews and surveys will be informed about the purpose of the study and their right to withdraw at any time.
- **Data Confidentiality:** Data will be anonymized and stored securely. Only aggregated results will be shared.
- **Bias Mitigation:** Researcher bias during qualitative coding will be addressed through multiple coders and inter-rater reliability checks. Analytical models will be tested for fairness across gender, job roles, and departments.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

H. Limitations and Delimitations

While this study seeks to adapt and extend Project Oxygen into healthcare, it acknowledges certain limitations:

- The behavior indicators may vary across healthcare roles (e.g. surgeons vs. nurses)
- Organizational culture and resource constraints may affect implementation
- Data access limitations could restrict analysis depth in some hospitals

Nevertheless, these constraints are addressed through adaptive design and stakeholder input at each phase.

B. Proposed Framework/Implementation Plan

To successfully replicate **Google's Project Oxygen** in the **healthcare industry,** we need to apply a structured framework that ensures a data-driven approach to identifying what makes a great manager in this sector. The framework used for this purpose is the **HR Analytics Value Chain,** supported by insights from the **People Analytics Maturity Model.**

i. Overview of the Framework

The **HR Analytics Value Chain (Bassi, 2011)** is a step-by-step framework that helps organizations move from raw data collection to deriving actionable insights that influence strategic decisions. It involves five stages:

1. **Data Collection**
2. **Data Reporting**
3. **Advanced Analytics**
4. **Insight Generation**
5. **Decision Making and Action**

Let us understand how this can be applied to our project replicating **Project Oxygen** in healthcare.

ii. Step-by-Step Implementation Plan using the HR Analytics Value Chain

A. Data Collection (descriptive analytics)

In this stage, we collect both **quantitative and qualitative data** from multiple sources in the healthcare organization.

Key data points to collect:

- **Managerial Demographics:** Age, gender, experience, education level, department (nursing, administration, diagnostics, etc.)
- **Employee Feedback on Managers:** 360-degree feedback, satisfaction surveys, engagement scores, communication skills.
- **Performance Metrics:** Team turnover rates, absenteeism, patient satisfaction under that team, error rates, time to resolve patient issues.
- **Managerial Behaviors:** Initiative taken, team collaboration, coaching efforts, decision-making ability, adaptability to stress.
- **Training History:** Participation in leadership or soft-skills programs

Why it matters: These data points help identify behavioral patterns of successful managers and areas for improvement, especially in a high-pressure, life-impacting industry like healthcare.

B. Data reporting (Operational reporting)

In this phase, we clean the data, process it, and generate **dashboards and summaries** for HR leaders to view basic trends.

Tools and Outputs:

- **Dashboards** showing average satisfaction scores across different managerial roles.
- **Tables/Charts** comparing performance of teams led by different managers.
- **Heat maps**showing problem areas like departments with high attrition or low engagement

Example Output: A chart may show that departments with higher managers who conduct weekly check-ins have higher employee engagement and lower burnout.

C. Advanced Analytics (Diagnostic and Predictive Analytics)

Once basic trends are clear, we apply analytics techniques to find **patterns and relationships** in the data.

Techniques used:

- **Regression Analysis:** To find which manager behaviors (e.g. clear communication, regular feedback) predict high performance or satisfaction.
- **Clustering:** Grouping managers based on similar behavioral styles.
- **Sentiment Analysis:** On employee feedback/comments about managers.
- **Predictive Modeling:** Forecast which manager traits lead to high retention or productivity.

Outcome: For example, we may find that managers who are approachable and delegate tasks effectively reduce employee stress and turnover in emergency units.

D. Insight Generation

Here, we interpret the results of the analytics to identify **what really matters** in making a healthcare manager effective.

Key insights may include:

- Top-performing managers exhibit behaviors like empathy, active listening, and team empowerment.
- Managerial transparency and regular communication directly affect job satisfaction.
- Units with strong leadership see fewer patient complaints and lower staff absenteeism.

This insight becomes the foundation for training and policy decisions.

E. Decision Making and Action (Prescriptive Analytics)

Based on the insights, HR can now implement structured interventions to improve overall managerial effectiveness.

Actions to Take:

- **Redesign Leadership Training:** Focus more on emotional intelligence, communication, and conflict resolution.
- **Create a Manager Scorecard:** Include behavioral KPIs like team satisfaction, feedback frequency, patient satisfaction scores.
- **Build a Talent Pipeline:** Identify high-potential employees with similar traits and groom them for future managerial roles.
- **Feedback Systems:** Institutionalize regular 360-degree feedback loops.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in
Cover Page

Example: If you find that nurse managers who mentor junior staff have the most engaged teams, build mentorship KPIs into performance appraisals.

iii. Alignment with People Analytics Maturity Model

To assess whether a healthcare organization or any other organization is ready to replicate Project Oxygen, we adopt the **PwC People Analytics Maturity Model.** This model helps in understanding the organization's current stage in people analytics and what capabilities need to be built.

People Analytics Maturity Model (PwC): Application to this research

Level	Name	Description	Application to Project Oxygen Replication
1	Pre-Foundational	No formal people analytics function. Data is fragmented and limited to compliance/reporting.	Most organizations in this stage **cannot replicate** Project Oxygen yet. They need to first centralize HR data and build basic capabilities.
2	Foundational	HR data is being collected but mostly used for reporting. Some dashboards exist.	These organizations can begin to **explore basic patterns** like attrition rates or employee satisfaction by manager – early steps toward Project Oxygen.
3	Aspiring	Some analytics capabilities exist. There are efforts to link HR data to business outcomes.	Project Oxygen-style research becomes **viable** here. Managers' behaviors can be linked to team outcomes using statistical techniques.
4	Mature	People analytics is embedded in HR. Predictive modeling is used to make strategic decisions.	These organizations can **fully replicate and expand** Project Oxygen, using predictive analytics to identify the impact of leadership on outcomes like performance retention, or engagement.
5	Leading	Analytics is integrated across the business. Real-time dashboards, machine learning, and AI tools guide strategy.	Project Oxygen can be taken to the next level – like **real-time analysis, AI-driven coaching tools**, and linking manager behavior with **organizational performance metrics**.

Why this Framework?

- It helps evaluate the **feasibility** of such a project based on an organization's data and analytics maturity.
- It serves as a **guideline for the required tools, data, and techniques** at each stage.
- It aligns well with industry best practices and global benchmarks for people analytics capability development.

Cover Page

iv. Implementation Timeline (tentative)

Phase	Time Frame	Activities
Phase 1: Planning	Month 1	Define roles, tools, secure data access
Phase 2: Data Collection	Month 2-3	Surveys, interviews, HR data mining
Phase 3: Analytics	Month 4-5	Clean data, run regressions, and clustering
Phase 4: Insights and Reports	Month 6	Build dashboards, summarize key findings
Phase 5: Action Plan	Month 7-8	Training design, scorecard development
Phase 6: Evaluation	Month 9 onwards	Track performance post-intervention

v. Stakeholder Involvement

Successful implementation requires collaboration across departments:

- **HR team:** Leading the design and execution.
- **Department Heads:** Providing access to team data and participating in evaluations.
- **IT Team:** Supporting data collection and dashboard creation.
- **Employees:** Providing feedback and participating in surveys.

vi. Benefits of using this framework

- **Structured Approach:** Helps move from raw data to real decisions.
- **Data-driven culture:** Encourages decisions based on facts, not assumptions.
- **Alignment with Business Goals:** Helps HR contribute strategically to healthcare outcomes.
- **Scalable:** can be expanded to multiple units or adapted to other roles also.

IV. Data Analysis and Findings

A. Introduction to the Dataset and Context

To carry out a replicable version of Project Oxygen in a different industry, such as healthcare, it is essential to first understand the kinds of variables that affect employee retention and workplace satisfaction. Due to the unavailability of healthcare-specific employee datasets, this analysis leverages the publicly available **IBM HR Analytics Employee Attrition & Performance** dataset as a suitable proxy. While this dataset originates from the technology sector it includes several employee-related attributes that are commonly relevant across industries, such as job satisfaction, work-life balance, job role, and overtime – all of which hold significant importance in sectors like healthcare.

The goal is not to directly infer healthcare-specific conclusions, but rather to **demonstrate an analytical framework** that can be **applied to healthcare HR data,** provided similar data points are collected. This aligns with the spirit of Project Oxygen, where the focus is on using data-driven insights to guide managerial and HR decisions.

The dataset consists of **1,470 records** with various features such as:

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- **Demographics** (Age, Gender, Marital Status)
- **Job roles and performance** (Department, Job Role, Job Level, Performance Rating)
- **Work-related behavior** (OverTime, Distance from Home, Total Working Years)
- **Compensation and benefits** (Monthly Income, Stock Options, etc.)
- **Satisfaction metrics** (Job Satisfaction, Environment Satisfaction, Work-Life Balance)
- **Attrition status** (target variable: whether the employee left or stayed)

Basic cleaning was performed to remove redundant columns (e.g., EmployeeNumber, Over18), and categorical variables were encoded for modeling purposes.

B. Methodological Alignment to the Research Question

The research question revolves around **replicating Project Oxygen in a different industry** – specifically, healthcare – by identifying relevant data points and conducting an evidence-based analysis to uncover what factors influence employee experience, performance, and attrition.

To approach this, the methodology mirrors the foundational steps of Google's Project Oxygen:

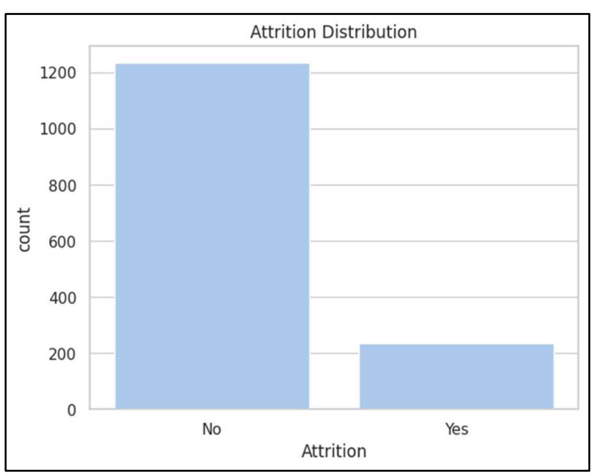
- **"Collecting people-related data:** The dataset includes employee-level information on demographics, compensation, performance metrics, job roles, satisfaction scores, and attrition."
- **"Analyzing patterns:** Through exploratory analysis, correlations, and predictive modeling (Logistic Regression and Random Forest), the study seeks to identify patterns that affect whether employees stay or leave."
- **"Drawing actionable insights:** By identifying the most influential features, the goal is to guide what healthcare HR teams might need to monitor and measure if they were to replicate this model."

While the dataset is from IBM and not a healthcare company, the methodology is **generalizable.** The same steps – collecting satisfaction scores, job roles, manager interaction levels, etc. – can be mirrored in hospitals or clinics, allowing decision-makers in healthcare to **apply similar people analytics frameworks.**

C. Univariate Analysis

Univariate analysis allows us to understand the basic distribution and frequency of individual variables within the dataset. This forms the foundation for identifying early signals of potential drivers of attrition – critical for designing effective people analytics strategies similar to Google's Project Oxygen but tailored to the healthcare industry context.

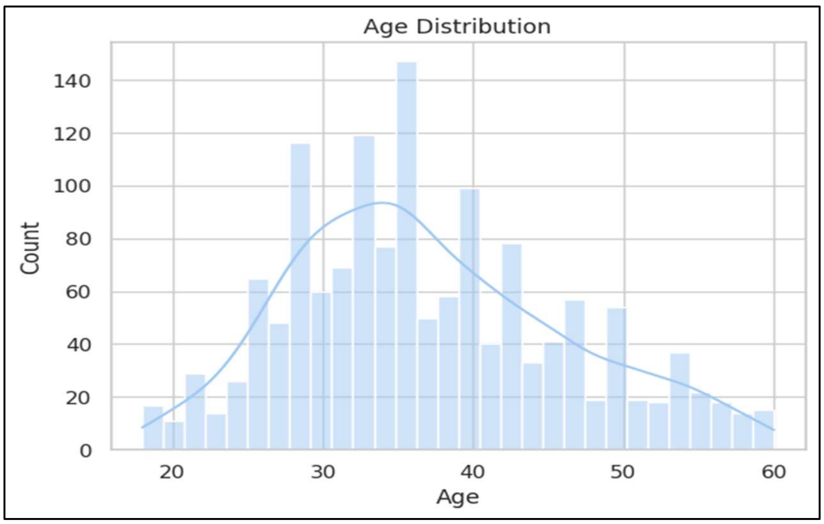
C.1 Attrition Distribution



The bar chart above shows a highly imbalanced distribution between employees who stayed with the organization and those who left. A significant majority (over 80%) of the employees did **not** leave the organization, while a relatively small portion (under 20%) did.

This imbalance reflects a real-world scenario common in organizational settings – most employees tend to stay unless significant issues arise. However, for people analytics and managerial effectiveness frameworks, even this minority group of attrited employees warrants serious investigation. High-potential employee turnover, especially in healthcare where skill shortages are prevalent, can drastically impact service quality and team performance.

C.2 Age Distribution



The distribution of age shows a concentration between **30 and 40 years,** with a tail extending into the 50s and early 60s. This pattern aligns with a typical workforce demographic, particularly in technical and operational roles, such as those found in the healthcare sector (e.g. nurses, technicians, mid-career clinicians).

Insight: Understanding the age profile is crucial when replicating Project Oxygen in healthcare. For instance:

- Younger employees may value learning opportunities and flexible hours.
- Mid-career employees may expect clear pathways for progression.
- Older employees might prioritize job security or work-life balance.

Each of these needs could impact attrition differently, which supports the rationale for collecting age as a data point in a healthcare-context version of Project Oxygen.

How it connects to the research question

Project Oxygen originally aimed to determine what makes a good manager. But to replicate such a framework in healthcare, we must also understand **what types of employees are leaving, their demographic characteristics,** and **how organizational factors may interact with these characteristics.**

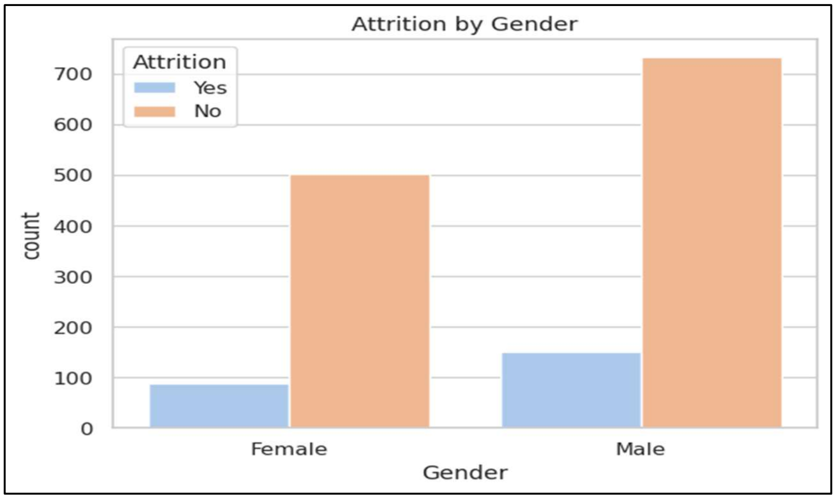
The univariate analysis provides an initial lens into the dataset's structure and surfaces key employee attributes that can help design relevant people-focused interventions. These foundational findings will be further enriched in the next sections through bivariate and multivariable analyses.

Cover Page
INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

D. Bivariate Analysis

Bivariate analysis aims to explore the relationship between the dependent variable (Attrition) and key independent variables in the dataset. This analysis helps us understand how different employee attributes are associated with attrition behavior. Such insights are crucial when attempting to replicate frameworks like Google's Project Oxygen in other industries, such as healthcare.

D.1 Attrition by Gender

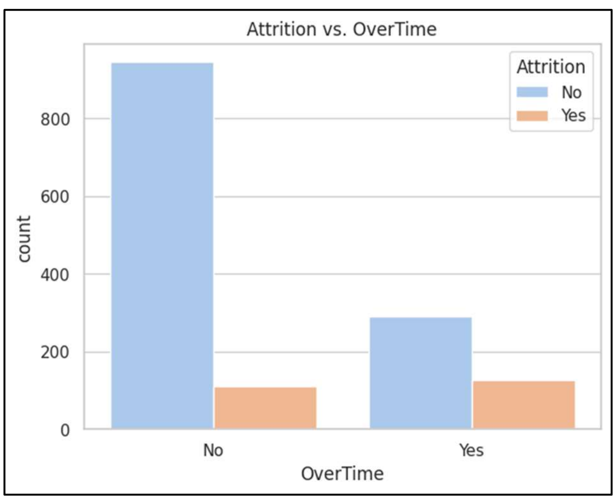


From the above plot, we observe that while the number of males and females differs, attrition exits in both groups. Notably:

- A higher absolute number of males are leaving compared to females.
- However, when proportionally analyzed, females also exhibit significant attrition.

Implication: When designing managerial effectiveness programs in the healthcare sector, gender dynamics should be taken into account. Differences in work expectations, support systems, or job roles might affect how men and women perceive organizational support and job satisfaction.

D.2 Attrition vs. OverTime



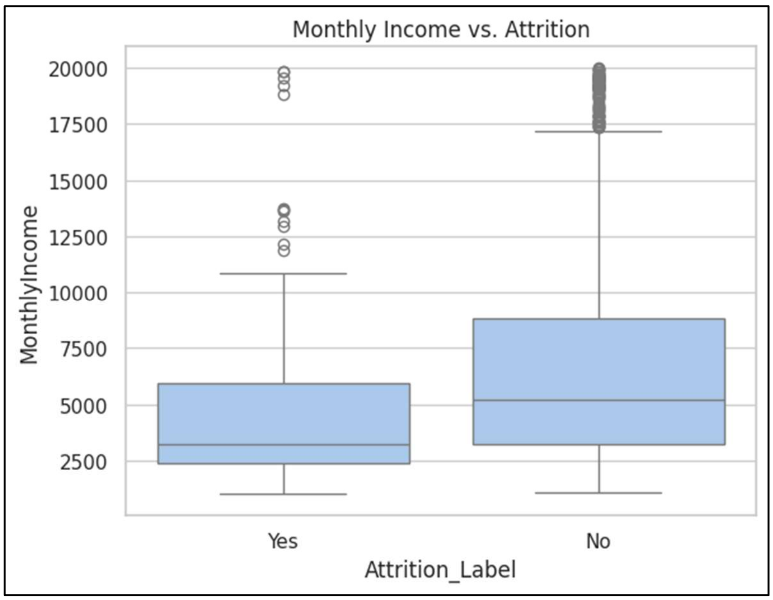
INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

Employees working overtime show a much higher rate of attrition compared to those who don't:

- Among employees who do not work overtime, attrition is relatively low.
- Conversely, those required to work extra hours demonstrate a much higher likelihood of leaving.

Implication: This insight is highly relevant for high-pressure environments like hospitals and banking, where overtime can be routine. In replicating Project Oxygen, one recommendation could be to focus on work-life balance and managerial support in scheduling.

D.3 Monthly Income vs. Attrition



This boxplot reveals a clear trend:

- Employees who left the company typically had lower monthly incomes.
- Those who stayed tended to earn higher salaries, with a wider range of income distribution.

Implication: Compensation and financial recognition are key retention factors. In the healthcare industry, frontline staff like nurses or technicians might experience burnout combined with relatively lower pay, leading to higher attrition. Managerial policies that support pay equity and career progression could help reduce turnover.

Connecting Back to Project Oxygen

These findings align with the core principles of Project Oxygen, which emphasized manager effectiveness in employee engagement and retention. While Project Oxygen focused on qualitative behaviors (e.g, 'is a good coach', 'empowers the team'), our bivariate analysis shows that **quantitative factors like pay, workload, and demographic nuances also play a critical role** in shaping attrition patterns:

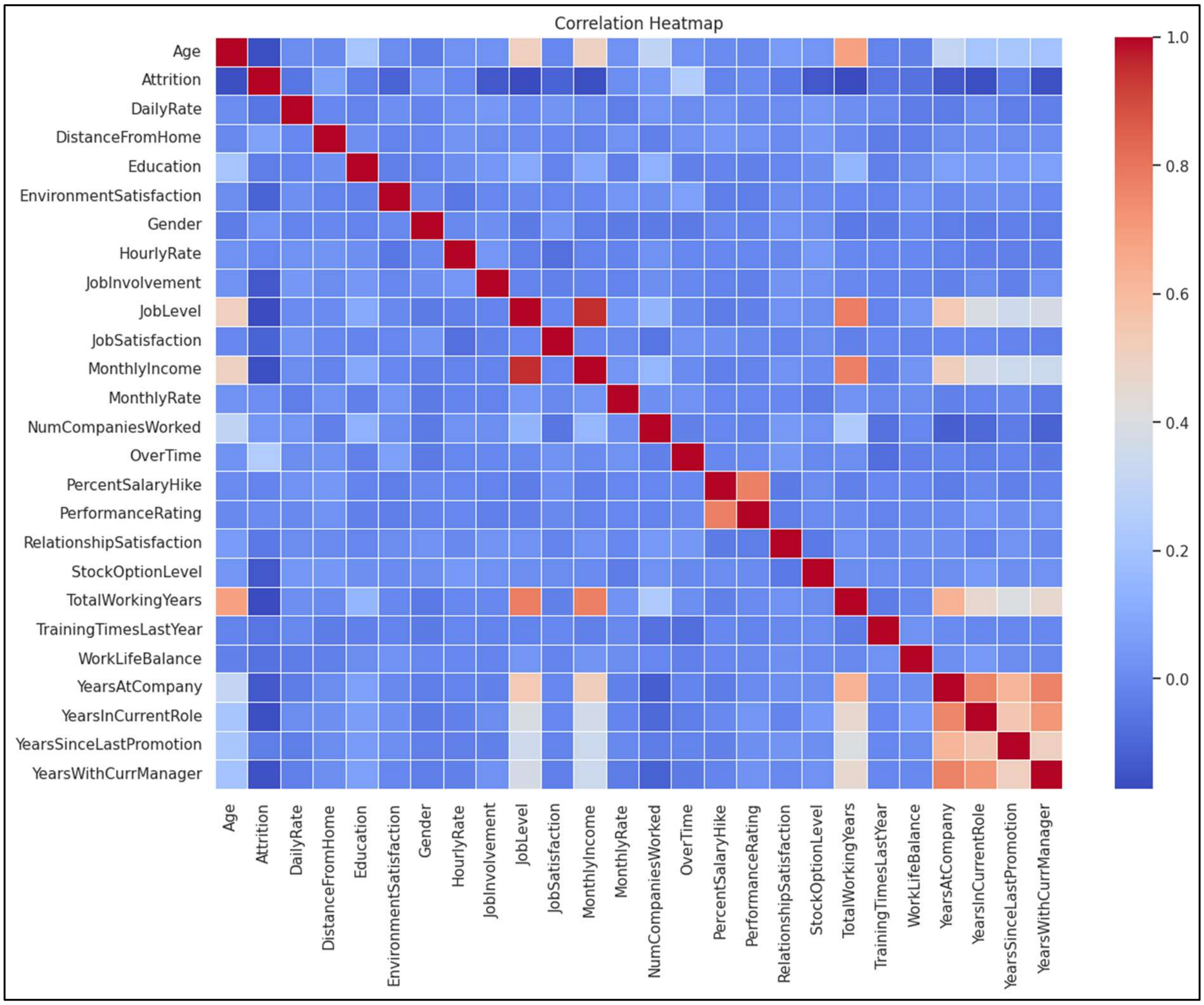
When applying this framework to industries like healthcare:

- It becomes essential to supplement managerial coaching with data-driven policies on compensation, equitable workloads, and individualized support.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- Managers must be equipped not only with interpersonal skills but also with awareness of data-backed organizational stressors.

E. Correlation Analysis



To identify key numerical factors influencing employee attrition, a Pearson correlation matrix was computed. Several important insights emerged:

- **OverTime (encoded):** Showed a **strong positive correlation** with attrition, suggesting that employees required to work overtime are more likely to leave. This supports the understanding that excessive workload or work-life imbalance is a critical factor in turnover.
- **MonthlyIncome:** Demonstrated a **moderate negative correlation** with attrition. Employees with lower salaries were more likely to leave, highlighting the role of compensation in retention.
- **Age and TotalWorkingYears:** Both showed **negative correlations,** indicating that younger and less experienced employees tend to leave more frequently. This is consistent with industry trends where early-career professionals often explore opportunities more actively.
- **YearsAtCompany:**Also, negatively correlation, reinforcing that tenure is associated with greater organizational loyalty and lower turnover.


Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- **YearsWithCurrManager:** Notably, this variable exhibited a **negative correlation** with attrition, suggesting that employees who have spent more time with their current manager are less likely to quit. This may imply that strong managerial relationships and consistent leadership positively impact retention. It is possible that effective managers cultivate a supportive environment, improving employee satisfaction and reducing turnover risk.

These correlations align with HR research and organizational behavior literature, emphasizing that both structural (e.g. compensation, overtime) and interpersonal (e.g. managerial quality) factors are critical in understanding attrition dynamics.

F. Predictive Modeling: Logistic Regression

F.1 Introduction and Rationale

Logistic Regression is a fundamental classification algorithm widely used in human resources analytics to model binary outcomes such as employee attrition. It estimates the probability of an employee leaving the organization (Attrition = 1) based on a combination of independent variables (e.g. income, job satisfaction, overtime). Its interpretability and robustness make it a suitable choice for initialpredictive modelling.

In this context, logistic regression was applied to assess how well organizational and demographic factors can predict employee attrition.

F.2 Model Performance and Evaluation

The logistic regression model achieved an overall **accuracy of 86.5%** on the test dataset. However, a deeper look into performance metrics for each class reveals an important aspect:

Logistic Regression Accuracy: 0.8684807256235828

	precision	recall	f1-score	support
0	0.88	0.98	0.93	380
1	0.58	0.18	0.28	61
accuracy			0.87	441
macro avg	0.73	0.58	0.60	441
weighted avg	0.84	0.87	0.84	441

- **High performance for Class 0 (No Attrition):**

The model correctly identifies the majority of employees who stay, with a very high recall of 98%. This means it rarely misclassifies employees who actually stay as leaving.

- **Low performance for Class 1 (Attrition):**

While the precision is moderate (0.58), the recall is **very low (0.18).** This means that **82% of employees who actually left were not correctly identified** by the model.

F.3 Interpretation and Challenges

The imbalance in performance across classes suggests that the model is biased toward predicting dominant class (No attrition), which is expected given the **class imbalance** in the dataset.

Despite the high accuracy, the model **struggles to detect actual cases of attrition,** which are the most critical for HR interventions. This emphasizes the need for:

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

- **Resampling techniques** (e.g. SMOTE, under-sampling)
- **Cost-sensitive modeling**
- **Alterative algorithms** that can better handle imbalanced classification (e.g. Random Forest, XGBoost)

G. Random Forest Classification Results

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and merges them together to get a more accurate and stable prediction. It helps reduce overfitting and improves generalization. This makes it particularly effective for handling classification problems like predicting employee attrition, where various features interact in complex, non-linear ways.

G.1 Model Performance

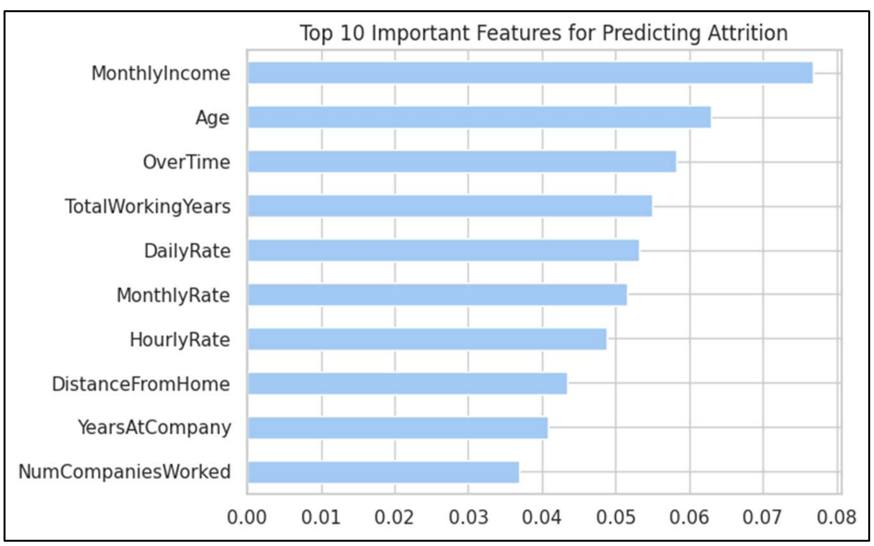
```
Random Forest Accuracy: 0.8684807256235828
              precision    recall  f1-score   support

           0       0.87      0.99      0.93       380
           1       0.64      0.11      0.19        61

    accuracy                           0.87       441
   macro avg       0.76      0.55      0.56       441
weighted avg       0.84      0.87      0.83       441
```

While the model performs very well in identifying employees who are not likely to leave the company (class 0), it struggles with correctly identifying who are likely to leave (class 1), as evidenced by the low recall score of 0.11. This suggests the model is highly biased toward the majority class, which is a common issue in imbalanced datasets like this one.

Despite this, the overall accuracy is high. However, in real-world applications, especially in attrition prediction, **recall for the "Yes" class is critical,** since the business value lies in identifying employees at risk of leaving.

G.2 Feature Importance Analysis and Business Implication

The bar plot illustrates the **top 10 features** that the Random Forest model found most important for predicting attrition.



Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

The findings from the Random Forest model offer significant value not just in prediction, but in driving strategic HR and managerial decisions – aligning directly with the core objective of this research.

- The **high importance of Monthly Income, OverTime, and Age** suggests that compensation policies, workload balance, and generational engagement strategies need careful attention. Managers should work closely with HR to ensure competitive salaries, fair workload distribution, and tailored career development programs, particularly for younger employees.
- The **importance of feature like YearsAtCompany and TotalWorkingYears** indicates that early career experiences and employee tenure can influence retention. Managers should consider onboarding quality, mentorship programs, and early recognition efforts to nurture longer-term commitment.
- Interestingly, features such as **DistanceFromHome and NumCompaniesWorked**imply that external life factors and prior work behavior also play ar ole. While these may be beyond direct managerial control, acknowledging these during hiring or engagement planning can help reduce attrition risk.
- These results reinforce the importance of **managerial awareness:** leaders and managers are in the best position to observe signs of burnout (e.g. consistent overtime), dissatisfaction, or disengagement. Therefore, the insights should not just stay with the data team – they must be **translated into everyday management practices,** such as regular one-on-ones, workload audits, and retention-focused policy design.

Ultimately, this analysis supports the broader goal of this research: to equip organizations with data-driven insights that allow managers to play an active role, thereby ensuring organizational stability and employee satisfaction.

V. Discussion

This research was inspired by Google's *Project Oxygen,* which identified key managerial behaviors that contribute to team performance and employee satisfaction. Our guiding research question was:

'If you were replicating Project Oxygen in a different industry (say, banking or healthcare), how would you carry out the experiment and what data points would you collect – any why?'

Although we did not have access to real-world data from the healthcare or banking sectors, we approached this analysis using a **proxy dataset from a corporate environment** to demonstrate the process and methodology we would adopt if such data were available. The purpose was not to solve attrition in this specific dataset alone, but to **prototype a data-driven framework** that could be adapted to other industries.

To begin with, we focused on **understanding the organizational context,** employee demographics, and the underlying causes of attrition. This was essential before jumping to managerial evaluation – because in any industry, **effective managerial qualities depend on the broader cultural and operational dynamics.** For instance, what works in tech may not work in healthcare, where stress factors, hierarchical structures, and employee roles are drastically different.

We started with **univariate analysis** to get a sense of how features like age, income, and attrition levels are distributed. Then, through **bivariate analysis,** we explored the relationships between key variables (e.g. gender vs. attrition, overtime vs. attrition) to identify patterns that may hint at root causes. This helped us understand which factors may indirectly reflect **managerial influence,** such as workload management or recognition practices.

Next, we conduced **correlation analysis** and feature importance evaluation, which offered insights into what variables are most predictive of attrition. Interestingly, features like *Years with Current Manager* were negatively correlation with attrition – suggesting that managerial stability and strong working relationships play a role in retention. This echoes the goals of Project Oxygen – to identify what makes managers effective.

Finally, we applied **logistic regression and random forest models** to test how well attrition could be predicted using the available features. While the overall model accuracy was high, the performance on the minority class (employees who left) was lower, reflection the **real-world challenge of predicting rare events** like voluntary attrition.

Through this process, we demonstrated a **scalable analytical approach** that could be applied in industries like banking or healthcare. If we were to replicate this in those sectors, we would aim to collect similar data points – age, job satisfaction, income, tenure, department, overtime, manager tenure, etc. – while also incorporating **industry-specific factors,** such as patient load (in healthcare) or client interaction levels (in banking). The goal would remain the same: **to identify which managerial behaviors and organizational factors are linked to lower attrition,** and use that insight to build better leadership frameworks.

In summary, this project serves as a **proof of concept.** While we worked with a general corporate dataset, the methodology is adaptable and aligned with our research objective – using data to uncover what makes a great manager in a given industry context.

VI. Conclusion

This project set out to explore how data-driven methods can help replicate and adapt the essence of Google's *Project Oxygen* – understanding what makes a great manager – in other industries such as healthcare or banking. Although we did not work with data from those specific sectors,our analysis on a corporate attrition dataset served as a **demonstrative framework** for how such an investigation could be conducted.

Through detailed exploratory data analysis, correlation assessments, and predictive modelling, we uncovered key insights into the factors that influence employee attrition. Features such as Monthly Income, OverTime, Age, and notably Years with current manager emerged as significant predictors – pointing to the broader influence of managerial support and work-life balance on employee retention.

Despite the high overall accuracy of our models, we observed challenges in predicting actual attrition cases (the minority class), highlighting a common limitation in real-world HR analytics – class imbalance and the subtlety of human decision-making.

Nonetheless, our findings demonstrate that a structured, data-centric approach can offer valuable signals about organizational health and managerial effectiveness. If replicated in the healthcare or banking sectors with contextualized data, this methodology could help leaders better understand their teams, reduce attrition, and improve the overall workplace environment.

Ultimately, our project underscores that **great managers don't just emerge – they are understood through data, refined by feedback, and supported by organizational insight.** By continuing to invest in data-backed leadership analysis, organizations can foster healthier, more resilient workplaces across any industry.

References (APA Style)

1. Bock, L., & Google Inc. (2011). *Google's Innovation Factory: Testing, Culture, and Infrastructure*. Retrieved from https://research.google.com/pubs/archive/41672.pdfresearch.google.com+3research.google.com+3research.google.com+3
2. Ghazi, A., Fallucchi, F., & Rakhra, H. K. (2022). Predicting Employee Attrition Using Logistic Regression with Feature Selection. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 6(4). Retrieved from https://jurnal.polgan.ac.id/index.php/sinkron/article/download/11783/1139Jurnal Politeknik Ganesha Medan
3. Chen, B. Z., & Tsinghua University. (2023). Factors of Employee Attrition: A Logistic Regression Approach. *ResearchGate*. Retrieved from

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :9.014(2025); IC VALUE:5.16; ISI VALUE:2.286
PEER REVIEWED AND REFEREED INTERNATIONAL JOURNAL
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:14, Issue:10(3), October, 2025
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed : Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available : www.ijmer.in

https://www.researchgate.net/publication/373896134_Factors_of_Employee_Attrition_A_Logistic_Regression_ApproachResearchGate
4. ResearchGate. (2023). A Study on Understanding Employee Attrition: Causes and Consequences in the Workplace. Retrieved from https://www.researchgate.net/publication/374813270_A_Study_on_Understanding_Employee_Attrition-Causes_and_Consequences_in_the_WorkplaceResearchGate
5. Investopedia. (2023). What Is Attrition in Business? Meaning, Types, and Benefits. Retrieved from https://www.investopedia.com/terms/a/attrition.aspInvestopedia
6. Liberty University. (2022). Exploring Factors for Employee Attrition. *Doctoral Dissertations and Projects*. Retrieved from https://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=5773&context=doctoral
7. Bassi, L. (2011). *Raging Debates in HR Analytics*. HR Executive.
8. Minbaeva, D. (2018). *Building credible human capital analytics for organizational competitive advantage*. Human Resource Management.
9. Bersin by Deloitte (2017). *People Analytics Maturity Model*.
10. Harvard Business Review (2013). *How Google Sold Its Engineers on Management*.
11. PWC (2023). *Future of Work and Skills Report*.
12. SHRM Reports on Managerial Effectiveness (2021, 2022).