CHAPTER -I

INTRODUCTION

1.1. DEFINITIONS AND CONCEPTS:

1.1.1. Introduction to employee performance prediction:

Employee performance analysis is a systematic evaluation of the contribution of an individual's performance of a person to the organization, measuring their efficiency, productivity, engagement, and resource utilization. Across organizations, business analytics, machine learning, and business intelligence tools are used to measure workforce effectiveness, streamline performance, and align employee pursuits with enterprise strategy.

Today the business environment has become data-driven and to make decisions organisations need to transcend subjective assessments and rely on quantitative and analytical approaches. Using predictive analytics, statistical models, and performance metrics allow businesses to manage workforce productivity, strengthen employee development programs, and utilize organizational resources more effectively.

1.1.2. Understanding Employee Performance Analysis

Analysis of employee performance is the process of systematically measuring and evaluating employee performance in relation to the goals of the organization. It takes into account the following several parameters:

Task Completion Rates – Ability to meet deadlines and complete required tasks in an efficient manner

Quality of Work– The level and correctness of output produced by an employee.

Teamwork & Collaboration – Contribution to groups projects and the ability to work well with peers.

Leadership & Initiative – The capacity to take on responsibility, make decisions, and drive projects.

Problem Solving – The ability to recognize and solve problems at work efficiently.

These elements aid in determining high-achieving workers, domains that need enhancement, and options for upskilling. That, together, helps organizations then, significant training, career development plans and performance incentives can be offered to improve the effectiveness of the workforce.

1.1.3. Importance of employee performance Analysis:

Analysing employee performance is important for businesses since it aids in the decision-making process

Optimizing workforce productivity:

With awareness of individual efforts, companies can effectively establish performance expectations that are within reach and formulate plans to boost productivity.

Data-Driven HR Decisions:

Removes the elements of bias, which ensures that promotions, rewards, and recognition are derived from quantifiable metrics and not opinions.

Employee Engagement & Motivation:

Regular performance tracking and recognition of achievements creates clarity and drives motivation among employees, helping them understand their future trajectory as well.

Use of Strategic Resources:

Organization of performance analysis clarifies over how people use all available resources, making it easy to allocate resources efficiently and save money, too.

Culture of Continual Improvement:

Promotes a mindset of being open to new ideas and methods and a commitment to continuous improvement through training and feedback which makes employees deliver their best.

Through the integration of business analytics with performance evaluation models, organizations develop a continuing path to employee growth and organizational success

1.1.4. Business analytics in employee performance evaluation:

Business analytics (BA) can be a critical aid in measuring how well your workforce is performing. HR Analytics is the process of gathering, processing, and analysing employee data to draw insights and improve decision making.

1.Descriptive Analytics:

Is based on historical performance patterns.

This includes analysing historical productivity, attendance, and engagement to identify behavioural patterns.

It helps organizations to identify high and low performers

2. **Predictive Analytics:**

Forecasting Trends with Data Insights

Relies on historical data to predict how performance trends will unfold in the future.

Finds employees who could be drifting away from their roles and predicts what factors might boost their motivation.

Assists HR departments in crafting forward-thinking training and development strategies.

3. prescriptive analytics

it really dives into actionable insights. Action-oriented analytics that guide decision-making.

Offers practical recommendations derived from data findings.

Advises on tailored training sessions, motivational rewards, and tweaks to workloads to boost employee performance.

4. Understanding employee sentiment:

Looks at what employees say in feedback, emails, and surveys to gauge how satisfied and engaged they are.

Employs NLP methods to evaluate how employees feel about their work environment.

1.1.5. Machine learning in employee performance analysis:

Employee performance analysis is improved with ML technologies such as pattern recognition, trend prediction, and decision automation. Applications include:

Regression Models:

Analyse the correlation of independent variables like training and engagement with dependent variables like work output and quality.

Clustering Techniques:

Cluster employees according to their performance, skills, or degree of engagement.

Facilitates effective and customized career development plans, as well as focused HR actions.

Random Forest:

Random Forests, an ensemble machine learning model, builds numerous decisions trees and combines their outputs. Like other ensemble approaches, Rf has improved accuracy in both training and validation phases and has predictive performance in estimating employee performance by training and education as well as prior performance (Raza et al., 2022). It also helps alleviate overfitting issues.

Neural Networks:

In particular, deep learning models excel at handling complex and nonlinear data. Neural networks are capable of learning intricate connections and patterns from numerous input features to predict employee performance (Fallucchi, 2020). More sophisticated algorithms, such as multi-layer perceptron, other deep learning frameworks, and less complex algorithms are able to capture subtleties more straightforward algorithms would not understand.

Support vector machine:

Support vector machines are effective in performing multiclass employee performance prediction (e.g., Low, medium, High) because they can be used for activity classification. SVM aims to detect the feature space hyperplane which optimally divides the different classes. SVM uses the gap between categories to provide detailed performance prediction, especially when there are complex decision boundaries (Fallucchi, 2020).

1.1.6. Business intelligence in employee performance analysis:

business intelligence (BI) solutions are essential for analysing and displaying worker performance data.

Importance of business intelligence in employee performance analysis:

Centralized Data Integration:

which combines information from performance reviews, employee surveys, and HR systems

Business Intelligence (BI) tools like Power BI play a crucial role in visualizing and analysing workforce performance data.

Interactive Dashboards:

Real-time performance data on KPIs such as engagement, productivity, and training efficacy are provided

Advanced Data Filtering:

Analysis by department, job role, or seniority level is made possible by advanced data filtering.

1.1.7. Integration of Business Analytics, Machine Learning, and Business Intelligence

Business intelligence (BI), machine learning (ML), and business analytics (BA) are all combined in a thorough data-driven strategy to revolutionize workforce management and decision-making procedures. Organizations may improve overall performance by utilizing these cutting-edge technologies to increase efficiency, and strategic insights.

predicting performance trends and workforce risks:

To find patterns and trends in employee performance, machine learning algorithms examine historical data. By anticipating workforce issues like poor engagement, high turnover risk, or productivity drops, this predictive capacity enables firms to take preventive measures.

Developing personalized training and career development:

In order to create personalized learning paths, AI-driven analytics evaluate workers' skills, shortcomings, and career goals. These programs improve employee retention and happiness, maximize skill development, and match training activities with corporate objectives.

Automating employee performance with dashboards:

business intelligence solutions automate performance tracking, which lessens subjectivity and bias in assessments. With the use of real-time dashboards that offer thorough performance insights, managers and HR leaders can make data-driven choices about position changes, rewards, and promotions.

Strategic planning:

Key workforce KPIs are instantly accessible to HR and leadership teams thanks to integrated BI tools, which combine and show data from many sources. Better resource allocation, talent management, and policy modifications that support corporate objectives are made possible by these insights.

Organizations may optimize workforce planning, streamline HR processes, and promote a continuous improvement culture by combining business intelligence, machine learning, and business analytics. In addition to increasing team and individual productivity, this data-driven strategy makes workforce management more strategic, predictive, and adaptable, which promotes long-term organizational success.

1.1.8. A Structured Approach to Employee Performance Analysis

Steps to Implement a Data-Driven Performance Framework

Step 1: Collecting data:

Using business analytics and machine learning to gather pertinent data will be the first step in predicting employee success. This information should incorporate training records, performance data from the past, job-related data, demography of the workforce and any other pertinent elements that could affect output. Verifying the integrity and accessibility of data is essential to building a strong basis for precise forecasts.

Step 1: Data preprocessing:

It is necessary to preprocess the collected data before implementing machine learning algorithms. This includes handling missing values, removing duplicates or incorrect entries, and cleansing the data. Additionally, feature engineering Techniques can be used to glean valuable characteristics from the available data, perhaps enhancing prediction models' accuracy.

Label Encoding: The numerical representation of categorical data was changed using the label encoding procedure. Within the object data type columns, a distinct integer value was assigned to every distinct category. Adding categorical variables to the ML models so that the algorithms could analyse and comprehend the data was made simpler by this change.

Managing missing and negative values:

To maintain the integrity and dependability of the dataset, both missing and negative values were methodically treated.

Step 3: Data profiling:

Data profiling will involve using Python's Pandas package to assess the employee performance dataset's basic statistics. One can learn more about a number of things by conducting data profiling, including the variables present in the dataset .

Step 4: feature selection:

A crucial step in identifying the most significant and pertinent elements influencing worker performance is feature selection. To identify the most important variables for prediction, techniques Feature selection ensures the model only uses the most relevant variables, reducing complexity and improving accuracy.

Methods used: Random Forest classifier & regression– Identifies which independent variables (Leadership, Training, etc.) have the highest impact on Target Completion and raking the departments based on their performance.

ANOVA (Analysis of Variance) – Tests if different departments have significantly different performance levels.

Step 5: Model selection:

It is possible to predict employee performance using a variety of machine learning methods. The type of problem and the information at hand determine which model is best.

Random Forest Regression – Used to predict Target Completion and Engagement & Motivation.

Random Forest Classifier – Used to rank departments based on performance levels.

ANOVA & Tukey HSD – Used to check whether departments significantly differ and by how much

Step 6: training and testing data:

The dataset should be divided into testing and training data in order to assess the effectiveness of the selected machine learning framework. The testing data is used to assess the model's capacity to generalize and predict employee performance precisely, while the training data is used to train and equip the framework on historical data. Both sets should have a sufficient amount of data in order to prevent the model from being overfit or underfit.

The dataset is split into **training (80%) and testing (20%)** sets:

- **Training Set** – Used to train the model on historical data.
- **Testing Set** – Used to assess how well the model predicts employee performance.

Step 7: Model Optimization:

The goal of model optimization is to improve performance by fine-tuning the hyperparameters of the selected machine learning algorithms. To investigate a variety of potential parameter configurations and select the best configuration, techniques like grid search, cross-validation, and hyperparameter optimization algorithms can be used. Reducing potential biases and raising the model's overall accuracy in forecasting employee performance are the objectives.

Step 8: Business Intelligence & Visualization (Power BI):

After model predictions and department rankings, Power BI is used to visualize and interpret the results:

Interactive Dashboards – Display performance trends, feature importance, and department rankings.

Step 9: Control and monitoring:

A predictive framework that has been developed and refined can be used to predict new hire performance in a practical setting. Maintaining accuracy and dependability requires constant monitoring and assessment of the framework's performance in comparison to actual employee performance data. Consistent refinement and updates may be essential to adjust the framework to evolving business requirements and changing organizational dynamics

1.2.Introduction to United Techno info systems:

United Techno is driven by the collective power of collaboration and innovation. The company envisions bringing together customers and partners to forge powerful connections, tackling challenges head-on and creating transformative solutions.

Starting from humble beginnings in a small office with just five dedicated individuals, United Techno initially served a single retail client, offering AS400 legacy technology services. Through unwavering commitment and a strong work ethic, the company has grown to a global presence, spanning five countries and employing over 500 people. A testament to their exceptional service and loyalty, the company's first client continues to be their top revenue contributor, even after 15 years.

United Techno's journey has evolved beyond legacy systems to expand its expertise into data engineering, analytics, cloud integration, and automation. The company prides itself on being employee-centric, with customer success always at the forefront. United Techno isn't just a business—it's a united family, embodying simplicity, reliability, and a steadfast commitment to delivering powerful solutions for their customers.

United techno Service Offerings:

- Enterprise Application Development & Maintenance
- Implementation Enterprise Systems (Salesforce, Blue Yonder, SAP)
- Quality Assurance Engineering & Automation
- Cloud Integration
- Data Engineering & Analytics
- Regulatory Compliance & Audits
- Legacy Modernization
- Managed Services

Industries they specialize:

- Life Sciences
- Retail
- Logistics
- Manufacturing
- Technology

UT specialize in :

- Salesforce - Sales Cloud, Service Cloud, Marketing Cloud, Finance Cloud, Health Cloud, Data Cloud, Tableau
- Integration - Boomi, MuleSoft, Celego, Workato, Informatica, Microsoft Logi Capp/ADF
- Data – Snowflake, Databricks, Microsoft Fabric, Matillion, Informatica, DBT, Nexla
- Analytics – Gen AI, Open AI, LLAMA, MicroStrategy, PowerBI, Tableau Cloud – Azure, AWS, GCP

Website	:	https://unitedtechno.com/
Industry	:	IT Services and IT Consulting
Company size	:	201-500 employees
Headquarters	:	Pleasanton, California
Type	:	Privately Held
Founded	:	2011

1.3: Scope of united techno info systems:

United Techno enables businesses to prosper in the digital age. Customers across a variety of business verticals, including retail, logistics & supply chain, life sciences, healthcare, and more, have benefited greatly from our extensive range of data management and artificial intelligence solutions.

Application support &Maintenance:

In order to improve applications and guarantee efficiency, security, and ongoing development, modern enterprises use digital transformation, AI, and ML. While automated performance reviews improve labour management, generative AI-driven development spurs creativity.

Enterprise applications:

Using BI and data analytics, enterprise solutions optimize Retail & Warehouse ERP, boost consumer interaction, and increase supply chain visibility, guaranteeing smooth operations and prompt decision-making.

Cloud solutions:

Cloud services integrate EDI, automation, and security to improve productivity while offering managed, advising, and consultancy services. Value-added services maximize enterprise performance, risk management, and technology adoption.

Quality engineering & automation:

Software dependability is guaranteed via automated testing, security, and compliance. While QA consulting promotes performance enhancements, Continuous Integration & Delivery (CI/CD) permits quicker releases.

Data quality:

Reliable analytics are ensured by upholding data completeness, integrity, and validation. Decision-making is enhanced by profiling, cleaning, and audits, and actionable insights are guaranteed by dashboard validation.

1.4. Global &Indian scenario with statistics:

United Techno is ideally situated to take advantage of the expanding need for data solutions and digital transformation in both the Indian and international markets. The demand for United Techno's services is driven by the quick adoption of cloud technologies, data governance, and process optimization in industries including manufacturing, finance, and e-commerce in India. Businesses throughout the world are putting more emphasis on data security, cloud integration, and artificial intelligence (AI), which makes United Techno a major player in providing specialized solutions to large enterprises. United Techno assists clients in overcoming difficult obstacles and achieving long-term success by fusing knowledge of regional market dynamics with international best practices.

The following statistics highlights key growth trends and projections across various global industries, driven by digital transformation and innovation. These sectors, including retail, insurance, oil & gas, travel, logistics, and e-commerce, are all experiencing significant market expansion.

Retail: With a predicted compound annual growth rate (CAGR) of 13.0% from 2021 to 2028, the global retail digital transformation market is estimated to be worth US$5,662.6 million in 2021.

Insurance: At a compound annual growth rate (CAGR) of 9.6%, the specialty insurance market is projected to reach US$130.1 billion by 2027 from US$81.5 billion in 2022

Oil and Gas: It is anticipated that the global upstream sector will continue to invest US$580 billion in hydrocarbons in 2023 and produce more than US$800 billion in free cash flows in 2024.

Logistics: At a compound annual growth rate (CAGR) of 4.4%, the worldwide freight and logistics market is projected to reach US$18.69 billion by 2026.

E-commerce: By the end of 2023, it is anticipated that online retail sales would total US$6.51 trillion, or 22.3% of all retail sales.

Travel & tourist: With a projected market value of US$1,016.00 billion and 1,333 million hotel customers, internet sales are predicted to account for 74% of the worldwide travel & tourism industry's total income by 2027.

Customers:

United Techno has the expertise, resources, and experience necessary to manage numerous integration projects worldwide as a certified cloud integration service partner. Complete full-cycle services are offered, ranging from managed services and support to architecture and design.

Ipass (integration platform as a service) partners:

1.boomi

2.celigo

3.workato

4.mulesoft

5.Azure logic apps

Cloud partners:

1.aws

2.azure

3.google cloud

1.5: Swot analysis:

Strengths: United Techno has outstanding R&D, cutting-edge product offerings, and sophisticated technological capabilities.

Weaknesses: Dependency on particular suppliers and market diversification provide difficulties for the organization.

Opportunities: There is a lot of room for expansion by entering emerging markets and creating sustainable solutions.

Threats: Risks include fierce competition, economic downturns, and quick technical advancements.

1.6. Trends and future:

With an emphasis on growing its proficiency in automation, data engineering, and cloud integration, United Techno is well-positioned for future expansion and innovation. The business will probably raise its investment in AI, machine learning, and next-generation solutions while growing its global footprint as the need for cutting-edge technologies rises. In order to remain competitive in the ever-changing IT sector, United Techno will continue to nurture its employee-centric culture and cultivate a creative, cooperative workforce while upholding its customer-centric approach by providing customized, innovative solutions to satisfy changing market needs.

1.7. Study on company profile:

United Techno is expected to grow by investing more in AI and machine learning and broadening its knowledge of automation, data engineering, and cloud integration. In order to satisfy changing market expectations, the company will prioritize worldwide expansion, cultivating a creative staff, and upholding a customer-centric strategy with customized, innovative solutions.

Fundamental principle of united techno:

Employee-Centric: To prevent any stakeholder from making unwarranted assumptions, we have set up clear lines of communication.

Customer Success: Our system for providing constructive criticism consistently helps our team members develop.

Dependability and Quality: A committed effort from every team member.

Global Excellence: No novel concept is rejected without first being tested.

Cooperation: Members of a team work together to accomplish projects.

Mission:

We innovate because we have a common goal and dedication! The best industry talent has been consistently attracted, inspired, developed, acknowledged, and rewarded thanks to it. Our people are inspired to present themselves in the best possible light by the atmosphere it fosters. In an environment that fosters equality of opportunity and a feeling of community, team members flourish and progressively establish prosperous careers.

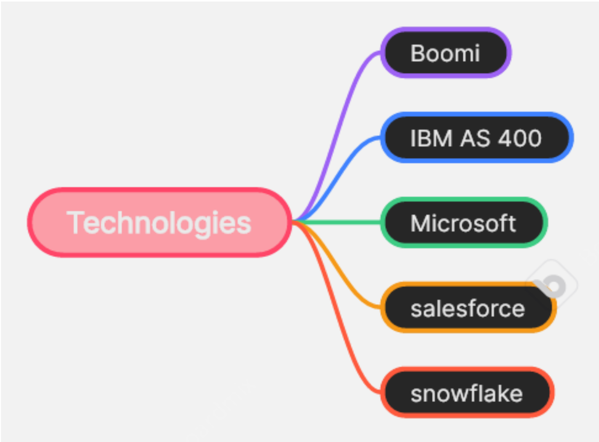
Leadership team of UT:

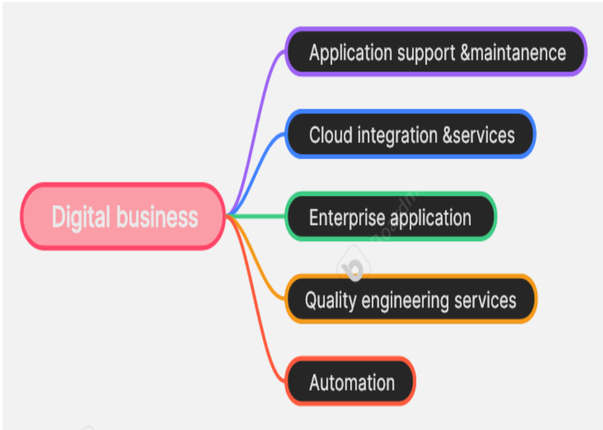
- Moorthy Subbiah - Founder & CEO
- Vijayalakshmi Devarajan- Senior Vice President (Human Resources & Administration)
- Mohan Boopalan-Global Practice Head (Cloud Integration)
- Gurusivaraman B-Associate Vice President – Global Delivery
- Ajay Natarajan-Vice President (Architecture & Integration Solutions)
- Fiona Fischer-Managing Director (United Techno Australia)
- Goutham Reddy Katangur Director – Delivery (Cloud Integration)

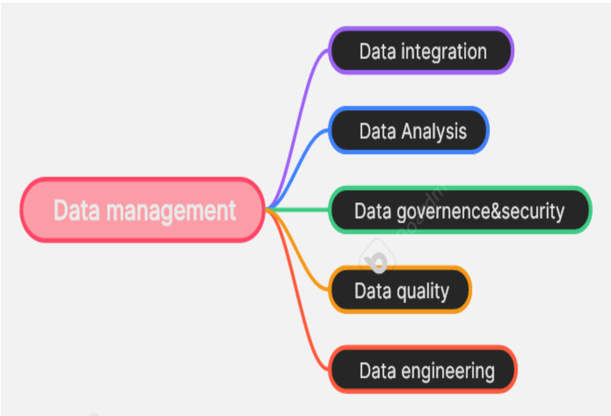
Locations:

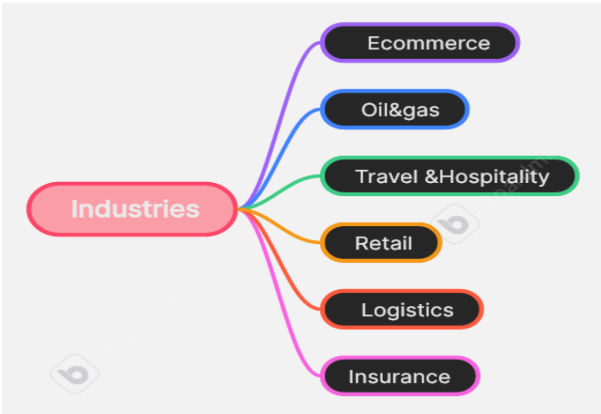
- United states
- India
- Australia
- Singapore
- Canada

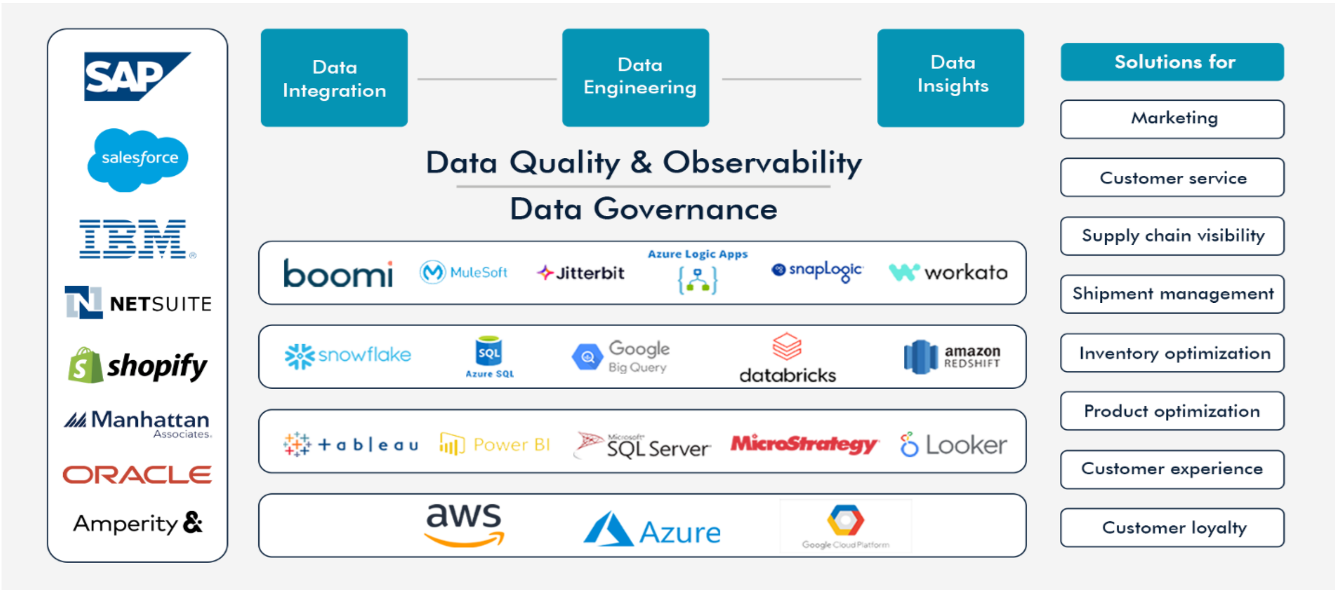
1.8: United techno services:











1.09. Problem statement:

- Employees are essential to an organization's success, yet it can be difficult to measure performance and resource usage. Conventional performance reviews have the potential to be uneven and subjective, which can result in ineffective resource allocation and lost chances for staff development. Businesses spend money on leadership development, training, and workplace supplies, but often find it difficult to assess if these expenditures actually improve performance in the absence of concrete insights.

- In order to assess how workers contribute to the company and how well they use the resources at their disposal, this research intends to create a data-driven methodology utilizing business analytics, machine learning.

1.10. Objective of the study:

- Determine the main elements that affect employee performance, such as training, leadership, and working circumstances.
- Evaluate how successfully employees make use of company resources, such as leadership assistance and training initiatives.
- Create a framework for performance analysis by utilizing machine learning and business analytics to evaluate departmental performance differences
- Implement power BI dashboards to visualize performance insights and aiding decision making

1.11. Scope of the study:

- focuses on how firms use their resources and how well their employees perform.
- employs Power BI, machine learning, and business analytics to provide data-driven insights.
- evaluates performance patterns and forecasts future worker output.
- aids businesses in making better decisions and increasing worker productivity**.**

CHAPTER II

REVIEW OF LITERATURE

2.0:Review of literature:

1. Hasan, MD Rokibul, et al. “Employee Performance Prediction: An Integrated Approach of Business Analytics and Machine Learning.” Journal of Business and Management Studies, vol. 6, no. 1, Al-Kindi Centre for Research and Development, Feb. 2024

2. John, Shylu, et al. “An Integrated Approach to Renew Software Contract Using Machine Learning.” Journal of Business Analytics, vol. 4, no. 1, Informa UK Limited, Dec. 2020, pp. 14 25.

3. Ujlayan, Anshul, and Manisha Sharma. “An Analysis of Employee Skills and Potency Using Machine Learning.” International Journal of Business and Data Analytics, vol. 2, no. 1, Inder science Publishers, 2023.

4. Tanasescu, L.G.; Vines, A.; Bologa, A.R.; Vîrgolici, O. Data Analytics for Optimizing and Predicting Employee Performance. *Appl. Sci.* 2024.

5. Peter, G.S., Amit, C.B., Deokar, V., Patel, N.R. (2023). Machine Learning for Business Analytics: Concepts, Techniques and Applications in RapidMiner. New York: Wiley

6. Suri, N., & Lakhanpal, P. (2024). People Analytics Enabling HR Strategic Partnership: A Review. South Asian Journal of Human Resources Management,

7. Utomo, W., Suryono, W., Jimmi, J., Santosa, T. A., & Agustina, I. (2023). Effect of STEAM-Based Hybrid Based Learning Model on Students’ Critical Thinking Skills. Jurnal Penelitian Pendidikan IPA, 9(9), 742–750.

8. Chen, Y.; Lin, Z. Business Intelligence Capabilities and Firm Performance: A Study in China. *Int. J. Inf. Manag.* **2021**, *57*, 102232

9. Avrahami, D., Pessach, D., Singer, G. & Ben-Gal, H.C. (2022) A human resources analytics and machine-learning examination of turnover: implications for theory and practice. International Journal of Manpower

10. The Role of Machine Learning Algorithms in Business Intelligence: Transforming Data into Strategic Insights(2024)

11. Conboy K, Mikalef P, Dennehy D, Krogstie J. Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. European Journal of Operational Research. 2020 Mar 16;281(3):656-72.

12. Watson, J., Hutyra, C. A., Clancy, S. M., Chandiramani, A., Bedoya, A., Ilangovan, K., . . . Poon, E. G. (2020). Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? JAMIA open, 3(2), 167-172.

13. Rahman, A. (2023). AI revolution: shaping industries through artificial intelligence and machine learning. Journal Environmental Sciences and Technology, 2(1), 93-105. [63] Sarker, I. H. (2021). Data science and analytics: an overview

14. Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H. C., Shmueli, E., & Ben-Gal, I. (2020). Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. Decision Support Systems, 134, 113290.

15. Nowicka, J., Pauliuchuk, Y., Ciekanowski, Z., Fałda, B., & Sikora, K. (2024). The use of data analytics in human resource management. European Research Studies Journal, 27(2), 203-215.

CHAPTER III

RESEARCH METHODOLOGY

3.0: Research Methodology:

3.1: Research Design:

This study follows the descriptive and predictive research design. the descriptive study focuses on analyzing employee performance based on certain factors.

The predictive aspect utilizes machine learning techniques (Random Forest classifier, Random Forest regression)

3.2: Type of Research study:

Causal-Comparative Study:

Anaysing the impact of independent variables on dependent variable

Uses ANOVA and Tukey HSD to determine is there are any significance differences between departments.

Predictive analytics study:

Applies random forest classifier to rank departments based on their performance.

Uses Random Forest regression to determine the impact of independent variables.

3.3. Sampling Methodology:

3.3.1. Population:

The population of this study consists of all employees working at United Techno, across various departments and job roles.

3.3.2. Frame:

The employees of united techno who are participated in the questionnaire survey

3.3.3. Method:

Convenience sampling method was used in the study, participants were chosen based on their availability and willingness to complete United Techno's Employee Performance Analytics Survey.

3.3.4. Sample Size:

To determine the appropriate sample size Slovin's formula used:

Formula:

$n = N / (1 + Ne^2)$

where:

- n = sample size
- N = population size (110 employees)
- e = margin of error (0.05)

Calculation:

$n=110 / 1+(0.05)^2$

$n=110/1+(0.0025)$

$n=110/1.275$

n~86

The minimum requires sample is 86 employees, to improve the reliability of the study 93 employee's responses was used.

3.4: Data Collection methodology:

3.4.1: Data collection method:

The primary data collection method utilizes in this study. Employee responses at United Techno were gathered for the study using a survey-based data gathering method. A standardized questionnaire was used to conduct the survey. Relevant existing research papers reviewed to support study findings.

3.4.2: Sources of Data:

Primary data source:

Primary data was collected directly from employees through the structured questionnaire created on google form.

Secondary data source:

To offer a theoretical framework and support the study's conclusions, pertinent research papers, journals, and articles about business analytics, employee performance analysis, and machine learning in HR analytics were examined

3.4.3.Data collection instrument:

The structured questionnaire was used the data collection. This questionnaire consists of multiple choice and Likert scale questions, the questionnaire was designed to assess the employee's perception on various factors that influencing employee performance (ex: Working conditions, leadership, Training opportunities) the responses were later numerically coded for quantitative analysis.

3.5: Reliability and Validity Analysis:

3.5.1: Reliability Analysis (Cronbach's Alpha)

Reliability refers to the consistency of a measurement instrument. A commonly used measure for reliability in survey research is Cronbach Alpha which evaluates the internal consistency of a questionnaire item.

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_T^2}\right)$$

n = number of items

σ_i^2=Variance of each individual item

σ_T^2= Variance of the total score

Source code for Cronbach's Alpha:

```
import pandas as pd
import NumPy as np
df= pd. read_excel("reliability.xlsx")
def cronbach_alpha(df):
    items_scores=df.values.T
    item_vars=items_scores.var(axis=1,ddof=1)
```

```
 total_var=items_scores.sum(axis=0).var(ddof=1)
 n_items=items_scores.shape[0]
return(n_items/(n_items-1))*(1-sum(item_vars)/total_var)
alpha= cronbach_alpha(df)
print (f ”Cronbach’s Alpha:{alpha:.4f}")
if alpha>0.07:
   print ("good reliability")
elif  0.6<=alpha<0.7:
   print ("acceptable reliability")
else:
   print ("Low reliability")
```

Output:

Cronbach's Alpha:0.7137

good reliability

Reliability Interpretation:

Table 3.5.1

Cronbach’s Alpha (α)	Reliability Interpretation
α >=0.7	Good Reliability
0.6 <=α<0.7	Acceptable reliability
α < 0.6	Low reliability

Since α=0.7137, this indicates that the questionnaire items are internally consistent and reliable for measuring employee factors.

3.5.2: Validity Analysis:

Exploratory Factor Analysis (EFA) was performed to uncover the underlying structure of the survey variables and to group related items into common factors.

To assess whether the dataset was suitable for factor analysis, Kaiser-Meyer-Olkin (KMO) Test and Bartlett's Test of Sphericity were conducted. A KMO value above 0.6 and a significant Bartlett's test ($p < 0.05$) confirmed the adequacy of the sample and the interrelationships among variables.

Source code :

```
import pandas as pd
from factor_analyzer import FactorAnalyzer
from factor_analyzer.factor_analyzer import calculate_kmo, calculate_bartlett_sphericity
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_excel("reliability.xlsx")
# Drop rows with missing values if any
df.dropna(inplace=True)
# Step 3: KMO and Bartlett's Test
kmo_all, kmo_model = calculate_kmo(df)
chi_square_value, p_value = calculate_bartlett_sphericity(df)
print(f"\nKMO Test Score: {kmo_model:.4f}")
print(f"Bartlett's Test: Chi-Square = {chi_square_value:.2f}, p-value = {p_value:.4f}")
if kmo_model >= 0.6 and p_value < 0.05:
    print("Data is suitable for Factor Analysis.")
else:
    print("Data may not be suitable for Factor Analysis")
```

output:

KMO Test Score: 0.6569

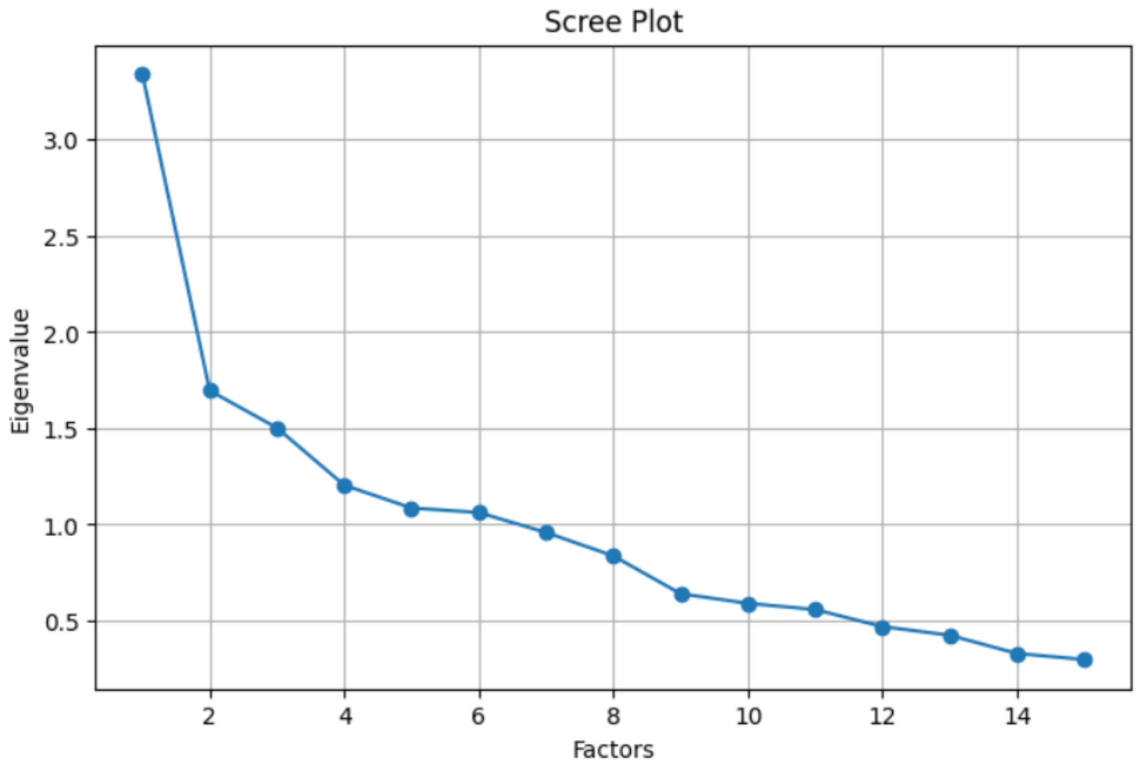
Bartlett's Test: Chi-Square = 278.71, p-value = 0.0000

Data is suitable for Factor Analysis.

```python
# Step 4: Exploratory Factor Analysis (EFA)
# First check eigenvalues to determine number of factors
fa = FactorAnalyzer(n_factors=df.shape[1], rotation=None)
fa.fit(df)
ev, v = fa.get_eigenvalues()

# Plot eigenvalues to see how many factors to keep (Scree plot)
plt.figure(figsize=(8, 5))
plt.plot(range(1, len(ev)+1), ev, marker='o')
plt.title("Scree Plot")
plt.xlabel("Factors")
plt.ylabel("Eigenvalue")
plt.grid()
plt.show()

# Keep factors with eigenvalue > 1
n_factors = sum(ev > 1)
print(f"\nNumber of factors with eigenvalue > 1: {n_factors}")

# Step 5: Run Factor Analysis with the chosen number of factors
fa = FactorAnalyzer(n_factors=n_factors, rotation='varimax')
fa.fit(df)

# Get and display factor loadings
loadings = pd.DataFrame(fa.loadings_, index=df.columns)
print("\nFactor Loadings:\n")
print(loadings.round(2))
```

output:

figure 3.5.1:



Number of factors with eigenvalue > 1: 6

Factor Loadings:

Table 3.5.2:

	0	1	2	3	4	5
Working condition	0.82	0.28	-0.19	0.12	0.16	-0.07
T.Working condition	0.17	0.53	0.37	-0.01	-0.26	-0.00
C.Leadership	0.13	0.05	-0.05	0.17	0.60	0.08
E.Leadership	0.13	0.18	0.04	0.47	0.21	-0.14
career development	0.31	0.36	0.34	-0.02	0.57	0.09
co-workers	-0.03	0.52	0.03	0.13	0.11	-0.01
Training	0.12	0.41	0.21	0.24	0.15	-0.10
Learning	0.05	0.14	0.07	0.78	0.01	0.24
problem solving	0.65	-0.00	0.35	0.26	0.21	0.05
Learning&dev	0.01	0.22	0.39	0.12	0.05	0.00
conflicts	0.24	0.37	-0.16	0.16	0.27	0.28

d.individual contribution	0.49	-0.10	0.11	-0.09	0.02	0.44
d.target completion	0.05	0.03	0.44	0.08	-0.23	0.03
d.quality of work	-0.00	-0.01	0.08	0.08	0.07	0.66
d.engagement&motivation	-0.01	0.02	0.52	-0.05	0.12	0.07

Interpretation:

Factor 1: Contribution & Working Conditions

highly correlated with individual contribution (0.49), problem solving (0.65), and working conditions (0.82). This component shows how employees' sense of contribution is influenced by their ability to solve problems and have good working conditions.

Factor 2: Team Support & Development

High loadings on T. Conflicts (0.37), Training (0.41), Coworkers (0.52), and Working Conditions (0.53). This element highlights how crucial team chemistry and support are to employee growth.

Factor 3: Performance & Motivation

most impacted by Learning & Development (0.39), Target Completion (0.44), and Engagement & Motivation (0.52), demonstrating the connection between learning opportunities, motivation, and goal achievement.

Factor 4: Learning Orientation & Leadership

Learning Orientation & Leadership, which is dominated by learning (0.78) and E. Leadership (0.47).

Factor 5: Communicative Leadership & Career Growth

Open leadership improves career support, as seen by the highest loadings on C. Leadership (0.60) and Career Development (0.57).

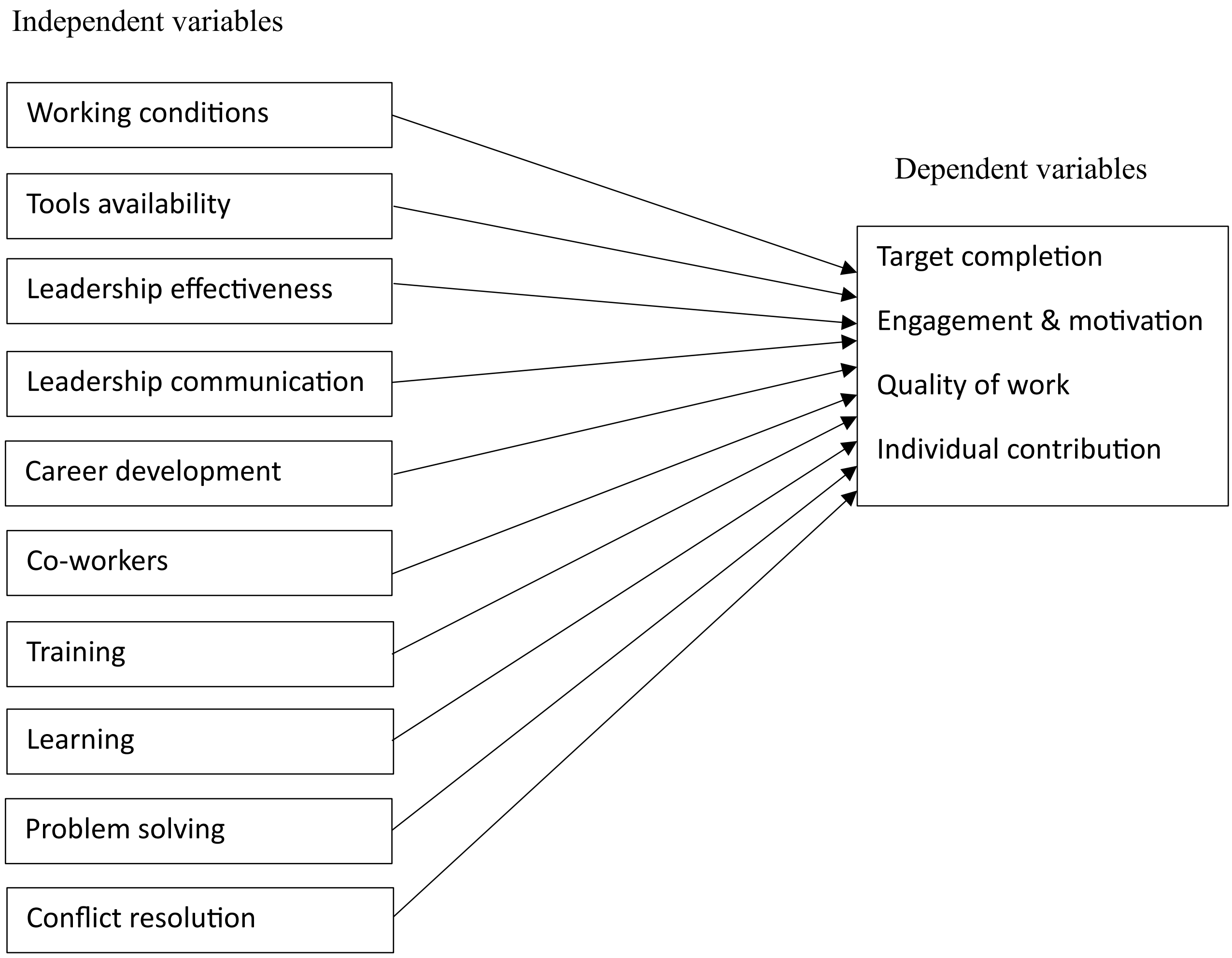
Factor 6: Quality & Recognition

defined by Individual Contribution (0.44) and Quality of Work (0.66), demonstrating the connection between contribution recognition and the caliber of work output.

Six distinct factors that grouped together related survey variables were identified by the factor analysis, suggesting a robust underlying structure in the data. These elements include: Quality & Recognition; Learning Orientation & Leadership; Communicative Leadership & Career Growth; Team Support & Development; Performance & Motivation; and Work Conditions & Contribution. The survey items successfully capture different but significant aspects of employee perception and performance, as evidenced by the strong loadings for each factor for pertinent variables.

Based on these statistical tests and the structure of the factor loadings, it is concluded that the questionnaire is both reliable and valid, making it suitable for further analysis in this research.

3.6. Independent and dependent variables:



3.7. standard deviation value:

3.7.1.Source code ;

```
import pandas as pd
import os
os.chdir(r"C:\Users\Sahana Dhandapani\OneDrive\Documents")
data=pd.read_excel("coded responses.xlsx")
columns = ['Working condition', 'T.Working condition', 'C.Leadership',
           'E.Leadership', 'career development', 'co-workers', 'Training',
           'Learning', 'problem solving', 'Learning&dev', 'conflicts',
           'd.individual contribution', 'd.target completion', 'd.quality of work',
           'd.engagement&motivation']
for col in columns:
    std_dev = data[col].std(ddof=0)
    mean_val = data[col].mean()
    print(f"{col:30} | Mean: {mean_val:.2f} | Std Dev: {std_dev:.2f}")
```

output:

Figure 3.7.1

```
Working condition              | Mean: 3.77 | Std Dev: 0.44
T.Working condition            | Mean: 3.67 | Std Dev: 0.49
C.Leadership                   | Mean: 3.67 | Std Dev: 0.47
E.Leadership                   | Mean: 3.83 | Std Dev: 0.38
career development             | Mean: 3.75 | Std Dev: 0.43
co-workers                     | Mean: 3.86 | Std Dev: 0.35
Training                       | Mean: 3.85 | Std Dev: 0.36
Learning                       | Mean: 3.77 | Std Dev: 0.42
problem solving                | Mean: 3.76 | Std Dev: 0.47
Learning&dev                   | Mean: 3.72 | Std Dev: 0.45
conflicts                      | Mean: 3.76 | Std Dev: 0.45
d.individual contribution      | Mean: 3.72 | Std Dev: 0.47
d.target completion            | Mean: 3.65 | Std Dev: 0.54
d.quality of work              | Mean: 3.72 | Std Dev: 0.47
d.engagement&motivation        | Mean: 3.70 | Std Dev: 0.50
```

3.8. Hypotheses of the Study:

Null Hypothesis (H_0):

There is no significant relationship between organizational factors (working conditions, leadership, etc.) and employee performance metrics (target completion, quality of work, engagement & motivation).

Alternative Hypothesis (H_1):

There is a significant relationship between organizational factors and employee performance metrics

3.9.statistical techniques:

- Descriptive statistics -mean, standard deviation
- Random forest regression
- Random forest classifier
- ANOVA(analysis of variance)
- Tukey HSD
- k-means clustering
- sentiment analysis

3.10.statistical package used:

1.Python 3.13 (64-bit) (Jupyter Notebook)

- **Libraries:**

  pandas, NumPy, scikit-learn, stats models, seaborn, matplotlib

2.Power BI for data visualization and dashboard creation

3.11. limitations of study:

1.Since United Techno was the only company included in the study, the results may not be as applicable to other contexts or sectors.

2.The results' general relevance may be impacted if the sample size is not representative of the diversity of the total workforce.

3.Self-reported responses were used to gather the data, and these can be skewed by social desirability or personal bias.

4.Economic issues or personal struggles were not taken into account in the analysis of external influences on employee performance.

5.The research focused on selected organizational factors (e.g., working conditions, problem solving, training) and did not incorporate other potentially impactful variables such as compensation, managerial style, or job security.

CHAPTER IV

DATA ANALYSIS AND INTERPRETATION

4.0: Data Analysis and interpretation:

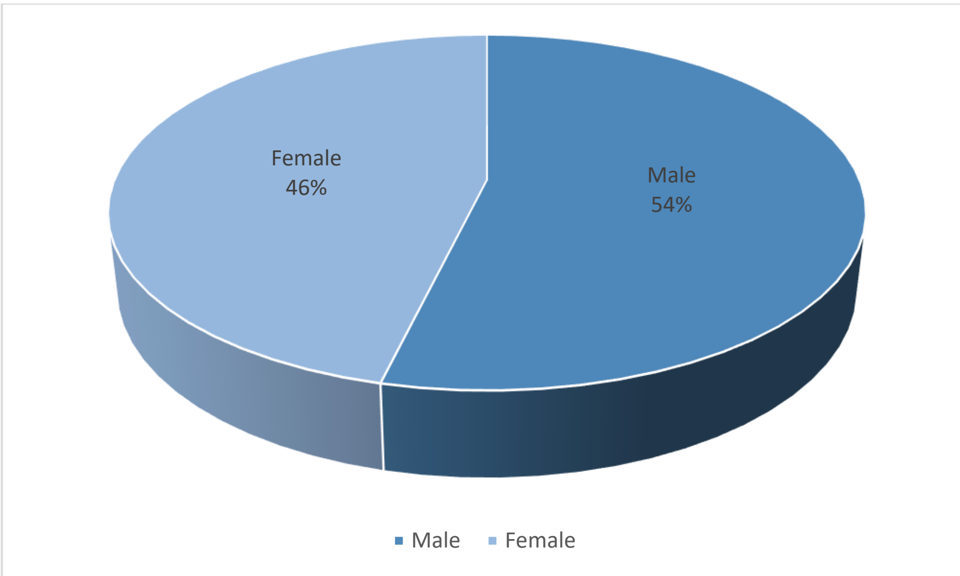
4.1: percentage Analysis:

1)gender

Table 4.1.1:

Gender	Count	Percentage
Male	50	46
Female	43	54
Total	93	100

Chart 4.1.1:



Interpretation:

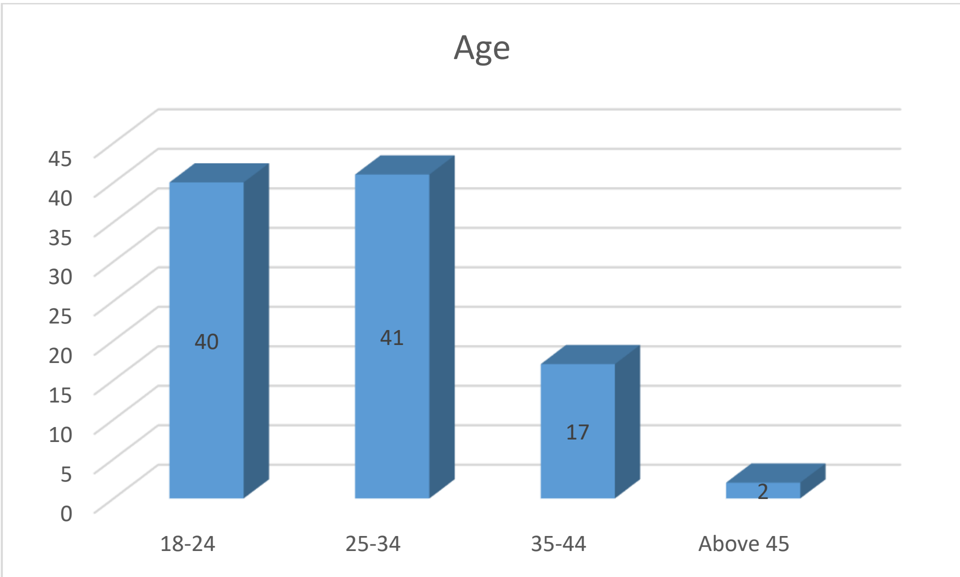
The above diagram represents the gender distribution of the survey respondents, with 54% males and 46% females.

2)Age

Table 4.1.2:

Age	Count	Percentage
18-24	37	40
25-34	38	41
35-44	16	17
Above 45	2	2
Total	93	100

Chart 4.1.2:



Interpretation:

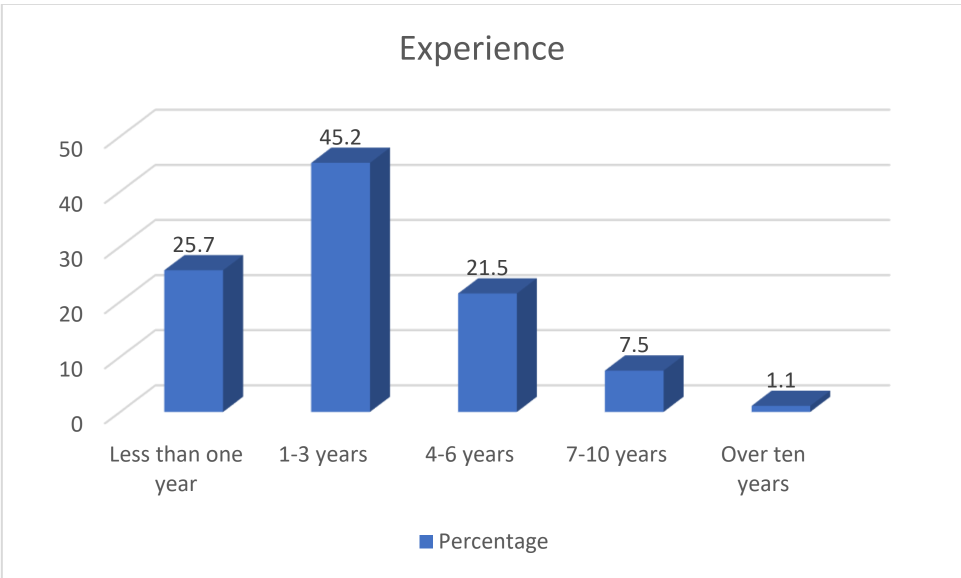
The above bar chart represents the age distribution of the employee survey respondents, indicating that the majority belong to the 18-34 age group (81%), followed by 35-44 (17%), and a smaller proportion of respondents are above 45 (2%), reflecting the overall demographic composition of the workforce in the study.

3)Experience:

Table 4.1.3

Experience	Count	Percentage
Less than one year	23	25.7
1-3 years	42	45.2
4-6 years	20	21.5
7-10 years	7	7.5
Over ten years	1	1.1
Total	93	100

Chart 4.1.3:



Interpretation:

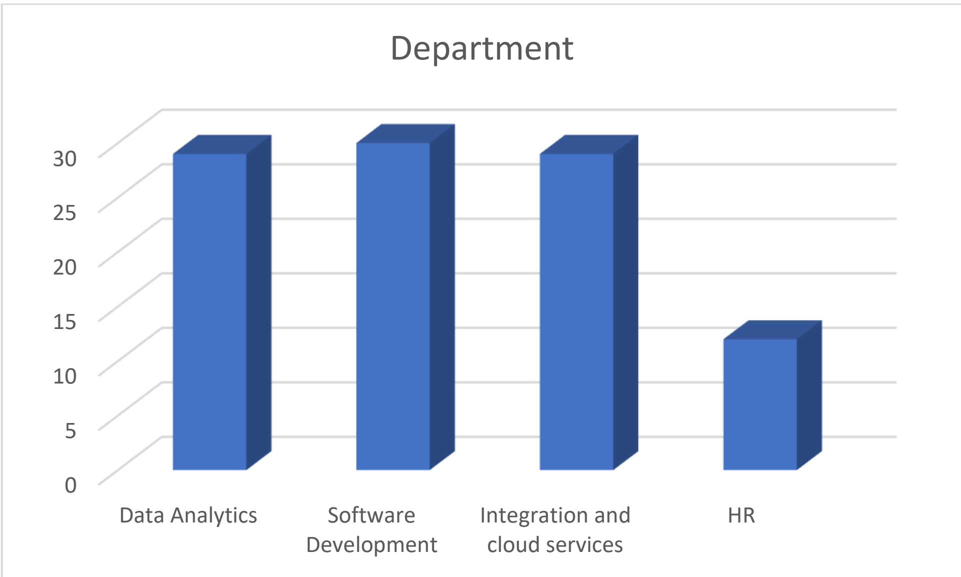
The above column chart represents the experience distribution of the 93 respondents, showing that most have 1-3 years of experience, followed by those with less than one year and 4-6 years, while a smaller number have 7-10 years and over ten years of experience, indicating a workforce with a majority of early-career professionals.

4) Department

Table 4.1.4:

Department	Count	Percentage
Data Analytics	27	29
Software Development	28	30
Integration and cloud services	27	29
HR	11	12
Total	93	100

Chart 4.1.4:



Interpretation :

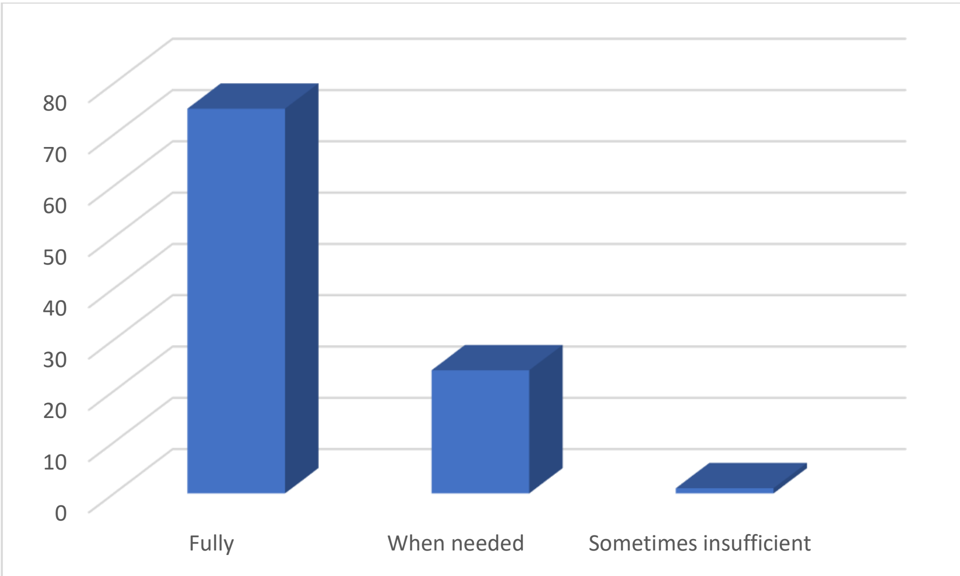
The above diagram shows a significant emphasis on technical roles is evident from the fact that the majority of respondents are employed in the fields of software development (30%), data analytics (29%), and integration & cloud services (29%). Comparing departments may be impacted by the HR department's lowest representation (12%). Results will mostly represent technical teams' experiences, therefore HR-related insights must be carefully interpreted.

5) To what extent do you use the tools and resources provided to complete your work?

Table 4.1.5:

Options	Count	Percentage
Fully	70	75
When needed	22	24
Sometimes insufficient	01	1
Total	93	100

Chart 4.5.1:



Interpretation:

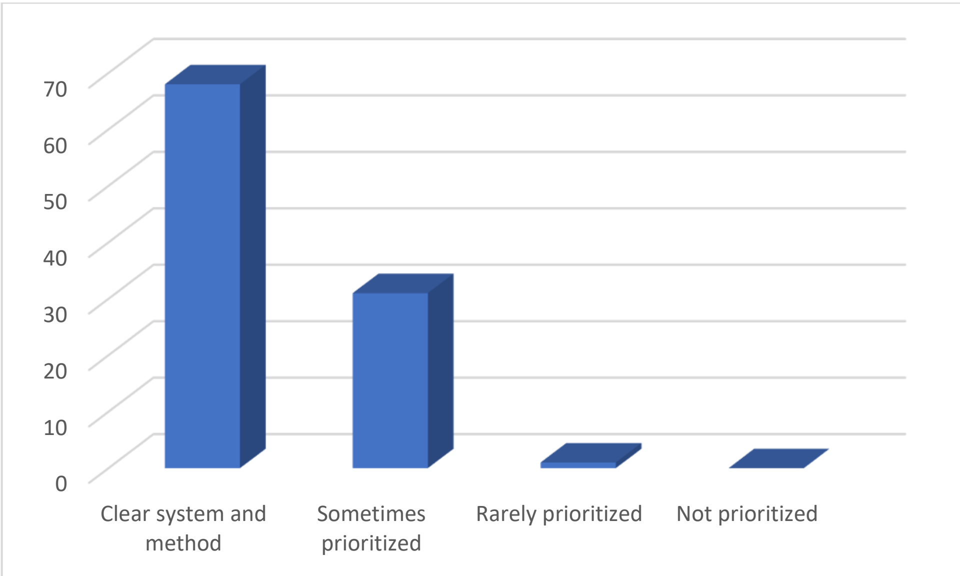
Strong organizational support is demonstrated by the majority of employees (75%) who believe that resources are entirely sufficient. 24% use resources only when necessary, indicating sporadic reliance on availability according on demand. Minimal resource restrictions are implied by the fact that just 1% claim deficiency. Although there are a few small areas for improvement, this shows a well-managed resource allocation system.

6) What is your approach to prioritizing tasks in your role?

Table 4.1.6

Options	Count	Percentage
Clear system and method	63	68
Sometimes prioritized	29	31
Rarely prioritized	01	1
Not prioritized	0	0
Total	93	100

Chart 4.1.6:



Interpretation:

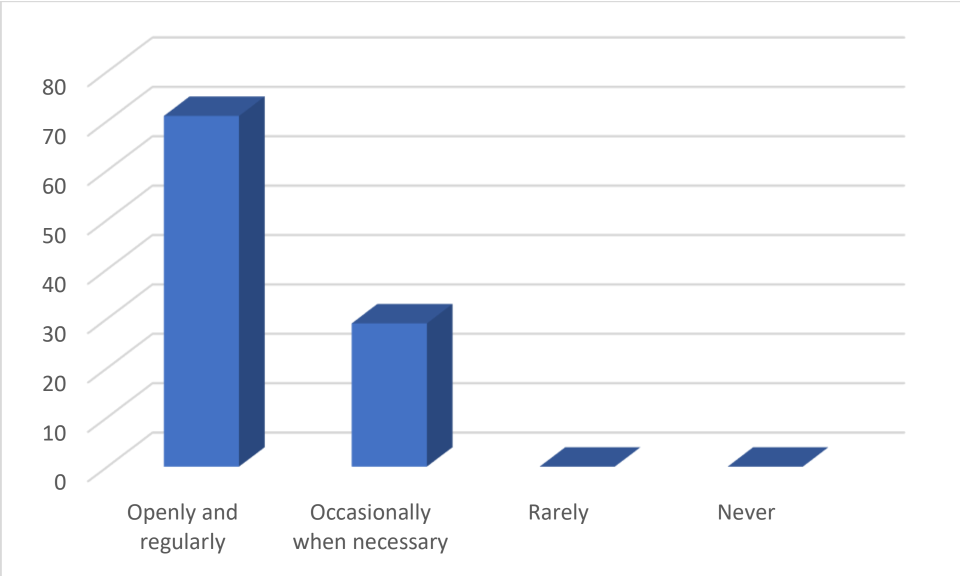
The above diagram shows 31% of respondents think prioritization only occurs occasionally, while 68% of respondents think there is a clear system and procedure in place. Just 1% of respondents said they rarely prioritize, and none said they never prioritize it. This implies that even though there are established procedures in place, they may be improved to guarantee uniform prioritizing in every situation.

7)To what extent do you communicate your needs and concerns to your manager?

Table 4.1.7:

Options	Count	Percentage
Openly and regularly	66	71
Occasionally when necessary	27	29
Rarely	0	0
Never	0	0
Total	93	100

Chart 4.1.7:



Interpretation :

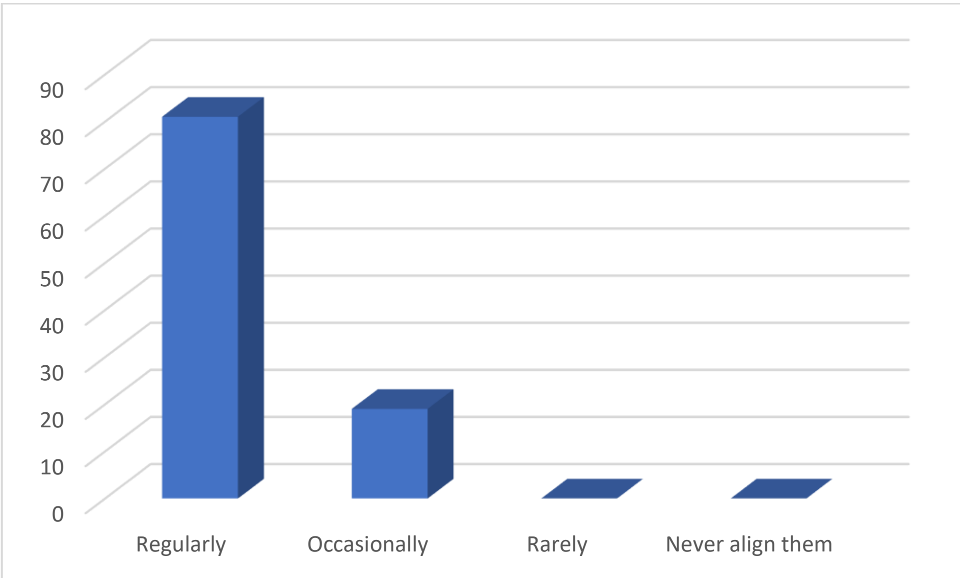
The above column chart shows 71% of respondents, according to the research, speak openly and frequently, whilst 29% only do so when required. Interestingly, none reported communication that was infrequent or non-existent. Although some employees might only participate, when necessary, this implies a strong culture of open communication and points to possible areas for development in proactive talks.

8) How do you ensure that your goals are in line with your manager's expectations?

Table 4.1.8

Options	Count	Percentage
Regularly	75	81
Occasionally	18	19
Rarely	0	0
Never align them	0	0
Total	93	100

Chart 4.1.8:



Interpretation:

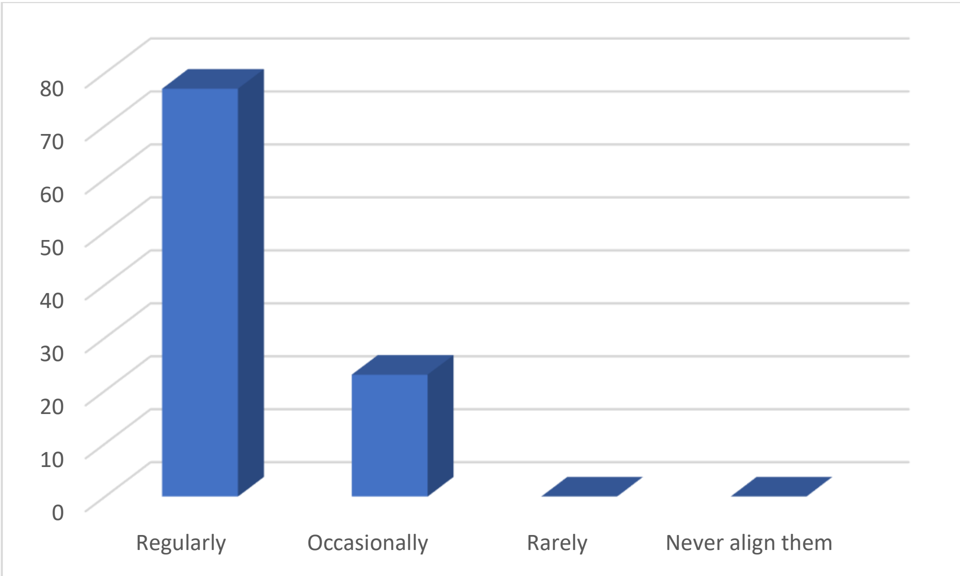
The above diagram shows 81% of respondents coordinate their tasks on a regular basis, compared to 19% who do it sometimes, according to the statistics. Interestingly, none said they aligned their tasks infrequently or never. Although some individuals could need more motivation to maintain consistency, this shows a solid adherence to structured work alignment.

9) How often do you align your current role with your long-term career goals?

Table 4.1.9:

Options	Count	Percentage
Regularly	72	77
Occasionally	21	23
Rarely	0	0
Never align them	0	0
Total	93	100

Chart 4.1.9:



Interpretation :

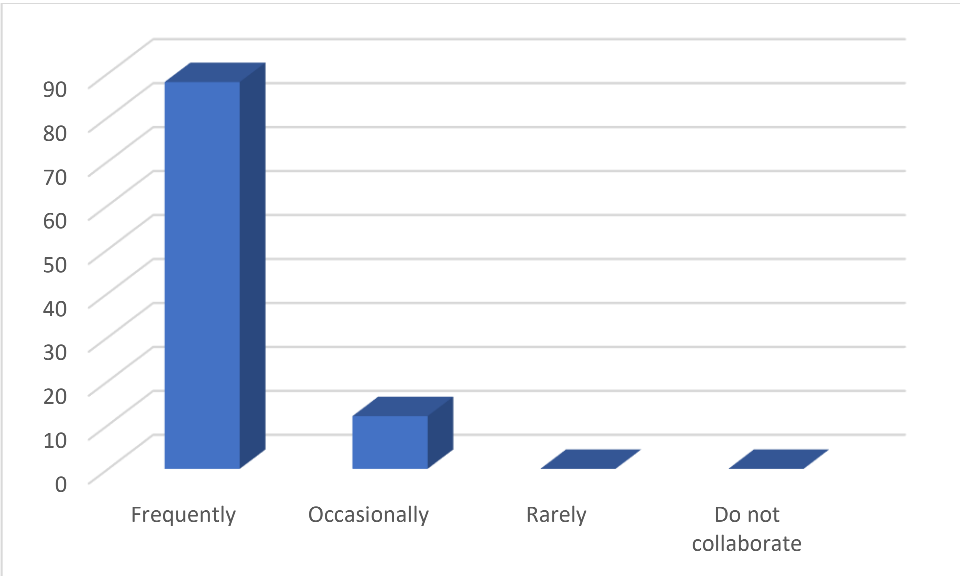
According to the above findings , 23% of respondents coordinate their duties occasionally, whilst 77% of respondents do so routinely. None of the respondents said they aligned duties infrequently or never. This points to a long-standing task alignment practice, albeit some staff members could need extra assistance to stay consistent.

10) How would you describe your collaboration with co-workers on team goals?

Table 4.1.10:

Options	Count	Percentage
Frequently	81	88
Occasionally	12	12
Rarely	0	0
Do not collaborate	0	0
Total	93	100

Chart 4.1.10



Interpretation :

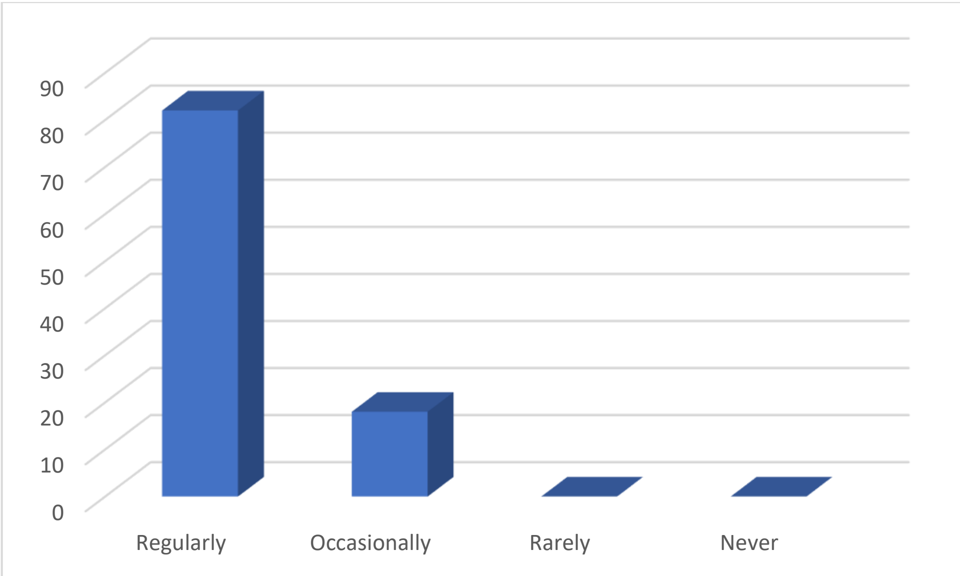
The above diagram shows , 12% of respondents only rarely collaborate, compared to 88% who do so regularly. None of the respondents said they collaborated infrequently or never. Although some staff might need more support or resources to collaborate more frequently, this shows a good culture of teamwork.

11) How often do you apply knowledge gained from training to your tasks?

Table 4.1.11

Options	Count	Percentage
Regularly	76	82
Occasionally	17	18
Rarely	0	0
Never	0	0
Total	93	100

Chart 4.1.11



Interpretation:

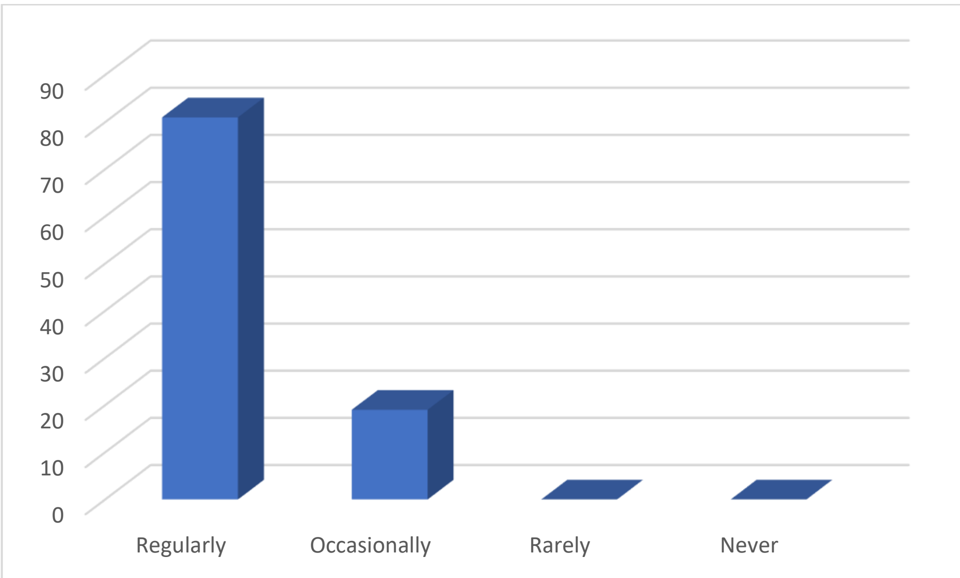
According to the diagram , 82% of respondents participate in the activity on a regular basis, while 18% do so infrequently. None of the respondents said they participated infrequently or never. This points to a long-standing practice, but there is room to promote more regular participation from those who only occasionally participate.

12) How frequently do you utilize available learning resources (e.g., workshops, courses)?

Table 4.1.12

Options	Count	Percentage
Regularly	75	81
Occasionally	18	19
Rarely	0	0
Never	0	0
Total	93	100

Chart 4.1.12:



Interpretation :

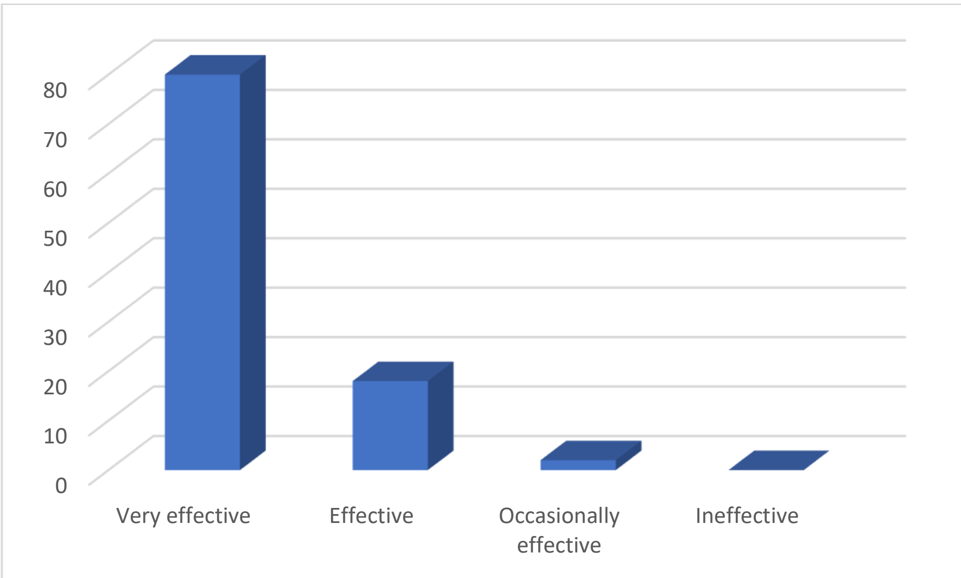
According to the data 19% of respondents only occasionally participate, whilst 81% do so on a regular basis. None of the respondents said they engaged infrequently or never. Although it might inspire sporadic people to participate more often, this shows a great dedication to the activity.

13) How effective are you in identifying and solving problems that arise within your team?

Table 4.1.13:

Options	Count	Percentage
Very effective	74	80
Effective	17	18
Occasionally effective	2	2
Ineffective	0	0
Total	93	100

Chart 4.1.13:



Interpretation :

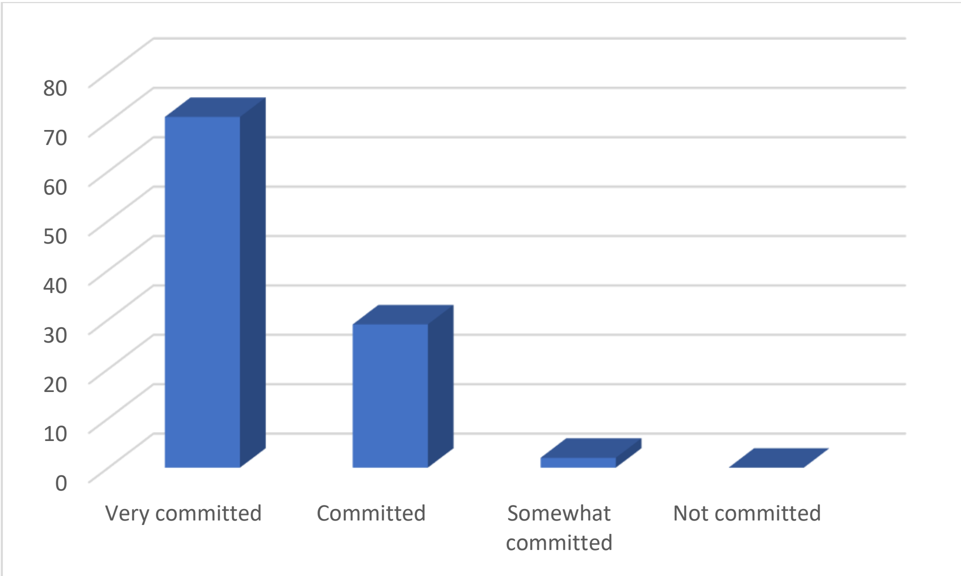
According to the data, 18% of respondents believe the method is effective, and 80% find it to be extremely effective. None regarded it as unsuccessful, while just 2% believe it is only sometimes effective. Although there are a few small areas for improvement to increase consistency, this indicates a high degree of confidence in the process.

14) How committed are you to learning new skills to improve your performance?

Table 4.1.14:

Options	Count	Percentage
Very committed	66	71
Committed	27	29
Somewhat committed	0	0
Not committed	0	0
Total	93	100

Chart 4.1.14:



Interpretation :

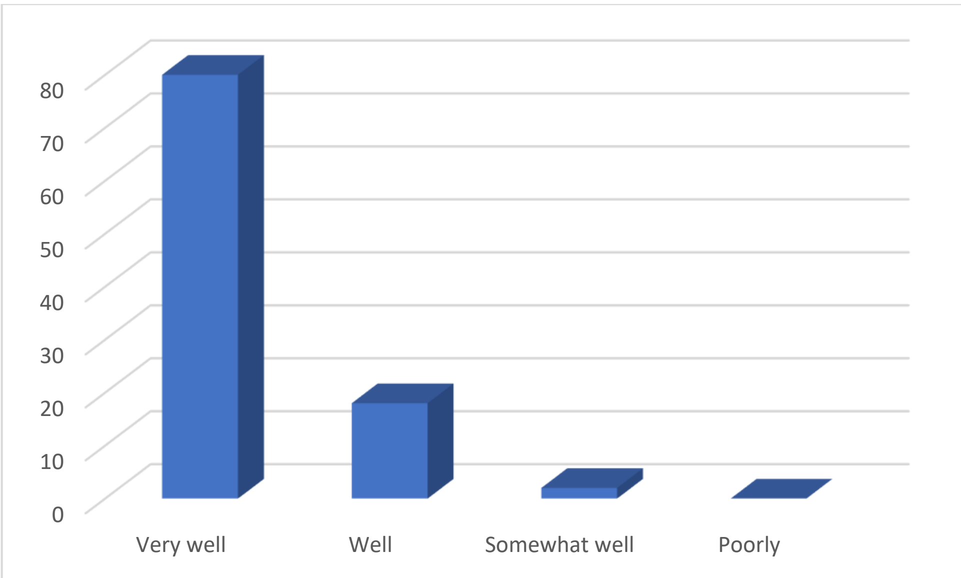
The above chart shows , 29% of employees are devoted, 71% are very committed, 2% are somewhat committed, and none are not committed. This suggests that United Techno employees are very dedicated. The lack of "Not Committed" answers points to a productive workplace. High levels of commitment can improve engagement, productivity, and the performance of the organization as a whole.

15) How well do you manage conflicts or disagreements within your team?

Table 4.1.15:

Options	No.of respondents	Percentage
Very well	74	80
Well	17	18
Somewhat well	02	2
Poorly	0	0
Total	93	100

Chart 4.1.15:



Interpretation :

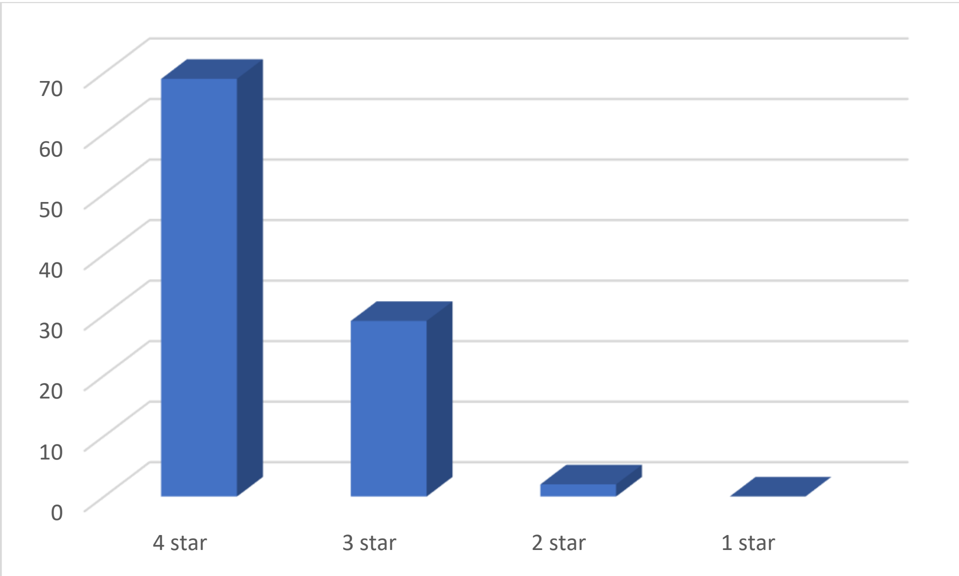
According to the aforementioned data, 80% of workers do very well, 18% do well, 2% do moderately well, and none do poorly. At United Techno, this suggests a high degree of overall performance. An excellent work environment and efficient organizational support are suggested by the lack of subpar performance.

16) Rate your efficiency in completing tasks and meeting deadlines

Table 4.1.16:

Options	No.of respondents	Percentage
4 star	64	69
3 star	27	29
2 star	02	2
1 star	0	0
Total	93	100

Chart 4.1.16:



Interpretation :

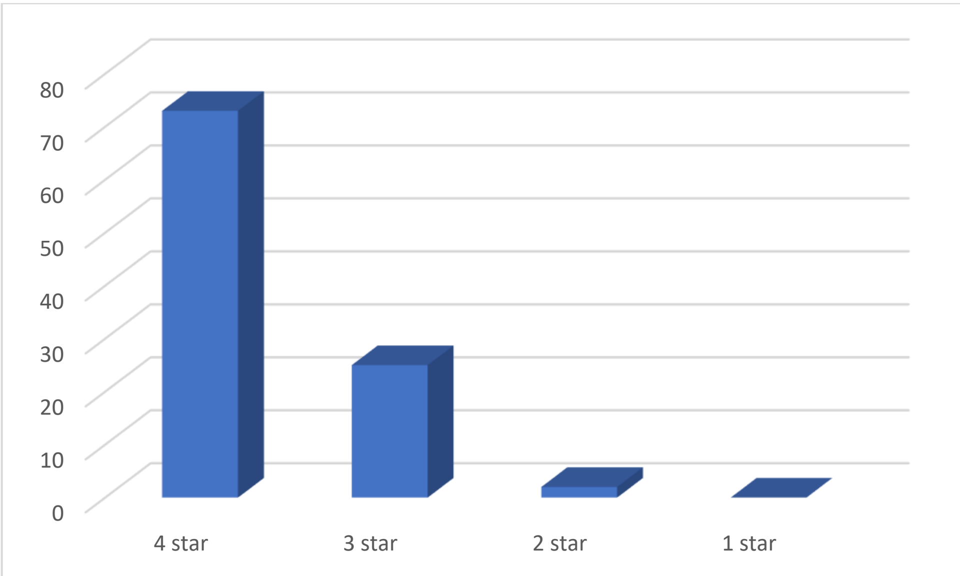
According to the above figure, 69% of employees gave it a 4-star rating, 29% gave it a 3-star rating, 2% gave it a 2-star rating, and none gave it a 1-star rating. This suggests that workers are generally very satisfied. The lack of one-star reviews points to a satisfying workplace and employee experience at United Techno.

17) Rate your contribution to the team's success and achievements

Table 4.1.17:

Options	No.of respondents	Percentage
4 star	68	73
3 star	23	25
2 star	02	2
1 star	0	0
Total	93	100

Chart 4.1.17:



Interpretation :

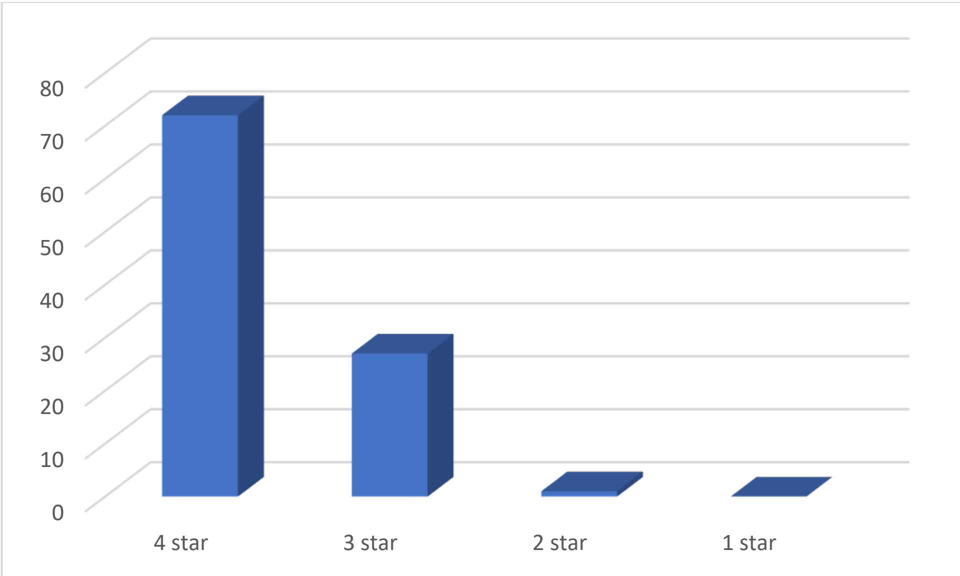
According to the above figure, 73% of employees gave it a 4-star rating, 25% gave it a 3-star rating, 2% gave it a 2-star rating, and no one gave it a 1-star rating. This suggests that workers are generally very satisfied. The lack of negative evaluations points to a satisfying work environment and experience for United Techno employees.

18) Rate the quality of your work in terms of accuracy and attention to detail

Table 4.1.18

Options	No.of respondents	Percentage
4 star	67	72
3 star	25	27
2 star	01	1
1 star	0	0
Total	93	100

Chart 4.1.18:



Interpretation :

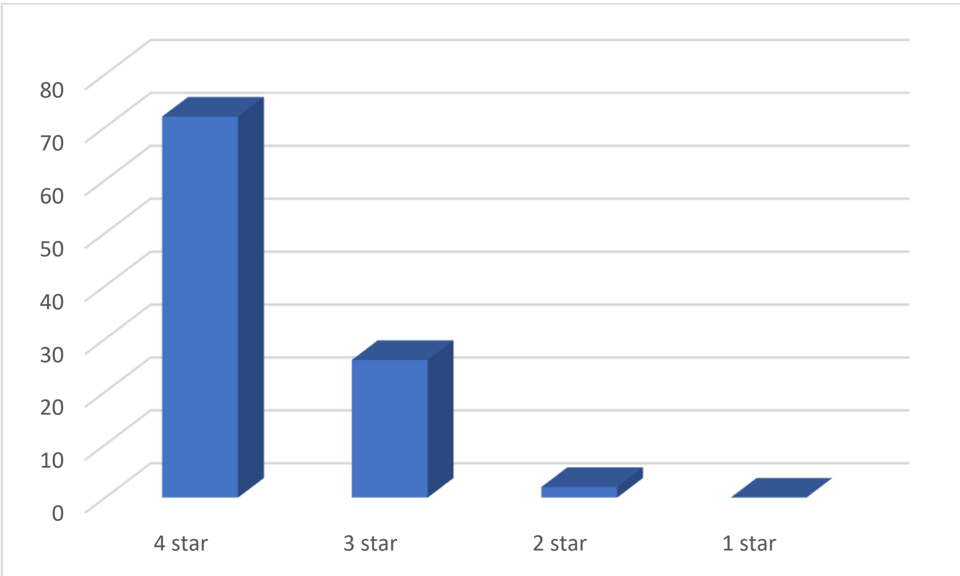
According to the above figure, 72% of employees gave it a 4-star rating, 27% a 3-star rating, 1% a 2-star rating, and no 1-star rating. This suggests that workers are generally very satisfied. The lack of negative evaluations points to a great workplace culture and satisfying employee experience at United Techno.

19) Rate your level of engagement and motivation to perform at your best in your role

Table 4.1.19

Options	No.of respondents	Percentage
4 star	67	72
3 star	24	26
2 star	02	2
1 star	0	0
Total	93	100

Chart 4.1.19:



Interpretation :

According to the aforementioned data, 72% of employees gave it a 4-star rating, 26% a 3-star rating, 2% a 2-star rating, and no 1-star rating. This suggests that workers are generally very satisfied. The lack of negative evaluations points to a great workplace culture and satisfying employee experience at United Techno.

4.2: Random forest classifier:

To assess how particular behavioural and performance-related factors affect workers' individual contributions across several departments, a Random Forest Classifier was used. To categorize employee contribution levels, the study employed three major predictors: problem-solving, learning and development, and conflict management.

To optimize the model, **GridSearchCV** was used for hyperparameter tuning, enhancing accuracy and performance.

4.2.1.Source code :

```
# Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score, classification_report
from sklearn.model_selection import cross_val_score, GridSearchCV
from imblearn.over_sampling import RandomOverSampler
import os

# Load dataset
os.chdir(r"C:\Users\Sahana Dhandapani\OneDrive\Documents")
data = pd.read_excel("coded responses.xlsx")

# Strip whitespace from column names and rename them properly
data.columns = data.columns.str.strip()
data = data.rename(columns={
    'problem solving': 'Problem_Solving',
    'Learning&dev': 'Learning_Dev',
    'conflicts': 'Conflict',
```

```
    'Department': 'Department',
    'd.individual contribution': 'Individual_Contribution'
})

# Encoding Department (Converting categorical to numeric)
data['Department'] = data['Department'].astype(str).str.strip()
label_encoder = LabelEncoder()
data['Department'] = label_encoder.fit_transform(data['Department'])

# Display department encoding
department_mapping = dict(zip(label_encoder.transform(label_encoder.classes_),
label_encoder.classes_))
print("Corrected Department Encoding Mapping:")
for code, dept in department_mapping.items():
    print(f"{code} → {dept}")

# Defining independent (X) and dependent (y) variables
X = data[['Problem_Solving', 'Learning_Dev', 'Conflict' ,'Department']]
y = data['Individual_Contribution']

# Handling class imbalance using Random OverSampling
ros = RandomOverSampler(random_state=42)
X_resampled, y_resampled = ros.fit_resample(X, y)

# Splitting dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.2,
random_state=42)

# Initialize Random Forest Classifier
clf_rf = RandomForestClassifier(n_estimators=100, random_state=42)
clf_rf.fit(X_train, y_train)
```

```
# Make predictions
y_pred_rf = clf_rf.predict(X_test)

# Evaluate the model
accuracy_rf = accuracy_score(y_test, y_pred_rf)
print(f"Random Forest Accuracy: {accuracy_rf:.2f}")
print(classification_report(y_test, y_pred_rf))

# Cross-validation to check performance consistency
cv_scores = cross_val_score(clf_rf, X_resampled, y_resampled, cv=5)
print("Cross-validation scores:", cv_scores)
print("Average cross-validation accuracy:", cv_scores.mean())

# Hyperparameter tuning using GridSearchCV
param_grid = {
    'n_estimators': [50, 100, 200],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}

# Performing Grid Search
grid_search = GridSearchCV(RandomForestClassifier(random_state=42), param_grid, cv=5,
scoring='accuracy', n_jobs=-1)
grid_search.fit(X_resampled, y_resampled)

# Train best model
best_rf = grid_search.best_estimator_
best_rf.fit(X_train, y_train)

# Make predictions using optimized model
y_pred_best_rf = best_rf.predict(X_test)
```

```
# Evaluate optimized model
accuracy_best_rf = accuracy_score(y_test, y_pred_best_rf)
print(f"Optimized Random Forest Accuracy: {accuracy_best_rf:.2f}")
print(classification_report(y_test, y_pred_best_rf))

# Decode department names
X_resampled_df = pd.DataFrame(X_resampled, columns=X.columns)
X_resampled_df['Department'] =
label_encoder.inverse_transform(X_resampled_df['Department'])

# Create a DataFrame with department and predicted contribution
data_resampled = pd.DataFrame({'Department': X_resampled_df['Department'], 'Predicted
Contribution': y_resampled})

# Rank departments based on average predicted contribution
dept_ranking = data_resampled.groupby("Department")["Predicted
Contribution"].mean().sort_values(ascending=False)

# Display department ranking
print("Department Ranking (Best to Worst):")
print(dept_ranking)

#plot department ranking
plt.figure(figsize=(5, 5))
dept_ranking.plot(kind='bar', color='skyblue', edgecolor='black')
plt.xlabel("Department")
plt.ylabel("Average Predicted Contribution")
plt.title("Department Performance Ranking")
plt.xticks(rotation=45)
plt.grid(axis="y", linestyle="--", alpha=0.7)
plt.show()
```

4.2.2.output :

i)Corrected Department Encoding Mapping:

0 → Data Analytics

1 → HR

2 → Integration and cloud services

3 → Software development

ii) Random Forest Accuracy: 0.90

Table 4.2.1

	Precision	Recall	F1-score	Support
2	1.00	1.00	1.00	14
3	1.00	0.71	0.83	14
4	0.76	1.00	0.87	13
accuracy			0.90	41
macro avg	0.92	0.90	0.90	41
weighted avg	0.93	0.90	0.90	41

iii)Cross-validation scores: [0.73170732 0.73170732 0.7804878 0.92682927 0.85]

Average cross-validation accuracy: 0.8041463414634146

iv)Optimized Random Forest Accuracy: 0.90

Table 4.2.2:

	Precision	Recall	F1-score	Support
2	1.00	1.00	1.00	14
3	1.00	0.71	0.83	14
4	0.76	1.00	0.87	13
accuracy			0.90	41
macro avg	0.92	0.90	0.90	41
weighted avg	0.93	0.90	0.90	41

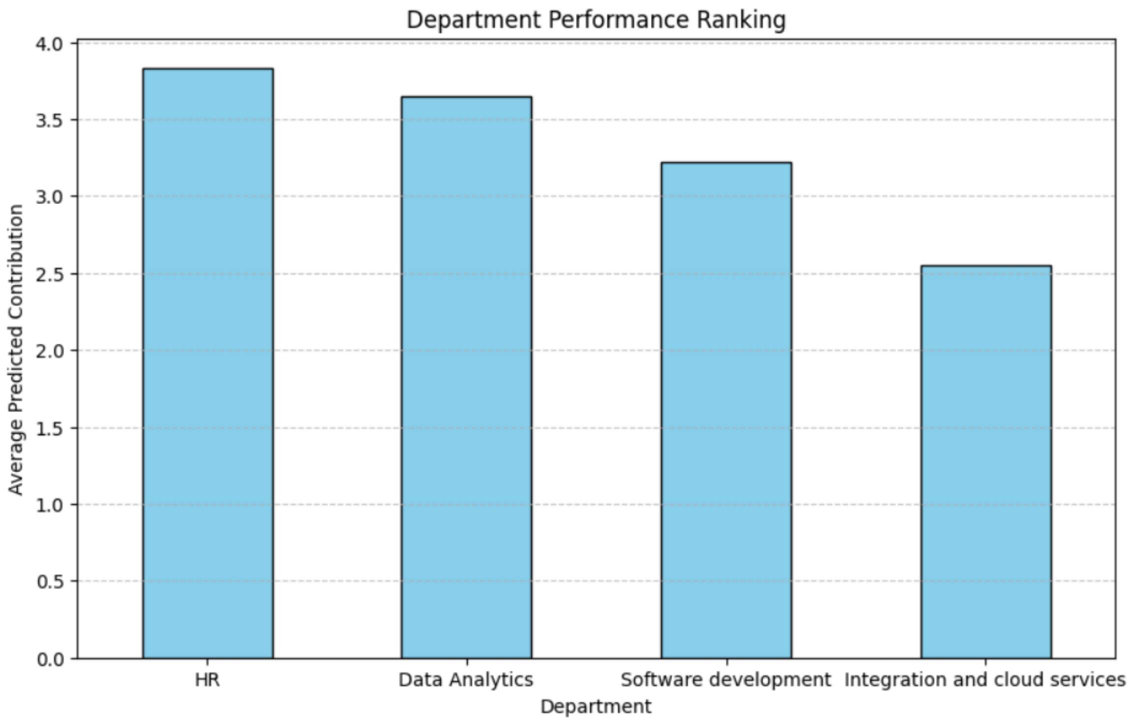
v) Department Ranking (Best to Worst):

Table 4.2.3:

Departments	
HR	3.833333
Data Analytics	3.647059
Software development	3.224138
Integration and cloud services	2.550000

vi)

figure 4.2.1



Interpretation:

Accuracy:

90% The model correctly predicted individual contribution levels 90% of the time.

Precision, Recall & F1-score:

For class 2 and 4, performance is very high (F1 > 0.85), showing strong prediction reliability

Cross validation:

Cross-validation accuracy ranges from **73% to 92%**, with an average of **80.4%**, indicating the model is stable and generalizes well.

Department Ranking (Based on Predicted Contribution):

1. **HR** – Highest predicted individual contribution
2. **Data Analytics**
3. **Software Development**
4. **Integration and Cloud Services** – Lowest predicted contribution

4.3 :ANOVA

A one-way ANOVA test was performed to confirm the department ranking derived from the Random Forest Classifier. It's crucial to statistically verify if the observed variations in performance between departments are meaningful or may have happened by accident, even though the Random Forest Classifier predicted department-wise contributions based on survey data.

The statistical method known as ANOVA (Analysis of Variance) is used to ascertain whether the means of three or more independent groups differ in any way that is statistically significant. The Predicted Contribution score is the dependent variable in this instance, while the groupings stand in for departments.

4.3.1.Source code :

```
import scipy.stats as stats

# Perform ANOVA test
anova_result = stats.f_oneway(
    data_resampled[data_resampled["Department"] == "HR"]["Predicted Contribution"],
    data_resampled[data_resampled["Department"] == "Data Analytics"]["Predicted 
Contribution"],
```

```
    data_resampled[data_resampled["Department"] == "Software development"]["Predicted Contribution"],
    data_resampled[data_resampled["Department"] == "Integration and cloud services"]["Predicted Contribution"]
)

# Display ANOVA result
print("ANOVA Test Result:")
print(f"F-statistic: {anova_result.statistic:.4f}, P-value: {anova_result.pvalue:.4f}")

# Interpret the result
alpha = 0.05  # Significance level
if anova_result.pvalue < alpha:
    print("Conclusion: There is a significant difference between departments' performances.")
else:
    print("Conclusion: No significant difference found between departments' performances.")
```

4.3.2.output:

ANOVA Test Result:

F-statistic: 33.7754, P-value: 0.0000

Conclusion: There is a significant difference between departments' performances.

Interpretation :

Given that the p-value (0.0000) is less than 0.05, we reject the null hypothesis and come to the conclusion that the anticipated performance contribution varies statistically significantly among departments. This demonstrates that the department-wise variation found using the Random Forest model is actually significant and not just random.

4.4.Tukey HSD:

After establishing with ANOVA that there are significant differences in performance among departments, a Tukey HSD (Honestly Significant Difference) test was conducted to identify which specific department pairs differ significantly in their predicted contribution scores.

4.4.1.Source code :

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd

tukey_result = pairwise_tukeyhsd(
    endog=data_resampled["Predicted Contribution"],  # Target variable
    groups=data_resampled["Department"],  # Grouping variable
    alpha=0.05  # Significance level
)


# Print the summary
print(tukey_result.summary())
```

4.4.2.output:

Figure 4.4.1:

```
                         Multiple Comparison of Means - Tukey HSD, FWER=0.05
==============================================================================================
           group1                          group2             meandiff p-adj  lower   upper  reject
----------------------------------------------------------------------------------------------
             Data Analytics                              HR    0.1863 0.8422 -0.3981  0.7707  False
             Data Analytics Integration and cloud services  -1.0971    0.0 -1.4426 -0.7515   True
             Data Analytics            Software development  -0.4229 0.0205 -0.7989  -0.047   True
                         HR Integration and cloud services  -1.2833    0.0 -1.8151 -0.7516   True
                         HR            Software development  -0.6092 0.0241 -1.1612 -0.0572   True
Integration and cloud services          Software development   0.6741    0.0  0.3869  0.9614   True
----------------------------------------------------------------------------------------------
```

Interpretation :

The biggest difference is between HR and Integration & Cloud Services, with HR receiving a far higher score.

When compared to every other department, Integration & Cloud Services continuously performs worse.

Data analytics is somewhat superior than software development and far superior to integration.

There was no discernible difference between HR and Data Analytics, suggesting that their performance levels were comparable

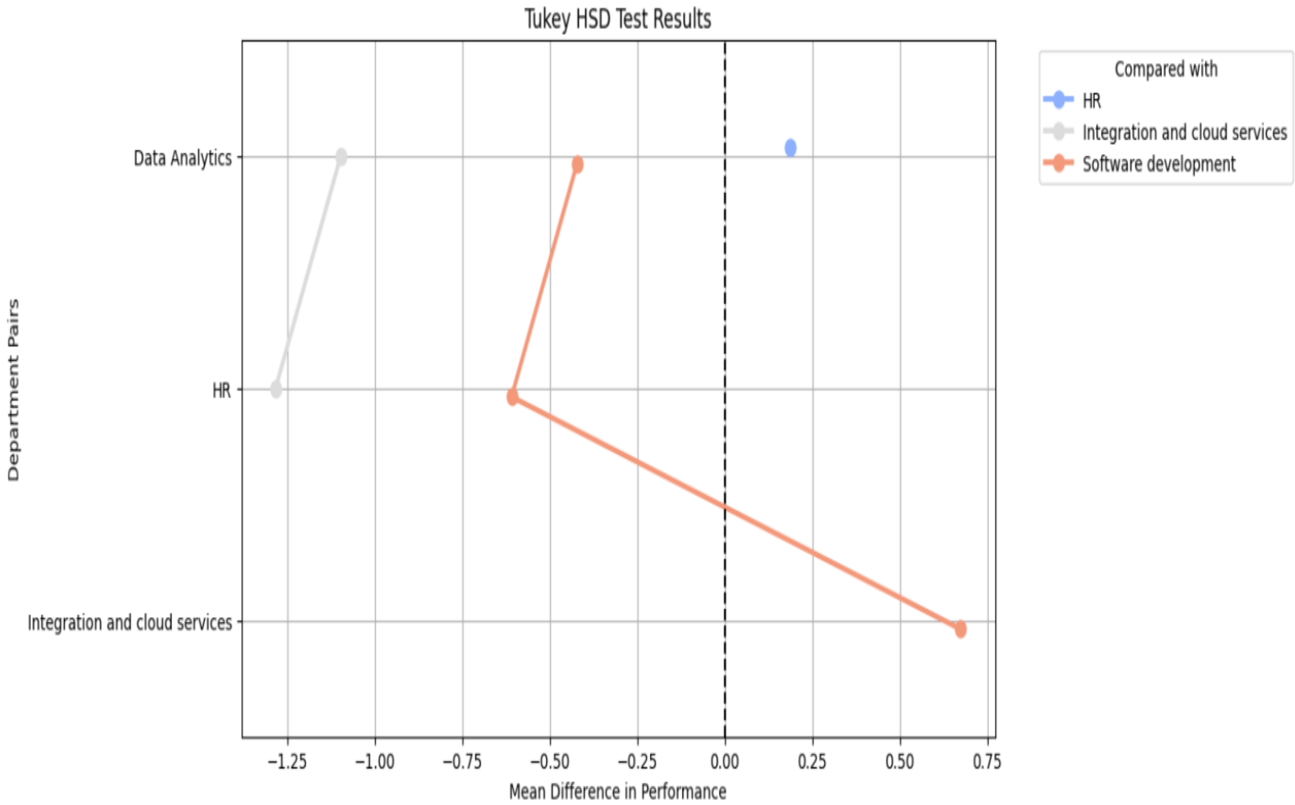
4.4.3.Visualization of Tukey Results

```
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Perform Tukey HSD test
tukey_result = pairwise_tukeyhsd(
    data_resampled["Predicted Contribution"],  # Target variable
    data_resampled["Department"],  # Grouping variable
    alpha=0.05  # Significance level
)

# Convert Tukey test result to DataFrame
tukey_df = pd.DataFrame(data=tukey_result.summary().data[1:],
columns=tukey_result.summary().data[0])

# Plot the confidence intervals
plt.figure(figsize=(10, 6))
sns.pointplot(
    x="meandiff",
    y="group1",
    hue="group2",
    data=tukey_df,
    dodge=True,
    markers="o",
    palette="coolwarm"
)
```

```
plt.axvline(x=0, color='black', linestyle='--')  # Add a vertical reference line at 0
plt.xlabel("Mean Difference in Performance")
plt.ylabel("Department Pairs")
plt.title("Tukey HSD Test Results")
plt.legend(title="Compared with", bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()
```

output :

figure 4.4.2:



interpretation :

Compared to all other departments, the Integration and Cloud Services department did much worse.

Additionally, software development performed worse than data analytics and human resources.

There was no discernible distinction between the HR and data analytics departments.

This demonstrates that departmental performance disparities are actual and not coincidental.

Tukey HSD confirms that department-wise differences are real — this supports the Random Forest classification results

4.5: Random forest regression :

Random Forest Regression is an ensemble machine learning method that builds multiple decision trees and combines their outputs to improve prediction accuracy. It is well-suited for complex, non-linear relationships and can handle multicollinearity and interaction effects among variables. In this study, Random Forest was applied to predict the "Target Completion" of employees based on key organizational factors such as leadership, learning, working conditions, and training.

4.5.1.Source code :

```
import pandas as pd
import os
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression, Ridge, Lasso, ElasticNet
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cross_decomposition import PLSRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import seaborn as sns


os.chdir(r"C:\Users\Sahana Dhandapani\OneDrive\Documents")
data=pd.read_excel("coded responses.xlsx")
data.columns = data.columns.str.strip()

# Define Independent Variables (IVs) - Employee Utilization of Resources
X = data[["Working condition", "E.Leadership", "career development", "co-workers",
"Training", "Learning"]]

# Define Dependent Variables (DVs) - Employee Performance
y1 = data["d.target completion"]
```

```
# Split data into training and testing sets (80% train, 20% test)
X_train, X_test, y1_train, y1_test = train_test_split(X, y1, test_size=0.2, random_state=42)

def train_random_forest(X_train, X_test, y_train, y_test, target_name):
    # Initialize Random Forest Regressor
    rf = RandomForestRegressor(n_estimators=100, random_state=42)

    # Train the model
    rf.fit(X_train, y_train)

    # Make predictions
    y_pred = rf.predict(X_test)

    # Evaluate Model Performance
    mse = mean_squared_error(y_test, y_pred)
    r2 = r2_score(y_test, y_pred)

    print(f"Random Forest Regression for {target_name}:")
    print(f"Mean Squared Error (MSE): {mse:.4f}")
    print(f"R-squared Score (R²): {r2:.4f}\n")

    # Feature Importance
    feature_importance = pd.DataFrame({"Feature": X.columns, "Importance":
rf.feature_importances_})
    feature_importance = feature_importance.sort_values(by="Importance", ascending=False)

    # Plot Feature Importance
    plt.figure(figsize=(8, 5))
    sns.barplot(x=feature_importance["Importance"], y=feature_importance["Feature"],
color="skyblue")
    plt.xlabel("Importance Score")
    plt.ylabel("Features")
    plt.title(f"Feature Importance for {target_name} (Random Forest)")
```

```
    plt.show()

return rf
# Train Random Forest for Each Dependent Variable
rf_target_completion = train_random_forest(X_train, X_test, y1_train, y1_test, "Target Completion")
```

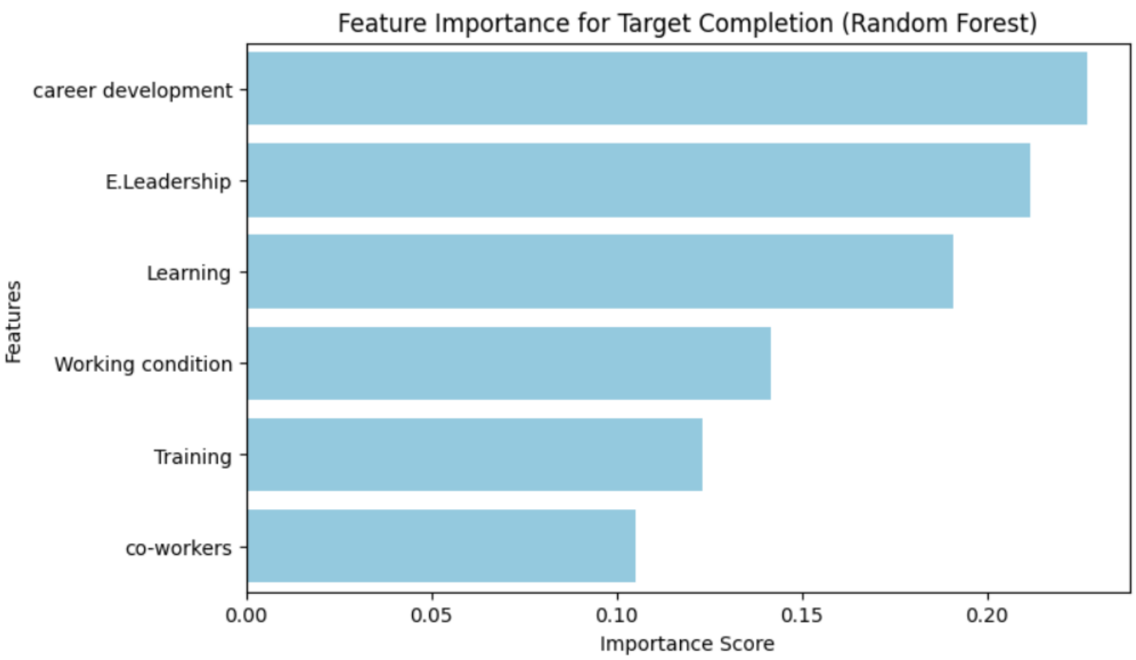
4.5.2.output:

Random Forest Regression for Target Completion:

Mean Squared Error (MSE): 0.1706

R-squared Score (R^2): 0.1204

Figure 4.5.1



The Random Forest Regression model yielded an R^2 score of 0.1204, indicating that approximately 12% of the variance in Target Completion can be explained by the selected independent variables:
Working Condition, E. Leadership, Career Development, Co-workers, Training, and Learning.

The Mean Squared Error (MSE) of 0.1706 represents the average squared difference between the actual and predicted values. While the R^2 value is relatively low, it still provides initial insights into how organizational factors influence an employee's ability to meet their targets.

The feature importance chart highlights that:

- Career Development, E. Leadership, and Learning are the top three contributors to predicting Target Completion.
- This implies that employees are more likely to complete their targets when they have access to growth opportunities, effective leadership, and a strong learning culture

4.6 : k-means clustering:

An unsupervised machine learning technique called K-means clustering divides a collection of data points into a predetermined number of clusters (K). Every employee (data point) is categorized into a cluster according to how similar their survey answers are. Finding hidden patterns or organic groupings in the collection without any prior labels is the main goal.

Why K-Means Clustering Was Used:

the use of K-Means Clustering to divide up the workforce according to performance metrics and survey responses.

To comprehend various employee profiles according to their contribution to corporate success, leadership views, degree of satisfaction, and conflict resolution skills.

To identify staff groups with specific needs or capabilities in order to deliver actionable insights for organizational development

Determination of Optimal Clusters – Elbow Method:

The optimal number of clusters (K) was ascertained using the Elbow Method. The number of clusters is plotted versus the Within-Cluster Sum of Squares (WCSS).

The overall variation within each cluster is measured by WCSS.

Because each cluster is more exact and smaller, WCSS automatically falls as the number of clusters grows.

The ideal number of clusters is indicated by the "elbow point"—where the curve bends.

Cluster Center and Cluster Formation:

K-means clustering was used after K was determined. For every cluster, the algorithm determined the average value (centroid) of every survey variable.

The average survey responses of the employees in that cluster are shown in each row of the cluster center table

Visualization Using Heatmap

To see cluster profiles across survey variables, a heatmap was made.

Higher average values (such as greater agreement or better results) are represented by redder hues.

Lower average values (such as poorer results or weaker agreement) are represented by bluer hues.

It offers a clear comparison between variables and clusters.

4.6.1.Source code :

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load your data
data = pd.read_excel("coded responses.xlsx")



# Step 2: Define the survey columns to include in clustering
features = ['Working condition', 'T.Working condition', 'C.Leadership', 'E.Leadership',
        'career development', 'co-workers', 'Training', 'Learning', 'problem solving', 'Learning&dev',
        '  conflicts ', 'd.individual contribution', 'd.target completion    ', 'd.quality of work ', 'd.engagement&motivation ']

X = data[features]

# Step 3: Scale the data
scaler = StandardScaler()
```

```
X_scaled = scaler.fit_transform(X)

# Step 4: Use the Elbow Method to find the optimal number of clusters
wcss = []
for k in range(1, 10):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

# Step 5: Plot the Elbow curve
plt.plot(range(1, 10), wcss, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('WCSS')
plt.grid(True)
plt.show()

# Step 6: Fit K-Means with chosen k (example: 3)
kmeans = KMeans(n_clusters=3, random_state=42)
data['Cluster'] = kmeans.fit_predict(X_scaled)



# Step 7: Analyze the clusters
cluster_summary = data.groupby('Cluster')[features].mean()
print(cluster_summary)

# Step 8: Visualize cluster profiles with heatmap
sns.heatmap(cluster_summary, annot=True, cmap='coolwarm')
plt.title('Cluster Profiles Based on Survey Variables')
plt.show()
```

4.6.2. output :

Elbow method :

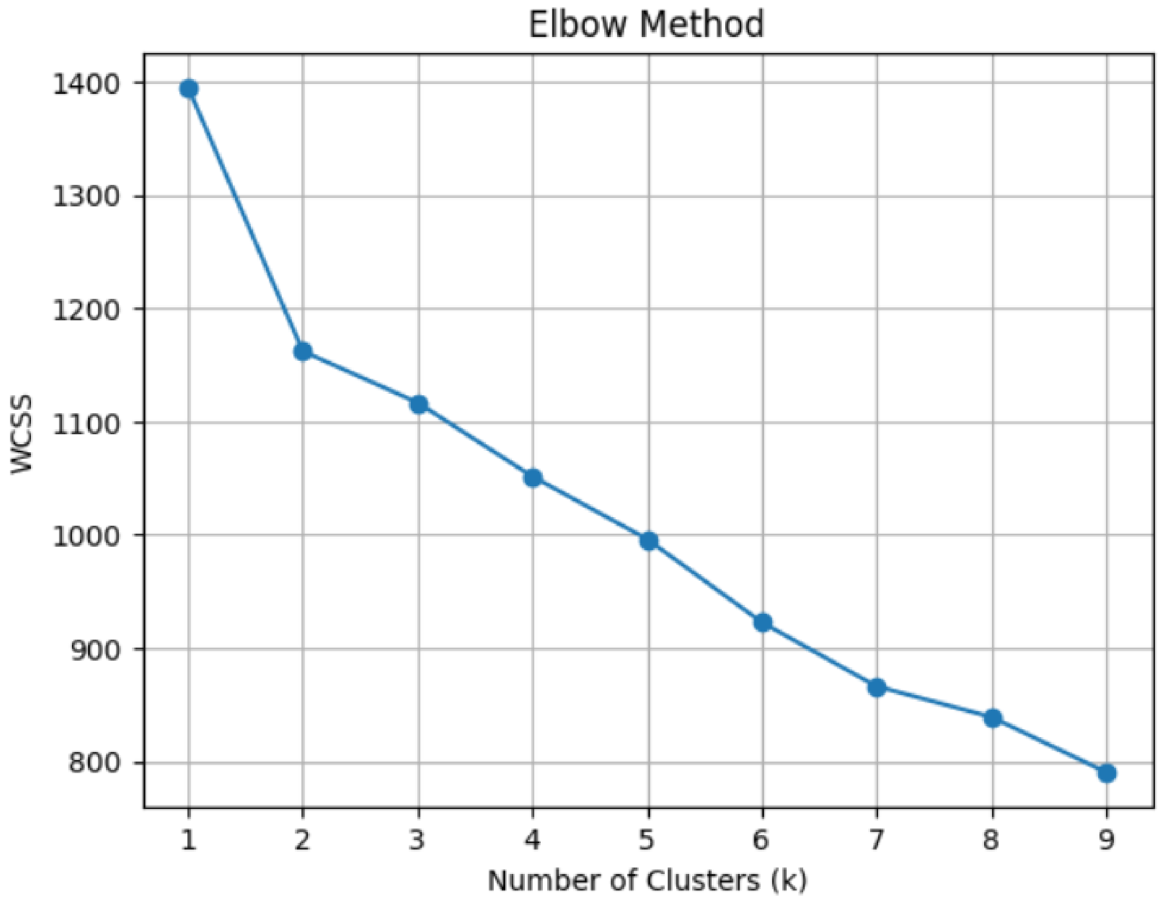


figure 4.6.1

Interpretation for elbow method :

Between clusters one and three, the WCSS drastically decreases.

The decline in WCSS is substantially slower after K=3.

Consequently, three clusters were selected for the best staff segmentation

Average score for each cluster :

Table 4.6.1

Variables	Cluster 0	Cluster 1	Cluster 2
Working condition	3.2777	3.9242	3.6666
Tools availability	3.4444	3.7272	3.6666
Leadership communication	3.2777	3.8030	3.4444
Leadership effectiveness	3.4444	3.9393	3.7777
Career development	3.2777	3.9242	3.4444
Co-workers	3.7222	3.9242	3.6666
Training	3.4444	3.9545	3.8888
Learning	3.6111	3.8636	3.4444
Problem solving	3.3333	3.9242	3.4444
Learning & development	3.5555	3.7575	3.7777
Conflicts	3.5555	3.9242	3.0000
Individual contribution	3.5000	3.8030	3.5555
Target completion	3.7777	3.6515	3.3333
Quality of work	3.8888	3.7878	2.8888
Engagement & motivation	3.5555	3.7575	3.5555

Interpretation :

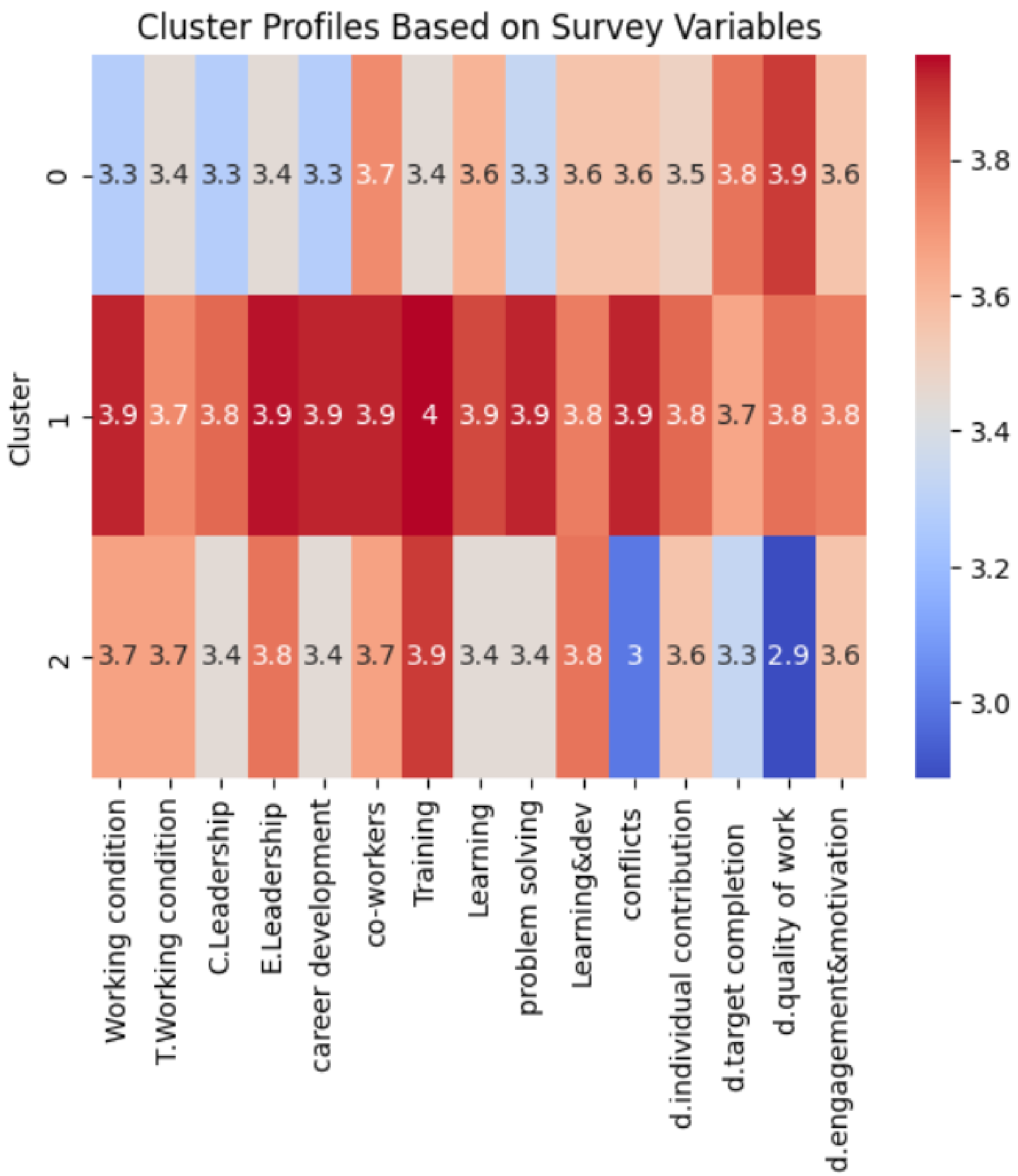
Cluster 1 represents **highly satisfied and highly performing employees** — leadership, training, career development are their strengths.

Cluster 0 represents **moderately satisfied employees** — performing well but could benefit from leadership improvements and stronger development programs.

Cluster 2 represents **employees at risk** — facing challenges in conflict management, quality of work, and performance metrics.

Heatmap :

Chart 4.6.2:



Interpretation :

The heatmap clearly shows that Cluster 1 employees exhibit the highest satisfaction across leadership, conflict management, and learning opportunities, while Cluster 2 employees face challenges related to conflict resolution, quality of work, and target achievement. Cluster 0 employees represent an intermediate group with moderate satisfaction

4.7. Sentiment analysis:

A natural language processing (NLP) method called sentiment analysis is used to identify if textual input has a neutral, negative, or positive sentiment. It assists businesses in comprehending the emotional tone of open-ended survey replies, customer reviews, and employee feedback.

In order to assess general sentiment trends, sentiment analysis was used in this study on open-ended employee responses. We employed a sentiment score technique called VADER (Valence Aware Dictionary for Sentiment Reasoning), which is ideal for evaluating brief, informal texts like survey responses.

The most often used terms in employee feedback are shown graphically in a word cloud. Each word's size reveals its emphasis or frequency.

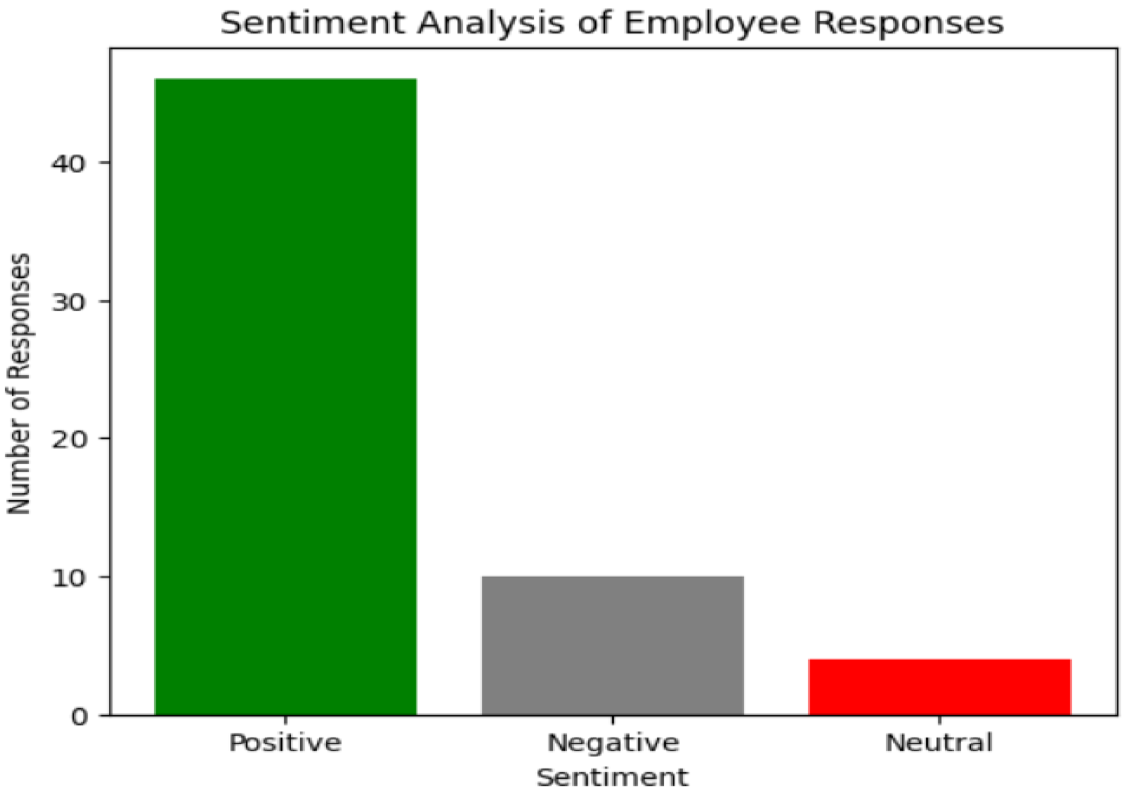
4.7.1. Source code:

```
import pandas as pd
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
# Download VADER lexicon (only once)
nltk.download('vader_lexicon')
df = pd.read_excel("unique_employee.xlsx")
response_column = ' Response'
df.columns = df.columns.str.strip()
print(df.columns)
response_column = 'Response'

# Initialize sentiment analyzer
sia = SentimentIntensityAnalyzer()
```

```
# Define function to classify sentiment
def get_sentiment(score):
    if score >= 0.05:
        return 'Positive'
    elif score <= -0.05:
        return 'Negative'
    else:
        return 'Neutral'

# Apply sentiment analysis
df['Sentiment_Score'] = df[response_column].apply(lambda x:
sia.polarity_scores(str(x))['compound'])
df['Sentiment'] = df['Sentiment_Score'].apply(get_sentiment)
sentiment_counts = df['Sentiment'].value_counts()
plt.figure(figsize=(6,5))
plt.bar(sentiment_counts.index, sentiment_counts.values, color=['green', 'grey', 'red'])
plt.title('Sentiment Analysis of Employee Responses')
plt.xlabel('Sentiment')
plt.ylabel('Number of Responses')
plt.show()

# --- Generate Word Cloud ---
all_text = ' '.join(df[response_column].astype(str))

wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(all_text)

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Word Cloud of Employee Responses')
plt.show()
```

4.7.2. Output:

Chart 4.7.1



Interpretation:

Positive Sentiment: Positive was the classification given to the majority of respondents . This suggests that the majority of workers were happy, appreciative, or hopeful about certain elements of their jobs, leadership, or work environment**.**

Negative sentiment was voiced by a smaller group of respondents These could be an indication of discontent with the workload, management style, or lack of room for advancement.

Neutral Sentiment: A small number of comments were categorized as neutral since they provided factual or emotionally balanced input devoid of overtly favourable or negative wording.

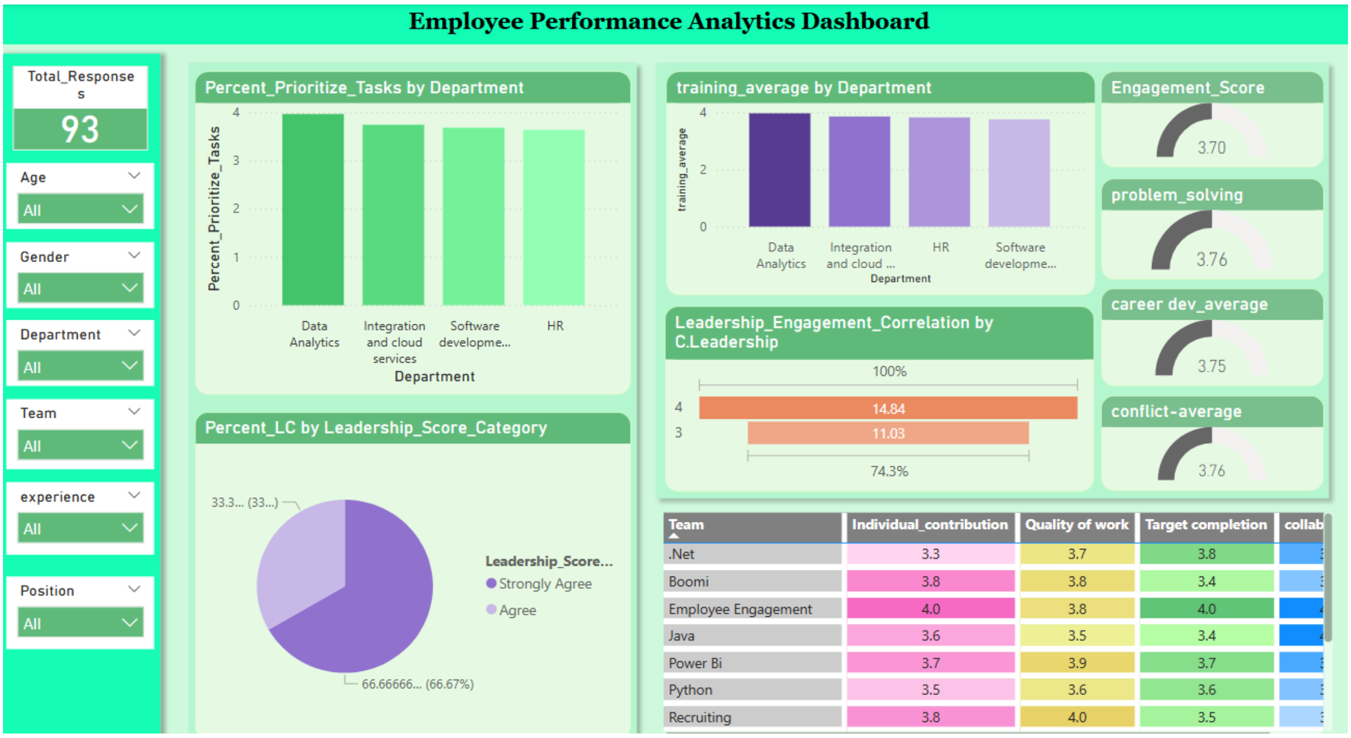
Chart 4.7.2

Interpretation :

According to the word cloud analysis of their response’s employees' feelings, their work, their role, and the team environment are just a few of the emotional and team-related elements that have a significant impact on them, Positive themes that suggested a proactive approach and a desire for change were prevalent, including motivation, growth, and support. Nonetheless, issues including stress, workload, and ambiguity were also identified, emphasizing crucial areas that require managerial attention. These findings imply that although workers are typically engaged, performance and satisfaction could be further increased by strengthening leadership, communication, and task management.

4.8. Power BI Dashboard:

figure 4.8.1:



Objective of the dashboard:

In order to assess how employees take advantage of organizational possibilities and how leadership affects individual performance, United Techno conducted a poll, the results of which are displayed in this Power BI dashboard. It links important factors like engagement, leadership, learning, and problem-solving with performance outcomes like contribution, target completion, and work quality via visual analytics.

Dashboard Components and Their Interpretations:

i)Top Left Panel – Filters and Total Responses:

Total Responses: Indicates that 93 workers took part in the poll.

Users can filter data using slicers (Age, Gender, Department, Team, Experience, Position) to examine performance by particular demographics or groupings.

ii)Percent_Prioritize_Tasks by Department

Type of Visual: Column Chart

Insight: Displays the proportion of workers in each division who gave themselves excellent marks for job prioritization.

Interpretation: The data analytics department is marginally ahead in task prioritization, indicating improved time management or more defined goals.

iii)Training average by Department:

Type of Visual: Bar Chart

Insight: Shows the average rating for each department's training and development.

Interpretation: All departments scored similarly (~4), indicating that training initiatives are consistent across departments.

iv)Leadership_Engagement_Correlation by Leadership:

Visual Type: Funnel chart

Insight: measures the relationship between:

How employees rated leadership (C.Leadership)

How employees rated their own engagement & motivation (d.engagement&motivation)

Interpretation: Shows if strong leadership increases employee engagement using a numeric average of all products (Leadership × Engagement)

DAX used:

```
Leadership_Engagement_Correlation =
AVERAGEX(
    EmployeeData,
    EmployeeData[C.Leadership] * EmployeeData[d.engagement&motivation ])
```

v) Percent_LC by Leadership_Score_Category

Type of Visual: Pie Chart

Insight: Indicates the proportion of respondents who assigned each leadership score.

DAX for category:

```
Leadership_Score_Category = SWITCH (TRUE (),
   EmployeeData[C.Leadership] = 4, "Strongly Agree",
   EmployeeData[C.Leadership] = 3, "Agree",
   EmployeeData[C.Leadership] = 2, "Disagree",
   EmployeeData[C.Leadership] = 1, "Strongly Disagree",
   "No Response")
```

DAX for Percentage:

```
Percent_LC =
VAR TotalResponses = CALCULATE(COUNT(EmployeeData[C.Leadership]),
ALL(EmployeeData))
VAR CountPerCategory = COUNT(EmployeeData[C.Leadership])
RETURN DIVIDE(CountPerCategory, TotalResponses, 0)*100
```

Interpretation:

66.67% of employees “Strongly Agree” that leadership is effective. 33.33% “Agree”. No negative responses were recorded, suggesting positive leadership perception.

vi) KPI gauges Averages :

The visual shows overall averages of the survey (on a scale 1 to 4)

Metrics like engagement score ,problem solving ,career development average ,conflict average.

vii)Team performance table :

visual : Table

metrics :

- individual contribution
- quality of work
- target completion
- collaboration

interpretation:

Example: The Employee Engagement team scored highest (4.0) in contribution and target completion. The Boomi and Power BI teams scored well in contribution and quality of work. This comparison helps identify high-performing teams.

The dashboard's main conclusions are as follows:

Leadership Perception is Positive: Two-thirds of employees strongly agree that leadership is effective

Leadership and Engagement Are Correlated: Greater motivation is associated with higher leadership scores

Team-Based Performance Insights: Teams such as "Employee Engagement" and "Power BI" exhibit strong performance metrics; and Training, Task Prioritization, and Problem Solving are consistently strong across departments.

CHAPTER V

SUMMARY OF FINDINGS SUGGESTIONS AND CONCLUSION

5.0. summary finding, suggestions, conclusion:

5.1. findings:

i)Contribution by Department (Random Forest Classifier):

The model's 90% accuracy rate validates that variables like problem-solving, learning, and conflict resolution may be used to accurately forecast employee contribution levels.

Data analytics, software development, integration, and cloud services were the next most anticipated contributions, after human resources.

ii)Significant Departmental Differences (ANOVA & Tukey HSD)

ANOVA revealed that departmental performance differed significantly ($p < 0.05$).
Tukey's HSD test verified that employee contributions vary greatly across a number of departments, particularly HR and Integration & Cloud Services.

iii) Impact of Resource Utilization on Performance (Random Forest Regression)

Training and career development were found to have the greatest effects on target completion by Random Forest Regression. These observations emphasize how crucial it is to concentrate on chances for learning and development in order to improve worker performance.

iv) sentiment analysis:

Responses were categorized using VADER Sentiment Analysis as follows:

- positive
- neutral
- negative

The majority of workers voiced positive or constructive opinions, particularly for leadership and teamwork.

Dominant phrases including "feels," "work," "team," "role," and "make" were found in the Word Cloud, indicating an emphasis on job satisfaction and team dynamics.

v) k-means clustering:

Cluster 1-High Performers: Workers in this group are always productive and extremely happy. They are the organization's greatest assets because of their superior leadership, training, and career development skills.

Moderate Performers (Cluster 0): These workers exhibit consistent performance and a moderate level of pleasure. Even though they are typically successful, they might gain from more leadership involvement and more organized development initiatives.

Cluster 2: Employees at Risk:
This group has issues with involvement, work quality, and conflict resolution. To increase performance and lower the danger of disengagement, they urgently need assistance through leadership, mentorship, and training.

vi)Business insights via power BI dashboard:

Percent_Prioritize_Department-specific tasks: highest in software development, integration & cloud, and data analytics.

Data analytics has the highest training average, indicating a culture that prioritizes learning.

The average score for engagement, problem-solving, career development, and conflict resolution is about 3.7, which indicates a modest level of participation with potential for improvement.

The correlation between leadership and engagement is clearly positive, indicating that effective leadership increases employee engagement.

Team-Wise Performance Matrix: Indicates which teams are the best at collaboration, target completion, and individual contribution.

5.2:suggestions :

Make Training and Career Development Investments Pay close attention to organized learning programs since they have a big impact on performance and goal accomplishment.

Develop Your Leadership Capabilities in All Teams
Fill up the leadership voids, particularly for clusters of employees that perform moderately or poorly.

Assist Workers Who Are at Risk (Cluster 2)
To increase their participation, offer performance reviews, dispute resolution assistance, and focused coaching.

Gain Knowledge from Departments That Contribute a Lot
Throughout the company, adopt the best practices from high-achieving departments like Cloud Services and HR.

Re-engage Cluster 0 through growth opportunities and incentive, retain Cluster 1 by acknowledging their efforts and involving them in mentorship, and assist Cluster 2 personnel with leadership direction and development plans in order to increase overall performance.

Respond to Sentiment Analysis
To resolve underlying unhappiness and boost morale, follow up on neutral and negative comments.

Utilize Power BI to Conduct Continuous Monitoring
Use your dashboard often to monitor developments, identify problems early, and inform HR choices.

5.3. conclusion:

This study demonstrates how the integration of machine learning and business analytics may yield insightful information about worker engagement, performance, and contribution. Using techniques like Random Forest, ANOVA, K-Means Clustering, and Sentiment Analysis, we were able to pinpoint the main factors that contribute to employee performance, including leadership, career growth, and training.

HR and Cloud Services are two high-performing departments that serve as excellent models for others. While sentiment analysis provided a greater grasp of the viewpoints of the employees, clustering techniques assisted in classifying them for focused improvement. All of the findings were combined in the Power BI dashboard, giving decision-makers instantaneous, clear insights.

Organizations may greatly enhance employee outcomes and overall performance by implementing data-driven initiatives to boost engagement, foster growth, and strengthen leadership.

Bibliography :

1.Breiman, L. (2001). Random forests. Machine Learning, 45(1), 532.

https://link.springer.com/article/10.1023/A:1010933404324

2. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning.

3. McKinsey & Company. (2020). How to use people analytics to drive business performance

https://www.mckinsey.com/

4.Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135

https://doi.org/10.1561/1500000011

5. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

6. Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications

7. Miner, G., Elder IV, J., Hill, T., Nisbet, R., Delen, D., & Fast, A. (2012). Practical text mining and statistical analysis for non-structured text data applications. Academic Press

8.Gandomi, A., & Haider, M. (2015). *Beyond the hype: Big data concepts, methods, and analytics*. International Journal of Information Management, 35(2), 137–144

https://doi.org/10.1016/j.ijinfomgt.2014.10.007

9.IBM. (2021). *Employee sentiment analysis using natural language processing (NLP)*. Retrieved from

https://www.ibm.com/blogs/

10.Kaggle. (n.d.). *Employee performance dataset*. Retrieved from https://www.kaggle.com/

(for practice)

ANNEXURE:

Plagiarism report :

- The plagiarism check for this research report was conducted using the online tool SEO Mega Tools Plagiarism Checker (URL: https://seomegatools.com/plagiarism-checker) on May 3, 2025.
  Total Rows Analysed: 836
- The plagiarism analysis was performed on the theoretical content and source code , excluding tables, figures, diagrams, and screenshots.
- The result indicates 100% unique content, with 0.0% similarity detected.

Below is the screenshot of the plagiarism result:

5/4/25, 10:08 PM Advance free Plagiarism Checker | SEO MEGA TOOLS

Advance free Plagiarism Checker

Plagiarism Checker

Paste (Ctrl + V) your article below then click Check for Plagiarism!

100% Unique Content

#	String	Uniqueness
1	ganization, measuring their efficiency, productivity,	Good
2	efficiency, productivity, engagement, and resource utilization.	Good
3	oductivity, engagement, and resource utilization. Across	Good
4	engagement, and resource utilization. Across organizations,	Good
5	tilization. Across organizations, business analytics,	Good
6	anizations, business analytics, machine learning, and business	Good
7	analytics, machine learning, and business intelligence	Good
8	e learning, and business intelligence tools are used	Good
9	ectiveness, streamline performance, and align employee	Good
10	erformance, and align employee pursuits with enterprise	Good
11	approaches. Using predictive analytics, statistical	Good
12	analytics, statistical models, and performance metrics	Good
13	cal models, and performance metrics allow businesses	Good
14	oductivity, strengthen employee development programs,	Good
15	t programs, and utilize organizational resources more	Good

https://seomegatools.com/plagiarism-checker 1/36

Questionnaire :

1)Name :

2)Gender

A. Male

B. Female

3)Age

A.18-24

B.25-34

C.35-44

D.Above 45

4) What is your tenure at United Techno?

A. Less than one year

B. 1-3 years

C. 4-6 years

D. 7-10 years

E. Over 10 years

5)Department :

6)please specify the name of your team

7)Educational Qualification:

8)current position at united techno

9)To what extent do you use the tools and resources provided to complete your work?

A. Fully

B. When needed

C. Sometimes insufficient

D. Rarely

10)What is your approach to prioritizing tasks in your role?

A. Clear system and method

B. Sometimes prioritized

C. Rarely prioritized

D. Not prioritized

11)To what extent do you communicate your needs and concerns to your manager?

A. Openly and regularly

B. Occasionally when necessary

C. Rarely

D. Never

12)How do you ensure that your goals are in line with your manager's expectations?

A. Regularly discuss and align them

B. Occasionally align them

C. Rarely align them

D. Never align them

13)How often do you align your current role with your long-term career goals?

A. Regularly evaluate and align

B. Occasionally review alignment

C. Rarely consider it

D. Never align them

14)How would you describe your collaboration with co-workers on team goals?

A. Frequently collaborates

B. Occasionally collaborates

C. Rarely collaborates

D. Do not collaborate

15)How often do you apply knowledge gained from training to your tasks?

A. Regularly

B. Occasionally

C. Rarely

D. Never

16) How frequently do you utilize available learning resources (e.g., workshops, courses)?

A. Regularly

B. Occasionally

C. Rarely

D. Never

17)How effective are you in identifying and solving problems that arise within your team?

A. Very effective, I quickly identify and resolve issues

B. Effective, but I take time to find solutions

C. Occasionally effective

D. Ineffective

18)How committed are you to learning new skills to improve your performance?

A. Very committed

B. Committed

C. Somewhat committed

D. Not committed

19)How well do you manage conflicts or disagreements within your team?

A. Very well

B. Well

C. Somewhat well

D. Poorly

20)Rate your contribution to the team's success and achievements.(4 star)

21)Rate your efficiency in completing tasks and meeting deadlines(4 star)

22) Rate the quality of your work in terms of accuracy and attention to detail (4 star)

23) Rate your level of engagement and motivation to perform at your best in your role (4 star)

24) How do you feel about your overall experience working in this organization

(Note: Your response to this question will be used for sentiment analysis to understand employee emotions and perceptions.)