HR Data Analysis and Workforce Performance Insights

Sindh Integrated Emergency & Health Services (SIEHS)

Analysis Report

By Muhammad Yahya

Contents

1 Introduction

Workforce performance and attendance management are central components of operational efficiency in large service organizations. Sindh Integrated Emergency & Health Services (SIEHS) operates one of the largest emergency and health support networks in the region, deploying thousands of employees across stations, control rooms, and field units. Managing punctuality, daily attendance, leave patterns, and compliance at this scale requires more than administrative oversight; it requires a systematic, data-driven framework.

This analysis report presents a comprehensive examination of attendance and workforce efficiency metrics for June 2024. Using data extracted from HR systems—including attendance logs, approved and pending leave records, invalid time entries, and consolidated absence sheets—this study aims to identify patterns, inefficiencies, and actionable insights.

The report further consolidates these disparate data sources into a unified master dataset, enabling quantification of on-time arrival rates, absenteeism patterns, working-hour distributions, and departmental performance differences. The final dashboard visualizations highlight the structural strengths and bottlenecks across departments and locations, offering leadership a high-level and granular understanding of workforce performance.

The objective of this analysis is twofold:

- To evaluate the operational efficiency of SIEHS through a detailed examination of attendance and punctuality data.
- To construct a unified analytics framework capable of supporting ongoing workforce monitoring and performance improvement.

This report is structured to be both analytical and operational, offering executives clarity on what is happening, why it is happening, and where corrective actions should be prioritized.

2 Organizational Background

Sindh Integrated Emergency & Health Services (SIEHS) is responsible for emergency medical response and support operations across Sindh, including ambulance services, emergency medical personnel, call-center staff, and field support teams. With a workforce of nearly 1,800 employees operating across multiple districts and specialized departments, achieving consistent workforce discipline is crucial.

SIEHS operates in a complex environment:

- Employees work in rotating shifts.
- Field and control-room duties often run 24/7.
- Attendance, leave approvals, and schedule adherence vary significantly by department.
- Timekeeping is influenced by operational surges, emergency demands, and public holidays.

Given this operational landscape, HR leadership must rely on accurate, consolidated data to understand workforce reliability, identify departments facing punctuality challenges, and ensure staffing levels meet operational requirements. Prior to this analysis, much of the data existed in fragmented Excel files maintained independently.

This project centralizes, cleans, and integrates these sources into one consistent analytical environment.

3 Data Sources

Five independent Excel-based HR datasets were used for the June 2024 analysis:

1. **Attendance Sheet – June 2024** Daily in/out timestamps for each employee, with coded values defining present, absent, holidays, off-days, and special leave types.

2. **Approved Leaves – June 2024** Contains approved leave requests, including leave type, date range, and posting timestamps.

3. **Pending Leaves – June 2024** Includes leave requests submitted but not yet approved or declined.

4. **Consolidated Absent Sheet – June 2024** Summaries of absences, off-days, holidays, and working-hour irregularities.

5. **Invalid Entries – June 2024** Contains entries with missing or incorrect clock-in/out times.

Each dataset contains valuable but incomplete information. Significant cleaning, normalizing, and restructuring was required before these files could be used in analysis. A multi-stage cleaning pipeline was developed to convert raw HR logs into structured analytical datasets.

4 Data Cleaning and Preparation

The raw datasets were not directly usable due to inconsistent formats, decimal date encodings, missing values, and irregular timestamp structures. A systematic cleaning pipeline was implemented in R to ensure consistency and analytical validity.

4.1 4.1 Column Standardization

All datasets were processed using the `janitor::clean_names()` function to normalize inconsistent column names. Duplicate header rows appearing inside the datasets were removed manually, and the first valid row was set as the standardized header set.

4.2 4.2 Decimal Date and Time Conversion

Excel frequently stores dates and timestamps in decimal format. A custom R function was created to convert these into human-readable formats:

```
decimal_to_date <- function(decimal_date) {
    as.Date(decimal_date, origin = "1899-12-30")
}
```

Similarly, decimal working-hour values were converted into HH:MM format.

4.3 4.3 Missing Value Treatment

Rows containing structural NA values (e.g., missing employee identifiers) were removed. Non-critical NA fields (e.g., missing exit timestamps) were imputed with zeroed or neutral values.

4.4 4.4 Attendance Field Normalization

Daily attendance fields originally contained textual flags, embedded shift information, or invalid markers. These were converted into structured categories:

- Present
- Absent
- Holiday
- Off-Day
- Approved Leave

- Unapproved Leave
- Invalid Entry

This normalization step allowed for downstream aggregations.

4.5 4.5 Working Hours Reconstruction

Daily working hours were not explicitly available and had to be computed from two values stored in a single cell (time in, time out):

```
calculate_working_hours <- function(time_str) {
    times <- strsplit(time_str, "\n")[[1]]
    time_in <- hms::as_hms(paste0(times[1], ":00"))
    time_out <- hms::as_hms(paste0(times[2], ":00"))
    as.numeric(difftime(time_out, time_in, units = "hours"))
}
```

Irregular overnight shifts were adjusted by adding 24 hours where needed.

4.6 4.6 Master Dataset Construction

All datasets were merged using employee identifiers (`Emp Index`, `Emp ID`). The final master dataset contained:

- Attendance counts (present, absent)
- Total working hours
- Invalid entries
- Off-days and holidays
- Approved and unapproved leave totals
- Department and position information

This unified dataset formed the analytical foundation used in subsequent sections.

5 Analysis and Findings

This section presents the analytical results derived from the unified master dataset. The purpose of the analysis is to quantify workforce performance, evaluate punctuality, assess departmental and locational disparities, and identify patterns that influence operational efficiency across SIEHS.

The findings integrate descriptive statistics, aggregated KPIs, and visualization outputs that reveal organizational strengths and inefficiencies. All metrics in this section correspond to the consolidated June 2024 HR dataset.

5.1 5.1 Summary Workforce Metrics

The organization's overall attendance and performance indicators are summarized below:

- **Total Employees Analyzed:** 1,790
- **Attendance Percentage:** 91.24%
- **Absent Percentage:** 8.76%
- **Annual Leaves (Approved):** 2,384
- **Invalid Time Entries:** 478

These metrics indicate a generally strong attendance rate across the organization. However, the high number of invalid entries and substantial leave volume suggests operational inefficiencies, uneven scheduling practices, and possibly inconsistent HR data entry.

5.2 5.2 Dashboard Overview

The main Power BI dashboard used in this analysis consolidates all workforce KPIs into a single visualization. It provides leadership with real-time clarity on attendance behavior, departmental performance, and locational differences.

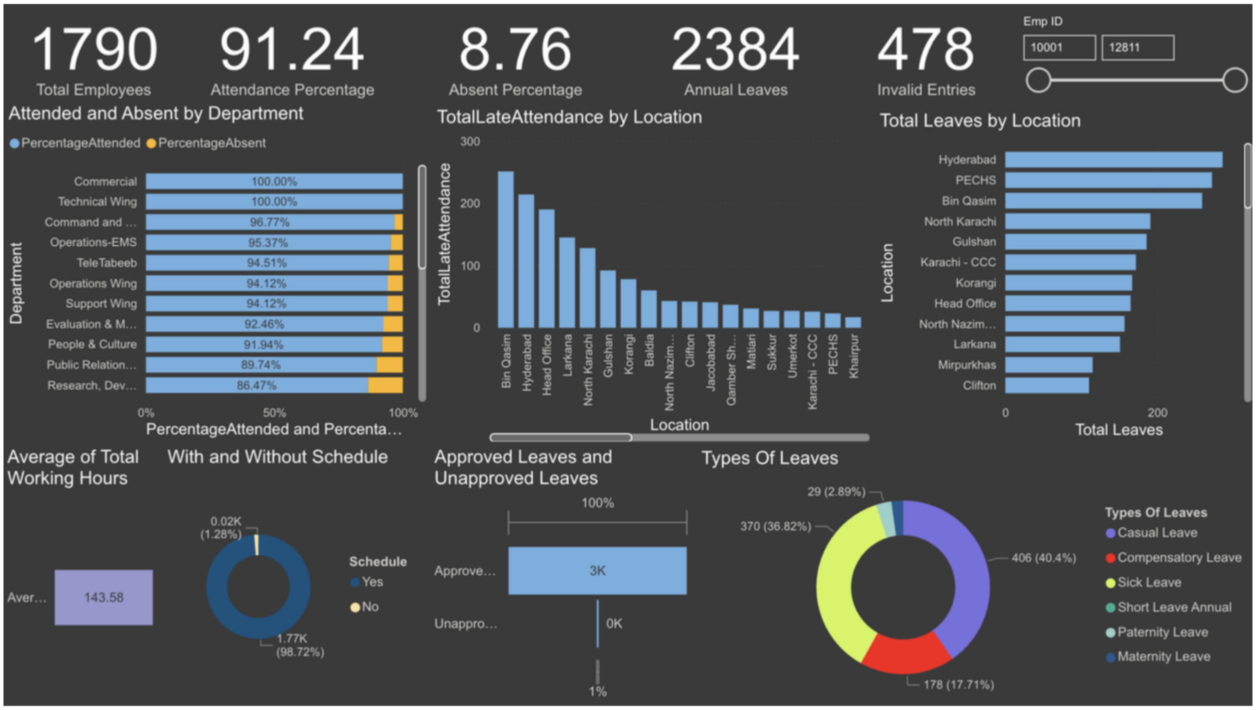

Figure 1: SIEHS HR Analytics Dashboard

The dashboard highlights critical workforce dynamics:

- High attendance but uneven distribution across departments.
- Large variances in late arrivals by location.
- Highly imbalanced leave distribution, with a few locations responsible for a majority of leave days.
- Very low proportion of “unscheduled” shifts, suggesting good rostering compliance.

5.3 5.3 Attendance and Absentee Patterns

Attendance rates were analyzed across departments and locations.

Department-Level Insights:

- The majority of departments recorded attendance rates between 90–100%.
- Commercial, Technical Wing, and Command & Control achieved the highest punctuality (96–100%).
- Research & Development recorded the lowest attendance at $\approx 86\%$.

The variation suggests differing workload distribution, scheduling rigidities, and managerial oversight across units. Consistently high attendance in operational departments indicates stronger enforcement of HR policies.

Location-Level Insights:

- Hyderabad, Bin Qasim, and PECHS recorded the highest volumes of late attendance.
- Locations such as Mirpurkhas and Clifton recorded comparatively lower late arrivals.

This indicates geographic operational challenges—traffic, shift timing, and regional workload patterns likely contribute to these disparities.

5.4 5.4 Invalid Entries and Data Quality Issues

The dataset contained 478 invalid attendance entries for June 2024. Invalid entries include:

- Missing time-in or time-out values.
- Duplicate entries.
- Entries containing unusable or ambiguous codes.

Invalid entries were aggregated per employee using the following logic:

```
invalid_entries_count <- invalid_entries_clean4 %>%
    group_by(`Emp Index`) %>%
    summarise(Invalid_Entries = n(), .groups = 'drop')
```

High invalid-entry counts often indicate:

- Device malfunction or inconsistent biometric records.
- Training gaps in timekeeping procedures.
- Manual intervention or corrections required by HR.

These errors directly affect payroll accuracy and monthly compliance reporting.

5.5 5.5 Working Hours Distribution

Working hours were computed for each employee by reconstructing time-in and time-out pairs. The average monthly working hours across all employees was:

143.58 hours per employee

However, individual variance was large. Some employees recorded fewer than 80 hours, while others exceeded 200 hours due to double shifts or emergency deployments.

This reinforces the operational reality of SIEHS: workload fluctuates based on emergency call volume, station demands, and shift availability.

5.6 5.6 Leave Analysis

Total approved leaves: **2,384** Most used leave types:

- Casual Leave (406 instances)
- Sick Leave (370 instances)
- Short Leave Annual (178 instances)

Less frequent:

- Compensatory Leave
- Paternity Leave
- Maternity Leave

The distribution is operationally aligned with organizational expectations—casual and sick leaves dominate due to emergency response workloads and variable shift pressure.

5.7 5.7 Scheduling Compliance

Only 1.28% of employees were flagged as "without schedule" (irregular or unassigned shifts).

This indicates:

- Effective roster management
- Minimal unscheduled duty occurrences
- Generally stable shift-planning practices

5.8 5.8 Summary of Key Findings

- Workforce attendance is strong overall but varies significantly by department and location.
- Some districts consistently record higher late arrivals, suggesting geographic operational constraints.
- Invalid entries represent a substantial bottleneck in HR data accuracy.
- Leave balances are heavily skewed toward casual and sick leave types.
- The vast majority of employees follow schedule assignments without deviation.

These findings highlight the operational strengths of SIEHS while also identifying opportunities to improve HR monitoring systems and workforce reliability.

6 Discussion

The analytical results from the June 2024 workforce dataset reveal several structural patterns that directly influence operational performance at SIEHS. This section interprets the findings in a broader organizational context, connecting patterns in attendance, lateness, scheduling, and leave behavior to systemic challenges and actionable insights.

6.1 6.1 Workforce Reliability and Operational Readiness

With an overall attendance rate of 91.24%, workforce participation remains strong. For an emergency and health services organization operating 24/7, consistent staffing is essential for service continuity. The data suggests that SIEHS maintains a stable and predictable labor force particularly within operational units such as EMS, Command & Control, and Operations Wing.

However, the eight-point spread between the highest-attending departments (100%) and the lowest (86%) signals structural imbalances. Departments with high on-ground operational responsibilities tend to maintain stronger attendance discipline, while knowledge-based or administrative units show greater variability. This pattern is typical in emergency-services institutions where direct service teams operate under tightly supervised shifts, whereas back-office units may operate under flexible routines.

6.2 6.2 Geographic Constraints and Late Attendance

The location-level analysis shows substantial differences in late attendance. For example:

- Bin Qasim, Hyderabad, and Head Office recorded the highest late arrivals.
- Smaller districts such as Clifton, Mirpurkhas, and Larkana reported minimal lateness.

Several structural factors may explain this:

1. **Urban congestion:** High-density locations face predictable delays due to traffic patterns and commute times.
2. **Shift synchronization:** Locations with round-the-clock operations may experience shift overlaps leading to timing inconsistencies.
3. **Workload pressure:** Busier centers may push employees toward arrival-time fatigue, increasing late entries over time.

This implies that HR interventions should be location-specific rather than organization-wide. For example, targeted transport stipends or flexible shift-start buffers may reduce lateness in affected zones.

6.3 6.3 Data Quality and Invalid Entries

The presence of 478 invalid entries represents a major data-quality concern. Invalid entries disrupt payroll, weaken HR analytics accuracy, and require manual correction. The distribution of invalid entries (often concentrated in small employee clusters) suggests three primary root causes:

1. **Device or biometric malfunction** at specific work sites.
2. **Manual overwriting of time stamps** by supervisors or HR clerks.
3. **Failure in routine attendance protocols** (e.g., missed check-in/out).

Recurring invalid entries indicate that technical and procedural improvements are needed. Enhanced biometric auditing or digital attendance-verification workflows could significantly reduce downstream errors.

6.4 6.4 Leave Behavior and Workforce Fatigue

Leave distribution is heavily skewed toward casual and sick leaves. While this is operationally expected, the volume—2,384 leave entries—suggests cumulative fatigue across certain units. Casual leave spikes often correlate with:

- High emergency-call volumes
- Consecutive long shifts
- Ramadan/Eid seasonal effects

Additionally, employees with high absenteeism also tend to accumulate more late arrivals, implying latent burnout or inconsistent shift alignment. Reviewing workloads and rebalancing shift rotations may reduce chronic absenteeism in affected clusters.

6.5 6.5 Scheduling Compliance and Workforce Discipline

Only 1.28% of employees lacked a formal schedule, indicating strong planning discipline. This is consistent with emergency-service environments where staffing precision is non-negotiable. The limited number of unscheduled shifts suggests efficient coordination between:

- HR
- Shift supervisors
- Operations command center

However, the small set of employees flagged for “No Schedule” need targeted review. These typically represent:

- Unassigned floaters
- Staff awaiting deployment
- Employees in temporary administrative states

Clarifying their scheduling structures would improve reporting transparency and KPI accuracy.

6.6 6.6 Interpretation of Departmental Efficiency

The sharp departmental attendance differences likely reflect:

1. **Management style variations**—departments with stricter supervision outperform others.
2. **Shift rigidity differences**—units with fixed shift times maintain better attendance than flexible ones.
3. **Workload imbalance**—units under high operational load may experience higher absence rates due to fatigue.

To achieve uniform performance, HR must implement standardized attendance protocols and supervise department-level adherence.

6.7 6.7 Organizational Implications

The results have several direct implications for SIEHS:

- High attendance ensures reliable emergency coverage.
- Geographic attendance disparities signal structural commute issues.
- Invalid entries highlight the need for digital systems reinforcement.
- Leave clustering indicates accumulating fatigue and departmental pressure.
- Strong schedule compliance validates HR’s operational planning frameworks.

Overall, the workforce is stable and reliable, but targeted optimizations can significantly increase operational resilience and HR reporting precision.

7 Conclusion

This analysis provides a structured and data-driven examination of workforce attendance, punctuality, scheduling behavior, and leave utilization across SIEHS for June 2024. By consolidating data from five independent Excel sources and constructing a unified master dataset, the report delivers a clear view of organizational performance patterns.

SIEHS demonstrates strong attendance and scheduling discipline across most departments and locations. Nonetheless, certain locations experience disproportionately high late arrivals, and a small number of employees generate a significant share of invalid entries. Leave utilization patterns show operational fatigue in high-pressure roles, emphasizing the need for proactive workforce well-being strategies.

The insights indicate that SIEHS is a well-structured and operationally stable emergency-health organization. Improvements in data quality controls, targeted interventions for high-lateness locations, and deeper load-balancing across units would further enhance workforce efficiency.

The results of this analysis equip HR, operations leadership, and senior management with actionable intelligence to refine scheduling, reduce operational bottlenecks, improve staff readiness, and maintain high-performance workforce standards across all regions.

8 Appendix: Selected R Code Snippets

This appendix provides key portions of the R code used during dataset cleaning, transformation, and feature extraction. The snippets illustrate the core logic that enabled the creation of a unified master dataset from multiple Excel files. Only essential blocks are included here for clarity; the full script can be made available upon request.

8.1 8.1 Importing and Inspecting Raw Files

```
library(readxl)
library(dplyr)
library(janitor)

approved_leaves <- read_excel("Approved␣Leaves␣-␣June␣2024.xlsx")
attendance_sheet <- read_excel("Attendance␣Sheet␣-␣June␣2024.xlsx")
invalid_entries <- read_excel("Invalid␣Entries␣-␣June␣2024.xlsx")

# Inspect structure and missing values
str(approved_leaves)
colSums(is.na(attendance_sheet))
```

This step validates that each dataset imported correctly and identifies data-quality issues before cleaning.

8.2 8.2 Standardizing Column Names

```
approved_leaves <- clean_names(approved_leaves)
attendance_sheet <- clean_names(attendance_sheet)
invalid_entries <- clean_names(invalid_entries)
```

Standardized column naming ensures consistency across all merged datasets.

8.3 8.3 Converting Excel Decimal Dates

```
decimal_to_date <- function(decimal_date) {
  as.Date(decimal_date, origin = "1899-12-30")
}

approved_leaves <- approved_leaves %>%
  mutate(from_date = decimal_to_date(as.numeric(from_date)))
```

Excel stores dates as serialized integers; this conversion is essential for correct time-based analysis.

8.4 8.4 Cleaning Invalid Time Entries

```
invalid_entries_clean <- invalid_entries %>%
  select(-1) %>%               # Remove junk first column
  filter(!is.na(emp_name))     # Remove malformed rows

invalid_entries_clean <- invalid_entries_clean %>%
  mutate(time_in = ifelse(is.na(time_in), "00:00", time_in),
         time_out = ifelse(is.na(time_out), "00:00", time_out))
```

This ensures all invalid entries remain analyzable for counting and aggregation.

8.5 8.5 Transforming Attendance to Long Format

```
attendance_long <- attendance_sheet %>%
  pivot_longer(cols = `1_june`:`30_june`,
               names_to = "date",
               values_to = "status")
```

Long format allows daily-level aggregation of presence, absence, leave type, and total working hours.

8.6 8.6 Computing Working Hours

```
calculate_working_hours <- function(time_str) {
  if (grepl("\n", time_str)) {
    times <- strsplit(time_str, "\n")[[1]]
    time_in <- hms::as_hms(paste0(times[1], ":00"))
    time_out <- hms::as_hms(paste0(times[2], ":00"))
    duration <- as.numeric(difftime(time_out, time_in, units="hours"
       ))
    if (duration < 0) duration <- duration + 24
    return(duration)
  }
  return(0)
}

attendance_long <- attendance_long %>%
  mutate(working_hours = sapply(status, calculate_working_hours))
```

This logic reconstructs daily working hours even when shifts span past midnight.

8.7 8.7 Aggregating KPI Metrics

```
attendance_summary <- attendance_long %>%
  group_by(emp_index) %>%
  summarise(
    attended = sum(status != "Absent"),
    absent = sum(status == "Absent"),
    invalid_entries = sum(status == "Invalid"),
    total_working_hours = sum(working_hours, na.rm = TRUE)
  )
```

This block forms the foundation of the master dataset by calculating employee-level performance indicators.

8.8 8.8 Integrating Leave Data

```
leave_summary <- approved_leaves %>%
  group_by(emp_index, leave_type) %>%
```

```
  summarise(days = sum(as.numeric(leave_days))) %>%
  pivot_wider(names_from = leave_type,
              values_from = days,
              values_fill = 0)

master <- master %>%
  left_join(leave_summary, by = "emp_index")
```

Leave types are aggregated and spread into separate columns for deeper analysis.

8.9 8.9 Final Master Dataset Assembly

```
master_sheet <- attendance_summary %>%
  left_join(attendance_sheet %>% select(emp_index, emp_id, employee_
     name),
            by = "emp_index") %>%
  left_join(leave_summary, by = "emp_index")
```

This produces the final dataset used for analysis, visualization, and dashboard creation.

Note: The full codebase, including all cleaning routines, pivot transformations, and validation logic, is preserved separately and can be shared if required for audit or replication.