ALI HYDER

Data Quality Audit Report

Table of Contents

CONTENTS

1. EXECUTIVE OVERVIEW

This report evaluates the operational dataset provided by CloudOps Analytics GmbH, focusing on data quality assessment, descriptive statistics, correlation analysis, and key operational insights.

The dataset contains 5,000 cloud project records covering budget allocation, usage cost, resource utilization, tagging completeness, and incident-related metrics.

Overall, the dataset demonstrates strong structural integrity with limited inconsistencies. While financial alignment appears stable, opportunities exist for improving budget control and operational efficiency.

2. DATA QUALITY ASSESSMENT

A structured data audit was conducted to evaluate completeness, consistency, and validity.

Identified Issues and Fixes

- The "Date" column required format correction from text to a valid date format to ensure analytical reliability.
- Budget and cost fields were reformatted to standardized currency format ($) to ensure numerical consistency and accurate financial interpretation during analysis.
- Data integrity checks were conducted using Excel-based validation formulas.
- Duplicates were identified using COUNTIF()
- Missing values were quantified using COUNTBLANK()
- Reversed dates were detected using conditional comparisons (IF(Start_Date > End_Date))
- Negative financial entries were validated using logical tests such as IF(Value < 0).
- This structured approach ensured reliable and transparent quality assessment.
- Fully empty columns (e.g., Environment, Department, Cloud Provider) suggest incomplete metadata capture, limiting governance visibility.
- Overall, the dataset is analysis-ready after cleaning and validation.

Key Findings

Duplicate records	0
Fully empty columns	9
Partially missing columns	3 (Usage_Cost, Resource_Utilization, Tagging_Completeness)
Negative financial values	None
Reversed date entries	None
Projects exceeding allocated budget	509 (10%)

3. DESCRIPTIVE & STATISTICAL INSIGHTS

Financial Overview

Average Usage Cost	$2,034
Average Budget Consumed	$2,004
Average Budget Utilization	81%
Budget Utilization Range	54%–119%
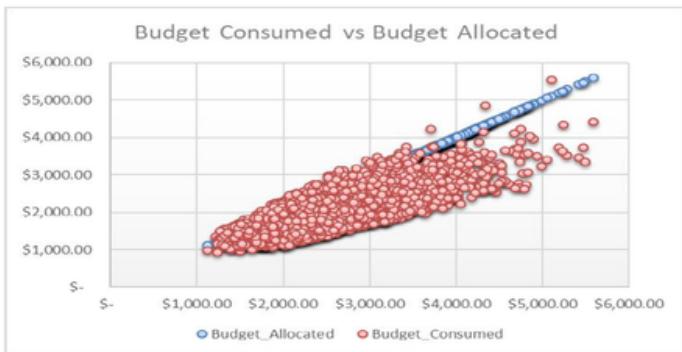
" *Most projects operate within planned budgets; however, approximately 10% exceed allocation.* "

Operational Overview

Average Resource Utilization	70%
Average Tagging Completeness	90%

" *Tagging governance appears strong, while resource utilization shows moderate variability, indicating potential optimization opportunities.* "
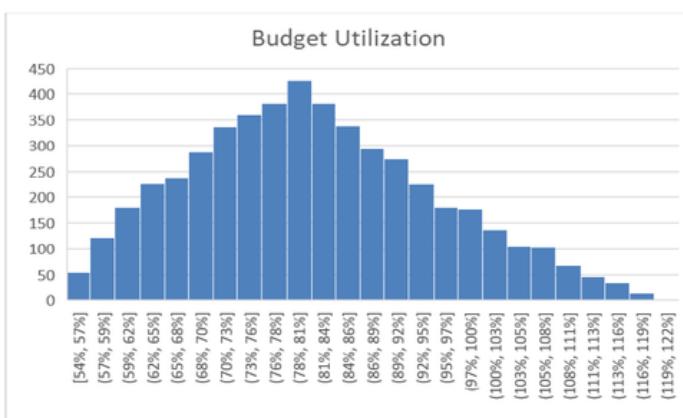
4. VISUAL INSIGHTS (CHARTS)

Budget Consumed vs Budget Allocated

1. Budget Allocated vs Budget Consumed (Scatter Plot)

The scatter plot demonstrates a clear positive linear relationship between allocated and consumed budgets. Most projects remain close to planned allocation, though a visible cluster exceeds 100% utilization.
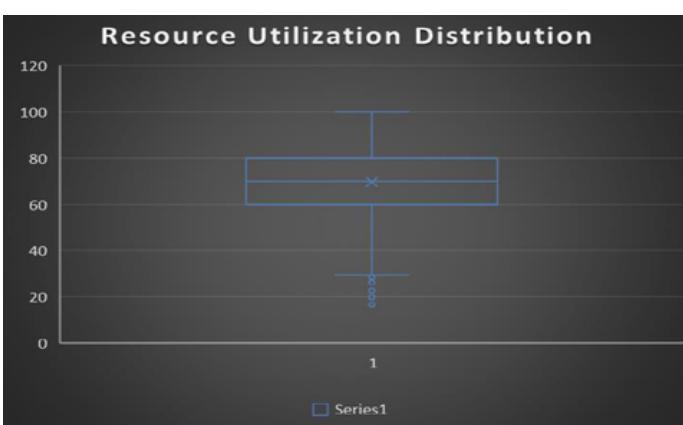
Insight: Financial tracking is consistent but requires monitoring for over-budget cases.

Budget Utilization

2. Budget Utilization Distribution (Histogram))

The majority of projects fall between 70%–95% utilization. A smaller right-tail distribution exceeds 100%.
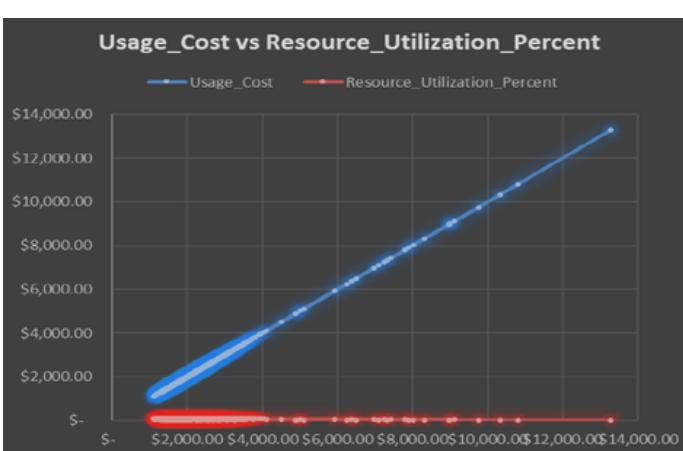
Insight: Spending control is generally effective; however, early-warning mechanisms could reduce overruns.

Resource Utilization Distribution

3. Resource Utilization Distribution (Boxplot)

Median utilization is approximately 70%, with lower-end outliers.

Insight: Some projects underutilize resources, indicating potential cost inefficiencies.

Usage_Cost vs Resource_Utilization_Percent

4. Usage Cost vs Resource Utilization (Scatter)

The relationship between cost and utilization is weak. Higher costs do not necessarily correspond to higher operational efficiency.

Insight: Cost drivers may be influenced more by pricing models than actual utilization levels.

5. CORRELATION SUMMARY

Relationship	Correlation	Interpretation
Usage_Cost vs Budget_Consumed	0.64	Moderate Positive
Budget_Consumed vs Budget_Utilization	0.3	Weak Positive
Usage_Cost vs Resource_Utilization	-0.1	Very Weak Negative
Tagging vs Resource_Utilization	0.02	No Meaningful Relationship

Financial variables show logical alignment, while operational efficiency metrics demonstrate limited dependency on spending levels.

6. RECOMMENDATIONS

Based on the analysis, the following improvements are recommended:

1. Implement Budget Early-Warning Alerts
Introduce automated alerts at 85–90% utilization to prevent budget overruns.

2. Improve Metadata Governance
Enforce mandatory tagging for currently empty columns such as Environment, Department, and Cloud Provider to strengthen accountability and reporting transparency.

3. Optimize Resource Allocation
Identify underutilized projects (below 50% utilization) and reallocate resources where possible.

4. Strengthen Cost-Efficiency Monitoring
Develop dashboards linking cost per utilization unit to identify inefficiencies beyond raw spending levels.

7. CONCLUSION

CloudOps Analytics GmbH demonstrates strong financial tracking and structured operational reporting. The dataset shows high integrity with minimal structural inconsistencies.

However, budget overruns in 10% of projects and underutilized infrastructure indicate opportunities for improved financial forecasting and operational optimization.

With enhanced governance enforcement and proactive monitoring mechanisms, CloudOps can further strengthen cost efficiency and operational performance.