CBSE Class 12 (2022): Attendance, Data Credibility & Risk Analysis

A reproducible Excel + Power Pivot analysis with integrity checks and coverage-weighted insights

Ritu Verma

Aspiring Data Analyst| Excel, SQL, Python| Reproducible Analytics

Executive Summary

Key Findings

- Overall appearance rate is high (~99.44), indicating strong system-level attendance.
- 99.96% of statistical weight comes from VALID_BASE row counts, making conclusions highly reliable.
- Anomalies affect only 0.14% of coverage, concentrated in a very small number of large-base cases.
- Structural zeros (22 cases) reflect institutional presence, not data errors.

Why this matters?

- Decisions based on this dataset are credibility safe.
- Risks are localized, not systemic.
- Over-reaction to small or perfect segments is avoided through base-aware modelling.

Dataset Overview

Scope

- Examination: CBSE Class 12
- Year: 2022
- Grain: Region x School Type
- Metrics: Registered, Appeared

What is not included?

- No student-level data
- No demographic or performance score
- No causality claims

This is a descriptive, integrity-aware operational analysis.

Methodology Overview

Workflow Architecture

1. Raw Data (01_DATA_MODEL_raw_wide): untouched source
2. Canonical Fact Table (02_FACT_APPEARANCE (VIEW)): all row-level logic
3. Integrity Audit Log (03_DATA_QUALITY_LOG (VIEW)): validation & diagnostics
4. Credibility & Coverage Analysis (04_CREDIBILITY)
5. Anomalies & Risk Analysis (05_ANALYSIS)
6. Volatility Analysis (06_VOLATILITY)
7. Executive Dashboard (07_DASHBOARD)
8. Assumptions & Thresholds (99_NOTES)

Design Principles

- Logic once, reuse everywhere
- Weight > Frequency
- Absence ≠ Failure
- Empty cells ≠ Missing data/error
- Thresholds must be documented

Data Integrity & Validation

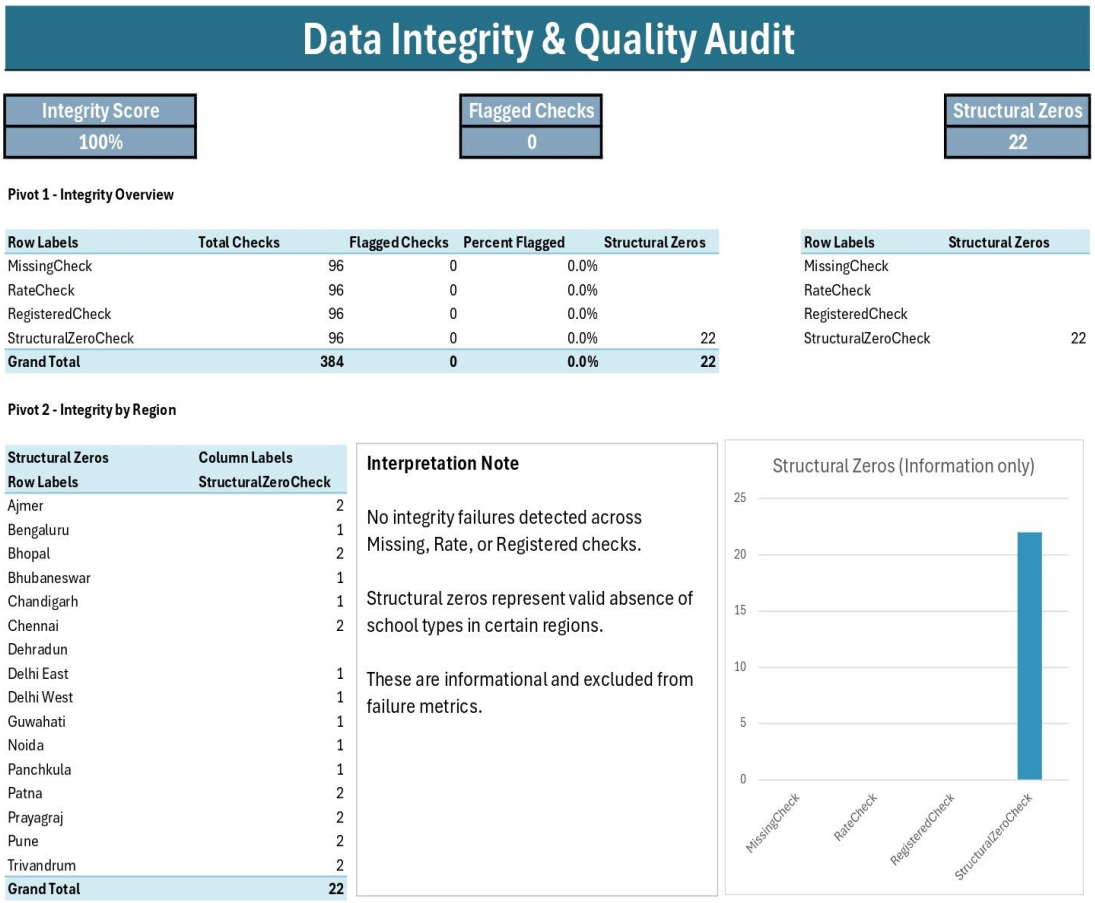
Integrity Checks Implemented

- Missing value checks
- Rate validity checks
- Registered vs Appeared consistency
- Structural zero identification

Results

- 0 integrity failures
- 22 structural zeros (valid absence of school types)
- Integrity Score: 100%

Structural zeros are informational, not excluded as errors.



Data Integrity & Quality Audit

Integrity Score	Flagged Checks	Structural Zeros
100%	0	22

Pivot 1 - Integrity Overview

Row Labels	Total Checks	Flagged Checks	Percent Flagged	Structural Zeros
MissingCheck	96	0	0.0%	
RateCheck	96	0	0.0%	
RegisteredCheck	96	0	0.0%	
StructuralZeroCheck	96	0	0.0%	22
Grand Total	**384**	**0**	**0.0%**	**22**

Row Labels	Structural Zeros
MissingCheck	
RateCheck	
RegisteredCheck	
StructuralZeroCheck	22

Pivot 2 - Integrity by Region

Structural Zeros	Column Labels
Row Labels	StructuralZeroCheck
Ajmer	2
Bengaluru	1
Bhopal	2
Bhubaneswar	1
Chandigarh	1
Chennai	2
Dehradun	
Delhi East	1
Delhi West	1
Guwahati	1
Noida	1
Panchkula	1
Patna	2
Prayagraj	2
Pune	2
Trivandrum	2
Grand Total	**22**

Interpretation Note

No integrity failures detected across Missing, Rate, or Registered checks.

Structural zeros represent valid absence of school types in certain regions.

These are informational and excluded from failure metrics.

Base Stability & Credibility

Base Classification

- ABSENT_BASE: Registered = 0
- SMALL_BASE: Registered < 100
- VALID_BASE: Registered > 100

Coverage Impact

- VALID_BASE rows: 65 out of 96
- VALID_BASE coverage: 99.96%
- SMALL_BASE rows appear frequently but carry negligible weight (~0.04%)

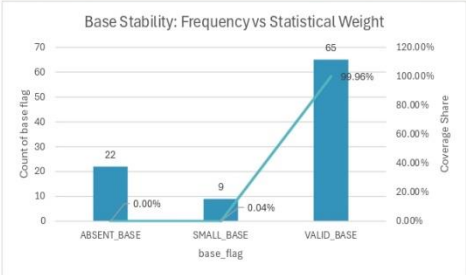
Statistical conclusions are driven by stable denominations.

Data Credibility & Coverage Weight

Usable Coverage Share	Valid Base Rows	Total Rows
99.96%	65	96

base_flag	Count of base_flag	Total Registered	Coverage Share
ABSENT_BASE	22	0	0.00%
SMALL_BASE	9	590	0.04%
VALID_BASE	65	14,38,227	99.96%
Grand Total	**96**	**14,38,817**	**100.00%**

Row Labels	Count of base_flag	Coverage Share
ABSENT_BASE	22	0.00%
SMALL_BASE	9	0.04%
VALID_BASE	65	99.96%
Grand Total	**96**	**100.00%**



Interpretation Note

VALID_BASE rows dominate statistical weight even if SMALL_BASE appears frequently, preventing fragile denominators from distorting conclusions.

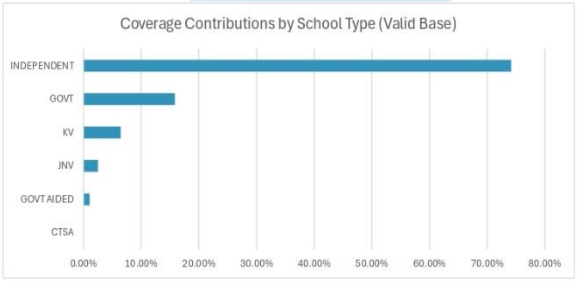
Analysis conclusions are driven primarily by VALID_BASE observations (99.96% of coverage).

base_flag	All

School Type	Total Registered	Coverage Share
CTSA	398	0.03%
GOVT	2,27,773	15.83%
GOVT AIDED	15,295	1.06%
INDEPENDENT	10,66,827	74.15%
JNV	35,946	2.50%
KV	92,578	6.43%
Grand Total	**14,38,817**	**100.00%**

base_flag	All

School Type	Coverage Share
CTSA	0.03%
GOVT AIDED	1.06%
JNV	2.50%
KV	6.43%
GOVT	15.83%
INDEPENDENT	74.15%
Grand Total	**100.00%**



Threshold Justification (Why 100 and 0.98?)

SMALL_BASE = 100

Evidence:

- Median non-zero registrations ≈ 4000 (≈4067.5)
- Only 9 rows below 100
- Below-100 bases show disproportionate volatility

Decision:

100 balances stability and coverage without over-excluding data.

LOW_RATE = 0.98

Evidence:

- Appearance rates cluster tightly near 1.00
- Only 2 rows below 98%
- Below-98% lies in extreme lower tail

Decision:

0.98 isolates statistically uncommon behavior without flagging normal variation.

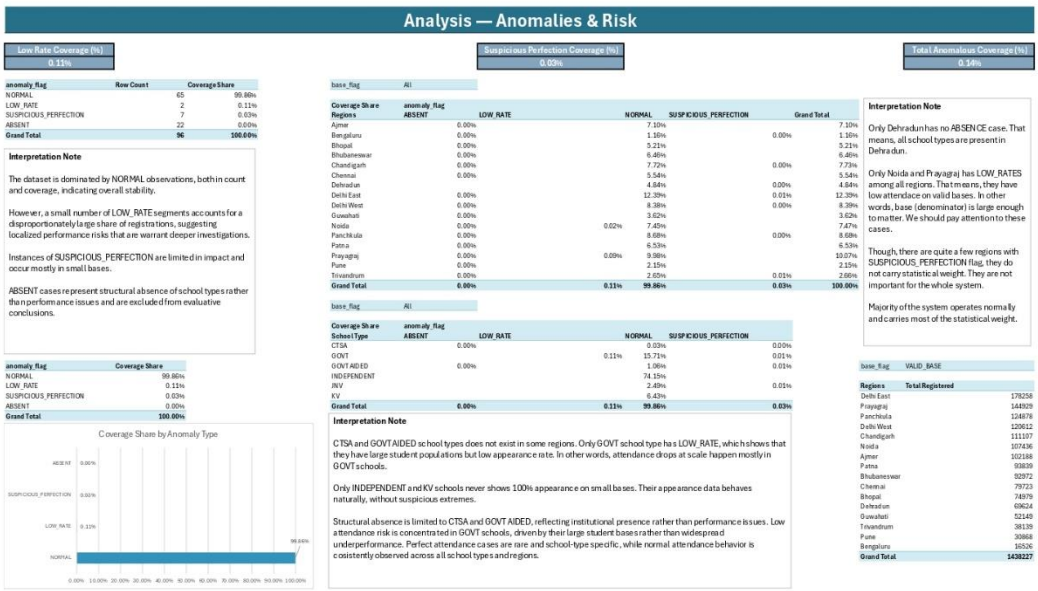
Anomalies & Risk Analysis

Anomaly Types

- NORMAL: Expected behavior
- LOW_RATE: Statistically uncommon under-attendance
- SUSPICIOUS_PERFECTION: Perfect rates on constrained bases
- ABSENT: Structural absence

Key Insights

- NORMAL dominates 99.86% coverage
- LOW_RATES affects only 0.11% coverage, but on large bases; deserves attention.
- Perfect attendance cases are rare and base-sensitive
- Seven small base “100%” cases: Bengaluru GOVT (70), Chandigarh GOVT AIDED (55), Dehradun GOVT AIDED (22), Delhi East JNV (81), Delhi West JNV (65), Panchkula CTSA (38), Trivandrum GOVT (82).
- ABSENT cases are excluded from evaluative conclusions
- Dehradun is uniquely complete (no structural absence)
- A large-base perfect case is normal, not suspicious. Example, Delhi West KV (2,872/2,872 = 100%) occurs on a valid base and is treated as NORMAL in the quality log.



Regional & School-Type Patterns

Regional

- LOW_RATE concentrated in very few regions
- Majority of regions operate normally
- Top 3 regions by registrations: Delhi East (178,339), Prayagraj (144,929), Panchkula (124,916). These three alone accounts for 31.15% of total registrations—prime targets for performance programs.

School Type

- GOVT school dominate both volume and risk (coverage = 15.83%)
- Exactly 2 cases fall below 98% appearance threshold, both GOVT: Prayagraj and Noida. Together they represent 0.11% of total registrations—large base and high risk.
- INDEPENDENT schools dominate coverage (74.15%) but show stable behavior
- Changes in INDEPENDENT or large GOVT clusters will impact overall metrices most.
- CTSA & GOVT AIDED structural absences reflect institutional presence

Risk follows scale, not school type quality.

Volatility Analysis

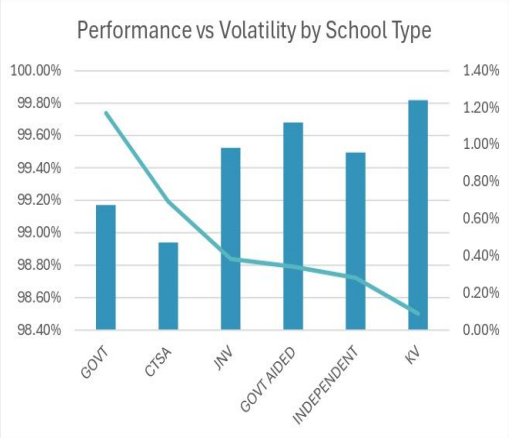
- Appearance rates are consistently high across school types
- Volatility decreases as base size increases
- High performance + low volatility = operational stability
- Volatility by school type shows KV is the most stable (std dev ≈ 0.09%) and GOVT least stable (std dev ≈ 1.17%).

Attendance Volatility by School Type

Overall Appearance Rate
99.44%

base_flag	All

School Type	Avg Appearance Rate	Appearance Rate Std Dev
GOVT	99.17%	1.17%
CTSA	98.94%	0.69%
JNV	99.52%	0.38%
GOVT AIDED	99.68%	0.34%
INDEPENDENT	99.50%	0.28%
KV	99.82%	0.09%
Grand Total	**99.49%**	**0.67%**



Interpretation Note

School types with high average appearance rates but low standard deviation indicate stable and predictable attendance behavior.

Higher volatility (standard deviation) suggests uneven attendance performance across regions, even if the average appears acceptable.

This helps distinguish consistently strong systems from those whose performance depends on local conditions.

In the chart, line represnts risks, while column represents performance.

What Should Be Acted On

Focus Areas

- Large base LOW_RATE segments
- GOVT school operational scale risks

What NOT to over-react to

- Perfect attendance on small bases
- Structural zeros

Limitations & Boundaries

- No casual interference
- No student-level diagnostics
- Thresholds are context-specific
- Results apply to 2022 snapshot only

Reproducibility & Auditability

- All logic documented
- Thresholds justified
- Measures separated from data
- Workbook fully slicer-driven