CLAIMS LEAKAGE & RECOVERY ANALYTICS

(General Insurance — Motor Claims)

0. EXECUTIVE OVERVIEW

Industry: General Insurance (Motor)
Company Type: Mid-size insurer
Role: Senior Data Analyst (Claims & Finance Analytics)
Stakeholders: CFO, Head of Claims, Risk & Operations

0.1 Project Background

Identify, quantify, and reduce claims leakage by analyzing internal claims, policy, payment, and provider data—then translate insights into recoverable cash, process fixes, and management controls.

0.2 Business Objective

- Detect overpayments, ineligible claims, and process gaps
- Quantify financial leakage (monetary value) and recovery potential
- Support CFO-level decisions with evidence-based insights

0.3 Scope & Assumptions

- Analysis covers Motor insurance claims only (auto/vehicle incidents)
- Dataset treated as sanitized internal operational data
- Conservative rule-based validation prioritized over aggressive detection
- Policy coverage limits (policy_csl) are encoded as categorical ranges

0.4 Tools & Technologies

- **SQL** (PostgreSQL / MySQL equivalent)
- **Data Warehouse concepts** (Fact–Dimension modeling)
- **BI Dashboard** (Power BI / Tableau / Looker-style)
- **Documentation** (internal analytics & stakeholder documentation)

0.5 Expected Business Impact

This analytics project is designed to enable **measurable financial and operational impact** by systematically identifying claims leakage across Motor insurance portfolios.

Based on preliminary data characteristics and industry benchmarks, the expected impacts are outlined as follows:

a) Claims Leakage Identification & Quantification

This project is designed to:

- Quantify **total claims leakage value** embedded in paid claims
- Measure leakage as:
    i) Absolute monetary value
    ii) Percentage of total claims paid
- Distinguish between:
    i) Structural leakage (process & control gaps)
    ii) Recoverable leakage (overpayment, mismatch, coverage breach)

Expected outcome:
Management gains **visibility into previously hidden financial loss** within claims operations.

b) Recovery Opportunity Enablement

Through rule-based validation and anomaly detection, the analysis is designed to:

- Identify claims with **high recovery potential**, such as:
    i) Payments exceeding policy coverage limits
    ii) Component mismatches (injury, property, vehicle)
    iii) Low-severity claims with disproportionate payouts
- Provide a **prioritized recovery list** based on:
    i) Amount
    ii) Likelihood of recovery
    iii) Operational feasibility

Expected outcome:

- Enable identification of short-term cash recovery opportunities, if present
- Improved recovery focus and ROI for claims teams

c) Loss Ratio & Profitability Improvement

Claims leakage directly inflates incurred claims cost.

By addressing identified leakage:

- Loss ratio can be improved without:
    i) Increasing premiums
    ii) Reducing customer benefits
- Financial improvement comes from **better execution**, not pricing changes

Expected outcome:

- Improved underwriting profitability
- Stronger financial control from CFO perspective

d) Operational Control & Process Improvement

The analysis highlights:

- Where current claim validation controls fail
- Which claim types, severities, or processing paths are most vulnerable
- Trade-offs between **SLA speed vs financial accuracy**

Expected outcome:

- More targeted controls
- Reduced rework and post-payment corrections
- Better alignment between Operations, Risk, and Finance

e) Management-Level Decision Support

The final deliverables (SQL outputs & dashboard) enable:

- Regular monitoring of leakage KPIs
- Early warning signals for control breakdowns
- Evidence-based discussions with:
    i) Claims teams
    ii) Providers
    iii) Internal audit

Expected outcome:

- Shift from reactive issue handling to proactive leakage prevention
- Stronger governance over claims spending

Summary Impact Statement

Overall, this project provides capability to:

- Reduce claims leakage through better detection and prevention
- Identify recoverable cash from historical claims, if present
- Strengthen financial discipline without compromising service quality

This positions analytics as a **direct profit-protection function**, not just a reporting tool.

1. BUSINESS CONTEXT & PROBLEM DEFINITION

1.1 Industry Context

In the general insurance industry, claims processing represents the **largest and most volatile cost component**, particularly for Motor insurance products. As claim volumes grow, insurers face increasing operational pressure to balance **speed of settlement (SLA)** with **financial accuracy and control**.

Mid-size insurers are especially exposed to claims leakage due to:

- High dependency on manual or semi-automated claim validation
- Complex policy coverage structures and benefit limits
- Multiple claim components (injury, property, vehicle)
- Fragmented oversight across Claims, Finance, and Operations

As a result, financial leakages often occur **within normal claim processing activities**, without being immediately visible as errors or fraud.

1.2 Business Problem Statement

The company currently lacks a **data-driven mechanism** to systematically identify, quantify, and monitor claims leakage across its Motor insurance portfolio.

While individual claim issues may be detected on a case-by-case basis, there is no consolidated view that answers:

- How much money is leaking from claims payments
- Where the leakage is occurring
- Which portion of leakage is recoverable versus structural

This results in hidden financial losses, inconsistent recovery efforts, and limited management visibility into claim cost drivers.

1.3 Why Claims Leakage Matters

Claims leakage directly increases incurred claims cost and negatively impacts the company's **loss ratio and underwriting profitability**.

Unlike pricing or premium growth initiatives, leakage reduction:

- Improves profitability **without increasing premiums**
- Does not require benefit reductions for customers
- Focuses on execution quality rather than market conditions

Even small percentage leakages, when applied across a large claims portfolio, can translate into **material financial losses** over time.

Therefore, controlling claims leakage represents one of the **highest ROI opportunities** for operational and financial improvement.

1.4 Stakeholders & Their Decisions

Chief Financial Officer (CFO)

Needs visibility into:

- Total financial exposure from claims leakage
- Impact on loss ratio and profitability
- Recovery potential and financial controls

Decisions supported:
- Financial recovery targets
- Control investment prioritization
- Provider or cost structure review

Head of Claims

Needs to understand:
- Where claim validation breaks down
- Which claim types or scenarios are most vulnerable
- Trade-offs between SLA speed and financial accuracy

Decisions supported:
- Process redesign
- Validation rule enhancements
- Team performance focus

Risk & Operations

Needs insight into:
- Policy coverage enforcement gaps
- Operational control weaknesses
- Systemic versus isolated issues

Decisions supported:
- Control framework strengthening
- Governance and audit alignment
- Preventive control implementation

1.5 Key Business Questions

This project is designed to answer the following core questions:
1. How much total claims leakage exists within paid Motor claims?
2. What percentage of total claims payments does this leakage represent?
3. Which claim types and incident severities contribute most to leakage?
4. Where do coverage limits and claim component mismatches occur?
5. Which leakage patterns indicate **recoverable opportunities**?
6. What operational controls could have prevented the identified leakages?

These questions are framed to support **decision-making**, not just reporting.

1.6 Success Criteria

This project is considered successful if it delivers:
- A clear and defensible **quantification of claims leakage** in monetary terms
- Identification of key **leakage drivers and root causes**
- Actionable insights that support **recovery and prevention**
- Executive-ready outputs (SQL metrics & dashboard)
- Business recommendations with **estimated financial impact**

The ultimate success of the project is measured by its ability to **support profitability improvement through better claims execution**, not by technical complexity.

2. DATA UNDERSTANDING & DATA QUALITY

2.1 Data Source Description

The dataset used in this project is sourced from the Mendeley Data repository under the title *Insurance Claims Dataset*. For portfolio and analytical purposes, the dataset is treated as **sanitized internal operational data** from a mid-size general insurance company.

The dataset contains historical information related to:

- Insurance policies
- Insured customers
- Claim incidents
- Claim payment components

Although publicly available, the structure and attributes closely resemble **real-world insurance claims data**, making it suitable for simulating internal claims leakage analysis.

2.2 Dataset Overview

The dataset is provided as a **single flat table**, where each row represents a **paid insurance claim**.

Key characteristics:

- One row per claim case
- Contains both policy-level and claim-level attributes
- Includes financial claim components: Injury claim, Property claim, Vehicle claim
- Includes contextual information: Incident details, Customer demographics, Vehicle attributes

While denormalized, the dataset captures sufficient detail to support **financial validation, leakage detection, and recovery analysis**.

2.3 Table & Field Business Meaning

The table below summarizes key fields and their business interpretation:

Field	Business Meaning
policy_number	Unique policy identifier
policy_bind_date	Policy effective date
policy_csl	Combined Single Limit (coverage limit)
policy_deductable*	Deductible borne by insured
policy_annual_premium	Annual premium value
umbrella_limit	Additional coverage beyond base limit
incident_date	Date of claim incident
incident_type	Type of incident (collision, theft, etc.)
incident_severity	Severity classification of incident
total_claim_amount	Total amount paid for the claim
injury_claim	Injury-related claim component
property_claim	Property damage claim component
vehicle_claim	Vehicle damage claim component
police_report_available	Availability of police report
auto_make, auto_model, auto_year	Vehicle information

Field	Business Meaning
fraud_reported	Post-investigation fraud indicator

**Note: Source data uses 'policy_deductable' spelling, preserved from original Mendeley dataset.*

These fields collectively support **coverage validation**, **cost reasonableness checks**, and **leakage identification**.

2.4 Data Quality Issues

A data quality assessment was conducted prior to analysis to ensure financial accuracy and avoid misleading conclusions.

2.4.1 Missing Values

Several fields contain missing values, particularly:

- collision_type
- police_report_available
- property_damage

Business interpretation:

- Missing values are often **contextual**, not errors
- For example, collision-related fields may be blank for non-collision incidents

Implication:

- Blind imputation may distort business meaning
- Missing values must be evaluated based on incident context

2.4.2 Duplicate Records

The dataset does not contain direct duplicate rows at the claim level. However, there is a **potential analytical risk** of double counting due to:

- Multiple claim components (injury, property, vehicle) within a single record

Risk: Inflated leakage estimation if components are aggregated incorrectly

Assessment:

- Each row is treated as **one claim**
- Component amounts are validated against the total claim amount

2.4.3 Inconsistent Values

Several inconsistencies were identified that may indicate claims leakage:

- total_claim_amount exceeding policy_csl
- Sum of claim components not matching total claim amount
- High claim amounts associated with low incident severity

Business implication:

- Potential overpayment
- Coverage enforcement failure
- Weak validation controls

These inconsistencies are preserved and flagged for analytical investigation rather than corrected.

2.4.4 Outliers

Outliers are observed in:

- total_claim_amount
- vehicle_claim values

Interpretation:

- Some outliers may represent legitimate high-cost incidents
- Others may indicate pricing errors, overvaluation, or control gaps

Outliers are **not removed**, but evaluated using rule-based thresholds aligned with business logic.

2.5 Data Cleaning & Handling Strategy

The following principles guide data handling:

- No overwriting of raw financial values
- Missing values retained when logically applicable
- Inconsistencies flagged using derived indicators
- No statistical imputation for monetary fields
- Business-rule validation prioritized over automated correction

This approach ensures **auditability and trust** in financial results.

2.6 Key Analytical Assumptions

The analysis is based on the following assumptions:

1. Each row represents a **paid claim**
2. total_claim_amount reflects actual cash outflow
3. Fraud labels are informational and not primary leakage drivers
4. Claims leakage includes non-fraud operational issues
5. Coverage limits and deductibles are contractually enforceable

These assumptions are documented to maintain transparency and analytical integrity.

3. DATA MODELING & ANALYTICAL DESIGN

3.1 Modeling Objective

The objective of the analytical data model is to enable **accurate, repeatable, and scalable analysis** of claims leakage and recovery opportunities across Motor insurance portfolios.

Specifically, the model is designed to:

- Quantify claims leakage in monetary terms
- Support root cause analysis across policy, incident, customer, and vehicle dimensions
- Enable aggregation and KPI reporting for management
- Serve as a stable foundation for SQL-based analysis and dashboards

The model prioritizes **financial integrity and business interpretability** over transactional completeness.

3.2 Modeling Approach

A **star schema** approach is adopted for this project.

Rationale:

- Claims leakage analysis is aggregation-heavy (sums, rates, trends)
- Management reporting requires fast, intuitive queries
- Separation of facts and dimensions improves clarity and auditability

This approach mirrors common data warehouse practices used in:

- Claims MIS
- Loss ratio analytics
- Finance and risk reporting within insurance companies

The model intentionally avoids excessive normalization to ensure usability by analysts and business users.

3.3 Fact Table Design

Fact Table: FACT_CLAIMS

Grain Definition: One row represents one paid insurance claim

This grain ensures:

- No double counting of financial amounts
- Clear linkage to policy coverage and claim context
- Accurate aggregation of claim costs and leakage

Key Fields:

Field	Description
claim_id	Surrogate key for claim
policy_id	Foreign key to policy dimension
customer_id	Foreign key to customer dimension
incident_id	Foreign key to incident dimension
vehicle_id	Foreign key to vehicle dimension

Field	Description
date_id	Foreign key to date dimension
claim_type	Motor
total_claim_amount	Total paid claim amount
injury_claim_amount	Injury claim component
property_claim_amount	Property claim component
vehicle_claim_amount	Vehicle claim component
policy_limit	Policy coverage limit
policy_deductible	Deductible amount
coverage_breach_flag	Coverage exceeded indicator
component_mismatch_flag	Component sum mismatch
severity_cost_mismatch_flag	Cost vs severity inconsistency
potential_leakage_amount	Estimated leakage value
recoverable_flag	Recoverable leakage indicator

3.4 Dimension Tables Design

DIM_POLICY

Captures contractual coverage context.

Attribute	Description
policy_id	Surrogate key
policy_number	Business identifier
policy_bind_date	Policy effective date
policy_state	Issuing region
policy_csl	Coverage limit
umbrella_limit	Additional coverage
policy_annual_premium	Annual premium
policy_deductible	Deductible

DIM_CUSTOMER

Captures insured characteristics relevant for segmentation.

Attribute	Description
customer_id	Surrogate key
age	Age
insured_sex	Gender
insured_education_level	Education
insured_occupation	Occupation
months_as_customer	Customer tenure in months
insured_relationship	Relationship to policyholder

DIM_INCIDENT

Provides context for claim occurrence.

Attribute	Description
incident_id	Surrogate key
incident_date	Date of incident
incident_hour_of_the_day	Hour of incident
incident_type	Type of incident
collision_type	Collision detail

Attribute	Description
incident_severity	Severity
incident_state	Location state
incident_city	Location city
police_report_available	Police report indicator

DIM_VEHICLE

Captures insured vehicle information.

Attribute	Description
vehicle_id	Surrogate key
auto_make	Vehicle make
auto_model	Vehicle model
auto_year	Vehicle year

DIM_DATE

Standard calendar dimension enabling time-based analysis.

Attribute	Description
date_id	Surrogate key
date	Calendar date
month	Month
quarter	Quarter
year	Year
weekday	Day of week

3.5 Derived Metrics & Flags

The following indicators are derived using business rules:

- **Coverage Breach Flag**
  total_claim_amount > policy_limit
- **Component Mismatch Flag**
  total_claim_amount ≠ (injury + property + vehicle)
- **Severity–Cost Mismatch Flag**
  Low incident severity with disproportionately high claim amount
- **Potential Leakage Amount**
  Estimated excess payment beyond expected or allowable values
- **Recoverable Flag**
  Indicates whether leakage is likely recoverable based on type and cause

These indicators form the basis of leakage quantification and prioritization.

3.6 Model Limitations

The model has the following intentional limitations:

- Does not perform predictive fraud detection

- Does not incorporate external repair or medical pricing benchmarks
- Assumes one claim per record (no partial payments over time)
- Does not account for legal dispute outcomes

These limitations are acceptable given the project's focus on **financial leakage detection and recovery analytics**.

4. SQL ANALYSIS & BUSINESS LOGIC

4.1 SQL Design Principles

The SQL layer in this project is designed following **production-grade analytics principles** to ensure accuracy, auditability, and reusability.

The key design principles are:

1. **Separation of Concerns** – SQL logic is separated into: Data validation, Business rule implementation, Leakage calculation, KPI aggregation. This prevents mixing validation logic with financial results.
2. **No Hard-Coded Business Assumptions** – Thresholds and rules are derived from policy attributes and claim context rather than arbitrary constants.
3. **Financial Integrity First** – Raw monetary values are never overwritten. All adjustments are expressed as derived fields or flags.
4. **Reproducibility & Auditability** – Queries are deterministic and can be rerun with consistent results, supporting audit and management review.
5. **Business-Readable Logic** – SQL is written with clear naming conventions and comments so that logic can be reviewed by Finance or Risk teams.

4.2 Data Validation Queries

These queries validate data integrity **before any leakage calculation is performed**.

Purpose:

- Ensure one row equals one paid claim
- Detect structural inconsistencies
- Prevent overstatement of leakage

4.2.1 Row Count & Claim Uniqueness

Validates dataset volume and confirms no unintended duplication exists.

```
-- Validate total number of claims
SELECT
    COUNT(*) AS total_claim_records,
    COUNT(DISTINCT policy_number) AS distinct_policies
FROM insurance_claims;
```

4.2.2 Component Sum Validation

Identifies claims where payment breakdown does not reconcile—a common source of leakage.

```
-- Check if claim components reconcile with total claim amount
SELECT
    COUNT(*) AS total_claims,
    SUM(
        CASE
            WHEN total_claim_amount <>
                 (injury_claim + property_claim + vehicle_claim)
            THEN 1
            ELSE 0
        END
    ) AS component_mismatch_count
```

```
FROM insurance_claims;
```

4.2.3 Coverage Limit Sanity Check

Highlights potential coverage enforcement failures.

```
-- Identify claims exceeding policy coverage limit
SELECT
    COUNT(*) AS coverage_breach_count
FROM insurance_claims
WHERE total_claim_amount > policy_csl;
```

4.2.4 Severity vs Cost Sanity Check

Flags potential overpayments masked by low severity classification. Threshold applied: Minor severity claims exceeding portfolio average of 52,762.

```
-- Detect high-cost claims with low incident severity
SELECT
    COUNT(*) AS low_severity_high_cost_claims
FROM insurance_claims
WHERE incident_severity = 'Minor'
  AND total_claim_amount > (
      SELECT AVG(total_claim_amount)
      FROM insurance_claims
  );
```

4.3 Business Rule Implementation

Business validation rules are translated into explicit SQL flags. These flags represent **control failures**, not fraud accusations.

```
-- Create claim validation flags
SELECT
    *,
    CASE
        WHEN total_claim_amount > policy_csl
        THEN 1
        ELSE 0
    END AS coverage_breach_flag,

    CASE
        WHEN total_claim_amount <>
             (injury_claim + property_claim + vehicle_claim)
        THEN 1
        ELSE 0
    END AS component_mismatch_flag,

    CASE
        WHEN incident_severity = 'Minor'
             AND total_claim_amount > (
                 SELECT AVG(total_claim_amount)
                 FROM insurance_claims
             )
        THEN 1
        ELSE 0
```

```
    END AS severity_cost_mismatch_flag
FROM insurance_claims;
```

4.4 Leakage Calculation Logic

Potential leakage is calculated conservatively to avoid overstatement. Leakage is estimated only where **clear financial deviation exists**.

```
-- Estimate potential leakage amount
SELECT
    *,
    CASE
        WHEN total_claim_amount > policy_csl
        THEN total_claim_amount - policy_csl

        WHEN incident_severity = 'Minor'
             AND total_claim_amount > (
                 SELECT AVG(total_claim_amount)
                 FROM insurance_claims
             )
        THEN total_claim_amount - (
            SELECT AVG(total_claim_amount)
            FROM insurance_claims
        )

        ELSE 0
    END AS potential_leakage_amount
FROM insurance_claims;
```

4.5 Aggregation & KPI Queries

These queries produce final management KPIs.

```
-- Aggregate claims leakage KPIs
SELECT
    COUNT(*) AS total_claims,
    SUM(total_claim_amount) AS total_claims_paid,
    SUM(potential_leakage_amount) AS total_potential_leakage,
    ROUND(
        SUM(potential_leakage_amount)
        / NULLIF(SUM(total_claim_amount), 0),
        4
    ) AS leakage_rate
FROM (
    SELECT
        *,
        CASE
            WHEN total_claim_amount > policy_csl
            THEN total_claim_amount - policy_csl
            ELSE 0
        END AS potential_leakage_amount
    FROM insurance_claims
) t;
```

CFO interpretation:

- Total financial exposure from leakage

- Leakage rate as % of claims paid
- Basis for recovery prioritization

Detailed SQL execution steps, validation checks, and data-type findings are documented separately in the Phase 4 execution log for auditability purposes.

5. ANALYSIS RESULTS & INSIGHTS

5.1 Overall Claims Leakage Overview

An end-to-end analysis was performed on 1,000 claims with a total paid value of 52,761,940, applying conservative and audit-safe analytical rules designed to identify operational claims leakage.

Executive Result:
The analysis did not identify material claims leakage within the reviewed portfolio. This outcome reflects stable claims payment controls rather than analytical gaps.

While a high number of claims exceeded nominal policy coverage indicators, further validation confirmed that coverage limits are encoded as categorical ranges rather than enforceable monetary ceilings. To prevent overstatement of financial risk, coverage breaches were therefore treated as risk indicators only and excluded from direct leakage quantification.

Management Interpretation:
Under conservative assumptions, the claims portfolio demonstrates control stability and financial discipline.

5.2 Leakage by Claim Type

Claims were analyzed within the Motor insurance category to evaluate operational leakage patterns.

Finding:
No leakage signals were observed across Motor claims. Cost behavior relative to claim severity was consistent across the portfolio.

Executive Insight:
Product mix does not appear to be a primary driver of operational leakage within the current data structure. Any future differentiation would require more granular cost benchmarks or behavioral dimensions.

5.3 Leakage Drivers Analysis

Key potential leakage drivers were assessed:

- **Policy Coverage Breach**
  999 claims were flagged by the coverage indicator due to categorical policy_csl encoding (string format such as "250/500"). These represent risk signals for monitoring purposes, not confirmed coverage breaches. Due to non-monetized coverage encoding, these flags were classified as risk indicators rather than financial leakage drivers.

- **Severity–Cost Reasonableness**
  No claims exhibited costs materially disproportionate to their assigned severity levels under conservative thresholds.

- **Component Reconciliation**
  Full reconciliation was observed between total claim amounts and component-level payments, indicating strong internal payment controls.

Executive Conclusion:
No active operational leakage drivers were identified based on available data and conservative analytical assumptions.

5.4 Recovery Opportunity

Given the absence of detected leakage, no immediate post-payment recovery opportunity could be quantified from the analyzed dataset.

However, the analysis highlights structural constraints that limit recovery analytics, including:

- Non-monetized policy coverage limits
- Coarse severity classification
- Lack of benchmark cost expectations by incident type

Management Implication:
Near-term recovery efforts are unlikely to yield material returns without upstream data enhancements.

5.5 Key Executive Insights

- Claims payment processes exhibit strong operational consistency
- No evidence of systemic overpayment was identified under conservative review
- Current data structures constrain deeper leakage and recovery analytics
- Value creation opportunities lie primarily in:
    i) Monetizing policy coverage data
    ii) Increasing severity classification granularity
    iii) Establishing benchmark cost baselines

Executive Takeaway:
The claims function appears well-controlled. Incremental financial value is more likely to be achieved through data and control enhancement rather than aggressive recovery initiatives.

6. DASHBOARD & STORYTELLING

6.1 Dashboard Objective

The objective of the Claims Leakage & Recovery Dashboard is to provide executive-level visibility into claims payment integrity, risk signals, and data readiness for leakage detection.

The dashboard is designed to:

- Monitor claims payment consistency
- Highlight risk indicators without overstating financial exposure
- Support CFO and Head of Claims in control assurance and prioritization decisions

The dashboard does not introduce new business logic; all metrics are sourced directly from validated SQL outputs in Section 4.

6.2 Target Users

Primary Users

- Chief Financial Officer (CFO)
- Head of Claims
- Risk & Operations Management

Key Decisions Supported

- Are claims payment controls functioning consistently?
- Where do risk indicators exist that warrant further review?
- Is the current data structure sufficient for advanced leakage analytics?
- Should resources be allocated to recovery or to data enhancement?

6.3 Key KPIs

The dashboard presents a focused set of KPIs derived directly from validated SQL outputs to ensure consistency and auditability.

Core KPIs:

- Total Claims Analyzed
- Total Claims Paid
- Claims with Coverage Breach (Risk Indicator)
- Claims with Detected Leakage
- Total Potential Leakage Amount

Supporting Metrics:

- Claim distribution by collision type
- Average claim cost by severity
- Component reconciliation status

KPI Design Principle: Metrics distinguish clearly between risk indicators and confirmed financial leakage to prevent misinterpretation at the executive level.

6.4 Dashboard Structure

The dashboard is organized into four main sections:

1. **Executive Overview** – High-level snapshot of portfolio size, total payments, and leakage status.
2. **Risk Indicator Panel** – Visualization of coverage breach flags as early warning signals, not monetary exposure.
3. **Claims Segmentation** – Breakdown by claim type, severity, and policy category to identify concentration patterns.
4. **Data Quality & Readiness** – Indicators highlighting data limitations that affect leakage and recovery analytics depth.

The layout is designed to be minimalist, CFO-oriented, and optimized for rapid interpretation.

6.5 Usage Scenarios

CFO:

- Validate claims cost discipline
- Decide whether focus should be on recovery actions or data investment

Head of Claims:

- Monitor operational consistency
- Identify segments requiring deeper review without triggering unnecessary recovery efforts

Risk & Operations:

- Use risk indicators to guide sampling, audits, and control enhancement initiatives

The dashboard supports control assurance and prioritization, rather than investigative or forensic analysis.

6.6 Dashboard Availability

The dashboard has been implemented in Power BI using validated SQL outputs.

Screenshots are intentionally excluded from this document to maintain focus on decision logic rather than static visuals.

7. BUSINESS RECOMMENDATIONS & IMPACT

7.1 Key Findings Summary

This analysis reviewed 1,000 insurance claims using a conservative, rule-based leakage detection framework. Key findings include:

- No material financial leakage detected under the applied conservative rules.
- 999 claims were flagged by the coverage indicator due to categorical policy_csl encoding (string format). These represent data format limitations rather than confirmed coverage breaches, and were treated as risk signals for monitoring purposes.
- Claim payment amounts were generally aligned with incident severity, suggesting reasonable cost distribution.
- No significant mismatches were identified between claim components and total claim amounts.

Overall, the results indicate that existing controls are effective at preventing obvious claim leakage under strict validation criteria.

7.2 Operational Recommendations

Although no material leakage was detected, several operational improvements are recommended to strengthen long-term controls and monitoring:

1. **Maintain Conservative Validation Rules as Baseline** – The current rule set should remain the foundation for claim validation to ensure continued financial discipline.
2. **Introduce Tiered Review for High-Severity Claims** – Claims categorized as *Major Damage* or *Total Loss* should undergo secondary review to detect subtle anomalies not captured by baseline rules.
3. **Enhance Coverage Breach Monitoring Visibility** – While coverage breaches are not leakage, improving reporting on these cases can help management assess policy adequacy and underwriting strategy.
4. **Standardize Claim Review Documentation** – Ensuring consistent documentation will support auditability and future analytical enhancements.

7.3 Financial Impact Estimation

Under the current conservative framework:

- Estimated recoverable leakage: 0
- Estimated potential leakage: 0

However, this does not imply the absence of financial value. Instead:

- The analysis demonstrates risk avoidance, confirming that current processes prevent material leakage.
- The financial benefit lies in loss prevention, reduced fraud exposure, and improved confidence in claims governance.
- Additional value could be unlocked by applying expanded or more aggressive detection rules, such as anomaly-based or peer-comparison methods.

In practice, the framework serves as a **preventive control mechanism** rather than a recovery-focused model.

7.4 Risk & Implementation Considerations

Several considerations should be addressed before expanding or modifying the leakage detection approach:

- **False Positives Risk** – Loosening detection rules may increase false alerts, impacting operational efficiency.
- **Data Quality Dependency** – Advanced detection methods rely heavily on consistent and accurate historical data.
- **Governance & Audit Sensitivity** – Any changes to detection logic must align with compliance and audit requirements to avoid disputes.
- **Change Management** – Operational teams may require training to interpret and act on more complex analytical outputs.

7.5 Next Steps

To further enhance analytical depth and business value, the following next steps are recommended:

- **Develop Tier-2 Leakage Rules** – Introduce secondary rules targeting outliers, severity-adjusted costs, and historical deviations.
- **Expand Analytical Dimensions** – Analyze claim behavior by combining severity, vehicle type, location, and historical averages.
- **Transition Toward Anomaly Detection** – Gradually move from static rules to semi-automated anomaly detection models.
- **Integrate Dashboard into Routine Reviews** – Use the Power BI dashboard as a regular management and audit review tool.

8. APPENDIX

A. Dataset Reference

This project utilizes a publicly available insurance claims dataset for analytical and educational purposes.

- Dataset Name: Insurance Claims Dataset
- Source: Mendeley Data
- Access Link: https://data.mendeley.com/datasets/992mh7dk9y/2
- Records Analyzed: 1,000 insurance claims
- Dataset Usage: Adapted for claims leakage and risk control analysis

Key adaptations applied:

- Policy coverage limits parsed into numeric values
- Conservative rule-based leakage indicators applied
- No synthetic fraud or leakage cases injected

B. SQL File Index

All analytical processing was executed using DuckDB to ensure reproducibility and auditability.

Primary SQL Objects:

- insurance_claims = Raw claims dataset
- claim_policy_parsed = Normalized policy coverage limits
- claim_validation_flags = Derived flags indicating potential leakage conditions
- claim_leakage = Final leakage assessment table

Visualization Output:
powerbi_claim_output.csv (Exported for Power BI dashboard visualization)

This approach ensured clear separation between data processing logic and visualization layers.

C. Metric Definitions

Metric	Definition
Total Claims	Total number of claims analyzed
Total Claims Paid	Sum of all claim payment amounts
Claims with Coverage Breach	Claims exceeding policy coverage limits
Detected Leakage Claims	Claims flagged under leakage detection rules
Total Potential Leakage	Estimated recoverable leakage amount
Coverage Breach Flag	Binary indicator (0/1) for claims triggering coverage threshold check
Severity Cost Mismatch Flag	Minor severity claims with unusually high cost

D. Analytical Assumptions & Scope Limitations

- Conservative rules were intentionally applied to minimize false positives
- Coverage breach does not automatically indicate financial leakage
- No predictive modeling or fraud classification was implemented
- Findings reflect control effectiveness rather than fraud prevalence

Appendix Note
The appendix serves as a technical reference to support transparency, reproducibility, and audit review, without impacting the readability of the main analytical narrative.