INAD Analysis Tool

User Documentation

Enhanced Version 2.0

For Legal Team Review

December 2024

Table of Contents

1. Introduction and Purpose

What is This Tool?

The INAD Analysis Tool is designed to identify airlines and routes with systematically elevated rates of inadmissible passengers (INADs). Its purpose is to support legal review by distinguishing between:

- Isolated incidents that may not require action
- Systemic patterns that warrant investigation or enforcement

The tool processes two data sources:
1. INAD records: Cases of passengers refused entry at Swiss borders
2. BAZL passenger data: Total passenger volumes by airline and route

By comparing INAD counts against passenger volumes, we can identify which airlines or routes have disproportionately high INAD rates relative to their traffic, suggesting potential systemic issues with passenger screening.

Who Should Use This Documentation?

This documentation is written for members of the legal team who will review the analysis results and make decisions about further investigation, warnings, or enforcement actions. Technical knowledge of statistics or programming is not required.

2. The Three-Step Analysis Process

The analysis follows a progressive filtering approach, each step narrowing down the focus to identify the most significant cases.

Step 1: Airline-Level Screening (Pruefstufe 1)

PURPOSE: Identify airlines with a meaningful number of INAD cases

HOW IT WORKS:
- Count the total number of INADs for each airline in the semester
- Flag airlines with 6 or more INADs (configurable threshold)
- Airlines below this threshold are excluded from further analysis

WHY 6 INADs?
A small number of INADs (1-5) could easily be random occurrences. Setting a minimum threshold ensures we focus on statistically meaningful patterns rather than isolated incidents.

Step 2: Route-Level Screening (Pruefstufe 2)

PURPOSE: From airlines identified in Step 1, identify specific routes with elevated INAD counts

HOW IT WORKS:
- For each airline that passed Step 1, count INADs by route (origin airport)
- Flag routes with 6 or more INADs
- Routes below this threshold are excluded from Step 3

WHY ANALYZE BY ROUTE?
An airline might have high total INADs but concentrated on one route. This helps identify specific problem origins rather than penalizing an airline's entire operation.

Step 3: Density Analysis (Pruefstufe 3)

PURPOSE: Compare INAD counts against passenger volumes to identify disproportionately high rates

HOW IT WORKS:
1. For each route from Step 2, retrieve the passenger count (PAX)
2. Calculate the INAD density: (INADs / PAX) x 1000
   This gives INADs per 1,000 passengers
3. Calculate the threshold (median of all densities)
4. Flag routes above the threshold for legal review

WHY USE DENSITY INSTEAD OF RAW COUNTS?
Raw INAD counts favor large airlines unfairly. A route with 20 INADs and 500,000 passengers (0.04 per mille) is performing better than a route with 10 INADs and 20,000 passengers (0.50 per mille). Density provides a fair, relative comparison across different traffic volumes.

3. Understanding the Enhanced Features

The enhanced version introduces several improvements to make the analysis more reliable and actionable for legal review.

3.1 Robust Threshold Calculation

PREVIOUS APPROACH: Used simple arithmetic mean (average) to calculate the threshold

PROBLEM: The mean is highly sensitive to outliers. A single extreme value can dramatically skew the threshold, leading to incorrect classifications. In actual data from 2016 H2, one route with incomplete passenger data raised the average from approximately 0.1 per mille to 22.9 per mille.

NEW APPROACH: Uses MEDIAN as the default threshold calculation
- The median is the middle value when all densities are sorted
- It is not affected by extreme outliers
- One bad data point cannot skew the entire analysis

3.2 Minimum Passenger Threshold

PREVIOUS APPROACH: All routes were included regardless of passenger volume

PROBLEM: Routes with very few passengers produce unreliable density calculations. 6 INADs on a route with only 100 passengers gives 60 per mille density, but this is statistically meaningless.

NEW APPROACH: Routes with fewer than 5,000 passengers are marked as UNRELIABLE
- These routes are still shown but flagged with a warning
- They are excluded from threshold calculation
- Legal team can see them but should interpret with caution

3.3 Priority Classification System

PREVIOUS APPROACH: Routes were simply marked as "above average" or "below average"

NEW APPROACH: Four-tier priority classification:

[RED] HIGH PRIORITY - Requires immediate legal review
  Criteria (ALL must be met):
  - Density >= 1.5x the threshold
  - Density >= 0.10 per mille (absolute minimum)
  - At least 10 INAD cases
  - Reliable data (>= 5,000 passengers)

[ORANGE] WATCH LIST - Should be monitored; may escalate
  - Density >= threshold
  - Does not meet all HIGH PRIORITY criteria

[GREEN] CLEAR - No action required
  - Density < threshold

[GREY] UNRELIABLE - Data quality concerns
  - Fewer than 5,000 passengers or incomplete data

3.4 Confidence Scoring

PREVIOUS APPROACH: No indication of how reliable each result was

NEW APPROACH: Each route receives a confidence score from 0-100%

Based on:
- INAD count (more cases = more confidence)
- Passenger volume (more passengers = more reliable denominator)

Interpretation:
- 0-30%: Low confidence, treat results with caution
- 30-60%: Medium confidence, results are indicative
- 60-100%: High confidence, results are reliable

3.5 Data Quality Checks

PREVIOUS APPROACH: No validation of input data quality

NEW APPROACH: Automatic data quality checks with warnings:

"Incomplete PAX data (2/6 months)"
- Passenger data exists for fewer than 4 of 6 months
- Density calculation may be inaccurate

"Low PAX volume (<5,000)"
- Total passengers too low for reliable statistics

"High variance in monthly PAX data"
- Unusual fluctuations in passenger numbers
- May indicate data recording issues

3.6 Systemic Case Detection

PREVIOUS APPROACH: Each semester was analyzed independently

NEW APPROACH: Multi-semester analysis identifies SYSTEMIC cases:

Definition: A route is SYSTEMIC if it appears on the WATCH LIST or HIGH PRIORITY in 2 or more consecutive semesters

Additional information provided:
- Total appearances across all semesters
- Maximum consecutive appearances
- Trend direction (IMPROVING or WORSENING)
- Percentage change in density over time

This helps legal team distinguish between one-time issues and persistent patterns.

3.7 Refusal Code Categorization

Refusal codes are now categorized and displayed for each flagged route:

Categories:

- Documentation: Missing or invalid travel documents
- Fraud: Forged or falsified documents
- Visa: Visa-related issues (expired, wrong type, overstay)
- Security: Security concerns, entry bans

This helps understand the NATURE of the problem and informs appropriate responses.

4. The Dashboard Interface

The dashboard provides six main tabs for navigating the analysis:

Tab 1: Overview

Quick summary of the entire analysis containing:
- Summary metrics (total INADs, counts by priority level)
- Priority distribution pie chart
- Confidence score distribution
- Data quality warnings
- Top routes by density
- Refusal category breakdown for flagged routes

Tab 2: Step 1 - Airlines

Review airline-level screening results:
- List of all airlines with INAD counts
- Color-coded status (Review / OK)
- Summary statistics
- Distribution histogram

Tab 3: Step 2 - Routes

Review route-level screening results:
- List of all routes with INAD counts
- Filter to airlines passing Step 1 only
- Top airports by INAD count

Tab 4: Step 3 - Priority Analysis

Detailed priority classification with full metrics:
- Classification criteria explanation
- Complete route list with density, confidence, and priority
- Interactive scatter plot (density vs. passengers)
- Summary of HIGH PRIORITY and WATCH LIST counts

Tab 5: Systemic Cases

Identify persistent patterns across multiple semesters:
- Count of confirmed systemic cases
- Trend analysis (improving/worsening)
- Historical chart of flagged routes over time
- Individual route history viewer

Tab 6: Legal Summary

Export-ready summary for legal team:
- Analysis parameters used
- List of HIGH PRIORITY routes with details
- List of WATCH LIST routes with details
- Data quality notes
- Export buttons for CSV downloads

5. Interpreting Results for Legal Review

Recommended Review Process

1. START WITH THE LEGAL SUMMARY TAB
   Review HIGH PRIORITY routes first. These have the strongest statistical basis for action. Note the confidence scores - higher is more reliable.

2. CHECK FOR SYSTEMIC CASES
   Routes appearing as systemic have persisted over time. These represent the strongest cases for enforcement. Review their trend - worsening trends are more concerning.

3. REVIEW WATCH LIST FOR CONTEXT
   These routes are above average but do not meet all HIGH criteria. May escalate to HIGH PRIORITY next semester. Consider for warning letters or monitoring.

4. NOTE DATA QUALITY WARNINGS
   Be cautious with UNRELIABLE classifications. If a route has quality warnings, results may not support action.

5. EXPORT FOR DOCUMENTATION
   Use export buttons to create records. Include analysis parameters in your documentation.

What Each Priority Level Means for Action

[RED] HIGH PRIORITY
   Statistical basis: Strong
   Recommended action: Formal investigation, potential warning or fine

[ORANGE] WATCH LIST
   Statistical basis: Moderate
   Recommended action: Monitor, consider informal outreach

[GREEN] CLEAR
   Statistical basis: Below threshold
   Recommended action: No action required

[GREY] UNRELIABLE
   Statistical basis: Insufficient data
   Recommended action: Do not take enforcement action

Important Caveats

1. Statistical indicators are not proof
   High density indicates a pattern worth investigating, but root cause analysis is still required before action.

2. Context matters
   Some origins may have inherently higher INAD rates. Consider geopolitical factors, visa regimes, etc.

3. Data limitations
   Passenger data may be incomplete for some routes. Always check data quality warnings.

4. Threshold is relative

Being above the median means above half of peers. It does not mean the absolute rate is necessarily problematic.

6. Why the New Approach is More Robust

Comparison: Old vs. New Methodology

ISSUE 1: Outlier Sensitivity
OLD: One route with bad data could skew the entire semester results
NEW: Median-based threshold ignores extreme values; data quality checks flag suspicious data; minimum PAX threshold excludes unreliable routes

ISSUE 2: Binary Classification
OLD: Routes were either "above average" or "below average" with no distinction
NEW: Four-tier system with clear criteria; proportionate response based on severity

ISSUE 3: Single Semester View
OLD: Each semester analyzed in isolation
NEW: Systemic case detection across all semesters; trend analysis shows improvement or deterioration

ISSUE 4: No Confidence Indication
OLD: All results presented with equal weight
NEW: Confidence scores based on sample size; clear warnings for low-confidence results

ISSUE 5: Hidden Data Problems
OLD: Data quality issues not detected or reported
NEW: Automatic quality checks; visible warnings in dashboard

Statistical Validity

The new approach follows established statistical best practices:

1. ROBUST STATISTICS - Using median instead of mean is standard when outliers may be present. It is used in scientific research, financial analysis, and regulatory contexts worldwide.

2. MINIMUM SAMPLE SIZE - The 5,000 passenger threshold ensures density calculations are based on sufficient data.

3. MULTI-PERIOD ANALYSIS - Looking at patterns across multiple semesters reduces the chance of acting on random fluctuations.

4. CONFIDENCE SCORING - Provides an intuitive measure of reliability that helps non-statisticians interpret results appropriately.

7. Configuration Parameters

All parameters can be adjusted in the dashboard sidebar:

Minimum INAD Threshold (Default: 6)

What it does: Sets the minimum number of INADs for Step 1 and Step 2

Recommendations:
- 6: Standard setting, balances sensitivity and specificity
- Lower (3-5): More sensitive, catches smaller patterns
- Higher (8-10): More conservative, only flags clear patterns

Minimum PAX for Reliable Data (Default: 5,000)

What it does: Routes with fewer passengers are marked UNRELIABLE

Recommendations:
- 5,000: Standard setting, reasonable statistical basis
- Lower: Include more routes but with less confidence
- Higher: More conservative, only highest-confidence results

Threshold Calculation Method (Default: Median)

Options:
- Median: Middle value, most robust against outliers (RECOMMENDED)
- Trimmed Mean: Removes top/bottom 10% then averages
- Mean: Simple average (NOT RECOMMENDED - sensitive to outliers)

Minimum Density for HIGH PRIORITY (Default: 0.10 per mille)

What it does: Even if above threshold, must exceed this absolute minimum

Recommendations:
- 0.10 per mille: Standard setting
- Lower: More routes qualify as HIGH PRIORITY
- Higher: Only the most severe cases flagged

HIGH PRIORITY Multiplier (Default: 1.5x)

What it does: Must be this multiple of threshold for HIGH PRIORITY

Recommendations:
- 1.5x: Standard setting (50% above threshold)
- Lower (1.2-1.3): More sensitive
- Higher (2.0+): Only extreme cases flagged

8. Glossary of Terms

BAZL

Bundesamt fuer Zivilluftfahrt (Swiss Federal Office of Civil Aviation). Source of passenger volume data.

Confidence Score

A 0-100% score indicating how reliable the density calculation is, based on INAD count and passenger volume.

Density (INAD-Dichtewert)

INADs per 1,000 passengers: (INAD count / PAX) x 1000. Expressed in per-mille.

HIGH PRIORITY

Routes requiring immediate legal review. Meet all criteria: high density, high INAD count, reliable data.

INAD

Inadmissible Passenger. A passenger refused entry at the border for various reasons.

Last Stop

The final departure airport before arriving in Switzerland. Used to identify the origin of INAD cases.

Median

The middle value in a sorted list. Unlike the mean, it is not affected by extreme values.

PAX

Passengers (abbreviation used in aviation industry).

Semester

Six-month period used for analysis. H1: January-June, H2: July-December.

Systemic Case

A route flagged in 2+ consecutive semesters. Indicates a persistent pattern.

Threshold

The density value used to separate "above average" from "below." Calculated using median of reliable route densities.

UNRELIABLE

Classification for routes with insufficient data. Should not be used as basis for enforcement action.

WATCH LIST

Routes above threshold but not meeting all HIGH PRIORITY criteria. Should be monitored.

Document Version: 2.0

Last Updated: December 2024

Tool Version: Enhanced INAD Analysis Tool 2.0

For technical support or questions about this documentation,

contact the data analysis team.