THE UNKNOWN: THE HAUNTING FROM A GHOST

Group 4 : A Data-Driven Look at 50 Years of Global Terrorism


"All the News That's Fit to Print"
The New York Times
Late Edition
VOL. CL . . No. 51,874
NEW YORK, WEDNESDAY, SEPTEMBER 12, 2001
75 CENTS
U.S. ATTACKED
HIJACKED JETS DESTROY TWIN TOWERS AND HIT PENTAGON IN DAY OF TERROR
A CREEPING HORROR
Buildings Burn and Fall as Onlookers Search for Elusive Safety
President Vows to Exact Punishment for 'Evil'
Detroit Free Press
ON GUARD FOR 181 YEARS
Tuesday 4.16.2013
www.freep.com
A GANNETT COMPANY
TERROR AT BOSTON MARATHON
WHY? WHO?
3 KILLED, 140-PLUS INJURED IN 2 BLASTS BY UNKNOWN ATTACKER
CHAOS
4 hours after race began, thousands flee
CASUALTIES
Child dies; many victims lose limbs
OBAMA
Perpetrator will 'feel the weight of justice'
The Washington Post
Assault on Paris: More than 140 killed
Militant's death a win in PR war
Trump recaptures spotlight with torrent of venom
High court to review abortion law in Tex.
The New York Times
PARIS TERRORIST ATTACKS KILL OVER 100; FRANCE DECLARES STATE OF EMERGENCY
Bursts of Chaos and Horror, Once Again
Series of Shootings and Blasts, Apparently Coordinated

> In 2013, an intelligence report missed a warning sign, simply because it was coded as 'Unknown'. A week later, a marketplace exploded, claiming dozens of lives. That data point wasn't just missing; it was a 'ghost' that returned to haunt us.

...HOW MANY SUCH 'GHOSTS' ARE HIDING IN OUR DATA?

OPERATION GHOST HUNT: THE INVESTIGATION

Our mission is to transform chaotic data into actionable intelligence by answering two critical questions:

1. CLEANSE THE CRIME SCENE: Data Preparation

How many "ghosts"—missed signs or miscoded warnings—are hidden in the raw data, and how do we purge them?

2. PORTRAIT OF THE INVISIBLE ENEMY: Clustering

After uncovering these ghosts, what do they reveal about the true faces of the enemy we are fighting?

FEAR IS LOUD. DATA IS QUIET.

We chose to listen to the data

OUR INVESTIGATION

SOURCE

The Global Terrorism Database

SCOPE

1970 - 2017

FOCUS

Systemic truths, not singular events

THE DATA CHALLENGE - THE "FOG OF WAR"

181,691
Raw Incidents

135
Features

MISSING INTELLIGENCE

45% of features are >90% empty

ILLOGICAL RECORDS

The Data Contradicts Itself: hours kidnapped < 0

INCONSISTENT CODES

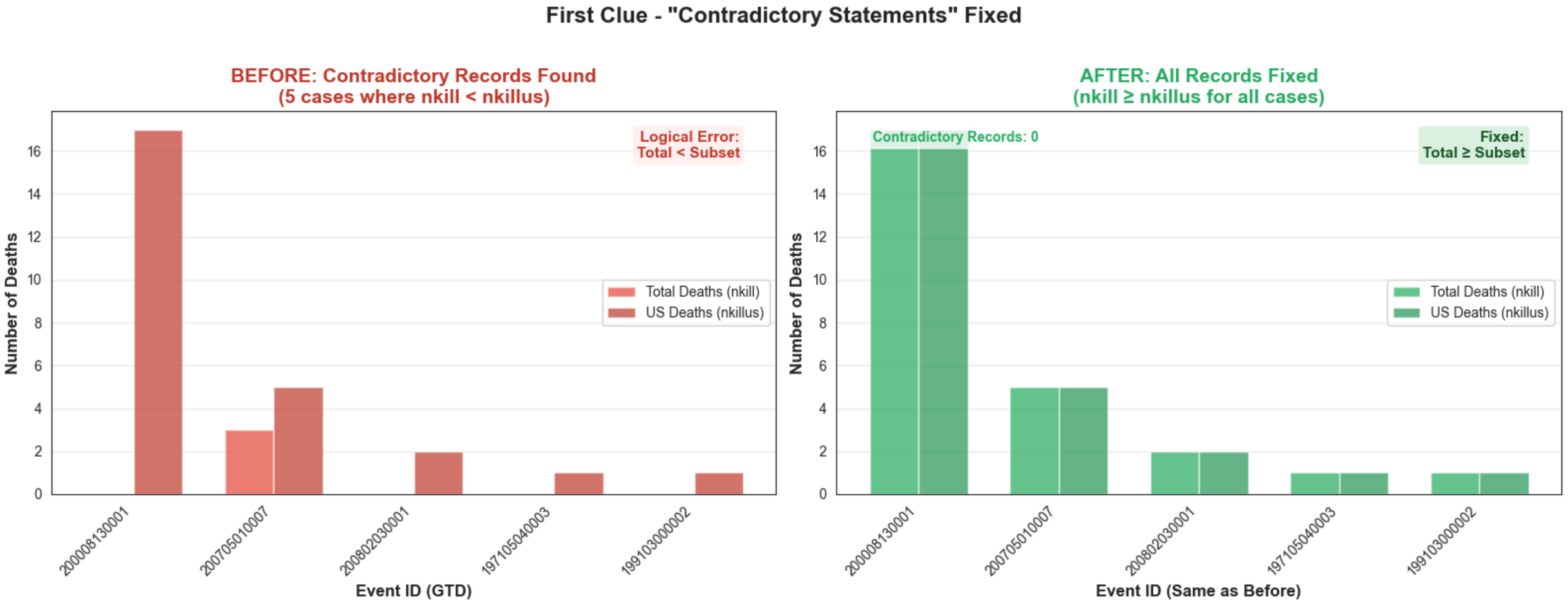
"Unknown" has many names: -9, -99, NaN, "Unknown", etc

INVALID INFORMATION

Impossible Dates: 0/5/1985, 14/0/2004

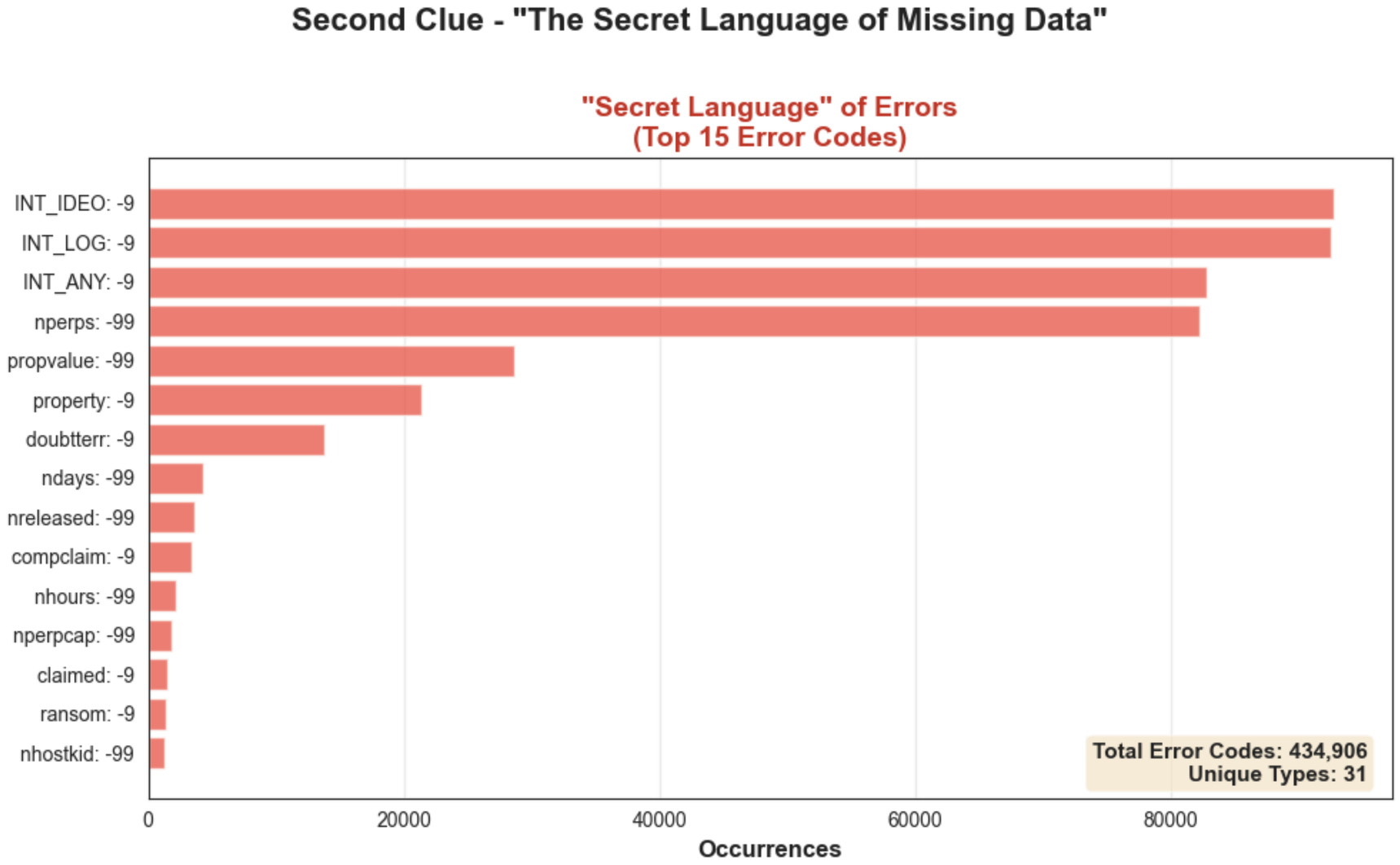
THE FIRST CLUE: CONTRADICTORY STATEMENTS

The Logical Error: Total Casualties cannot be less than U.S. Casualties (nkill < nkillus)



THE SECRET LANGUAGE OF OMISSIONS

Standardizing -9, -99, NaN, and Unknown into a Unified Term.



Before

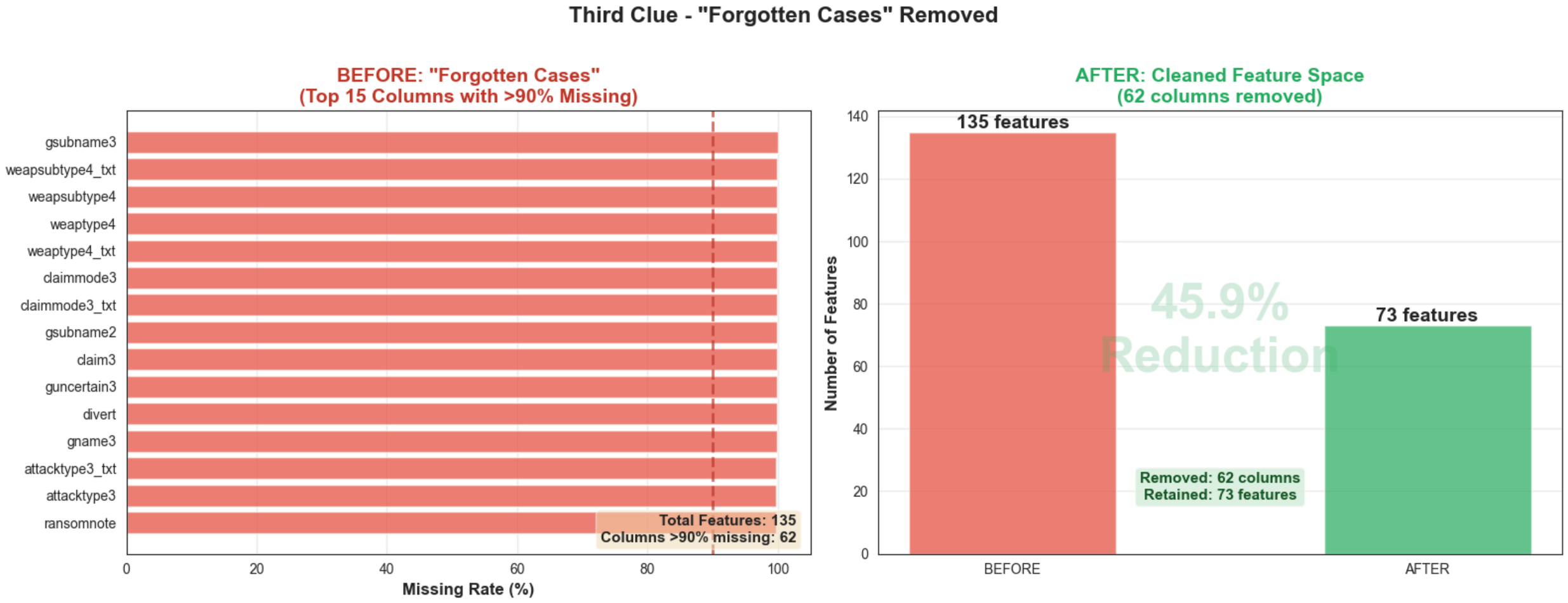
31 DIFFERENT ERROR CODE TYPES

After

1 UNIFIED (NAN)

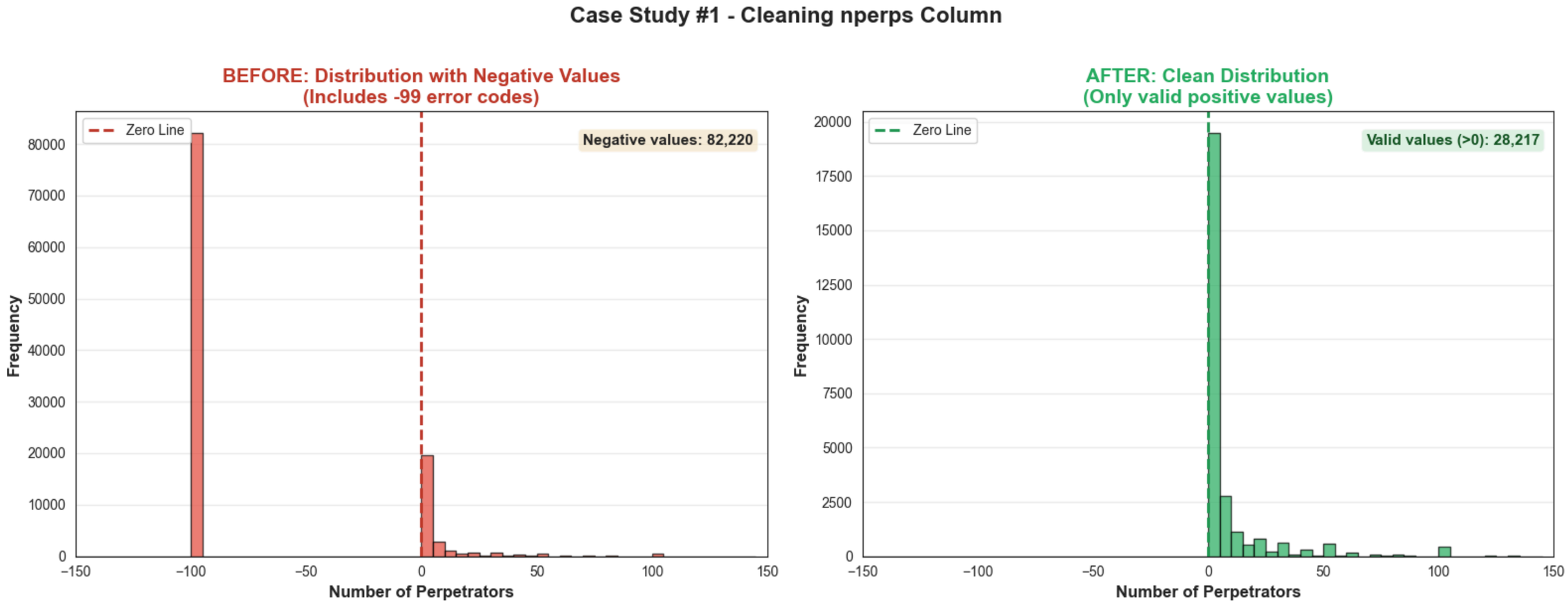
THE THIRD CLUE: THE FORGOTTEN CASES

Feature Reduction: Eliminating Columns with >90% Missing Data.



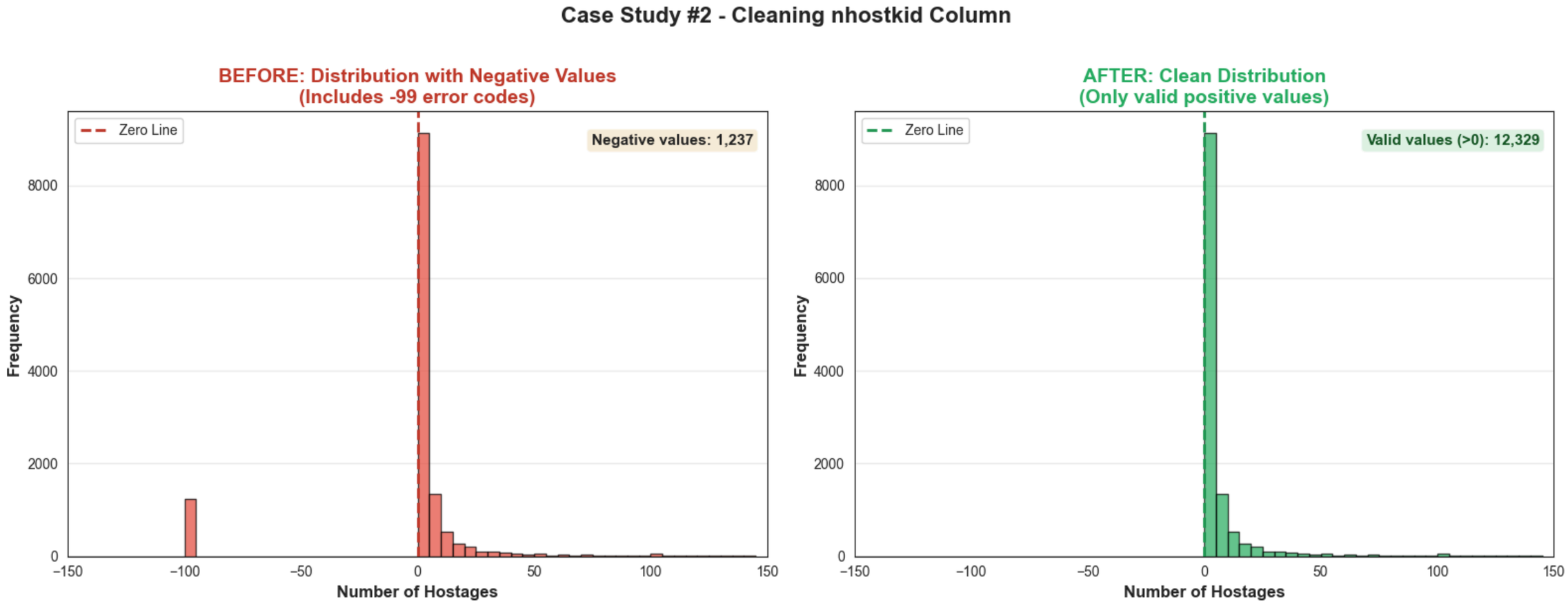
CASE STUDY 1: EXORCISING GHOST VALUE FROM NPERPS

Handle Errors Outliers: Replace non-positive with NaN



CASE STUDY 2: EXORCISING GHOST VALUE FROM NHOSTKID

Handle Errors Outliers: Replace non-positive with NaN



FROM CHAOS...

135+ TOTAL FEATURES
A sea of noise

5+ "MISSING" CODES
Ambiguous, error-prone

45% OF FEATURES >90% EMPTY
Unreliable & incomplete

1000+ ILLOGICAL RECORD
Fundamentally broken

TO CLARITY

70 RELEVANT FEATURE
Signal over noise

0 MISSING VALUES (NAN)
Complete for Analysis

CONSISTENT "UNKNOWN" LABEL
Clear & machine-readable

0 LOGICAL CONTRADICTIONS
Internally consistent & trustworthy

THE GHOST HUNT: DRAW A PORTRAIT OF INVISIBLE ENEMY

“UNKNOWN”

THIS ISN’T AN ANSWER.
IT’S AN INTELLIGENCE FAILURE

OUR APPROACH

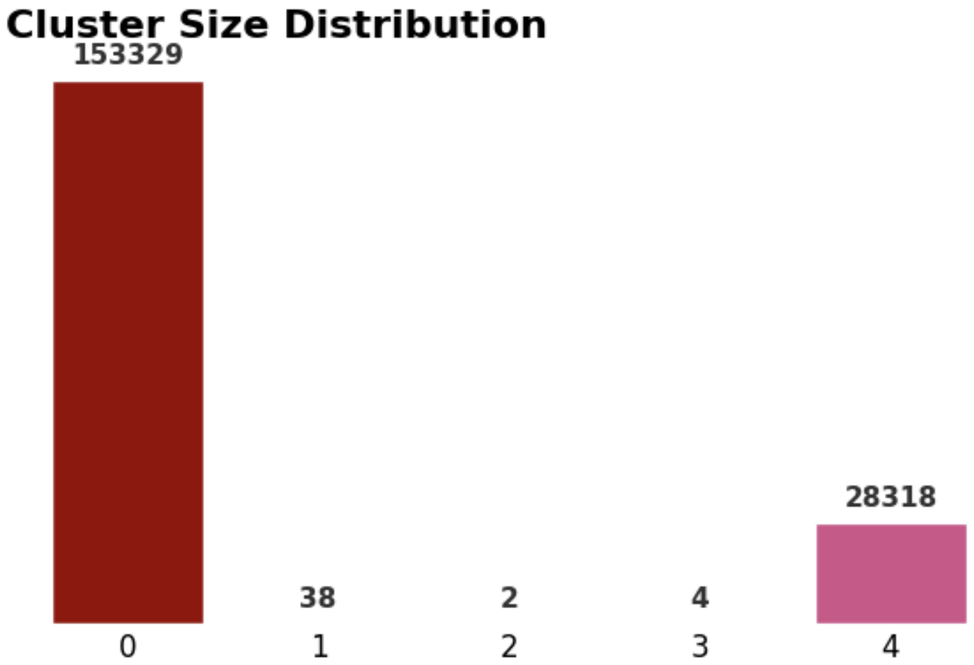
INSTEAD OF WHO THEY ARE
WE ASKED HOW THEY ACT
TO FIND PATTERNS

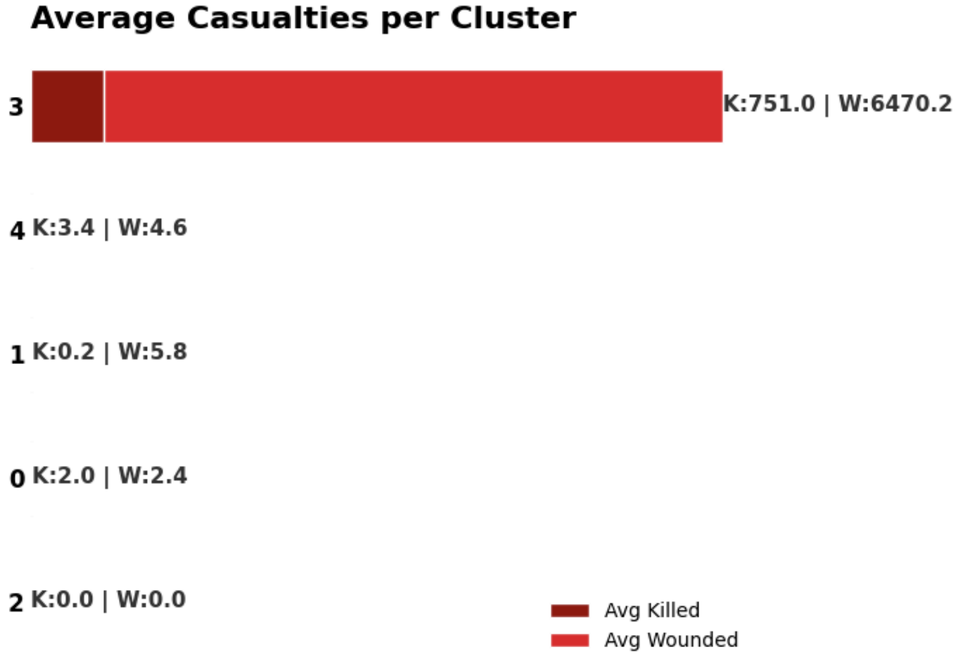
WEAPON

MACHINE LEARNING → HUNT THE GHOST HINDING IN THE DATA

THE FIRST ATTEMPT: CLUSTERING ON RAW DATA

WHEN THE MACHINE TRIES TO READ A CORRUPTED FILE, IT SEES NONSENSE.





A BROKEN RESULT: 99% OF THE DATA WAS FORCED INTO A SINGLE GROUP WHILE OTHERS ARE EMPTY

ILLOGICAL PROFILES: CLUSTER 3 SHOWS AN IMPOSSIBLE AVERAGE OF OVER 7,000 CASUALTIES PER ATTACK

THE BREAKTHROUGH: FIVE GHOSTS IN THE MACHINE

WITH CLEAN DATA, THE MACHINE COULD FINALLY SEE. THE "UNKNOWN" IS NOT ONE ENTITY – IT IS FIVE.

PROFILE 1: THE AGITATOR

The Low-Level Noise

PROFILE 2: THE GUERRILLA

The Hybrid Fighter

PROFILE 3: THE SPECIALIST

The Kidnapping Professional

PROFILE 4: THE MASS-CASUALTY BOMBER

The Hyper-Lethal Threat

PROFILE 5: THE ASSASSIN

The Precision Striker

THE LANDSCAPE: MAPPING THE VOLUME OF THE GHOSTS

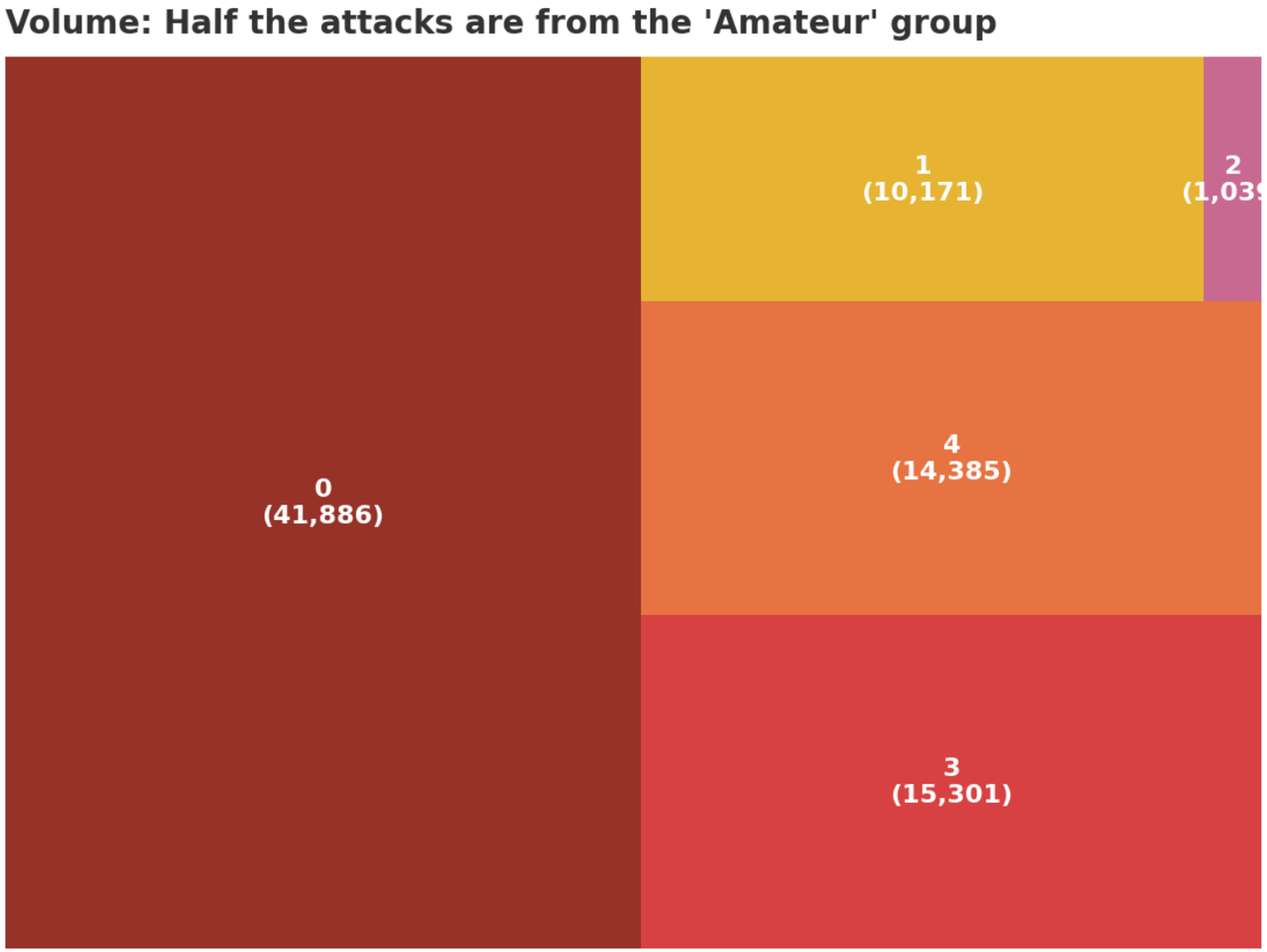
NEARLY 50% OF ALL 'UNKNOWN' INCIDENTS ARE LOW-IMPACT 'NOISE'.

CLUSTERE 0 – THE AGITATOR

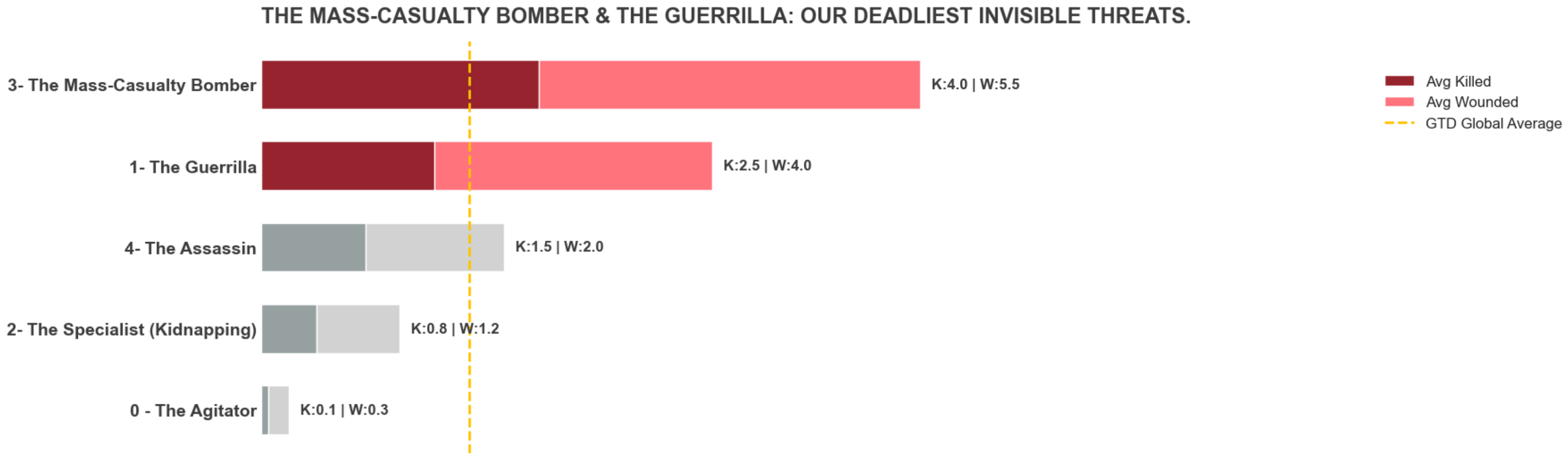
The dominant force by volume, creating the most 'noise'.

CLUSTERS 3 – THE MASS-CASUALTY BOMBER

The second most frequent, representing a significant volume of high-risk attacks.



THE MASS-CASUALTY BOMBER & THE GUERRILLA: OUR DEADLIEST INVISIBLE THREATS



50% OF THE DATA (CLUSTER 0 - THE AGITATOR) IS LOW-IMPACT 'NOISE'

they are resource-draining 'noise', not strategic threats.

CLUSTERS 3 AND 1 REPRESENT THE ACTUAL KILLING MACHINES

all counter-terrorism efforts must be re-focused on these two profiles.

EACH GHOST HAS A UNIQUE TACTICAL FINGERPRINT

THEIR METHODS ARE NOT RANDOM; THEY ARE SPECIALIZED. THIS IS THEIR OPERATIONAL DNA.

CLUSTER 2 – THE SPECIALIST

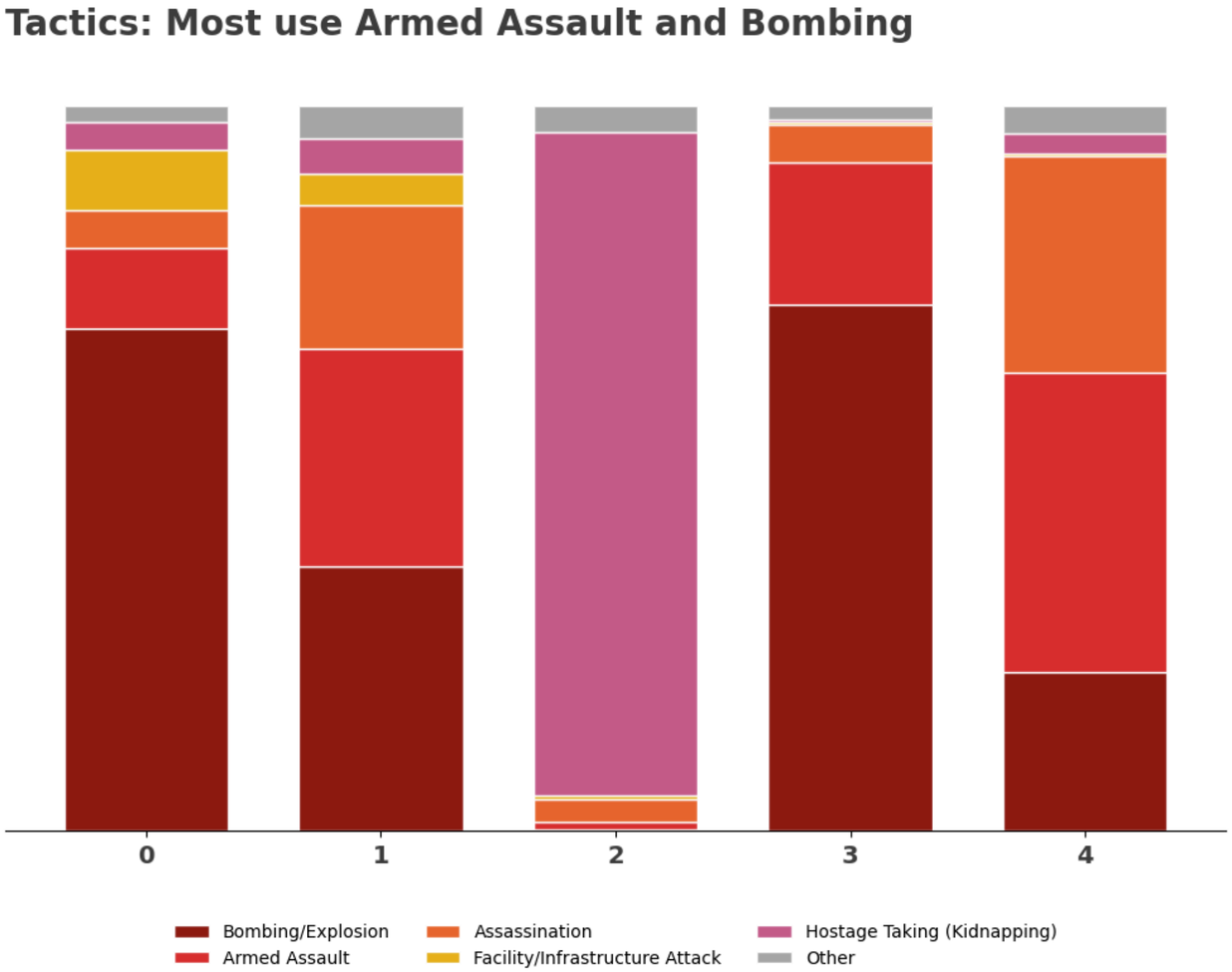
Over 90% of this group's activity is a single tactic: Hostage Taking (Kidnapping).
They are professionals, not generalists.

CLUSTER 4 – THE BRUTE FORCE

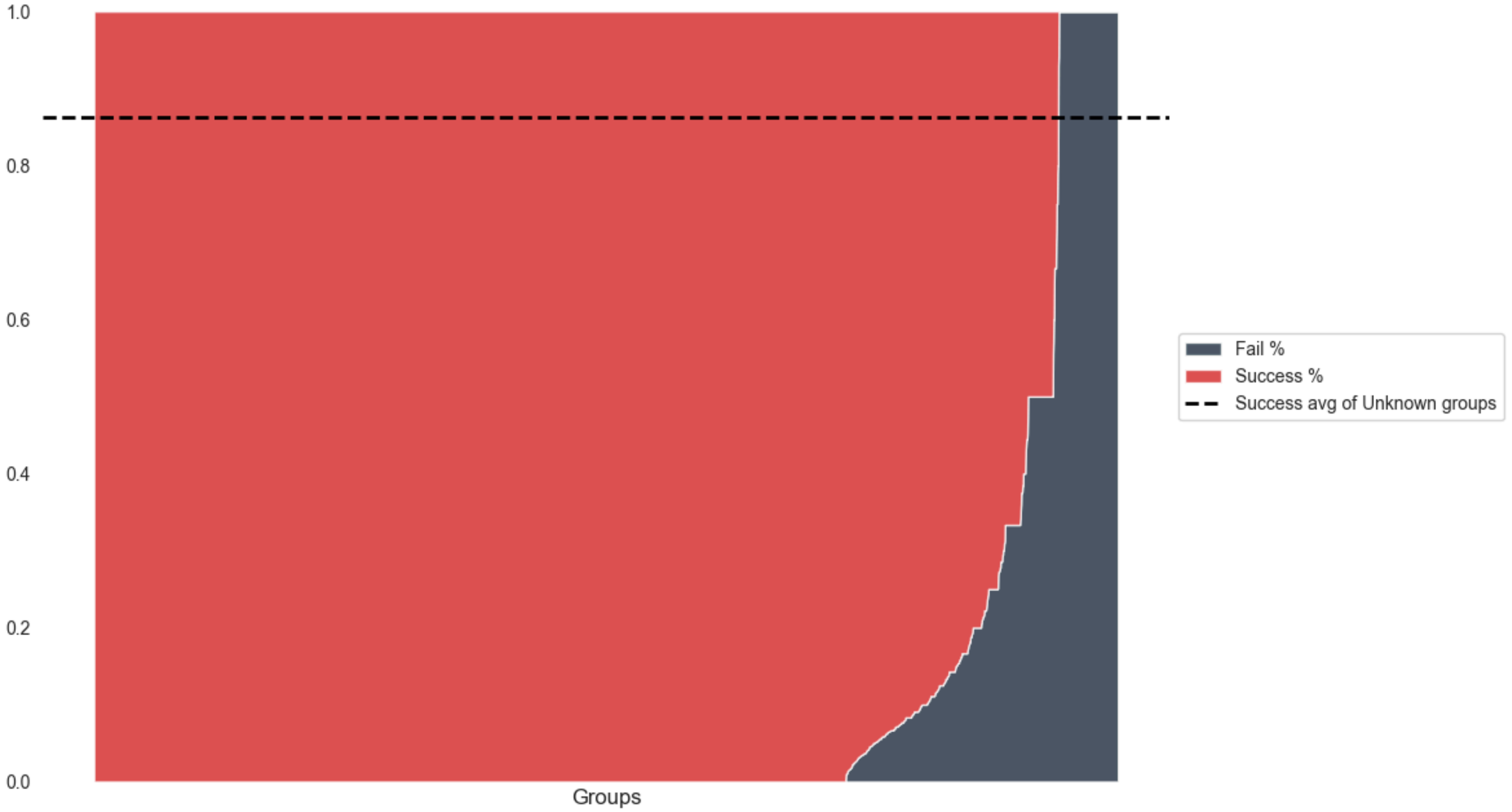
The Mass-Casualty Bomber relies almost exclusively on one method: Bombing/Explosion.
This is a high-impact, low-complexity strategy.

CLUSTER 1 – THE HYBRID FIGHTER

The Guerrilla shows the most tactical diversity, balancing Bombing, Armed Assault, and Assassination.
This suggests a more organized, flexible adversary.



DO NOT MISTAKE ANONYMITY FOR INEFFECTIVENESS



PROFILING IS NOT ENOUGH


The Washington Post

Wednesday, September 12, 2001

Terrorists Hijack 4 Airliners, Destroy World Trade Center, Hit Pentagon; Hundreds Dead

Bodies Pulled From Pentagon; Troops Patrol District Streets

Bush Promises Retribution; Military Put on Highest Alert


USA TODAY

Special Edition

'Tears most of the day'

Carnage in New York

Pentagon in flames

'Act of war'

Terrorists strike; death toll 'horrendous'

Bush to nation: U.S. 'saw evil'

Law enforcement eyes bin Laden


Detroit Free Press

AMERICA'S DARKEST DAY

UNMASKING THESE FIVE "GHOSTS" IS A CRITICAL BREAKTHROUGH. BUT IT IS ONLY THE FIRST STEP.

RAW DATA → PREDICTIVE INTELLIGENCE

KNOWING WHAT THEY LOOK LIKE IS *REACTIVE*. PREDICTING WHAT THEY WILL DO NEXT IS *PROACTIVE*.

01 02 03 04 05 06 07

THE BLUEPRINT: A PREDICTIVE EARLY WARNING SYSTEM

RAW INTEL FRAGMENT

↓

MACHINE LEARNING CLASSIFIER
(Trained on our 5 Profiles)

↓

THREAT ASSESSMENT & PROBABILITY SCORE

THE FUTURE OF SECURITY IS NOT FOUGHT WITH *BULLETS*. IT IS WON WITH *DATA*.

SHIFT THE PARADIGM: FROM REACTIVE TO PREDICTIVE

BUILD THE SYSTEM: AI-POWERED EARLY WARNING SYSTEM

PRIORITIZE DATA PREPARATION: ACCURACY OF ANY PREDICTIVE SYSTEM IS ENTIRELY DEPENDENT ON THE DATA QUALITY