

Chicago Crash ETL Pipeline Report

Generated: 2025-11-09 02:16:03

Project Information

Project Name:	Chicago Crash Data ETL Pipeline
Pipeline Type:	Medallion Architecture (Bronze → Silver → Gold)
Target Variable:	hit_and_run_i (Binary Classification)
Data Source:	City of Chicago Open Data Portal
Storage:	MinIO (Object Storage) + DuckDB (Analytics Database)
Orchestration:	RabbitMQ Message Queue

Gold Table Schema

Table Name: gold.crashes

Total Rows: 573

Latest Data Date: 2025-11-08

Latest Run: 2025-11-09 02:03:38.673922

Key Features for Hit & Run Prediction:

- Temporal: crash_hour, crash_day_of_week, is_weekend, hour_bin
- Location: latitude, longitude, grid_id, beat_of_occurrence
- Environmental: lighting_condition, weather_condition, roadway_surface_cond
- Context: posted_speed_limit, traffic_control_device, intersection_related_i
- Crash Details: num_units, injuries_total, crash_type, work_zone_i

Recent Pipeline Runs

Correlation ID	Rows	Status	Started	Ended
2025-11-09-02-03-35	298	Completed	2025-11-09 02:03:38	2025-11-09 02:03:38
2025-11-09-01-46-25	275	Completed	2025-11-09 01:46:29	2025-11-09 01:46:29

Error Summary

Correlation ID	Error Type	Message	Count
2025-11-09-02-03-35	None	No errors logged	0

Data Quality Metrics

Metric	Value
Missing Values %	0.00%
Duplicate Records	0
Hit & Run Rate	33.86%
Sample Size (for metrics)	573 rows