Analyzing NTSB Airplane Crash Reports

This tutorial demonstrates how to analyze National Transportation Safety Board (NTSB) airplane crash reports using PDF processing capabilities of certain LLM providers. We'll build a pipeline that extracts crash causes and synthesizes common patterns across incidents.

> **LLM Requirements**
>
> PDF processing is only supported with Claude (Anthropic) or Gemini (Google) models.

Dataset Overview

The dataset contains 689 NTSB airplane crash reports--the reports corresponding to fatal accidents after 2020. You can download it here: NTSB Airplane Crashes

Pipeline Overview

Our pipeline will:

1. Process PDF crash reports from the NTSB database to extract causes and recommendations
2. Synthesize common patterns across all analyzed crashes

Let's examine the pipeline structure:

```yaml
pipeline:
  steps:
    - name: analyze_crashes
      input: crashes
      operations:
        - extract_crash_cause # this is a map operation
        - synthesize_findings # this is a reduce operation
```

Full Pipeline Configuration

```yaml
datasets:
  crashes:
    type: file
    path: "fatal.json"

default_model: gemini/gemini-2.0-flash

operations:
  - name: extract_crash_cause
    type: map
    pdf_url_key: ReportUrl
    skip_on_error: true # Skip llm calls where the PDF is malformed or not
found
    output:
      schema:
        cause: str
        contributing_factors: "list[str]"
        recommendations: str
    prompt: |
      Analyze this NTSB airplane crash report and extract:
      1. The primary cause of the crash (2-3 sentences)
      2. Any contributing factors (list)
      3. Key safety recommendations made

  - name: synthesize_findings
    type: reduce
    reduce_key: _all
    output:
      schema:
        summary: str
    prompt: |
      Analyze the following airplane crash reports:

      {% for item in inputs %}
      Report {{loop.index}}:
      Cause: {{ item.cause }}
      Contributing Factors: {{ item.contributing_factors | join(", ") }}
      Recommendations: {{ item.recommendations }}

      {% endfor %}

      Generate a comprehensive analysis that:
      1. Identifies common causes across incidents
      2. Lists recurring contributing factors
      3. Synthesizes key safety recommendations
      4. Highlights any notable patterns

      Format your response as a structured report.

pipeline:
  steps:
    - name: analyze_crashes
      input: crashes
      operations:
        - extract_crash_cause
        - synthesize_findings
```

```
  output:
    type: file
    path: "crash_analysis.json"
    intermediate_dir: "checkpoints"
```

Sample Output

Here's the output we get from running the pipeline:

Sample Analysis Output

Airplane Crash Report Analysis:

1. Common Causes:

After analyzing the provided airplane crash reports, the most common primary causes include:

- Loss of Control (often due to aerodynamic stall, spatial disorientation, or pilot incapacitation)
- Engine Failure (often due to fuel exhaustion, mechanical issues, or improper maintenance)
- Controlled Flight Into Terrain (CFIT) (often in IMC or low visibility)
- Pilot Error (poor decision-making, failure to maintain airspeed, inadequate pre-flight planning).

2. Recurring Contributing Factors:

Several contributing factors recur across multiple reports:

- Improper Maintenance (inadequate inspections, incorrect repairs)
- Fuel Issues (fuel exhaustion, fuel contamination, improper fuel management)
- Adverse Weather Conditions (IMC, icing, turbulence, low visibility)
- Pilot Impairment (fatigue, alcohol/drug use, medical conditions)
- Failure to Maintain Airspeed (leading to stalls)
- Low Altitude Maneuvering
- Lack of Instrument Proficiency
- Poor Decision-Making (continuing flight into adverse conditions, improper risk assessment)
- Spatial Disorientation (particularly in IMC or at night)
- Inadequate Pre-flight Planning (weather, fuel, weight and balance).
- Exceeding Aircraft Limitations (weight, structural, etc.)

3. Synthesized Key Safety Recommendations:

Based on the analyzed reports, key safety recommendations can be synthesized:

- **Enhanced Pilot Training:**
  - Stall recognition and recovery techniques
  - Instrument meteorological conditions (IMC) flight procedures and spatial disorientation awareness.
  - Mountain flying techniques and high-density altitude operations

    - Emergency procedures training, particularly related to engine failures.
    - Aerobatic Maneuver training
- **Improved Maintenance Practices:**
    - Adherence to manufacturer's recommended maintenance schedules and procedures
    - Thorough inspections of critical components (fuel systems, control cables, engines)
    - Proper documentation of maintenance and repairs
    - Emphasis on proper installation and torquing of critical parts.
- **Robust Pre-flight Planning:**
    - Thorough weather briefings and in-flight weather monitoring
    - Accurate fuel planning and management
    - Weight and balance calculations
    - Familiarization with terrain, obstacles, and airport characteristics.
- **Sound Aeronautical Decision-Making:**
    - Avoid flying under the influence of alcohol or drugs
    - Avoid self-induced pressure to complete a flight
    - Recognize personal limitations and make conservative go/no-go decisions
    - Proper risk management
- **Effective Use of Technology:**
    - Installation and proper use of angle-of-attack indicators
    - Use of autopilot systems and electronic flight displays
    - Ensure aircraft has and is broadcasting ADS-B signals, and use traffic advisory systems when available
- **Awareness of Physiological Factors:**
    - Understanding of spatial disorientation and how to mitigate its effects
    - Awareness of the effects of fatigue, medical conditions, and medications on pilot performance.
    - Use of oxygen at night.
- **Adherence to Regulations and Procedures:**
    - Compliance with minimum safe altitudes and approach procedures
    - Proper use of checklists
    - Following air traffic control instructions.
- Maintain proper aircraft certification.

4. Notable Patterns:

- **VFR into IMC:** A significant number of accidents involve pilots without instrument ratings continuing visual flight into instrument meteorological conditions.
- **Loss of Control on Approach:** A recurring theme is loss of control during the approach phase, often related to stalls, wind shear, or unstable approaches.
- **Pilot Actions Under Stress:** Many accidents involve pilots making poor decisions under stressful situations, such as engine failures or adverse weather conditions.
- **Experimental Aircraft Issues:** Several reports involve experimental amateur-built aircraft, highlighting potential risks associated with construction, maintenance, and pilot familiarity.
- **Medical Incapacitation:** Several accidents were potentially caused by medical incapacitation of the pilot, suggesting that it may be necessary to have health safety standards, especially for older pilots.
- **Power Lines:** A concerning number of incidents involve collision with power lines during aerial application or low-altitude maneuvering.

The pipeline costs < $0.05 USD to run.