

Project Individual Review - Group 03

IPAI: Tiago Garcia - 54938

1. Original Data Sources & Integrated Schema

Clarity & Conciseness: The four data sources (NTSB, Kaggle traffic, Open-Meteo, aircraft production) are clearly described with attributes and preprocessing steps. However:

- Missing dataset sizes (e.g., row counts, unique models)
- No explanation of temporal resolution harmonization (hourly vs monthly data)
- Schema rationale for column removal is underjustified

Rating: 3/5

2. Adequacy of Selected Approaches

Methods are generally appropriate but lack validation:

- **Identity Resolution:** Robust blocking/similarity scoring but threshold (0.75) and 38 matches is probably an error
- **Weather Integration:** API use is efficient, but missing cleaning strategy for over-fetched data
- **Data Cleaning:** Manual date corrections are unscalable; automated methods preferred

Rating: 4/5

3. Project Weaknesses

Key Issues:

- Weakest paragraph: "For Open-Meteo API... no need for cleaning" (ignores missing geospatial data)
- Small matched sample (38 models) risks bias
- Temporal alignment between hourly/monthly data unexplained

Suggestions:

- Impute missing coordinates via airport IDs
- Validate matches with expert review
- Aggregate weather data to monthly averages

4. Project Strengths

- Comprehensive integration of 4 heterogeneous datasets
- Advanced identity resolution techniques (q-grams, multi-metric scoring)
- Policy-relevant focus on tourism/weather/manufacturing correlations

5. Improvements

- **New Data:** Add maintenance records/pilot reports
- **Automation:** Replace manual corrections with outlier detection
- **Validation:** Include precision/recall metrics for matches

6. Questions

- (a) How were hourly weather data aligned with monthly traffic timelines?
- (b) What percentage of NTSB aircraft models were matched (38 seems like a mistake)?
- (c) How do missing coordinates impact analysis validity?

7. Ratings

Component	Score (1-5)
Rigor	4/5
Complexity	4/5
Clarity	4/5
Overall	4/5

Final Evaluation (1-5): 4. I think this project has a very interesting topic and with the right modifications they can improve their work.