Pipeline Design Overview

The pipeline extracts customer journey data, processes it, sends it to the IHC API for attribution, and generates a marketing report. The core steps include:

1. **Check for Pre-Existing Data**
   - If `customer_journeys.json` exists, reuse it.
   - If `channel_reporting` exists, proceed to CSV generation.
   - Otherwise, send data to IHC API and populate `channel_reporting`.
2. **Customer Journey Extraction**
   - Extract session data for each conversion.
   - Store extracted journeys in `customer_journeys.json`.
3. **Sending Data to the IHC API**
   - Send journeys in chunks to comply with API limits.
   - Store results in `attribution_customer_journey`.
4. **Channel Reporting & Metrics**
   - Populate `channel_reporting` from relevant tables.
   - Compute **CPO** (Cost per Order) and **ROAS** (Return on Ad Spend).
5. **Final Report Generation**
   - Generate CSV with user confirmation.
   - Check data integrity using IHC sum condition.

Assumptions

- Database tables exist and have the required structure.
- Sessions before conversion timestamps are relevant.
- The API accepts and returns valid attribution results.
- Users confirm CSV generation before execution.

Potential Improvements

- Add better error handling and structured logging.
- Optimize database queries for performance.
- Implement retry mechanisms for API failures.
- Modularize code for better maintainability.
- Add validation to ensure data consistency.

These enhancements will improve the pipeline's robustness and efficiency.