## **Proposal: Structuring Event Data for Strategic Decision-Making at Chama**

In today’s data-driven economy, companies like Chama thrive on the ability to transform raw data into actionable insights. As Chama continues to scale its mobile application and backend infrastructure, the importance of high-quality, structured data becomes paramount. This proposal outlines a solution to a core data engineering challenge faced by the company: transforming semi-structured JSON event logs into structured tabular formats to support analytics, decision-making, and operational reporting.

The dataset provided for this project, case.json, contains a series of time-stamped app and backend event logs. Each entry includes an event type, a UTC timestamp, and a nested JSON string payload. While this format is efficient for logging and storage, it presents a barrier to downstream consumption by analysts, data scientists, and reporting systems. The solution proposed here is a pipeline that ingests this raw JSON, intelligently parses the payloads based on event type, and outputs three well-structured CSV files: CuratedOfferOptions.csv, DynamicPriceOption.csv, and DynamicPriceRange.csv.

The process begins with data ingestion and parsing. Each record’s Payload field is a stringified JSON object, requiring careful decoding before any meaningful transformation can begin. Based on the EventName field, each payload is routed to one of the three target outputs. Within each payload, relevant fields are extracted, renamed where necessary, and formatted according to strict data type and quoting conventions. This level of attention to detail ensures the CSVs can be immediately used in business intelligence tools or loaded into a relational database.

One of the more nuanced requirements of the project is handling the EnqueuedTimeUtc timestamp. As the data originates in UTC, but business operations are based in Brazil (UTC-3), it was critical to convert all timestamps to the correct local timezone. These were then formatted into the Brazilian standard date format (DD/MM/YYYY), ensuring consistency and clarity for all downstream users.

A key objective of this project was to demonstrate data quality and reproducibility. With this in mind, the transformation logic was written in clean, modular Python code. Each component of the pipeline—parsing, transformation, formatting, and export—was handled by dedicated functions, making the system easy to test, maintain, and extend. If new event types are introduced in the future, they can be added to the existing framework with minimal disruption.

In practical terms, this project serves as a foundational building block for Chama’s data platform. The structured outputs allow for immediate integration into analytics dashboards, pricing algorithm evaluation, and operational monitoring. Over time, this transformation pipeline could evolve into a scheduled ETL job or be incorporated into a broader data orchestration system such as Apache Airflow or dbt.

In conclusion, this data transformation project aligns directly with Chama’s mission to make metrics-driven decisions across all levels of the organization. By converting noisy, nested event logs into clean, consumable CSV files, we enable every department—product, operations, marketing, and executive leadership—to work with reliable data. This project is not just a technical exercise, but a strategic enabler of growth, optimization, and insight at Chama.