Kaggle Writeup Submission

**Project name**

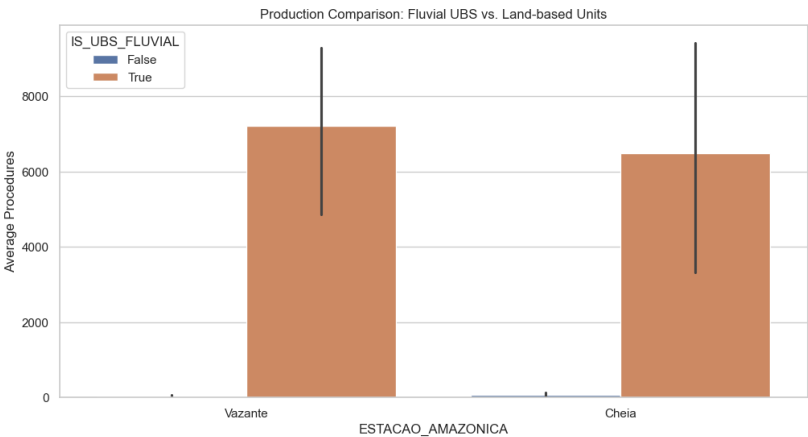**Sentinela Fluvial (Fluvial Sentinel): Predictive Health Monitoring in the Amazon with MedGemma**

**Your team**

- **Richardson Allan Ferreira de Souza** - Data Scientist & AI Engineer (ETL Pipeline, Prompt Engineering, and LLM Integration).
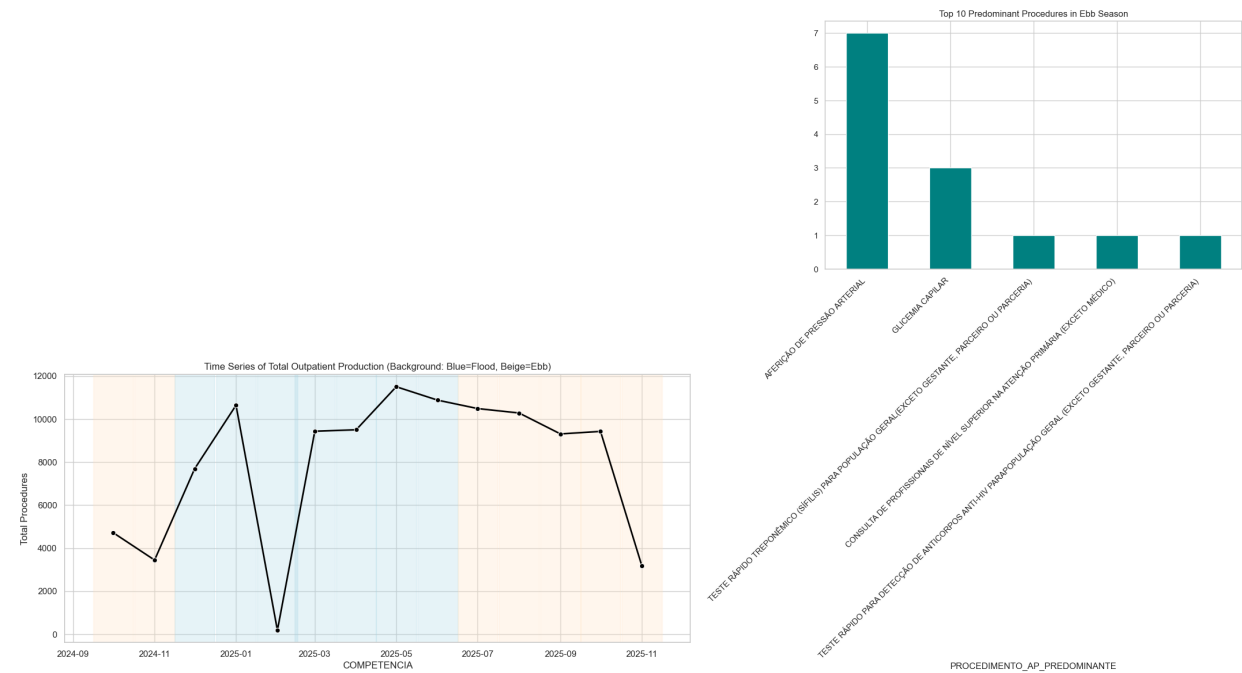
**Problem statement**

**Bridging the Amazonian Data Desert**
The Brazilian Amazon relies on Riverine Basic Health Units (UBSF) to reach isolated communities. However, traditional health dashboards fail here because they ignore the **Amazonian hydrological cycle**. As proven by our EDA, seasonality is a clinical barrier: ambulatory production experiences massive variance between *Vazante* (dry) and *Cheia* (flood) seasons. Fluvial units show extreme operational vulnerability compared to terrestrial units.



A 90% drop in clinical output in an urban center might suggest low demand; in the Amazon, our data shows it signals a geographical crisis. During extremes, markers like Capillary Glycemia and Rapid Testing plummet, blocking chronic patients from care and leaving outbreaks undetected. Currently, management is retroactive, analyzing data months late. *Sentinela Fluvial* shifts this to a predictive paradigm, translating tabular statistical anomalies—like the 98% drop in consultations detected in Manicoré—into immediate logistical interventions.

Top 10 Predominant Procedures in Ebb Season



Time Series of Total Outpatient Production (Background: Blue=Flood, Beige=Ebb)

By cross-referencing statistical drops with environmental constraints, MedGemma helps prevent logistical bottlenecks from evolving into public health tragedies, such as preventable hospitalizations (CSAP).

**Overall solution:**

**Context-Aware Clinical Logistics with MedGemma**
Amazonian health outposts operate in disconnected environments. We utilized the `google/medgemma-4b-it` model because its lightweight 4B architecture allows local offline deployment while possessing deep clinical instruction-tuning.

Our architecture is a hybrid RAG pipeline:

1. **Trigger:** A deterministic algorithm monitors DataSUS tables for statistical anomalies.
2. **Reasoning:** The anomaly is injected into a Few-Shot prompt enriched with the `ESTACAO_AMAZONICA` variable. MedGemma acts as a "Virtual Sanitary Doctor," resolving the friction between clinical necessity and logistical reality.
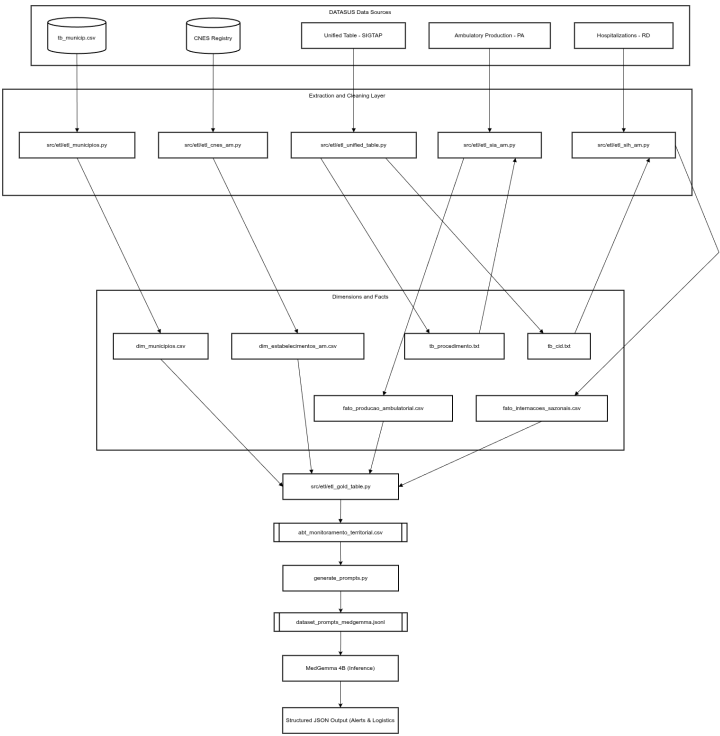
**HAI-DEF and MedGemma Capability** Generic rules cannot interpret logistical nuance. MedGemma generates specific, safe interventions outputted as **JSON objects**.

- **In Tefé (Severe Drought):** Instead of a generic "send a boat," MedGemma recommended *"remote monitoring via portable devices and local agents"* for isolated diabetics.
- **In Novo Aripuanã (Extreme Floods):** The model suggested *"deploying adapted draft boats to perform Rapid Tests in floating health posts"* due to submerged docks.
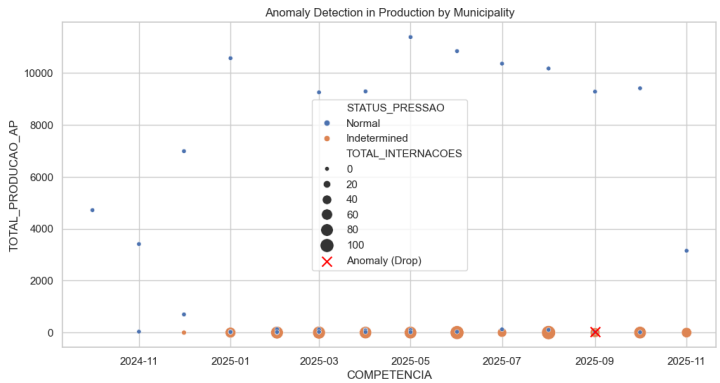
**Technical details**

**Product Feasibility, Architecture, and Execution**

*Sentinela Fluvial* combines deterministic data engineering with generative reasoning.

Our Bronze-Silver-Gold pipeline extracts DATASUS data to systematically generate Tabular RAG prompts.

**1. Data Engineering (Bronze to Gold):** Using Python/Pandas, we process raw CNES, SIA, and SIH files. The output is an Analytical Base Table (`abt_monitoramento_territorial.csv`) where we calculate Z-scores for anomaly detection, flag `IS_UBS_FLUVIAL`, and classify seasons.



Z-scores computationally isolate access failures (marked red). These anomalies are transformed into 'Computational Intelligence' within the prompt, ensuring focus and efficiency.

**2. Context Serialization & Tuning:** We developed a middleware (`generate_prompts.py`) that constructs Few-Shot Prompts including "90-day Clinical Memory" and "Epidemiological Intelligence." By using strict prompt engineering, we enforce MedGemma to output exclusively in JSON for frontend integration.

**3. Explainable AI (XAI):** Accountability is ensured through a `situational_analysis` field. Before outputting a strategy, MedGemma must write out its clinical reasoning (Chain of Thought), providing managers with an interpretable audit trail. This Edge AI approach guarantees patient privacy and high availability without cloud dependency.