D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

End-to-End AI & BI Workflow: From Collection to Insights

Overview

This project demonstrates a complete **end-to-end data workflow**, integrating data engineering, machine learning, and business intelligence. It automates data collection, processes sentiment analysis, and visualizes insights in a dynamic dashboard. The goal is to show what full-stack data work looks like—ensuring alignment between expectations and execution.

Many companies expect a **data engineer** to build models or a **data scientist** to manage infrastructure. The reality is, these roles are different—but if you need someone who can do it all, make sure they actually can. This project is my way of demonstrating that expertise.

Project Workflow

1. Data Collection

- **Source:** News headlines are collected daily using the **NewsAPI**.
- **Automation:** A scheduled pipeline using **Google Cloud Scheduler** triggers data collection.
- **Processing: Google Cloud Run** hosts a Python-based service that fetches and processes headlines.
- **Storage:** The data is stored in **Google BigQuery**, allowing for scalable processing.

2. Sentiment Analysis (AI/ML Processing)

- **Libraries Used:** `pandas`, `numpy`, `google.cloud.bigquery`, `nltk`, `nltk.sentiment.SentimentIntensityAnalyzer`
- **Tool:** Python-based sentiment analysis model using **Google Vertex AI**.
- **Processing:** The model analyzes the sentiment of each news headline.

- **Integration:** Results are pushed back to **BigQuery** for further analysis.

3. Data Storage & Analysis

- **Database:** BigQuery acts as the central data warehouse.
- **Transformation:** SQL queries process the sentiment-scored headlines.
- **Aggregation:** Headlines are categorized by sentiment and topic for visualization.

4. Data Visualization

- **Tool:** Looker Studio for interactive dashboards.
- **Live Updates:** Direct SQL connections ensure real-time insights.
- **User Experience:** The dashboard allows filtering by **time, sentiment, and category**.

Key Features

✅ **Automated Data Collection** – NewsAPI fetches real-time headlines. ✅ **Google Cloud Scheduler & Cloud Run** – Automates and scales data pipelines. ✅ **Scalable Sentiment Analysis** – Python model runs in **Vertex AI**. ✅ **BigQuery for Storage & Processing** – Ensures efficient data handling. ✅ **Looker Studio for Dynamic Insights** – Real-time, interactive dashboards.

Tech Stack

- **Data Engineering:** Python, Google Cloud Functions, Google Cloud Run, Google Cloud Scheduler, Google BigQuery
- **AI/ML:** Vertex AI, Natural Language Processing (NLP) with Python
- **Business Intelligence:** Looker Studio (Google Data Studio), SQL

Project Links

D. Saul Jameson
https://www.linkedin.com/in/dsauljameson/

🔗 **Code & Documentation**
📌 Engineering Code
📌 AI/ML Code
📌 Live Looker Dashboard
📌 Full Documentation