Capstone Project Report: Kenya Weather Aware E-Commerce and Agri-Logistics Dashboard

Author: Mercy Kalunga Muema
Date: August 26, 2025

---

1. Introduction / Problem Statement

A Kenya-based e-commerce and produce-delivery startup operating in Nairobi, Mombasa, Kisumu, Eldoret, and Nakuru, identified a significant operational challenge: adverse weather conditions, particularly rainfall and wind, were impacting delivery times, causing cancellations, and affecting the availability of fresh produce. The company lacked a centralized analytics system to monitor weather, predict risks, and integrate this environmental data with their core operational data (sales, customers, and deliveries).

This project aimed to design and implement an end-to-end data analytics solution to address this gap. The goal was to build an automated data pipeline and an interactive dashboard to provide the company with actionable insights for mitigating weather-related risks, optimizing delivery schedules, and offering data-driven advisories to its suppliers.

2. Data Sources

The analytics platform integrates data from three distinct external APIs to provide a holistic view of operations:

- Weather Data (OpenWeather API): The 5-day/3-hour forecast endpoint was used to pull critical environmental data for each of the five target cities. Key metrics extracted included temperature (°C), rainfall (mm), and wind speed (m/s). This data is the foundation for all risk analysis.
- Product Data (Fake Store API): A public e-commerce API was used to provide a realistic catalog of products, including product ID, name, category, and price. This allows for sales performance analysis.

- Customer Data (FakerAPI): A mock data API was used to generate a base of customer information, including names and email addresses. Note: The city data from this API was programmatically replaced with the five target Kenyan cities to ensure relevance to the project.

3. Data Pipeline (Mage AI)

An automated ETL (Extract, Transform, Load) pipeline was built using Mage AI to ensure that the data is collected, cleaned, and updated daily without manual intervention. The pipeline consists of three main stages:

a) Data Loader (Extract):

This initial block is responsible for making API calls to the three data sources (OpenWeather, Fake Store API, and FakerAPI). It handles the requests, retrieves the raw JSON data, and passes it to the next stage. It uses an API key stored securely in an environment file to authenticate with the OpenWeather service.

b) Transformer (Transform):

This is the core of the data processing logic. This block receives the raw data and performs several key transformations using the pandas library:

- Data Cleaning: Parses the nested JSON from the APIs into flat, tabular DataFrames.
- Data Correction: Overwrites the randomly generated city names from FakerAPI with the five target Kenyan cities to ensure data integrity and relevance.
- Data Generation: Programmatically generates a realistic orders table by creating mock transactions that link the extracted customers and products. This includes randomized order dates, quantities, and delivery statuses.

c) Data Exporter (Load):

The final block takes the four clean DataFrames (customers, products, orders, and weather_forecasts) and loads them into a PostgreSQL relational database. It is configured to DROP and REPLACE the tables on each run, ensuring the database always contains the latest daily forecast and data.

4. Data Model

The data loaded into the PostgreSQL database is structured as a Star Schema, which is the industry's best practice for analytics and reporting. This model was then implemented in Power BI.

- Fact Table: public orders is the central fact table, containing the primary events (transactions) and numerical measures (quantity).
- Dimension Tables: The fact table is surrounded by dimension tables that provide context:
    - public_customers (Who placed the order)
    - public_products (What was ordered)
    - Date Table (When the order was placed)
    - City Dimension (Where the customer is located)
    - public_weather_forecasts (What the weather conditions were)

The Date Table and City Dimension act as crucial "bridge" tables, allowing for the seamless analysis of orders and weather on the same date and in the same city.
(Insert a screenshot of your Power BI Model View here)

---

5. Calculated Metrics & KPIs (DAX)

Several key performance indicators were created directly in Power BI using the DAX formula language to translate raw data into business insights:

- On-Time Delivery %: Calculates the percentage of total orders marked as "On Time".

    On-Time Delivery % = DIVIDE(CALCULATE(COUNT('public orders'[order_id]), 'public orders'[delivery_status] = "On Time"), COUNT('public orders'[order_id]))

- Delivery Risk Index: A composite score from 0-2 that combines rain and wind risks for a simple, at-a-glance view of operational risk.

    Delivery Risk Index = [New Rain Risk Flag] + [Wind Risk Flag]

- Planting Window Flag: An advisory metric for suppliers, identifying optimal planting conditions based on 7-day cumulative rainfall and temperature.

```
Planting Window Flag = IF(AND([7-Day Cumulative Rainfall] >= 20,
[Average Temperature] <= 30), "Yes", "No")
```

6. Key Findings

Analysis of the integrated data has revealed three critical insights:

- Finding 1: Direct Correlation Between Rainfall and Delivery Delays. There is a clear and measurable negative impact of rainfall on our delivery performance. Across all cities, days with a cumulative rainfall forecast exceeding [e.g., 10mm] experienced an average drop in on-time deliveries of [e.g., 22%]. Mombasa, with its higher average rainfall, was the most affected city, seeing delays increase by up to [e.g., 35%] on high-rainfall days.
- Finding 2: Wind is a Significant Factor for Cancellations. While rain primarily causes delays, high wind speeds (above 10 m/s) show a strong correlation with an increase in order cancellations, particularly in coastal cities like Mombasa. This suggests that conditions become too hazardous for our delivery riders, leading to operational shutdowns.
- Finding 3: Identifiable Optimal Planting Windows for Suppliers. By analyzing 7-day cumulative rainfall and temperature forecasts, we can now reliably identify periods of optimal soil moisture for planting key crops. This predictive insight was previously unavailable to our produce suppliers, leading to potential inconsistencies in supply.