Bioproduction Data Platform

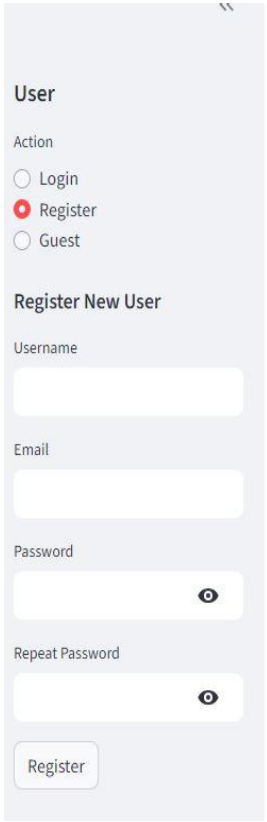
Sample Program

Shine Jose

PACKAGES USED

```python
# app.py
import streamlit as st
import pandas as pd
import numpy as np
import os
from datetime import datetime
import sqlite3
import hashlib
import secrets
import plotly.express as px
import plotly.graph_objects as go
import sys

# -------------------------
# Fix for PySpark on Windows
# -------------------------
os.environ["PYSPARK_PYTHON"] = sys.executable
os.environ["PYSPARK_DRIVER_PYTHON"] = sys.executable

from pyspark.sql import SparkSession

from src.synthetic import generate_multiple_runs
from src.etl import save_parquet, save_csv_for_powerbi
from src.quality import compute_quality_metrics
from src.models import train_anomaly_detector, detect_anomalies

# -------------------------
# Initialize Spark
# -------------------------
spark = SparkSession.builder \
    .appName("Bioproduction Data Platform") \
    .config("spark.driver.memory", "4g") \
    .getOrCreate()

"
```

REGISTRATION

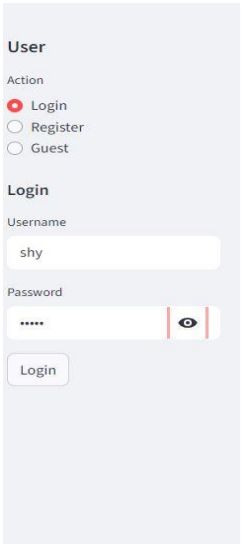


Bioproduction Data Platform

LOGIN

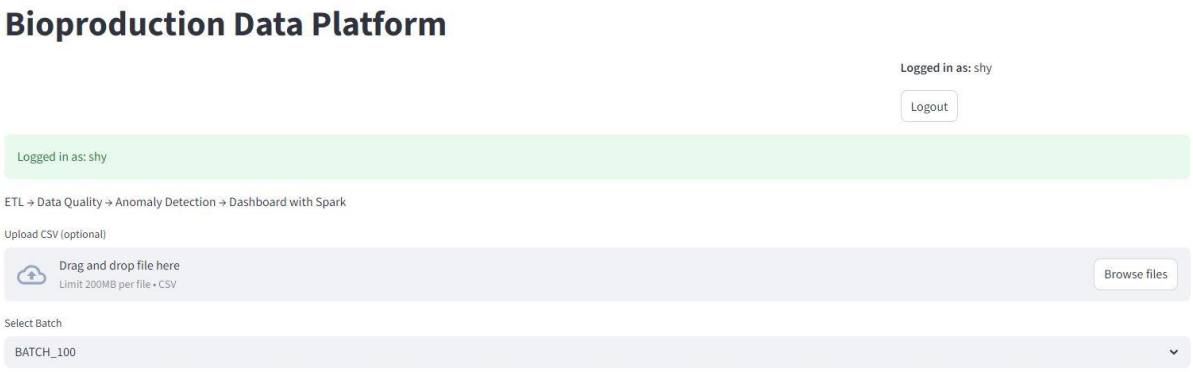

User
Action
Login
Register
Guest
Login
Username
Password
Login
Bioproduction Data Platform
Not logged in

LOGIN

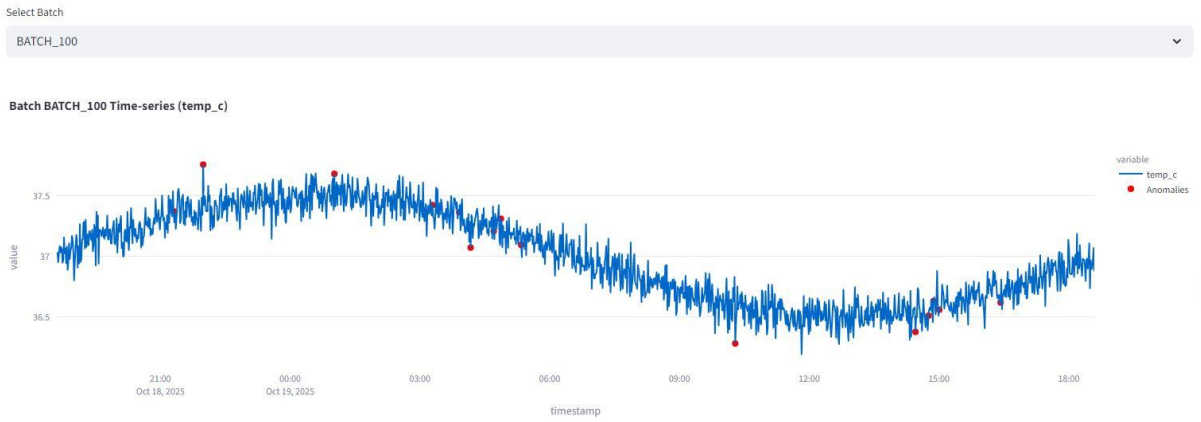


Bioproduction Data Platform

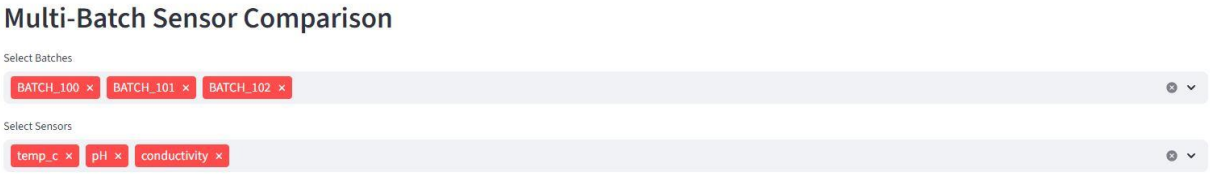
LOGIN


Bioproduction Data Platform
Logged in as: shy
Logout
Logged in as: shy
ETL → Data Quality → Anomaly Detection → Dashboard with Spark
Upload CSV (optional)
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Select Batch
BATCH_100

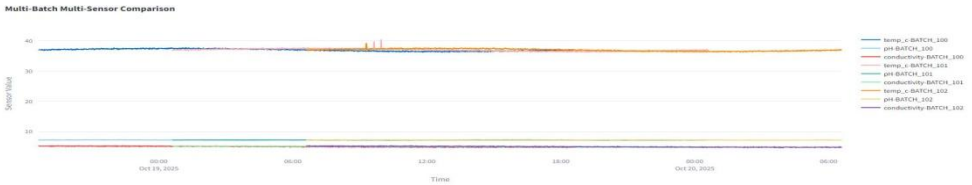
BATCH SELECTION



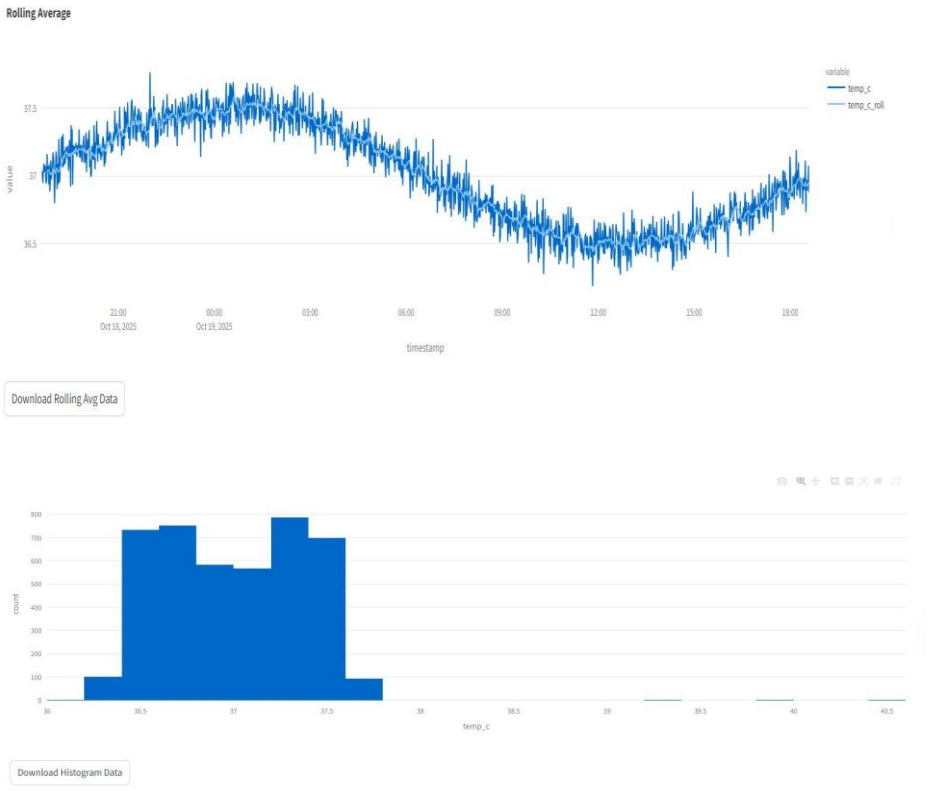
MULTI-BATCH SELECTION



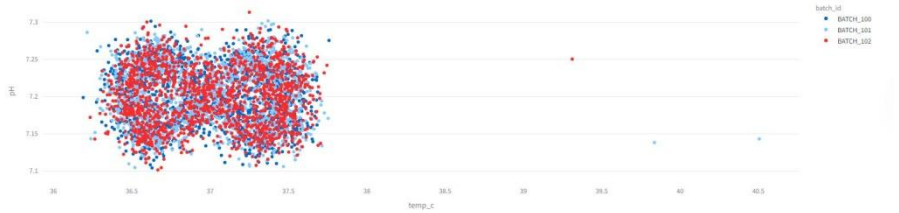
MULTI-BATCH SELECTION



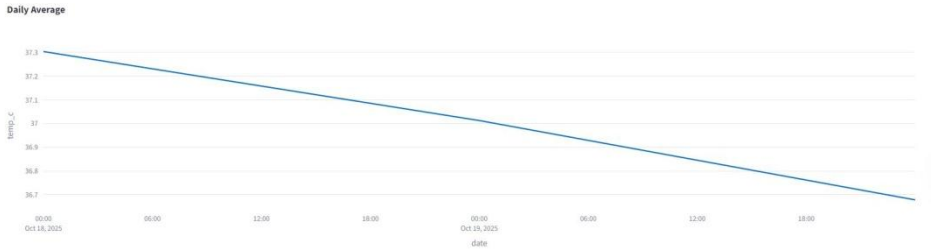
MULTI-BATCH SELECTION



MULTI-BATCH SELECTION



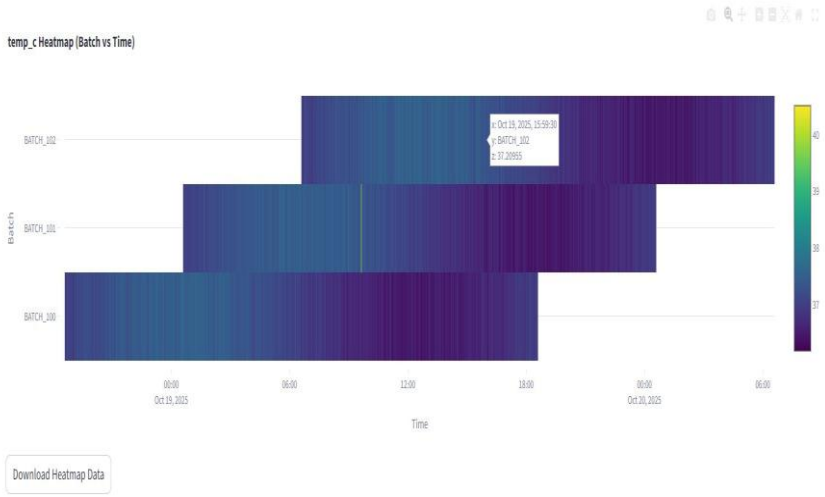
Download Scatter Data



Download Daily Avg Data

MULTI-BATCH SELECTION

MULTI-BATCH SELECTION



DOWNLOAD REPORTS

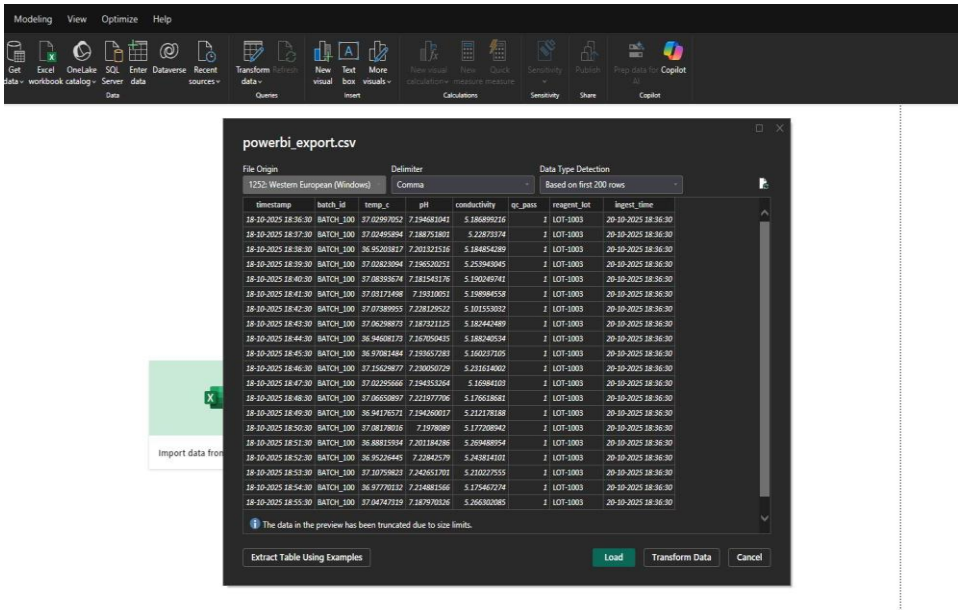
- Connect Reports To POWER BI

DOWNLOAD REPORTS

- Connect Reports To POWER BI

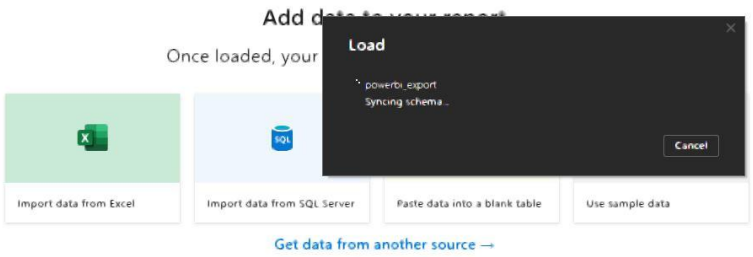
POWER BI

- Connect Reports To POWER BI



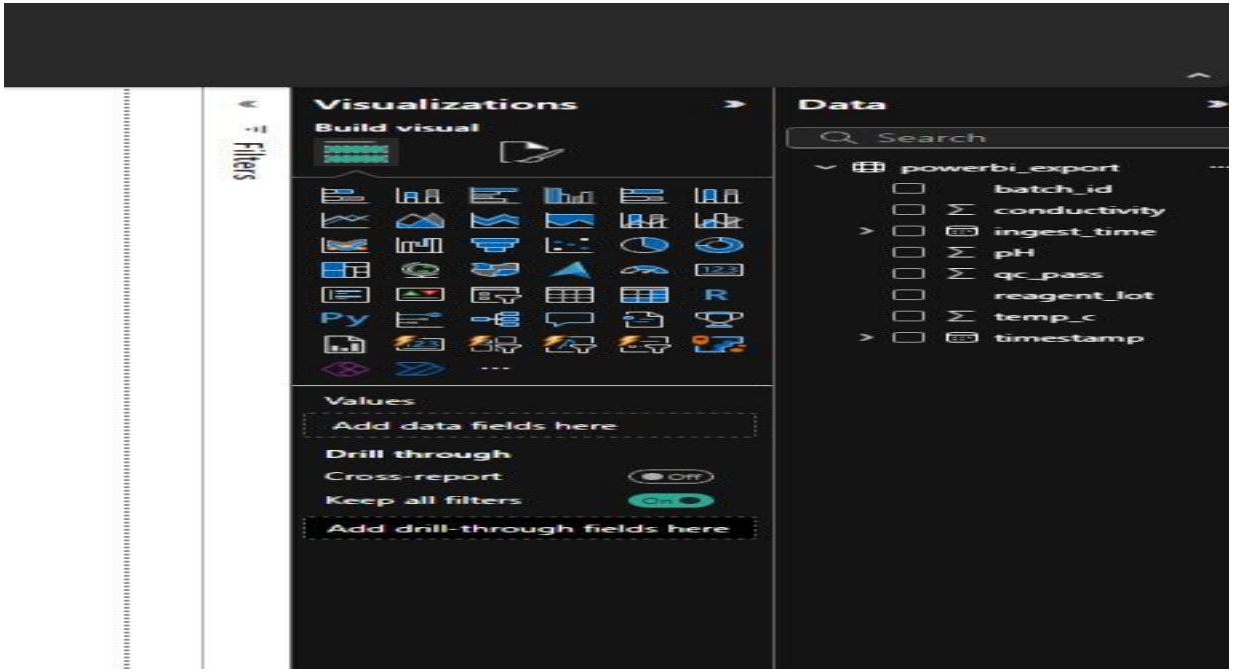
POWER BI

- LOADING DATA



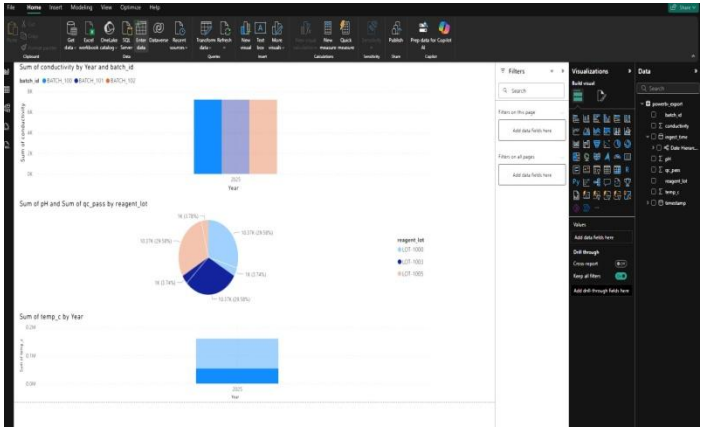
POWER BI

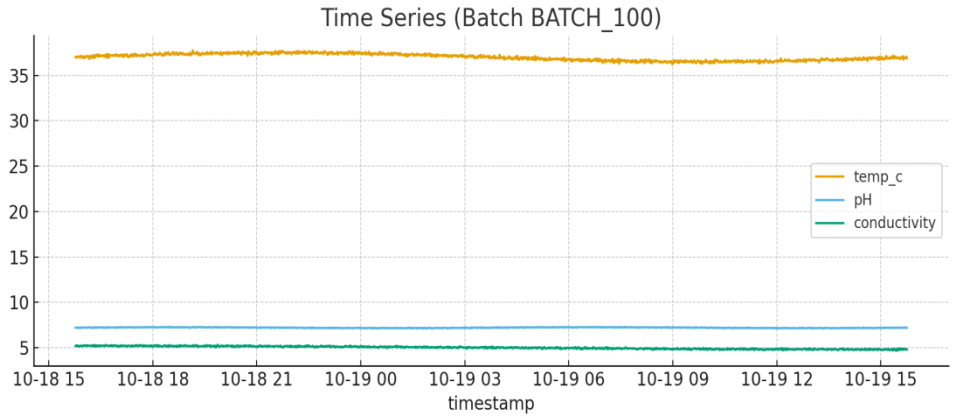
- LOADING DATA



POWER BI

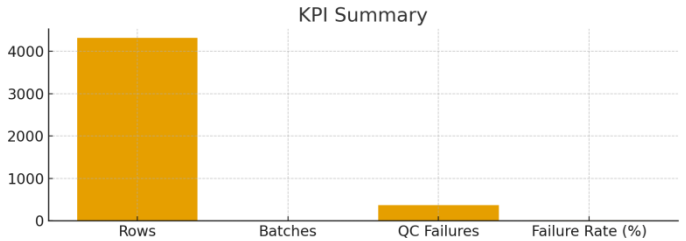
- DATA PRESENTATION IN POWER BI





Batch QC Summary

batch_id	rows	qc_failures
BATCH_100	1440	128
BATCH_101	1440	130
BATCH_102	1440	115



- Thank You