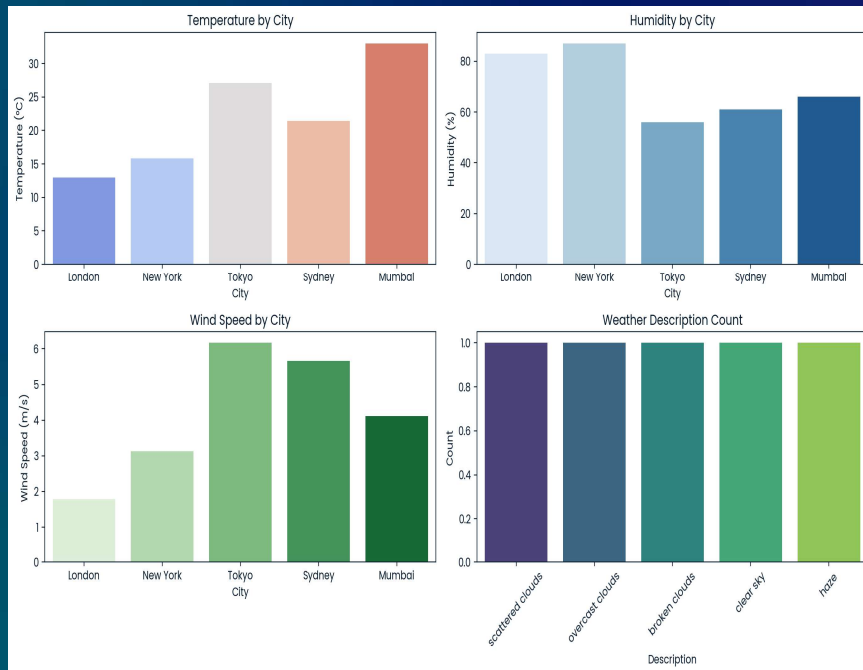


DATA INGESTION PIPELINE



TRANSFORMING WEATHER DATA

- A comprehensive ingestion pipeline

PRODUCT OVERVIEW



GOAL

To build an end-to-end data ingestion pipeline that fetches, processes, and stores weather data for further analysis and reporting



FOCUS

Automating the data collection from OpenWeather API, processing it, and storing it in a PostgreSQL database using Docker and Airflow.

MY ROLE

PROJECT PLANNING

Defined the project scope, objectives, and timeline

DOCUMENTATION

Created detailed documentation and a comprehensive README file

DEVELOPMENT

Implemented the data ingestion pipeline, developed Airflow DAGs, and configured Docker containers

TESTING AND DEBUGGING

Ensured data accuracy, reliability, and handled any technical issues.



TECHNICAL CHALLENGES

API LIMIT RATE

Managing API rate limits and ensuring data consistency

DOCKER CONFIGURATION

Setting up and configuring Docker containers for Airflow and PostgreSQL

ERROR HANDLING

Implementing robust error handling and logging within the data pipeline.

DATA TRANSFORMATION

Implementing robust error handling and logging within the data pipeline.

SOLUTION APPROACH



**DATA
INGESTION**



Used Python to fetch weather data from OpenWeather API

ORCHESTRATION



Employed Apache Airflow to schedule and manage the ETL pipeline.

CONTAINERIZATION



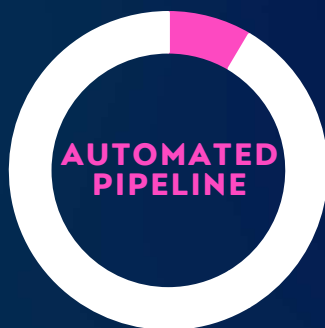
Utilized Docker to containerize the application, ensuring consistent environments.

DATA STORAGE



Stored processed data in a PostgreSQL database, ensuring it's readily accessible for analysis.

RESULT AND IMPACT



Successfully automated the data ingestion and processing pipeline



Built a scalable solution that can handle varying data loads



Provided easy access to processed weather data for analytics and reporting



Reduced manual intervention, allowing for continuous data updates and improved data accuracy



LESSONS LEARNED

DOCKER AND API INTEGRATION

- **Docker Mastery:** Gained in-depth understanding of Docker and container orchestration.
- **API Integration:** Learned best practices for integrating and handling third-party APIs.

ERROR HANDLING AND WORKFLOW MANAGEMENT

- **Error Handling:** Developed skills in implementing robust error handling and logging mechanisms.
- **Workflow Management:** Enhanced my ability to design and manage data workflows using Airflow.



ADAPTABILITY

EVOLVING DATA SOURCE

Designed the pipeline to accommodate additional data sources with minimal changes

DYNAMIC CONFIGURATIONS

Designed the pipeline to accommodate additional data sources with minimal changes

SCALABLE ARCHITECTURE

Built a solution that can scale horizontally to handle increased data volumes and complexity

TECHNICAL TOOLS

PYTHON



Used for data fetching and processing scripts

AIRFLOW



Managed and scheduled ETL workflows

DOCKER



Containerized the application for consistency and scalability

POSTGRESQL



Stored processed data for analysis and reporting.

OPENWEATHER API



Source of data

OpenWeather

GITHUB: [abaya12/weather-data-ingestion-pipeline \(github.com\)](https://github.com/abaya12/weather-data-ingestion-pipeline)





SUMMARY

- ❖ **Expertise:** Demonstrated strong abilities in handling complex data engineering challenges.
- ❖ **Innovation:** Applied innovative solutions to create a robust and scalable data ingestion pipeline.
- ❖ **Future-Ready:** Equipped to handle evolving data sources and project requirements.

SPONSORS



inngen

FACILITATOR

- DEREK DEGBEDZUI

THANK YOU

Ishmael Kabu Abayateye

+233 243 0692 37

ishmaelloabtkb19@gmail.com