Data Engineering Project

1. Data
2. Tools
3. Method
4. Documentation

Data: CSV files are available – Credit Score Analysis

Tools: Python, Data Lake, Data Warehouse, Dashboard, Hosting

1. Data Lake and Data Warehouse need databases. PostgreSQL is one widely used option.
2. Dashboard using plotly.
3. Hosting with Docker
4. Python to develop code
5. Airflow for orchestrating

PROGRESS SO FAR:

1. Docker compose with Airflow and PostgreSQL database is up and running
2. Data cleaning and transforming scripts are ready and working
3. Airflow DAGs have been developed and tested and they work fine!!!
4. End to End data engineering project is almost complete.
5. We need to add data visualization like a dashboard using Plotly.

Article related screenshots:

Airflow UI with all tasks successfully executed.

