Meets Specifications

Congrats student!!! 👏🏼👏🏼👏🏼👏🏼👏🏼

Your project meets all rubric points so I will mark it as ready to pass, you've done a great job building an ETL pipeline 💪🏼

All the best,

🎓

Write Up

The write up includes an outline of the steps taken in the project.

The purpose of the final data model is made explicit.

Good start!!, the write up includes an outline of the steps taken in the project. OK

The write up describes a logical approach to this project under the following scenarios:

- **The data was increased by 100x.**
- **The pipelines would be run on a daily basis by 7 am every day.**
- **The database needed to be accessed by 100+ people.**

Good work providing some ideas to address these three scenarios OK

As a general rule, public cloud services are a perfect match for these scenarios

The choice of tools, technologies, and data model are justified well.

Good!! Apache Airflow and Public Cloud services (like AWS) are a perfect choice for these kind of problems OK

Execution

All coding scripts have an intuitive, easy-to-follow structure with code separated into logical functions.

Naming for variables and functions follows the PEP8 style guidelines. The code should run without errors.

The project includes at least two data quality checks.

Good work implementing some quality checks OK

As you did learn, data quality is a critical factor in any BigData/ML/DL project

- **The ETL processes result in the data model outlined in the write-up.**
- **A data dictionary for the final data model is included.**
- **The data model is appropriate for the identified purpose.**

The project includes:

- **At least 2 data sources**
- **More than 1 million lines of data.**
- **At least two data sources/formats (csv, api, json)**

Good job working with with these data sources as these data sets meet this rubric point OK