Data Engineering Capstone

REVIEW | CODE REVIEW | HISTORY

Meets Specifications

Great job, you are ready to go! 💯 Clearly, you have acquired all the important concepts from this project. Wish you all the best for the upcoming projects! 🤞

Tip: If you are interested in knowing more about data modeling in the real world, please read this series of posts

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 job on explaining the goal for this data modeling.👍 Analyzing US visitors experience is a good example of this project!

✓ **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 answers for each question, well done! It's a good way to leverage cloud services to address each challenge. But I would encourage you to think more, for example, when the data was increased by 100x, do you store the data in the same way? If your project is heavy on reading over writing, how do you store the data in a way to meet this requirement? What if the requirement is heavy on writing instead?
Tip:

- If you are interested in knowing more about the heavy-read system design, I would suggest that you read this tutorial about Netflix system.
- I would suggest that you read this experience sharing about the lesson learned when implementing data pipeline

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

Good job on providing your reasonings about each tool you use and justifying the data model well👍

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 code is clean and easy to follow👍

✓ **The project includes at least two data quality checks.**

You correctly design the required data quality checks.💯

✓

- **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.**

Good job providing all the information in your readme file.👍

✓ **The project includes:**

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

Your project meets all the requirements.👏

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review