Requires Changes

**1 SPECIFICATION REQUIRES CHANGES**

You’ve made a great effort in your work in the project! Your submission needs some effort from your side in terms of meeting specifications, there are a few issues that need to be revised before the project can pass all rubric points. See below for some suggestions on revisions that can be made for your next project submission!

**ETL**

**The script, etl.py, runs in the terminal without errors. The script reads song\_data and load\_data from S3, transforms them to create five different tables, and writes them to partitioned parquet files in table directories on S3.**

The script, etl.py, runs in the terminal without errors. The script reads song\_data and load\_data from S3, transforms them to create five different tables, and writes them to partitioned parquet files in table directories on S3.

**Each of the five tables are written to parquet files in a separate analytics directory on S3. Each table has its own folder within the directory. Songs table files are partitioned by year and then artist. Time table files are partitioned by year and month. Songplays table files are partitioned by year and month.**

Please find the review comments in the code review section.

**Each table includes the right columns and data types. Duplicates are addressed where appropriate.**

Correct columns and datatypes are used for each table. Great job dropping duplicates where appropriate! To make this even better, consider how you’d want to handle duplicate entries with the same id but different values for other columns. For example, two user entries with the same information except for the level? This could happen when a user upgrades from a free to a paid level. Would you want to keep both entries or just the one with the most recently updated level?

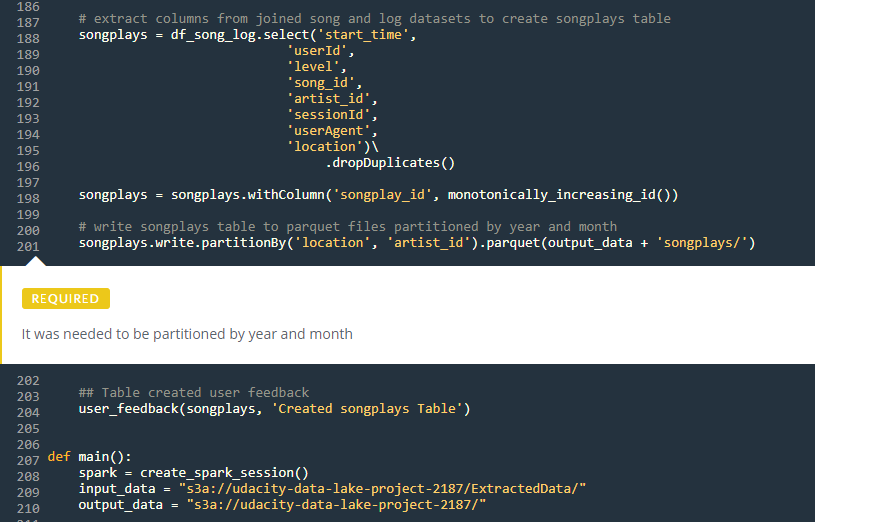
**Code Quality**

**The README file includes a summary of the project, how to run the Python scripts, and an explanation of the files in the repository. Comments are used effectively and each function has a docstring.**

Great job including information on the project, instructions, and files in your README file! Consider adding some examples in your README to demonstrate the use of your data lake. You can add code blocks or images that show example queries and results that the analytics team would find useful. Great job with your clear and concise docstrings and comments!

**Scripts have an intuitive, easy-to-follow structure with code separated into logical functions. Naming for variables and functions follows the PEP8 style guidelines.**

Code follows an organized l and clear variable and function names are used.



Meets Specifications

Dear Student,

Thank you for the extremely extraordinary effort you have put into this project. I'm impressed!  
I must tell you that this is one of the most well-formatted projects I ever happen to review.  
We look forward to no less than this great submission for the coming projects.  
Good luck and happy learning.

**Friendly Note:**  
Stay safe and take care of yourself and all your beloved ones in the current circumstances :)

**ETL**

**The script, etl.py, runs in the terminal without errors. The script reads song\_data and load\_data from S3, transforms them to create five different tables, and writes them to partitioned parquet files in table directories on S3.**

Runs with no single issue, thanks a lot! :)

**Each of the five tables are written to parquet files in a separate analytics directory on S3. Each table has its own folder within the directory. Songs table files are partitioned by year and then artist. Time table files are partitioned by year and month. Songplays table files are partitioned by year and month.**

You did exactly as requested by my fellow previous reviewer, thank you for implementing this functionality! :)

**Each table includes the right columns and data types. Duplicates are addressed where appropriate.**

Great effort here, thank you! :)

**Code Quality**

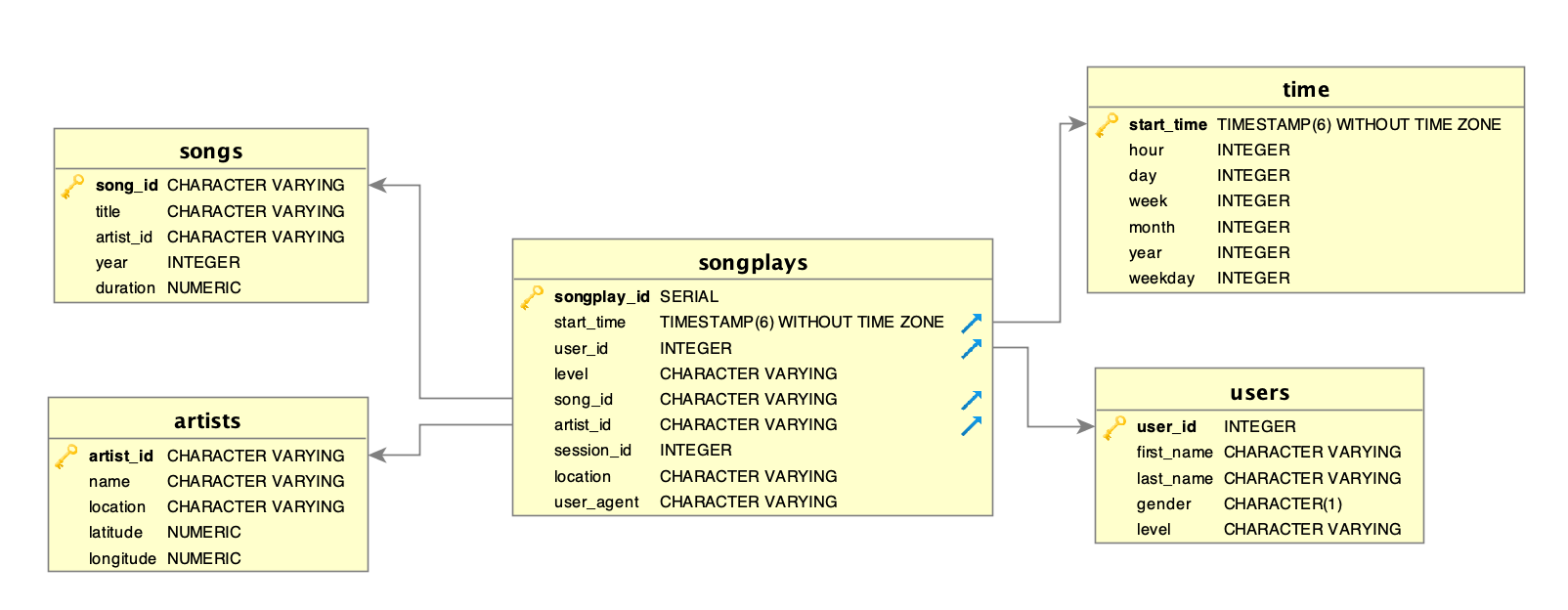
**The README file includes a summary of the project, how to run the Python scripts, and an explanation of the files in the repository. Comments are used effectively and each function has a docstring.**

Outstanding **README** file, you did exactly as requested by the rubric. Thank you for implementing that :)

Here are some general tips to follow whenever creating a **README** file for this project or for similar ones:

A nice README is a great way to showcase your project to potential employers. Following sections provide better readability:

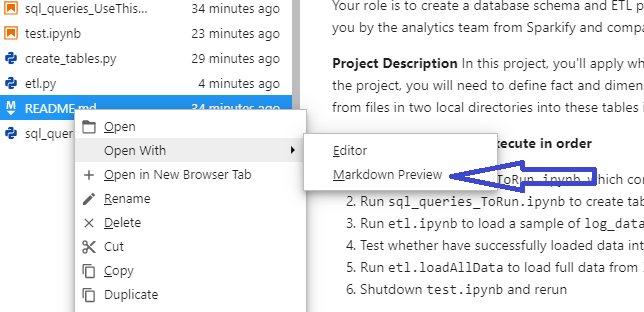
* Introduction - purpose of project; in this case, what is Sparkify, how this project is going to help it.
* Database schema design and ETL process
* Files in repository
* How to run the python scripts
* Add a screenshot or an image (ER Diagram) showing how the fact and dimension tables are connected. Example:

[](https://udacity-reviews-uploads.s3.us-west-2.amazonaws.com/_attachments/339318/1586016120/Song_ERD.png)

* Properly use the markdown language in your README, for headings, lists, subsections. Refer: <https://guides.github.com/features/mastering-markdown/>
* You can use an online markdown editor like <https://dillinger.io/>
* Refer good READMEs on web:

<https://github.com/matiassingers/awesome-readme>  
<https://bulldogjob.com/news/449-how-to-write-a-good-readme-for-your-github-project>  
<https://medium.com/@meakaakka/a-beginners-guide-to-writing-a-kickass-readme-7ac01da88ab3>

* You can preview your README in your Udacity workplace as below:

[](https://udacity-reviews-uploads.s3.us-west-2.amazonaws.com/_attachments/339318/1586016113/preview.png)

**Scripts have an intuitive, easy-to-follow structure with code separated into logical functions. Naming for variables and functions follows the PEP8 style guidelines.**

* Your project code is clean and for the most part follows the PEP8 style guidelines.
* I can clearly see your code structured into logical functions. Your function names clearly specify what your code is going to do. Great effort here!