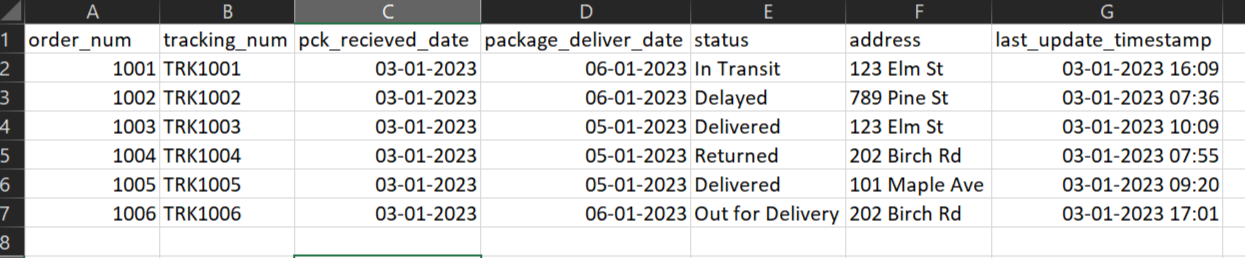
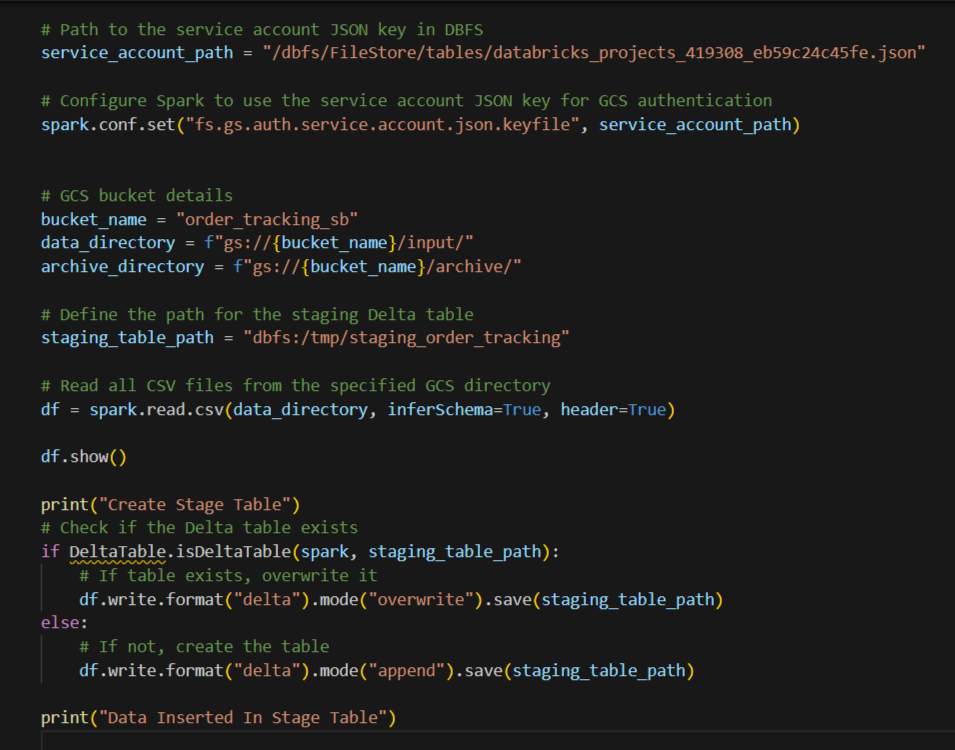
Order Tracking Incremental Load

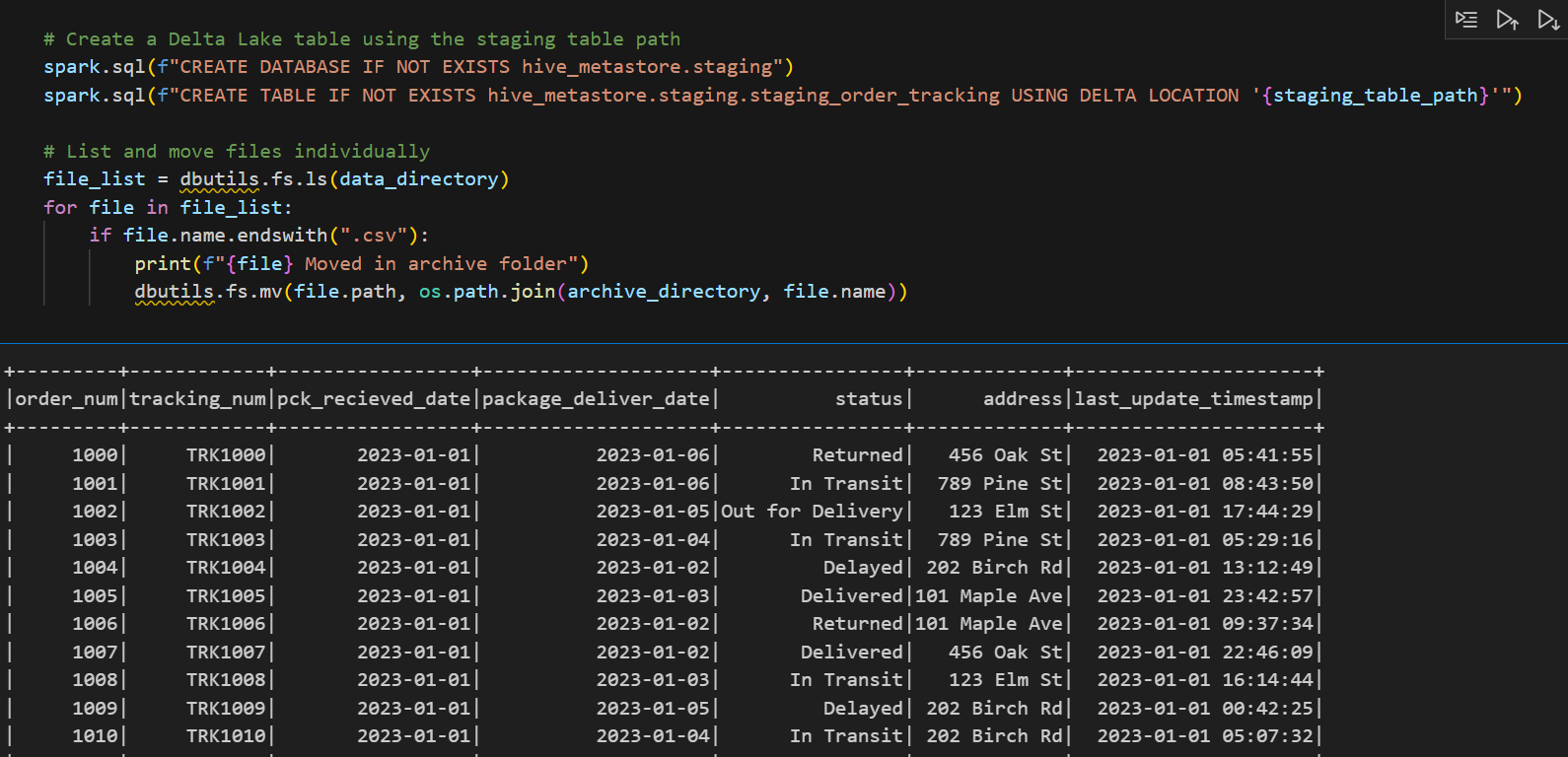
1. this project was about creating two different spark jobs in databricks. Once spark jobs were ready, I put them into a workflow job and scheduled that workflow. Basically on the execution of this scheduled job, we were moving data from staging table and doing upsert operation in target table.
2. We received daily csv file in gcs bucket input dir & from there we read data.



1. For 'stage\_logistic\_tracking\_job' spark script we had gcs bucket input storage and an archieve storage. Also we had a dbfs staging table path. Next step was to writing data in delta format at staging path if not exists or else just overwriting it.



1. This way data were inserted at staging path. Further, created delta table on the top of that delta location. At last, we were moving input source file to the gcs archieve folder once the staging process is done.



1. For 'target\_logistic\_tracking\_upsert\_job' spark script, we had paths for staging and target delta tables in dbfs. We were next reading data from staging delta file location. Then we were writing staging dataframe at target table path if that not existed and created delta table object from their target path. Further performed upsert from staging to target table using tracking\_num as key. At last, created delta table on the top of that target location.

