Data engineering assignment

Our fraud detection ML models expect transactions with amounts in USD as their input. However, transactions can have different currencies. Please implement a component that can convert the transaction amount to USD, for real-time as well as batch processing.

Consider this (simplified) data schema for transactions that we receive from our customers:

- Transaction_id: int
- Merchant_id: int
- Card_id: int
- Timestamp: int
- currency: str (ISO 4217 Alpha codes)
- Amount: float

You can find example transactions in the attached dataset.

Requirements:

- Only data matching the above-mentioned schema should be accepted
- Convert the *amount* to USD from the original as indicated by the *currency* field
- Use a monthly updating FX rate based on the *timestamp* of each transaction
- The conversion should work for real-time (<50 ms) as well as batch (> 100M rows) transactions.
- The real-time and batch conversion function should be identical, i.e. same inputs produce same outputs.
- Prefer to use Scala, Python, and Spark in your solution, if appropriate.

To get FX rates, you can use any service of your choice. We have a very limited free-tier subscription at currencylayer.com (max 250 requests / month) that you can use. As you only need monthly information, you should need no more than 50 requests. Our API key is fbc605d8dd59ec4995eec52c8bbefaa0

Scope of this assignment:

- Please don't spend more than 8 hours on this assignment.
- You are allowed to take "shortcuts" during development where appropriate.
- In particular, we do not expect you to implement an API. To demonstrate functionality, a command line interface or test cases would be sufficient.
- Please be prepared to describe any limitations of your solution during your presentation, as well as the design of a more mature alternative solution for an actual production system.
- Use your own judgment if you encounter any data quality issues.

Please submit the code of your solution to us in your preferred way. Following that, we can plan a date for a presentation with Q&A, If you have any questions regarding this assignment, please don't hesitate to contact us.