MG-GY 8411 Data Engineering HW1

Submitted by: Yogya Sharma (ys5250@nyu.edu)

Summary Report: Forex Real Time Data Collection and Analysis

1. Introduction

This document outlines the process of collecting Forex data from multiple sources in real time for 2 hours, storing it in both SQLite and MongoDB databases, exporting the data to CSV files, and conducting basic analysis including calculating average values.

2. Data Collection and Storage (Python file: fx_to_db.py)

  - Forex data was collected using the Polygon API using the get_real_time_currency_conversion method of the Polygon client.
  - Two databases were used for storage: SQLite and MongoDB.

- SQLite Database:
  - A SQLAlchemy ORM was used to define and interact with the database schema.
  - Forex data was stored in a table named `forex_data` in a database named "my_sql.db".
  - The data included columns for currency pair, timestamp, rate, and entry timestamp.

- MongoDB:
  - A MongoDB database was set up to store the same Forex data.
  - Data was stored in a collection named `fx` in a database named "DE".
  - Each document contained fields for currency pair, timestamp, rate, and entry timestamp.

3. Data Export to CSV (Python file: db_to_csv.py)

- Export Functionality:
  - Data from both SQLite and MongoDB was exported to CSV files.
  - CSV files were named `csv_sql.csv` and `csv_mongodb.csv`.
  - Each file contained columns for FX timestamp, pair, FX rate, and entry timestamp.

4. Average Rate Calculation CSV (Jupyter Notebook: calc_averages.ipynb)

- SQLite Average Calculation:
  - Pairwise averages were calculated for each currency pair in the SQLite database.
  - Additionally, the average rate across all data points was computed.

- MongoDB Average Calculation:
  - Pairwise averages were computed for each currency pair in the MongoDB collection.
  - Similarly, the average rate across all data points was calculated.

Averages:

```
SQLite Averages (pairwise): [('EUR,USD', 1.0847482929441774), ('GBP,CAD', 1.7200000000002393),
('USD,GBP', 0.7852517070560839)]
SQLite Average (all data points): 1.1966666666673498
MongoDB Averages (pairwise): {'EUR,USD': 1.08, 'USD,GBP': 0.79, 'GBP,CAD': 1.72}
MongoDB Average (all data points): 1.1966666666666668
```