

Design overview

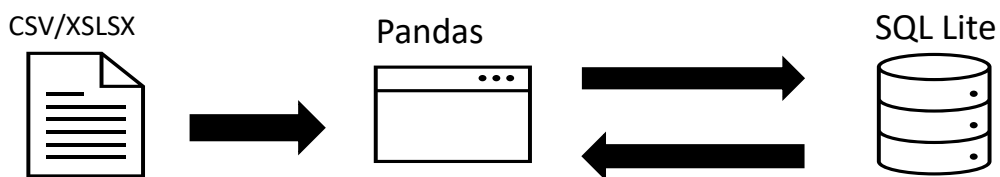
To perform the assigned tasks, I have used the following python libraries and will explain in detail on the design of the modules.

Libraries used: Pandas, Numpy, Flask, SQL Alchemy, Scipy, JSON, Argparse

Task 1:

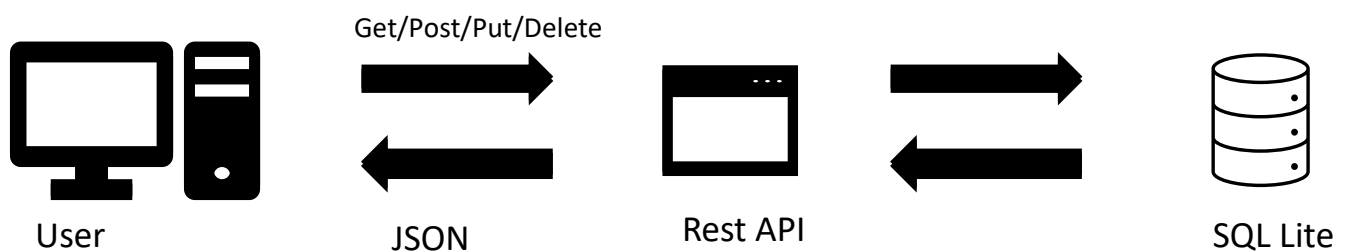
ETL Operation: I have separated this task into this module into two different operations in order to achieve the required output, the first operation is dedicated for the ETL process that's needed in order to server the data properly via the REST API, I am using Panda's library to slice and clean the data before using it for computations, the data from the CSV file is loaded onto the data frame and parsed for column names initially so that it can be fitted to the appropriate tables (Order, Fill, Route) in the SQL Lite database.

SQL Lite database was chosen as it is a lightweight, fully portable, Serverless and it comes inbuilt with Python which enables easy deployment and querying for the provided use case. And since this used case doesn't involve parallel access SQL lite was chosen as more appropriate DB. Also, this design enables the usage of the ETL process as a standalone module which can be called while processing similar files independent of the REST API.



Rest API: The Rest API was developed using Flask framework to interact with the SQL Lite DB using SQL Alchemy library as the SQL toolkit, Flask enables development and deployment of the API in a single python file. The framework is unopinionated and allows flexibility with respect design and components usage, the framework provides the ability to provide easy DB authentication.

SQL Alchemy ORM translates Python classes to tables for relational databases and automatically converts Pythonic SQL Alchemy Expression Language to SQL statements and henceforth SQL queries can easily be written within the Python syntax. SQLAlchemy also abstracts database connections and provides connection maintenance automatically. Together these features make SQLAlchemy a good choice in the context of this assignment.



Task 2:**Beta Calculation:**

I have used Scipy and Pandas to perform the beta calculation as part of this assignment, I have used pandas to capture the data in a data frame from the excel file and have manipulated it to perform the calculations. As suggested, I have used the winsorize function from mstats with 5% limit on both right and left tails.

Also, since the window is fixed to be 1 year and which is a long-time frame, I have chosen to run the calculation in logarithmic scale is better to show relative percentage changes and it can be extended to plot the data points to give a pictorial representation of the data at a later time.