Assignment 2 - Due by Monday 2/23

A major airline carrier hires you to decide the quantity of lunch boxes that they will prepare for each flight in the same leg of the New York to London flight. You are given historical counts of the number of actual passengers every flight in the file *hist_counts.csv*. Each lunchbox costs the airline \$8: a fresh part that cost \$5 and a snack (non-perishable) kit that cost \$3. Whenever the airline cannot serve a meal to a passenger due to running out of lunchboxes, they will issue a \$30 voucher per passenger. On the other hand, unused fresh parts will have to be disposed at an additional cost of \$2 due to composting and recycling fees. Moreover, every time a new order is placed there is a fixed cost of \$50 associated to transportation costs for the fresh part and \$10 for the snack part. The snack part can be reused on the next flights of the same leg (while the perishable cannot), but there is a holding cost of \$0.50 per kit part overnight.

Your model will be implemented as part of the decision making pipeline of the airline and will be run daily to decide the quantity of ordered lunchboxes for the day t, $\hat{q}_t^f$ and $\hat{q}_t^k$ representing the quantities ordered for fresh and kit part of the lunchbox for day t. The model has to be contained in a single script called *lunch_opt.py* with a function called *optimal_lunch_quantity*. The workflow is as follows:

- Every day t, you are given all the available historical passenger counts as a pandas series (indexed by date and with a single column named *passengers*) up to time $t-1$, as well as the final inventory of the kit part at time $t-1$, that is $I_{t-1}^k \geq 0$.
- The function *optimal_lunch_quantity* should return a pair of integers $\hat{q}_t^f, \hat{q}_t^k$ representing the quantities of the lunches to be ordered and prepared for the flight of day t.
- The total number of lunches at time t is represented as $\hat{q}_t = \min\{\hat{q}_t^f, \hat{q}_t^k + I_{t-1}^k\}$. Moreover, there is a physical limit on the number of kits that can be stored at 800, that is $\hat{q}_t^k + I_{t-1}^k \leq 800$.

To ensure that the script works as intended you are given a unit test script *unit_tests.py* ensuring that the output of your model runs within the infrastructure of the company. For the same reason, you are also constrained that the script can only use packages contained in the file *requirements.txt*.
The performance of your strategy is evaluated over out-of-sample future data continuing the time series in *hist_counts.csv*. Let q_t be the actual (future and unknown) count of passengers. The evaluation metric is defined as the out-of-sample metric:

$$\frac{1}{T}\sum_{t=1}^{T} 2(\hat{q}_t^f - \min(q_t, \hat{q}_t))^+ + 30(q_t - \hat{q}_t)^+ + 5\hat{q}_t^f + 3\hat{q}_t^k + \mathbf{1}\{\hat{q}_t^f > 0\}50 + \mathbf{1}\{\hat{q}_t^k > 0\}10 + 0.50 I_t$$

$$\text{s.t. } \hat{q}_t^k + I_{t-1}^k \leq 800 \text{ and } I_t^k = (I_{t-1}^k + \hat{q}_t^k - \min(q_t, \hat{q}_t))^+ \text{ with } I_0^k = 0.$$

Task: Come up with a model that performs best according to this metric.