Recently, the presence of businesses in the digital world has exploded, and the health crisis seems to have even accelerated this trend. It is in this situation of ultra competitiveness that our problem was born:

What dynamic pricing strategy should be adopted in a market where there are strategic customers with a product with a fixed maturity (finite or infinite)?

In order to solve this problem we chose to create a model with, on one hand a salesperson who sets prices intelligently and, on the other hand customers with different behaviours.
Our solution initially proposes to define two types of customers: naive and strategic. This consideration of customer behaviour is a determining element, as this point is not used by existing solutions. A naive customer has a predictable behaviour and no knowledge of the market.
Their willingness to buy is sometimes ruled by chance, but we have also defined a proportion of naive customers who have to make a purchase, which therefore determines a probability for each customer to buy.
A strategic customer, on the other hand, has a reflection on the market and the evolution of prices.
First of all, they have access to the historical evolution of the price of the product, which allows them to base their reflections on a given fact.

Next, the range of prices acceptable to the strategic customer must be defined. This range is determined by the average of the 20% highest and lowest prices. Then we define the *Willingness-To-Pay* which represents the customer's willingness to purchase when the price is within an acceptable price range. In concrete terms, we had to define a strong buying signal, based on the price evolution. This signal is a decrease compared to the last two prices. If the price of month M is lower than the prices of months M-1 and M-2, then the purchase is made.
We then model the seller using a Deep neural Network approximating the Q-function (DQN), which has the advantage over simple Q-learning of transmitting a large amount of data to the seller.
The data sent to the network are all the possible states of the environment. We have considered the following settings: date, price and number of bookings, but we could consider more to help the network learn more efficiently.
On this database, the seller learns the consequences of his actions, i.e. setting a price for a given state, thanks to a reward received, positive or negative. The seller's goal during learning is then to maximise this reward in the long term.
During the learning process, the salesperson first explores the environment and then uses this information. The actions taken and the consequences are recorded in the salesperson's memory as an experience, so all past experiences can be exploited to maximise the reward.

Once the seller has been modelled, we put them face to face with the market, and therefore with the different proportions of strategic and naive customers. We test all possible proportions of these two types of customers and find that our DQN simulated vendor manages to outperform all of them. To achieve this result we trained the DQN on every 20%

portion of strategic customers. DQN is most effective on the portion where there are 80-100% strategic customers, which is a very encouraging result for its real usefulness.

It seems to us that analysing customer behaviour is an important aspect to consider when dealing with pricing issues. Indeed, it would provide better accuracy allowing companies to obtain better results. Concerning the prospects of the project, we think it would be interesting to include more elements and also to optimise the learning parameters, in order to achieve greater accuracy.