We have processes that generated some time-series data. We know that one of these processes is easy to formulate mathematically (and the data is quite clean) while the other two are noisy. Also while all the problems are time series (non-iid data), for the different problems the amount of information we get from the recent past behavior can be quite different. For all of these we need to be able to explain to our stakeholders:

First, I would like to point you to my notebook, with full explanation about my thought:

<https://weslleymoura.github.io/olx/time_series_analysis.html>

And also, to my GitHub repo

<https://github.com/weslleymoura/olx>

1. Are there any periodic behaviors?

All series presents a sort of random season component, what makes me think there is no relevant information about seasonality.

1. Is there any seasonal or global trend in the data?

Dataset 1: No

Dataset 2: There is a global decreasing movement. In the most recent part of the series, there was an increase trend, and, in the very end, it has started a plateau movement. According the history of the series, I would think there will be a decrease movement moving forward (that such assumption really needs some business knowledge, which I don’t have).

Dataset 3: I did not see any relevant movement of global trend. When we zoom some parts of the series, we might find local trends, but they are not supported in long term, neither captured by the seasonal component.

1. What is the error in predicting, if we can use the first ¾ of the data for training and we need to predict the remaining ¼ ?

I am reporting RMSE:

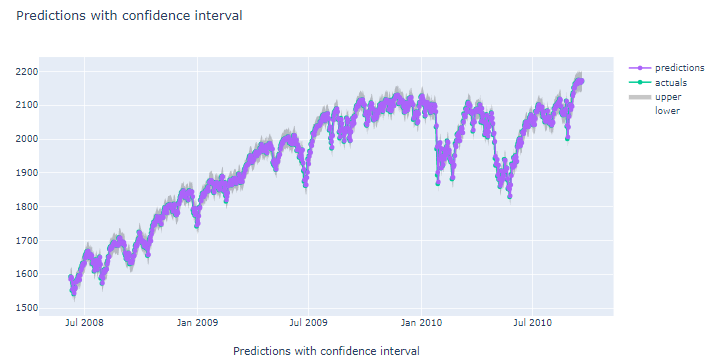
Dataset 1: Because the behavior is so regular, we can so exact predictions with no issue. Of course, any further change on the distribution of the data will require a new model.

Dataset 2: 16.427

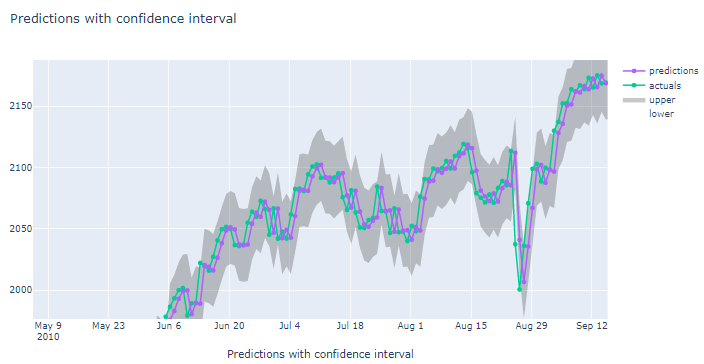
Dataset 3: ~0.98

1. Can we provide confidence intervals for our predictions (and respective plots)?

Yes, I have made predictions and built the 95% confidence interval on the test data of the second dataset as follow:



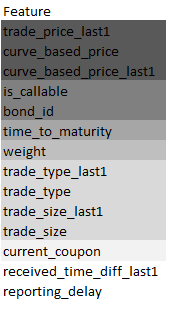
And this is a zoon view:



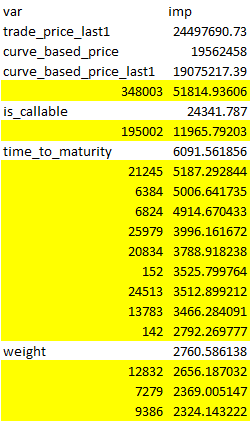
Please, find more comments on my notebook <https://weslleymoura.github.io/olx/time_series_analysis.html>

1. For the third dataset, which are the most important features? Are there any bonds or bond categories we should prefer to invest on?

I have done some manual work to highlight feature importance based on these results. All of these features play a certain level of importance, I built this heat map to show the most important ones (from top to bottom)



I have done made bond\_id one hot encoded, and some codes looks for relevant to the problem



Please, check my complete rational on my notebook  
<https://weslleymoura.github.io/olx/time_series_analysis.html>

1. Any other insights that you consider significant to communicate?

This assignment is quite long, so I had many things to communicate all over the task. I would like to share my notebook and ask your attention to walk through it <https://weslleymoura.github.io/olx/time_series_analysis.html>

Finally, make the model you have created accessible through an API: anything like tensorflow serving, Flask, lambda functions, etc. is ok as long as we can use the API to make predictions live (speed is not a big consideration for this task).

Notes: for the third dataset the trade price is the target variable and the curve\_based\_price is a naive baseline estimate, which we can include as a feature to our model.

Note: some of the questions are intentionally open ended and many correct answers are possible.

Please, find my API code and some documentation here:

<https://github.com/weslleymoura/olx/>

Let me know if you want that I make this repo private.

It is making online predictions here: <http://201.39.71.68:5000/prediction/>. I decided to create a very simple HTML response, so you can test it in your browser. However, you can make post requests, if you want. Please, find a postman collection on my GitHub to understand how to pass the right params.