# Description of the challenge

BNP Paribas is constantly trying to improve the service it offers to its customers and recent advances in Machine Learning offer a great opportunity to achieve this goal.

In particular, the offer in Corporate Bonds market is quite large with thousands of products that the customers may be interested to buy or sell at any point in time.

The goal of this challenge is to try to predict which Bond a given Customer is going to be most interested in buying or selling, so that our Sales team can issue the most relevant trade recommendation or send the most interesting research paper to the customer.

More precisely, we are asking the teams to estimate the probability that a given Customer is going to express an interest in buying or selling a given product in the week following on the 23rd of April. We will measure the performance of the model using the [Area Under the Receiver Operating Characteristic Curve](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html) metric.

The complexity of this challenge comes from the fact that we have very different customers, with different investment strategies as well as very different type of Bonds, some being very active, trading almost every day while some others will see a trade only once a month! Overall we are trying to capture fairly infrequent events, the historical probability of expressing an interest being around 1.5% for our perimeter.

The available information is split in 5 different databases:

* The Trade Database: this database shows the historical interaction between our Customers and BNP Paribas
* The Isin Database: the database shows various static data on each of the instruments
* The Customer Database: this database gives a few static information on the Customers
* The Market Database: for each Date and each Bond, we give 3 different way of looking at the price of a Bond: its Price, its Yield and the difference between this Yield and a standard risk-free curve (ZSpread)
* The Macro Market Database: in this database we give the time series of Price of various high level instruments like Equity Indices, Spot Exchange Rates between currencies and level of risk-free Yield Curve.