* What is the problem you want to solve?

The problem lies in getting the predicting the credibility of a person of getting a loan along with deciding the time-line when the loan would get paid off.

* Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn’t have done otherwise?

My client here is my company, who will be able to get the best lead who can be converted and given the best possible loan. Based on the analysis the company will be able to disburse the loan in the right manner with capability to decide the correct tenure and total sanction amount.

* What data are you using? How will you acquire the data?

The data is easily available from the company’s database. Will be referring to a collated data of the different attributes of a lead and then clean the data before processing. Since this is confidential data so changing the primary attributes (i.e.: email and phone number) with dummy data will act the input data set.

* Briefly outline how you’ll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

As data wrangling becomes the first step, so i will start collecting the data and as mentioned the data is already available so probably referring to a single table in the database can fetch me the input data.

Now, the data captured, is from multiple sources so the entire data needs to be cleaned before use. We need to transform it with standard acceptable values and keep distinct and non duplicate data.

Once cleaning is completed will apply different algorithms to come up with a trend and use python libraries to present a visualization of the data.

* What are your deliverable? Typically, this includes code, a paper, or a slide deck.

My deliverable would consist of:

1. A working model that can take the input data set and generate patterns based on the historical data.
2. Report generated after applying different statistical Methods for Data Analysis
3. Also a machine learning trained algorithm which will be able to train the model based on existing/new data sets.