**Project: Home Credit Default Risk – predict how capable each applicant is of repaying a loan**

*By Jasmine Zeng*

**Question/Need:**

* **What is the question behind your analysis or model and what practical impact will your work have?**

Assessing clients’ ability to repay the loan is important for any loan issuer – reliable customers are more likely to repay the loan by deadline. On the other side, if you give higher rate to unreliable customer, then there is a large chance that they couldn’t make the payment on time. Being able to predict/guess if a client can repay the loan will help bank to make decision whether to lend money to a client.

* **Who is your client and how will that client benefits from exploring this question or building this model/system?**

Home Credit bank is my client. They provide client level dataset with various features based on which I could built a binary classification model to predict whether a client can repay the loan.

**Data Description:**

* **What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you’re downloading, the site you’re scraping, etc.)**

I get it from Kaggle: https://www.kaggle.com/c/home-credit-default-risk/data

* **What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?**

One row represents one client.

* **What characteristics/features do you expect to work with? In other words, what are your columns of interest?**

Income and credit score are my columns of interest

* **If modeling, what will you predict as your target?**

Default or not. If default: 1; else: 0

**Tools:**

* **How do you intend to meet the tools requirement of the project?**

I plan to use the python packages for classification models (sklearn, xgboost, etc.) that were covered in class.

* **Are you planning in advance to need or use additional tools beyond those required?**

No plans as of now.

**MVP Goal:**

* **What would a**[**minimum viable product (MVP)**](https://app.thisismetis.com/courses/138/assignments/1031)**look like for this project?**

An ideal MVP would contain some key visuals of initial findings. In a more ideal MVP, a baseline model should be available. A confusion matrix will be included.