**Project Sketch**

Data Mining Project Sketch

Section A, Team - 13

Members: Gamini Singh, Helena Gao, Max Liang, Cindy Yu, Maxwell Zhang

**What is the exact business problem?**

There are lots of default actions happening in the use of credit card. In order to reduce the risk of not collecting payments and increasing the profits for the Lending Club, it should set specific criteria on approving loan applications based on analysis of past loan records.

**What is the use scenario?**

By analyzing the dataset, we can:

1. Find the characteristics of loan applicants who are more likely to default and be more cautious when approving loan applications from such loan applicants. Additionally, the purpose that each loan is applied for can also be related to how likely the loan applicant is going to default.
2. More comprehensive analysis based on current loan status and payment history of a borrower can be used revamp the existing credit rating system followed, and this will help reduce futures loans made to borrowers with lower credit ratings and/or degrading credit rating over a period of time.
3. Also, different potentially risky categories for purpose of loan can be identified. This would help the collectors be more vigilant in terms of collection and modifying economics of such loans.

**Which is the data source?**

The data source is the Lending Club Loan Data, the url: https://www.kaggle.com/wendykan/lending-club-loan-data

**What is a data instance/unit?**

One row per loan listing.

**What might be the target variable?**

The target variable is the loan\_status. We want to figure out some relationships and help the company to reduce offering loans that are more likely to default.

**What features would be useful?**

Id, annual\_inc, home\_ownership, purpose, loan\_amnt.

**What precisely is the data mining problem? Is it supervised or unsupervised?**

**Supervised:**

1. Predictive Analysis: Given the characteristics of the loan appliers, whether Lending Club should lend or not based on the probability of the default rate.

2. Classification: What loan request of the borrower has the higher rate of default?

3. Causal Modeling: How likely the member is going to default on the loan given that he/she defaulted on another loan?

**How exactly would it add business value? Etc.**

The Lending Club can reduce the default risk by providing loan to more qualified loan applicators and thus increase the company profits.