

*Sunday, December 5th, 2021*

# Predictive Model Challenge Proposal

Prepared by: \_\_\_\_\_

**Facts & Stats**

(Marie Reine Axalan,  
Marcos Fontes,  
Jordyn Gerstle-Goodman)

# Dataset used in our model

---

The data our team has selected to develop a predictive model for gives us a quantitative baseline for credit lenders. With the dataset, we are able to measure and transform lender perspective variables into features that will allow us to predict priceless insights when considering risk and reward among applicants.

## Business Question

---

The main business question that we are looking to answer is **if a customer will default on a loan or not given their financial and personal characteristics**. In order to identify which customers are more likely not to default, we need to be able to predict the likelihood of them doing so; those that have a lower likelihood of default are more likely to be approved as we know they will pay back.

## Variables that influence our model

---

Our outcome variable would be whether or not the customer defaulted on their loan indicated by the "LOAN\_DEFAULT" column in the data set. This is a binary predictor variable that classifies a 1 as "Yes the customer did default" and 0 as "No the customer did not default." Our predictor variables would mainly include demographic and behavioral information regarding the customer. We are mindful to remove features such as PIN, Drivers license, and UNIQUE ID which would not be good indicators of whether a customer will default.

We will do models with and without the bureau score to analyze the effect this has on the likelihood of default.

Another model would predict a customer's likelihood to be approved (approval odds) to help us better understand the customer. This could be a point of interest given that if you calculate the odds that someone would be approved would become a feature that could also be a strong predictor for a potential loan default. If the disbursed\_amount is 0, it would be an indicator that they were not approved, but any amount over 0 would indicate they were approved; we would feature engineer a binary outcome. The dataset has the potential to answer multiple questions, so it would be best utilized in this capacity.

## Predictive Model Hypothesis

Payment history, Credit care use, and derogatory marks are the highest impact features on credit score according to Credit Karma.

Age X sec.no.of.accts - We would look at this interaction because we assume that younger applicants would be more likely to have a co-applicant. When an applicant is younger, their co-applicant is more likely to be their parent, who has a long history of credit, giving this applicant a smaller chance of defaults.

- If a customer is younger with a secondary account attached to their loan, they are less likely to default on their loan.

Credit History Length X "PRI.NO.OF.ACCTS" - An applicant that has a long credit history and multiple loans taken could be less likely to default than an applicant with a very short credit history and a high number of loans taken. We define a long credit history as 10 years or more.

- If a customer has a short credit history length and has multiple open loans, they are more likely to default on their loan.