# Data Science Interview Task
# "Prequalification"

September 13, 2021

## Introduction

A type of loan called "merchant cash advance" is repaid in the following way: the lender takes a sweep, e.g., 20%, of each card transaction received by the merchant until the loan is repaid. Therefore, card transaction data, AKA payment data, is useful when determining who to give loans to.

The present task is to identify merchants who based on their payment data seem to qualify for a loan.

## Input

There are two input data files:

- "payments.csv" containing daily payments received by a set of merchants
- "prequalifications.csv" indicating whether merchants were prequalified (1) or not (0)

## Task

Using a programming language of your choice, complete the following steps:

1. Change the layout of the payment data so that each row corresponds to a merchant, and each column corresponds to a month. Name the columns in the format "YYYY-MM" (e.g, "2021-08" for August 2021)
2. Add these metric columns: "n_months_data" (number of months data each merchant has) and "mean_monthly_payment" (mean monthly payment for each merchant) to the data
3. Train a machine learning algorithm/model of your choice to predict whether each merchant prequalifies, using "n_months_data" and "mean_monthly_payment" as input features and the data in "prequalifications.csv" as target variable

4. Evaluate performance of the model, taking measures to prevent overfitting
5. The prequalification is based on rules. Identify the rules (and show how)

## Output

Please provide the following after completing the task:

- "transformed_payment_data.csv" containing the transformed data set, including the two added metric columns
- The code used to solve the task