Loan granting

Goal

Another area where data science and machine learning play a huge role is in choosing if granting a loan. This is a particularly hot field as many start-ups feel that bank loan models can be improved. Therefore, there is space to come up with better loan strategies that can benefit both the lender and the borrower.

In this challenge, you will have access to loan data from a bank and will have to improve their model.

Challenge Description

We have access to a specific bank loan data. We have data about all loans asked to the bank, whether the bank decided to grant it and, finally, whether the borrower managed to repay it. We also have info about the borrower at the moment she is asking for the loan.

You have to come up with a better strategy to grant loans. Specifically you should:

- Build a model which is better than the bank model. For simplicity, assume that:
  - If you grant the loan and the it doesn't get repaid, you lose 1
  - If you grant the loan and the it does get repaid, you gain 1
  - If you don't grant the loan, you gain 0
- Using the rules above, compare bank profitability vs your model profitability
- Describe the impact of the most important variables on the prediction. Also, focus on the variable "is_employed", which describes whether the borrower is employed when she asks for the loan. How does this variable impact the model? Explain why
- Are there any other variables, not in the data provided, that you'd have liked to include in the model?

Data

We have 2 table downloadable by clicking **here**.

The 2 tables are:

"loan_table" - general information about the loan

Columns:

- **loan_id** : the id of the loan. Unique by loan. Can be joined to loan id in the other table
- **loan_purpose** : the reason for asking the loan: *investment, business, emergency_funds, home, other,*
- **date** : when the loan was asked
- **loan_granted** : whether the loan was granted
- **loan_repaid** : whether the loan was repaid. NA means that the loan was not granted

---

"borrower_table" - information about the borrower

Columns:

- **loan_id** : the id of the the loan. Unique by loan. Can be joined to loan id in the other table
- **is_first_loan** : did she ask for any other loans in her lifetime?
- **fully_repaid_previous_loans** : did she pay on time all of her previous loans? If this is the first loan, it is NA
- **currently_repaying_other_loans** : is she currently repaying any other loans? If this is the first loan, it is NA
- **total_credit_card_limit** : total credit card monthly limit
- **avg_percentage_credit_card_limit_used_last_year** : on an average, how much did she use of her credit card limit in the previous 12 months. This number can be >1 since it is possible to go above the credit card limit
- **saving_amount** : total saving account balance when she asked for the loan
- **checking_amount** : total checking account balance when she asked for the loan
- **is_employed** : whether she is employed (1) or not (0)
- **yearly_salary** : how much she earned in the previous year
- **age** : her age
- **dependent_number** : number of people she claims as dependent

Example

```
    Let's check one loan:
```

head (loan_table,1)

Column Name	Value	Description
loan_id	19454	this is id of the loan
loan_purpose	investment	she needed the loan in order to make an investment
date	2012-03-15	the loan was asked on March, 15.
loan_granted	0	unfortunately, the bank didn't grant this loan
loan_repaid	NA	therefore, loan_repaid is NA since there was no loan granted in the first place.

```
    Let's now check the characteristics of that user who was denied the loan
```

subset (borrower_table, loan_id == 19454)

Column Name	Value	Description
loan_id	19454	same id as in the example above
is_first_loan	1	this is the first time this person asks for a loan
fully_repaid_previous_loans	NA	has to be NA since she never asked for a loan in the past
currently_repaying_other_loans	NA	has to be NA since she never asked for a loan in the past
total_credit_card_limit	8600	her monthly credit card limit is 8600 USD
avg_percentage_credit_card_limit_used_last_year	0.79	on an average, per month, she used 79% of her credit card limit in the previous yr.
saving_amount	1491	she has 1491 USD in her saving bank account
checking_amount	6285	she has 6285 USD in her checking account
is_employed	1	she has a job when she asked for the loan
yearly_salary	45200	and she was making 45200 USD a yr
age	42	she is 42 y/o
dependent_number	7	she claims 7 dependents