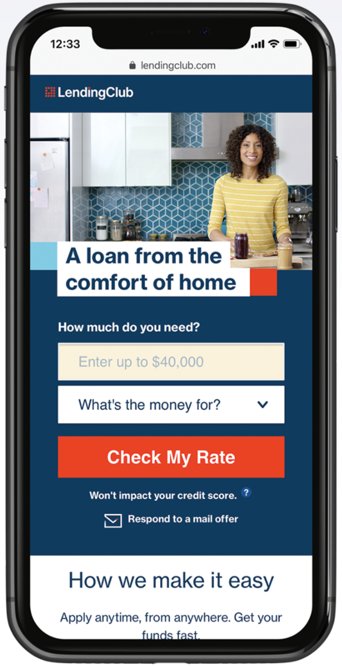
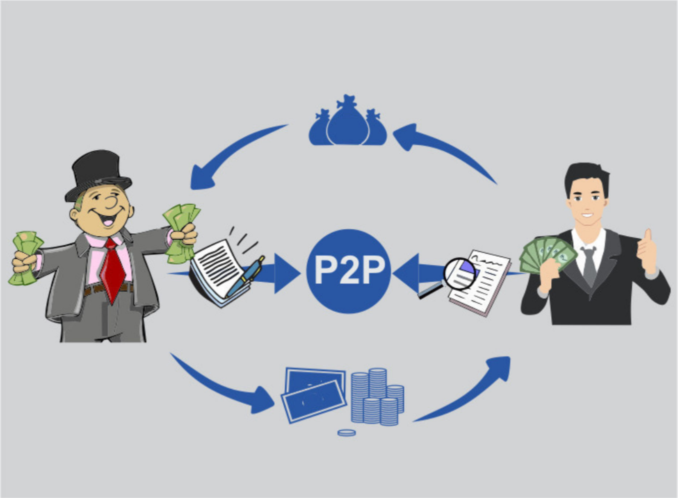


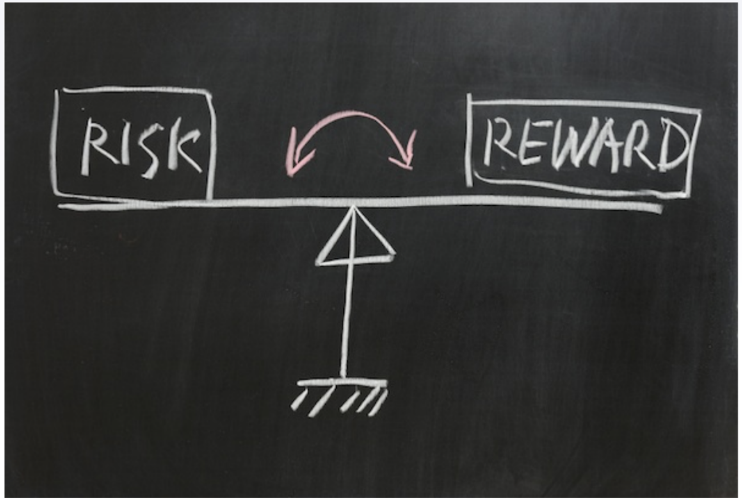
Portfolio Optimization for Investors in Lending Club Company

Team13

Jiatao Li 3035675153	Data Processing, Prediction Model Building, Optimization Model Building, Report Drafting
Jiantong Yang 3035675701	PowerPoint Drafting, Report Drafting, Data Collection, Model Building, Presentation
Linzuo Zou 3035675373	Optimization Model Building, Report Drafting, PowerPoint Drafting, Presentation
Yujing Huang 3035657814	PowerPoint Drafting, Model Building

P2P Lending Platform





- ➔ **match** lenders with borrowers through online system
- ➔ companies **low operating costs** give investors & borrowers **high return**

- ➔ **return**: 4% to 7% of P2P- V.S. 3% of one-year-T-Bills
- ➔ **default risk** at Lending Club company: 6% to 7%

1. "Peer to Peer Lending Vs Stocks." *Global P2P Lending*, www.globalp2plending.com/en/blog/entry/p2p-lending-vs-stocks-and-bonds/.
2. Renton, Peter, et al. "What Are the Risks of Peer to Peer Lending?" Lend Academy, 10 May 2011, www.lendacademy.com/what-are-the-risks-of-peer-to-peer-lending/.

Project Goal

Objective:

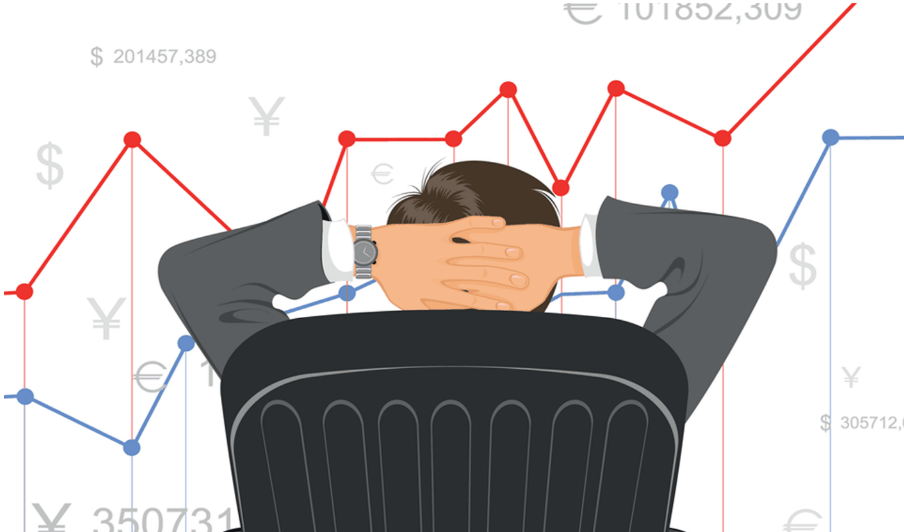
- ❑ Optimize the investment portfolio for investors in Lending Club Company
- ❑ Achieve the greatest return given a specific default risk

Models:

- ❑ LGB Model
- ❑ Optimization Model

Content

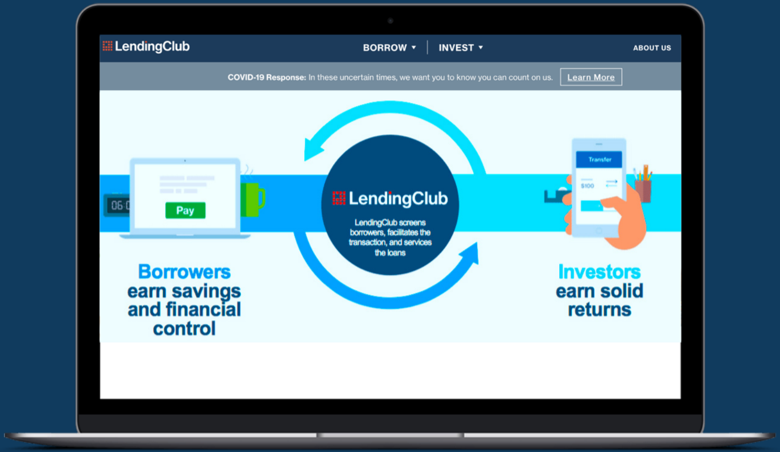
- Project Goal
- Lending Club Company
- Data Interpretation
- Prediction Model
- Optimization Model
- Conclusion



1. Lending Club Company

- ✔ Company background
- ✔ Loaning Process Mechanics

1.1 Lending Club is ...



America's largest online marketplace connecting borrowers and investors

Borrowers access low interest rate loans through a fast and easy online or mobile interface.

Investors provide the capital to enable the loans in exchange for earning interes

Lending Club operate fully **online** with no branch infrastructure and use technology to lower cost and deliver an amazing experience.

Lending Club **pass the cost savings to borrowers in the form of lower rates and investors in the form of attractive returns.**

1.2 Loaning Mechanism in Lending Club

1.2 Loaning Mechanism in Lending Club

Borrowers submit loan request

Lending Club evaluate the request based on underwriting criteria

Borrowers are assigned interest rate

The loan request is listed on the website

Investors browse and invest

Investors can access to all supplement information that borrowers provide such as borrowers' financial background in order for investors to make investment decisions

1.2 How do Investors Invest in Lending Club?

Investors browse active loan listings with borrowers' information on the website

⇩

Borrowers' loans are broken up and split into many **Notes**

⇩

Investors purchase Notes corresponding to fractions of loans to form portfolio

2. Data Interpretation

- ✓ Data Source
- ✓ Variable Interpretation
- ✓ Data Preprocessing

2.1 Data Source

id	member_	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title
68407277		3600	3600	3600	36 months	13.99	123.03	C	C4	leadman
68355089		24700	24700	24700	36 months	11.99	820.28	C	C1	Engineer
68341763		20000	20000	20000	60 months	10.78	432.66	B	B4	truck driver
66310712		35000	35000	35000	60 months	14.85	829.9	C	C5	Information Systems Officer
68476807		10400	10400	10400	60 months	22.45	289.91	F	F1	Contract Specialist
68426831		11950	11950	11950	36 months	13.44	405.18	C	C3	Veterinary Tecnician
68476668		20000	20000	20000	36 months	9.17	637.58	B	B2	Vice President of Recruiting O
67275481		20000	20000	20000	36 months	8.49	631.26	B	B1	road driver
68466926		10000	10000	10000	36 months	6.49	306.45	A	A2	SERVICE MANAGER
68616873		8000	8000	8000	36 months	11.48	263.74	B	B5	Vendor liaison
68356421		22400	22400	22400	60 months	12.88	508.3	C	C2	Executive Director
68426545		16000	16000	16000	60 months	12.88	363.07	C	C2	Senior Structural Designer
68338832		1400	1400	1400	36 months	12.88	47.1	C	C2	Logistics Manager
66624733		18000	18000	18000	60 months	19.48	471.7	E	E2	Software Manager
68466961		28000	28000	28000	36 months	6.49	858.05	A	A2	Senior Manager
68354783		9600	9600	9600	36 months	7.49	298.58	A	A4	tech
68466916		25000	25000	25000	36 months	7.49	777.55	A	A4	Sales Manager

Extract of the Dataset

Source: “Lending Club Dataset Provided by Nathan Geroge”, Kaggle

Dataset: 0.5 million pieces of raw data, 150 features in total

Selected Features: 9 features out of 150 given by the dataset

➔ relavent to borrowers’ default probability

1. **loan_status**: Variables with seven levels
2. **loan_amnt**: Total amount of loan taken
3. **int_rate**: Loan interest rate
4. **grade**: Grade of employment
5. **emp_length**: Duration of employment
6. **home_ownership**: Type of house ownership
7. **annual _inc**: Total annual income
8. **term**: 36-month or 60 month period
9. **issue_d**: the day that loan issued

2.2 Variable Interpretation

Target Variable: whether the borrower would default by identifying whether the loan is a good loan

- ➔ There are seven status in the loan status, we group the status into good loans and bad loans based on the description of each status provided by Lending Club
  - ◆ good loans: “Current”, “Fully paid”, “In Grace Period”
  - ◆ bad loans: “Late(16-30)”,”Late(31-120)”,”Default”, “Charged off”
- ➔ binary variable
  - ◆ 0: normal repayment
  - ◆ 1: default

- Current: Loan is up to date on all outstanding payments.
- In Grace Period: Loan is past due but within the 15-day grace period.
- Late (16-30): Loan has not been current for 16 to 30 days.
- Late (31-120): Loan has not been current for 31 to 120 days.
- Fully paid: Loan has been fully repaid, either at the expiration of the 3- or 5-year year term or as a result of a prepayment.
- Default: Loan has not been current for an extended period of time.
- Charged Off: Loan for which there is no longer a reasonable expectation of further payments.

1. “LendingClub: Peer-to-Peer Lending & Alternative Investing.” Peer to Peer Lending & Alternative Investing, www.lendingclub.com/.

2.2 Variable Interpretation

Selected Variables for the Prediction Model (7 variables)

➔ the amount of the loan, loans’ interest rate, the grade and the duration of the employment, the type of the house ownership, borrowers’ annual income and the loan term

2.3 Data Preprocessing

Fill up the missing values by the function of mean

```
# fill missing values
Missing = ['loan_amnt','int_rate','annual_inc']
raw_X[Missing] = raw_X[Missing].fillna(raw_X[Missing].mean())
```

Fill up the missing values by the function of mean

```
# set 'home_ownership'
# Only addr_state, grade and sec_app_earliest_cr_line are remained
for cat_feat in ['grade', 'term', 'emp_length', 'home_ownership']:
    dummies_df = pd.get_dummies(raw_X[cat_feat]).rename(columns=lambda x: cat_feat + str(x))
    X = pd.concat([raw_X, dummies_df], axis=1)
    X.drop(cat_feat,axis=1, inplace=True)
```

Split the dataset into training data and testing data

```
X_trainval = X[(data.issue_d < '2018-01-01 00:00:00')]
X_test = X[(data.issue_d >= '2018-01-01 00:00:00')]
y_trainval = y[(data.issue_d < '2018-01-01 00:00:00')]
y_test = y[(data.issue_d >= '2018-01-01 00:00:00')]
```

3. Prediction Model

LGB Model

3.1 LGB Model

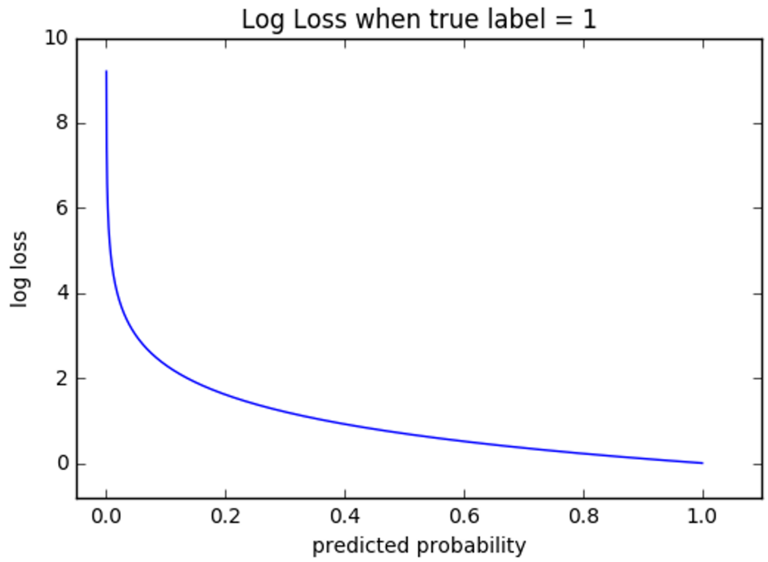
Definition: gradient boosting framework that uses tree-based learning algorithm

Goal: predict the default risk of each borrower's loan

Advantage:

- high training speed
- high efficiency with low memory consumption
- large scale data

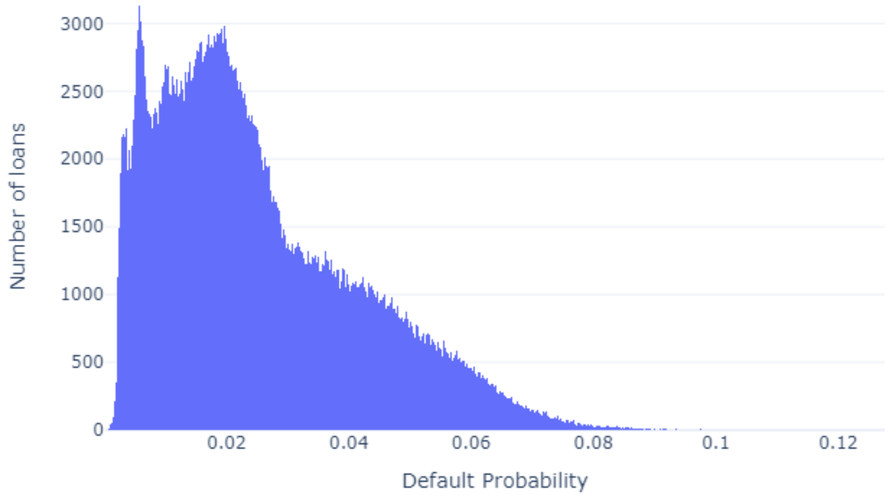
3.1 LGB Model



Log Loss Penalty Function

- measure the performance of the prediction model
- improve the accuracy of the model by punishing the prediction error

3.1 LGB Model



- The graph displays the distribution of the predicted default probability
- Most of the default probability are around 0.02

4. Optimization Model

Optimization

Model Performance - Example

4.1 Loan-Selection Optimization

Variables

Decision Variable: xi: whether the investor to choose to invest in loan **i**

ai: the amount of money that the loan **i** ask for

ri: annual lending rate

B: the amount of investor's principal

R: risk tolerance (rate)

n: the number of loans in loan pool that are waiting to be chosen

Ti: length of months

Di: default risk (probability)

4.1 Loan-Selection Optimization

Objective Function: maximize the total return of the portfolio

$$\sum_{i=1}^{n} x_i * a_i * (1 + r_i * T_i/12)$$

Constraint 1: the invested amount is less/equal to investor's amount of the principal

$$\sum_{i=1}^{n} a_i x_i \leq \text{B}$$

Constraint 2: the risk of the portfolio is less/equal to investor's risk tolerance

$$\sum_{i=1}^{n} \frac{a_i D_i x_i}{B} \leq \text{R}$$

Constraint 3: the requirement of money liquidity – 40% for 36-month loan and the other for 60-month

$$\sum_{i=1}^{n} a_i T_i x_i \leq (36 * 0.4 + 60 * 0.6)$$

Constraint 4: the upper limit of the number of selected loans is less/equal to 50

$$\sum_{i=1}^{n} x_i \leq 50$$

4.1 Loan-Selection Optimization

Create the model and define variables and objective function

```
import pyomo.environ as pe

# create a model
model = pe.ConcreteModel()

# define decision variables
n = 50
# B is the total amount
B = 500000
# R is the risk
R = 0.02
# whether to lend
model.x = pe.Var(range(n), domain = pe.Binary)

# declare objective
model.re = pe.Objective(expr = sum(model.x[i]*k[i] for i in range(n)), sense = pe.maximize)
```

4.1 Loan-Selection Optimization

Build the constraints and solve the model

```
# constraint: less than total amount
model.amt = pe.Constraint(expr = sum(sample.iloc[i,0]*model.x[i] for i in range(n)) <= B)

# constraint: risk limit
model.risk = pe.Constraint(expr = sum(sample.iloc[i,0]*sample.iloc[i,8]*model.x[i] for i in range(n))/B <= R)

# constraint: money liquidity
model.liquidity = pe.Constraint(
    expr = sum(int(sample.iloc[i,1]) * sample.iloc[i,0] * model.x[i] for i in range(n)) <= B*(36*0.4+60*0.6)
)

# constraint: upper limit of number of selected loans
model.total = pe.Constraint(
    expr = sum(model.x[i] for i in range(n)) <= 50
)

# solve the model
solver = pe.SolverFactory('gurobi')
results = solver.solve(model)
```

4.1 Loan-Selection Optimization

Results

```
print('the return rate is {}, and the ture risk is {}'.format(model.re()/B, model.risk()))
```

```
return rate is 1.5398509299999996, and the ture risk is 0.019674443108549496
```

```
loan=[]
for i in range(50):
    if model.x[i]() > 0:
        loan.append(i)

print(loan)
print(len(loan))
```

```
[1, 2, 3, 5, 6, 8, 9, 11, 12, 13, 14, 16, 17, 18, 20, 22, 23, 24, 25, 26, 27, 29, 31, 32, 35, 36, 37, 38, 39, 4
0, 41, 42, 44, 47, 49]
35
```

4.2 Model Performance - Example

- **Example:** An investor has 0.5 million dollars in principal, with a risk tolerance of is 0.02. Help the investor build the portfolio that generate greatest return given the default risk

- **Variables**
  - **B** (investor's principal)**:** 500,000
  - **R** (risk-tolerance): 0.02
  - **n**: 100
  - **n:** randomly select 100 borrowers who applied for loans on Lending Club in March 2018
- **Loans**

4.2 Model Performance - Example

Rate of Return: percentage of the investment's initial cost

model.objective() / B

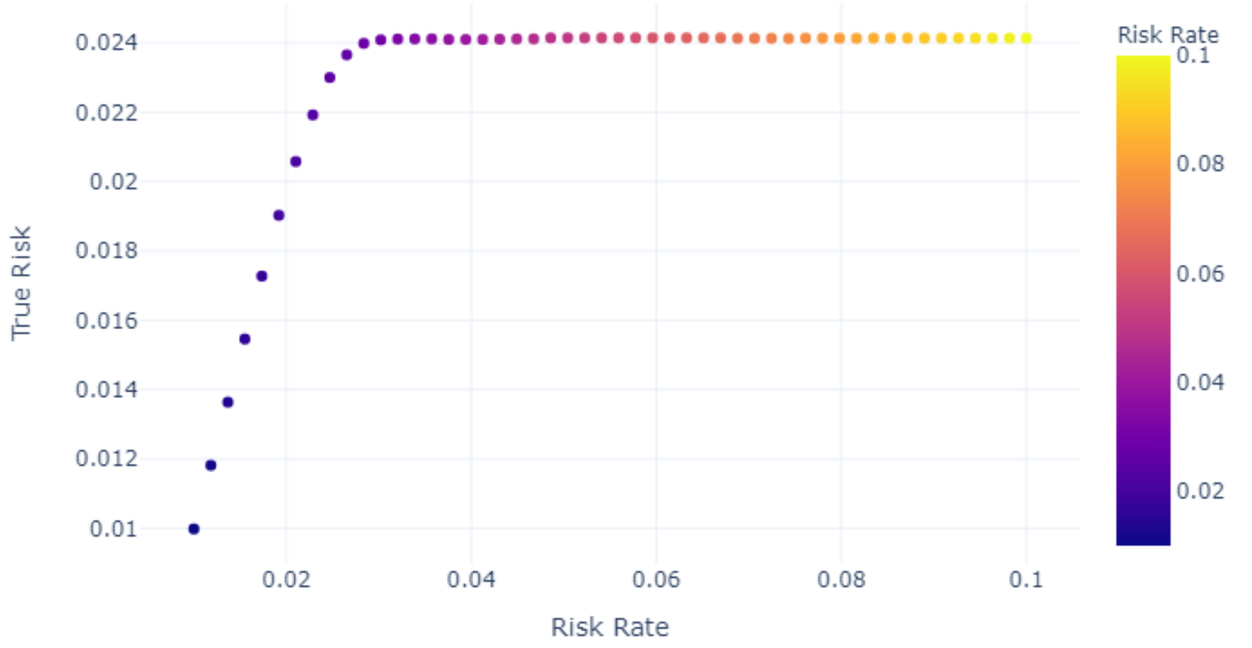
the rate of return for the example is 1.70117

True Risk: risk that the portfolio holds

model.risk()

the true risk for the example is 0.01998

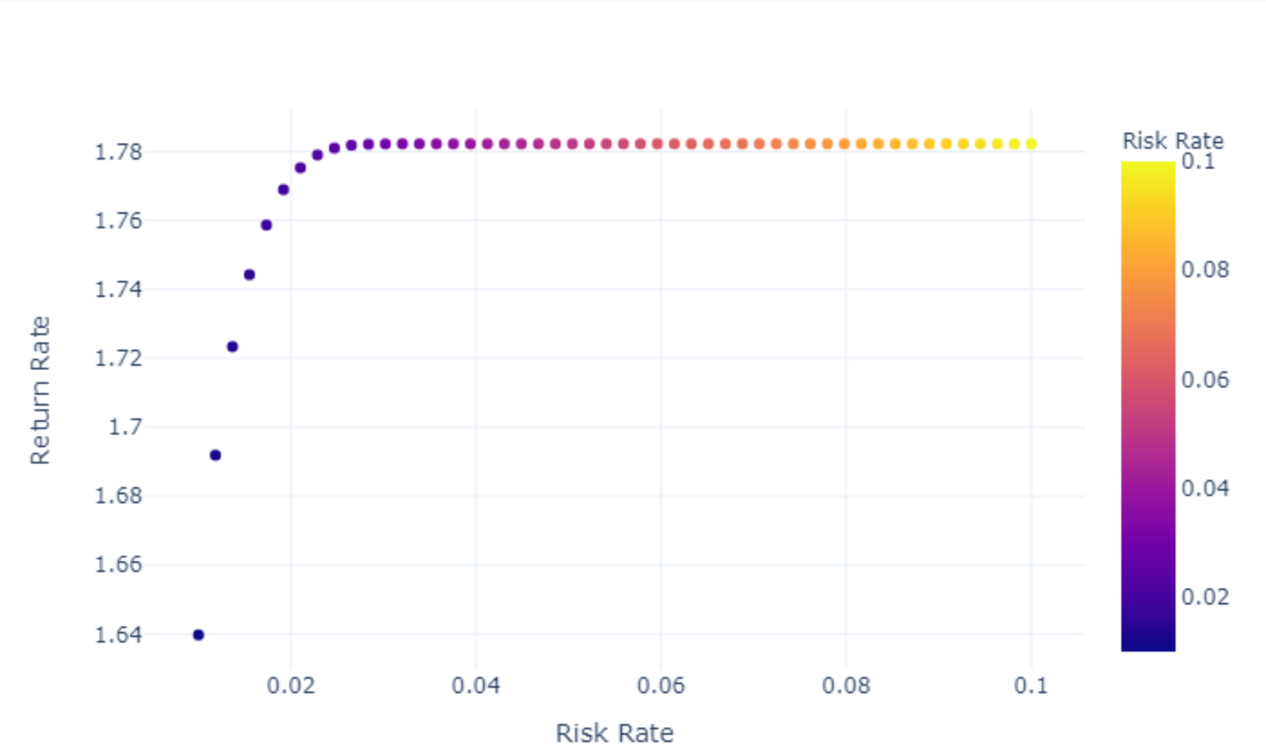
4.3 Model Simulation



- **Run 50 different risk rates**

  (from 0.01 to 0.1)
- **Simulate 100 times for each risk rate**
- **Compute expected true risk rate**

4.3 Model Simulation



- **Run 50 different risk rates**

  (from 0.01 to 0.1)
- **Simulate 100 times for each risk rate**
- **Compute expected rate of return**

Conclusion

- Our group expect that by using our optimization model, we are able to help investors in lending club to better form the portfolio in a goal of reaching the best total return given the default risks.
- Due to our capability, we realize that there are still factors that we fail to consider but would have impacts on the results of our optimization. For example, the time value of the money, the loan risk in different years and seasonality (i.e. loan risk in 2008 vs 2020, loans in Christmas Festival)
- Our group will try to continue improving our model by further doing simulation and comparing the real cases with our model results in order to offer better and more reliable recommendation for the investors.

Thank you