

資料探勘與知識發現

Final Project - Kiva

M064020025 劉育銘

B034020027 陳韻卉

B034020043 許曼軒

分析工具

以 R 為主要分析工具，python 為輔助工具。



研究主題

分析影響募款和捐款人意願的因素



如何找到重要的因素？



Text Mining

Latent Dirichlet allocation (LDA)

資料來源與分析欄位

資料來源 : **build.kiva**

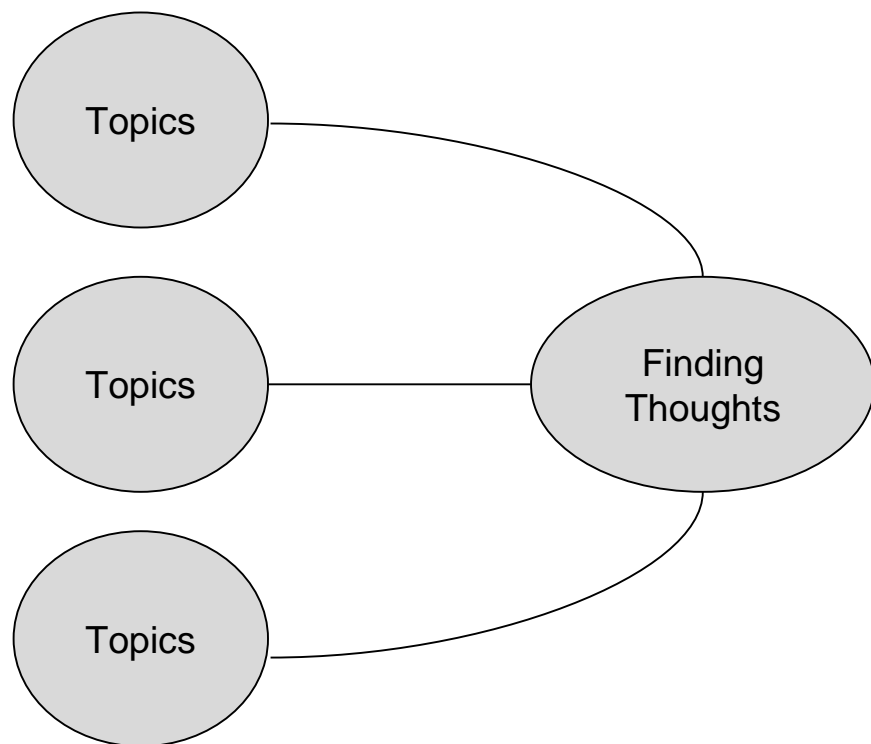
資料筆數 : **174327** 筆

分析欄位 : **loanbecause**

```
> # pre-processing:  
> reviews=lenders$LOAN_BECAUSE[lenders$LOAN_BECAUSE!='']  
> length(reviews)  
[1] 174327  
> |
```

	loanbecause
1	I believe its important to support others in their time of need
2	As the bible states: ~ ""Give someone a fish and you feed him...
3	CMS Website Development
4	I was introduced to Kiva on my recent visit to San Francisco b...
5	I need money for my jobs
6	small things together matter to make the world a little better
7	Nothing is more rewarding than watching people achieve the...
8	I want expand my business
9	It's the right thing to do. We live in a global community, and ...
10	People need help and I can help.
11	I desire this loan to establish a creche, daycare, nursery and pr...
12	I want to help make a difference in world.
13	You can't always do it on your own.
14	Captain america civil war stream

	loanbecause
1	I believe its important to support others in their time of need
2	As the bible states: ~ ""Give someone a fish and you feed him...
3	CMS Website Development
4	I was introduced to Kiva on my recent visit to San Francisco b...
5	I need money for my jobs
6	small things together matter to make the world a little better
7	Nothing is more rewarding than watching people achieve the...
8	I want expand my business
9	It's the right thing to do. We live in a global community, and ...
10	People need help and I can help.
11	I desire this loan to establish a creche, daycare, nursery and pr...
12	I want to help make a difference in world.
13	You can't alway's do it on your own.
14	Captain america civil war stream



進行斷詞

	loanbecause
1	I believe its important to support others in their time of need
2	As the bible states: ~ ""Give someone a fish and you feed him...
3	CMS Website Development
4	I was introduced to Kiva on my recent visit to San Francisco b...
5	I need money for my jobs
6	small things together matter to make the world a little better
7	Nothing is more rewarding than watching people achieve the...
8	I want expand my business
9	It's the right thing to do. We live in a global community, and ...
10	People need help and I can help.
11	I desire this loan to establish a creche, daycare, nursery and pr...
12	I want to help make a difference in world.
13	You can't alway's do it on your own.
14	Captain america civil war stream

doc.list	Large list (7731 elements, 7.8 Mb)
	: chr [1:12] "i" "believe" "its" "important" ...
	: chr [1:26] "as" "the" "bible" "states" ...
	: chr [1:3] "cms" "website" "development"
	: chr [1:67] "i" "was" "introduced" "to" ...
	: chr [1:6] "i" "need" "money" "for" ...
	: chr [1:11] "small" "things" "together" "matter" ...
	: chr [1:18] "nothing" "is" "more" "rewarding" ...
	: chr [1:5] "i" "want" "expand" "my" ...
	: chr [1:19] "its" "the" "right" "thing" ...
	: chr [1:7] "people" "need" "help" "and" ...
	: chr [1:20] "i" "desire" "this" "loan" ...
	: chr [1:9] "i" "want" "to" "help" ...
	: chr [1:8] "you" "cant" "always" "do" ...
	: chr [1:5] "captain" "america" "civil" "war" ...

資料前處理

```
# pre-processing:
reviews <- gsub("'", "", reviews) # remove apostrophes
reviews <- gsub("[[:punct:]]", " ", reviews) # replace punctuation with space
reviews <- gsub("[[:cntrl:]]", " ", reviews) # replace control characters with space
reviews <- gsub("[[:digit:]]", "", reviews) # remove number
reviews <- gsub("^[[:space:]]+", "", reviews) # remove whitespace at beginning of doc
reviews <- gsub("[[:space:]]+$", "", reviews) # remove whitespace at end of doc
```

```
# compute the table of terms:
term.table <- table(unlist(doc.list))
term.table <- sort(term.table, decreasing = TRUE)
# remove terms that are stop words or occur fewer than 5 times:

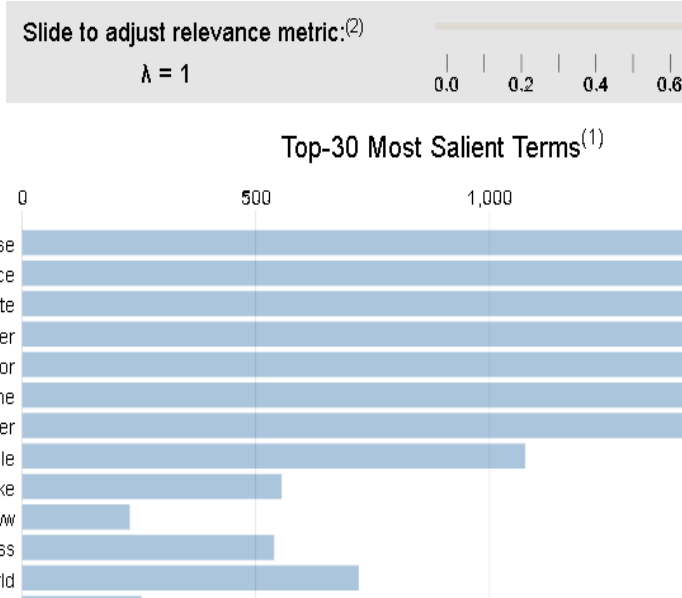
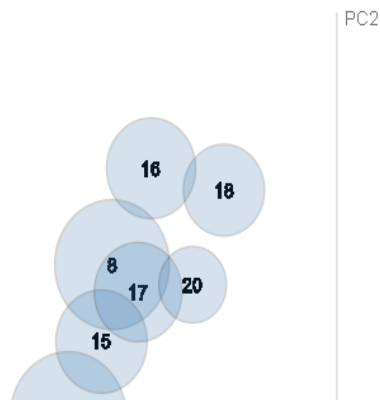
del <- (names(term.table) %in% stop_words | term.table < 5)
term.table <- term.table[!del]
vocab <- names(term.table)
```

Topic Model

LDA Visualization

Selected Topic:

Intertopic Distance Map (via multidimensional scaling)



Topic 1

I believe in the power of
small-scale entrepreneurship
to transform families,
communities, and economies.

I support access to higher
education and women's
empowerment, particularly in
developing countries.

I believe in the power of
community engagement as a
tool for economic and social
justice.

Topic 8

I believe in supporting
and empowering fellow
small business owners and
entrepreneurs to follow their
dreams.

I believe every entrepreneur
should have seed funding to
realize their dreams.

as a small business owner
I want to help other small
business owners achieve their
dreams and goals.

我們在文字探勘中 找到方向

性別



國家



開發中國家

用途

還款時間

借款額

資料探索

分析資料集

資料來源: **build.kiva**

資料筆數: **1419607** 筆

資料欄位: **34** 欄

```
In [74]: print(data.shape)
(1419607, 34)
```

```
Data columns (total 34 columns):
LOAN_ID                1419607 non-null int64
LOAN_NAME              1372049 non-null object
ORIGINAL_LANGUAGE      1376356 non-null object
DESCRIPTION            1376322 non-null object
DESCRIPTION_TRANSLATED 966929 non-null object
FUNDED_AMOUNT         1419607 non-null float64
LOAN_AMOUNT           1419607 non-null float64
STATUS                1419607 non-null object
IMAGE_ID              1376356 non-null float64
VIDEO_ID              610 non-null float64
ACTIVITY_NAME         1419607 non-null object
SECTOR_NAME           1419607 non-null object
LOAN_USE              1376334 non-null object
COUNTRY_CODE          1419598 non-null object
COUNTRY_NAME          1419607 non-null object
TOWN_NAME             1257849 non-null object
CURRENCY_POLICY       1419607 non-null object
CURRENCY_EXCHANGE_RATE 1098081 non-null float64
CURRENCY              1419607 non-null object
PARTNER_ID            1402817 non-null float64
POSTED_TIME           1419607 non-null object
PLANNED_EXPIRATION_TIME 1047773 non-null object
DISBURSE_TIME         1416794 non-null object
RAISED_TIME          1355325 non-null object
LENDER_TERM           1419583 non-null float64
NUM_LENDERS_TOTAL     1419607 non-null int64
NUM_JOURNAL_ENTRIES   1419607 non-null int64
NUM_BULK_ENTRIES       1419607 non-null int64
TAGS                  635553 non-null object
BORROWER_NAMES        1372049 non-null object
BORROWER_GENDERS       1376356 non-null object
BORROWER_PICTURED     1376356 non-null object
REPAYMENT_INTERVAL    1419607 non-null object
DISTRIBUTION_MODEL     1419607 non-null object
```

World Visualization

Map

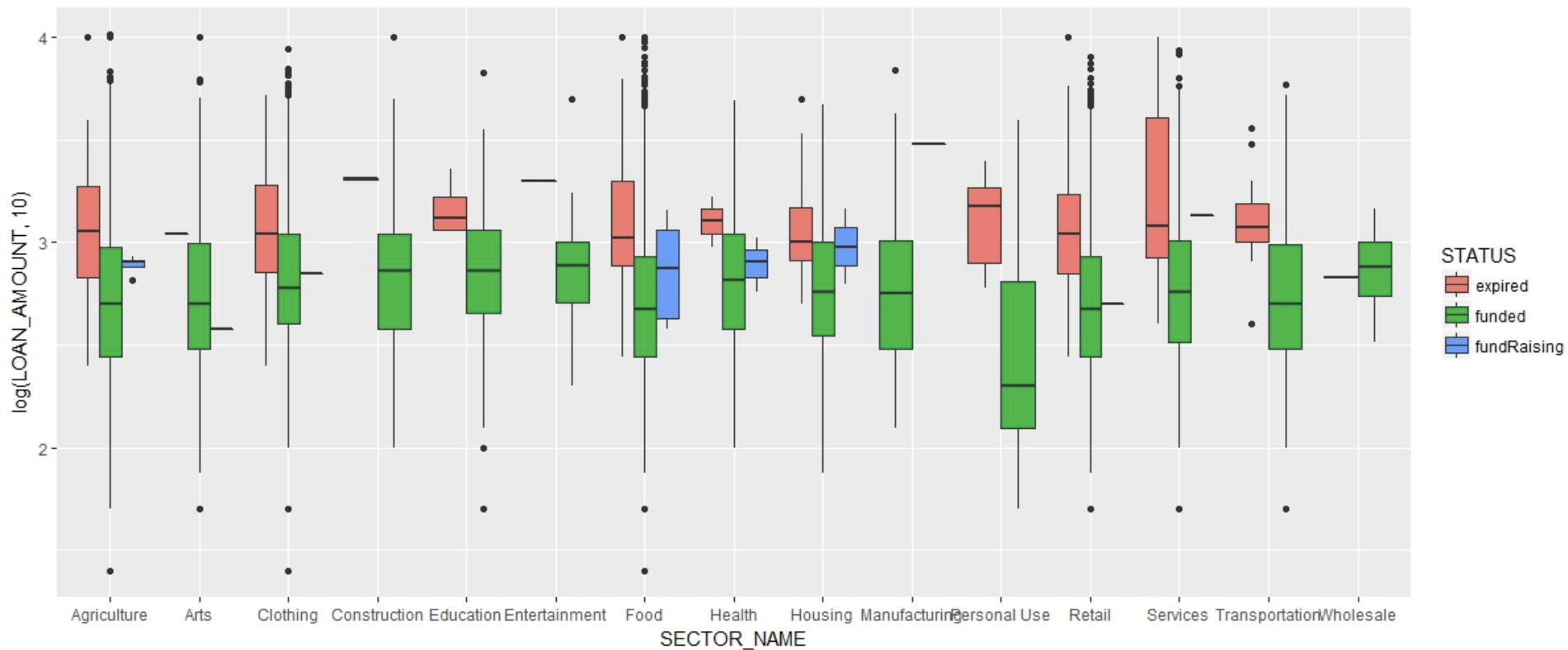
	COUNTRY_NAME	loan_num	funded_num	loan_avg	fund_ratio	2015_gdp	gdp_rank
1	Afghanistan	2337	2337	842.0839	1	569.5779	72
2	Albania	3075	3075	1400.7642	1	3934.8954	27
3	Armenia	13952	13952	1644.9595	1	3617.9357	31
4	Azerbaijan	10172	10172	1453.4629	1	5500.3104	18
5	Belize	218	218	688.8761	1	4849.9975	20
6	Benin	5946	5946	650.1556	1	783.9471	64

```
> dim(loan_all)
```

```
[1] 91 12
```

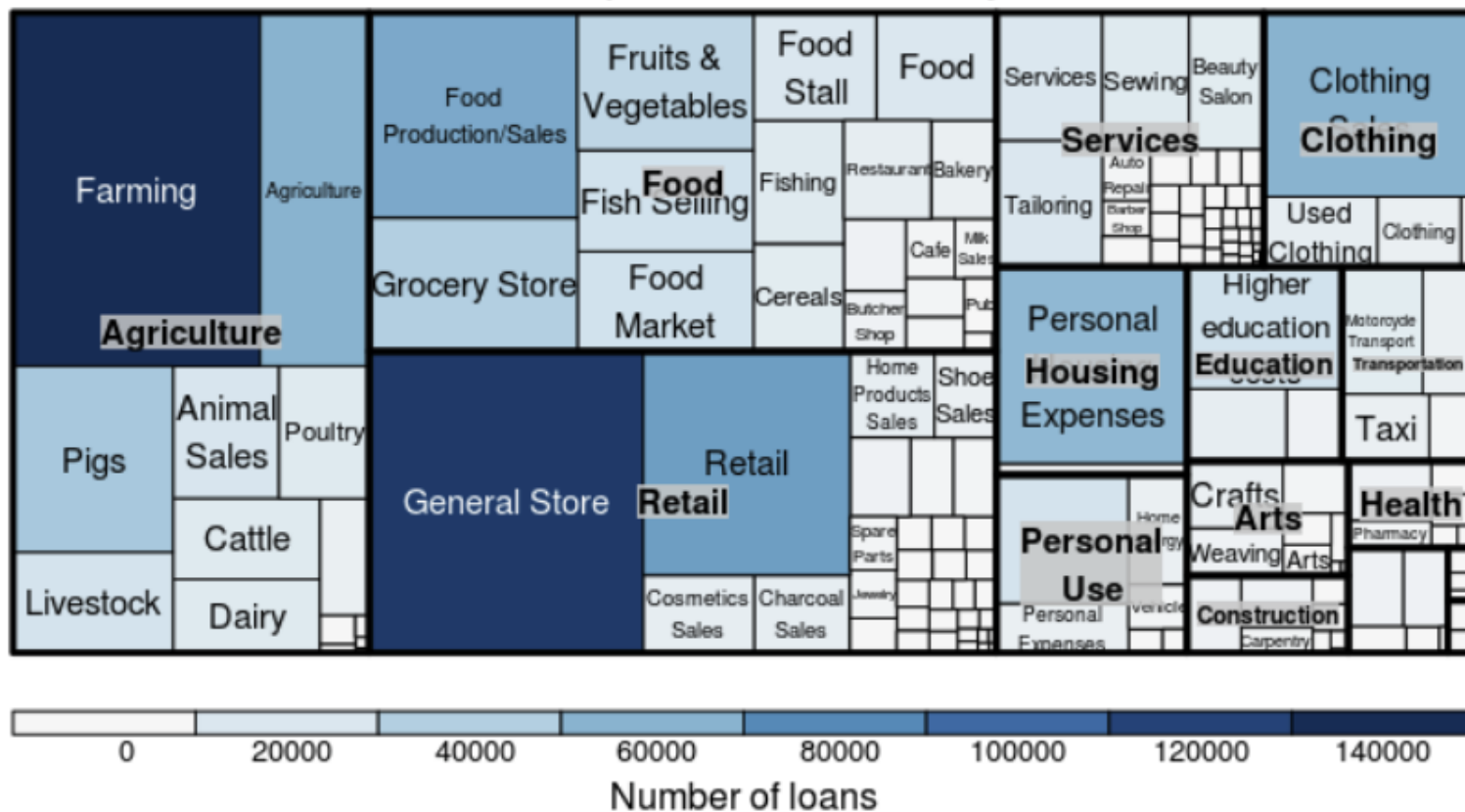
```
>
```

資料探索

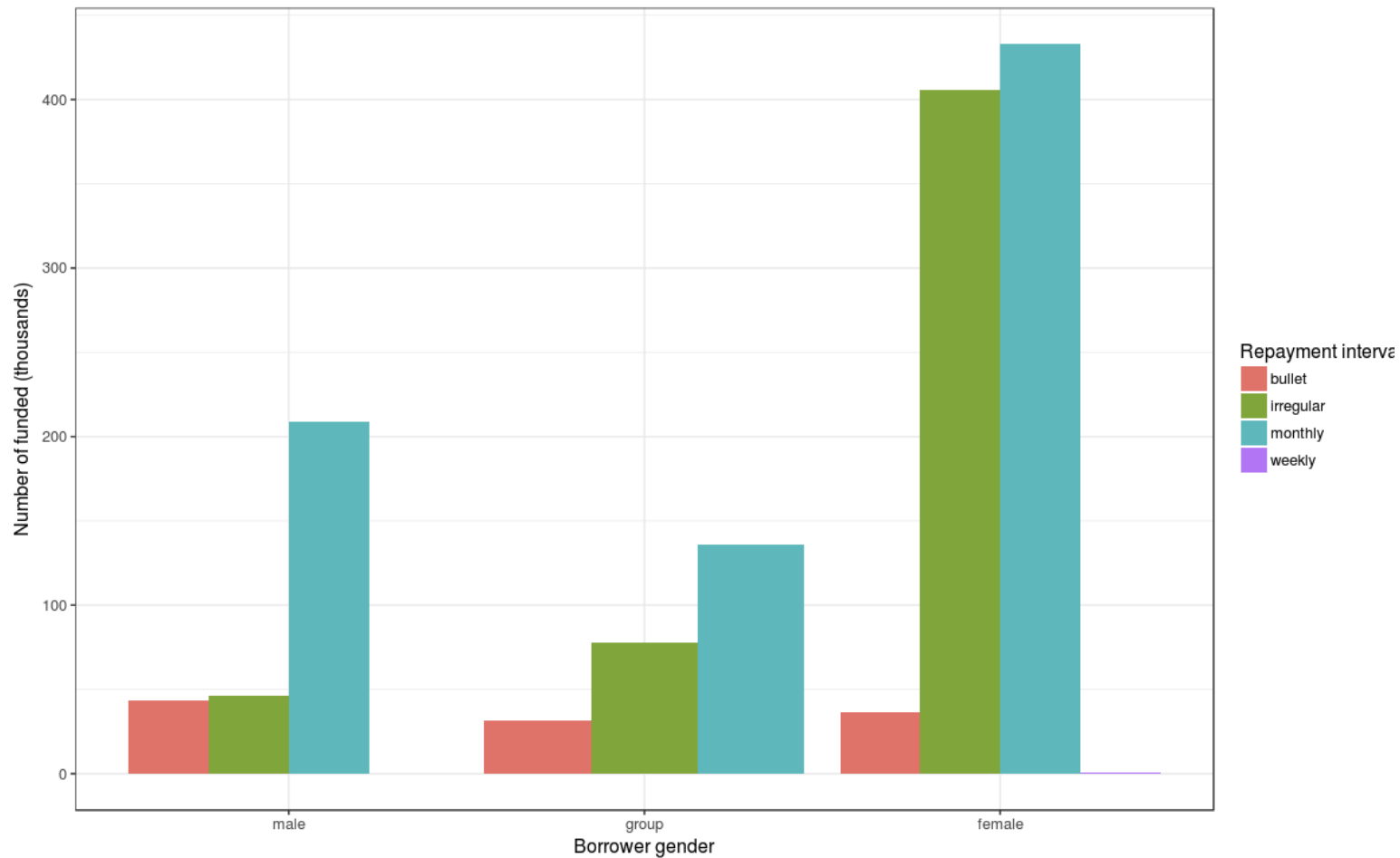


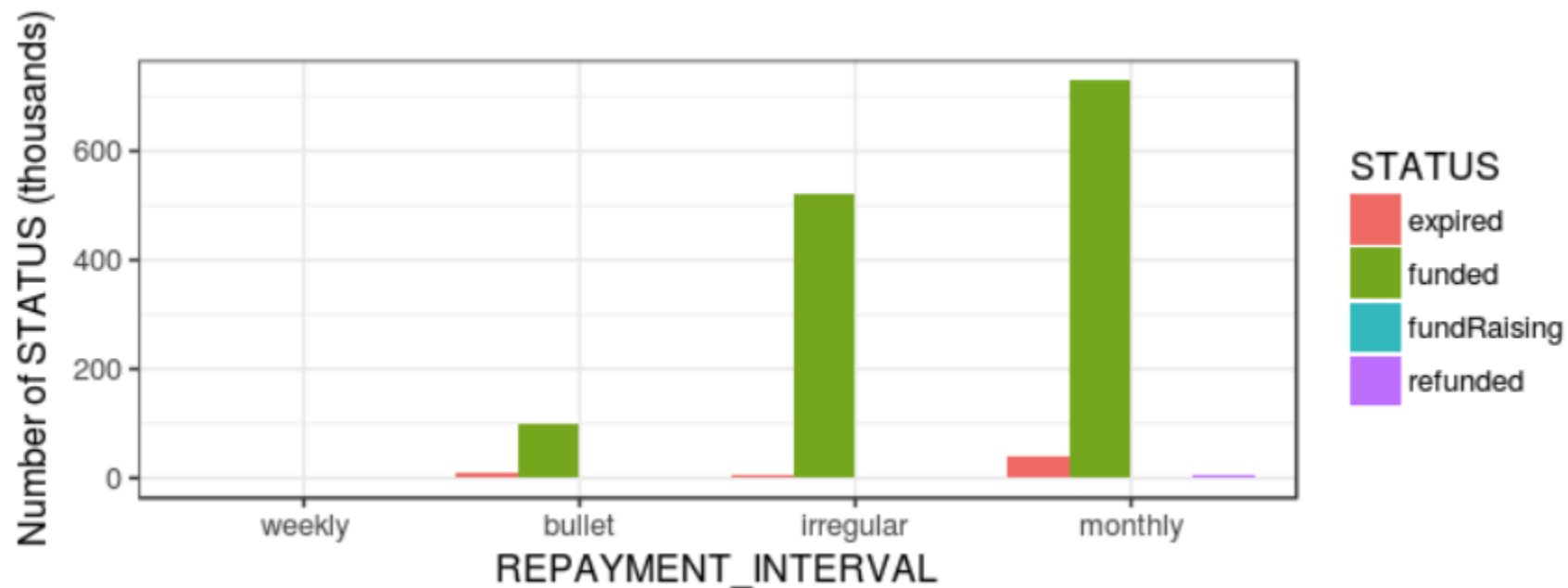
資料探索

Loans per sector and activity



資料探索





Decision Tree

預測變數 - STATUS

利用募款的成敗，來評估不同因素帶給募資案的影響。



Funded

Total loan: \$200
Powered by 7 lenders

Isah



Kaduna, Nigeria / Personal Expenses

Find a new loan



Expired

Total loan: \$1,500
Powered by 34 lenders

Mohamad



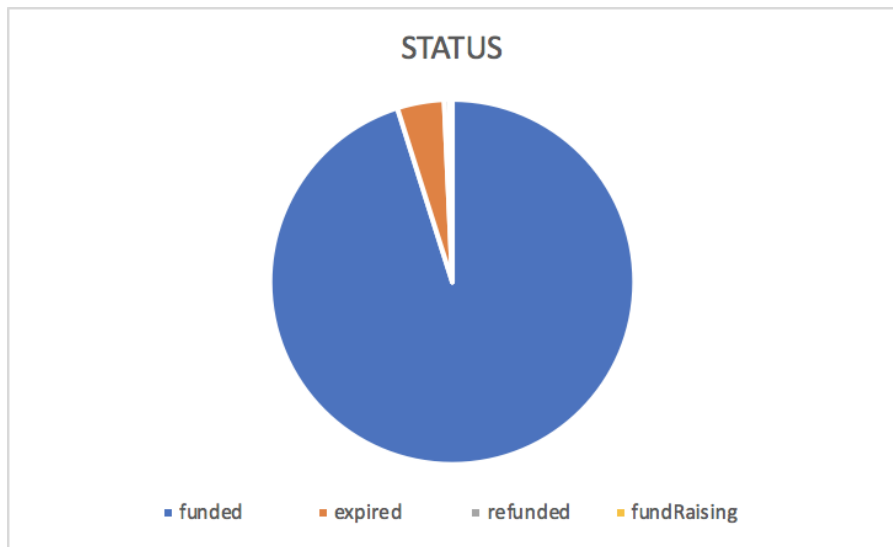
Ain, Lebanon / Agriculture

Find a new loan

資料抽樣

抽取10000筆資料，並調整funded和expired的比例為 1:1，

避免資料不平均，帶來預測準確率過高的狀況。

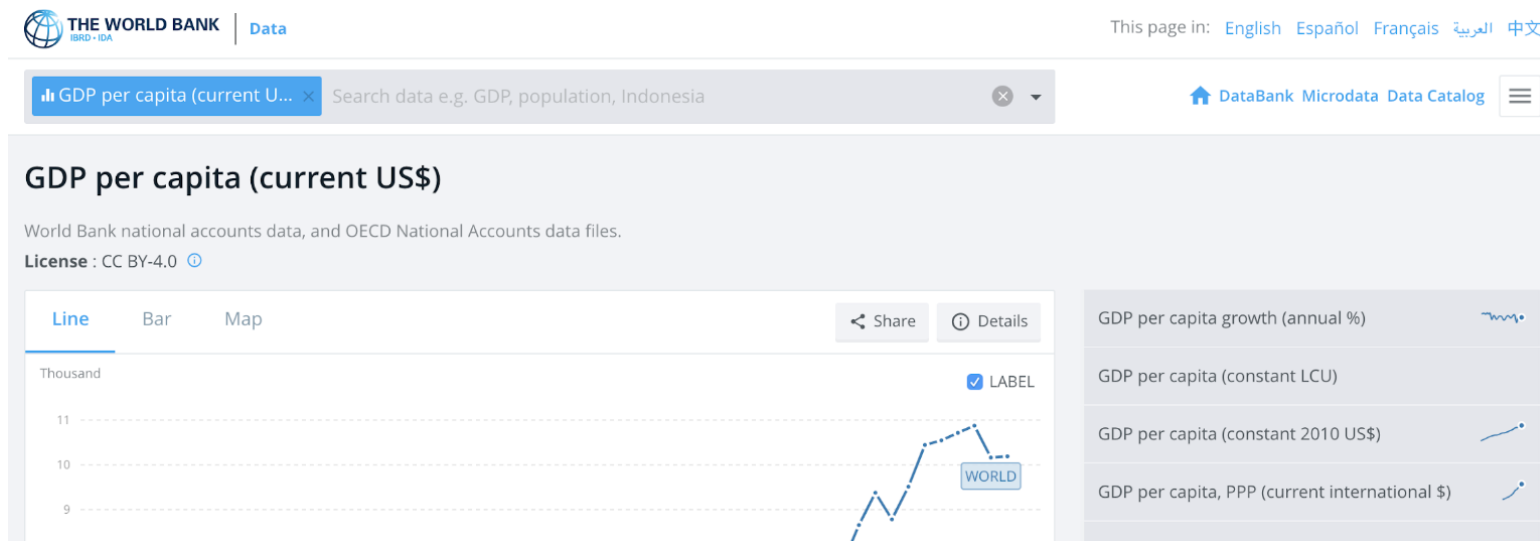


```
Out[19]:  
funded      0.951207  
expired     0.041618  
refunded    0.004634  
fundRaising 0.002542  
Name: STATUS, dtype: float64
```

加入外部資訊

外部資訊：人均國內生產總值 (GDP per capita)

資訊來源：The World Bank



屬性選擇

1. 過濾欄位：如LOAN_ID,VIDEO_ID和文字敘述型等欄位。
2. 選擇出的欄位：如性別、貸款目的、國家、捐款人數等欄位。

GDP 、gender
ORIGINAL_LANGUAGE
LOAN_AMOUNT
ACTIVITY_NAME
SECTOR_NAME
COUNTRY_CODE
CURRENCY_POLICY
LENDER_TERM
CURRENCY_EXCHANGE_COVERAGE_RATE
DISTRIBUTION_MODEL
NUM_LENDERS_TOTAL
REPAYMENT_INTERVAL



資料轉換

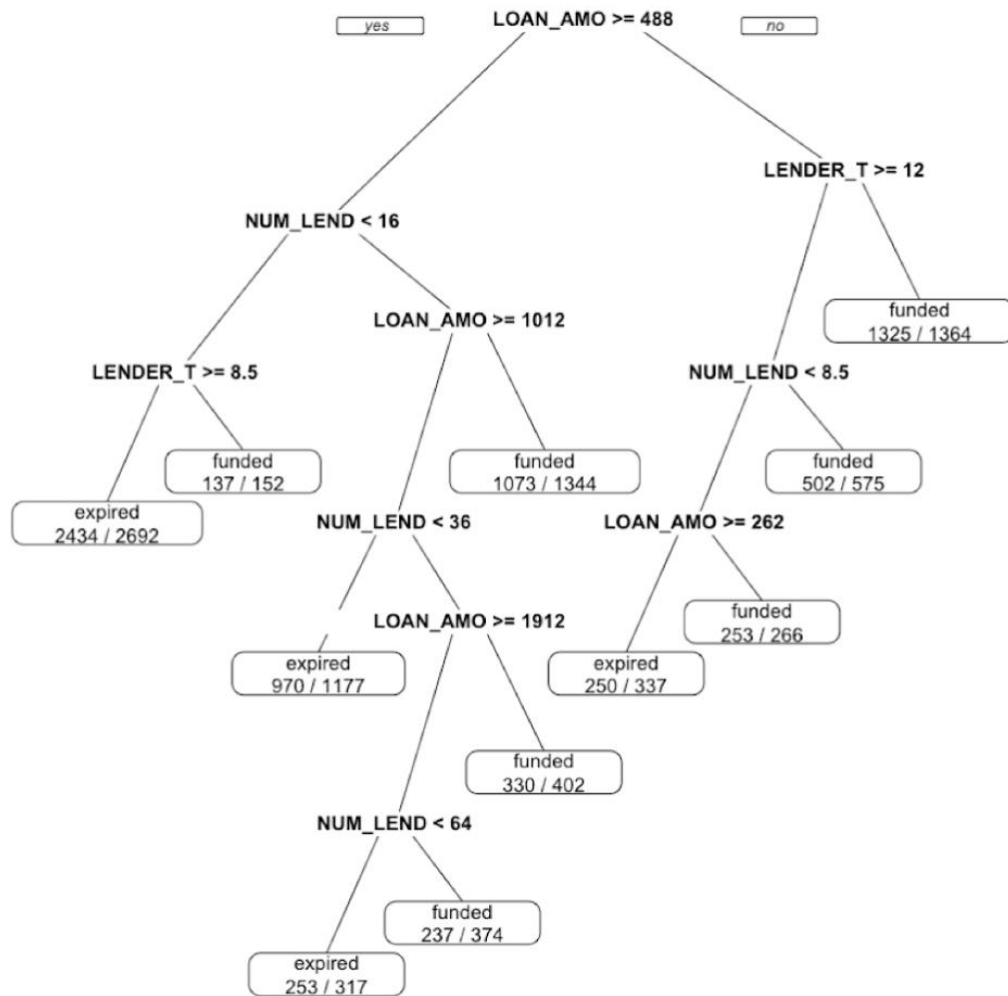
若BORROWER_GENDERS中，
同一個欄位有男性和女性，
便將其gender改變為**mix**。

female, female, female, male, female
female, female
male
male, male, male, female, female
female
female
female, male, male, male
female
female, female
female, female
female, female, female
female, female, female, female, female, female, female, female, female, fem:
female, female
female
female
female
female, female, female
female, female, female
female, female, female, female, female, female, female, female, female

決策樹

Variable importance

LOAN_AMOUNT	NUM_LENDERS_TOTAL	COUNTRY_CODE
39	24	15
LENDER_TERM	ACTIVITY_NAME	REPAYMENT_INTERVAL
10	6	3
gender	SECTOR_NAME	
1	1	



關聯法則

關聯法則

研究問題：

lender(貸款人)的个人因素會不會影響他選擇放貸的案件？

資料集：

每個row是一筆貸款資料，一個loan案件可能會有好幾筆

	lender	LOAN_ID	FUNDED_AMOUNT	LOAN_AMOUNT	STATUS	ACTIVITY
1	07272010	1176597	325	675	expired	Beauty Sal
2	Oli	87949	125	125	funded	Traveling S
3	100ofhumanity1199	1018272	700	1400	expired	Cosmetics
4	100ofhumanity1199	1161906	750	1500	expired	Grocery St
5	100ofhumanity1199	1210762	850	1225	expired	Clothing S
6	100ofhumanity1199	1233420	400	400	funded	Farming
7	100ofhumanity1199	1428215	175	475	expired	Sewing
8	123321	1095055	125	400	expired	Clothing S

關聯法則

Apriori

itemset :

{lender, 貸款人ID

SECTOR_NAME, 借款用途類別

loan_countryCode, 借款人國家

REPAYMENT_INTERVAL, 借款人還款時間

loan_gender, 借款人性別

lender_countryCode, 貸款人國家

lender_occupation, 貸款人職業

}

關聯法則

- 什麼樣的人會選擇農業用途借貸？
- rhs : SECTOR_NAME=Agriculture

$A \Rightarrow B$ (當發生A時，則會發生B)

support: {AB}項目集在全部資料出現的比例

confidence: 在{A}出現的所有記錄中，{B}也同時出現的比例

lift:support/((support(A)*(support(B)))

```
/ > rules.inspect(surrogate)
lhs                                     rhs                                     support confidence lift count
[1] {lender=themiissionbeltco}          => {SECTOR_NAME=Agriculture} 0.012789281 1.0000000 4.077815 63
[2] {lender_occupation=Making the coolest belts around. No Holes. Always Fits.} => {SECTOR_NAME=Agriculture} 0.012789281 1.0000000 4.077815 63
[3] {lender=themiissionbeltco,         => {SECTOR_NAME=Agriculture} 0.012789281 1.0000000 4.077815 63
    lender_occupation=Making the coolest belts around. No Holes. Always Fits.}
[4] {lender=themiissionbeltco,         => {SECTOR_NAME=Agriculture} 0.006496143 1.0000000 4.077815 32
    loan_gender=male}
[5] {lender=themiissionbeltco,         => {SECTOR_NAME=Agriculture} 0.012789281 1.0000000 4.077815 63
    lender_countryCode=US}
[6] {lender=themiissionbeltco,         => {SECTOR_NAME=Agriculture} 0.006496143 1.0000000 4.077815 32
    REPAYMENT_INTERVAL=monthly}
[7] {loan_gender=male,                 => {SECTOR_NAME=Agriculture} 0.006496143 1.0000000 4.077815 32
    lender_occupation=Making the coolest belts around. No Holes. Always Fits.}
```

Kiva Lender Mission Belt Co



Location: Orem, Utah, United States

Occupation: Making the coolest belts around. No Holes. Always Fits.

I loan because: Our MISSION is to fight global poverty by empowering people with the tools need become self-sufficient.

About me: Inspired by modern, simplistic design and intuitive functionality, we gave this clas accessory a fresh start. We don't like belts... we LOVE them. We are a BELT compo

Check out: <http://www.missionbelt.com>

Member Since: Jul 5, 2012

Send message

Send a Kiva Card

Portfolio distribution

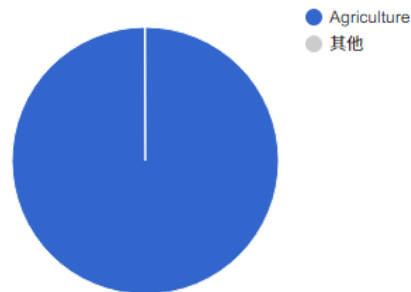
By gender

By country

By sector

By partner

Sector



Mission Belt Co's loans

1-28 of 101828 loans

關聯法則

- 什麼樣的人會選擇教育用途借貸？
- rhs : SECTOR_NAME=Education

	lhs	rhs	support	confidence	lift	count
[1]	{lender_occupation=Philosophy Professor}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[2]	{lender=christine21386582}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[3]	{lender=christine21386582, lender_occupation=Philosophy Professor}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[4]	{loan_gender=female, lender_occupation=Philosophy Professor}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[5]	{lender_countryCode=US, lender_occupation=Philosophy Professor}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[6]	{REPAYMENT_INTERVAL=monthly, lender_occupation=Philosophy Professor}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5
[7]	{lender=christine21386582, loan_gender=female}	=> {SECTOR_NAME=Education}	0.001015022	1	46.91429	5

Kiva Lender Chris



Location: Cambridge, MA, United States

Occupation: Philosophy Professor

I loan because: I lend to support the cause of women's education.

Check out: <http://www.people.fas.harvard.edu/~korsgaar/>

Member Since: Jan 12, 2014

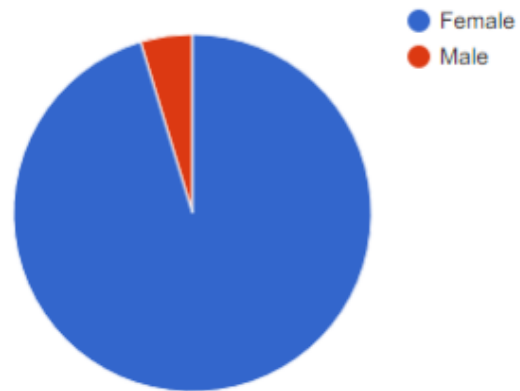
Send message

Send a Kiva Card

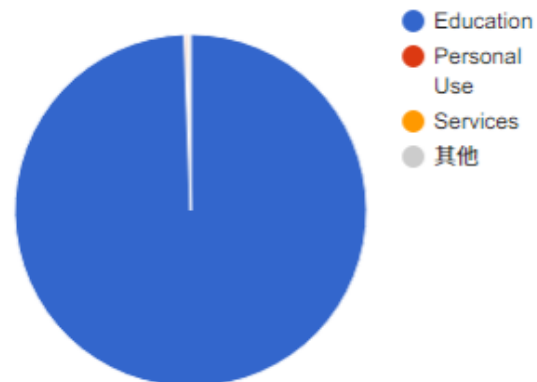
Chris's loans

1-28 of 2560 loans

Gender



Sector



Kiva Lender Good Dogg



Location: Front Porch

Occupation: www.linkedin.com/in/peacekeeper

I loan because: Jesus Christ is my Lord and Savior and He is good.

About me: For where your treasure is, there your heart will be also. Matthew 6:21

Check out: <http://www.youtube.com/watch?v=oUzBr49xqjk&feature=g-all-u&context=G270e2ccFAAAAAAAAAACAA>

Member Since: Jun 5, 2007

[Send message](#)

[Send a Kiva Card](#)

Good Dogg's loans

1-28 of 189675 loans

Kiva Lender Koa



Location: New Jersey

Occupation: Super Dog

I loan because: The dog told me to do it.

Member Since: Aug 17, 2006

[Send message](#)

[Send a Kiva Card](#)

Koa's loans

1-28 of 27875 loans

問卷回饋

(符合)

假設：sector為Agriculture裡，activity為Farming的比例是最高的

答案選項	填答次數	百分比
■ 幫助Santita購買葡萄、豆子，以及其他需要種植葡萄的物品	17	50%
■ 幫助Noriniso開拓家畜繁殖的事業與購買黃牛	12	35.2%
■ 幫助Lidia購買家畜及清潔用品	5	14.7%

(符合)

假設：GDP值低(題目有給國家GDP資料)，funded比例高

答案選項	填答次數	百分比
<input checked="" type="checkbox"/> 在烏干達的Damalie是位教師，她想為學校建築做裝修	21	61.7%
<input checked="" type="checkbox"/> 在巴基斯坦的Sana開創了學校，想新建學校建築	13	38.2%

(符合)

假設：貸款金額低，funded比例高

答案選項	填答次數	百分比
 Andi想貸款550美元，為他的社區提供行動支付服務	28	82.3%
 Francisco想貸款9750美元，創建一個跨國購物網站	6	17.6%

(不符合)

假設：借貸人為女生，funded比例高



答案選項	填答次數	百分比
■ 幫助Seferino購買農業用品來種植玉米，支持他的孩子學習與前進	19	55.8%
■ 幫助Francisca購買椰子來出售，支持她與她的孩子	15	44.1%

(不符合)

假設：案件目前貸款人多，funded比例高

答案選項	填答次數	百分比
 Aizaada想購買家畜，目前願意貸款給他的人有8位	23	67.6%
 Burul想購入馬匹，目前願意貸款給他的人有97位	11	32.3%

(不符合)

假設：一次付清，funded比例高

答案選項	填答次數	百分比
<p>幫助Rosmery購買農業用品如肥料</p> <p>■ 等，改善農田的收穫。她將不定期還款。</p>	19	55.8%
<p>幫助Ramatoulaye維持他的生意營</p> <p>■ 運，貸款能使她拓展銷售。她將一次付清她的貸款。</p>	15	44.1%

Thanks for listening!

QA