

The background features several abstract blue geometric shapes. On the left, there is a large, light blue rounded shape and a smaller, darker blue triangle. On the right, there is a large, dark blue triangle. A thin blue line runs diagonally from the bottom left towards the top right. There are also several small, light blue squares and circles scattered across the page.

Personal Finance Concerns

Content Analysis Based on the Reddit Data

Jinfei Zhu

MACSS 1st Year

CONTENTS

01

Introduction

Project and Reddit Data introduction

02

Counting words & Phrases

Counting, Normalization, Conditional Distribution, POS analysis

03

Clustering & Topic Modeling

K-means, Hierarchical clustering, Dynamic topic modeling

04

Word Embedding

Word2Vec, Doc2Vec, Projection

Project Introduction

- In 2017, 40% of American adults wouldn't be able to cover a \$400 emergency with cash
- What are the most frequent personal finance concerns?
- Do the topics of these concerns change over time?
- Do people talk differently?

Project Introduction

- In 2017, 40% of American adults wouldn't be able to cover a \$400 emergency with cash
- What are the most frequent personal finance concerns?
- Do the topics of these concerns change over time?
- Do people talk differently?

Corpora

PRAW: The Python Reddit API Wrapper



Personal Finance

r/personalfinance

About Community



Learn about budgeting, saving, getting out of debt, credit, investing, and retirement planning. Join our community, read the PF Wiki, and get on top of your finances!

14.4m

Members

14.9k

Online

Created Feb 9, 2009



wallstreetbets

r/wallstreetbets

About Community



Like 4chan found a bloomberg terminal

9.6m

Degenerates

255k

Buying FDs

Created Jan 31, 2012



Lose money with friends!

r/investing

About Community



Welcome to investing

1.8m

Members

9.0k


Online

Created Mar 15, 2008

Sample Texts & Top words

Title: “If you can't get your emergency fund to grow because of emergencies that keep coming up, you're still doing a good job.”

Text: “Over the summer I made a steadfast commitment to getting my 3 month emergency fund built, which is only about 15k. I'm saving \$750 a month, which is exactly 15% of my family's post-tax income. In the 3 months since I made that change, I've had \$1.8k in car repairs, \$600 in vet bills, and \$250 to cover a friend who got towed from our guest parking (our fault). Needless to say, the needle hasn't moved as I wanted it to, and I have to keep reassuring myself that, had I not made this commitment, I'd be in real trouble covering these costs. The end goal will come eventually. EDIT: Just to clarify - this is a two person budget!.”

Personal Finance			Wall Street Bets			Investing		
rank	word	count	rank	word	count	rank	word	count
1	\$	2236	1	\$	295	1	\$	1677
2	money	1153	2	gme	285	2	>	902
3	credit	1006	3	shares	234	3	market	854
4	time	872	4		217	4	company	629
5	pay	834	5	short	191	5	stock	549
6	edit	811	6	people	163	6	price	536
7	account	809	7	edit	155	7	said	470
8	like	771	8	market	155	8	year	462
9	know	694	9	like	148	9	people	455
10	people	692	10	buy	147	10	like	434
11	work	627	11	stock	139	11	time	407
12	years	617	12	money	135	12	=	372
13	year	590	13	price	131	13	money	360
14	job	581	14	know	119	14	billion	358
15	2	580	15	fucking	118	15	companies	355
16	car	572	16	time	112	16	years	329
17	going	569	17	going	101	17	shares	319
18	bank	549	18	sell	93	18	value	318
19	got	541	19	hedge	90	19	short	307
20	want	539	20	want	88	20	stocks	295

r/personalfinance



r/wallstreetbets

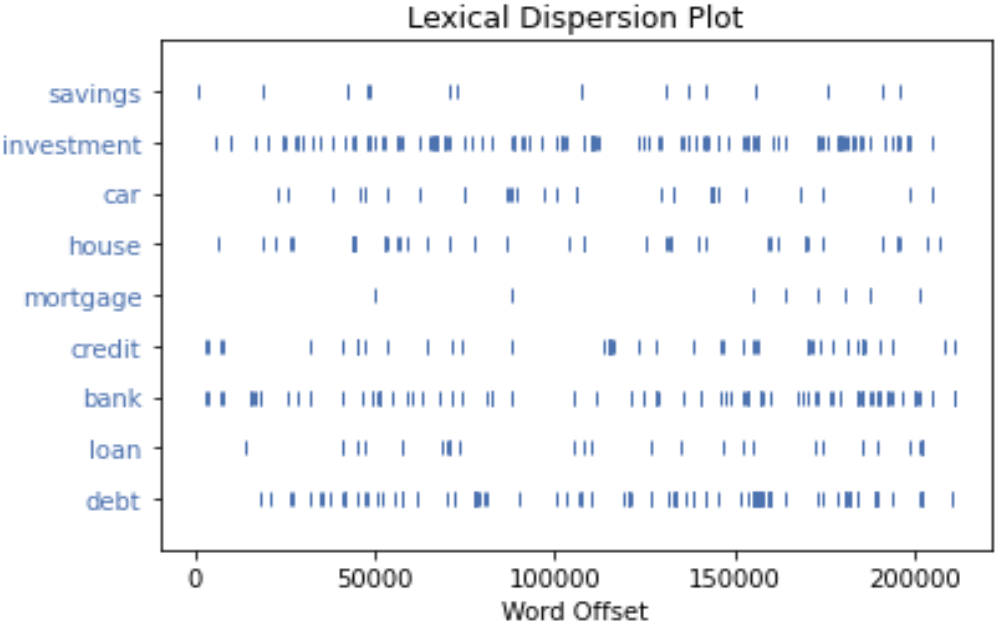


r/investing



Dispersion Plot & n-gram

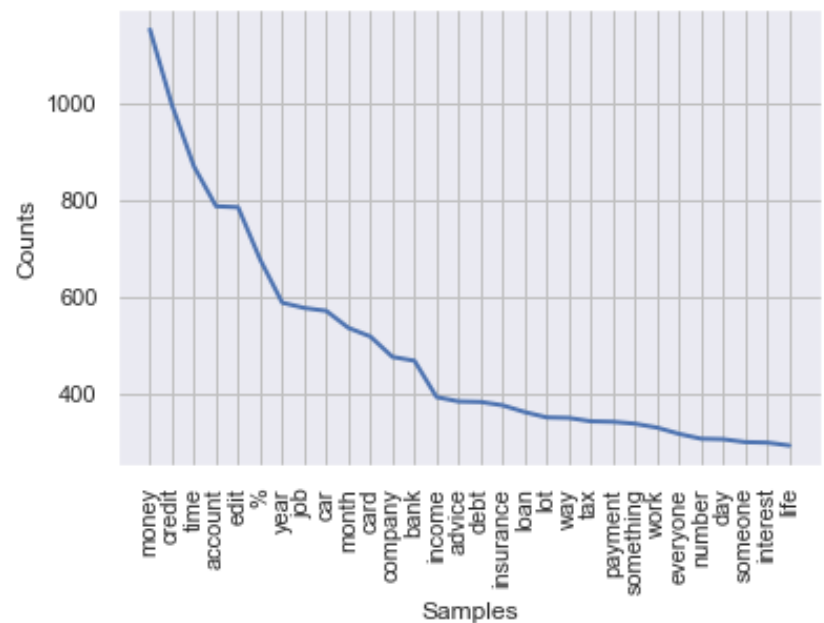
id my own way had about \$ 20,000 in student loans when i graduated first year o
ncy-tops-166-billion?srnd=premium > student loan delinquencies surged last year
ren 's expenses my personal opinion student debt will drag down the economy ove
ehold debt with strong increases in student loan and auto debt while mortgage d
ryone i 'm a mechanical engineering student with a hobby interest in finance i
ple to take credit card advances or student loans to buy tesla stock at \$ 600 +
\$ 2.31 billion of that cash advance student loan and summer job money during th
the act was calling for changes to student loans to allow borrowers to refinan
ghed in on the dangers of exploding student debt this debt is stopping a growin
g-2017 05 22 it 's easy to see that student loans put economic pressure on mill
des small spending behavior changes student loans are killing spending altogeth
t data](https://studentloanhero.com student loan debt statistics/ 44,000,000 am
atistics/ 44,000,000 americans with student loan debt average loan delinquency
40,000,000,000 estimated total u.s. student loan debt in 2012 71 percent of stu
duating from four year colleges had student loan debt represents 1.3 million st



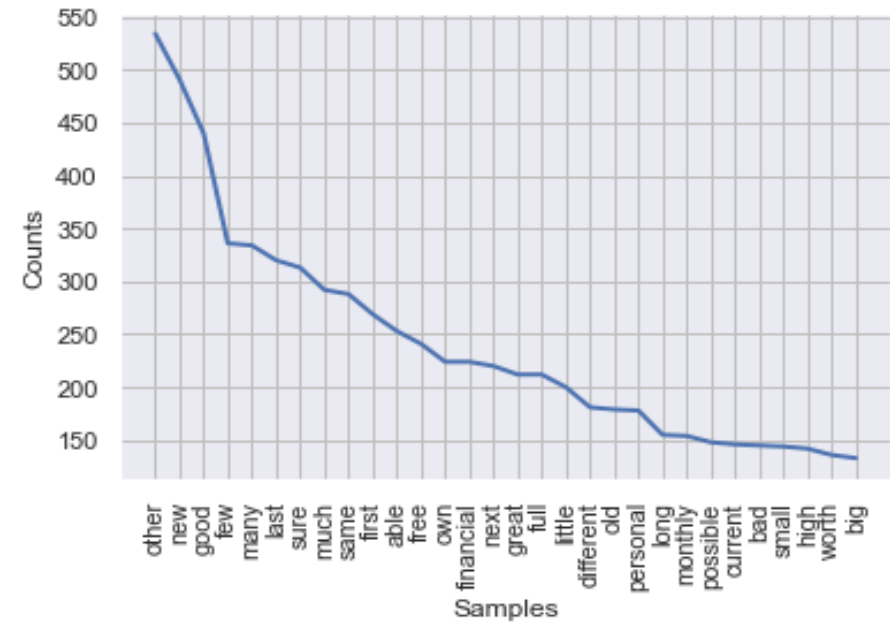
bigram	likelihood	bigram	t	trigram	t
(credit, card)	2643.3	(credit, card)	18.4	(credit, card, debt)	6.2
(r, personalfinance)	1550.0	(student, loan)	12.5	(r, personalfinance, wiki)	5.8
(student, loan)	1459.9	(\$, month)	12.3	(domain, core, finance)	5.5
(emergency, fund)	867.2	(feel, like)	10.6	(finance, domain, core)	5.5
(wells, fargo)	819.7	(r, personalfinance)	10.4	(economic, finance, domain)	4.8
(feel, like)	772.1	(year, ago)	9.6	(pay, credit, card)	4.7
(credit, score)	678.8	(edit, thank)	9.4	(use, credit, card)	4.4
(year, ago)	640.5	(credit, score)	9.3	(, \$)	4.2
(\$, month)	628.4	(m, sure)	9.3	(long, story, short)	4.1
(debit, card)	571.5	(bank, account)	9.2	(credit, card, company)	4.1

POS Conditional Distribution

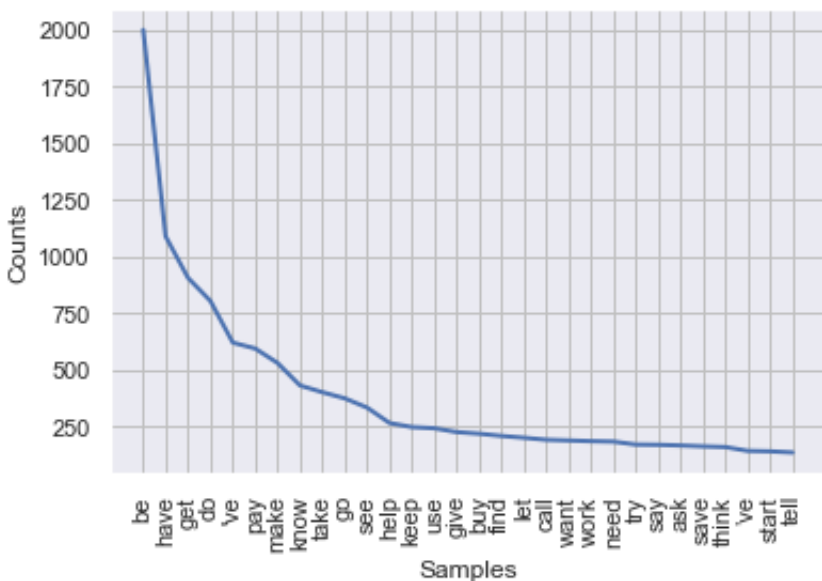
Nouns



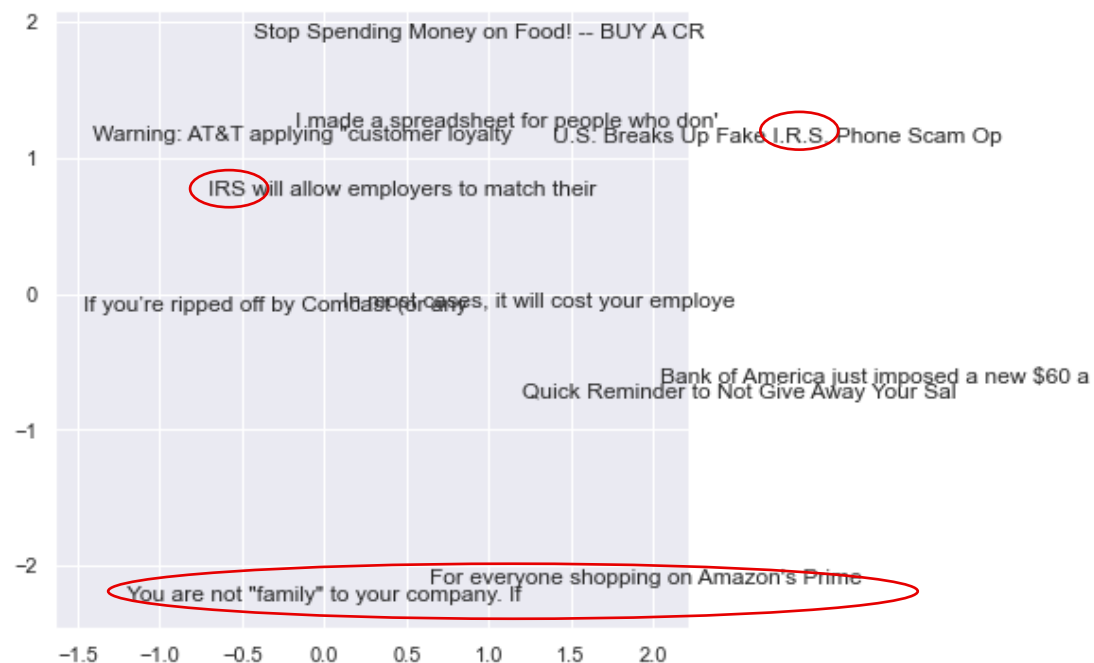
Adjectives



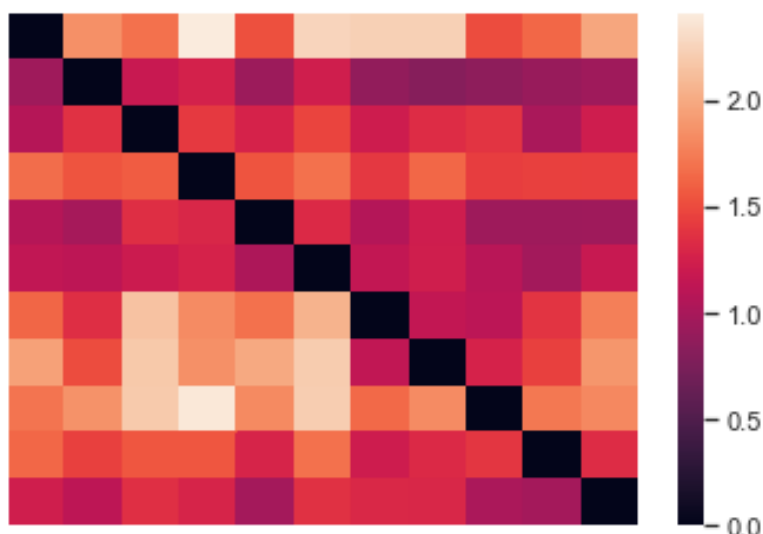
Verbs



Distributional Distances



You are not "family" to your company. If
Warning: AT&T applying "customer loyalty"
If you're ripped off by Comcast (or any
U.S. Breaks Up Fake I.R.S. Phone Scam Op
I made a spreadsheet for people who don'
Stop Spending Money on Food! -- BUY A CR
Quick Reminder to Not Give Away Your Sal
Bank of America just imposed a new \$60 a
For everyone shopping on Amazon's Prime
In most cases, it will cost your employee
IRS will allow employers to match their



You are not "family" to your company. If
Warning: AT&T applying "customer loyalty"
If you're ripped off by Comcast (or any
U.S. Breaks Up Fake I.R.S. Phone Scam Op
I made a spreadsheet for people who don'
Stop Spending Money on Food! -- BUY A CR
Quick Reminder to Not Give Away Your Sal
Bank of America just imposed a new \$60 a
For everyone shopping on Amazon's Prime
In most cases, it will cost your employee
IRS will allow employers to match their

'For everyone shopping on Amazon\'s Prime Day:
"savings" from sales aren\'t savings if you weren\'t
already planning on buying the item.'

'You are not "family" to your company. If you have an
opportunity to better yourself, take it.
They will do the same when it comes to cutting ties with you.'

K-means Clustering

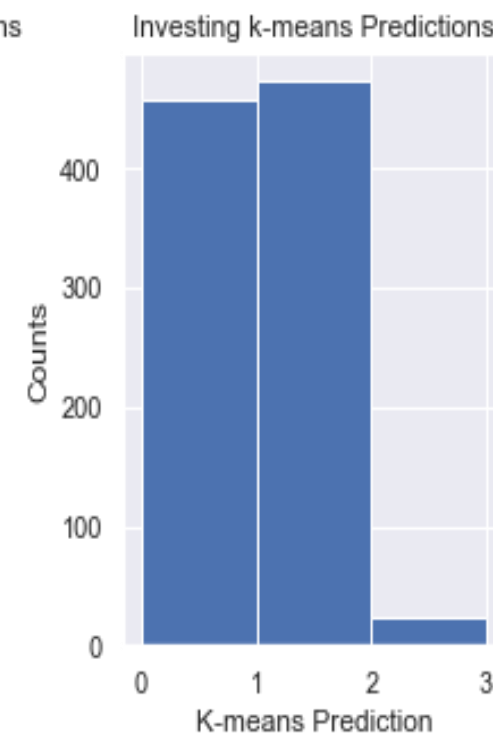
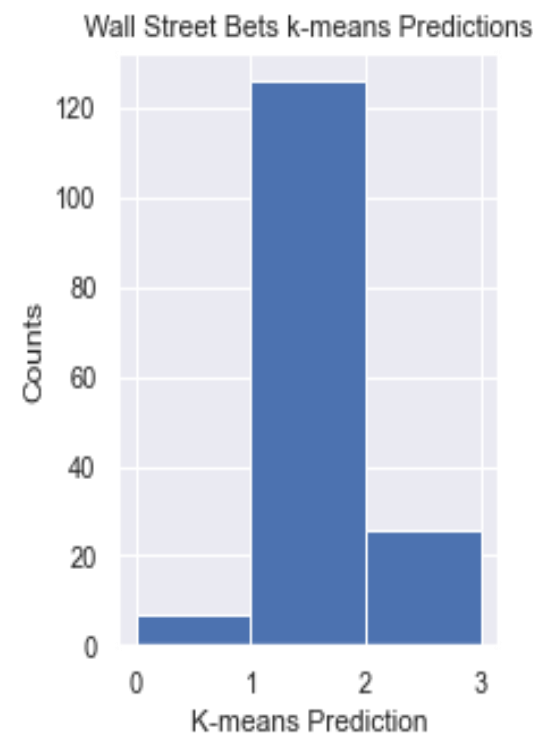
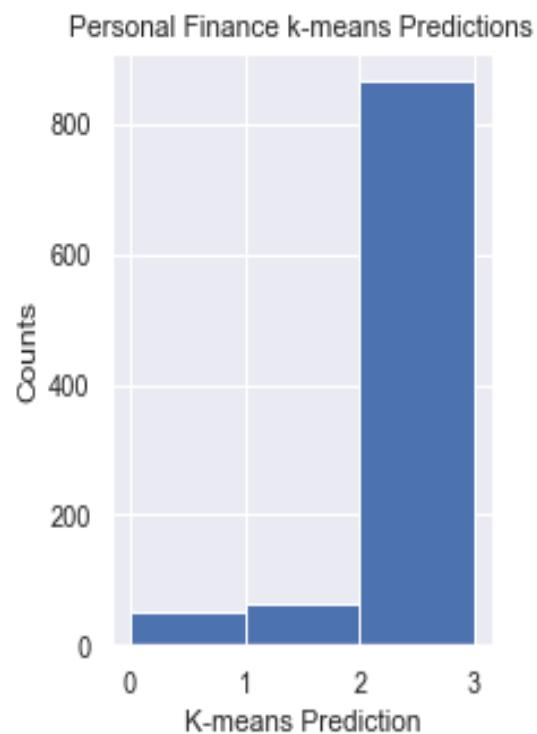
Homogeneity: 0.504

Completeness: 0.428

V-measure: 0.463

Adjusted Rand Index: 0.513

Cluster 0	Cluster 1	Cluster 2
com	market	just
https	stock	money
www	https	credit
cnbc	com	account
html	gme	edit
2020	stocks	ve
2019	price	don't
news	year	job
http	people	card



K-means Clustering

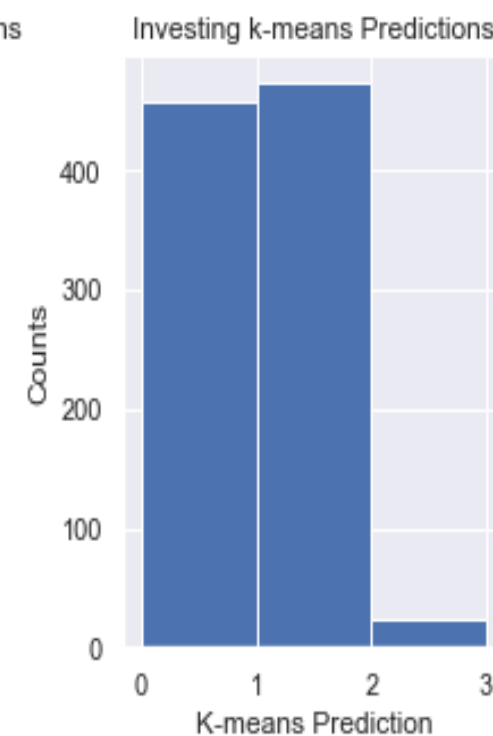
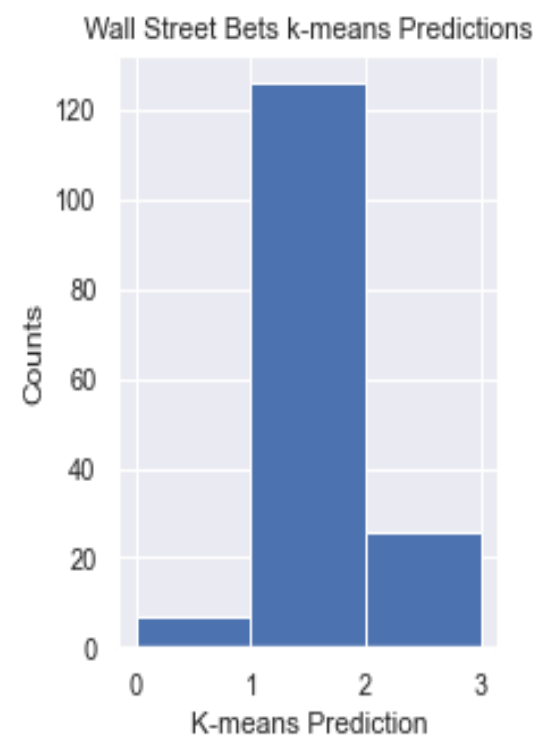
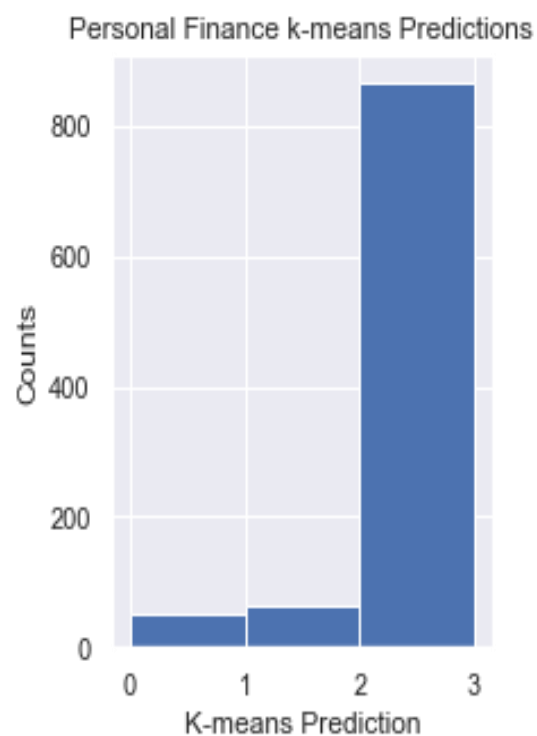
Homogeneity: 0.504

Completeness: 0.428

V-measure: 0.463

Adjusted Rand Index: 0.513

Cluster 0	Cluster 1	Cluster 2
com	market	just
https	stock	money
www	https	credit
cnbc	com	account
html	gme	edit
2020	stocks	ve
2019	price	don't
news	year	job
http	people	card

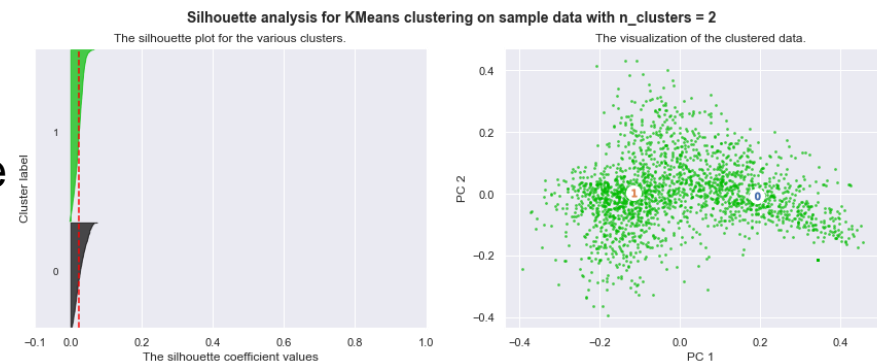


True & Predicted Clusters

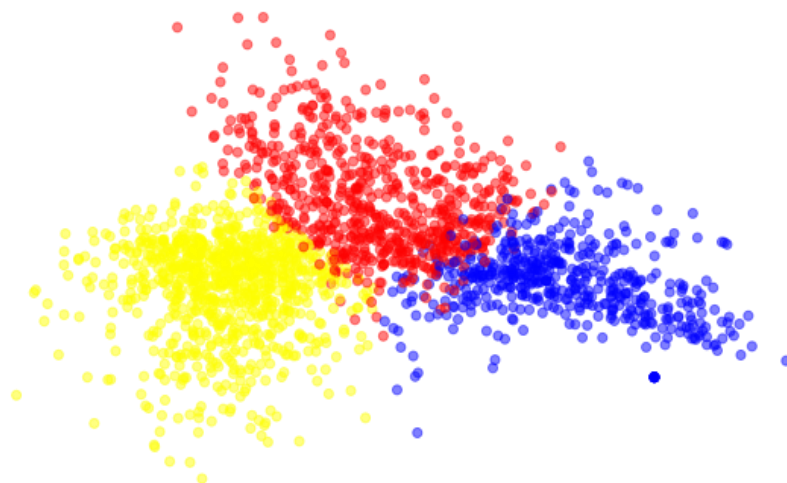
True Class



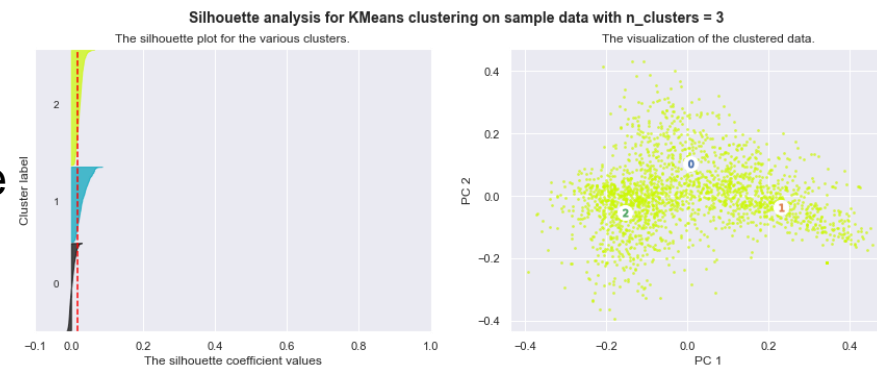
Silhouette Score
0.025



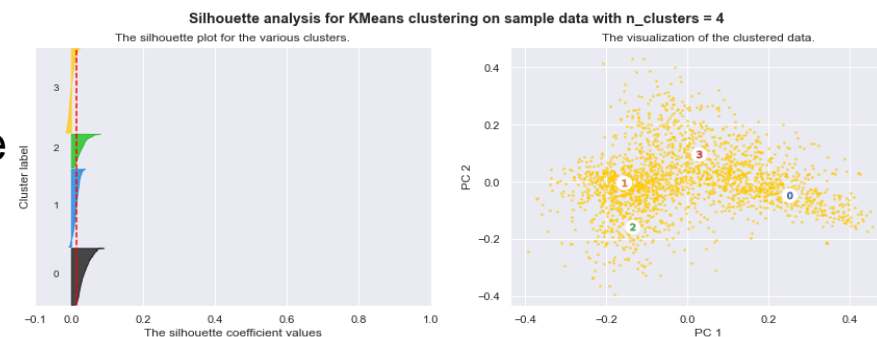
Predicted Clusters



Silhouette Score
0.019

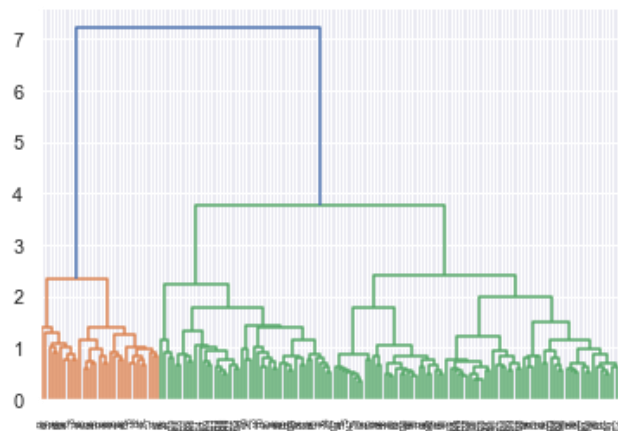


Silhouette Score
0.016

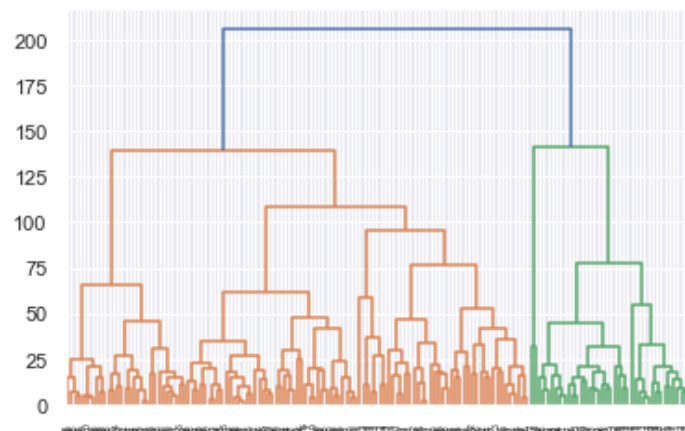


Hierarchical clustering

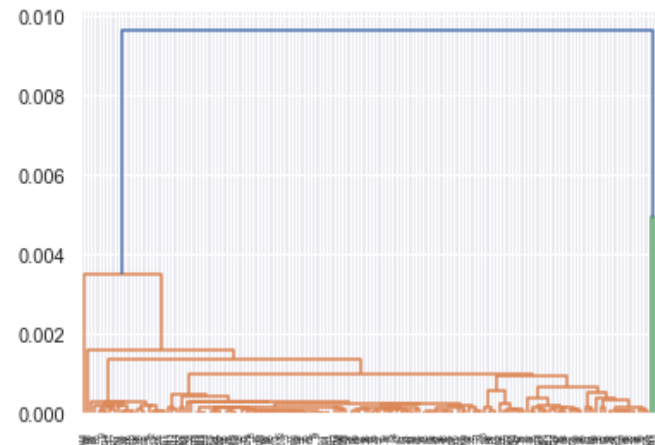
Euclidean



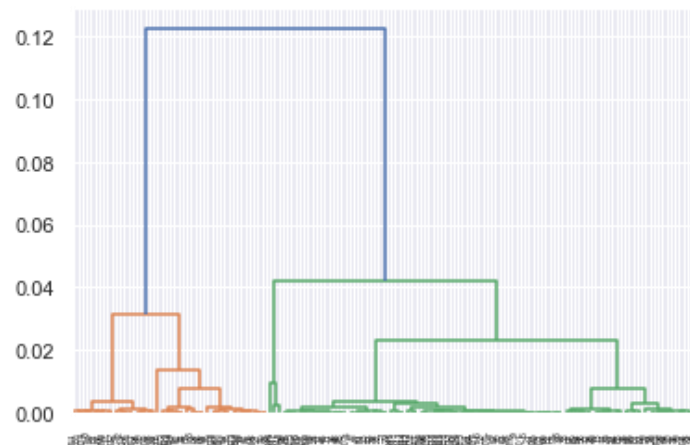
Average



Correlation

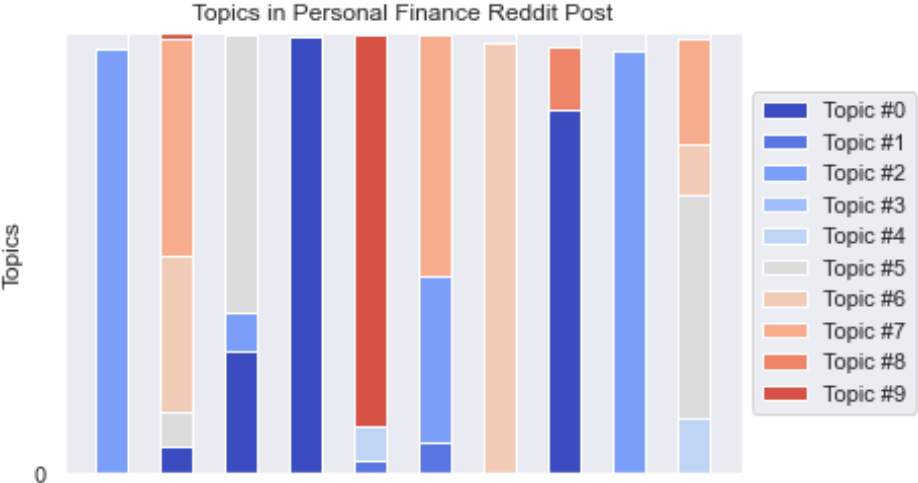


Cosine

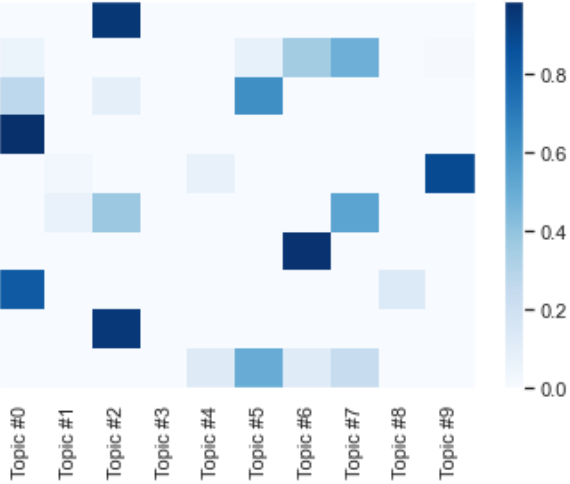


	k-means	Ward
Homogeneity	0.504	0.304
Completeness	0.428	0.268
V-measure	0.463	0.285
Adjusted Rand Index	0.513	0.367

Topic Modeling



he same when it comes to cutting ties with you.]
alty speed upgrades" without customer consent']
company), here's how to get your money back.]
: sentenced for up to 20 yrs, 32 in India indicted']
heet for people who don't know how to budget!"]
pending Money on Food! -- BUY A CROCKPOT']
vay Your Salary Requirement in a Job Interview']
their previously free personal savings account.]
u weren't already planning on buying the item.]
i than it would to give you a raise. So ask firmly.]"



Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
credit	year	money	pay	pay	pay	work	pay	accou nt	credit
pay	loan	pay	car	year	work	accou nt	month	card	pay
accou nt	know	need	tell	car	year	year	accou nt	money	card
edit	work	edit	ask	money	time	pay	day	credit	year
know	time	year	work	work	job	time	know	pay	like
year	pay	bank	say	cost	compa ny	card	time	bank	money
loan	want	car	time	say	think	tell	people	financ e	accou nt
bank	month	like	money	need	money	try	edit	year	want
job	money	work	year	payme nt	say	thank	money	spend	time
money	payme nt	say	month	incom e	thank	say	loan	time	month

Dynamic Topic Modeling

Topic 1

2015	2016	2017	2018	2019	2021
=	=	job	job	job	job
job	job	=	=	=	=
people	people	people	people	people	people
work	work	work	time	time	time
time	time	time	work	work	work
ask	ask	ask	ask	ask	ask
company	company	company	like	offer	offer
like	like	like	company	like	company
offer	offer	offer	offer	company	like
know	know	know	know	know	know

Topic 3

2015	2016	2017	2018	2019	2021
delete	delete	delete	delete	delete	delete
fund	fund	fund	remove	remove	remove
remove	remove	remove	fund	fund	fund
stock	stock	stock	stock	stock	stock
sell	sell	sell	sell	sell	sell
buy	buy	buy	buy	buy	buy
market	market	market	market	market	market
share	share	share	share	vanguard	vanguard
vanguard	vanguard	vanguard	vanguard	share	share
portfolio	portfolio	portfolio	portfolio	portfolio	portfolio

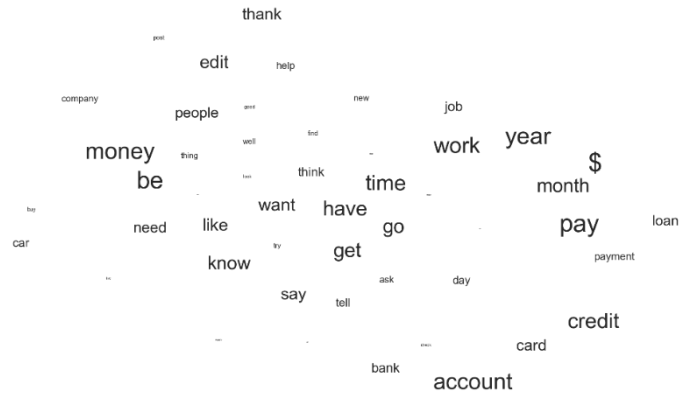
Topic 2

2015	2016	2017	2018	2019	2021
\$	\$	\$	\$	\$	\$
pay	pay	pay	pay	pay	pay
loan	loan	loan	loan	loan	car
year	year	year	year	car	year
month	month	month	month	year	loan
car	car	car	car	month	month
debt	debt	debt	debt	debt	debt
payment	payment	payment	payment	work	work
work	work	work	work	payment	payment
house	live	live	live	get	get

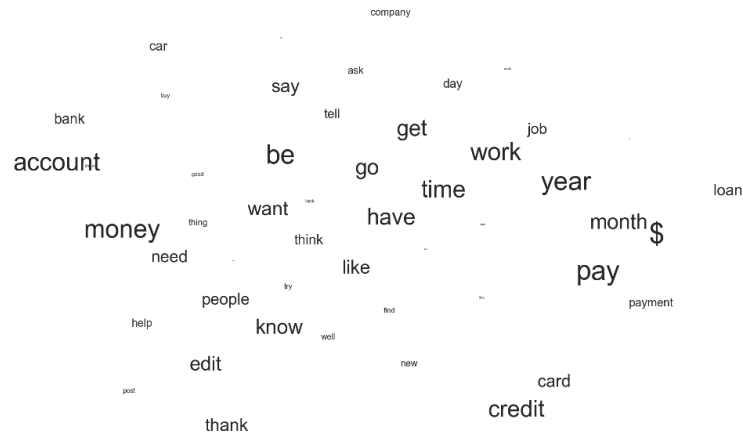
Topic 4

2015	2016	2017	2018	2019	2021
\$	\$	\$	\$	\$	\$
year	year	year	year	year	year
money	money	account	account	account	account
account	account	money	money	money	money
tax	tax	tax	tax	tax	tax
saving	saving	saving	ira	ira	saving
401k	ira	ira	401k	401k	401k
income	401k	401k	saving	saving	ira
ira	income	income	invest	invest	invest
invest	invest	invest	roth	roth	roth

Word Embedding: word2vec



CBOW



Skip gram

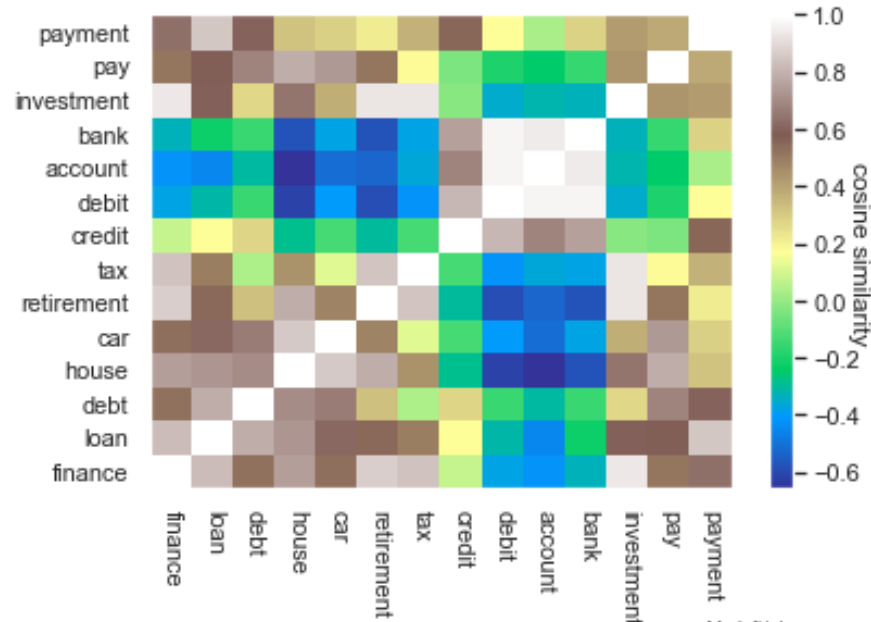
Most similar word to “loan”

CBOW	SG
(pay, 0.9886571168899536)	(student, 0.947920560836792)
(interest, 0.986750602722168)	(forgiveness, 0.9250237345695496)
(payment, 0.9842690229415894)	(program, 0.9044044613838196)
(student, 0.982388973236084)	(borrower, 0.8913569450378418)
(month, 0.9769452810287476)	(qualify, 0.8813320398330688)
(\$, 0.9758726358413696)	(graduate, 0.8758156299591064)
(debt, 0.975852370262146)	(forgive, 0.8746172189712524)
(rate, 0.9507907629013062)	(consolidate, 0.8732489347457886)
(year, 0.9503719806671143)	(repay, 0.8646396398544312)
(car, 0.9470028281211853)	(discharge, 0.8627040982246399)

Most similar word to “house”

CBOW	SG
(home, 0.9981556534767151)	(cheap, 0.9140413999557495)
(make, 0.9973364472389221)	(home, 0.90831458568573)
(buy, 0.9973288774490356)	(own, 0.8955202102661133)
(expense, 0.9969913959503174)	(sell, 0.893168032169342)
(college, 0.9969626665115356)	(rent, 0.8855429887771606)
(old, 0.9969006180763245)	(buy, 0.8843915462493896)
(plan, 0.996694803237915)	(nice, 0.8827567100524902)
(school, 0.9964942336082458)	(clothe, 0.879166841506958)
(price, 0.9963560700416565)	(car, 0.8713674545288086)
(cost, 0.9963399171829224)	(reliable, 0.8700470924377441)

Doc2Vec



Positive cosine similarity:
Payment – credit, tax, debt

Negatively cosine similarity:
Retirement – turning 18

Reddit is advertising really bad IRA advice. Please do not follow their suggested adds.

My lyft/uber expenses were nearly \$750 every month so I instead bought a bike for \$900 and now my travel expenses are <\$140

This article perfectly shows how Uber and Lyft are taking advantage of drivers that don't understand the real costs of the business.

: The federal government just filed suit against Navient, claiming they scammed millions of borrowers between 2010-2015 to the tune of \$4 billion. This is huge.

Your parents took decades to furnish their house

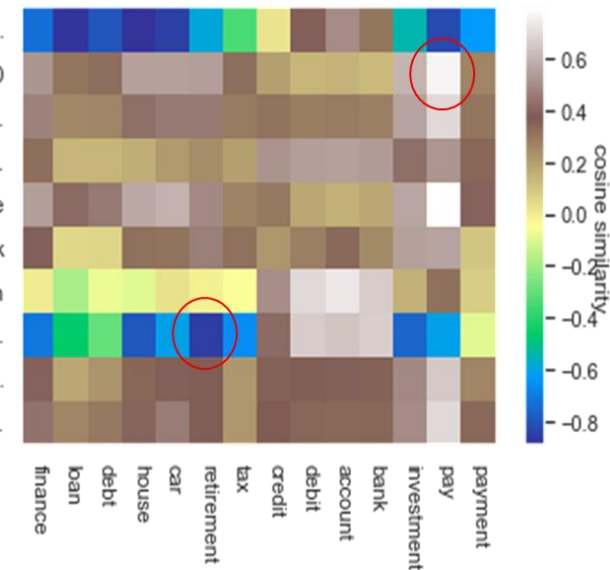
Don't look at your 401k balances today; it has no impact on your long-term investment outlook

If you start suddenly getting email/spam "bombed" there's probably a reason

To all the graduating high school seniors and those turning 18 - Get a bank account that's only in your name.

Please, for the sake of your wallet, meal prep.

If you can't get your emergency fund to grow because of emergencies that keep coming up, you're still doing a good job.



Projection

retirement: 'retirement', '401k', 'ira', 'young', 'old', 'plan'

Dimensions					
family	'mom', 'parent', 'father', 'dad', 'mother'	'son', 'daughter', 'child'	ira 401k young old retirement	family	plan old retirement 401k ira
	'loan', 'debt', 'poor', 'unemployed'	'saving', 'investment', 'rich', 'job'	plan		young

Projection

retirement: 'retirement', '401k', 'ira', 'young', 'old', 'plan'

Dimensions					
family	'mom', 'parent', 'father', 'dad', 'mother'	'son', 'daughter', 'child'	ira 401k young old retirement	family	plan old retirement 401k ira
	'loan', 'debt', 'poor', 'unemployed'	'saving', 'investment', 'rich', 'job'	plan		young