

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one partially covering the green one.

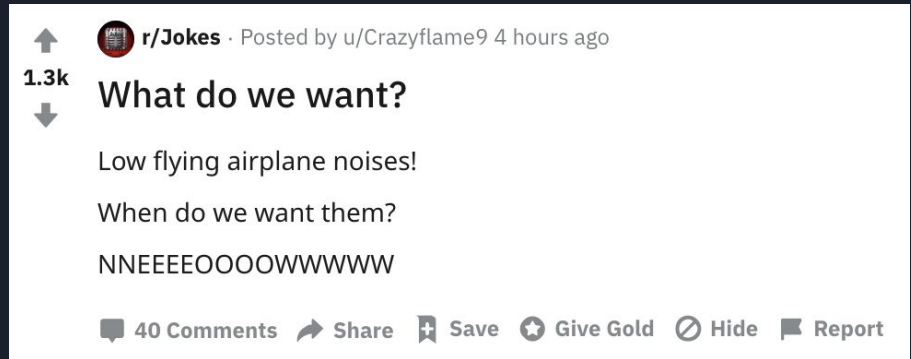
Classifying Reddit Posts

Charley Dixon

Problem Statement

“What characteristics of a post on Reddit contribute most to what subreddit it belongs to?”

When you throw out the packaging of a microwave dinner and immediately forget how long to microwave it for



Reddit API

```
def get_reddit_posts(sub_list, website = 'reddit.com'): # pass in list of subreddits
    headers = {'User-agent': 'Chuck Bob MAC 5.5'}
    posts = [] # list of dicts for each feature in a reddit post
    for subreddit in sub_list:
        url = 'http://www.{}/r/{}/.json'.format(website, subreddit)
        after = None # each subreddit must start w 'after = None' to start retrieving most recent
        # start range at 1 to make status message work properly (can't multiply by 0)
        for i in range(1, 41): # to get maximum of 1000 posts
            if after == None:
                params = {}
            else:
                params = {'after': after} # prevents duplication of get requests
            res = requests.get(url, params = params, headers = headers)
            if res.status_code == 200:
                json = res.json()
                # create a list of dicts for each post
                posts.extend([post['data'] for post in json['data']['children']]) # json['data']
                after = json['data']['after']
            else:
                print('Oops... something went wrong! Status Code: {}'.format(res.status_code))
                break
        if i != 40:
            print('For subreddit, {}, about {} percent of posts have been retrieved'.format(subreddit, (i-1)*2.5))
        else:
            print('Subreddit {} has been retrieved'.format(subreddit))

        # SAVE DATA TO CSV AFTER EACH API CALL
        pd.DataFrame(posts).to_csv('posts.csv', index = False)
        time.sleep(3)
    return posts
```

Subreddits



r/personalfinance

13.2m

Subscribers

10.0k

Online

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!



r/financialindependence

421k

Subscribers

1.7k

Online

This is a place for people who are or want to become Financially Independent (FI), which means not having to work for money. Financial Independence is closely related to the concept of Early Retirement/Retiring Early (RE) - quitting your job/career and pursuing other activities with your time. At its core, FI/RE is about maximizing your savings rate (through less spending and/or higher income) to achieve FI and have the freedom to RE as fast as possible.

Typical Financial Independence Post



114



r/financialindependence · Posted by u/mathmagician3000 22 hours ago

I've only been working for one year and I'm already thinking about FIRE. Is that weird?

While I didn't know FIRE was a thing until recently, I actually came up with the idea just one month into my first job. I realized that with enough money, I can basically live off of investment profits forever. With some quick calculations, I found \$1 million to be the sweet spot, which I can reach in 10-15 years. I thought it was crazy that I barely begun working and am already thinking about early retiring at the age of 40. I never told anybody about my plans. But apparently I'm not alone. Nice to meet you all.



68 Comments



Share



Save



Give Gold



Hide



Report

79% Upvoted

Typical Personal Finance Post

↑
1.9k
↓

 **r/personalfinance** · Posted by u/Notelephone122 1 day ago

Retirement **Why is the Roth IRA so special?**

Hello,

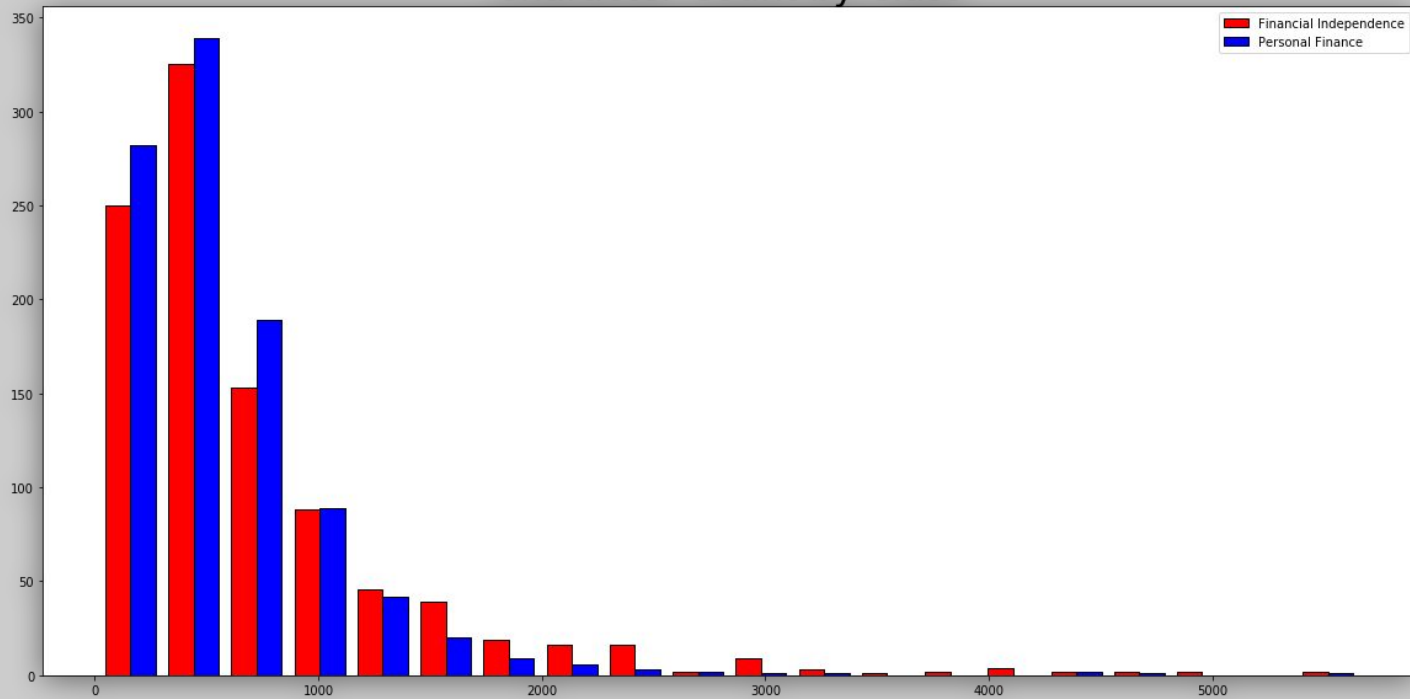
I am having trouble grasping the idea that the Roth IRA is so special. I know that the Roth IRA is taxed as you contribute money into and not taxed when you take the money out, but why wouldn't you just skip that step and just invest the money in stocks in a brokerage account and pay the ~15% capital gains tax instead of paying the ~24% income tax on the Roth IRA contributions? I know I am missing something. Is it because if you invest the money separately from a Roth IRA you are paying the income tax and the capital gains tax? Can someone please explain to me what I am not getting. Thank you.

 **554 Comments**  **Share**  **Save**  **Give Gold**  **Hide**  **Report**

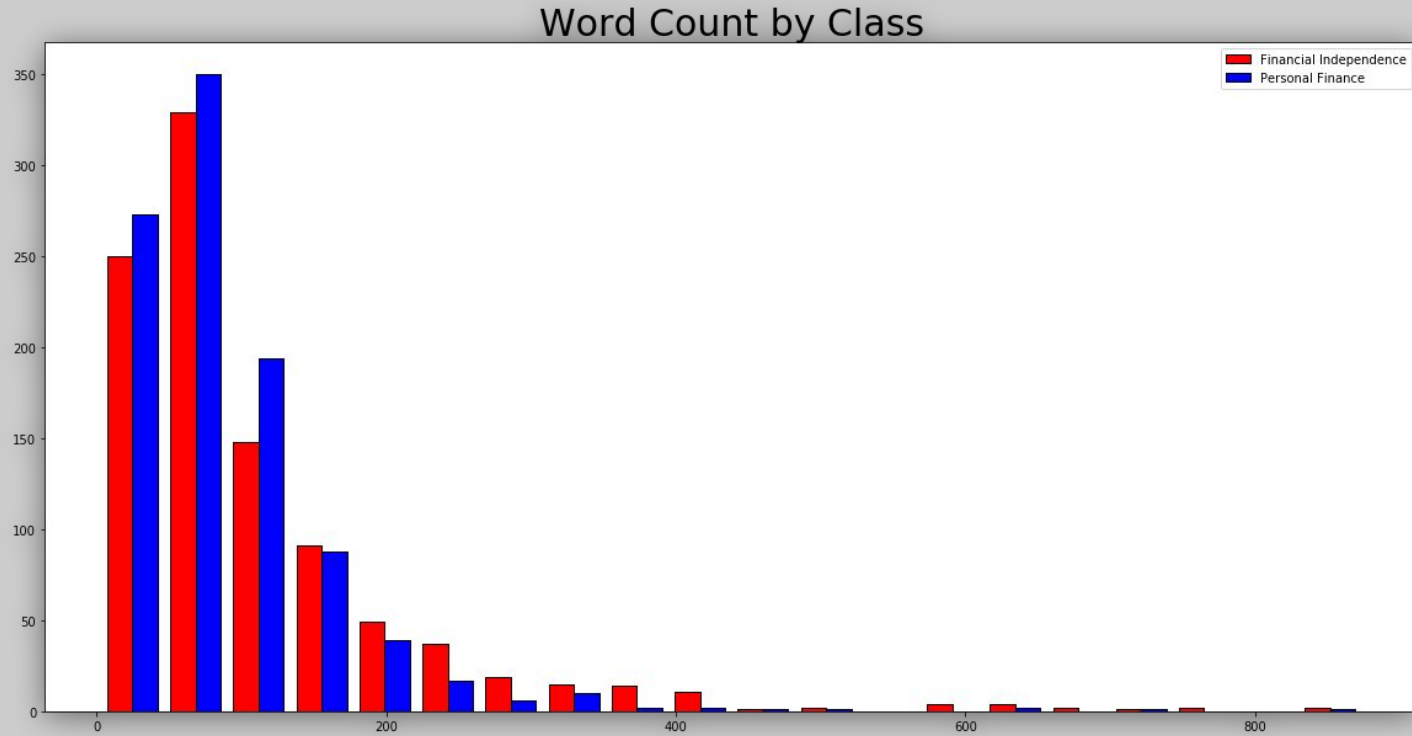
92% Upvoted

Exploratory Analysis

Character Count by Class



Exploratory Analysis





Count Vectorizer

CREDIT

CARD

WORK

LOAN

POST

RETIREMENT

CAR

INCOME

TERM	FINANCIAL INDEPENDENCE	PERSONAL FINANCE	DIFFERENCE
credit	42	808	766
fire	517	5	512
year	1383	879	504
card	40	504	464
fi	383	1	382
work	600	226	374
loan	148	507	359
post	399	63	336
account	285	540	255
retirement	393	138	255
car	156	410	254
income	496	249	247
life	358	116	242
payment	118	355	237
time	574	354	220
people	306	89	217
debt	157	368	211
pay	304	510	206
like	589	383	206
http	250	52	198



Modeling

- Models
 - Multinomial NB
 - Random Forest Classifier
- Methods
 - Pipeline
 - GridSearchCV
- Evaluation

		<i>Actual</i>	
		FI (0)	PF (1)
<i>Predictions</i>	FI (0)	232	21
	PF (1)	21	220

Confusion Matrix