

```
In [136... import praw
import pandas as pd
```

```
In [137... reddit = praw.Reddit(
    client_id='#',
    client_secret='#',
    user_agent='assignment',
)
```

```
In [138... reddit.read_only
```

```
Out[138]: True
```

Collecting post and comments from r/python

```
In [139... posts = reddit.subreddit('python').top(time_filter='week', limit=10)
```

```
In [142... from praw.models import MoreComments

post_list, comment_list = [], []

def fetch_comments(comments_obj, post_id):

    q = []
    for i, comment in enumerate(comments_obj):
        if i > 20:
            break
        if isinstance(comment, MoreComments):
            continue

        q.append((comment, 0))

    while(q):
        comment, lvl = q.pop()

        comment_list.append({
            'post_id': post_id,
            'comment_id': comment.id,
            'parent_id': comment.parent_id,
            'body': comment.body,
            'created_on': comment.created_utc,
            'upvotes': comment.score,
            'author_id': comment.author.id if comment.author and hasattr(c
            'author_name': comment.author.name if comment.author and hasat

        })

        if lvl > 0:
            continue

        for i, neigh_comment in enumerate(comment.replies):
            if isinstance(neigh_comment, MoreComments):
                continue
            if i > 10:
```

```

        break
    q.append((neigh_comment, lvl+1))

def fetch_posts(subreddit, post_types = ['rising', 'hot', 'top week'], limit=
visited = set()
for post_type in post_types:
    if len(post_type.split(' ')) == 2 and post_type.split(' ')[0] == 'top
        posts = reddit.subreddit(subreddit).top(time_filter=post_type
    elif post_type == 'top':
        posts = reddit.subreddit(subreddit).top(time_filter='week', li
    elif post_type == 'rising':
        posts = reddit.subreddit(subreddit).rising(limit=limit)
    elif post_type == 'hot':
        posts = reddit.subreddit(subreddit).hot(limit=limit)

    for post in posts:
        if post.id in visited:
            continue
        visited.add(post.id)
        post_list.append({
            'id': post.id,
            'author': post.author.name if post.author and hasattr(post
            'author_id': post.author.id if post.author and hasattr(pos
            'total_comments': post.num_comments,
            'upvote': post.score,
            'post_type': post_type,
            'title': post.title,
            'downvote': post.score*(1-post.upvote_ratio),
            'url': post.url,
            'created_on': post.created_utc,
            'subreddit': subreddit,
            'subreddit_id': post.subreddit_id
        })

        fetch_comments(post.comments, post.id)

```

In [143... `fetch_posts('python', ['top'])`

In [144... `post_list`

```

Out[144]: [{ 'id': '12glkw4',
  'author': 'TheBodyPolitic1',
  'author_id': 'von3w6y2',
  'total_comments': 319,
  'upvote': 588,
  'post_type': 'top',
  'title': "Why didn't Python become popular until long after its creati
on?",
  'downvote': 35.280000000000003,
  'url': 'https://www.reddit.com/r/Python/comments/12glkw4/why_didnt_pyth
hon_become_popular_until_long_after/',
  'created_on': 1681051909.0,
  'subreddit': 'python',
  'subreddit_id': 't5_2qh0y'},
{ 'id': '12hj9oc',
  'author': 'MetonymyQT',
  'author_id': 'shnqm',
  'total_comments': 73,
  'upvote': 510,
  'post_type': 'top',
  'title': 'Free course: Build a modern API with FastAPI and Python',
  'downvote': 25.500000000000002,
  'url': 'https://www.reddit.com/r/Python/comments/12hj9oc/free_course_b
uild_a_modern_api_with_fastapi_and/',
  'created_on': 1681134311.0,
  'subreddit': 'python',
  'subreddit_id': 't5_2qh0y'},
{ 'id': '12egsoz',
  'author': '2broke2code',
  'author_id': 'rs4dqilj',
  'total_comments': 105,
  'upvote': 465,
  'post_type': 'top',
  'title': "I trained a RoastBot on >120,000 faces and >0.5 million comm
ents and it's a menace 🐱.",
  'downvote': 32.549999999999997,
  'url': 'https://www.reddit.com/r/Python/comments/12egsoz/i_trained_a_r
oastbot_on_120000_faces_and_05/',
  'created_on': 1680863778.0,
  'subreddit': 'python',
  'subreddit_id': 't5_2qh0y'},
{ 'id': '12ffsif',
  'author': 'midnitte',
  'author_id': '3gad9',
  'total_comments': 64,
  'upvote': 385,
  'post_type': 'top',
  'title': 'EP 684: A Per-Interpreter GIL Accepted',
  'downvote': 11.550000000000001,
  'url': 'https://discuss.python.org/t/pep-684-a-per-interpreter-gil/195
83/42',
  'created_on': 1680942397.0,
  'subreddit': 'python',
  'subreddit_id': 't5_2qh0y'},
{ 'id': '12fzdu2',
  'author': 'aeluro1',
  'author_id': '88efmodhn',
  'total_comments': 12,
  'upvote': 367,
  'post_type': 'top',

```

rious number of reasons, though more recently it seems that they've become another small part of [the pile](https://pile.eleuther.ai), a dataset that quite a few language models are trained on.",

```
'created_on': 1680646662.0,
'upvotes': 20,
'author_id': '8um8szao',
'author_name': 'pointmetoyourmemory'},
{'post_id': '12bl2uj',
'comment_id': 'jeywttj',
'parent_id': 't3_12bl2uj',
'body': 'your github link is not working as of 1612 central time\n\nEdit: 1900 central. Still getting 404\n\nEdit: this works -> https://github.com/SuperflowsAI/enron-sentiment-analysis\n\nCredit to u/ShadowDocket',
'created_on': 1680642784.0,
'upvotes': 22,
'author_id': '89sylldx',
'author_name': 'WhyDoIHaveAnAccount9'},
{'post_id': '12bl2uj',
'comment_id': 'jfv8f8',
'parent_id': 't1_jeywttj',
'body': 'Whoops! Sorry about that. Looks like you found the right one!\n\nAccidently sent the private repo I was working from. Have edited the post to update this',
'created_on': 1680678914.0,
'upvotes': 2,
'author_id': 'alzl8hco',
'author_name': '0k-Craft-9908'},
{'post_id': '12bl2uj',
'comment_id': 'jfv0txb',
'parent_id': 't1_jeywttj',
'body': '[deleted]',
'created_on': 1680660368.0,
'upvotes': 8,
'author_id': 'deleted',
'author_name': 'deleted'}]
```

```
In [146... from tinydb import TinyDB

post_db, comment_db = TinyDB('post.json'), TinyDB('comment.json')
```

```
In [147... post_db.insert_multiple(post_list)
comment_db.insert_multiple(comment_list)
```