

Process Users

1. How many tweets did I sample in [73]? Seeing how these tweets were used, why such a large sample?
2. Why did I do this: `tweets = list(cursor_sample)`?
3. What is `all_keys` — type and substance?
4. What is `keep` — type and substance?
5. Why did I do this: `users = users[keep].copy()`? Why the copy?
6. Consider this:

```
cursor_sample = collection_reference.aggregate([
    sample,
    filter_on_unprocessed
])
```

Why would it be important to sample then filter as opposed to filter then sample? What complications arise if I sample then filter? How was this addressed in the definition of `get_n_tweets`?

7. Originally, I wrote `get_n_tweets` to use `id` instead of `_id`. It was about ten times slower. Why?
8. How does `process_users` scale with `n`?