



# Basic Reddit Analysis with PRAW

PRANAV SALUNKE

# Why this and what were the goals?



- ▶ Reddit is a very widely used website and I have always wondered about the usage of it
- ▶ I had begun thinking about this a while ago
- ▶ Goals:  
wanted to learn:
  - ▶ When the most posts were made
  - ▶ Which subreddits gets the most posts
  - ▶ Which users post the most often
  - ▶ If there are many reposts? (same title posted multiple times)
  - ▶ What amount of Karma (Reddit points) do users have
- ▶ Ran the code over multiple days on a raspberry pi and collected over 100k posts in total
- ▶ Unexpected finding: I was able to find bot accounts by the frequency at which they posted

# APIs/modules/libraries



- ▶ `praw` (Python Reddit API Wrapper)
  - ▶ The most important, the key to scraping Reddit posts
- ▶ Saving and reading data
  - ▶ Saving data as csv by writing to files
  - ▶ `csv` module to read it, `pandas` to work with it
- ▶ Visuals
  - ▶ `matplotlib`, `bokeh`
- ▶ Misc.
  - ▶ `os`, `sys`, `time`, `datetime`, `traceback`, `pickle`



# praw commands

other important ones will be mentioned while going over the code

- ▶ `praw.Reddit()` to create the reddit instance. Everything uses this
- ▶ `reddit.subreddit(sub).new(limit=n)` get `n` posts from a subreddit (`sub`)
- ▶ `submission` object (called `post` in this example)
  - ▶ `Post.id` get the post's unique id
  - ▶ `post.subreddit.display_name` get the subreddit name the post was submitted to
  - ▶ `post.title` get the post's title
  - ▶ `post.created_utc` get the time the post was created in UTC
  - ▶ `post.author` get the user the post was made by
  - ▶ `post.author.link_karma`, `post.author.comment_karma` get the user's karma amount