

Week 7 Assignment

This week we are collecting some data from Reddit and doing some basic EDA on it. You should:

- create your Reddit account and API keys
- collect data from a subreddit of your choice
 - at a minimum, collect the posts from the subreddit; optionally collect comments on the posts
- save the data to a SQLite3 database
- perform some basic EDA on the data
 - create at least 2 plots
- write a short analysis at the end describing the process and results
- turn in the Jupyter Notebook and PDF printout or export to the week 7 dropbox

Optional advanced section

- Practice SQL queries and select a subsection of the posts you collected
- Modify your code to collect data beyond the 1000 item limit
- Collected comments from the posts for analysis next week and do some EDA on the comments (e.g. who is the top commenter, which commenters have the most up and down votes or most controversial posts, etc)
- examine n-grams (bigrams, trigrams) or collocations

Note: There is no solution file for this week.

```
In [1]: import praw
import pandas as pd
import credentials
import matplotlib.pyplot as plt
import seaborn as sns
```

Creating subreddit account

```
In [2]: reddit = praw.Reddit(client_id=credentials.client_id,
                             client_secret=credentials.client_secret,
                             user_agent=credentials.user_agent)
```

Collecting data from Subreddit

```
In [3]: co_subreddit = reddit.subreddit('snowfall').hot(limit=10)

In [4]: for post in co_subreddit:
         print(post.title, post.num_comments)
```

Snowfall S06xE10 | Sins of the Father | Episode Discussion 4896
I couldn't stand Louie 18
Skyzoo brilliant Snowfall based album is now Deluxe! Peep cover is changed to the now Alcoholic homeless Franklin 2
Franklin Saint's downfall wasn't his fault. 78
Holy fuck 39
What exactly would have happened if Franklin had someone anonymously call the cops.. 7
Franklin is Jord shiva 7
Don't let Teddy see this 33
Why was season 1 so sorry and mid? It's hard to put this show onto a friend 29
Franklin offscreen on the plane back to LA, in his head, after he left his cane in Mel classroom at the end of S4 2

```
In [5]: reddit_data = {'title': [],
                      'link': [],
                      'author': [],
                      'n_comments': [],
                      'score': [],
                      'text': []}

co_subreddit = reddit.subreddit('snowfall').hot(limit=None)

for post in list(co_subreddit):
    reddit_data['title'].append(post.title)
    reddit_data['link'].append(post.permalink)
    if post.author is None:
        reddit_data['author'].append('')
    else:
        reddit_data['author'].append(post.author.name)

    reddit_data['n_comments'].append(post.num_comments)
    reddit_data['score'].append(post.score)
    reddit_data['text'].append(post.selftext)
```

```
In [6]: co_df = pd.DataFrame(reddit_data)

In [7]: co_df
```

```
Out[7]:
```

		title	link	author	n_comments	score	text
0	Snowfall S06xE10 Sins of the Father Episod...	/r/SnowFall/comments/12sdjgn/snowfall_s06xe10_...	md28usmc	4896	487	The discussion/comments below assume you have ...	
1	I couldn't stand Louie	/r/SnowFall/comments/1b77j4f/i_couldnt_stand_l...	2wavyGuyy	18	24	I just finished Snowfall and man I was heated ...	
2	Skyzoo brilliant Snowfall based album is now D...	/r/SnowFall/comments/1b77nny/skyzoo_brilliant_...	smashadamspel	2	9		
3	Franklin Saint's downfall wasn't his fault.	/r/SnowFall/comments/1b6u47/franklin_saints_d...	Xaixiu	78	24	Franklin would've been a top player and CEO fo...	
4	Holy fuck	/r/SnowFall/comments/1b6e7uu/holy_fuck/	Expert_Dog_9069	39	132	just finished the show, yes i'm a bit late \n...	
...	
994	Saw it on TikTok	/r/SnowFall/comments/15w407/saw_it_on_tiktok/	compromisedplot	8	15	the cinematography is insane for some no brand...	
995	Learn the best methods for CPNs, EDD, and more...	/r/SnowFall/comments/15vukv/learn_the_best_me...	OrdinaryOutrageous40	0	20		
996	¿Que opinás de la mañana de la gran nevada en ...	/r/SnowFall/comments/15vwbh2/que_opinás_de_la_...	ushuaiafindelmundo	1	3		
997	I asked Michael Hyatt who plays Cissy Saint if...	/r/SnowFall/comments/15utaycl/asked_michael_h...	tfengbrah	14	55	Kill Teddy regardless of what he said and she ...	
998	Markice Moore aka ray ray thru the years	/r/SnowFall/comments/15ujpdp/markice_moore_aka...	Particular_Minute_67	20	122	From his insta https://www.instagram.com/reel/...	

999 rows × 6 columns

Saving data to SQLite3

```
In [8]: import sqlite3

con = sqlite3.connect("co_reddit.sqlite")
co_df.to_sql('posts', con, if_exists='replace', index=False)
```

```
Out[8]: 999

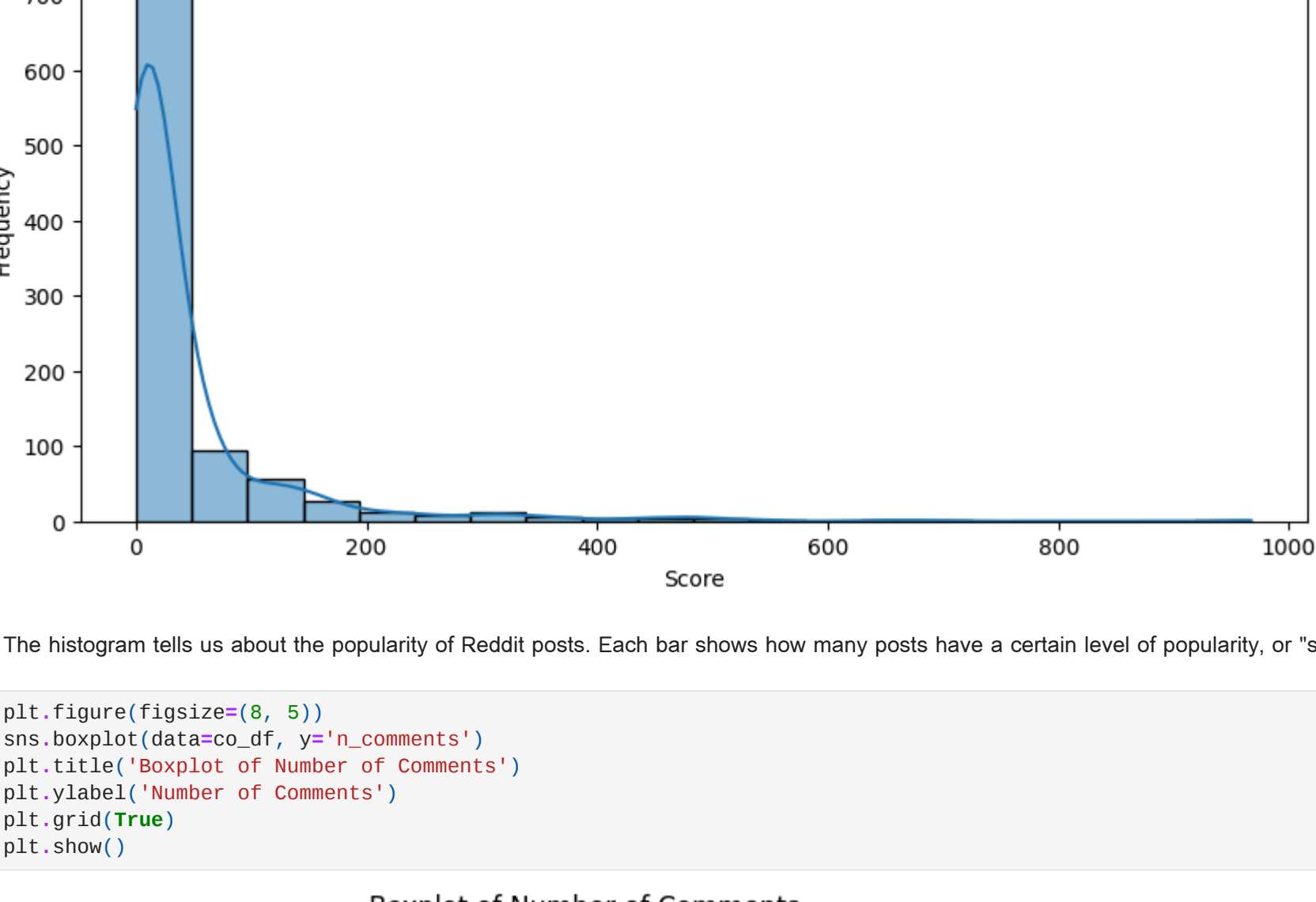
In [9]: co_df_check = pd.read_sql_query('SELECT * FROM posts;', con)
con.close()
co_df_check
```

```
Out[9]:
```

		title	link	author	n_comments	score	text
0	Snowfall S06xE10 Sins of the Father Episod...	/r/SnowFall/comments/12sdjgn/snowfall_s06xe10_...	md28usmc	4896	487	The discussion/comments below assume you have ...	
1	I couldn't stand Louie	/r/SnowFall/comments/1b77j4f/i_couldnt_stand_l...	2wavyGuyy	18	24	I just finished Snowfall and man I was heated ...	
2	Skyzoo brilliant Snowfall based album is now D...	/r/SnowFall/comments/1b77nny/skyzoo_brilliant_...	smashadamspel	2	9		
3	Franklin Saint's downfall wasn't his fault.	/r/SnowFall/comments/1b6u47/franklin_saints_d...	Xaixiu	78	24	Franklin would've been a top player and CEO fo...	
4	Holy fuck	/r/SnowFall/comments/1b6e7uu/holy_fuck/	Expert_Dog_9069	39	132	just finished the show, yes i'm a bit late \n...	
...	
994	Saw it on TikTok	/r/SnowFall/comments/15w407/saw_it_on_tiktok/	compromisedplot	8	15	the cinematography is insane for some no brand...	
995	Learn the best methods for CPNs, EDD, and more...	/r/SnowFall/comments/15vukv/learn_the_best_me...	OrdinaryOutrageous40	0	20		
996	¿Que opinás de la mañana de la gran nevada en ...	/r/SnowFall/comments/15vwbh2/que_opinás_de_la_...	ushuaiafindelmundo	1	3		
997	I asked Michael Hyatt who plays Cissy Saint if...	/r/SnowFall/comments/15utaycl/asked_michael_h...	tfengbrah	14	55	Kill Teddy regardless of what he said and she ...	
998	Markice Moore aka ray ray thru the years	/r/SnowFall/comments/15ujpdp/markice_moore_aka...	Particular_Minute_67	20	122	From his insta https://www.instagram.com/reel/...	

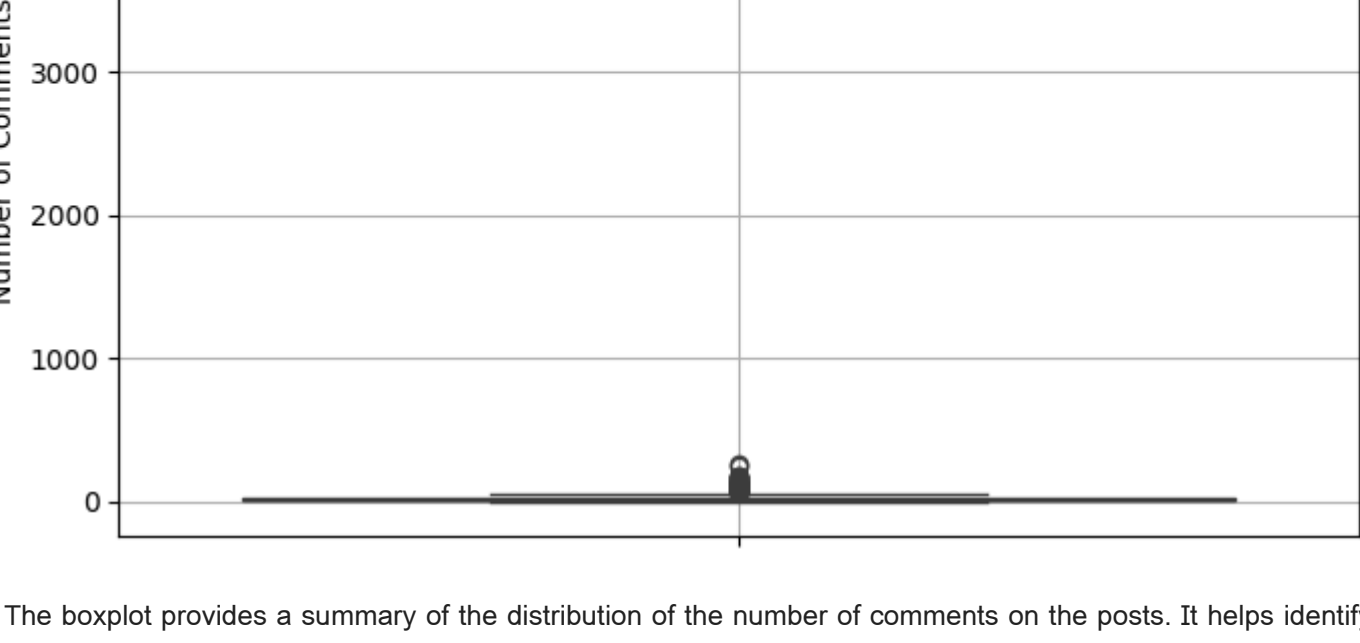
999 rows × 6 columns

```
In [10]: plt.figure(figsize=(10, 5))
sns.histplot(data=co_df, x='score', bins=20, kde=True)
plt.title('Distribution of Scores')
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.show()
```



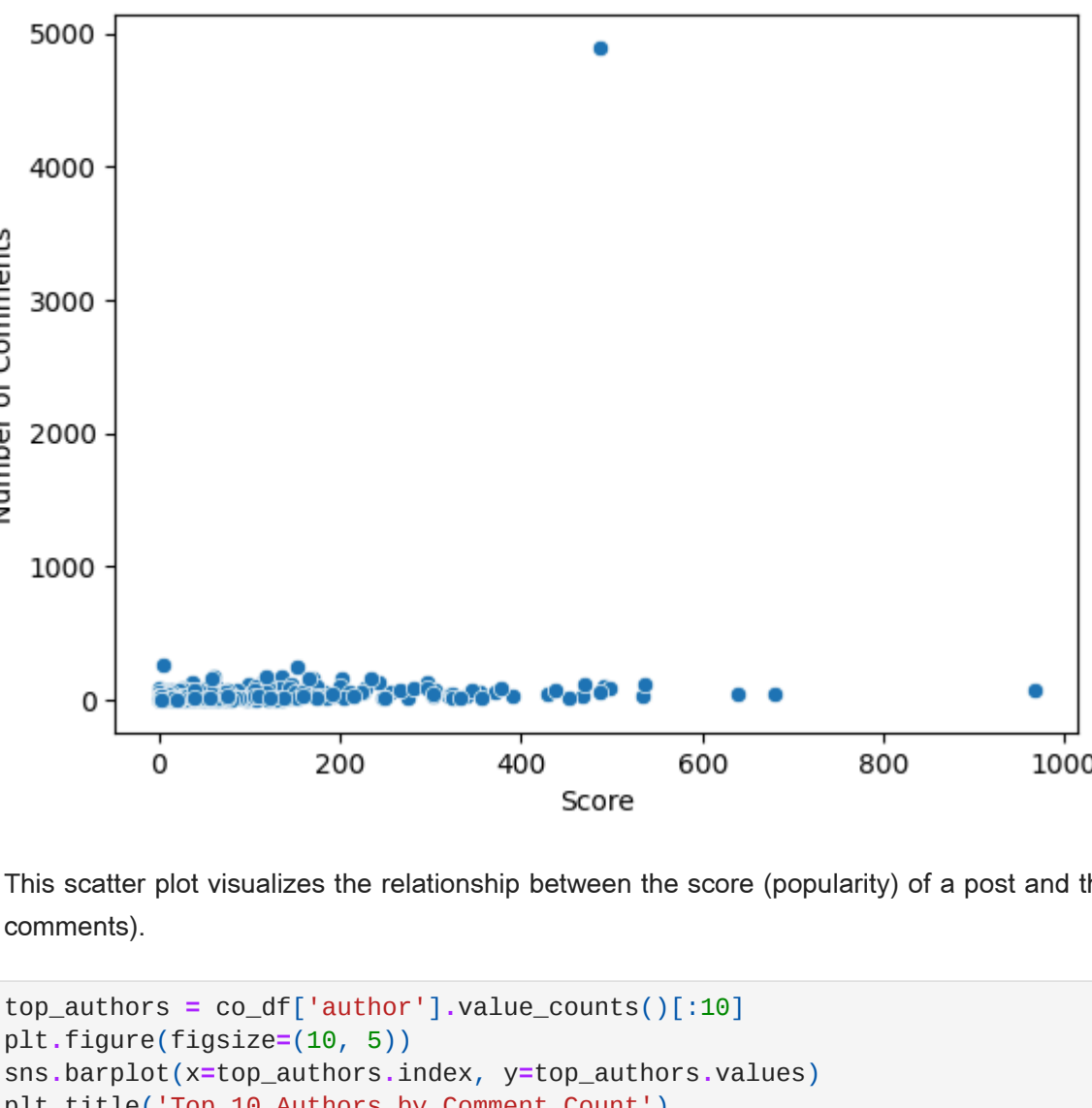
The histogram tells us about the popularity of Reddit posts. Each bar shows how many posts have a certain level of popularity, or "score." This helps us see if most posts are very popular, not so popular, or somewhere in between.

```
In [11]: plt.figure(figsize=(8, 5))
sns.boxplot(data=co_df, y='n_comments')
plt.title('Boxplot of Number of Comments')
plt.ylabel('Number of Comments')
plt.grid(True)
plt.show()
```



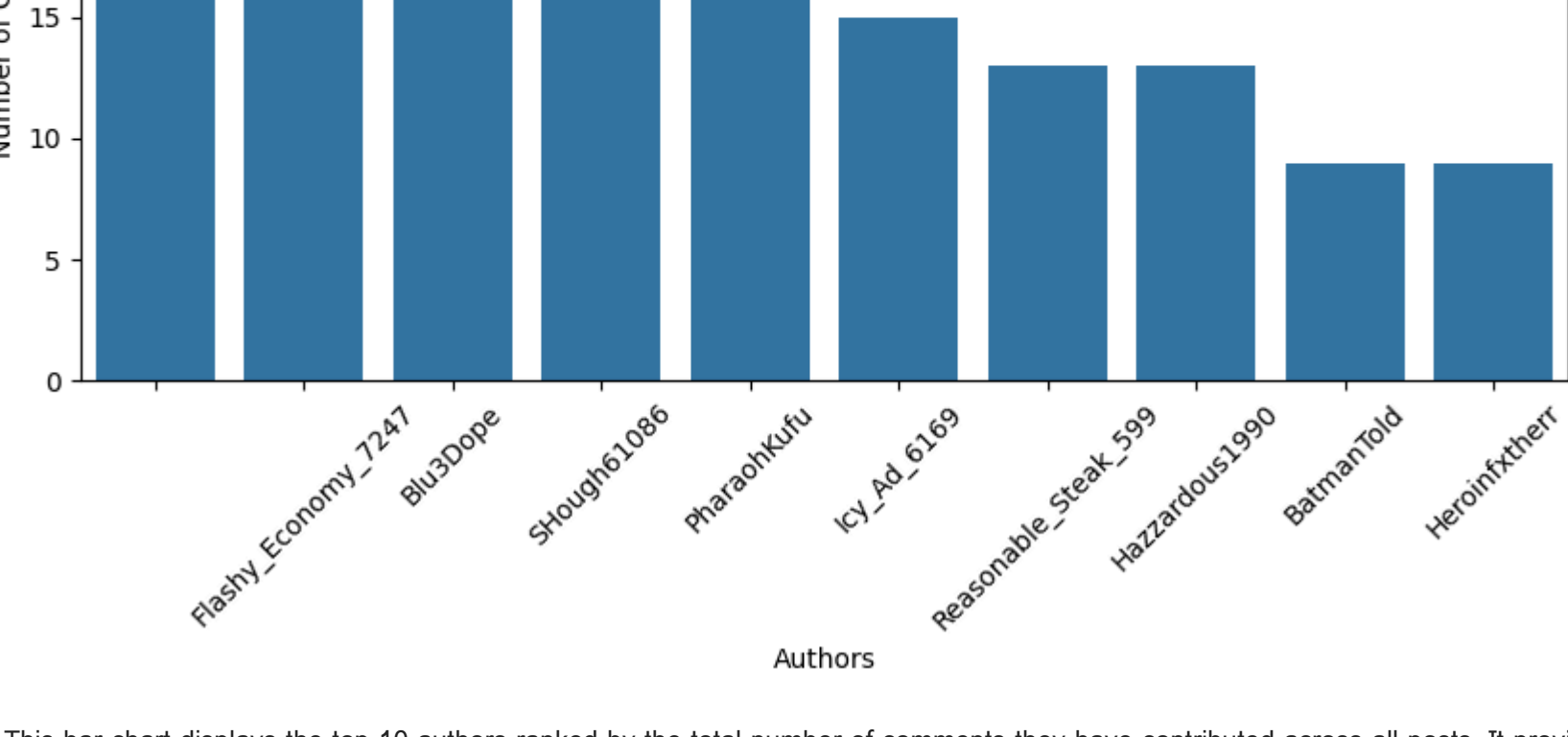
The boxplot provides a summary of the distribution of the number of comments on the posts. It helps identify the median, quartiles, and outliers in the comment counts. This visualization aids in understanding the spread and variability in the number of comments across the posts.

```
In [12]: sns.scatterplot(data=co_df, x='score', y='n_comments')
plt.title('Score vs. Number of Comments')
plt.xlabel('Score')
plt.ylabel('Number of Comments')
plt.show()
```



This scatter plot visualizes the relationship between the score (popularity) of a post and the number of comments it has received. It helps in analyzing if there's any correlation between a post's score and the level of discussion it generates (number of comments).

```
In [13]: top_authors = co_df['author'].value_counts()[:10]
plt.figure(figsize=(10, 5))
sns.barplot(x=top_authors.index, y=top_authors.values)
plt.title('Top 10 Authors by Comment Count')
plt.xlabel('Authors')
plt.ylabel('Number of Comments')
plt.xticks(rotation=45)
plt.show()
```



This bar chart displays the top 10 authors ranked by the total number of comments they have contributed across all posts. It provides insights into which authors are most active in the subreddit, contributing to discussions and interactions.

```
In [14]: co_df[co_df['author'] == '']

Out[14]:
```

		title	link	author	n_comments	score	text
355	S1 Onwards Question...	/r/SnowFall/comments/190t8gu/s1_onwards_question/		8	7	Just finished Season 1, loved it apart from th...	
642	How y'all think the story would have gone if F...	/r/SnowFall/comments/17jnp6/show_yall_think_th...		11	44		
678	Idc what noone says. Season 4 was PEAK snowfall.	/r/SnowFall/comments/17cfcy/dc_what_noone_sa...		30	45	I never understood why people act like It's th...	
688	Which one of these Characters were written/kill...	/r/SnowFall/comments/17bzms0/which_one_of_thes...		10	2	I think Ima go with Ray Ray on this one. Bring...	
696	The Kane storyline is so stupid	/r/SnowFall/comments/17aburk/the_kane_storyfin...		35	32	When I saw those shots being fired at Franklin...	
698	What's y'all's fave songs from the series?	/r/SnowFall/comments/17a2m3/whats_yalls_fave_...		19	13	Mine is Notorious	
701	Crip vs Crip Showdown PT 1. Who Wins?	/r/SnowFall/comments/17aecd/crip_vs_crip_show...		3	2	ManBoy (Compton) vs Big Deon (PJ Watts)	
706	Which One Of These Characters Should've Had Mo...	/r/SnowFall/comments/17a2m3/which_one_of_thes...		3	2	\n\n[View Poll](https://www.reddit.com/poll/17...	
707	I think Wanda's character was irrelevant and I...	/r/SnowFall/comments/17a6c96/i_think_wandas_ch...		26	4	She didn't really add much to the storyline b...	
709	Idky, but I couldn't take Peaches and her s...	/r/SnowFall/comments/173er1w/ldky_but_i_couldn...		11	12	It felt so weird watching them on there. Like...	
714	Franklin lowkey is my intrusive thoughts manif...	/r/SnowFall/comments/178pmv5/franklin_lowkey_i...		20	17		
717	Skully	/r/SnowFall/comments/178iqh/skully/		9	9	Did they try to make Skully an antagonist, or ...	
724	Season 6 Discussion.	/r/SnowFall/comments/1784ilc/season_6_discussio...		1	1	The scene where all the Snowfall Allstars com...	
746	Cissy and Leon Prison Scene	/r/SnowFall/comments/173x0q/cissy_and_leon_pr...		11	12	[https://www.youtube.com/watch?v=rwq15o7lCOK]...	
754	Alot of people say Kane was a Crip OG but he g...	/r/SnowFall/comments/172j6jv/alot_of_people_sa...		2	1	Most of his goons drove red and gray cars, and...	
817	What if....	/r/SnowFall/comments/16sayqr/what_if/		14	4	They added a little "end of credits" scene wher...	
854	Just finished 6x10	/r/SnowFall/comments/16i3zoh/just_finished_6x10/		6	14	I'm sad now. I'm lowkey too sad to even consid...	
860	I think the Cici hate Is more trendy than Genui...	/r/SnowFall/comments/16k5zql/i_think_the_cici_...		29	29	Now on the official last episode of Snowfall a...	
875	Wait a minute. (06x03)	/r/SnowFall/comments/168607/wait_a_minute_06x03/		5	7	Since when did Franklin hate Deon? Franklin an...	
883	Was Teddy actually surprised about Avi?	/r/SnowFall/comments/16he7bq/was_teddy_actuall...		4	11	When Franklin called Teddy on the payphone and...	
890	Was trying to figure out what didn't sit right...	/r/SnowFall/comments/16gcogr/was_trying_to_fg...		17	0	I made a previous post about how I thought Ted...	
898	Louie vs Franklin	/r/SnowFall/comments/16ex57d/louie_vs_franklin/		3	3	I like how Franklin threatened Louie In 5x3 si...	
920	Jerome (Season 5)	/r/SnowFall/comments/16cf3zi/jerome_season_5/		5	6	Why Is Jerome so unhinged and Angry In Season ...	
938	Did anyone else think the Teddy storyline was ...	/r/SnowFall/comments/16919m/did_anyone_else_t...		23	46	We start off introduced to Teddy as this like ...	
939	Season 4	/r/SnowFall/comments/169fxb/season_4/		6	3	Now on 4x2. How did manboy go from terrified o...	
941	Can't believe I held this show off for so long.	/r/SnowFall/comments/168oec8/cant_believe_i_he...		1	10	Being a fan of breaking bad and better call Sa...	
946	Mel	/r/SnowFall/comments/167ri2t/mel/		5	10	Just finished 3x7. I just have to say....fuck...	
948	what's so appealing about Franklin?	/r/SnowFall/comments/166yqg/whats_so_appealin...		39	14	I mean, I know he's the main character and we'...	
990	Did I miss something at the end of season 2?	/r/SnowFall/comments/15xkquy/did_i_miss_someth...		6	4	So I'm currently on season 3 of this show, so ...	

Summary:

This Python code uses a library called PRAW to gather information from the "snowfall" subreddit on Reddit. It collects various details about the posts there, like their titles, links, authors, how many comments they have, their scores, and what they say. This data is then organized into a table using pandas, a tool for handling data in Python, and saved in a way that makes it easy to analyze later on.

After saving the data, the code creates some pictures, or visualizations, to help us understand what's happening in the subreddit. One picture is like a bar graph that shows how many posts have different scores. Another picture looks like a box and whisker plot, which tells us how many comments each post has and helps us spot any unusual numbers. There's also a scatter plot, which shows if there's a connection between a post's score and how many comments it gets. Finally, there's a bar chart that highlights the top 10 people who have left the most comments in the subreddit.

These pictures are helpful because they give us clues about how active and popular the subreddit is. We can see which posts are getting lots of attention and who's contributing the most to discussions. By saving the data in a special place, like a database, it's easier to ask questions about the posts in the future.

In short, this code helps us peek into what's going on in the "snowfall" subreddit. It uses pictures to show us how many people are talking about different posts and who's leading the conversation. And by saving all this information, we can dig deeper into the subreddit's activity whenever we want.