Total 24 Columns:

* `comment_id`: A unique identifier for each comment.
* `score`: The score assigned to each comment.
* `self_text`: The content of the comment.
* `subreddit`: The subreddit where the comment was posted.
* `created_time`: The timestamp of when the comment was created.
* `post_id`: The unique identifier of the post associated with the comment.
* `author_name`: The username of the comment author.
* `controversiality`: Indicates whether the comment is deemed controversial or not.
* `ups`: The number of upvotes received by the comment.
* `downs`: The number of downvotes received by the comment.
* `user_is_verified`: Indicates whether the user's account is verified.
* `user_account_created_time`: The timestamp of when the user's account was created.
* `user_awardee_karma`: The karma of the user as an awardee.
* `user_awarder_karma`: The karma of the user as an awarder.
* `user_link_karma`: The link karma of the user.
* `user_comment_karma`: The comment karma of the user.
* `user_total_karma`: The total karma of the user.
* `post_score`: The score assigned to the post associated with the comment.
* `post_self_text`: The self-text content of the post associated with the comment.
* `post_title`: The title of the post associated with the comment.
* `post_upvote_ratio`: The ratio of upvotes to total votes received by the post.
* `post_thumbs_ups`: The number of thumbs up received by the post.
* `post_total_awards_received`: The total number of awards received by the post.
* `post_created_time`: The timestamp of when the post associated with the comment was created.

The dataset also includes some summary statistics, such as the number of unique values for certain columns, and the count of comments within certain date ranges. Additionally, there are some columns with labels that are not immediately clear without additional context, such as `Label` and `Count`.

Based on this information, some potential questions that could be explored with this dataset include:

* What are the most common subreddits where comments about the Russia-Ukraine conflict are posted?
* How has the volume of comments about the conflict changed over time?
* Are there any patterns in the sentiment of comments based on the subreddit they were posted in?
* How do the karma scores of users who comment on the conflict compare to users who do not comment on the conflict?
* Are there any correlations between the controversiality score of a comment and its upvote/downvote ratio?

* How do the self-text content and titles of posts associated with comments about the conflict differ from posts not associated with the conflict?
* What are some common topics or themes that emerge in comments about the conflict?
* How do the demographics (e.g., account age, verified status) of users who comment on the conflict compare to users who do not comment on the conflict?
* Are there any notable trends in the use of awards (e.g., gold, silver, platinum) in comments about the conflict?
* How do the upvote/downvote ratios of comments about the conflict compare to comments not about the conflict?

Columns Rejected and Reasons:

Post_self_text : 85% column empty
Post_total_award_recieved : all 0
Created_time : comment created time was left out because there is no context
Ups: for new comment there will be no upvotes, so not used due the final goal of project, that is to predict controversiality of a new comment
Downs: all zero
User_account_created_time: not necessary for project (can be used but not necessary, I think)
user_awardee_karma, user_awarder_karma, user_link_karma, user_comment_karma : not used to avoid overfitting, used total karma instead which is sum of all 4
Post_thumbs_up: same as post score
Post_total_awards_recieved: All zero
Post_created_time: will make more sense if there was context

Columns Used for Analysis :

Comment_id, score, self_text, subreddit, post_id, Author_name, controversiality, user_is_verified, user_total_karma, post_score, post_title, post_upvote_ratio

Columns That could be Used :

Author_name: not sure how to use (famous authors post may have more chance of controversy)
Post_created_time and created_time: can be used to check engagement time for post, like when post was created, when 1st comment came, over what time period post was popular by checking comment frequency, when most popular posts were created, when most popular comment under each post was created…..