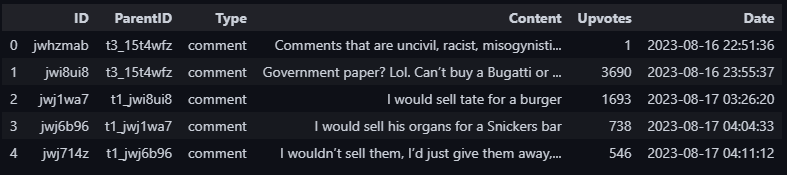
**Reddit Data Documentation**

**Tate\_all.csv**



**Type –** comment or reply

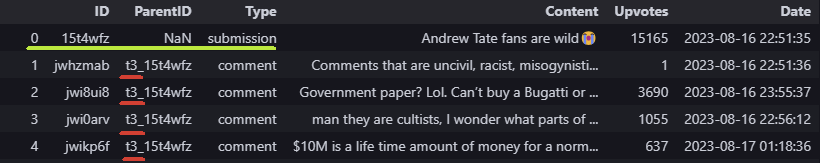
**Content –** Title or text of the submission / comment

**Upvotes –** Comment Score

**Date -** %Y-%m-%d %H:%M:%S

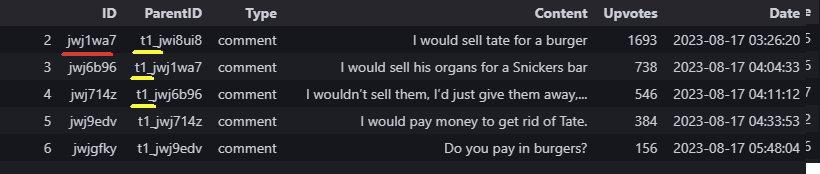
**How the Comment Thread is Listed:**

* **Type** = submission
* **Type** = comment & **ParentID** corresponds to **ID**: top-level comment
* **Type** = comment and **ParentID** corresponds **ParentID**: This is a reply to another comment.



[0] is a submission

[1-4] are replying to the submission (t3\_)



[2] is a reply (t1\_)

[3] is replying to [2]

**Code**

tempdf = pd.read\_csv('Tate\_all.csv')

submission\_id = '15t4wfz'

# submission rows

submission\_row = tempdf[tempdf['ID'] == submission\_id]

#top-level comments for submission

top\_level\_comments = tempdf[tempdf['ParentID'] == 't3\_' + submission\_id]

# IDs of top-level comments

top\_level\_comment\_ids = top\_level\_comments['ID'].tolist()

# replies to top-level comments

replies\_to\_top\_level\_comments = tempdf[tempdf['ParentID'].isin(top\_level\_comment\_ids)]

# Concatenate the submission row, top-level comments, and their replies

result\_df = pd.concat([submission\_row, top\_level\_comments, replies\_to\_top\_level\_comments])

# Reset the index

result\_df = result\_df.reset\_index(drop=True)

result\_df

