Untitled2

October 27, 2021

```python
[1]: import pandas as pd
```

```python
[2]: var = pd.read_csv("/home/sudnya_mali2/DS1.csv")
     var.head()
```

```
[2]:          Id  PostTypeId  AcceptedAnswerId  ParentId          CreationDate  \
     0    927358           1          927386.0       NaN  2009-05-29 18:09:14
     1   2003505           1         2003515.0       NaN  2010-01-05 01:12:15
     2   5767325           1         5767357.0       NaN  2011-04-23 22:17:18
     3  16956810           1        16957078.0       NaN  2013-06-06 08:06:45
     4   2906582           1         2906586.0       NaN  2010-05-25 16:39:47

       DeletionDate  Score  ViewCount  \
     0          NaN  23348   10062790
     1          NaN  18514    9285139
     2          NaN   9931    8937271
     3          NaN   6304    8229811
     4          NaN   2306    7805008

                                                     Body  OwnerUserId  …  \
     0  <p>I accidentally <strong>committed the wrong …      89904.0  …
     1  <p>I want to delete a branch both locally and …      95592.0  …
     2  <p>I have an array of numbers and I'm using th…     364969.0  …
     3  <p>I'm trying to find a way to scan my entire …     954986.0  …
     4  <p>I would like to create an HTML button that …      48523.0  …

              LastEditDate      LastActivityDate  \
     0  2021-06-30 05:07:30  2021-10-05 13:26:47
     1  2021-10-06 22:38:35  2021-10-06 22:38:35
     2  2020-05-21 04:20:03  2021-09-24 08:51:22
     3  2017-05-21 11:40:16  2021-08-30 13:56:02
     4  2021-04-21 16:46:06  2021-09-22 19:42:43

                                                    Title  \
     0  How do I undo the most recent local commits in…
     1  How do I delete a Git branch locally and remot…
     2    How can I remove a specific item from an array?
```

```
3  How do I find all files containing specific te…
4  How to create an HTML button that acts like a …


                                                 Tags AnswerCount CommentCount  \
0            <git><version-control><git-commit><undo>          98           10
1  <git><version-control><git-branch><git-push><g…          41           10
2                                 <javascript><arrays>         112            5
3              <linux><text><grep><directory><find>           53           10
4                  <html><button><hyperlink><anchor>           37            9


   FavoriteCount  ClosedDate   CommunityOwnedDate   ContentLicense
0         7322.0         NaN  2013-03-16 10:08:31   CC BY-SA 4.0
1         5727.0         NaN                  NaN   CC BY-SA 4.0
2         1507.0         NaN                  NaN   CC BY-SA 4.0
3         3036.0         NaN                  NaN   CC BY-SA 3.0
4          503.0         NaN                  NaN   CC BY-SA 4.0

[5 rows x 23 columns]
```

```python
var['Body']=var['Body'].str.replace(r'<[^<>]*>',"",regex=True)
var['Body']=var['Body'].str.replace(r'([^\w])'," ", regex=True)
var['Body']=var['Body'].str.replace(r'\n'," ",regex=True)
var['Body']=var['Body'].str.replace(r'\d+'," ",regex=True)
var
```

```
[3]:             Id  PostTypeId  AcceptedAnswerId  ParentId          CreationDate  \
0        927358           1         927386.0       NaN  2009-05-29 18:09:14
1       2003505           1        2003515.0       NaN  2010-01-05 01:12:15
2       5767325           1        5767357.0       NaN  2011-04-23 22:17:18
3      16956810           1       16957078.0       NaN  2013-06-06 08:06:45
4       2906582           1        2906586.0       NaN  2010-05-25 16:39:47
...         ...         ...              ...       ...                  ...
49995  30862162           1              NaN       NaN  2015-06-16 08:14:39
49996   1067016           1        1067021.0       NaN  2009-07-01 02:23:49
49997  16587585           1       21623752.0       NaN  2013-05-16 12:34:42
49998   1017621           1        1017937.0       NaN  2009-06-19 12:12:06
49999   8618374           1        8618395.0       NaN  2011-12-23 16:30:59


       DeletionDate  Score  ViewCount  \
0               NaN  23348   10062790
1               NaN  18514    9285139
2               NaN   9931    8937271
3               NaN   6304    8229811
4               NaN   2306    7805008
...             ...    ...        ...
49995           NaN     51     127759
49996           NaN    204     127758
```

```
49997              NaN    31    127755
49998              NaN   327    127755
49999              NaN    28    127754


                                                 Body  OwnerUserId  … \
0       I accidentally committed the wrong files to Gi…      89904.0  …
1       I want to delete a branch both locally and rem…      95592.0  …
2       I have an array of numbers and I m using the  …     364969.0  …
3       I m trying to find a way to scan my entire Lin…     954986.0  …
4       I would like to create an HTML button that act…      48523.0  …
…                                                     …            …  …
49995   I am working on an Android App  in which the s…    1903022.0  …
49996   Are JOIN queries faster than several queries  …      95135.0  …
49997   I am using an editable primefaces selectOneMen…    2094488.0  …
49998   I have always thought that functional programm…     120410.0  …
49999   I need to show the name of the currently selec…     979470.0  …


            LastEditDate     LastActivityDate  \
0       2021-06-30 05:07:30  2021-10-05 13:26:47
1       2021-10-06 22:38:35  2021-10-06 22:38:35
2       2020-05-21 04:20:03  2021-09-24 08:51:22
3       2017-05-21 11:40:16  2021-08-30 13:56:02
4       2021-04-21 16:46:06  2021-09-22 19:42:43
…                        …                    …
49995   2020-09-28 11:53:11  2021-08-17 06:03:24
49996   2011-02-10 15:39:38  2019-03-28 10:05:55
49997                   NaN  2019-02-07 22:10:31
49998   2017-05-23 12:34:41  2019-11-25 18:50:25
49999   2011-12-23 19:25:40  2018-07-02 11:10:17


                                                 Title  \
0       How do I undo the most recent local commits in…
1       How do I delete a Git branch locally and remot…
2         How can I remove a specific item from an array?
3       How do I find all files containing specific te…
4       How to create an HTML button that acts like a …
…                                                     …
49995   OTP (token) should be automatically read from …
49996                     JOIN queries vs multiple queries
49997                           selectOneMenu ajax events
49998   Why isn't Python very good for functional prog…
49999           Escaping backslash in string - javascript


                                                 Tags  AnswerCount  \
0              <git><version-control><git-commit><undo>           98
1       <git><version-control><git-branch><git-push><g…           41
2                               <javascript><arrays>          112
```

3

```
3                       <linux><text><grep><directory><find>              53
4                         <html><button><hyperlink><anchor>              37
...                                                      ...             ...
49995                         <android><one-time-password>               9
49996        <mysql><database><join><query-optimization>              14
49997                              <ajax><jsf><primefaces>               4
49998                     <python><functional-programming>               9
49999      <javascript><file><input><escaping><backslash>               5

       CommentCount  FavoriteCount            ClosedDate    CommunityOwnedDate  \
0                10         7322.0                   NaN   2013-03-16 10:08:31
1                10         5727.0                   NaN                   NaN
2                 5         1507.0                   NaN                   NaN
3                10         3036.0                   NaN                   NaN
4                 9          503.0                   NaN                   NaN
...             ...            ...                   ...                   ...
49995             8           24.0                   NaN                   NaN
49996             4           45.0                   NaN                   NaN
49997             0            7.0                   NaN                   NaN
49998             1          199.0   2018-06-16 04:50:17                 NaN
49999             2            4.0                   NaN                   NaN

        ContentLicense
0         CC BY-SA 4.0
1         CC BY-SA 4.0
2         CC BY-SA 4.0
3         CC BY-SA 3.0
4         CC BY-SA 4.0
...                ...
49995     CC BY-SA 4.0
49996     CC BY-SA 2.5
49997     CC BY-SA 3.0
49998     CC BY-SA 2.5
49999     CC BY-SA 3.0

[50000 rows x 23 columns]
```

[4]:
```python
var.to_csv("/home/sudnya_mali2/DS1_clean.csv", index=False)
```

[5]:
```python
var = pd.read_csv("/home/sudnya_mali2/DS2.csv")
var.head()
```

[5]:
```
         Id  PostTypeId  AcceptedAnswerId  ParentId          CreationDate  \
0    287404           1          287586.0       NaN   2008-11-13 16:25:07
1   6119774           1         6119798.0       NaN   2011-05-25 05:04:38
2  21502264           1               NaN       NaN   2014-02-01 19:06:45
3  39288997           1        39290979.0       NaN   2016-09-02 09:28:39
```

```
4   32039476            1        32039561.0       NaN   2015-08-16 20:13:30


   DeletionDate  Score  ViewCount  \
0           NaN     32     127752
1           NaN     50     127747
2           NaN     26     127747
3           NaN     94     127743
4           NaN     67     127737


                                                Body  OwnerUserId  … \
0  <p>So I've got a big text file which looks lik…      25371.0  …
1  <p>1: Is there a way to log in to an AWS insta…     139070.0  …
2  <p>I try to convert a SysVintit script used on…    1949357.0  …
3  <p>Hi all I am trying to create schema Test. <…    1931890.0  …
4  <p>I'm generating an SSH key but I don't know …    4457169.0  …


          LastEditDate      LastActivityDate  \
0  2015-09-17 18:18:40   2016-05-09 06:29:28
1                  NaN   2021-03-30 11:53:31
2  2018-11-09 08:48:20   2018-11-27 22:45:24
3  2019-07-20 23:07:26   2021-04-30 06:00:42
4  2021-02-23 18:48:53   2021-02-23 18:48:53


                                               Title  \
0  Using regular expressions to do mass replace i…
1                 SSH to AWS Instance without key pairs
2  How to use a parameter in ExecStart command line?
3  Elasticsearch : Root mapping definition has un…
4  Copy file contents to the clipboard in Linux t…


                             Tags AnswerCount CommentCount FavoriteCount  \
0         <regex><vim><notepad++>          16            0          11.0
1     <ssh><amazon-ec2><ssh-keys>           7            2          36.0
2                 <systemd><sysv>           2            6           5.0
3         <elasticsearch><mapping>           6            0          19.0
4     <linux><git><github><ssh>           5            2           8.0


            ClosedDate  CommunityOwnedDate  ContentLicense
0                  NaN                 NaN    CC BY-SA 3.0
1                  NaN                 NaN    CC BY-SA 3.0
2                  NaN                 NaN    CC BY-SA 4.0
3                  NaN                 NaN    CC BY-SA 4.0
4  2020-07-05 14:59:32                 NaN    CC BY-SA 4.0

[5 rows x 23 columns]
```

```
var['Body']=var['Body'].str.replace(r'<[^<>]*>',"",regex=True)
var['Body']=var['Body'].str.replace(r'([^\w])'," ", regex=True)
var['Body']=var['Body'].str.replace(r'\n'," ",regex=True)
var['Body']=var['Body'].str.replace(r'\d+'," ",regex=True)
var
```

[6]:

	Id	PostTypeId	AcceptedAnswerId	ParentId	CreationDate	\
0	287404	1	287586.0	NaN	2008-11-13 16:25:07	
1	6119774	1	6119798.0	NaN	2011-05-25 05:04:38	
2	21502264	1	NaN	NaN	2014-02-01 19:06:45	
3	39288997	1	39290979.0	NaN	2016-09-02 09:28:39	
4	32039476	1	32039561.0	NaN	2015-08-16 20:13:30	
...	...	...	...	...	...	
49995	5159065	1	5159405.0	NaN	2011-03-01 19:01:51	
49996	6509600	1	6509769.0	NaN	2011-06-28 16:12:08	
49997	13468286	1	13468456.0	NaN	2012-11-20 07:17:56	
49998	998662	1	NaN	NaN	2009-06-15 22:06:02	
49999	13836848	1	13836933.0	NaN	2012-12-12 09:40:49	

	DeletionDate	Score	ViewCount	\
0	NaN	32	127752	
1	NaN	50	127747	
2	NaN	26	127747	
3	NaN	94	127743	
4	NaN	67	127737	
...	...	...	...	
49995	NaN	50	74792	
49996	NaN	250	74790	
49997	NaN	48	74786	
49998	NaN	48	74786	
49999	NaN	4	74785	

	Body	OwnerUserId	...	\
0	So I ve got a big text file which looks like t…	25371.0	…	
1	Is there a way to log in to an AWS instance…	139070.0	…	
2	I try to convert a SysVintit script used on De…	1949357.0	…	
3	Hi all I am trying to create schema Test PU…	1931890.0	…	
4	I m generating an SSH key but I don t know how…	4457169.0	…	
...	...	...	…	
49995	Looking to add in vertical space between plott…	632729.0	…	
49996	How can I resolve this warning WARN warn…	19875.0	…	
49997	Consider the following Linux kernel dump stack…	1031417.0	…	
49998	I hear this term sometimes and am wondering wh…	31327.0	…	
49999	My CSS a_x visibility hidden wi…	1102574.0	…	

	LastEditDate	LastActivityDate	\
0	2015-09-17 18:18:40	2016-05-09 06:29:28	

6
```
```

```
1                        NaN  2021-03-30 11:53:31
2      2018-11-09 08:48:20  2018-11-27 22:45:24
3      2019-07-20 23:07:26  2021-04-30 06:00:42
4      2021-02-23 18:48:53  2021-02-23 18:48:53
…                      …                    …
49995  2011-03-01 19:15:05  2019-07-08 18:43:03
49996  2018-01-27 16:43:50  2019-07-03 11:10:20
49997  2021-08-04 18:58:09  2021-08-04 18:59:03
49998  2010-05-30 22:39:01  2016-07-17 07:54:14
49999  2019-12-22 21:29:21  2019-12-22 21:29:21


                                                   Title  \
0          Using regular expressions to do mass replace i…
1                      SSH to AWS Instance without key pairs
2          How to use a parameter in ExecStart command line?
3          Elasticsearch : Root mapping definition has un…
4          Copy file contents to the clipboard in Linux t…
…                                                        …
49995  Need to add space between SubPlots for X axis …
49996  compilation warning: no rule to process file f…
49997  How to read, understand, analyze, and debug a …
49998                      What is image hashing used for?
49999                         Show/Hide div, with plain JS


                                                    Tags AnswerCount  \
0                            <regex><vim><notepad++>          16
1                         <ssh><amazon-ec2><ssh-keys>           7
2                                   <systemd><sysv>           2
3                          <elasticsearch><mapping>           6
4                          <linux><git><github><ssh>           5
…                                                   …           …
49995                           <python><matplotlib>           2
49996  <objective-c><xcode><cocoa-touch><compiler-con…          5
49997        <c><linux><debugging><linux-kernel><panic>           3
49998                        <algorithm><image><hash>           5
49999                    <javascript><jquery><html><css>           6


       CommentCount FavoriteCount            ClosedDate CommunityOwnedDate  \
0                 0          11.0                   NaN                NaN
1                 2          36.0                   NaN                NaN
2                 6           5.0                   NaN                NaN
3                 0          19.0                   NaN                NaN
4                 2           8.0  2020-07-05 14:59:32                NaN
…                 …             …                     …                  …
49995             2          10.0                   NaN                NaN
49996             1          43.0                   NaN                NaN
49997             1          32.0                   NaN                NaN
```

```
49998             0        30.0                NaN                    NaN
49999             4         NaN                NaN                    NaN

       ContentLicense
0         CC BY-SA 3.0
1         CC BY-SA 3.0
2         CC BY-SA 4.0
3         CC BY-SA 4.0
4         CC BY-SA 4.0
...                ...
49995     CC BY-SA 2.5
49996     CC BY-SA 3.0
49997     CC BY-SA 4.0
49998     CC BY-SA 2.5
49999     CC BY-SA 4.0

[50000 rows x 23 columns]
```

[7]:
```
var.to_csv("/home/sudnya_mali2/DS2_clean.csv", index=False)
```

[8]:
```
var = pd.read_csv("/home/sudnya_mali2/DS3.csv")
var.head()
```

[8]:
```
         Id  PostTypeId  AcceptedAnswerId  ParentId          CreationDate  \
0  44436856           1        44437064.0       NaN   2017-06-08 13:17:32
1   8799660           1         8801409.0       NaN   2012-01-10 06:57:19
2  16698644           1        16698962.0       NaN   2013-05-22 18:00:55
3  21811789           1        21812046.0       NaN   2014-02-16 13:31:29
4   1234163           1         1237018.0       NaN   2009-08-05 15:57:54

  DeletionDate  Score  ViewCount  \
0          NaN     35      74784
1          NaN     31      74784
2          NaN     13      74783
3          NaN     21      74783
4          NaN     77      74782

                                                Body  OwnerUserId  … \
0  <p>I have a dataset in the following way:</p>\…    1134753.0  …
1  <p>I am trying to configure Tomcat 6 using SSL…    1129903.0  …
2  <p>I am writing a method that can take the <co…    1276534.0  …
3  <p>I'm trying to connect my android applicatio…    2493252.0  …
4  <p>I am using Internet Explorer 8, not by choi…     106634.0  …

         LastEditDate       LastActivityDate  \
0                 NaN  2019-08-13 13:35:38
1  2012-01-10 11:09:43  2018-02-23 06:02:28
```

```
2  2020-07-02 16:24:26  2020-07-02 16:24:26
3                   NaN  2020-04-05 03:46:54
4  2010-01-21 21:11:31  2015-01-23 08:46:24


                                         Title  \
0              Explode array data into rows in spark
1       TOMCAT SSL Error: Alias name does not identify…
2       VBA How to copy the content of a cell without …
3                     Android connection to localhost
4           IE 8 Developer Tools not working In Windows?


                                     Tags AnswerCount CommentCount  \
0                    <apache-spark><pyspark>           3            4
1                      <tomcat><ssl><keytool>           1            0
2                     <excel><vba><excel-2003>           3            2
3          <java><android><json><localhost><wamp>        9            0
4  <internet-explorer><internet-explorer-8><ie-de…     23            0


   FavoriteCount          ClosedDate  CommunityOwnedDate  ContentLicense
0           16.0  2017-06-08 13:31:38                 NaN     CC BY-SA 3.0
1            6.0                 NaN                 NaN     CC BY-SA 3.0
2            4.0                 NaN                 NaN     CC BY-SA 4.0
3            8.0                 NaN                 NaN     CC BY-SA 3.0
4           24.0                 NaN                 NaN     CC BY-SA 2.5

[5 rows x 23 columns]
```

```python
var['Body']=var['Body'].str.replace(r'<[^<>]*>',"",regex=True)
var['Body']=var['Body'].str.replace(r'([^\w])'," ", regex=True)
var['Body']=var['Body'].str.replace(r'\n'," ",regex=True)
var['Body']=var['Body'].str.replace(r'\d+'," ",regex=True)
var
```

```
[9]:             Id  PostTypeId  AcceptedAnswerId  ParentId          CreationDate  \
0       44436856           1        44437064.0       NaN  2017-06-08 13:17:32
1        8799660           1         8801409.0       NaN  2012-01-10 06:57:19
2       16698644           1        16698962.0       NaN  2013-05-22 18:00:55
3       21811789           1        21812046.0       NaN  2014-02-16 13:31:29
4        1234163           1         1237018.0       NaN  2009-08-05 15:57:54
...          ...         ...               ...       ...                  ...
49995     767551           1          767577.0       NaN  2009-04-20 09:29:04
49996   12467794           1               NaN       NaN  2012-09-17 22:10:18
49997   18156122           1        18156912.0       NaN  2013-08-09 21:44:57
49998   44595433           1        44595725.0       NaN  2017-06-16 18:04:13
49999    4389232           1         4389290.0       NaN  2010-12-08 15:33:50


       DeletionDate  Score  ViewCount  \
```

```
0                 NaN     35      74784
1                 NaN     31      74784
2                 NaN     13      74783
3                 NaN     21      74783
4                 NaN     77      74782
…                   …      …        …
49995             NaN     52      53348
49996             NaN     22      53348
49997             NaN     16      53348
49998             NaN     13      53347
49999             NaN     52      53347


                                                    Body   OwnerUserId  …  \
0       I have a dataset in the following way    FieldA…     1134753.0  …
1       I am trying to configure Tomcat    using SSL wi…    1129903.0  …
2       I am writing a method that can take the Target…    1276534.0  …
3       I m trying to connect my android application t…     2493252.0  …
4       I am using Internet Explorer    not by choice …     106634.0  …
…                                                       …          …   …
49995   I have a VirtualBox process hanging around whi…       76663.0  …
49996   As shown on Apple s September    keynote   the s…   1491531.0  …
49997   Hi I have a simple use case for ng repeat star…     2316221.0  …
49998   I m very new to Java    lambdas and stuff    I …    5402618.0  …
49999   I have a piece of Java code in a String    Str…      470184.0  …


              LastEditDate      LastActivityDate  \
0                      NaN   2019-08-13 13:35:38
1      2012-01-10 11:09:43   2018-02-23 06:02:28
2      2020-07-02 16:24:26   2020-07-02 16:24:26
3                      NaN   2020-04-05 03:46:54
4      2010-01-21 21:11:31   2015-01-23 08:46:24
…                        …                     …
49995  2020-01-13 13:51:21   2020-02-06 06:12:10
49996  2012-09-17 22:58:21   2013-06-20 09:23:16
49997                  NaN   2014-02-26 03:18:24
49998  2021-06-15 09:09:06   2021-06-15 09:09:06
49999  2010-12-08 15:36:50   2020-04-09 01:53:42


                                                  Title  \
0                    Explode array data into rows in spark
1        TOMCAT SSL Error: Alias name does not identify…
2        VBA How to copy the content of a cell without …
3                        Android connection to localhost
4            IE 8 Developer Tools not working In Windows?
…                                                       …
49995    How to stop 'uninterruptible' process on Linux?
49996        How to set status bar tint color on iOS 6?
```

```
49997  angularjs: ng-repeat-start and ng-repeat-end w…
49998  Java 8: How to write lambda stream to work wit…
49999              Run piece of code contained in a String

                                                    Tags AnswerCount  \
0                          <apache-spark><pyspark>          3
1                           <tomcat><ssl><keytool>          1
2                         <excel><vba><excel-2003>          3
3              <java><android><json><localhost><wamp>         9
4       <internet-explorer><internet-explorer-8><ie-de…        23
…                                                    …          …
49995   <linux><signals><sigabrt><kill-process><sigkill>       6
49996                                    <ios><ios6>          6
49997                  <angularjs><angularjs-ng-repeat>       3
49998              <java><lambda><java-8><java-stream>         3
49999                             <java><reflection>          9

       CommentCount FavoriteCount            ClosedDate CommunityOwnedDate  \
0                 4          16.0  2017-06-08 13:31:38                NaN
1                 0           6.0                  NaN                NaN
2                 2           4.0                  NaN                NaN
3                 0           8.0                  NaN                NaN
4                 0          24.0                  NaN                NaN
…               …             …                    …                  …
49995             1           7.0                  NaN                NaN
49996             7           9.0                  NaN                NaN
49997             1           3.0                  NaN                NaN
49998             1           3.0                  NaN                NaN
49999             6          16.0                  NaN                NaN

       ContentLicense
0         CC BY-SA 3.0
1         CC BY-SA 3.0
2         CC BY-SA 4.0
3         CC BY-SA 3.0
4         CC BY-SA 2.5
…                  …
49995     CC BY-SA 4.0
49996     CC BY-SA 3.0
49997     CC BY-SA 3.0
49998     CC BY-SA 4.0
49999     CC BY-SA 2.5

[50000 rows x 23 columns]
```

```
[10]: var.to_csv("/home/sudnya_mali2/DS3_clean.csv", index=False)
```

```
[11]: var = pd.read_csv("/home/sudnya_mali2/DS4.csv")
      var.head()
```

```
[11]:          Id  PostTypeId  AcceptedAnswerId  ParentId           CreationDate  \
      0  12130653           1        12130885.0       NaN  2012-08-26 13:49:18
      1  31223189           1        31223247.0       NaN  2015-07-04 17:08:56
      2    369147           1          369174.0       NaN  2008-12-15 17:40:18
      3  13846050           1        13846183.0       NaN  2012-12-12 18:15:17
      4   9947073           1               NaN       NaN  2012-03-30 16:40:13


        DeletionDate  Score  ViewCount  \
      0          NaN     36      53346
      1          NaN     32      53345
      2          NaN     18      53345
      3          NaN     18      53345
      4          NaN     47      53344


                                                    Body  OwnerUserId  ...  \
      0  <p>Java Gurus,</p>\n\n<p>I am pretty new for <...     185041.0  ...
      1  <p>I am using Laravel 5 for developing an app...    2627842.0  ...
      2  <p>I have a paragraph of text in a javascript ...      74552.0  ...
      3  <blockquote>\n  <p><strong>Possible Duplicate:...    1898723.0  ...
      4  <p>I have a NSMutableDictionary, and i have ad...    1180990.0  ...


                LastEditDate        LastActivityDate  \
      0  2012-08-26 14:10:19  2020-08-29 07:33:45
      1  2020-02-03 07:53:19  2020-07-27 00:57:25
      2                  NaN  2017-02-01 09:33:32
      3  2017-05-23 10:29:21  2012-12-12 18:43:17
      4  2016-10-03 13:12:07  2018-03-19 19:47:37


                                                   Title  \
      0  Custom annotation as Interceptor for a method ...
      1  In Laravel 5, How to disable VerifycsrfToken m...
      2  javascript regex to extract anchor text and UR...
      3                          Python 3.3 Mysql Connector
      4                 Saving NSDictionary to NSUserDefaults


                                                    Tags AnswerCount CommentCount  \
      0      <java><logging><annotations><interceptor>           3            3
      1               <php><laravel><laravel-5><csrf>           4            2
      2                   <javascript><regex><anchor>           6            1
      3                                       <python>           2            2
      4      <iphone><objective-c><ios><cocoa-touch>           2            5


         FavoriteCount        ClosedDate  CommunityOwnedDate  ContentLicense
      0           24.0               NaN                 NaN     CC BY-SA 3.0
```

```
1           3.0                 NaN            NaN    CC BY-SA 4.0
2          13.0                 NaN            NaN    CC BY-SA 2.5
3           4.0  2012-12-12 23:18:37           NaN    CC BY-SA 3.0
4          11.0                 NaN            NaN    CC BY-SA 3.0

[5 rows x 23 columns]
```

```
[12]: var['Body']=var['Body'].str.replace(r'<[^<>]*>',"",regex=True)
      var['Body']=var['Body'].str.replace(r'([^\w])'," ", regex=True)
      var['Body']=var['Body'].str.replace(r'\n'," ",regex=True)
      var['Body']=var['Body'].str.replace(r'\d+'," ",regex=True)
      var
```

```
[12]:              Id  PostTypeId  AcceptedAnswerId  ParentId           CreationDate  \
      0      12130653           1      12130885.0           NaN  2012-08-26 13:49:18
      1      31223189           1      31223247.0           NaN  2015-07-04 17:08:56
      2        369147           1        369174.0           NaN  2008-12-15 17:40:18
      3      13846050           1      13846183.0           NaN  2012-12-12 18:15:17
      4       9947073           1             NaN           NaN  2012-03-30 16:40:13
      ...         ...         ...             ...           ...                  ...
      49995   1705525           1       1705638.0           NaN  2009-11-10 03:53:15
      49996  13705087           1      13705282.0           NaN  2012-12-04 14:24:49
      49997  34117191           1      34121884.0           NaN  2015-12-06 12:03:41
      49998   1673544           1       1673574.0           NaN  2009-11-04 12:40:20
      49999  42080123           1      42080223.0           NaN  2017-02-07 01:21:32

            DeletionDate  Score  ViewCount  \
      0              NaN     36      53346
      1              NaN     32      53345
      2              NaN     18      53345
      3              NaN     18      53345
      4              NaN     47      53344
      ...            ...    ...        ...
      49995          NaN      8      41426
      49996          NaN      5      41426
      49997          NaN     22      41425
      49998          NaN     23      41425
      49999          NaN      9      41425

                                                         Body  OwnerUserId  …  \
      0      Java Gurus   I am pretty new for annotations a…     185041.0  …
      1      I am using Laravel   for developing an app  My…    2627842.0  …
      2      I have a paragraph of text in a javascript var…      74552.0  …
      3          Possible Duplicate    MySQL db lib for Pyth…    1898723.0  …
      4      I have a NSMutableDictionary  and i have added…    1180990.0  …
      ...                                                 ...          ...  …
      49995  I m using jQuery to drive my AJAX UI  I have a…      10140.0  …
```

13

```
49996  I downloaded a Java project that I want to stu…    1675976.0  …
49997  I am currently trying to build my ts files int…   1332999.0  …
49998  Suppose I have a string that contains Ü  How w…     31610.0  …
49999  I was wondering how to justify how much space …   5354898.0  …


                 LastEditDate     LastActivityDate  \
0       2012-08-26 14:10:19  2020-08-29 07:33:45
1       2020-02-03 07:53:19  2020-07-27 00:57:25
2                       NaN  2017-02-01 09:33:32
3       2017-05-23 10:29:21  2012-12-12 18:43:17
4       2016-10-03 13:12:07  2018-03-19 19:47:37
…                         …                    …
49995   2009-11-10 18:27:51  2017-08-23 11:18:56
49996   2012-12-04 14:40:45  2019-03-01 01:33:32
49997                   NaN  2019-08-10 09:21:29
49998                   NaN  2019-03-04 08:08:12
49999   2018-02-19 16:27:13  2019-10-30 15:15:20


                                                Title  \
0           Custom annotation as Interceptor for a method …
1           In Laravel 5, How to disable VerifycsrfToken m…
2           javascript regex to extract anchor text and UR…
3                              Python 3.3 Mysql Connector
4                      Saving NSDictionary to NSUserDefaults
…                                                       …
49995   Drag and drop sorting of table columns with jQ…
49996   "The import java.awt cannot be resolved" and "…
49997     tsconfig.json - Only build ts files from folder
49998   How do I detect unicode characters in a Java s…
49999   Controlling the amount of space in justify-con…


                                       Tags AnswerCount CommentCount  \
0           <java><logging><annotations><interceptor>           3            3
1                    <php><laravel><laravel-5><csrf>           4            2
2                       <javascript><regex><anchor>           6            1
3                                        <python>           2            2
4           <iphone><objective-c><ios><cocoa-touch>           2            5
…                                             …           …            …
49995   <jquery><jquery-ui><sorting><drag-and-drop>           7            2
49996                          <java><eclipse>           3            9
49997                            <typescript>           4            0
49998                   <java><string><unicode>           6            3
49999                     <html><css><flexbox>           5            2


     FavoriteCount          ClosedDate  CommunityOwnedDate  ContentLicense
0            24.0                 NaN                 NaN   CC BY-SA 3.0
1             3.0                 NaN                 NaN   CC BY-SA 4.0
```

14

```
2              13.0                    NaN                    NaN    CC BY-SA 2.5
3               4.0   2012-12-12 23:18:37                    NaN    CC BY-SA 3.0
4              11.0                    NaN                    NaN    CC BY-SA 3.0
…               …                      …                      …          …
49995           5.0                    NaN                    NaN    CC BY-SA 2.5
49996           NaN                    NaN                    NaN    CC BY-SA 3.0
49997           0.0                    NaN                    NaN    CC BY-SA 3.0
49998           4.0                    NaN                    NaN    CC BY-SA 2.5
49999           1.0                    NaN                    NaN    CC BY-SA 3.0

[50000 rows x 23 columns]
```

[13]:
```python
var.to_csv("/home/sudnya_mali2/DS4_clean.csv", index=False)
```