

## Task A: Shell Commands

### Question Description:

Are you a fan of sports? Have you paid attention to the Olympic games held in Tokyo - Japan

earlier this year? In this task, you are required to explore and wrangle the data in the file

“Olympics\_tweets.csv”, which contains tweets related to the Tokyo Olympic 2021 that were

collected from Twitter. The file contains different variables to describe the collected tweets,

e.g., the id of a tweet, the text content of a tweet, the screen name of a user who posted a

tweet (i.e., user\_screen\_name), and so on. Please note that you are only allowed to use shell

commands as you would run in Linux shell, Mac terminal, or Cygwin, to tackle this task.

Using other utilities or tools such as PowerShell is NOT allowed.

1. The first step is to explore the contents of the file. Please write commands to display the first 15 rows and the last 15 rows of the data file.

### First 15 rows:

#### First way:

#### Code Explanation:

Open the csv file | split by ‘,’ and select all column between 1 to 15 | open in editor

#### Code:

cat Olympics\_tweets.csv | awk -F',' 'NR>=1 && NR<= 15' | less

#### Output:

```
id,text,user_screen_name,user_location,retweet_count,favorited,favorite_count,user_description,user_created_at,user_followers
user_friends,date,language
141896231718678e4,Mirabai Chanu's maiden Olympic silver helped India clinch joint-12th spot on the medals tally on Day 1 atã
https://t.co/ksy80vaJr1,ITGDSports,Noida- India,1,FALSE,1,Live cricket scores- news- analysis and fun facts on all your fav
ourite sports.,26/09/2017 10:57,9333,28,24/07/2021 15:52,en
141896235133692e4,ãMirabai Chanu brings home first silver medal on Day 1. DSEU congratulates her on this achievement. A brill
iant stãhttps://t.co/aIvBeuJfH,dseu_official,NA,0,FALSE,1,The official Twitter account for Delhi Skill and Entrepreneurs
hip University.,12/11/2020 9:01,1168,4,24/07/2021 15:52,en
1418962410679689900,Heartiest congratulations to Mirabai Chanu for starting the medal tally for India in the Tokyo Olympics
2020 by wãhttps://t.co/NukYganXe4,SutejKumar4,New Delhi- India,0,FALSE,0,Believe in yourselfã80Yã*,20/04/2020 16:56,72,
143,24/07/2021 15:52,en
141896246682283e4,Hearty congratulations to ace Indian weightlifter @mirabai_chanu for winning a Silver in the womenãs #Wei
ghtliftingãhttps://t.co/QJqHGstelZ,NsChandapaka,Nalgonda- Telangana,1,FALSE,0,Govt Teacher | YouTuber | Ambedkarist | Indi
an | Cricket Lover | Founder of @808NAMI | Social Activist | Kargil Vijay Diwas-8/08/2017 1:26:5907:60,24/07/2021 15:53,en
1418962529569670100,Congratulations to @mirabai_chanu for winning silver medal in weight lifting. This is first medal for Ind
ia in Tokyoãhttps://t.co/NhVphfT0IA,PrabhatFatehpur,Fatehpur-Uttar Pradesh,0,FALSE,1,Nationalist |Engineer |Plastics Techno
logist| 'CIPET'ian,14/11/2016 17:13,2721,353,24/07/2021 15:53,en
141896254617485e4,ãNarendramodi @Tokyo2020 @mirabai_chanu Noble Group congratulates India's Pride Mirabai Chanu on her winni
ng 1st tãhttps://t.co/4B1zUPL60n,JaiHanu32567563,Panchkula-India,0,FALSE,0,MBA(MDU)- FORMER GM HARYANA,29/07/2017 16:19,1,
84,24/07/2021 15:53,en
1418962573756550100,Mirabai Chanu's maiden Olympic silver helped India clinch the joint-12th spot on the medals tally on Day
1 atãhttps://t.co/TPneDQFzbi,RS_02,NA,0,FALSE,0,Sports fanatic (crazy abt cricket)- media pro- traveler and big time foodi
e.,25/12/2009 15:50,61,405,24/07/2021 15:53,en
141896264720307e4,Congratulations #India #olympics https://t.co/SAPwOJumgR,YSR4Ever,United States,1,FALSE,2,YSRCP USA Leader-
APNRTS USA Regional Coordinator- Advisor- Investor & IT Consultant. Re-tweets are not endorsements.,30/11/2010 16:05,7247,18
5,24/07/2021 15:53,en
14189627232555e4,Congratulations to Mirabai Chanu getting first medal for INDIA on the first day of the #Olympics at Tokyo-
gettingãhttps://t.co/8370zB4rhg,Pravabhavani,NA,0,FALSE,0,NA,31/12/2017 6:33,180,9,24/07/2021 15:54,en
141896284790895e4,ãIndia is on the board on day 1 at #Olympics . Huge congratulations and really proud of you .@mirabai_chanu
. You anãhttps://t.co/WpRHlqfQw,SharatGanesh,San Francisco Bay Area,0,FALSE,0,Product Marketing Leader @Tanium / Advisory
Board @tssafesecurity/ Ex-@FireEye / Alumnì @CUBoulder / Proud Rescue Cat Dad/ 6x Half-Marathoner,1/05/2009 18:45,771,1071,
24/07/2021 15:54,en
141896291031601e4,Tokyo Olympics: As Team India marches holding the Tricolour in their hands- PM Modi stands up to cheer for
themãhttps://t.co/Zi3Uf0B0ux,R90679775Rao,NA,0,FALSE,0,Psychologist-Counselor-motivational speaker,22/07/2021 16:02,0,0,24
/07/2021 15:54,en
141896293736922e4,Olympics: Mirabai lifts weight of a nation- puts India on medal table https://t.co/91AoaF6f1l,TOITopStories
,NA,0,FALSE,0,Breaking news- Home page stories- special coverage and much more from http://t.co/u04cby761l,9/04/2009 10:05,34
62,3534,24/07/2021 15:55,en
141896317847451e4,ãWe congratulate #MirabaiChanu for making the nation proud by winning India's first medal at the Tokyo Olym
pics 202ãhttps://t.co/PeZw2bNcAE,FlakcoIndia,New Delhi- India,0,FALSE,0,NA,17/07/2017 10:49,6,1,24/07/2021 15:55,en
1418963345806329900,Proud to take a Bow and Congratulate @mirabai_chanu on the maiden Olympic Medal For Indian in 2021 Olympi
cs.ãhttps://t.co/iYN9FajRj3,naveenius,Coimbatore,0,FALSE,0,Technologist | Entrepreneur Head of Ventures @LadderMinds Dir
ector - Technology Services @Acodeidea,1/07/2009 13:12,224,873,24/07/2021 15:56,en
```

### Second way:

#### Code Explanation:

Open the csv file | select first 15 rows using head command | open in a editor

Code:

cat Olympics\_tweets.csv | head -n15 | less

Output:

```
id,text,user_screen_name,user_location,retweet_count,favorited,favorite_count,user_description,user_created_at,user_followers
,user_friends,date,language
141896231718678e4,Mirabai Chanu's maiden Olympic silver helped India clinch joint-12th spot on the medals tally on Day 1 at a#
| https://t.co/ksy80vJrJ,ITGDsports,NOIDA- India,1,FALSE,1,Live cricket scores- news- analysis and fun facts on all your fav
ourite sports.,26/09/2017 10:57,9333,28,24/07/2021 15:52,en
141896235133692e4,@mirabai_chanu brings home first silver medal on Day 1. DSEU congratulates her on this achievement. A bril
liant at# | https://t.co/AlTv8eUjFf,dseu_official,NA,0,FALSE,1,The official Twitter account for Delhi Skill and Entrepreneurs
hip University.,12/11/2020 9:01,1168,4,24/07/2021 15:52,en
1418962410673689900,Heartiest congratulations to Mirabai Chanu for starting the medal tally for India in the Tokyo Olympics
2020 by wri# | https://t.co/WukYganXeA,SutejKumar4,New Delhi- India,0,FALSE,0,Believe in yourself#0Y#*,20/04/2020 16:56,72,
143,24/07/2021 15:52,en
141896246682283e4,Hearty congratulations to ace Indian weightlifter @mirabai_chanu for winning a Silver in the women#s #de#
ghtlifting# | https://t.co/OJqHGstE12,NsChandapaka,Malgond- Telangana,1,FALSE,0,Govt Teacher | YouTuber | Ambedkarist | Indi
an | Cricket Lover | Founder of #BORN4WIN | Social Activist | Kargil Vijay Diwas,8/08/2017 1:26,5907,60,24/07/2021 15:53,en
1418962529569670100,Congratulations to @mirabai_chanu for winning silver medal in weight lifting. This is first medal for Ind
ia in Tokyo# | https://t.co/NhVphFI0IA,PrabhatFatehpur,Fatehpur-Uttar Pradesh,0,FALSE,1,Nationalist |Engineer |Plastics Techno
logist| 'CIPE'T'ian,14/11/2016 17:13,2721,353,24/07/2021 15:53,en
141896254617485e4,@narendramodi @Tokyo2020 @mirabai_chanu Noble Group congratulates India's Pride Mirabai Chanu on her winni
ng ist 1# | https://t.co/4BtzuPL60n,JaiHanu3256763,Panchkula-India,0,FALSE,0,MBA(MDU)- FORMER GM HARYANA,29/07/2017 16:19,1,
84,24/07/2021 15:53,en
141896257375650100,Mirabai Chanu's maiden Olympic silver helped India clinch the joint-12th spot on the medals tally on Day
1 at# | https://t.co/1PneDQZb1,RS_02,NA,0,FALSE,0,Sports fanatic (crazy abt cricket)- media pro- traveler and big time Foodi
e,25/12/2009 15:50,61,405,24/07/2021 15:53,en
141896264720307e4,Congratulations #india #olympics https://t.co/Sapw0jurnR,YSR4Ever,United States,1,FALSE,2,YSRCP USA Leader-
APRKS USA Regional Coordinator- Advisor- Investor & IT Consultant. Re-tweets are not endorsements.,30/11/2010 16:05,7247,18
5,24/07/2021 15:53,en
14189627233255e4,Congratulations to Mirabai Chanu getting first medal for INDIA on the first day of the #Olympics at Tokyo-
getting# | https://t.co/8370zB4rhg,Pravabhavani,NA,0,FALSE,0,NA,31/12/2017 6:33,180,9,24/07/2021 15:54,en
141896284790895e4,India is on the board on day 1 at #Olympics . Huge congratulations and really proud of you .@mirabai_chanu
. You an# | https://t.co/WpRhlqITfw,SharatGanesh,San Francisco Bay Area,0,FALSE,0,Product Marketing Leader @Tanium / Advisory
Board @ItssafeSecurity/ Ex-@FireEye / Alumni @CUBoulder / Proud Rescue Cat Dad/ 6x Half-Marathoner ,1/05/2009 18:45,771,1071,
24/07/2021 15:54,en
141896291031601e4,Tokyo Olympics: As Team India marches holding the Tricolour in their hands- PM Modi stands up to cheer for
them# | https://t.co/Z13JUF0Boux,R9067977SRao,NA,0,FALSE,0,Psychologist-Counselor-motivational speaker,22/07/2021 16:02,0,0,24
/07/2021 15:54,en
141896293736922e4,Olympics: Mirabai lifts weight of a nation- puts India on medal table https://t.co/91AoaF6f1l,TOTopStories
,NA,0,FALSE,0,Breaking news - Home page stories- special coverage and much more from http://t.co/u04cbY761l,9/04/2009 10:05,34
461,8534,24/07/2021 15:55,en
141896317847451e4,We congratulate #MirabaiChanu for making the nation proud by winning India's first medal at the Tokyo Olym
pics 202# | https://t.co/Pezw2NCAE,FlakcoIndia,New Delhi- India,0,FALSE,0,NA,17/07/2017 10:49,6,1,24/07/2021 15:55,en
141896334806329900,Proud to take a Bow and Congratulate @mirabai_chanu on the maiden Olympic Medal for Indian in 2021 Olympi
cs. # | https://t.co/1yN9FajR3J,naveen#eus,Coimbatore,0,FALSE,0,Technologist | Entrepreneur Head of Ventures @LadderMinds Dir
ector - Technology Services @Acadidea,1/07/2009 13:12,224,673,24/07/2021 15:56,en
```

Last 15 rows:

Code Explanation:

Open the csv file | select bottom 15 rows | open in a editor

Code:

cat Olympics\_tweets.csv | tail -n15 | less

Output:

```
142157846137203e4,You can do archery @U008F# as well? You should be in the Olympics,orange#cher,England- Unite
d Kingdom,0,FALSE,0,I live with my bossy cat NA #Candace- I want to rem#*,17/04/2010 9:16,1386,519,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,1,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21:45,en
14215787920905e4,@huuggj#s that#s what i#e#m saying# #TurnerAndHoock #TokyoOlympics #LoveIsLandUK #GhostsAft#
rIfe# | https://t.co/7ip7Jc#H7g,glssk#j,sb to unfollow,0,FALSE,0,moved to @koosginbap moved to @koosginbap moved to @koosgi
nbap moved to @koosginbap moved to @koosginbap moved to @koosginbap,30/10/2020 19:44,2993,2810,31/07/2021 21
```

Code Explanation:

Display the size of the file and number of line inside the file

Code:

`wc Olympics_tweets.csv`

output:

lines, word, byte counts

```
114214 3394061 34715688 Olympics_tweets.csv
```

3. What is the field separator in the file? Write a command to display the number of variables (i.e., the columns) in the data and a command to display the column names.

**What is the field separator?**

The field operator is ','

Example:

```
id,text,user_screen_name,user_location,retweet_count,favorited,favorite_count,user_description,user_created_at,user_friends,date,language
141896231718678e4,Mirabai Chanu's maiden Olympic silver helped India clinch joint-12th spot on the medals
https://t.co/ksy80vaJr1,ITGDsports,Noida- India,1,FALSE,1,Live cricket scores- news- analysis and fun fa
ourite sports.,26/09/2017 10:57,9333,28,24/07/2021 15:52,en
```

**Number of variables**

Code Explanation:

display the number of variables

Code:

`awk -F, '{ print NF; exit }' Olympics_tweets.csv`

Output:

```
$ awk -F, '{ print NF; exit }' Olympics_tweets.csv
13
```

**Columns names**

Code Explanation:

The first line is the column names and head -n 1 means display the first line

Code:

`head -n 1 Olympics_tweets.csv`

Output:

```
id,text,user_screen_name,user_location,retweet_count,favorited,favorite_count,user_description,user_created_at,user_followers
,user_friends,date,language
```

4. Write commands to count and then remove lines with an id that is not a number of 19 digits long, i.e., id values that contain anything other than numbers OR are of a length more/less

than 19. Store the filtered set in a file named “filtered\_tweets\_1.csv” and use this file for the following questions in Task A.

Code Explanation:

Split by ‘,’ if values’ length in the first line is equal to 19 or it’s the column names line, print. Save all selected values in a new file named filtered\_tweets\_1.csv | open the new file | count lines and words and bytes count.

Code:

```
awk -F',' '{if (length($1) == 19 || NR==1) print }' Olympics_tweets.csv > filtered_tweets_1.csv | cat filtered_tweets_1.csv | head -n -1 | wc
```

Output:

Lines, words, bytes count

```
awk -F',' '{if (length($1) == 19 || NR==1) print }' Olympics_tweets.csv > filtered_tweets_1.csv | cat filtered_tweets_1.csv | head -n -1 | wc
21390 631362 6501252
```

5. What is the date range of the tweets? Please note that the file is not guaranteed to be sorted and Nulls (NA and empty values) should not be considered.

Code Explanation:

Open the csv file | split by ‘,’ if column 12 is not null value and its not the first line, print column 12 | sort the column | print the first and last dates.

Code:

```
cat filtered_tweets_1.csv | awk -F',' '{if($12!=NULL && NR > 1) print $12}' | sort | (head -n1 && tail -n1)
```

output:

```
24/07/2021 15:52
31/07/2021 9:59
```

6. When was the first mention of the term “Japan” in the dataset? What about “Australia”? Please note that the first mention of a term refers to the creation time of the tweet that contained the specific term for the first time in the dataset and the terms to be searched are case sensitive.

**Japan:**

Code Explanation:

Open the csv file | find the word Japan | find the first column | print column 12

Code:

```
$ cat filtered_tweets_1.csv | grep 'Japan' | head -n1 | awk -F',' '{print $12}'
output:
```

```
$ cat filtered_tweets_1.csv | grep 'Japan' | head -n1 | awk -F',' '{print $12}'
25/07/2021 16:21
```

Australia:

Code Explanation:

Open the csv file | find the word Australia | select first line | print column 12

Code:

```
cat filtered_tweets_1.csv | grep 'Australia' | head -n1 | awk -F"," '{print $12}'
```

Output:

```
25/07/2021 16:25
```

7. How many unique “user\_screen\_name” values are there in the dataset? Can you list the top 10 most frequent user\_screen\_name in the dataset?

Total unique values:

Code Explanation:

Open the csv file | split by ‘,’ print third column | remove first line | sort values | find unique values | sort values | count total rows

Code:

```
cat filtered_tweets_1.csv | awk -F',' '{print $3}' | tail -n+2 | sort | uniq -c | sort -nr | wc -l
```

Output:

```
18425
```

Top 10 most frequent value:

Code Explanation:

Open the csv file | split by ‘,’ print column 3 | unselect first line | count all unique values | sort rows | print first 10 lines

Code:

```
cat filtered_tweets_1.csv | awk -F',' '{print $3}' | tail -n+2 | sort | uniq -c | sort -nr | head -n10
```

Output:

```
45 kegan61438051
36 dev_discourse
32 allsports70
25 nippon_ja
25 olympics
19 nbcchicago
19 NBCLA
18 godinhumanform4
16 mniskhoka
16 olympics_india_
```

8. Hashtags are commonly used in Twitter (e.g., #olympics2021). Write commands to count the number of hashtags in the “text” column of the data. Please notice that hashtags always start with #. If a hashtag appears twice in a text value, it should be counted twice.

Code Explanation:

Open the csv file | split by ',' print column 2 | find and count character '#'

Code:

```
cat filtered_tweets_1.csv | awk -F',' '{print $2}' | grep -c '#'
```

Output:

4821

9. In the following, let's focus on certain columns contained in the file. Write commands to only keep columns that are in the following list:

- id
- user\_screen\_name
- user\_created\_at
- user\_followers
- user\_friends
- date

In particular, please only keep the tweets with user\_friends and user\_followers each larger than 1000. Export the selected data to a new file named "filtered\_tweets\_2.csv". Please ensure that the file "filtered\_tweets\_2.csv" contains data from the selected columns as well as the column names.

Code Explanation:

Split the file by ',' | if value in column 11 is greater than 1000 and value in column 10 is greater than 1000, print column 1, 3, 9, 10, 11, 12 and save in a new csv file called filtered\_tweets\_2.csv

Code:

```
awk -F',' '{if ($11 > 1000 && $10 > 1000) print $1, "$3", "$9", "$10", "$11", "$12}' filtered_tweets_1.csv > filtered_tweets_2.csv
```

10. In the file "filtered\_tweets\_2.csv", how many tweets are with a NA value in the last column (i.e., the column "date")? How many user accounts were created prior to 2020. Please note that the column "user\_created\_at" specifies when a user account was created and one Twitter user might have produced multiple tweets and you are supposed to count the accounts.

**How many tweets with a NA in the date column:**

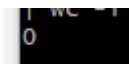
Code Explanation:

Open the csv file | split by ',' if the column 6 is null or equal to "NA" print the line | count rows

Code:

```
cat filtered_tweets_2.csv | awk -F',' '{if ($6 == Null || $6 == "NA") print}' | wc -l
```

output:

0

**How many user accounts were created prior to 2020:**

Code Explanation:

Open the csv file | sort and remove duplicate id | split by ',' print column 3 | split by " and print column 1 | split by '/' select rows if column 3 is smaller than 2020 | remove first line | count total lines

Code:

```
cat filtered_tweets_2.csv | sort -u -t, -r -k1,1 | awk -F ',' '{print $3}' | awk -F '"' '{print $1}' | awk -F '/' '{if ($3 < 2020) print}' | tail -n+2 | wc -l
```

output:

A terminal window showing the output of the command, which is the number 4077.