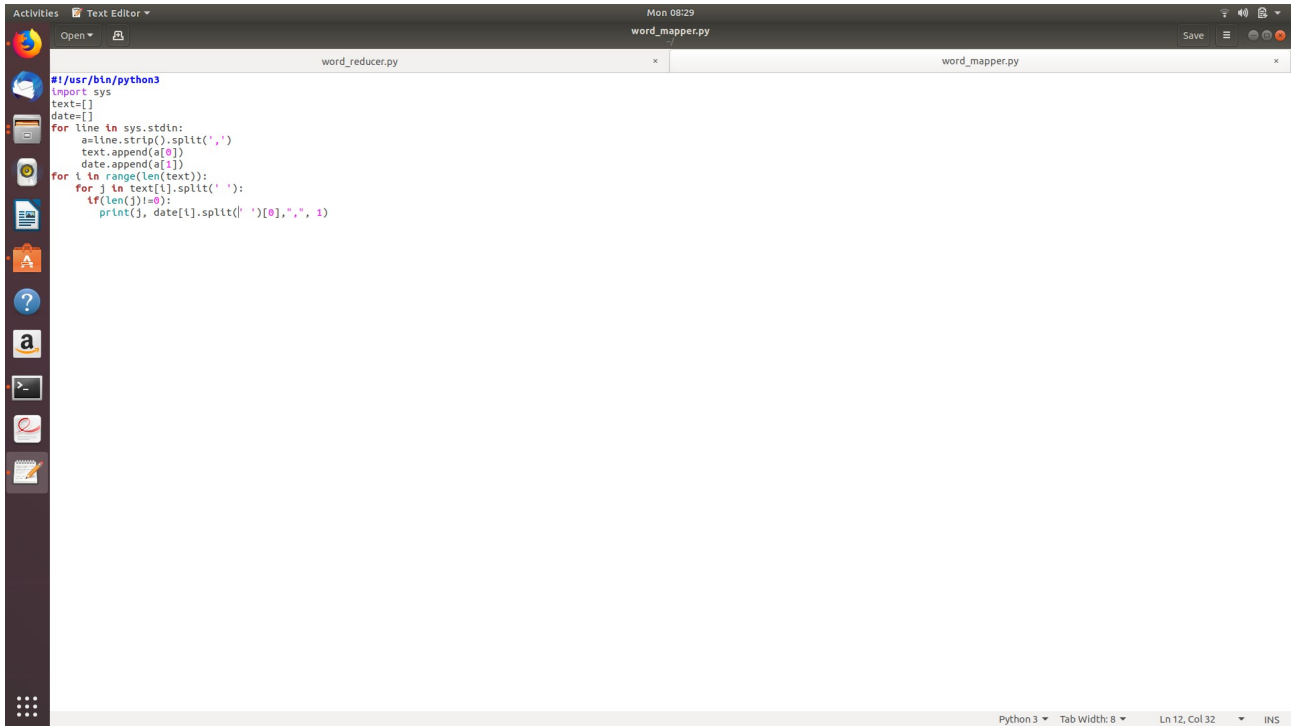


ASSIGNMENT-3(b)

Mapper code for not null words with date:

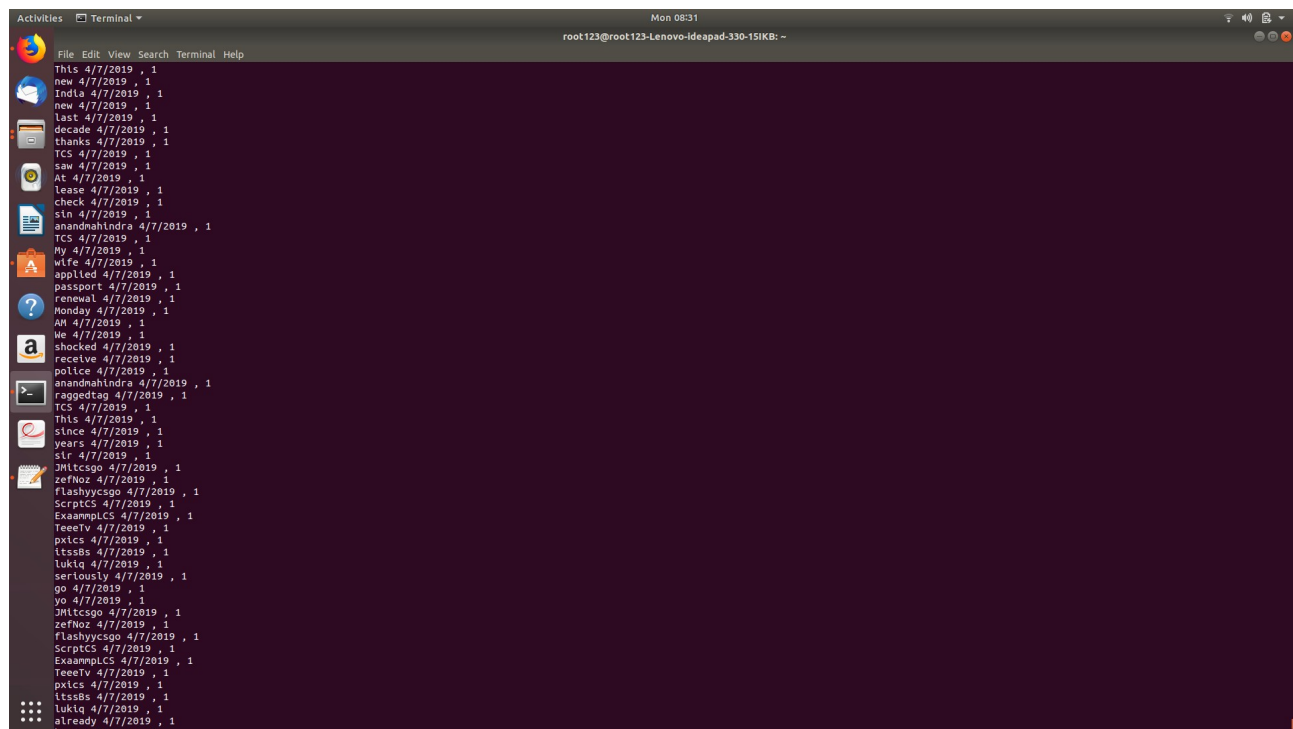


The screenshot shows a text editor window titled 'word_mapper.py' with the following Python code:

```
#!/usr/bin/python3
import sys
text=[]
date=[]
for line in sys.stdin:
    a=line.strip().split(',')
    text.append(a[0])
    date.append(a[1])
    for i in range(len(text)):
        for j in text[i].split(' '):
            if(len(j)!=0):
                print(j, date[i].split(' ')[0],",", 1)
```

The editor interface includes a sidebar with application icons, a top menu bar with 'Open', and a status bar at the bottom showing 'Python 3', 'Tab Width: 8', 'Ln 12, Col 32', and 'INS'.

hadoop fs -cat /input/tweets.csv | '/home/root123/word_mapper.py'

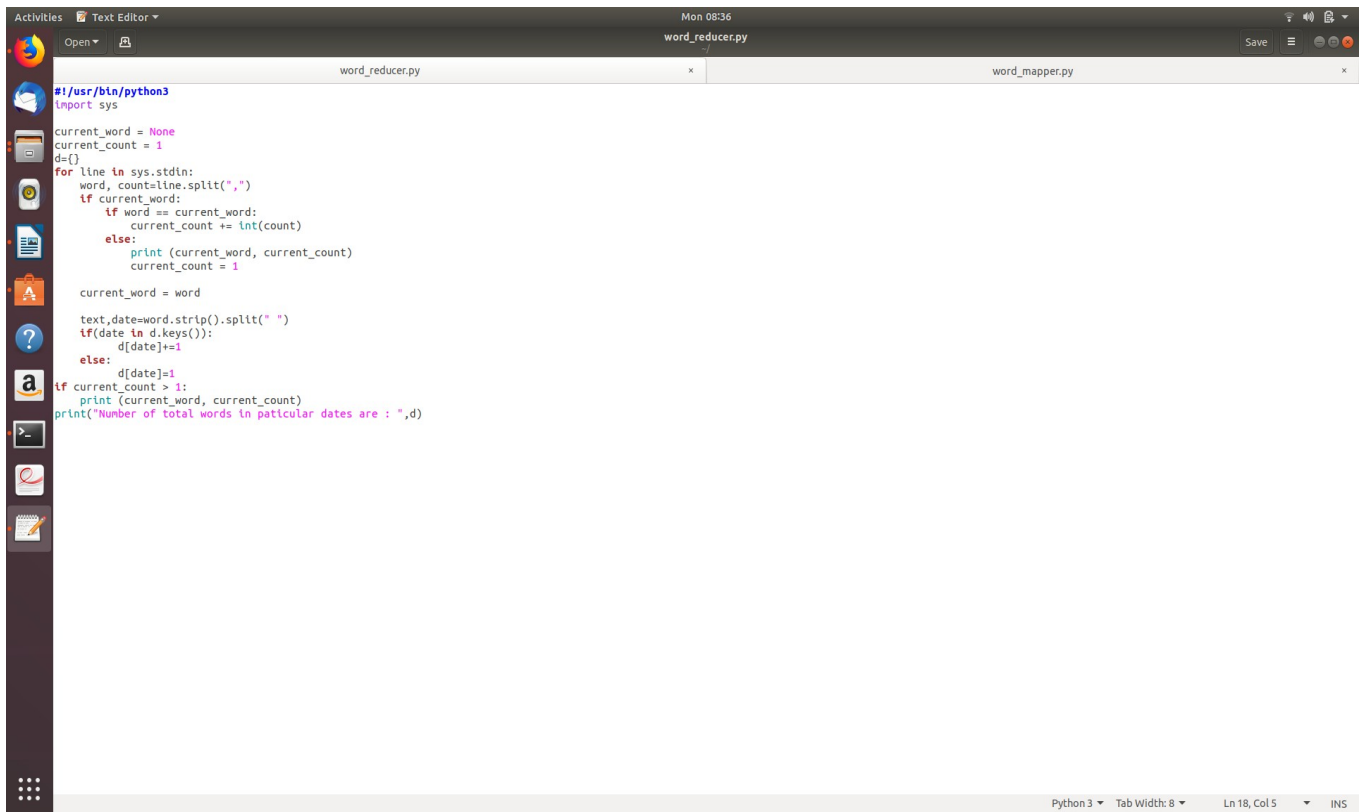


The screenshot shows a terminal window with the output of the word_mapper.py script. The output consists of multiple lines, each containing a word, a date, and a count. The words are: This, new, India, new, last, decade, thanks, TCS, saw, At, lease, check, sin, anandnahlndra, TCS, My, wife, applied, passport, renewal, Monday, AM, Me, shocked, receive, police, anandnahlndra, raggedtag, TCS, This, since, years, str, Jhktsgo, zefNoz, flashyycsgo, ScryptCS, ExamplCS, TeeTv, pxlcs, itsBs, lukiq, seriously, go, yo, Jhktsgo, zefNoz, flashyycsgo, ScryptCS, ExamplCS, TeeTv, pxlcs, itsBs, lukiq, already, before.

```
This 4/7/2019 , 1
new 4/7/2019 , 1
India 4/7/2019 , 1
new 4/7/2019 , 1
last 4/7/2019 , 1
decade 4/7/2019 , 1
thanks 4/7/2019 , 1
TCS 4/7/2019 , 1
saw 4/7/2019 , 1
At 4/7/2019 , 1
lease 4/7/2019 , 1
check 4/7/2019 , 1
sin 4/7/2019 , 1
anandnahlndra 4/7/2019 , 1
TCS 4/7/2019 , 1
My 4/7/2019 , 1
wife 4/7/2019 , 1
applied 4/7/2019 , 1
passport 4/7/2019 , 1
renewal 4/7/2019 , 1
Monday 4/7/2019 , 1
AM 4/7/2019 , 1
Me 4/7/2019 , 1
shocked 4/7/2019 , 1
receive 4/7/2019 , 1
police 4/7/2019 , 1
anandnahlndra 4/7/2019 , 1
raggedtag 4/7/2019 , 1
TCS 4/7/2019 , 1
This 4/7/2019 , 1
since 4/7/2019 , 1
years 4/7/2019 , 1
str 4/7/2019 , 1
Jhktsgo 4/7/2019 , 1
zefNoz 4/7/2019 , 1
flashyycsgo 4/7/2019 , 1
ScryptCS 4/7/2019 , 1
ExamplCS 4/7/2019 , 1
TeeTv 4/7/2019 , 1
pxlcs 4/7/2019 , 1
itsBs 4/7/2019 , 1
lukiq 4/7/2019 , 1
seriously 4/7/2019 , 1
go 4/7/2019 , 1
yo 4/7/2019 , 1
Jhktsgo 4/7/2019 , 1
zefNoz 4/7/2019 , 1
flashyycsgo 4/7/2019 , 1
ScryptCS 4/7/2019 , 1
ExamplCS 4/7/2019 , 1
TeeTv 4/7/2019 , 1
pxlcs 4/7/2019 , 1
itsBs 4/7/2019 , 1
lukiq 4/7/2019 , 1
already 4/7/2019 , 1
before 4/7/2019 , 1
```

The terminal window has a title bar with 'Activities', 'Terminal', and 'Mon 08:31'. The prompt is 'root123@root123-Lenovo-Ideapad-330-15IKB: ~'.

Reducer code:



The screenshot shows a text editor window titled 'word_reducer.py' with the following Python code:

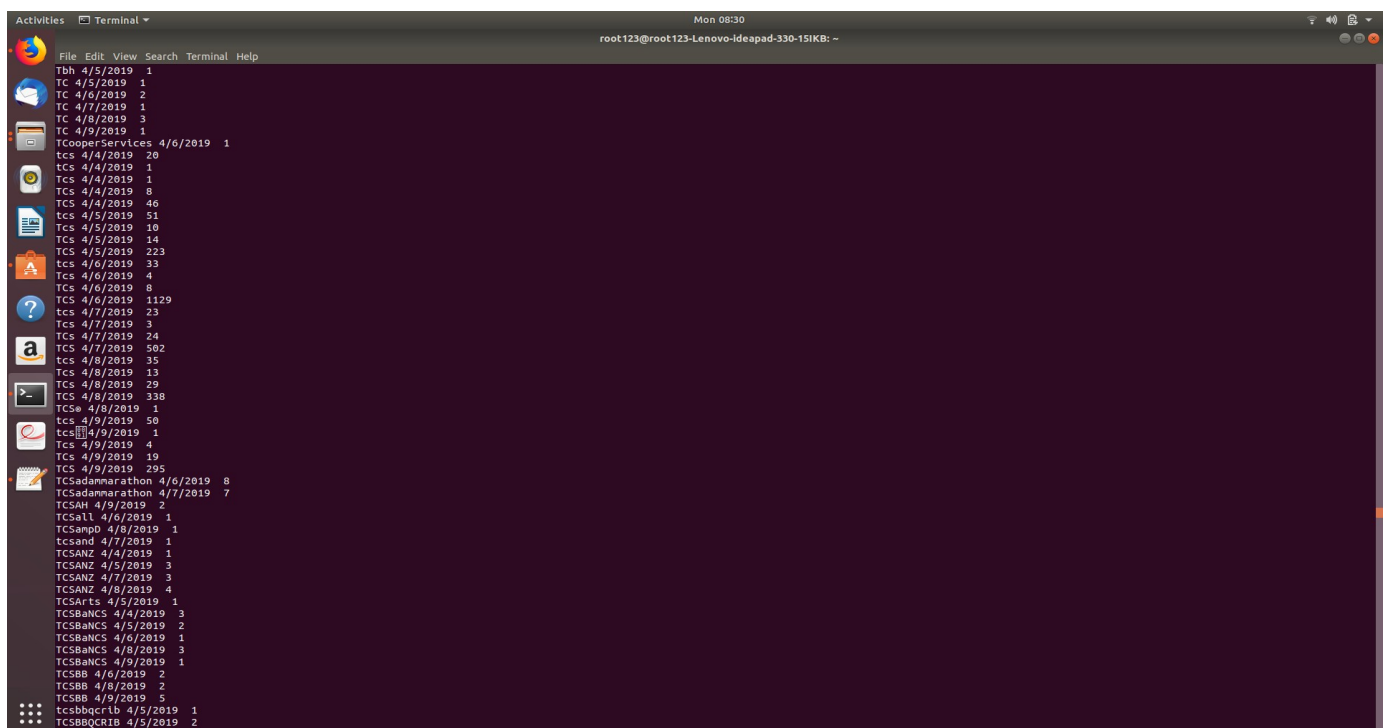
```
#!/usr/bin/python3
import sys

current_word = None
current_count = 1
d={}
for line in sys.stdin:
    word, count=line.split(",")
    if current_word:
        if word == current_word:
            current_count += int(count)
        else:
            print (current_word, current_count)
            current_count = 1
    current_word = word

text,date=word.strip().split(" ")
if(date in d.keys()):
    d[date]+=1
else:
    d[date]=1
if current_count > 1:
    print (current_word, current_count)
print("Number of total words in paticular dates are : ",d)
```

The status bar at the bottom indicates 'Python 3', 'Tab Width: 8', 'Ln 18, Col 5', and 'INS'.

hadoop fs -cat /input/tweets.csv | '/home/root123/word_mapper.py' | sort |
'/home/root123/word_reducer.py'



The screenshot shows a terminal window with the following output:

```
Tbh 4/5/2019 1
TC 4/5/2019 1
TC 4/6/2019 2
TC 4/7/2019 1
TC 4/8/2019 3
TC 4/9/2019 1
TCoopersServices 4/6/2019 1
tcs 4/4/2019 20
tcs 4/4/2019 1
Tcs 4/4/2019 1
Tcs 4/4/2019 8
TCS 4/4/2019 46
tcs 4/5/2019 51
Tcs 4/5/2019 10
TCS 4/5/2019 14
TCS 4/5/2019 223
tcs 4/6/2019 33
Tcs 4/6/2019 4
Tcs 4/6/2019 8
TCS 4/6/2019 1129
tcs 4/7/2019 23
Tcs 4/7/2019 3
TCS 4/7/2019 24
TCS 4/7/2019 502
tcs 4/8/2019 35
Tcs 4/8/2019 13
TCS 4/8/2019 29
TCS 4/8/2019 338
TCS 4/8/2019 1
tcs 4/9/2019 50
tcs 4/9/2019 1
Tcs 4/9/2019 4
TCS 4/9/2019 19
TCS 4/9/2019 295
TCSadamarathon 4/6/2019 8
TCSadamarathon 4/7/2019 7
TCSAH 4/9/2019 2
TCSall 4/6/2019 1
TCSanpb 4/8/2019 1
tcsand 4/7/2019 1
TCSANZ 4/4/2019 1
TCSANZ 4/5/2019 3
TCSANZ 4/7/2019 3
TCSANZ 4/8/2019 4
TCSArts 4/5/2019 1
TCSBANCS 4/4/2019 3
TCSBANCS 4/5/2019 2
TCSBANCS 4/6/2019 1
TCSBANCS 4/8/2019 3
TCSBANCS 4/9/2019 1
TCSBB 4/6/2019 2
TCSBB 4/8/2019 2
TCSBB 4/9/2019 5
TCSbbqcr1b 4/5/2019 1
TCSbbqcr1b 4/5/2019 2
TCSBBANCS 4/4/2019 1
```



```
Activities Terminal * Mon 08:30
root123@root123-Lenovo-Ideapad-330-15IKB: ~

File Edit View Search Terminal Help

VuckFou 4/9/2019 1
Vujjing 4/4/2019 2
Vujjing 4/5/2019 1
YunnL 4/5/2019 1
yung 4/8/2019 1
Yup 4/6/2019 1
yup 4/7/2019 1
Yup 4/7/2019 1
yurrrrrt 4/7/2019 1
Yyou 4/5/2019 1
Z 4/6/2019 1
Z 4/7/2019 3
Zaha 4/7/2019 1
ZatdZamanHaid 4/5/2019 1
zaklavelllv 4/5/2019 1
Zales 4/4/2019 1
zall 4/9/2019 1
ZANTO 4/8/2019 1
zaphare 4/9/2019 1
ZarlashtFaisal 4/9/2019 3
zarneenaay 4/5/2019 1
Zayn 4/5/2019 1
zeal 4/6/2019 1
Zee 4/7/2019 1
ZeeNewsHindi 4/7/2019 1
zeffnoz 4/7/2019 2
ZehriliBaloch 4/9/2019 1
ZerlinaMaxwell 4/5/2019 1
ZerlinaMaxwell 4/8/2019 1
zero 4/6/2019 1
zero 4/7/2019 1
zerocancerrun 4/7/2019 2
zerodhaonline 4/5/2019 1
ZF 4/5/2019 1
Zhang 4/4/2019 2
Zhang 4/5/2019 1
Zla 4/5/2019 2
zip 4/5/2019 1
zkid 4/6/2019 1
ZoahHS 4/7/2019 1
zolevirginia 4/8/2019 1
zonato 4/5/2019 1
ZomatoIN 4/8/2019 1
zone 4/4/2019 1
Zone 4/4/2019 1
zone 4/8/2019 1
Zoners 4/5/2019 1
zorbaly 4/8/2019 1
zuber 4/7/2019 1
Zumba 4/5/2019 1
Zumba 4/8/2019 2
Zurich 4/9/2019 1
ZURICHOPENAIR 4/5/2019 2
...
Number of total words in particular dates are : {'4/4/2019': 2347, '4/5/2019': 7555, '4/6/2019': 13998, '4/7/2019': 8198, '4/8/2019': 9824, '4/9/2019': 9187, 'date': 1}
root123@root123-Lenovo-Ideapad-330-15IKB: ~
```