

Solution to Laboratory Week 7.

The files for this lab can be found in the folder

/KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining/ /code-and-scripts/parking-time.sh
/data/parking-small.csv

1.Using PUTTY with X11 forwarding, open up a bash window to jupiter.csit.rmit.edu.au and log in. You can find detailed instructions in the Canvas shell in the basic unix guide or in the first part of the recording of lecture 6.

First make sure you have activated Xming on your PC as shown in figure 1. If you do not have Xming on your laptop, you can download and install it through the following link:

For windows:

<https://xming.en.softonic.com/?ex=BB-527.0>

For mac user:

<https://www.xquartz.org/>

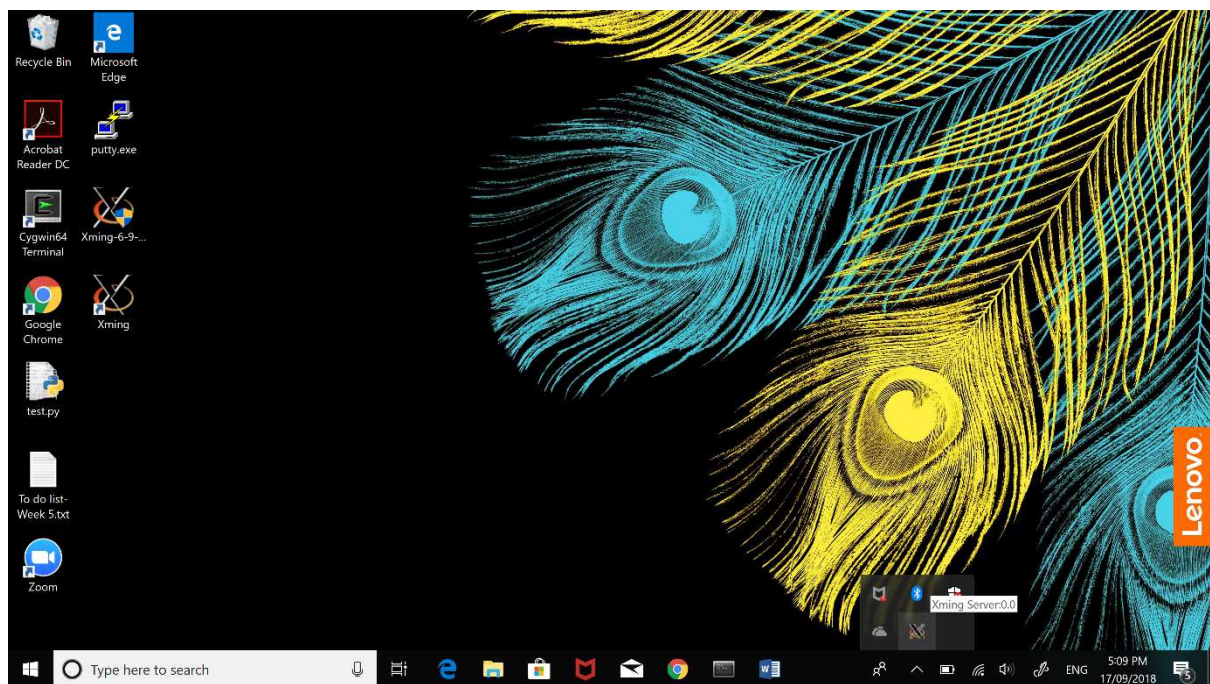


Figure 1. Activate Xming.

After making sure about activation of Xming, you can log in to any of RMIT's linux servers(titan or Jupiter) using Putty. If you do not have Putty on your PC, you can download and install it through the following link:

<https://www.putty.org/>

First you need to enable X11 on Putty from SSH-> X11 and then checking the Enable X11 forwarding as shown in the following figures.

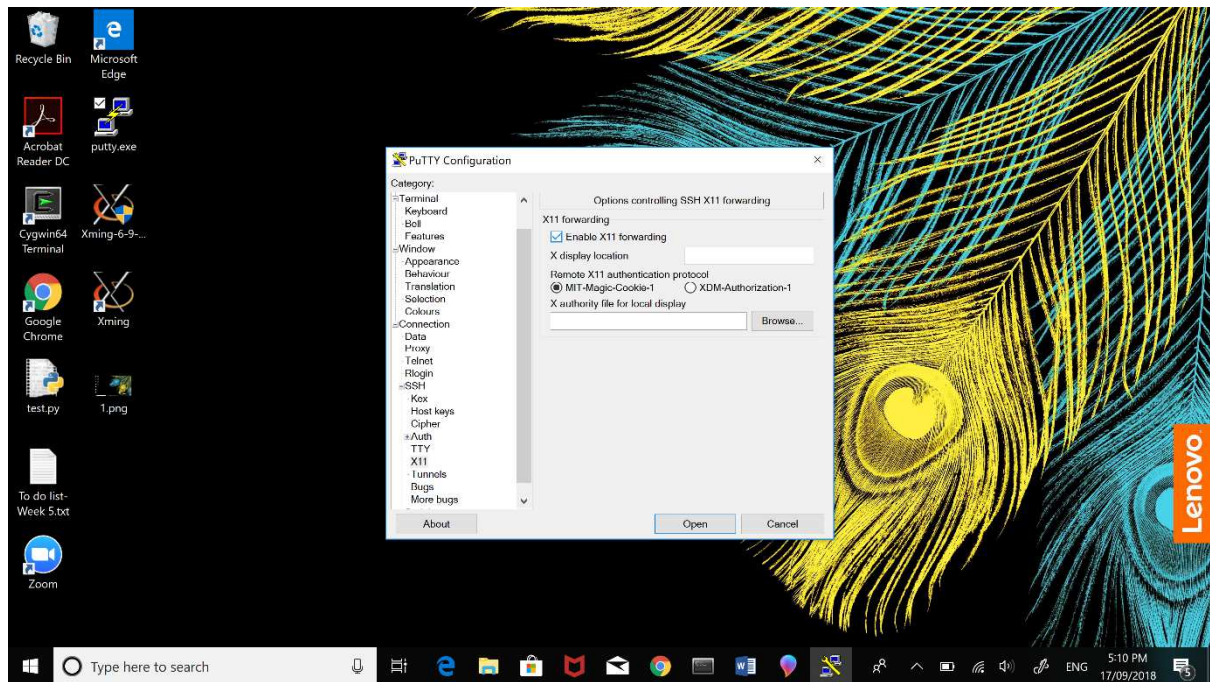


Figure 2. Enable X11 on Putty.

Then, you can go back to session and log in to server by typing “titan.csit.rmit.edu.au “ or Jupiter.csit.rmit.edu.au” in Host Name (or IP address) tab and click open.

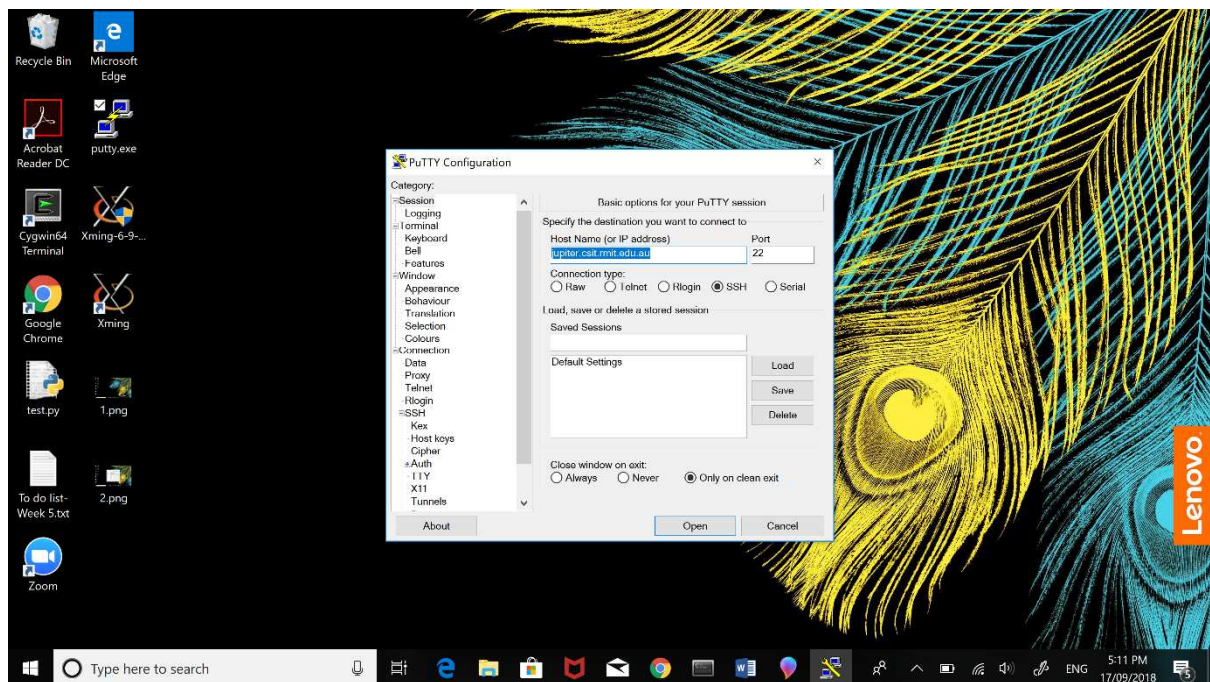


Figure 4. Key in name of RMIT’s linux server (titan or Jupiter) on Putty to log in to the server remotely.

Log in to sever for mac users:

First activate XQuartz.

If you are using mac, you can connect to the servers through your shell command by typing:

```
ssh -Y studentnumber@titan.csit.rmit.edu.au ,
```

where -Y enables X11.

If everything is going alright, you should be able to see the following command shell and type your RMIT user name and then your password, then click on yes and then connect to the server as shown in the following figures.

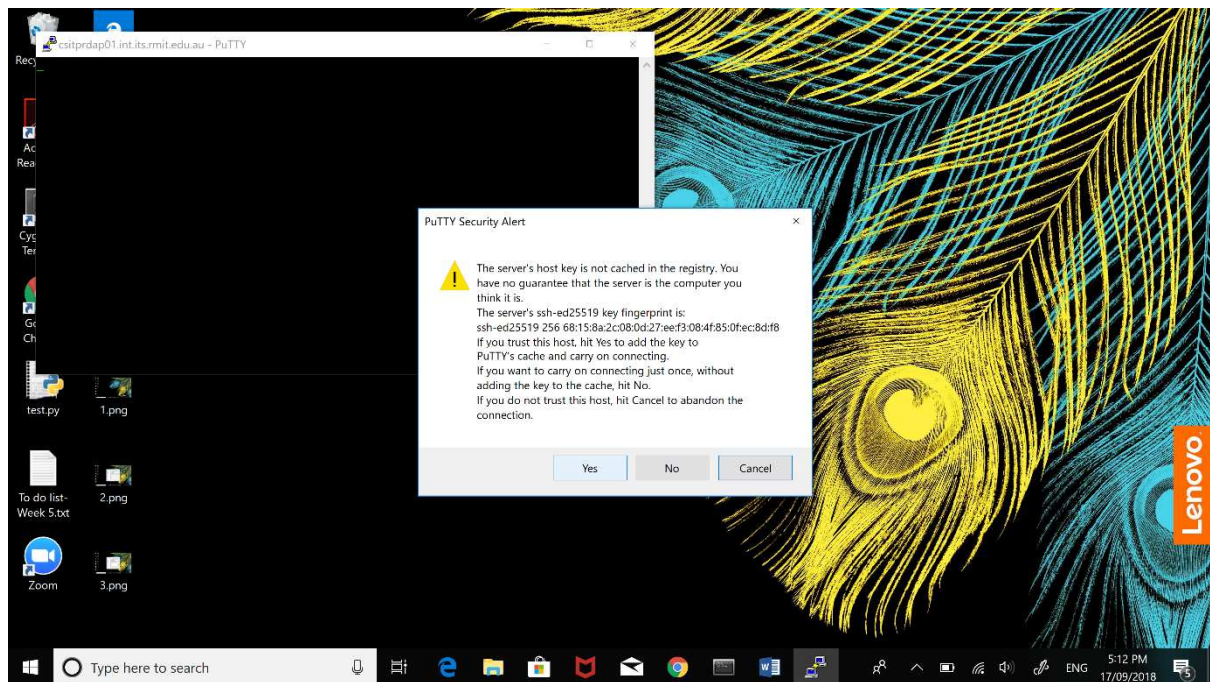


Figure 5.

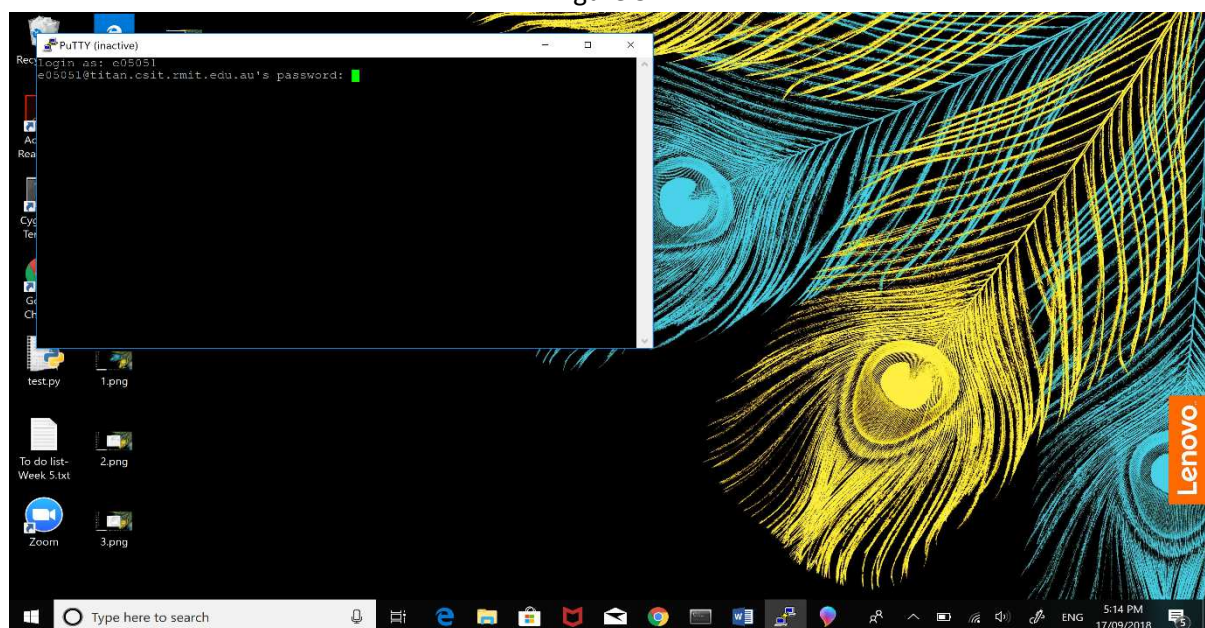


Fig 6.

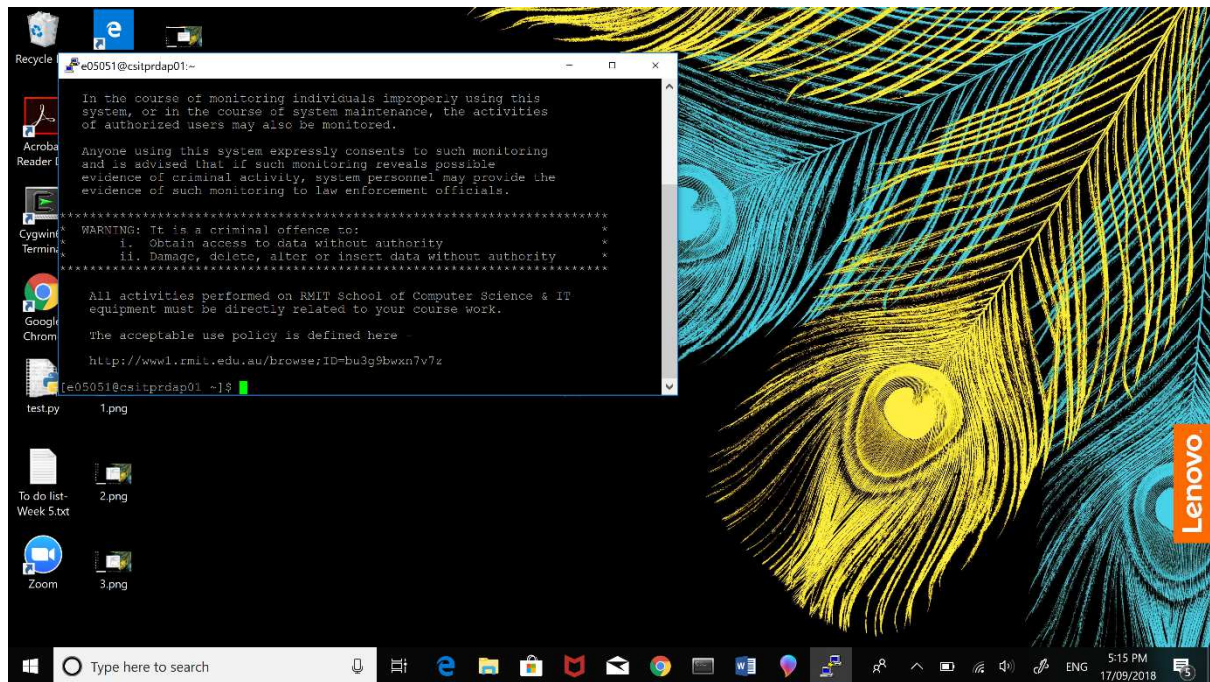


Figure 5-7. Snapshots of connecting to the Jupiter server.

STRONGLY RECOMMENDED: After you have logged in to the server key in “xeyes”& or “xclock”. If this command runs without error you have a working Xwindows connection.

2. If you haven’t already done so make a folder on HDrive for Data Mining, eg,

cd HDrive

mkdir DM

cd DM

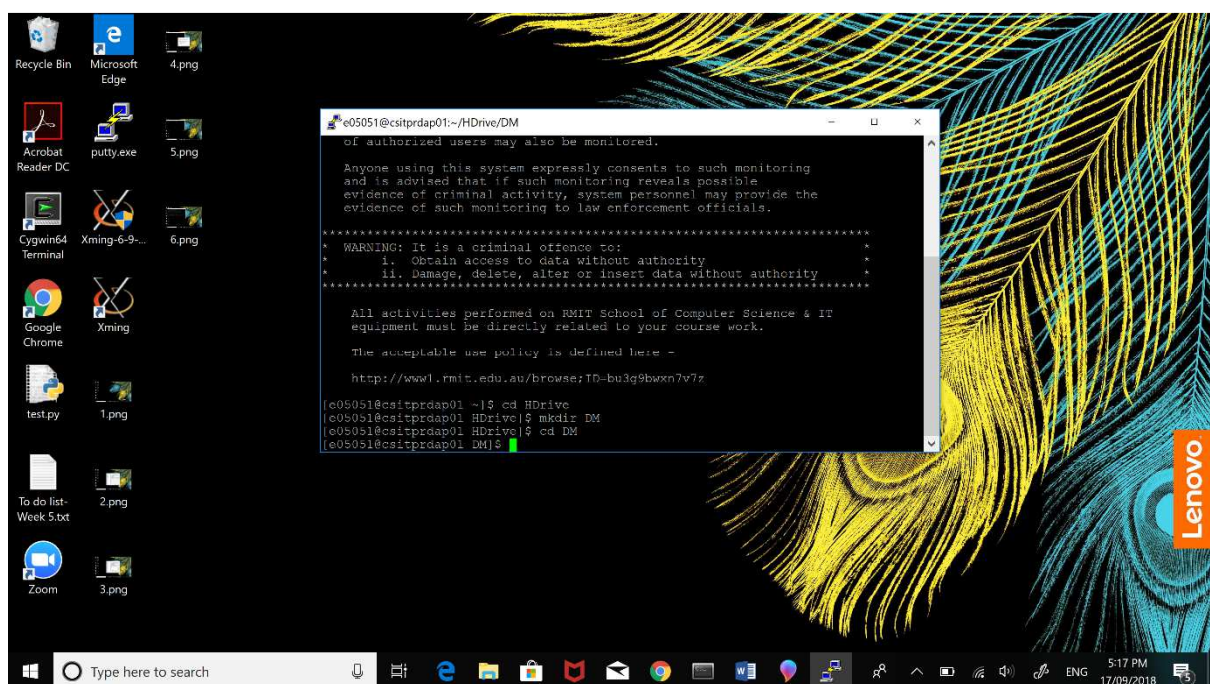


Figure 8. Change directory to HDrive, creating a new directory named DM in HDrive and changing the current directory to the new directory (DM).

Using mkdir DM command you have created a new folder with the name of DM, you can check it in your HDrive as shown in the following figure.

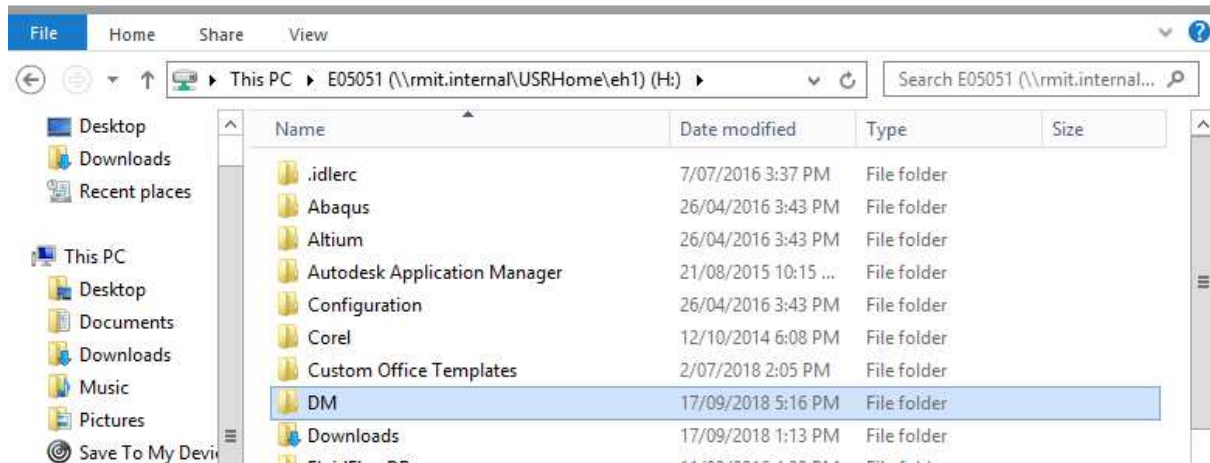


Figure 9. created directory as DM using above commands in HDrive.

3. Copy in the two files above: [Note: If you are cutting and pasting from a pdf of this file, be aware that the ~ character might not be cut out properly]

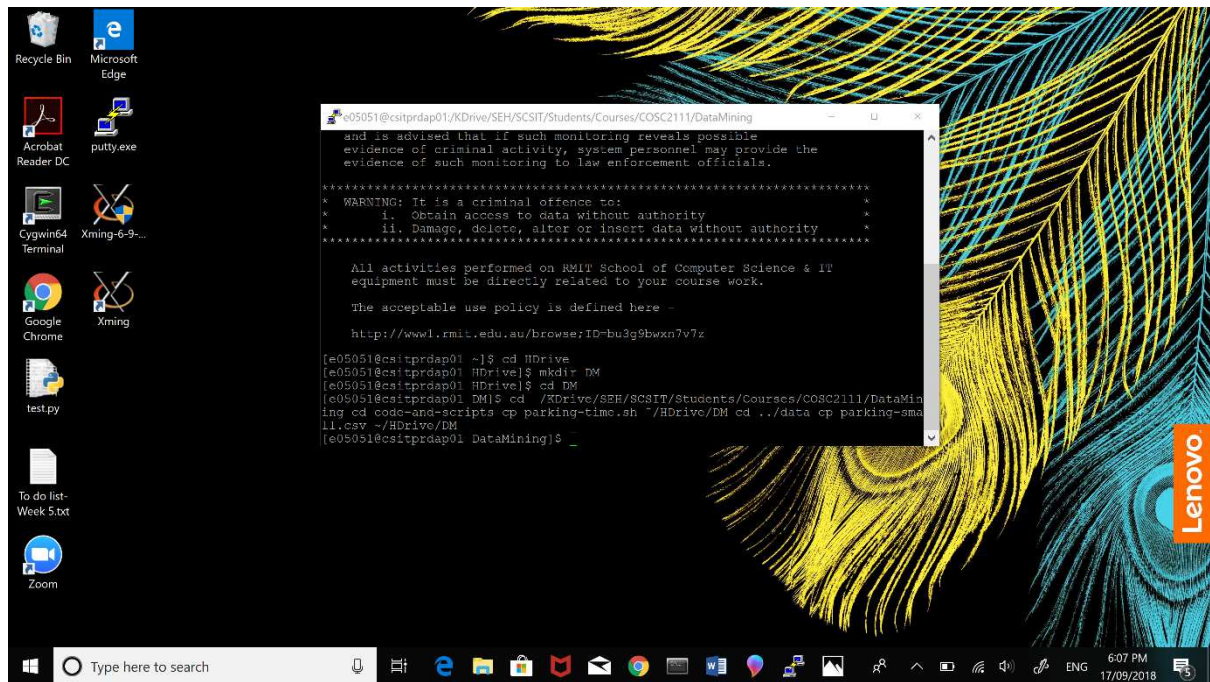
```
cd /KDrive/SEH/SCSIT/Students/Courses/COSC2111/DataMining
```

```
cd code-and-scripts
```

```
cp parking-time.sh ~/HDrive/DM c
```

```
d ../data
```

```
cp parking-small.csv ~/HDrive/DM
```



4. Navigate to your DM folder on the HDrive

`cd`

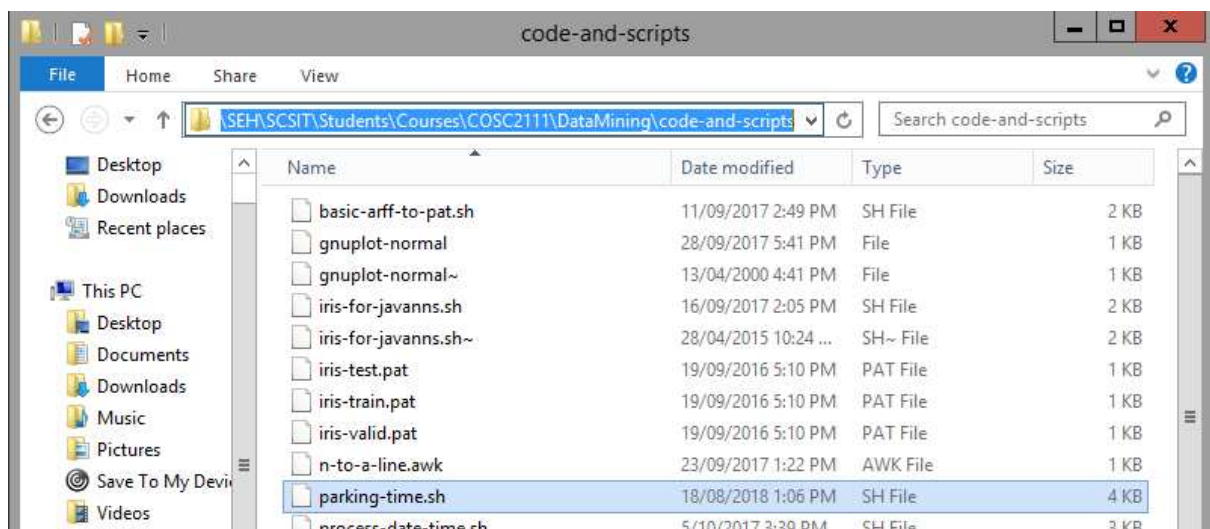
`ls`

`cd HDrive`

`ls`

`cd DM`

5. Open up parking-time.sh with an editor and inspect it. If you don't know any other editor, try nano.



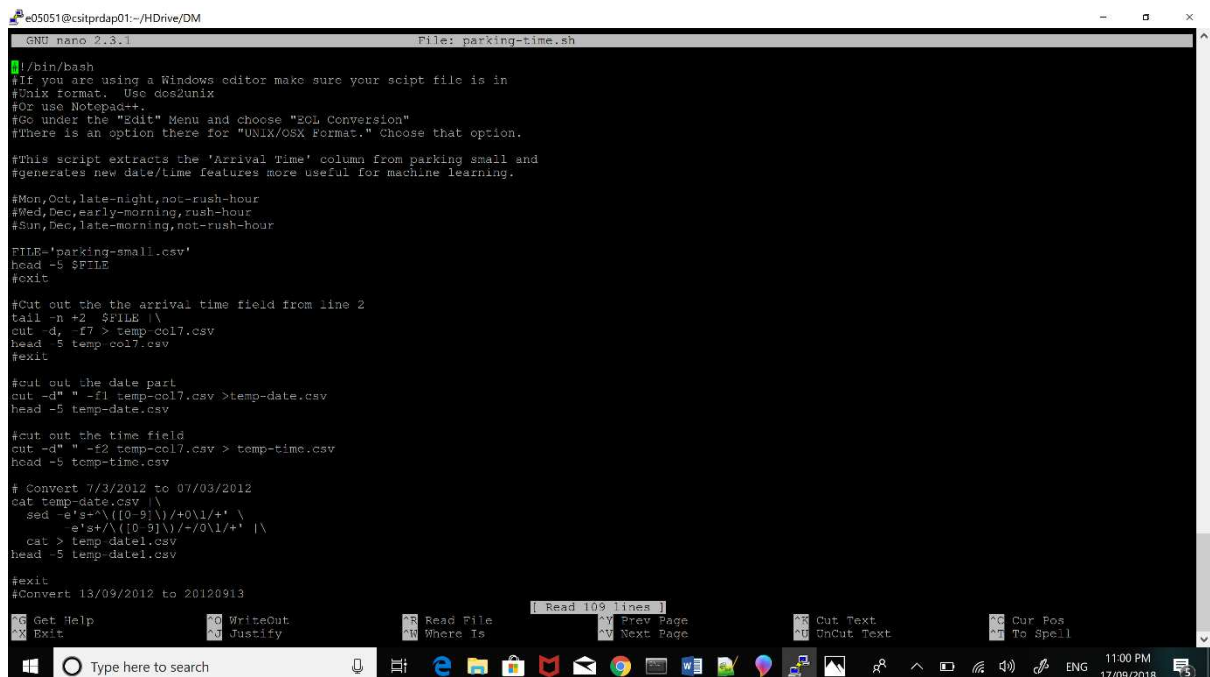
Open parking-time.sh using Notepad++ or any other editor you like.

```
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
parking-time.sh
1 #!/bin/bash
2 #If you are using a Windows editor make sure your script file is in
3 #Unix format. Use dos2unix
4 #Or use Notepad++.
5 #Go under the "Edit" Menu and choose "EOL Conversion"
6 #There is an option there for "UNIX/OSX Format." Choose that option.
7
8 #This script extracts the 'Arrival Time' column from parking small and
9 #generates new date/time features more useful for machine learning.
10
11 #Mon,Oct,late-night,not-rush-hour
12 #Wed,Dec,early-morning,rush-hour
13 #Sun,Dec,late-morning,not-rush-hour
14
15 FILE='parking-small.csv'
16 head -n 5 $FILE
17 #exit
18
19 #Cut out the the arrival time field from line 2
20 tail -n +2 $FILE | \
21 cut -d, -f7 > temp-col7.csv
22 head -n 5 temp-col7.csv
23 #exit
24
25 #cut out the date part
26 cut -d" " -f1 temp-col7.csv >temp-date.csv
27 head -n 5 temp-date.csv
28
29 #cut out the time field
30 cut -d" " -f2 temp-col7.csv > temp-time.csv
31 head -n 5 temp-time.csv
32
33 # Convert 7/3/2012 to 07/03/2012
34 cat temp-date.csv | \
35 sed -e's/^([0-9]\+)/\1/+' \
36 -e's/^([0-9]\+)/\1/+' | \
37 cat > temp-datel.csv
38 head -n 5 temp-datel.csv
39
40 #exit
41 #Convert 13/09/2012 to 20120913
42 cat temp-datel.csv | \
43 sed -e's/^([0-9]\+)/\1/+' \
44 cat > temp-date2.csv
45 head -n 5 temp-date2.csv
46 #exit
```

Or open through nano by typing the following command in the linux command shell:

nano parking-time.sh

Now, you can see what is inside parking-time.sh script as shown in the following figure.



```
GNU nano 2.3.1 File: parking-time.sh
#!/bin/bash
#If you are using a Windows editor make sure your script file is in
#Unix format. Use dos2unix
#Or use Notepad++.
#Go under the "Edit" Menu and choose "EOL Conversion"
#There is an option there for "UNIX/OSX Format." Choose that option.

#This script extracts the 'Arrival Time' column from parking small and
#generates new date/time features more useful for machine learning.

#Mon,Oct,late-night,not-rush-hour
#Wed,Dec,early-morning,rush-hour
#Sun,Dec,late-morning,not-rush-hour

FILE='parking-small.csv'
head -n 5 $FILE
#exit

#Cut out the the arrival time field from line 2
tail -n +2 $FILE | \
cut -d, -f7 > temp-col7.csv
head -n 5 temp-col7.csv
#exit

#cut out the date part
cut -d" " -f1 temp-col7.csv >temp-date.csv
head -n 5 temp-date.csv

#cut out the time field
cut -d" " -f2 temp-col7.csv > temp-time.csv
head -n 5 temp-time.csv

# Convert 7/3/2012 to 07/03/2012
cat temp-date.csv | \
sed -e's/^([0-9]\+)/\1/+' \
-e's/^([0-9]\+)/\1/+' | \
cat > temp-datel.csv
head -n 5 temp-datel.csv

#exit
#Convert 13/09/2012 to 20120913
```

6. To run the script: sh parking-time.sh

```
e05051@csitprdap01:~/HDrive/DM
[e05051@csitprdap01:~/HDrive/DM]$ cd DM
[e05051@csitprdap01:~/HDrive/DM]$ ls
parking-small.csv parking-time.sh
[e05051@csitprdap01:~/HDrive/DM]$ sh parking-time.sh
Area Name,Street Name,Between Street 1,Between Street 2,Side Of Street,Street Ma
rker,Arrival Time,Departure Time,Duration of Parking Event (in seconds),Sign,in
Violation?,Street ID,Device ID
Victoria Market,THERRY STREET,QUEEN STREET,ELIZABETH STREET,4,70218,24/08/2012 1
1:34,24/08/2012 12:49,4473,LZ 15M M-SAT 7:30-19:30,1,1346,3770
Courtney,FEL STREET,O'CONNELL STREET,QUEENSBERRY STREET,2,5398E,17/03/2012 13:0
7,17/03/2012 13:10,127,1/2P A RPE M-SUN 7:30-23:00,0,1101,3472
Victoria Market,FRANKLIN STREET,QUEEN STREET,ELIZABETH STREET,1,C6624,17/02/2012
13:54,17/02/2012 14:20,1543,2P MTR M-SAT 7:30-20:30,0,681,2805
Chinatown,LONSDALE STREET,SWANSTON STREET,RUSSELL STREET,3,2888N,27/11/2011 15:0
3,27/11/2011 16:02,3562,1P SUN 7:30-18:30,0,894,1770
24/08/2012 11:34
17/03/2012 13:07
17/02/2012 13:54
27/11/2011 15:03
21/04/2012 15:08
24/08/2012
17/03/2012
17/02/2012
27/11/2011
21/04/2012
11:34
13:07
13:54
15:03
15:08
24/08/2012
17/03/2012
17/02/2012
27/11/2011
21/04/2012
20120824
20120317
20120217
20111127
20120421
Fri Aug 24 00:00:00 AEST 2012
Sat Mar 17 00:00:00 AEDT 2012
Fri Feb 17 00:00:00 AEDT 2012
Sun Nov 27 00:00:00 AEDT 2011
Sat Apr 21 00:00:00 AEST 2012
Area Name,Street Name,Between Street 1,Between Street 2,Side Of Street,Street Ma
```

7. To understand the script, we recommend using two windows, one for the editor and one for running the script. For each #exit

(a) Work out what the code immediately prior is attempting to do.

```
15:08
24/08/2012
17/03/2012
17/02/2012
27/11/2011
21/04/2012
20120824
20120317
20120217
20111127
20120421
Fri Aug 24 00:00:00 AEST 2012
Sat Mar 17 00:00:00 AEDT 2012
Fri Feb 17 00:00:00 AEDT 2012
Sun Nov 27 00:00:00 AEDT 2011
Sat Apr 21 00:00:00 AEST 2012
Area Name,Street Name,Between Street 1,Between Street 2,Side Of Street,Street Ma
rker,Arrival Time,Departure Time,Duration of Parking Event (
in seconds),Sign,in Violation?,Street ID,Device ID,DOW,Month,Qtime:Ru
shHour
Victoria Market,THERRY STREET,QUEEN STREET,ELIZABETH STREET,4,70218,24
/08/2012 11:34,24/08/2012 12:49,4473,LZ 15M M-SAT 7:30-19:30,1,1346,3/
70,Fri,24,late morning,not-rush-hour
Courtney,FEL STREET,O'CONNELL STREET,QUEENSBERRY STREET,2,5398E,17/03
/2012 13:07,17/03/2012 13:10,127,1/2P A RPE M-SUN 7:30-23:00,0,1101,34
72,Sat,17,afternoon,not-rush-hour
Victoria Market,FRANKLIN STREET,QUEEN STREET,ELIZABETH STREET,1,C6624,
17/02/2012 13:54,17/02/2012 14:20,1543,2P MTR M-SAT 7:30-20:30,0,681,2
805,Fri,17,afternoon,not-rush-hour
Chinatown,LONSDALE STREET,SWANSTON STREET,RUSSELL STREET,3,2888N,27/11
/2011 15:03,27/11/2011 16:02,3562,1P SUN 7:30-18:30,0,894,1770,Sun,27,
afternoon,not-rush-hour
10000 temp-col7.csv
10000 temp-datel.csv
10000 temp-date2.csv
10000 temp-date.csv
10001 temp-dow.csv
10000 temp-full-date.csv
10001 temp-month.csv
10000 temp-time24.csv
10000 temp-time.csv
10001 temp-time-qual01.csv
10001 temp-time-qual02.csv
10001 output.csv
120005 total
[e05051@csitprdap01:~/HDrive/DM]$
```

```
e05051@csitprdap01:~/HDrive/DM
GNU nano 2.3.1 File: parking-time.sh

#Mon,Oct,late-night,not-rush-hour
#Wed,Dec,early-morning,rush-hour
#Sun,Dec,late-morning,not-rush-hour

FILE='parking-small.csv'
head -5 $FILE
#exit

#Cut out the the arrival time field from line 2
tail -n +2 $FILE \
cut -d, -f7 > temp-col7.csv
head -5 temp-col7.csv
#exit

#cut out the date part
cut -d" " -f1 temp-col7.csv >temp-date.csv
head -5 temp-date.csv

#cut out the time field
```

(b) Remove the # and run the script.

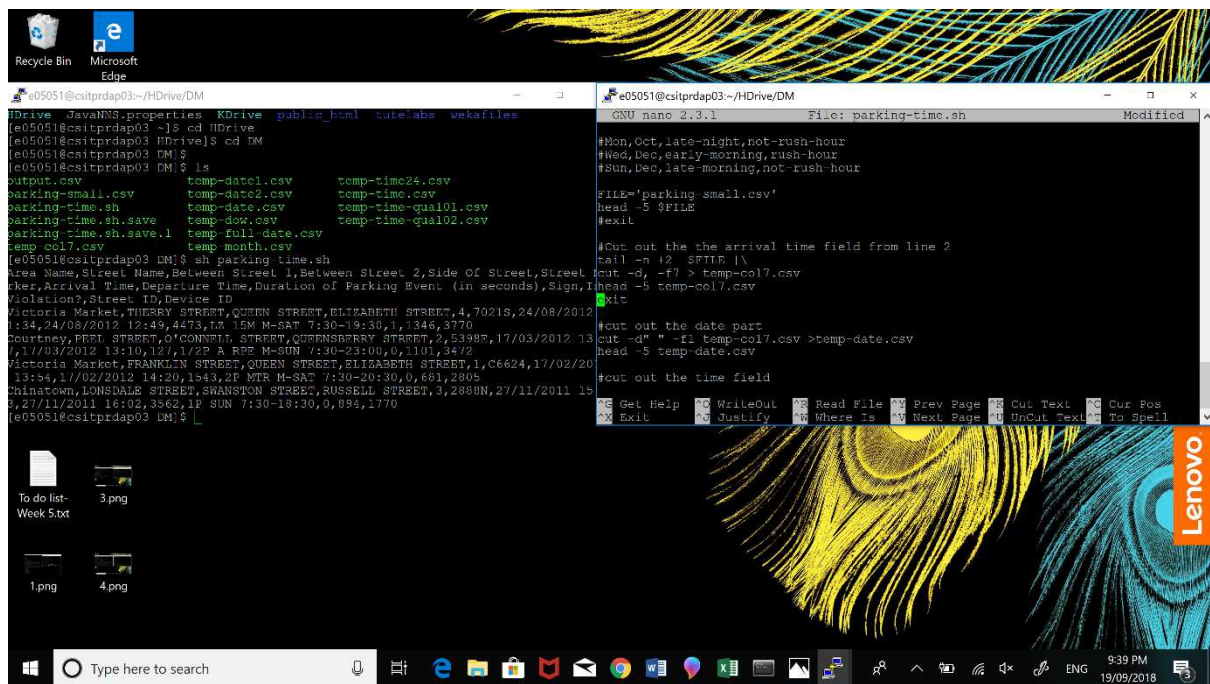
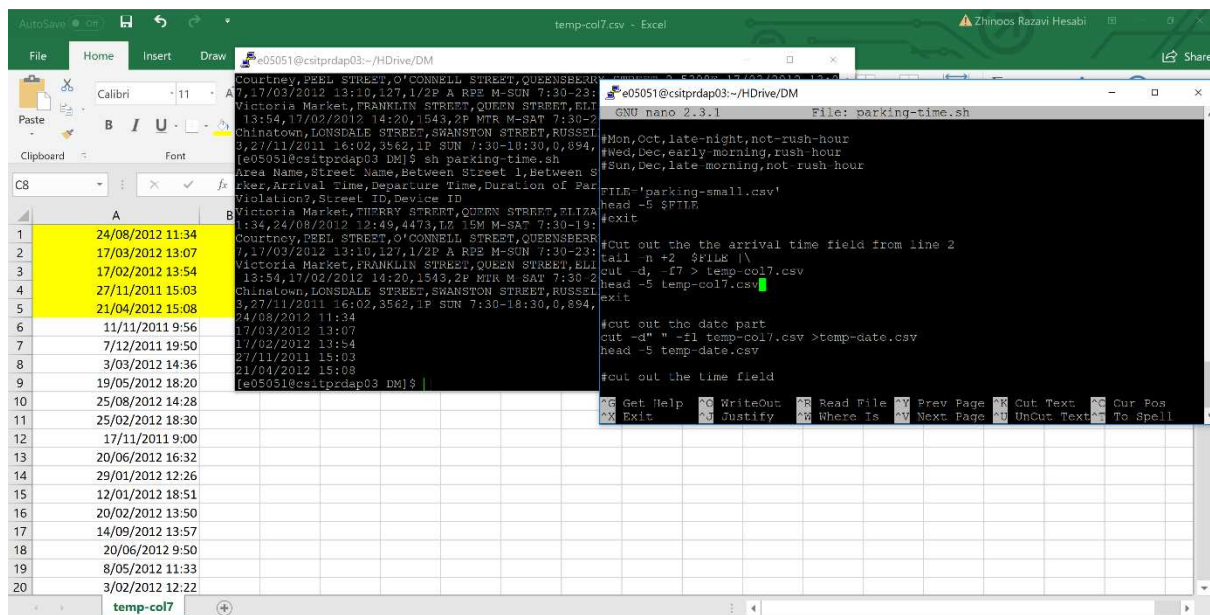


Figure . Removing # from the first exit and the output of running bash script after removing # in the left hand side.

(c) When it stops verify that the output and the temporary files written are what you expect.



As you can see from the above screenshot, the script exit after running the following set of commands written in the script:

#Cut out the the arrival time field from line 2

tail -n +2 \$FILE | \

cut -d, -f7 > temp-col7.csv

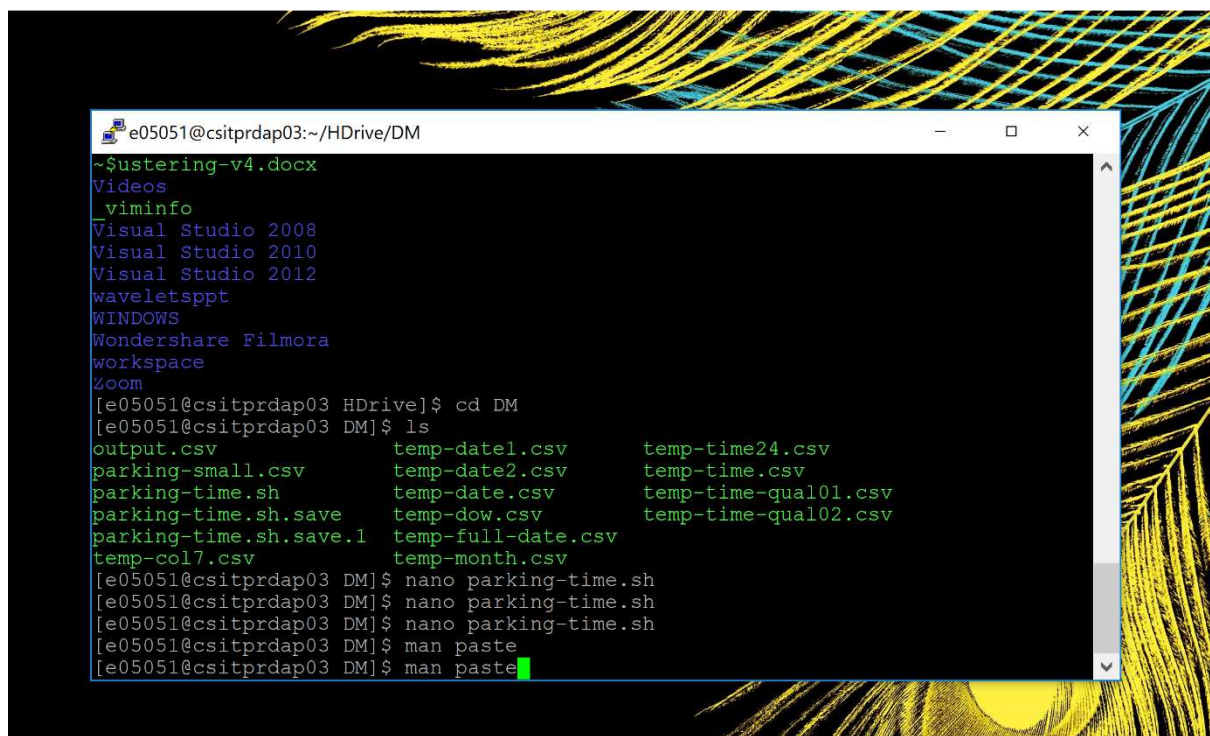
head -5 temp-col7.csv

exit

and write the results on temp-col7.csv. When you open the output file of temp-col7.csv, you can match the first 5 rows of csv file are the same as the output of the middle shell command in the above figure.

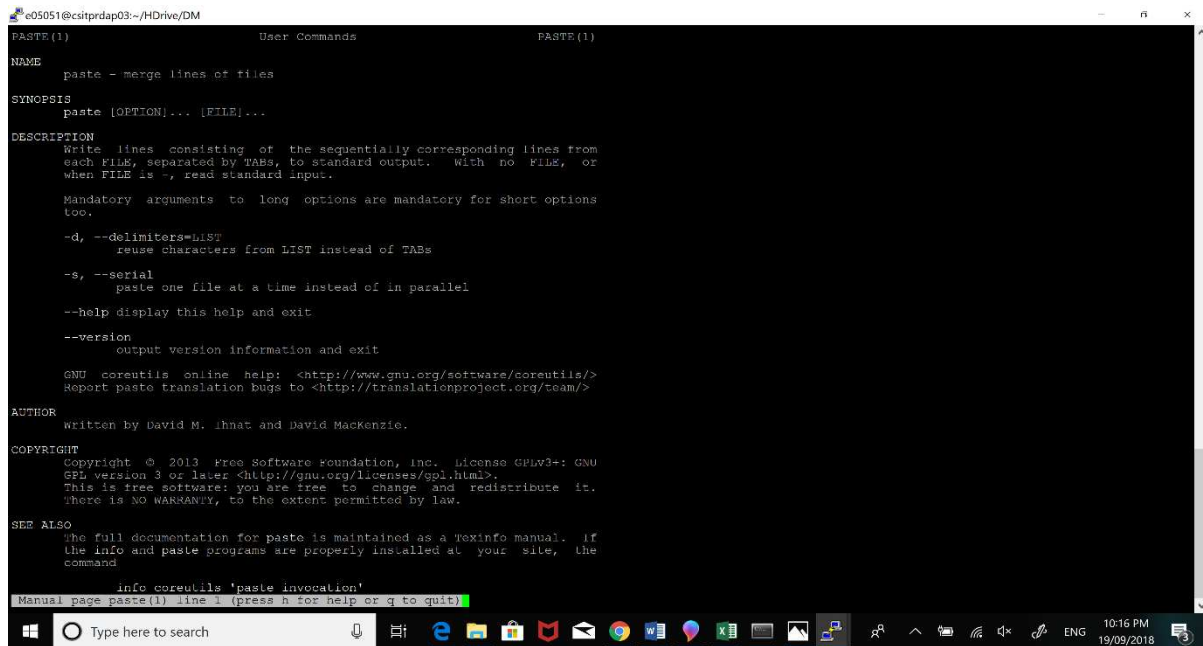
You can do the same for the rest of script to see what's happening within this script.

8. To get basic help on any program, use man, eg : man paste

A screenshot of a terminal window titled "e05051@csitprdap03:~/HDrive/DM". The terminal shows a directory listing of files and folders, including "Videos", "viminfo", "Visual Studio 2008", "Visual Studio 2010", "Visual Studio 2012", "wavelets ppt", "WINDOWS", "Wondershare Filmora", "workspace", and "Zoom". Below the listing, several commands are entered: "cd DM", "ls", "nano parking-time.sh", and "man paste". The output of "ls" shows a list of files and folders, including "output.csv", "temp-date1.csv", "temp-time24.csv", "parking-small.csv", "temp-date2.csv", "temp-time.csv", "parking-time.sh", "temp-date.csv", "temp-time-qual01.csv", "parking-time.sh.save", "temp-dow.csv", "temp-time-qual02.csv", "parking-time.sh.save.1", "temp-full-date.csv", "temp-col7.csv", and "temp-month.csv". The terminal window has a yellow and blue abstract background pattern on the right side.

```
e05051@csitprdap03:~/HDrive/DM
~$usterling-v4.docx
Videos
viminfo
Visual Studio 2008
Visual Studio 2010
Visual Studio 2012
wavelets ppt
WINDOWS
Wondershare Filmora
workspace
Zoom
[e05051@csitprdap03 HDrive]$ cd DM
[e05051@csitprdap03 DM]$ ls
output.csv          temp-date1.csv      temp-time24.csv
parking-small.csv   temp-date2.csv      temp-time.csv
parking-time.sh      temp-date.csv        temp-time-qual01.csv
parking-time.sh.save temp-dow.csv         temp-time-qual02.csv
parking-time.sh.save.1 temp-full-date.csv
temp-col7.csv        temp-month.csv
[e05051@csitprdap03 DM]$ nano parking-time.sh
[e05051@csitprdap03 DM]$ nano parking-time.sh
[e05051@csitprdap03 DM]$ nano parking-time.sh
[e05051@csitprdap03 DM]$ man paste
[e05051@csitprdap03 DM]$ man paste
```

after typing `man paste`, you need to press enter and you would get the following result:



```
e05051@csltrdp03: ~/HDrive/DM
PASTE(1)                                User Commands                                PASTE(1)

NAME
    paste - merge lines of files

SYNOPSIS
    paste [OPTION]... [FILE]...

DESCRIPTION
    Write lines consisting of the sequentially corresponding lines from
    each FILE, separated by TABs, to standard output. With no FILE, or
    when FILE is -, read standard input.

    Mandatory arguments to long options are mandatory for short options
    too.

    -d, --delimiters=LIST
        reuse characters from LIST instead of TABs

    -s, --serial
        paste one file at a time instead of in parallel

    --help
        display this help and exit

    --version
        output version information and exit

    GNU coreutils online help: <http://www.gnu.org/software/coreutils/>
    Report paste translation bugs to <http://translationproject.org/team/>

AUTHOR
    written by David M. Ihnat and David Mackenzie.

COPYRIGHT
    Copyright © 2013 Free Software Foundation, Inc. License GPLv3+: GNU
    GPL version 3 or later <http://gnu.org/licenses/gpl.html>.
    This is free software; you are free to change and redistribute it.
    There is NO WARRANTY, to the extent permitted by law.

SEE ALSO
    The full documentation for paste is maintained as a Texinfo manual. If
    the info and paste programs are properly installed at your site, the
    command

        info coreutils 'paste invocation'

    Manual page paste(1) line 1 (press h for help or q to quit)
```

9. Run the script, view the temporary files, look at where they are produced in the script and understand the that code that produces them.

10. Extend the script to generate a new column, “Weekday” in the output file where “yes” indicates a week day and “no” indicates a Saturday or Sunday.