Output_Part 1

```
start manager using: ./sbin/start-master.sh
start slave using: ./sbin/start-slave.sh spark://comp529-VirtualBox:7077
Execute the part1.py code using: ./bin/spark-submit --master local[*] part1.py
```

Initial Dataframe given

```
comp529@comp529-VirtualBox: /usr/local/spark
The Initial Dataframe given
                                        Latitude|
                                                                                       Longitude|AllZero|
                                                                                                                                                                                     Altitude|
                                                                                                                                                                                                                                               Timestamp|
                                                                                                                                                                                                                                                                                                                                       Datel
                                                                                                                                                                                                                                                                                                                                                                        Timel
             100 | 39.974408918 | 116.303522101 | 100 | 39.9743987078 | 116.303526932 | 100 | 39.973982524 | 116.303621837 | 100 | 39.973943291 | 116.303632641 | 100 | 39.973937148 | 116.303638648 | 100 | 39.973916715 | 116.30363848 | 100 | 39.973892264 | 116.303644867 | 100 | 39.973867401 | 116.303647142 | 100 | 39.973836462 | 116.30365019
                                                                                                                                                 0|480.287355643045|40753.5306944444|2011-07-29 00:00:00|12:44:12|0|480.121151574803|40753.5307060185|2011-07-29 00:00:00|12:44:13|0|478.499455380577|40753.5307291667|2011-07-29 00:00:00|12:44:15|
                                                                                                                                                0|479.176988188976|40753.5307407407|2011-07-29|00:00:00|12:44:16|
0|479.129432414698|40753.5307523148|2011-07-29|00:00:00|12:44:17|
0|479.615278871391|40753.530763889|2011-07-29|00:00:00|12:44:18|
0|480.506026902887|40753.530775463|2011-07-29|00:00:00|12:44:19|
0|481.38750984252|40753.530787037|2011-07-29|00:00:00|12:44:20|
             100 39.973836462 | 116.30365019 |
100 39.973821199 | 116.303649412 |
100 | 39.973807136 | 116.303642951 |
                                                                                                                                                 0|482.008727034121|40753.5307986111|2011-07-29 00:00:00|12:44:21
0|482.325816929134|40753.5308101852|2011-07-29 00:00:00|12:44:22
0|482.289422572178|40753.5308217593|2011-07-29 00:00:00|12:44:23
             100|39.973807136|116.303642951|
100|39.973791514|116.303626231|
100|39.973782219|116.303626231|
100|39.973774037|116.303619373|
100|39.9737740604|116.303612174|
100|39.973754251|116.303612502|
100|39.973736535|116.303615942|
100|39.973776454|116.303615942|
100|39.973776454|116.303615942|
100|39.973701448|116.303622536|
                                                                                                                                                0|482.314353674541|40753.5308333333]2011-07-29 00:00:00|12:44:24|
0|482.362713254593|40753.53083439074|2011-07-29 00:00:00|12:44:25|
0|482.472345800525|40753.5308564815|2011-07-29 00:00:00|12:44:26|
0|482.716797900262|40753.5308680556|2011-07-29 00:00:00|12:44:27|
0|482.782529527559|40753.5308796296|2011-07-29 00:00:00|12:44:28|
                                                                                                                                                 0|483.396486220472|40753.5308912037|2011-07-29 00:00:00|12:44:29|
0|483.396486220472|40753.5309027778|2011-07-29 00:00:00|12:44:30|
0|483.624166666667|40753.5309143519|2011-07-29 00:00:00|12:44:31|
              100|39.973701448|116.303622536
                                                                                                                                                  0 | 484.271414041995 | 40753.5309259259 | 2011-07-29 | 00:00:00 | 12:44:32
    nly showing top 20 rows
              UserID: integer (nullable = true)
Latitude: double (nullable = true)
Longitude: double (nullable = true)
Altitude: double (nullable = true)
Timestamp: double (nullable = true)
Date: timestamp (nullable = true)
Time: string (nullable = true)
```

Q1: Convert all dates/times from GMT to Beijing time

```
comp529@comp529-VirtualBox: /usr/local/spark
      **** Q1: The dataframe modified to Beijing time
                                                                                                                                                                                                                                                                                                                                                                                                                                                               Timestamp|
                                                                                     Latitude|
                                                                                                                                                                                 Longitude|
                                                                                                                                                                                                                                                                                                                                Altitude|
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        Date|
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            Time|added_hours|added_date| added_Timestamp|
                       100| 39.974408918|116.303522101| 480.287355643045| 40753.5306944444| 2011-07-29| 12:44:13| 100| 39.974397078| 116.30352932| 480.121151574803| 40753.5307604185| 2011-07-29| 12:44:13| 100| 39.973982524| 116.303621837| 478.499455380577| 40753.5307291667| 2011-07-29| 12:44:15| 100| 39.97394251| 116.303632641| 479.176988188976| 40753.5307407407| 2011-07-29| 12:44:16| 100| 39.973937148| 116.30363864| 479.129432414098| 40753.5307523148| 2011-07-29| 12:44:18| 100| 39.973916715| 116.30363848| 479.615278871391| 40753.530763889| 2011-07-29| 12:44:18| 100| 39.973892264| 116.303634867| 480.506026902887| 40753.5307536889| 2011-07-29| 12:44:18| 100| 39.973867401| 116.303647142| 481.38750984252| 40753.530787037| 2011-07-29| 12:44:20| 100| 39.973886462| 116.30365019| 482.008727034121| 40753.530787037| 2011-07-29| 12:44:21| 100| 39.9738821199| 116.303649412| 482.325816929134| 40753.530810852| 2011-07-29| 12:44:22| 100| 39.973873136| 116.303649412| 482.325816929134| 40753.530810852| 2011-07-29| 12:44:22| 100| 39.973873136| 116.303649612| 482.325816929134| 40753.530810852| 2011-07-29| 12:44:22| 100| 39.9738731514| 116.303638069| 482.314353674541| 40753.5308333333| 2011-07-29| 12:44:25| 100| 39.973782219| 116.303662231| 482.362713254593| 40753.5308849074| 2011-07-29| 12:44:25| 100| 39.973780437| 116.30363937| 482.472345800525| 40753.530866815| 2011-07-29| 12:44:25| 100| 39.973774037| 116.303619373| 482.472345800525| 40753.530866815| 2011-07-29| 12:44:25| 100| 39.973774037| 116.303619373| 482.472345800525| 40753.530866815| 2011-07-29| 12:44:25| 100| 39.973774037| 116.303619373| 482.472345800525| 40753.530866815| 2011-07-29| 12:44:26|
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        20:44:12 | 2011-07-29 | 40753.8640244444; 20:44:13 | 2011-07-29 | 40753.8640260185; 20:44:15 | 2011-07-29 | 40753.8640591667 | 20:44:15 | 2011-07-29 | 40753.8640591667 | 20:44:16 | 2011-07-29 | 40753.8640823148 | 20:44:18 | 2011-07-29 | 40753.8640938889 | 20:44:19 | 2011-07-29 | 40753.864105463 | 20:44:20 | 2011-07-29 | 40753.864117037 | 20:44:21 | 2011-07-29 | 40753.86411852 | 20:44:22 | 2011-07-29 | 40753.864157593 | 20:44:23 | 2011-07-29 | 40753.8641633333 | 20:44:25 | 2011-07-29 | 40753.864157593 | 20:44:25 | 2011-07-29 | 40753.864179074 | 20:44:26 | 2011-07-29 | 40753.864179074 | 20:44:26 | 2011-07-29 | 40753.864151852 | 20:44:25 | 2011-07-29 | 40753.8641583333 | 20:44:25 | 2011-07-29 | 40753.864179074 | 20:44:26 | 2011-07-29 | 40753.8641864815 |
                         100|39.973774037|116.303626231|482.302713254593|40753.53085449074|2011-07-29|12:44:25|
100|39.973774037|116.303619373|482.472345800525|40753.5308564815|2011-07-29|12:44:26|
100|39.973764604|116.303612174|482.716797900262|40753.5308680556|2011-07-29|12:44:26|
100|39.973754251|116.303615089|482.782529527559|40753.53089796296|2011-07-29|12:44:28|
100|39.973736535|116.303612592|483.086404199475|40753.5308912037|2011-07-29|12:44:29|
100|39.973776284|116.303615942|483.0396486229472|40753.5309927778|2011-07-29|12:44:31|
100|39.973717545|116.303614266|483.624166666667|40753.5309143519|2011-07-29|12:44:31|
100|39.973701448|116.303622536|484.271414041995|40753.5309259559|2011-07-29|12:44:32|
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            20:44:26 | 2011-07-29 | 40753 . 8641864815
20:44:27 | 2011-07-29 | 40753 . 8641980556
20:44:28 | 2011-07-29 | 40753 . 8642906296
20:44:29 | 2011-07-29 | 40753 . 8642212037
20:44:30 | 2011-07-29 | 40753 . 8642327778
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            20:44:31|2011-07-29|40753.8642443519
20:44:32|2011-07-29|40753.8642559259
        nly showing top 20 rows
                                                                                     Latitude|
                                                                                                                                                                                                                                                                                                                                Altitudel
|UserID|
                                                                                                                                                                          Longitudel
                                                                                                                                                                                                                                                                                                                                                                                                                                Timel
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Datel
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Timestampl
                       100 | 39.974408918 | 116.303522101 | 480.287355643045 | 20:44:12 | 2011-07-29 | 40753.8640244444 | 100 | 39.974397078 | 116.303526932 | 480.121151574803 | 20:44:13 | 2011-07-29 | 40753.8640360185 | 100 | 39.973982524 | 116.30362041 | 479.176988188976 | 20:44:15 | 2011-07-29 | 40753.86408591667 | 100 | 39.973937148 | 116.303632641 | 479.176988188976 | 20:44:16 | 2011-07-29 | 40753.8640823148 | 100 | 39.973937148 | 116.303639667 | 479.129432414698 | 20:44:17 | 2011-07-29 | 40753.8640823148 | 100 | 39.973937148 | 116.30363848 | 479.615278871391 | 20:44:18 | 2011-07-29 | 40753.8640823148 | 100 | 39.973892264 | 116.303648467 | 480.506026902887 | 20:44:19 | 2011-07-29 | 40753.864105463 | 100 | 39.973867401 | 116.303647142 | 481.38750984252 | 20:44:20 | 2011-07-29 | 40753.864105463 | 100 | 39.973867401 | 116.303654914 | 481.38750984252 | 20:44:20 | 2011-07-29 | 40753.8641186111 | 100 | 39.973807136 | 116.303649412 | 482.325816929134 | 20:44:21 | 2011-07-29 | 40753.8641186111 | 100 | 39.973807136 | 116.30364951 | 482.289422572178 | 20:44:22 | 2011-07-29 | 40753.864157593 | 100 | 39.973807136 | 116.303638069 | 482.31435674541 | 20:44:24 | 2011-07-29 | 40753.8641633333 | 100 | 39.973782219 | 116.303638691 | 482.362713254593 | 20:44:25 | 2011-07-29 | 40753.864173793 | 100 | 39.973782219 | 116.3036380691 | 482.362713254593 | 20:44:25 | 2011-07-29 | 40753.8641749074 | 100 | 39.973782219 | 116.303630631 | 482.362713254593 | 20:44:26 | 2011-07-29 | 40753.8641749074 | 100 | 39.973782219 | 116.303639373 | 482.362713254593 | 20:44:25 | 2011-07-29 | 40753.8641749074 | 100 | 39.973774037 | 116.303619373 | 482.362713254593 | 20:44:26 | 2011-07-29 | 40753.864184815 |
                                                        39.973774037|116.303619373|482.472345800525|20:44:26|2011-07-29|40753.8641864815
39.973764604|116.303612174|482.716797900262|20:44:27|2011-07-29|40753.8641980556
```

Q2: The number of days the data recorded for each person after converting to Beijing Time

```
comp529@comp529-VirtualBox: /usr/local/spark
**** Q2: The number of days the data recorded Number of unique users
 UserID|count|
                47 |
27 |
22 |
     102
103
     104|
105|
106|
               117
8
3
     108
109
     110 |
111 |
112 |
113 |
114 |
115 |
               23 |
36 |
109 |
19 |
16 |
116 |
     116
117
118
                 3|
6|
5|
      119
                46
 nly showing top 20 rows
The top 5 user IDs and their total number of days
|UserID|count|
               910|
               178
117
116
      126
     104|
115|
               109
 nly showing top 5 rows
```

Q3: The count for each user for each date when more than 100 data points were recorded

```
comp529@comp529-VirtualBox: /usr/local/spark
 **** Q3: The count for each dates when more than 100 data points recorded
|UserID|
                     Date
                                ctl
     112 | 2008 - 04 - 07 | 112 | 2008 - 06 - 28 | 115 | 2008 - 07 - 21 | 122 | 2009 - 09 - 18 |
                               1140
                               1877
                               2328
      128|2009-04-01|
128|2009-04
                                798
                                436
      126 | 2008 - 07 - 19 |
                                 766
      128 | 2007 - 07 - 07 |
                               1629
     128 | 2009 - 10 - 04 |
128 | 2009 - 10 - 29 |
128 | 2010 - 03 - 25 |
128 | 2010 - 10 - 29 |
                               5891
                               1039
                               1001
                                 753
      126 | 2008 - 05 - 04 |
      126 | 2008 - 08 - 10 | 11890
      128 | 2009 - 04 - 14 |
128 | 2009 - 11 - 04 |
                               1432
                               1430
      104|2008-09-07|
112|2008-04-26|
                                215
                                 170
      115 2008-05-29
      128 | 2008 - 10 - 28 |
only showing top 20 rows
The number of days for each user when more than 100 data points recorded
|UserID|count|
                 31
21
22
      101|
102|
      103
      104
                 86
      105
                 8 |
3 |
2 |
2 |
16 |
20 |
      106
      108
      109 |
110 |
                 98
                 19
15
      114
115
               100
      116
```

The highest number of days for each user(in ascending order of count)

```
comp529@comp529-VirtualBox: /usr/local/spark
The number of days for each user in ascending order of count
|UserID|count|
    128 l
           873|
    126
           171
    115
           100|
    112
            98|
    104
            86|
    125
            45
    119
    101
            31|
    122
    103
    130
only showing top 20 rows
```

Q4: The top 5 users who reached the highest Altitude and the date when they reached there for the first time

```
comp529@comp529-VirtualBox: /usr/local/spark
**** Q4: The highest Altitude users have reached
he top 5 highest Altitudes reached by users and the respectve dates
Idl
              MaxAlt|
                            Date
128|
            107503.3 | 2009-11-02 |
106|36581.3648293963|2007-10-09|
             25259.2 | 2008-09-12 |
103 l
101|
             24806.4 | 2008-03-28 |
             19432.4 | 2008-06-22 |
126|
nly showing top 5 rows
```

Traverse through the entire data and during each iteration perform the below:

- A new dataframe is created with all the data from that respective UserID using groupby function
- From this new dataframe, found the maximum altitude and created other dataframe with those values alone
- o In this dataframe, found the minimum date and store it to new dataframe
- Append the resultant dataframe for each UserIDs and display the output by sorting based on Altitude.

```
comp529@comp529-VirtualBox: /usr/local/spark
 **** 05: The timespan of the observation by users
                    cntl
|128| 1426.294375000005|
|114| 963.8455902778005|
|111| 838.7832175925942|
|115| 506.6880439814995|
|126| 325.831643518497|
|104|317.21797453710315|
|101| 212.6959490740992|
|112|175.85527777780226|
|125| 156.0722453703056|
122 | 142.05542824079748 |
|110|110.82624999999825|
|102| 75.07249999999476|
|119| 70.11480324080185|
|129| 56.27479166669946|
|103| 24.20990740739944|
|113|24.029664351794054|
|118|13.820243055604806|
|117|12.981909722198907|
|100|11.129791666695382|
|127| 9.506712962996971|
only showing top 20 rows
The top 5 observation time by users
I idl
                   cntl
|128|1426.294375000005|
|114|963.8455902778005|
|111|838.7832175925942|
|115|506.6880439814995|
|126| 325.831643518497|
only showing top 5 rows
```

Q6: The distance travelled by users

comp529@comp529-VirtualBox:~\$ sudo update-alternatives --config java [sudo] password for comp529:

There are 2 choices for the alternative java (providing /usr/bin/java).

| * 0 | /usr/lib/jvm/java-11-openjdk-amd64/bin/java | 1111 | auto mode |
|-----|--|------|-------------|
| 1 | /usr/lib/jvm/java-11-openjdk-amd64/bin/java | 1111 | manual mode |
| 2 | /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java | 1081 | manual mode |

Priority Status

Press <enter> to keep the current choice[*], or type selection number: 2

update-alternatives: using /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java to provide /usr/bin/java (java) in manual mode

comp529@comp529-VirtualBox:~\$ echo %JAVA_HOME%

%JAVA HOME%

Selection Path

comp529@comp529-VirtualBox:~\$

comp529@comp529-VirtualBox:~\$ export JAVA_HOME=/usr/

comp529@comp529-VirtualBox:~\$ cd downloads

bash: cd: downloads: No such file or directory

comp529@comp529-VirtualBox:~\$ cd Downloads

comp529@comp529-VirtualBox:~/Downloads\$ Is

spark-2.4.6-bin-hadoop2.7.tgz

comp529@comp529-VirtualBox:~/Downloads\$ tar -xzf spark-2.4.6-bin-hadoop2.7.tgz

comp529@comp529-VirtualBox:~/Downloads\$ sudo tar -xzf spark-2.4.6-bin-hadoop2.7.tgz

[sudo] password for comp529:

comp529@comp529-VirtualBox:~/Downloads\$ Is

spark-2.4.6-bin-hadoop2.7 spark-2.4.6-bin-hadoop2.7.tgz

comp529@comp529-VirtualBox:~/Downloads\$

comp529@comp529-VirtualBox:~/Downloads\$ sudo mv spark-2.4.6-bin-hadoop2.7 /usr/local/spark

comp529@comp529-VirtualBox:~/Downloads\$

comp529@comp529-VirtualBox:~/Downloads\$ cd /usr/local

comp529@comp529-VirtualBox:/usr/local\$ ls

bin etc games include lib man sbin share spark src

comp529@comp529-VirtualBox:/usr/local\$ whereis python3

python3: /usr/bin/python3 /usr/bin/python3.6m /usr/bin/python3.6 /usr/bin/python3.6-config

/usr/bin/python3.6m-config /usr/lib/python3 /usr/lib/python3.7 /usr/lib/python3.8 /usr/lib/python3.6 /etc/python3

/etc/python3.6 /usr/local/lib/python3.6 /usr/include/python3.6m /usr/include/python3.6 /usr/share/python3

/usr/share/man/man1/python3.1.gz

comp529@comp529-VirtualBox:/usr/local\$

comp529@comp529-VirtualBox:/usr/local\$

comp529@comp529-VirtualBox:/usr/local\$ sudo apt-get install python3-pip -y

Reading package lists... Done

Building dependency tree

Reading state information... Done

python3-pip is already the newest version (9.0.1-2.3~ubuntu1.18.04.5).

0 to upgrade, 0 to newly install, 0 to remove and 0 not to upgrade.

comp529@comp529-VirtualBox:/usr/local\$

comp529@comp529-VirtualBox:/usr/local\$

comp529@comp529-VirtualBox:/usr/local\$

comp529@comp529-VirtualBox:/usr/local\$ echo \$SPARK_HOME

```
/usr/local/spark
comp529@comp529-VirtualBox:/usr/local$
comp529@comp529-VirtualBox:/usr/local$ cd /usr/local/spark
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ ls
bin examples kubernetes logs R
                                                     yarn
conf input.txt LICENSE NOTICE README.md spark-2.4.6-bin-hadoop2.7
           licenses python RELEASE work
data jars
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ ls
bin examples kubernetes logs R
                                      sbin
                                                     yarn
conf input.txt LICENSE NOTICE README.md spark-2.4.6-bin-hadoop2.7
data jars
           licenses python RELEASE work
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ export SPARK_HOME=/usr/local/spark
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ export PATH=$PATH:$SPARK_HOME/bin
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ export PYSPARK_PYTHON=/usr/bin/python3
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ export PYSPARK_DRIVER_PYTHON=/usr/bin/python3
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ source~/.bashrc
bash: source~/.bashrc: No such file or directory
comp529@comp529-VirtualBox:/usr/local/spark$ source ~/.bashrc
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ ./sbin/start-master.sh
starting org.apache.spark.deploy.master.Master, logging to /usr/local/spark/logs/spark-comp529-
org.apache.spark.deploy.master.Master-1-comp529-VirtualBox.out
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ hostname
comp529-VirtualBox
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$./sbin/start-slave.sh spark://comp529-VirtualBox:7077
starting org.apache.spark.deploy.worker.Worker, logging to /usr/local/spark/logs/spark-comp529-
org.apache.spark.deploy.worker.Worker-1-comp529-VirtualBox.out
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:~$ cd /usr/local/spark
comp529@comp529-VirtualBox:/usr/local/spark$ pyspark
/usr/local/spark/bin/pyspark: line 45: python: command not found
Python 3.6.9 (default, Jan 26 2021, 15:33:00)
[GCC 8.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
22/01/07 13:02:47 WARN Utils: Your hostname, comp529-VirtualBox resolves to a loopback address: 127.0.1.1;
using 10.0.2.15 instead (on interface enp0s3)
22/01/07 13:02:47 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
22/01/07 13:02:48 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
```

To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel). Welcome to

Using Python version 3.6.9 (default, Jan 26 2021 15:33:00)

SparkSession available as 'spark'.

>>>

>>>

comp529@comp529-VirtualBox:/usr/local/spark\$

comp529@comp529-VirtualBox:/usr/local/spark\$./bin/spark-submit --master local[*] part2.py

22/01/10 14:57:33 WARN Utils: Your hostname, comp529-VirtualBox resolves to a loopback address: 127.0.1.1;

using 10.0.2.15 instead (on interface enp0s3)

22/01/10 14:57:33 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address

22/01/10 14:57:33 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties

22/01/10 14:57:34 INFO SparkContext: Running Spark version 2.4.6

22/01/10 14:57:34 INFO SparkContext: Submitted application: part2.py

22/01/10 14:57:34 INFO SecurityManager: Changing view acls to: comp529

22/01/10 14:57:34 INFO SecurityManager: Changing modify acls to: comp529

22/01/10 14:57:34 INFO SecurityManager: Changing view acls groups to:

22/01/10 14:57:34 INFO SecurityManager: Changing modify acls groups to:

22/01/10 14:57:34 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with

view permissions: Set(comp529); groups with view permissions: Set(); users with modify permissions:

Set(comp529); groups with modify permissions: Set()

22/01/10 14:57:35 INFO Utils: Successfully started service 'sparkDriver' on port 42531.

22/01/10 14:57:35 INFO SparkEnv: Registering MapOutputTracker

22/01/10 14:57:35 INFO SparkEnv: Registering BlockManagerMaster

22/01/10 14:57:35 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

22/01/10 14:57:35 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

22/01/10 14:57:35 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-fd135201-fec1-4a61-ba7d-1a7bc924102b

22/01/10 14:57:35 INFO MemoryStore: MemoryStore started with capacity 413.9 MB

22/01/10 14:57:35 INFO SparkEnv: Registering OutputCommitCoordinator

22/01/10 14:57:35 INFO Utils: Successfully started service 'SparkUI' on port 4040.

22/01/10 14:57:35 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://10.0.2.15:4040

22/01/10 14:57:36 INFO Executor: Starting executor ID driver on host localhost

22/01/10 14:57:36 INFO Utils: Successfully started service

'org.apache.spark.network.netty.NettyBlockTransferService' on port 46373.

22/01/10 14:57:36 INFO NettyBlockTransferService: Server created on 10.0.2.15:46373

22/01/10 14:57:36 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

22/01/10 14:57:36 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 10.0.2.15, 46373, None)

22/01/10 14:57:36 INFO BlockManagerMasterEndpoint: Registering block manager 10.0.2.15:46373 with 413.9 MB RAM, BlockManagerId(driver, 10.0.2.15, 46373, None)

22/01/10 14:57:36 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 10.0.2.15, 46373, None)

22/01/10 14:57:36 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 10.0.2.15, 46373, None) 22/01/10 14:57:37 INFO SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/usr/local/spark/spark-warehouse').

22/01/10 14:57:37 INFO SharedState: Warehouse path is 'file:/usr/local/spark/spark-warehouse'. 22/01/10 14:57:38 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint The Initial Dataframe given

| UserID | | | | Time |
|--|-----------------|--------------------|--------------|------------------------|
| 100 39.974408918 116.303522101 | | - | | 2011-07-29 |
| 00:00:00 12:44:12 | 01400424454 | F74002 407F2 F26 | 220004051 | 2044 07 20 |
| 100 39.974397078 116.303526932 | 0 480.121151 | 574803 40753.530 |)/060185 2 | 2011-07-29 |
| 00:00:00 12:44:13 | 01479 400455 | 200577140752 520 | 77016671 | 2011 07 20 |
| 100 39.973982524 116.303621837 00:00:00 12:44:15 | 0 4/8.499455 | 380577 40753.530 |)/29100/ 2 | 2011-07-29 |
| 100 39.973943291 116.303632641 | 01/70 176088 | 188976 40753.530 | 740740713 | 0011-07-20 |
| 00:00:00 12:44:16 | 0 473.170300 | 100570 40755.550 | 77407407 2 | 2011-07-23 |
| 100 39.973937148 116.303639667 | 01479 129432 | 414698 40753.530 | 752314812 | 2011-07-29 |
| 00:00:00 12:44:17 | 0 173.123 132 | 111030 10733.330 | 7732311012 | 1011 07 23 |
| 100 39.973916715 116.30363848 | 01479.6152788 | 71391 40753.530 | 763888912 | 011-07-29 |
| 00:00:00 12:44:18 | | | | |
| 100 39.973892264 116.303644867 | 01480.506026 | 902887 40753.53 | 077546312 | 011-07-29 |
| 00:00:00 12:44:19 | ., | , | | |
| 100 39.973867401 116.303647142 | 0 481.387509 | 84252 40753.530 | 787037 20 | 011-07-29 00:00:00 12: |
| 100 39.973836462 116.30365019 | • | 34121 40753.530 | • | · |
| 00:00:00 12:44:21 | · | · | • | |
| 100 39.973821199 116.303649412 | 0 482.325816 | 929134 40753.530 | 08101852 2 | 2011-07-29 |
| 00:00:00 12:44:22 | | | | |
| 100 39.973807136 116.303642951 | 0 482.289422 | 572178 40753.530 | 08217593 2 | 2011-07-29 |
| 00:00:00 12:44:23 | | | | |
| 100 39.973791514 116.303638069 | 0 482.314353 | 674541 40753.530 | 08333333 2 | 2011-07-29 |
| 00:00:00 12:44:24 | | | | |
| 100 39.973782219 116.303626231 | 0 482.362713 | 254593 40753.530 | 08449074 2 | 2011-07-29 |
| 00:00:00 12:44:25 | | | | |
| 100 39.973774037 116.303619373 | 0 482.472345 | 800525 40753.530 | 08564815 2 | 2011-07-29 |
| 00:00:00 12:44:26 | | | | |
| 100 39.973764604 116.303612174 | 0 482.716797 | 900262 40753.530 | 08680556 2 | 2011-07-29 |
| 00:00:00 12:44:27 | | | | |
| 100 39.973754251 116.303615089 | 0 482.782529 | 527559 40753.530 | 08796296 2 | 2011-07-29 |
| 00:00:00 12:44:28 | | | | |
| 100 39.973736535 116.303612592 | 0 483.086404 | 199475 40753.530 | 08912037 2 | 2011-07-29 |
| 00:00:00 12:44:29 | | | | |
| 100 39.973726284 116.303615942 | 0 483.396486 | 220472 40753.530 |)9027778 2 | 2011-07-29 |
| 00:00:00 12:44:30 | 01.00.00 | | | |
| 100 39.973717545 116.303614266 | 0 483.624166 | 666667 40753.530 |)9143519 2 | 2011-07-29 |
| 00:00:00 12:44:31 | 01404 274 44 4 | 044005140752 524 | 0025025012 | 2011 07 20 |
| 100 39.973701448 116.303622536 00:00:00 12:44:32 | U 484.2/1414 | 041995 40753.530 | 19259259 2 | 2011-07-29 |

```
|-- Latitude: double (nullable = true)
|-- Longitude: double (nullable = true)
|-- Altitude: double (nullable = true)
|-- Timestamp: double (nullable = true)
|-- Date: timestamp (nullable = true)
|-- Time: string (nullable = true)
**** Q1: The dataframe modified to Beijing time
|UserID| Latitude| Longitude| Altitude| Timestamp| Date| Time|added_hours|added_date|
added_Timestamp|
100|39.974408918|116.303522101|480.287355643045|40753.5306944444|2011-07-29|12:44:12|
20:44:12|2011-07-29|40753.8640244444|
100|39.974397078|116.303526932|480.121151574803|40753.5307060185|2011-07-29|12:44:13|
20:44:13 | 2011-07-29 | 40753.8640360185 |
100|39.973982524|116.303621837|478.499455380577|40753.5307291667|2011-07-29|12:44:15|
20:44:15 | 2011-07-29 | 40753.8640591667 |
100|39.973943291|116.303632641|479.176988188976|40753.5307407407|2011-07-29|12:44:16|
20:44:16|2011-07-29|40753.8640707407|
100|39.973937148|116.303639667|479.129432414698|40753.5307523148|2011-07-29|12:44:17|
20:44:17|2011-07-29|40753.8640823148|
100|39.973916715| 116.30363848|479.615278871391|40753.5307638889|2011-07-29|12:44:18|
20:44:18 | 2011-07-29 | 40753.8640938889 |
100|39.973892264|116.303644867|480.506026902887|40753.530775463|2011-07-29|12:44:19|
20:44:19|2011-07-29|40753.864105463|
100|39.973867401|116.303647142|481.38750984252|40753.530787037|2011-07-29|12:44:20|
20:44:20|2011-07-29|40753.864117037|
100|39.973836462|116.30365019|482.008727034121|40753.5307986111|2011-07-29|12:44:21|
20:44:21|2011-07-29|40753.8641286111|
100|39.973821199|116.303649412|482.325816929134|40753.5308101852|2011-07-29|12:44:22|
20:44:22|2011-07-29|40753.8641401852|
100|39.973807136|116.303642951|482.289422572178|40753.5308217593|2011-07-29|12:44:23|
20:44:23 | 2011-07-29 | 40753.8641517593 |
100|39.973791514|116.303638069|482.314353674541|40753.5308333333|2011-07-29|12:44:24|
20:44:24|2011-07-29|40753.8641633333|
100|39.973782219|116.303626231|482.362713254593|40753.5308449074|2011-07-29|12:44:25|
20:44:25 | 2011-07-29 | 40753.8641749074 |
100|39.973774037|116.303619373|482.472345800525|40753.5308564815|2011-07-29|12:44:26|
20:44:26|2011-07-29|40753.8641864815|
100|39.973764604|116.303612174|482.716797900262|40753.5308680556|2011-07-29|12:44:27|
20:44:27|2011-07-29|40753.8641980556|
100|39.973754251|116.303615089|482.782529527559|40753.5308796296|2011-07-29|12:44:28|
20:44:28 | 2011-07-29 | 40753.8642096296 |
100|39.973736535|116.303612592|483.086404199475|40753.5308912037|2011-07-29|12:44:29|
20:44:29|2011-07-29|40753.8642212037|
100|39.973726284|116.303615942|483.396486220472|40753.5309027778|2011-07-29|12:44:30|
20:44:30|2011-07-29|40753.8642327778|
100|39.973717545|116.303614266|483.624166666667|40753.5309143519|2011-07-29|12:44:31|
20:44:31|2011-07-29|40753.8642443519|
100|39.973701448|116.303622536|484.271414041995|40753.5309259259|2011-07-29|12:44:32|
20:44:32|2011-07-29|40753.8642559259|
```

```
|UserID| Latitude | Longitude | Altitude | Time | Date |
+-----+
100|39.974408918|116.303522101|480.287355643045|20:44:12|2011-07-29|40753.8640244444|
100|39.974397078|116.303526932|480.121151574803|20:44:13|2011-07-29|40753.8640360185|
 100|39.973982524|116.303621837|478.499455380577|20:44:15|2011-07-29|40753.8640591667|
  100|39.973943291|116.303632641|479.176988188976|20:44:16|2011-07-29|40753.8640707407|
  100|39.973937148|116.303639667|479.129432414698|20:44:17|2011-07-29|40753.8640823148|
 100|39.973916715| 116.30363848|479.615278871391|20:44:18|2011-07-29|40753.8640938889|
  100 | 39.973892264 | 116.303644867 | 480.506026902887 | 20:44:19 | 2011-07-29 | 40753.864105463 |
  100|39.973867401|116.303647142|481.38750984252|20:44:20|2011-07-29|40753.864117037|
  100|39.973836462|116.30365019|482.008727034121|20:44:21|2011-07-29|40753.8641286111|
  100 | 39.973821199 | 116.303649412 | 482.325816929134 | 20:44:22 | 2011-07-29 | 40753.8641401852 |
  100|39.973807136|116.303642951|482.289422572178|20:44:23|2011-07-29|40753.8641517593|
  100|39.973791514|116.303638069|482.314353674541|20:44:24|2011-07-29|40753.8641633333|
  100 | 39.973782219 | 116.303626231 | 482.362713254593 | 20:44:25 | 2011-07-29 | 40753.8641749074 |
  100 | 39.973774037 | 116.303619373 | 482.472345800525 | 20:44:26 | 2011-07-29 | 40753.8641864815 |
  100|39.973764604|116.303612174|482.716797900262|20:44:27|2011-07-29|40753.8641980556|
  100 | 39.973754251 | 116.303615089 | 482.782529527559 | 20:44:28 | 2011-07-29 | 40753.8642096296 |
  100|39.973736535|116.303612592|483.086404199475|20:44:29|2011-07-29|40753.8642212037|
  100|39.973726284|116.303615942|483.396486220472|20:44:30|2011-07-29|40753.8642327778|
  100 | 39.973717545 | 116.303614266 | 483.624166666667 | 20:44:31 | 2011-07-29 | 40753.8642443519 |
  100|39.973701448|116.303622536|484.271414041995|20:44:32|2011-07-29|40753.8642559259|
 .----+-----+----+-----+------+
only showing top 20 rows
**** Q6: The distance travelled by users
Traceback (most recent call last):
File "/usr/local/spark/part2.py", line 76, in <module>
  prnt(df.show(5))
NameError: name 'prnt' is not defined
comp529@comp529-VirtualBox:/usr/local/spark$ nano part2.py
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ ./bin/spark-submit --master local[*] part2.py
22/01/10 14:59:02 WARN Utils: Your hostname, comp529-VirtualBox resolves to a loopback address: 127.0.1.1;
using 10.0.2.15 instead (on interface enp0s3)
22/01/10 14:59:02 WARN Utils: Set SPARK LOCAL IP if you need to bind to another address
22/01/10 14:59:03 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
22/01/10 14:59:04 INFO SparkContext: Running Spark version 2.4.6
22/01/10 14:59:04 INFO SparkContext: Submitted application: part2.py
22/01/10 14:59:04 INFO SecurityManager: Changing view acls to: comp529
22/01/10 14:59:04 INFO SecurityManager: Changing modify acls to: comp529
22/01/10 14:59:04 INFO SecurityManager: Changing view acls groups to:
22/01/10 14:59:04 INFO SecurityManager: Changing modify acls groups to:
22/01/10 14:59:04 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with
view permissions: Set(comp529); groups with view permissions: Set(); users with modify permissions:
Set(comp529); groups with modify permissions: Set()
22/01/10 14:59:04 INFO Utils: Successfully started service 'sparkDriver' on port 43973.
```

22/01/10 14:59:04 INFO SparkEnv: Registering MapOutputTracker

22/01/10 14:59:04 INFO SparkEnv: Registering BlockManagerMaster

22/01/10 14:59:04 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information

22/01/10 14:59:04 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up

22/01/10 14:59:05 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-e9ee1ad7-6248-4f23-85cc-3728801e8535

22/01/10 14:59:05 INFO MemoryStore: MemoryStore started with capacity 413.9 MB

22/01/10 14:59:05 INFO SparkEnv: Registering OutputCommitCoordinator

22/01/10 14:59:05 INFO Utils: Successfully started service 'SparkUI' on port 4040.

22/01/10 14:59:05 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://10.0.2.15:4040

22/01/10 14:59:06 INFO Executor: Starting executor ID driver on host localhost

22/01/10 14:59:06 INFO Utils: Successfully started service

00:00:00|12:44:23|

'org.apache.spark.network.netty.NettyBlockTransferService' on port 46355.

22/01/10 14:59:06 INFO NettyBlockTransferService: Server created on 10.0.2.15:46355

22/01/10 14:59:06 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy

22/01/10 14:59:06 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 10.0.2.15, 46355, None)

22/01/10 14:59:06 INFO BlockManagerMasterEndpoint: Registering block manager 10.0.2.15:46355 with 413.9 MB RAM, BlockManagerId(driver, 10.0.2.15, 46355, None)

22/01/10 14:59:06 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 10.0.2.15, 46355, None)

22/01/10 14:59:06 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 10.0.2.15, 46355, None)

22/01/10 14:59:07 INFO SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/usr/local/spark/spark-warehouse').

22/01/10 14:59:07 INFO SharedState: Warehouse path is 'file:/usr/local/spark/spark-warehouse'.

22/01/10 14:59:08 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint The Initial Dataframe given

| | | + | _ | 1 | | | |
|--------|--------------------------|----------------------|----------------|------------------|------------|-------------|------------------|
| UserID | Latitude | Longitude AllZero | Altitude | Timestamp | Date | Time | |
| 100 39 | 9.974408918 | + 8 116.303522101 | | | | 2011-07-29 | |
| 100 39 | | 8 116.303526932 | 0 480.121151 | .574803 40753.53 | 07060185 2 | 2011-07-29 | |
| 100 39 | | 4 116.303621837 | 0 478.499455 | 380577 40753.53 | 07291667 2 | 2011-07-29 | |
| 100 39 | | 1 116.303632641 | 0 479.176988 | 3188976 40753.53 | 07407407 2 | 2011-07-29 | |
| 100 39 | | 8 116.303639667 | 0 479.129432 | 414698 40753.53 | 07523148 2 | 2011-07-29 | |
| 100 39 | | 5 116.30363848 | 0 479.615278 | 871391 40753.530 | 07638889 2 | 011-07-29 | |
| 100 39 | | 4 116.303644867 | 0 480.506026 | 902887 40753.53 | 30775463 2 | 011-07-29 | |
| | 12:44:19 9.97386740: | 1 116.303647142 | 0 481.387509 | 984252 40753.53 | 0787037 20 |)11-07-29 0 | 0:00:00 12:44:20 |
| | 9.973836462 12:44:21 | 2 116.30365019 | 0 482.008727 | 034121 40753.530 | 07986111 2 | 011-07-29 | |
| • | 9.973821199 12:44:22 | 9 116.303649412 | 0 482.325816 | 5929134 40753.53 | 08101852 2 | 2011-07-29 | |
| | | 6 116.303642951 | 0 482.289422 | 2572178 40753.53 | 08217593 2 | 2011-07-29 | |

| 100 39.973791514 116.303638069 | 0 482.314353674541 40753.5308333333 2011-07-29 |
|--------------------------------|--|
| 00:00:00 12:44:24 | |
| 100 39.973782219 116.303626231 | 0 482.362713254593 40753.5308449074 2011-07-29 |
| 00:00:00 12:44:25 | |
| 100 39.973774037 116.303619373 | 0 482.472345800525 40753.5308564815 2011-07-29 |
| 00:00:00 12:44:26 | |
| 100 39.973764604 116.303612174 | 0 482.716797900262 40753.5308680556 2011-07-29 |
| 00:00:00 12:44:27 | |
| 100 39.973754251 116.303615089 | 0 482.782529527559 40753.5308796296 2011-07-29 |
| 00:00:00 12:44:28 | |
| 100 39.973736535 116.303612592 | 0 483.086404199475 40753.5308912037 2011-07-29 |
| 00:00:00 12:44:29 | |
| 100 39.973726284 116.303615942 | 0 483.396486220472 40753.5309027778 2011-07-29 |
| 00:00:00 12:44:30 | |
| 100 39.973717545 116.303614266 | 0 483.624166666667 40753.5309143519 2011-07-29 |
| 00:00:00 12:44:31 | |
| 100 39.973701448 116.303622536 | 0 484.271414041995 40753.5309259259 2011-07-29 |
| 00:00:00 12:44:32 | |
| ++ | ++ |
| only showing top 20 rows | |

root

- |-- UserID: integer (nullable = true)
- |-- Latitude: double (nullable = true)
- |-- Longitude: double (nullable = true)
- |-- Altitude: double (nullable = true)
- |-- Timestamp: double (nullable = true)

20:44:19|2011-07-29|40753.864105463|

20:44:20|2011-07-29|40753.864117037|

- |-- Date: timestamp (nullable = true)
- |-- Time: string (nullable = true)

**** Q1: The dataframe modified to Beijing time |UserID| Latitude| Longitude| Altitude| Timestamp| Date| Time|added_hours|added_date| added_Timestamp| +----+ 100|39.974408918|116.303522101|480.287355643045|40753.5306944444|2011-07-29|12:44:12| 20:44:12|2011-07-29|40753.8640244444| | 100|39.974397078|116.303526932|480.121151574803|40753.5307060185|2011-07-29|12:44:13| 20:44:13|2011-07-29|40753.8640360185| 100|39.973982524|116.303621837|478.499455380577|40753.5307291667|2011-07-29|12:44:15| 20:44:15|2011-07-29|40753.8640591667| 100|39.973943291|116.303632641|479.176988188976|40753.5307407407|2011-07-29|12:44:16| 20:44:16|2011-07-29|40753.8640707407| 100|39.973937148|116.303639667|479.129432414698|40753.5307523148|2011-07-29|12:44:17| 20:44:17|2011-07-29|40753.8640823148| 100|39.973916715| 116.30363848|479.615278871391|40753.5307638889|2011-07-29|12:44:18| 20:44:18|2011-07-29|40753.8640938889| 100|39.973892264|116.303644867|480.506026902887|40753.530775463|2011-07-29|12:44:19|

| 100|39.973867401|116.303647142|481.38750984252|40753.530787037|2011-07-29|12:44:20|

```
100|39.973821199|116.303649412|482.325816929134|40753.5308101852|2011-07-29|12:44:22|
20:44:22|2011-07-29|40753.8641401852|
100|39.973807136|116.303642951|482.289422572178|40753.5308217593|2011-07-29|12:44:23|
20:44:23 | 2011-07-29 | 40753.8641517593 |
100|39.973791514|116.303638069|482.314353674541|40753.5308333333|2011-07-29|12:44:24|
20:44:24|2011-07-29|40753.8641633333|
100|39.973782219|116.303626231|482.362713254593|40753.5308449074|2011-07-29|12:44:25|
20:44:25|2011-07-29|40753.8641749074|
100|39.973774037|116.303619373|482.472345800525|40753.5308564815|2011-07-29|12:44:26|
20:44:26|2011-07-29|40753.8641864815|
100|39.973764604|116.303612174|482.716797900262|40753.5308680556|2011-07-29|12:44:27|
20:44:27 | 2011-07-29 | 40753.8641980556 |
100|39.973754251|116.303615089|482.782529527559|40753.5308796296|2011-07-29|12:44:28|
20:44:28|2011-07-29|40753.8642096296|
100|39.973736535|116.303612592|483.086404199475|40753.5308912037|2011-07-29|12:44:29|
20:44:29|2011-07-29|40753.8642212037|
100|39.973726284|116.303615942|483.396486220472|40753.5309027778|2011-07-29|12:44:30|
20:44:30 | 2011-07-29 | 40753.8642327778 |
\mid \ 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 116.303614266 \mid 483.624166666667 \mid 40753.5309143519 \mid 2011-07-29 \mid 12:44:31 \mid 100 \mid 39.973717545 \mid 39.973717545117545117545117545117545117545117541175411754117541175411754117541175411754117541175411754117541175411754117541175411754117541175
20:44:31|2011-07-29|40753.8642443519|
100|39.973701448|116.303622536|484.271414041995|40753.5309259259|2011-07-29|12:44:32|
20:44:32|2011-07-29|40753.8642559259|
only showing top 20 rows
|UserID| Latitude | Longitude | Altitude | Time | Date | Timestamp |
+-----+
   100|39.974408918|116.303522101|480.287355643045|20:44:12|2011-07-29|40753.8640244444|
100|39.974397078|116.303526932|480.121151574803|20:44:13|2011-07-29|40753.8640360185|
100|39.973982524|116.303621837|478.499455380577|20:44:15|2011-07-29|40753.8640591667|
100|39.973943291|116.303632641|479.176988188976|20:44:16|2011-07-29|40753.8640707407|
     100 \mid 39.973937148 \mid 116.303639667 \mid 479.129432414698 \mid 20:44:17 \mid 2011-07-29 \mid 40753.8640823148 \mid 20:44:17 
     100 | 39.973916715 | 116.30363848 | 479.615278871391 | 20:44:18 | 2011-07-29 | 40753.8640938889 |
     100|39.973892264|116.303644867|480.506026902887|20:44:19|2011-07-29|40753.864105463|
     100|39.973867401|116.303647142|481.38750984252|20:44:20|2011-07-29|40753.864117037|
     100 | 39.973836462 | 116.30365019 | 482.008727034121 | 20:44:21 | 2011-07-29 | 40753.8641286111 |
     100|39.973821199|116.303649412|482.325816929134|20:44:22|2011-07-29|40753.8641401852|
     100|39.973807136|116.303642951|482.289422572178|20:44:23|2011-07-29|40753.8641517593|
     100 | 39.973791514 | 116.303638069 | 482.314353674541 | 20:44:24 | 2011-07-29 | 40753.8641633333 |
     100|39.973782219|116.303626231|482.362713254593|20:44:25|2011-07-29|40753.8641749074|
     100|39.973774037|116.303619373|482.472345800525|20:44:26|2011-07-29|40753.8641864815|
     100 | 39.973764604 | 116.303612174 | 482.716797900262 | 20:44:27 | 2011-07-29 | 40753.8641980556 |
     100|39.973754251|116.303615089|482.782529527559|20:44:28|2011-07-29|40753.8642096296|
     100 | 39.973736535 | 116.303612592 | 483.086404199475 | 20:44:29 | 2011-07-29 | 40753.8642212037 |
     100 | 39.973726284 | 116.303615942 | 483.396486220472 | 20:44:30 | 2011-07-29 | 40753.8642327778 |
     100|39.973717545|116.303614266|483.624166666667|20:44:31|2011-07-29|40753.8642443519|
100|39.973701448|116.303622536|484.271414041995|20:44:32|2011-07-29|40753.8642559259|
          -+-----+
only showing top 20 rows
```

**** Q6: The distance travelled by users

+----+----

| ++ 100 2011-07-29 966 100 2011-07-30 1010 |
|---|
| 100 2011-07-31 483 100 2011-08-09 2670 100 2011-08-03 1128 ++ |
| None 1142 |
| ++ UserID Date count ++ |
| 101 2008-02-02 721 101 2008-06-29 2469 101 2008-03-13 95 101 2008-04-18 79 101 2007-12-15 112 |
| only showing top 5 rows |
| None 1142 |
| ++ UserID Date count ++ |
| 102 2011-11-01 156 102 2011-10-22 215 102 2011-12-24 16 102 2011-11-14 562 102 2011-11-11 288 ++ |
| None |
| 1142 ++ UserID Date count ++ |
| 103 2008-08-28 1579 103 2008-09-15 1473 103 2008-09-11 2401 103 2008-09-12 3236 103 2008-09-20 729 |
| only showing top 5 rows |
| None 1142 |
| ++ UserID Date count |

+----+

```
104 | 2008 - 03 - 24 | 19 |
  104 | 2008-09-07 | 215 |
| 104|2008-06-01| 274|
| 104|2008-05-09| 502|
 104 | 2008-10-05 | 1401 |
+----+
only showing top 5 rows
None
1142
+----+
|UserID| Date|count|
+----+
| 105|2007-10-05| 247|
 105 | 2007-10-03 | 390 |
| 105|2007-10-08| 153|
| 105|2007-10-07| 287|
| 105|2007-10-10| 220|
+----+
only showing top 5 rows
None
1142
+----+
|UserID| Date|count|
+----+
| 106|2007-10-10| 457|
| 106|2007-10-09| 776|
| 106|2007-10-08| 752|
+----+
None
1142
+----+
|UserID| Date|count|
+----+
| 107|2007-10-02| 81|
| 107|2007-10-07| 69|
| 107|2007-10-06| 49|
+----+
None
^CTraceback (most recent call last):
File "/usr/local/spark/part2.py", line 78, in <module>
  ct=df2.head()[0]
File "/usr/local/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.py", line 1255, in head
File "/usr/local/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.py", line 1257, in head
File "/usr/local/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.py", line 574, in take
File "/usr/local/spark/python/lib/pyspark.zip/pyspark/sql/dataframe.py", line 535, in collect
File "/usr/local/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 1255, in __call__
File "/usr/local/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 985, in send_command
File "/usr/local/spark/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 1152, in send_command
File "/usr/lib/python3.6/socket.py", line 586, in readinto
```

```
return self._sock.recv_into(b)
File "/usr/local/spark/python/lib/pyspark.zip/pyspark/context.py", line 270, in signal_handler
KeyboardInterrupt
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ cd /home/comp529/Desktop
comp529@comp529-VirtualBox:~/Desktop$
comp529@comp529-VirtualBox:~/Desktop$
comp529@comp529-VirtualBox:~/Desktop$ cp part1.py /usr/local/spark
comp529@comp529-VirtualBox:~/Desktop$
comp529@comp529-VirtualBox:~/Desktop$
comp529@comp529-VirtualBox:~/Desktop$ cd /usr/local/spark
comp529@comp529-VirtualBox:/usr/local/spark$ ls
             Desktop input.txt kubernetes licenses NOTICE part2.py R
bin data
                                                                            RELEASE work
conf dataset.txt examples jars
                                 LICENSE logs
                                                  part1.py python README.md sbin yarn
comp529@comp529-VirtualBox:/usr/local/spark$ ./bin/spark-submit --master local[*] part1.py
22/01/10 15:07:23 WARN Utils: Your hostname, comp529-VirtualBox resolves to a loopback address: 127.0.1.1;
using 10.0.2.15 instead (on interface enp0s3)
22/01/10 15:07:23 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
22/01/10 15:07:23 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-
java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
22/01/10 15:07:24 INFO SparkContext: Running Spark version 2.4.6
22/01/10 15:07:24 INFO SparkContext: Submitted application: part1.py
22/01/10 15:07:24 INFO SecurityManager: Changing view acls to: comp529
22/01/10 15:07:24 INFO SecurityManager: Changing modify acls to: comp529
22/01/10 15:07:24 INFO SecurityManager: Changing view acls groups to:
22/01/10 15:07:24 INFO SecurityManager: Changing modify acls groups to:
22/01/10 15:07:24 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with
view permissions: Set(comp529); groups with view permissions: Set(); users with modify permissions:
Set(comp529); groups with modify permissions: Set()
22/01/10 15:07:25 INFO Utils: Successfully started service 'sparkDriver' on port 40705.
22/01/10 15:07:25 INFO SparkEnv: Registering MapOutputTracker
22/01/10 15:07:25 INFO SparkEnv: Registering BlockManagerMaster
22/01/10 15:07:25 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for
getting topology information
22/01/10 15:07:25 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
22/01/10 15:07:25 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-de80f248-b1fc-4f1e-a2e4-
cbf59dd83a83
22/01/10 15:07:25 INFO MemoryStore: MemoryStore started with capacity 413.9 MB
22/01/10 15:07:25 INFO SparkEnv: Registering OutputCommitCoordinator
22/01/10 15:07:25 INFO Utils: Successfully started service 'SparkUI' on port 4040.
22/01/10 15:07:26 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://10.0.2.15:4040
22/01/10 15:07:26 INFO Executor: Starting executor ID driver on host localhost
22/01/10 15:07:26 INFO Utils: Successfully started service
'org.apache.spark.network.netty.NettyBlockTransferService' on port 36013.
22/01/10 15:07:26 INFO NettyBlockTransferService: Server created on 10.0.2.15:36013
22/01/10 15:07:26 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block
replication policy
```

22/01/10 15:07:26 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 10.0.2.15, 36013,

None)

22/01/10 15:07:26 INFO BlockManagerMasterEndpoint: Registering block manager 10.0.2.15:36013 with 413.9 MB RAM, BlockManagerId(driver, 10.0.2.15, 36013, None)

22/01/10 15:07:26 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 10.0.2.15, 36013, None)

22/01/10 15:07:26 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 10.0.2.15, 36013, None) 22/01/10 15:07:27 INFO SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/usr/local/spark/spark-warehouse').

22/01/10 15:07:27 INFO SharedState: Warehouse path is 'file:/usr/local/spark/spark-warehouse'. 22/01/10 15:07:28 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint The Initial Dataframe given

| ++ | + |
|---|---|
| UserID Latitude Longitude AllZero Altitu | ude Timestamp Date Time |
| ++ | |
| 100 39.974408918 116.303522101 0 480.28 | 87355643045 40753.5306944444 2011-07-29 |
| 00:00:00 12:44:12 | 24454574000 40750 5007050405 2044 07 20 |
| | 21151574803 40753.5307060185 2011-07-29 |
| 00:00:00 12:44:13 | 20455200577140752 520720466712044 07 20 |
| · · · · · · · · · · · · · · · · · · · | 99455380577 40753.5307291667 2011-07-29 |
| 00:00:00 12:44:15 100 39.973943291 116.303632641 0 479.1 | 76988188976 40753.5307407407 2011-07-29 |
| 00:00:00 12:44:16 | 70300100370 40733.3307407407 2011-07-23 |
| · | 29432414698 40753.5307523148 2011-07-29 |
| 00:00:00 12:44:17 | 25452414050 40755.5507525140 2011-07-25 |
| · | 5278871391 40753.5307638889 2011-07-29 |
| 00:00:00 12:44:18 | 32/00/1331 10/33.330/030003 2011 0/ 23 |
| · | 06026902887 40753.530775463 2011-07-29 |
| 00:00:00 12:44:19 | |
| · | 8750984252 40753.530787037 2011-07-29 00:00:00 12:44:20 |
| | 8727034121 40753.5307986111 2011-07-29 |
| 00:00:00 12:44:21 | |
| 100 39.973821199 116.303649412 0 482.33 | 25816929134 40753.5308101852 2011-07-29 |
| 00:00:00 12:44:22 | |
| 100 39.973807136 116.303642951 0 482.2 | 89422572178 40753.5308217593 2011-07-29 |
| 00:00:00 12:44:23 | |
| · · · · · · · · · · · · · · · · · · · | 14353674541 40753.5308333333 2011-07-29 |
| 00:00:00 12:44:24 | |
| | 62713254593 40753.5308449074 2011-07-29 |
| 00:00:00 12:44:25 | |
| · · · · · · · · · · · · · · · · · · · | 72345800525 40753.5308564815 2011-07-29 |
| 00:00:00 12:44:26 | 167070003631407F3 F309690FF613011 07 30 |
| 100 39.973764604 116.303612174 0 482.73 00:00:00 12:44:27 | 16797900262 40753.5308680556 2011-07-29 |
| · | 82529527559 40753.5308796296 2011-07-29 |
| 00:00:00 12:44:28 | 32323327333 40733.3300730230 2011 07 23 |
| · | 86404199475 40753.5308912037 2011-07-29 |
| 00:00:00 12:44:29 | 50 10 1255 175 1075515500512057 2022 07 25 |
| · | 96486220472 40753.5309027778 2011-07-29 |
| 00:00:00 12:44:30 | |
| · | 24166666667 40753.5309143519 2011-07-29 |
| 00:00:00 12:44:31 | |
| 100 39.973701448 116.303622536 0 484.2 | 71414041995 40753.5309259259 2011-07-29 |
| 00:00:00 12:44:32 | |
| ++ | + |

```
root
|-- UserID: integer (nullable = true)
|-- Latitude: double (nullable = true)
|-- Longitude: double (nullable = true)
|-- Altitude: double (nullable = true)
|-- Timestamp: double (nullable = true)
|-- Date: timestamp (nullable = true)
|-- Time: string (nullable = true)
|-- Time: String (nullable = true)
```

20:44:30|2011-07-29|40753.8642327778|

|-- Timestamp: double (nullable = true) **** Q1: The dataframe modified to Beijing time |UserID| Latitude| Longitude| Altitude| Timestamp| Date | Time | added_hours | added_date | added_Timestamp| 100|39.974408918|116.303522101|480.287355643045|40753.5306944444|2011-07-29|12:44:12| 20:44:12 | 2011-07-29 | 40753.8640244444 | 20:44:13 | 2011-07-29 | 40753.8640360185 | 100|39.973982524|116.303621837|478.499455380577|40753.5307291667|2011-07-29|12:44:15| 20:44:15 | 2011-07-29 | 40753.8640591667 | 100|39.973943291|116.303632641|479.176988188976|40753.5307407407|2011-07-29|12:44:16| 20:44:16|2011-07-29|40753.8640707407| 100|39.973937148|116.303639667|479.129432414698|40753.5307523148|2011-07-29|12:44:17| 20:44:17 | 2011-07-29 | 40753.8640823148 | 100|39.973916715| 116.30363848|479.615278871391|40753.5307638889|2011-07-29|12:44:18| 20:44:18|2011-07-29|40753.8640938889| 100|39.973892264|116.303644867|480.506026902887|40753.530775463|2011-07-29|12:44:19| 20:44:19|2011-07-29|40753.864105463| | 100|39.973867401|116.303647142| 481.38750984252| 40753.530787037|2011-07-29|12:44:20| 20:44:20|2011-07-29|40753.864117037| 100|39.973836462|116.30365019|482.008727034121|40753.5307986111|2011-07-29|12:44:21| 20:44:21 | 2011-07-29 | 40753.8641286111 | 100|39.973821199|116.303649412|482.325816929134|40753.5308101852|2011-07-29|12:44:22| 20:44:22|2011-07-29|40753.8641401852| 100|39.973807136|116.303642951|482.289422572178|40753.5308217593|2011-07-29|12:44:23| 20:44:23 | 2011-07-29 | 40753.8641517593 | 100|39.973791514|116.303638069|482.314353674541|40753.5308333333|2011-07-29|12:44:24| 20:44:24|2011-07-29|40753.8641633333| 100|39.973782219|116.303626231|482.362713254593|40753.5308449074|2011-07-29|12:44:25| 20:44:25|2011-07-29|40753.8641749074| 20:44:26|2011-07-29|40753.8641864815| 100|39.973764604|116.303612174|482.716797900262|40753.5308680556|2011-07-29|12:44:27| 20:44:27 | 2011-07-29 | 40753.8641980556 | 100|39.973754251|116.303615089|482.782529527559|40753.5308796296|2011-07-29|12:44:28| 20:44:28|2011-07-29|40753.8642096296| 100|39.973736535|116.303612592|483.086404199475|40753.5308912037|2011-07-29|12:44:29| 20:44:29|2011-07-29|40753.8642212037|

100|39.973726284|116.303615942|483.396486220472|40753.5309027778|2011-07-29|12:44:30|

```
100|39.973717545|116.303614266|483.624166666667|40753.5309143519|2011-07-29|12:44:31|
20:44:31|2011-07-29|40753.8642443519|
100|39.973701448|116.303622536|484.271414041995|40753.5309259259|2011-07-29|12:44:32|
20:44:32|2011-07-29|40753.8642559259|
+----+
only showing top 20 rows
|UserID| Latitude | Longitude | Altitude | Time | Date | Timestamp |
+-----+
100|39.974408918|116.303522101|480.287355643045|20:44:12|2011-07-29|40753.8640244444|
100|39.974397078|116.303526932|480.121151574803|20:44:13|2011-07-29|40753.8640360185|
100|39.973982524|116.303621837|478.499455380577|20:44:15|2011-07-29|40753.8640591667|
100|39.973943291|116.303632641|479.176988188976|20:44:16|2011-07-29|40753.8640707407|
    100 | 39.973937148 | 116.303639667 | 479.129432414698 | 20:44:17 | 2011-07-29 | 40753.8640823148 |
    100|39.973916715| 116.30363848|479.615278871391|20:44:18|2011-07-29|40753.8640938889|
    100|39.973892264|116.303644867|480.506026902887|20:44:19|2011-07-29| 40753.864105463|
   100|39.973867401|116.303647142| 481.38750984252|20:44:20|2011-07-29| 40753.864117037|
    100 \mid 39.973836462 \mid 116.30365019 \mid 482.008727034121 \mid 20:44:21 \mid 2011-07-29 \mid 40753.8641286111 \mid 20:44:21 \mid
    100|39.973821199|116.303649412|482.325816929134|20:44:22|2011-07-29|40753.8641401852|
    100 | 39.973807136 | 116.303642951 | 482.289422572178 | 20:44:23 | 2011-07-29 | 40753.8641517593 |
    100 | 39.973791514 | 116.303638069 | 482.314353674541 | 20:44:24 | 2011-07-29 | 40753.8641633333 |
    100|39.973782219|116.303626231|482.362713254593|20:44:25|2011-07-29|40753.8641749074|
   100|39.973774037|116.303619373|482.472345800525|20:44:26|2011-07-29|40753.8641864815|
    100|39.973764604|116.303612174|482.716797900262|20:44:27|2011-07-29|40753.8641980556|
    100|39.973754251|116.303615089|482.782529527559|20:44:28|2011-07-29|40753.8642096296|
    100|39.973736535|116.303612592|483.086404199475|20:44:29|2011-07-29|40753.8642212037|
    100|39.973726284|116.303615942|483.396486220472|20:44:30|2011-07-29|40753.8642327778|
    100|39.973717545|116.303614266|483.624166666667|20:44:31|2011-07-29|40753.8642443519|
    100|39.973701448|116.303622536|484.271414041995|20:44:32|2011-07-29|40753.8642559259|
+-----+
only showing top 20 rows
**** Q2: The number of days the data recorded
Number of unique users
+----+
|UserID|count|
+----+
| 100| 5|
| 101| 47|
| 102 | 27 |
| 103| 22|
```

104 | 117 | 105 | 8 | 106 | 3 | 107|

108

109|

| 110| 23| 111 | 36 | 112 | 109 | 113 | 19 | 114 | 16 |

3|

7|

4|

```
| 115| 116|
  116 3
| 117|
       6|
| 118| 5|
| 119| 46|
+----+
only showing top 20 rows
The top 5 user IDs and their total number of days
+----+
|UserID|count|
+----+
| 128| 910|
| 126| 178|
| 104| 117|
| 115| 116|
| 112| 109|
+----+
only showing top 5 rows
None
**** Q3: The count for each dates when more than 100 data points recorded
+----+
|UserID| Date| ct|
+----+
| 112|2008-04-07| 864|
| 112|2008-06-28| 1140|
| 115|2008-07-21|1877|
| 122|2009-09-18| 2328|
| 128|2008-12-08| 798|
| 128|2009-04-01| 436|
| 126|2008-07-19| 766|
| 128|2007-07-07| 1629|
| 128|2009-10-04|5891|
| 128|2009-10-29| 1039|
| 128|2010-03-25| 1001|
| 128|2010-10-29| 753|
| 126|2008-05-04| 709|
| 126|2008-08-10|11890|
| 128|2009-04-14| 1432|
| 128|2009-11-04| 1430|
| 104|2008-09-07| 215|
| 112|2008-04-26| 170|
| 115|2008-05-29|2890|
| 128|2008-10-28| 938|
+----+
only showing top 20 rows
The number of days for each user when more than 100 data points recorded
+----+
|UserID|count|
+----+
| 100| 5|
```

```
101| 31|
  102 | 21 |
 103 | 22 |
  104 | 86 |
  105
        8|
  106|
        3|
  108
        2|
  109|
        2|
  110 | 16 |
 111 | 20 |
| 112| 98|
 113 | 19 |
| 114| 15|
| 115| 100|
  116 | 3 |
| 118| 2|
| 119| 45|
| 120| 2|
| 121| 7|
+----+
only showing top 20 rows
The number of days for each user in ascending order of count
+----+
|UserID|count|
+----+
| 128| 873|
  126 | 171 |
 115 | 100 |
| 112| 98|
 104 | 86 |
  125 | 48 |
  119 | 45 |
| 101| 31|
| 122| 24|
| 103| 22|
| 102| 21|
| 130| 20|
  111 | 20 |
  113 | 19 |
| 110| 16|
| 114| 15|
  105|
        8|
| 121|
        7|
| 127| 7|
 124 6
+----+
only showing top 20 rows
**** Q4: The highest Altitude users have reached
```

```
+---+-----+
|128|
        107503.3 | 2009-11-02 |
|106|36581.3648293963|2007-10-09|
        25259.2 | 2008-09-12 |
|103|
|101|
        24806.4 | 2008-03-28 |
|126|
        19432.4 | 2008-06-22 |
+---+----+
only showing top 5 rows
None
**** Q5: The timespan of the observation by users
+---+
| id|
          cnt
+---+
|128| 1426.294375000005|
|114| 963.8455902778005|
|111| 838.7832175925942|
|115| 506.6880439814995|
|126| 325.831643518497|
|104|317.21797453710315|
|101| 212.6959490740992|
|112|175.85527777780226|
|125| 156.0722453703056|
|122|142.05542824079748|
|110|110.82624999999825|
|102| 75.07249999999476|
|119| 70.11480324080185|
|129| 56.27479166669946|
|103| 24.20990740739944|
|113|24.029664351794054|
|118|13.820243055604806|
|117|12.981909722198907|
|100|11.129791666695382|
|127| 9.506712962996971|
+---+-----+
only showing top 20 rows
The top 5 observation time by users
+---+
| id|
        cnt|
+---+-----+
|128|1426.294375000005|
|114|963.8455902778005|
|111|838.7832175925942|
|115|506.6880439814995|
|126| 325.831643518497|
+---+
only showing top 5 rows
None
**** Q6: The distance travelled by users
+----+
```

|UserID| Date|count|

+----+ 100 | 2011-07-29 | 966 | | 100|2011-07-30| 1010| | 100|2011-07-31| 483| | 100|2011-08-09| 2670| | 100|2011-08-03|1128| +----+ None Distinct dates 5 +----+ |UserID| Date|count| +----+ | 101|2008-02-02| 721| | 101|2008-06-29|2469| | 101|2008-03-13| 95| | 101|2008-04-18| 79| | 101|2007-12-15| 112| +----+ only showing top 5 rows None Distinct dates 47 +----+ |UserID| Date|count| +----+ | 102|2011-11-01| 156| | 102|2011-10-22| 215| | 102|2011-12-24| 16| | 102|2011-11-14| 562| | 102|2011-11-11| 288| +----+ only showing top 5 rows None Distinct dates 27 +----+ |UserID| Date|count| +----+ | 103|2008-08-28|1579| | 103|2008-09-15| 1473| | 103|2008-09-11|2401| | 103|2008-09-12|3236| | 103|2008-09-20| 729| +----+ only showing top 5 rows

```
+----+
|UserID| Date|count|
+----+
| 104|2008-03-24| 19|
| 104|2008-09-07| 215|
| 104|2008-06-01| 274|
| 104|2008-05-09| 502|
| 104|2008-10-05| 1401|
+----+
only showing top 5 rows
None
Distinct dates
117
+----+
|UserID| Date|count|
+----+
| 105|2007-10-05| 247|
| 105|2007-10-03| 390|
| 105|2007-10-08| 153|
| 105|2007-10-07| 287|
| 105|2007-10-10| 220|
+----+
only showing top 5 rows
None
Distinct dates
8
+----+
|UserID| Date|count|
+----+
| 106|2007-10-10| 457|
| 106|2007-10-09| 776|
| 106|2007-10-08| 752|
+----+
None
Distinct dates
+----+
|UserID| Date|count|
+----+
| 107|2007-10-02| 81|
| 107|2007-10-07| 69|
| 107|2007-10-06| 49|
+----+
None
Distinct dates
+----+
|UserID| Date|count|
+----+
```

```
| 108|2007-10-02| 20|
| 108|2007-10-03| 14|
| 108|2007-10-06| 103|
| 108|2007-10-04| 260|
| 108|2007-10-09| 10|
+----+
only showing top 5 rows
None
Distinct dates
7
+----+
|UserID| Date|count|
+----+
| 109|2007-12-01| 305|
| 109|2007-11-30| 2|
| 109|2007-12-03| 262|
| 109|2007-12-04| 26|
+----+
None
Distinct dates
4
+----+
|UserID| Date|count|
+----+
| 110|2007-12-02| 65|
| 110|2008-01-16| 9|
| 110|2008-01-19| 1041|
| 110|2007-12-06| 147|
| 110|2007-12-30| 194|
+----+
only showing top 5 rows
None
Distinct dates
23
+----+
|UserID| Date|count|
+----+
| 111|2007-04-16| 38|
| 111|2007-04-19| 134|
| 111|2009-07-28| 814|
| 111|2009-07-18| 58|
| 111|2007-04-13| 30|
+----+
only showing top 5 rows
None
Distinct dates
36
+----+
```

|UserID| Date|count|

+----+ | 112|2008-04-07| 864| | 112|2008-06-28| 1140| | 112|2008-04-26| 170| | 112|2008-07-02| 687| | 112|2008-06-05| 440| +----+ only showing top 5 rows None Distinct dates 109 +----+ |UserID| Date|count| +----+ | 113|2010-06-03| 1231| | 113|2010-06-02|1115| | 113|2010-05-25| 1222| | 113|2010-05-12| 565| | 113|2010-05-11| 4296| +----+ only showing top 5 rows None Distinct dates 19 +----+ |UserID| Date|count| +----+ | 114|2010-06-08| 761| | 114|2010-06-03| 731| | 114|2010-06-09| 579| | 114|2010-05-26| 873| | 114|2010-05-11| 1750| +----+ only showing top 5 rows None Distinct dates 16 +----+ |UserID| Date|count| +----+ | 115|2008-07-21|1877| | 115|2008-04-21| 23| | 115|2008-05-29|2890| | 115|2008-09-22| 1506| | 115|2008-06-04|2277| +----+ only showing top 5 rows

```
116
+----+
|UserID| Date|count|
+----+
| 116|2011-08-05| 812|
| 116|2011-08-03| 1014|
| 116|2011-08-02| 688|
+----+
None
Distinct dates
3
+----+
|UserID| Date|count|
+----+
| 117|2007-07-04| 67|
| 117|2007-06-21| 24|
| 117|2007-06-25| 64|
| 117|2007-06-29| 75|
| 117|2007-06-27| 47|
+----+
only showing top 5 rows
None
Distinct dates
6
+----+
|UserID| Date|count|
+----+
| 118|2007-05-20| 1086|
| 118|2007-05-21| 49|
| 118|2007-05-19| 3|
| 118|2007-05-26| 3|
| 118|2007-05-13| 219|
+----+
None
Distinct dates
+----+
|UserID| Date|count|
+----+
| 119|2008-08-21| 297|
| 119|2008-08-23|2442|
| 119|2008-10-06| 3413|
| 119|2008-07-30|2940|
| 119|2008-09-27| 2258|
+----+
only showing top 5 rows
None
```

```
+----+
|UserID| Date|count|
+----+
| 120|2009-09-19| 7445|
| 120|2009-09-20| 3345|
+----+
None
Distinct dates
+----+
|UserID| Date|count|
+----+
| 121|2009-10-09| 797|
| 121|2009-10-12| 527|
| 121|2009-10-04| 1168|
| 121|2009-10-11| 1930|
| 121|2009-10-10| 175|
+----+
only showing top 5 rows
None
Distinct dates
+----+
|UserID| Date|count|
+----+
| 122|2009-09-18| 2328|
| 122|2009-08-01| 632|
| 122|2009-09-11| 161|
| 122|2009-08-24| 2126|
| 122|2009-09-03|3423|
+----+
only showing top 5 rows
None
Distinct dates
24
+----+
|UserID| Date|count|
+----+
| 123|2009-09-24| 2875|
| 123|2009-09-28| 50|
| 123|2009-09-23| 216|
| 123|2009-09-27| 3527|
| 123|2009-09-25| 2160|
+----+
None
Distinct dates
5
+----+
|UserID| Date|count|
```

+----+ | 124|2008-10-07|3674| | 124|2008-10-03|23333| | 124|2008-10-01|39434| | 124|2008-10-02|45137| | 124|2008-10-04| 3189| +----+ only showing top 5 rows None Distinct dates 6 +----+ |UserID| Date|count| +----+ | 125|2008-09-03|3880| | 125|2008-05-24| 2815| | 125|2008-09-09|1608| | 125|2008-05-12| 114| | 125|2008-09-26| 408| +----+ only showing top 5 rows None Distinct dates 50 +----+ |UserID| Date|count| +----+ | 126|2008-07-19| 766| | 126|2008-05-04| 709| | 126|2008-08-10|11890| | 126|2008-09-12|3572| | 126|2008-04-22|4098| +----+ only showing top 5 rows None Distinct dates 178 +----+ |UserID| Date|count| +----+ | 127|2008-10-05| 435| | 127|2008-10-02|4505| | 127|2008-10-03| 2| | 127|2008-10-01|7202| | 127|2008-09-28| 90| +----+ only showing top 5 rows

| 9 |
|-------------------------|
| ++ |
| UserID Date count |
| ++ |
| 128 2007-07-07 1629 |
| 128 2008-12-08 798 |
| 128 2009-04-01 436 |
| 128 2009-10-04 5891 |
| 128 2009-10-29 1039 |
| ++ |
| only showing top 5 rows |