

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
+-----+-----+-----+-----+-----+-----+-----+-----+
|UserID| Latitude| Longitude|AllZero| Altitude| Timestamp| Date| Time|
+-----+-----+-----+-----+-----+-----+-----+-----+
| 100|39.974408918|116.303522101|0|480.287355643045|40753.5306944444|2011-07-29 00:00:00|12:44:12|
| 100|39.974397078|116.303526932|0|480.121151574803|40753.5307060185|2011-07-29 00:00:00|12:44:13|
| 100|39.973982524|116.303621837|0|478.499455380577|40753.5307291667|2011-07-29 00:00:00|12:44:15|
| 100|39.973943291|116.303632641|0|479.176988188976|40753.5307407407|2011-07-29 00:00:00|12:44:16|
| 100|39.973937148|116.303639667|0|479.129432414698|40753.5307523148|2011-07-29 00:00:00|12:44:17|
| 100|39.973916715|116.30363848|0|479.615278871391|40753.5307638889|2011-07-29 00:00:00|12:44:18|
| 100|39.973892264|116.303644867|0|480.506026902887|40753.530775463|2011-07-29 00:00:00|12:44:19|
| 100|39.973867401|116.303647142|0|481.38750984252|40753.530787037|2011-07-29 00:00:00|12:44:20|
| 100|39.973836462|116.30365019|0|482.008727034121|40753.5307986111|2011-07-29 00:00:00|12:44:21|
| 100|39.973821199|116.303649412|0|482.325816929134|40753.5308101852|2011-07-29 00:00:00|12:44:22|
| 100|39.973807136|116.303642951|0|482.289422572178|40753.5308217593|2011-07-29 00:00:00|12:44:23|
| 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)
|-- 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
+-----+-----+-----+-----+-----+-----+-----+-----+
|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|
+-----+-----+-----+-----+-----+-----+-----+-----+

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|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|
```

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|
+-----+-----+
| 100|    5|
| 101|   47|
| 102|   27|
| 103|   22|
| 104|  117|
| 105|    8|
| 106|    3|
| 107|    3|
| 108|    7|
| 109|    4|
| 110|   23|
| 111|   36|
| 112|  109|
| 113|   19|
| 114|   16|
| 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
```

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|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|
```

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| 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 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
the top 5 highest Altitudes reached by users and the respective dates
+-----+-----+-----+
|Id|          MaxAlt|      Date|
+-----+-----+-----+
|128|      107503.3|2009-11-02|
|106|36581.3648293963|2007-10-09|
|103|      25259.2|2008-09-12|
|101|      24806.4|2008-03-28|
|126|      19432.4|2008-06-22|
+-----+-----+-----+
only 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
- 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.

```
for p in range(100, 130):
    df7 = df6.filter((df6.UserID==p))
    q=df7.agg({'Altitude': 'max'})
    q1=q.head()[0]
    df8 = df7.filter((df6.Altitude==q1))
    val=df8.agg({'Date': 'min'})
    val=val.head()[0]
    newRow = spark.createDataFrame([(p,q1,val)], schema=schema)
    appended = df0.union(newRow)
    df1=appended
    df0=df1
```

Q5: The timespan of the observation by each user and the top 5 users with maximum observation time

```
comp529@comp529-VirtualBox: /usr/local/spark

**** 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| 175.8552777780226|
|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
```

Q6: The distance travelled by users

```
comp529@comp529-VirtualBox: /usr/local/spark

**** 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
```

```
comp529@comp529-VirtualBox:~$ sudo update-alternatives --config java
[sudo] password for comp529:
There are 2 choices for the alternative java (providing /usr/bin/java).
```

Selection	Path	Priority	Status
* 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

```
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%
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$ ls
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$ ls
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      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$ 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 PYSARK_PYTHON=/usr/bin/python3
comp529@comp529-VirtualBox:/usr/local/spark$
comp529@comp529-VirtualBox:/usr/local/spark$ export PYSARK_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".
```


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	Latitude	Longitude	AllZero	Altitude	Timestamp	Date	Time
+-----+-----+-----+-----+-----+-----+-----+-----+							
100	39.974408918	116.303522101	0	480.287355643045	40753.5306944444	2011-07-29	00:00:00 12:44:12
100	39.974397078	116.303526932	0	480.121151574803	40753.5307060185	2011-07-29	00:00:00 12:44:13
100	39.973982524	116.303621837	0	478.499455380577	40753.5307291667	2011-07-29	00:00:00 12:44:15
100	39.973943291	116.303632641	0	479.176988188976	40753.5307407407	2011-07-29	00:00:00 12:44:16
100	39.973937148	116.303639667	0	479.129432414698	40753.5307523148	2011-07-29	00:00:00 12:44:17
100	39.973916715	116.30363848	0	479.615278871391	40753.5307638889	2011-07-29	00:00:00 12:44:18
100	39.973892264	116.303644867	0	480.506026902887	40753.530775463	2011-07-29	00:00:00 12:44:19
100	39.973867401	116.303647142	0	481.38750984252	40753.530787037	2011-07-29	00:00:00 12:44:20
100	39.973836462	116.30365019	0	482.008727034121	40753.5307986111	2011-07-29	00:00:00 12:44:21
100	39.973821199	116.303649412	0	482.325816929134	40753.5308101852	2011-07-29	00:00:00 12:44:22
100	39.973807136	116.303642951	0	482.289422572178	40753.5308217593	2011-07-29	00:00:00 12:44:23
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)
-- 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

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|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 '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

+-----+-----+-----+-----+-----+-----+-----+-----+							
UserID	Latitude	Longitude	AllZero	Altitude	Timestamp	Date	Time
+-----+-----+-----+-----+-----+-----+-----+-----+							
100	39.974408918	116.303522101	0	480.287355643045	40753.5306944444	2011-07-29	00:00:00 12:44:12
100	39.974397078	116.303526932	0	480.121151574803	40753.5307060185	2011-07-29	00:00:00 12:44:13
100	39.973982524	116.303621837	0	478.499455380577	40753.5307291667	2011-07-29	00:00:00 12:44:15
100	39.973943291	116.303632641	0	479.176988188976	40753.5307407407	2011-07-29	00:00:00 12:44:16
100	39.973937148	116.303639667	0	479.129432414698	40753.5307523148	2011-07-29	00:00:00 12:44:17
100	39.973916715	116.30363848	0	479.615278871391	40753.5307638889	2011-07-29	00:00:00 12:44:18
100	39.973892264	116.303644867	0	480.506026902887	40753.530775463	2011-07-29	00:00:00 12:44:19
100	39.973867401	116.303647142	0	481.38750984252	40753.530787037	2011-07-29	00:00:00 12:44:20
100	39.973836462	116.30365019	0	482.008727034121	40753.5307986111	2011-07-29	00:00:00 12:44:21
100	39.973821199	116.303649412	0	482.325816929134	40753.5308101852	2011-07-29	00:00:00 12:44:22
100	39.973807136	116.303642951	0	482.289422572178	40753.5308217593	2011-07-29	00:00:00 12:44:23

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)
|-- 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|
```

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

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|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

```
+-----+-----+-----+
```

```
|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

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|
+-----+-----+-----+
```

only showing top 5 rows

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
bin  data      Desktop  input.txt  kubernetes  licenses  NOTICE  part2.py  R      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	Altitude	Timestamp	Date	Time
100	39.974408918	116.303522101	0	480.287355643045	40753.5306944444	2011-07-29	00:00:00 12:44:12
100	39.974397078	116.303526932	0	480.121151574803	40753.5307060185	2011-07-29	00:00:00 12:44:13
100	39.973982524	116.303621837	0	478.499455380577	40753.5307291667	2011-07-29	00:00:00 12:44:15
100	39.973943291	116.303632641	0	479.176988188976	40753.5307407407	2011-07-29	00:00:00 12:44:16
100	39.973937148	116.303639667	0	479.129432414698	40753.5307523148	2011-07-29	00:00:00 12:44:17
100	39.973916715	116.30363848	0	479.615278871391	40753.5307638889	2011-07-29	00:00:00 12:44:18
100	39.973892264	116.303644867	0	480.506026902887	40753.530775463	2011-07-29	00:00:00 12:44:19
100	39.973867401	116.303647142	0	481.38750984252	40753.530787037	2011-07-29	00:00:00 12:44:20
100	39.973836462	116.30365019	0	482.008727034121	40753.5307986111	2011-07-29	00:00:00 12:44:21
100	39.973821199	116.303649412	0	482.325816929134	40753.5308101852	2011-07-29	00:00:00 12:44:22
100	39.973807136	116.303642951	0	482.289422572178	40753.5308217593	2011-07-29	00:00:00 12:44:23
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)
|-- 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|
+-----+-----+-----+-----+-----+-----+-----+-----+
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|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
```

**** 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| 3|
| 108| 7|
| 109| 4|
| 110| 23|
| 111| 36|
| 112| 109|
| 113| 19|
| 114| 16|
```

	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

The top 5 highest Altitudes reached by users and the respective dates

+---+-----+-----+

| Id| MaxAlt| Date|

```

+---+-----+-----+
|128|    107503.3|2009-11-02|
|106|36581.3648293963|2007-10-09|
|103|    25259.2|2008-09-12|
|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

None

Distinct dates

22

+-----+-----+-----+		
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

3

+-----+-----+-----+		
UserID	Date	count
+-----+-----+-----+		
107	2007-10-02	81
107	2007-10-07	69
107	2007-10-06	49
+-----+-----+-----+		

None

Distinct dates

3

+-----+-----+-----+		
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

None

Distinct dates

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

5

+-----+-----+-----+		
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

Distinct dates

46

UserID	Date	count
120	2009-09-19	7445
120	2009-09-20	3345

None

Distinct dates

2

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

7

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

None

Distinct dates

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

None

Distinct dates