Big Data Project 1

Group Members:

Fernando Martinez
Kevin Fernandez
Nicholas Wai
Naseem Machlovi

NY Parking Violations

The NYC Department of Finance collects data on every parking ticket issued in NYC (10M per year!). This data is made publicly available to aid in ticket resolution and to guide policy-makers.

1. • When are tickets most likely to be issued?
2. • What are the most common years and types of cars to be ticketed?
3. • Where are tickets most commonly issued?
4. • Which color of the vehicle is most likely to get a ticket?

1. When are tickets most likely to be issued?

Conclusion: 08 AM is the time tickets are most likely to be issued.

MapReduce Process:

Mapper : Using Regular Expression we have filtered the data for the violation time which has A/P after the given time. Setting violation time as key and assigning a value of count 1 to corresponding each entry.

```
import sys
import re
rule = re.compile('[^\s]{4}[A|P]')
for line in sys.stdin:
 line=line.strip(',').split(',')
 line_len = len(line)
 if line_len ==43:
   vtime = line[19]
   match = rule.search(vtime)
   if match:
     vtime=vtime[:2]+vtime[-1:]# Only printing the Violation time
     print('%s\t%s' % (vtime, '1'))
 else:
    continue
```

Reducer: Passing the output from mapper which has only values for the violation time as key and counter as value. Sorting all the values with the same key and combining them by adding their counters gives us the final count for each key value.

```
import sys
from operator import itemgetter
dict_vtime_count = {}
for line in sys.stdin:
   vtime, count = line.split('\t',1)
   try:
      count = int(count)
      dict_vtime_count[vtime] = dict_vtime_count.get(vtime, 0) + count
   except ValueError:
      pass
```

```
sorted_dict_vtime_count = sorted(dict_vtime_count.items(),
key=itemgetter(1))[::-1]
for vtime, count in sorted_dict_vtime_count:
   print ('%s\t%s' % (vtime, count))
```

Test.sh: Only output the key value with maximum count value using head -1 in tesh.sh file on hadoop.

```
2022-04-06 03:15:25,473 INFO mapreduce.Job: Running job: job_1649214890424_0001
2022-04-06 03:15:36,660 INFO mapreduce.Job: Job job_1649214890424_0001 running in uber mode : false
2022-04-06 03:15:36,662 INFO mapreduce.Job:  map 0% reduce 0%
2022-04-06 03:16:11,979 INFO mapreduce.Job:  map 4% reduce 0%
2022-04-06 03:16:12,987 INFO mapreduce.Job:  map 9% reduce 0%
2022-04-06 03:16:18,025 INFO mapreduce.Job:  map 12% reduce 0%
2022-04-06 03:16:19,032 INFO mapreduce.Job:  map 17% reduce 0%
2022-04-06 03:16:23,061 INFO mapreduce.Job:  map 23% reduce 0%
2022-04-06 03:16:24,068 INFO mapreduce.Job:  map 26% reduce 0%
2022-04-06 03:16:25,074 INFO mapreduce.Job:  map 32% reduce 0%
2022-04-06 03:16:29,110 INFO mapreduce.Job:  map 39% reduce 0%
2022-04-06 03:16:30,130 INFO mapreduce.Job:  map 46% reduce 0%
2022-04-06 03:16:31,138 INFO mapreduce.Job:  map 57% reduce 0%
2022-04-06 03:16:35,182 INFO mapreduce.Job:  map 65% reduce 0%
2022-04-06 03:16:36,188 INFO mapreduce.Job:  map 66% reduce 0%
2022-04-06 03:16:41,228 INFO mapreduce.Job:  map 74% reduce 0%
2022-04-06 03:16:42,241 INFO mapreduce.Job:  map 76% reduce 0%
2022-04-06 03:16:44,253 INFO mapreduce.Job:  map 79% reduce 0%
2022-04-06 03:16:47,269 INFO mapreduce.Job:  map 82% reduce 0%
2022-04-06 03:16:48,275 INFO mapreduce.Job:  map 83% reduce 0%
2022-04-06 03:16:50,287 INFO mapreduce.Job:  map 95% reduce 0%
2022-04-06 03:16:51,292 INFO mapreduce.Job:  map 100% reduce 31%
2022-04-06 03:16:57,323 INFO mapreduce.Job:  map 100% reduce 70%
2022-04-06 03:17:03,356 INFO mapreduce.Job:  map 100% reduce 86%
2022-04-06 03:17:08,387 INFO mapreduce.Job:  map 100% reduce 100%
2022-04-06 03:17:09,400 INFO mapreduce.Job: Job job_1649214890424_0001 completed successfully
2022-04-06 03:17:09,513 INFO mapreduce.Job: Counters: 56
        File System Counters
                FILE: Number of bytes read=79764870
                FILE: Number of bytes written=163678362
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1884594574
                HDFS: Number of bytes written=393
                HDFS: Number of read operations=47
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Killed map tasks=2
                Launched map tasks=16
                Launched reduce tasks=1
                Data-local map tasks=15
                Rack-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=893438
                Total time spent by all reduces in occupied slots (ms)=36015
                Total time spent by all map tasks (ms)=893438
                Total time spent by all reduce tasks (ms)=36015
                Total vcore-milliseconds taken by all map tasks=893438
                Total vcore-milliseconds taken by all reduce tasks=36015
                Total megabyte-milliseconds taken by all map tasks=914880512
                Total megabyte-milliseconds taken by all reduce tasks=36879360
        Map-Reduce Framework
                Map input records=9980450
                Map output records=9970608
                Map output bytes=59823648
                Map output materialized bytes=79764948
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=46
                Reduce shuffle bytes=79764948
                Reduce input records=9970608
                Reduce output records=46
                Spilled Records=19941216
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=4211
                CPU time spent (ms)=138380
                Physical memory (bytes) snapshot=4534337536
                Virtual memory (bytes) snapshot=41843462144
```

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1884593202
        File Output Format Counters
                Bytes Written=393
2022-04-06 03:17:09,519 INFO streaming.StreamJob: Output directory: /Part1/output/
08A     917303
Deleted /Part1/input
Deleted /Part1/output
Stopping namenodes on [instance-0.c.big-data-339500.internal]
Stopping datanodes
Stopping secondary namenodes [instance-0]
Stopping nodemanagers
10.128.0.4: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
10.128.0.3: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager
WARNING: Use of this script to stop the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon stop" instead.
root@instance-0:/BigData_Project/part1/Part1#
```

2. What are the most common years and types of cars to be ticketed?

Conclusion: The combination of type of cars and its most common year is **SUBN 2020** has the maximum number of tickets issued during the whole fiscal year.

MapReduce Process:

Mapper : Using Regular Expression we have filtered the data where car details have alphabetical name and year has numeric value. Setting car description (**car_type+' '+year**) as key with value value 1 as counter.

```
import sys
import re
rule = re.compile('[A-Z]\s[^\s]{3}')
for line in sys.stdin:
  line=line.strip(',').split(',')
  line_len = len(line)
  if line_len ==43:
    car_type= line[6]
    year= line[35]
    car_descrp=car_type+' '+year
    match = rule.search(car_descrp)
    if match:
      print('%s\t%s' % (car_descrp, '1'))
    else:
      continue
```

Reducer: Passing the output from mapper which has only values for the car description as key and counter as value. Sorting all the values with the same key and combining them by adding their counters gives us the final count for each key value.

```
import sys
from operator import itemgetter
dic_car_count = {}
for line in sys.stdin:
  car_descp, count = line.split('\t',1)
  try:
    count = int(count)
    dic_car_count [car_descp] = dic_car_count .get(car_descp, 0) + count
  except ValueError:
    pass
sorted_dict_vtime_count = sorted(dic_car_count .items(), key=itemgetter(1))[::-1]
for car_descp, count in sorted_dict_vtime_count:
  print ('%s\t%s' % (car_descp, count))
```

Test.sh: Only output the key value with maximum count value using head -10 in tesh.sh file on hadoop. It shows **SUBN 2020** has the maximum number of tickets issued.

```
on_1649214067923_0001/
2022-04-06 03:01:42,591 INFO mapreduce.Job: Running job: job_1649214067923_0001
2022-04-06 03:01:52,824 INFO mapreduce.Job: Job job_1649214067923_0001 running in uber mode : false
2022-04-06 03:01:52,826 INFO mapreduce.Job:  map 0% reduce 0%
2022-04-06 03:02:28,146 INFO mapreduce.Job:  map 3% reduce 0%
2022-04-06 03:02:29,152 INFO mapreduce.Job:  map 8% reduce 0%
2022-04-06 03:02:34,184 INFO mapreduce.Job:  map 10% reduce 0%
2022-04-06 03:02:35,194 INFO mapreduce.Job:  map 16% reduce 0%
2022-04-06 03:02:39,221 INFO mapreduce.Job:  map 21% reduce 0%
2022-04-06 03:02:40,228 INFO mapreduce.Job:  map 24% reduce 0%
2022-04-06 03:02:41,236 INFO mapreduce.Job:  map 32% reduce 0%
2022-04-06 03:02:45,264 INFO mapreduce.Job:  map 37% reduce 0%
2022-04-06 03:02:46,270 INFO mapreduce.Job:  map 40% reduce 0%
2022-04-06 03:02:47,277 INFO mapreduce.Job:  map 43% reduce 0%
2022-04-06 03:02:48,310 INFO mapreduce.Job:  map 45% reduce 0%
2022-04-06 03:02:49,319 INFO mapreduce.Job:  map 50% reduce 0%
2022-04-06 03:02:50,326 INFO mapreduce.Job:  map 57% reduce 0%
2022-04-06 03:02:51,332 INFO mapreduce.Job:  map 62% reduce 0%
2022-04-06 03:02:52,340 INFO mapreduce.Job:  map 65% reduce 0%
2022-04-06 03:02:57,370 INFO mapreduce.Job:  map 71% reduce 0%
2022-04-06 03:02:58,377 INFO mapreduce.Job:  map 75% reduce 0%
2022-04-06 03:03:03,404 INFO mapreduce.Job:  map 81% reduce 0%
2022-04-06 03:03:04,410 INFO mapreduce.Job:  map 83% reduce 0%
2022-04-06 03:03:05,416 INFO mapreduce.Job:  map 86% reduce 0%
2022-04-06 03:03:09,444 INFO mapreduce.Job:  map 88% reduce 0%
2022-04-06 03:03:10,450 INFO mapreduce.Job:  map 95% reduce 19%
2022-04-06 03:03:11,456 INFO mapreduce.Job:  map 100% reduce 19%
2022-04-06 03:03:16,481 INFO mapreduce.Job:  map 100% reduce 68%
2022-04-06 03:03:22,524 INFO mapreduce.Job:  map 100% reduce 88%
2022-04-06 03:03:26,543 INFO mapreduce.Job:  map 100% reduce 100%
2022-04-06 03:03:26,551 INFO mapreduce.Job: Job job_1649214067923_0001 completed successfully
2022-04-06 03:03:26,645 INFO mapreduce.Job: Counters: 56
        File System Counters
                FILE: Number of bytes read=111348983
                FILE: Number of bytes written=226846573
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1884594574
                HDFS: Number of bytes written=57298
                HDFS: Number of read operations=47
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Killed map tasks=2
                Launched map tasks=16
                Launched reduce tasks=1
                Data-local map tasks=15
                Rack-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=931746
                Total time spent by all reduces in occupied slots (ms)=35013
                Total time spent by all map tasks (ms)=931746
                Total time spent by all reduce tasks (ms)=35013
                Total vcore-milliseconds taken by all map tasks=931746
                Total vcore-milliseconds taken by all reduce tasks=35013
                Total megabyte-milliseconds taken by all map tasks=954107904
                Total megabyte-milliseconds taken by all reduce tasks=35853312
        Map-Reduce Framework
                Map input records=9980450
                Map output records=8120608
                Map output bytes=95107761
                Map output materialized bytes=111349061
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=4891
                Reduce shuffle bytes=111349061
                Reduce input records=8120608
                Reduce output records=4891
                Spilled Records=16241216
                Shuffled Maps =14
```

```
        Map-Reduce Framework
                Map input records=9980450
                Map output records=8120608
                Map output bytes=95107761
                Map output materialized bytes=111349061
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=4891
                Reduce shuffle bytes=111349061
                Reduce input records=8120608
                Reduce output records=4891
                Spilled Records=16241216
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=3904
                CPU time spent (ms)=147900
                Physical memory (bytes) snapshot=4564312064
                Virtual memory (bytes) snapshot=41948041216
                Total committed heap usage (bytes)=3051356160
                Peak Map Physical memory (bytes)=335028224
                Peak Map Virtual memory (bytes)=2837843968
                Peak Reduce Physical memory (bytes)=287166464
                Peak Reduce Virtual memory (bytes)=2818715648
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1884593202
        File Output Format Counters
                Bytes Written=57298
2022-04-06 03:03:26,646 INFO streaming.StreamJob: Output directory: /Part2/output/
SUBN 2020       445567
SUBN 2021       442986
SUBN 2019       436701
SUBN 2018       312857
SUBN 2017       252550
4DSD 2017       206021
SUBN 2016       197947
4DSD 2018       192873
4DSD 2019       192509
SUBN 2015       185126
cat: Unable to write to output stream.
Deleted /Part2/input
Deleted /Part2/output
Stopping namenodes on [instance-0.c.big-data-339500.internal]
Stopping datanodes
Stopping secondary namenodes [instance-0]
Stopping nodemanagers
10.128.0.3: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
10.128.0.4: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager
WARNING: Use of this script to stop the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon stop" instead.
root@instance-0:/BigData_Project/part1/Part2# cd /BigData_Project/part1/
root@instance-0:/BigData_Project/part1# cd Part1
```

3. Where are tickets most commonly issued?

Conclusion: The final output shows that **Broadway** has the maximum number of tickets issued that is 119050.

MapReduce Process:

Mapper : Using regex we have filtering the data where streets have alpha numeric names and setting it as key of value value 1 as counter.

```
import sys
import re
rule = re.compile('[A-Za-z0-9|/|-]')
for line in sys.stdin:
 line=line.strip(',').split(',')
 line_len = len(line)
 if line_len ==43:
   street= line[24]
   match = rule.search(street)
   if  match:
     print('%s\t%s' % ( street, '1'))
 else:
   continue
```

Reducer: Passing the output from mapper which has only values for the street name as key and counter as value. Sorting all the values with the same key and combining them by adding their counters gives us the final count for each key value.

```
import sys
from operator import itemgetter
dic_street_count = {}
for line in sys.stdin:
   street, count = line.split('\t',1)
   try:
      count = int(count)
      dic_street_count [street] = dic_street_count .get(street, 0) + count
   except ValueError:
      pass
sorted_dict_vtime_count = sorted(dic_street_count .items(), key=itemgetter(1))[::-1]
```

```
for street, count in sorted_dict_vtime_count:
   print ('%s\t%s' % (street, count))
```

Test.sh: Only output the key value with maximum count value using head -10 in the tesh.sh file on hadoop. It shows **Broadway** has the maximum number of tickets issued that is 119050.

```
2022-04-06 03:25:24,134 INFO conf.Configuration: resource-types.xml not found
2022-04-06 03:25:24,137 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-04-06 03:25:24,697 INFO impl.YarnClientImpl: Submitted application application_1649215488661_0001
2022-04-06 03:25:24,754 INFO mapreduce.Job: The url to track the job: http://instance-0:8088/proxy/applicati
on_1649215488661_0001/
2022-04-06 03:25:24,756 INFO mapreduce.Job: Running job: job_1649215488661_0001
2022-04-06 03:25:35,102 INFO mapreduce.Job: Job job_1649215488661_0001 running in uber mode : false
2022-04-06 03:25:35,103 INFO mapreduce.Job:  map 0% reduce 0%
2022-04-06 03:26:12,423 INFO mapreduce.Job:  map 3% reduce 0%
2022-04-06 03:26:13,430 INFO mapreduce.Job:  map 8% reduce 0%
2022-04-06 03:26:18,462 INFO mapreduce.Job:  map 11% reduce 0%
2022-04-06 03:26:19,469 INFO mapreduce.Job:  map 17% reduce 0%
2022-04-06 03:26:21,483 INFO mapreduce.Job:  map 19% reduce 0%
2022-04-06 03:26:22,491 INFO mapreduce.Job:  map 23% reduce 0%
2022-04-06 03:26:24,504 INFO mapreduce.Job:  map 25% reduce 0%
2022-04-06 03:26:25,510 INFO mapreduce.Job:  map 33% reduce 0%
2022-04-06 03:26:27,522 INFO mapreduce.Job:  map 36% reduce 0%
2022-04-06 03:26:28,529 INFO mapreduce.Job:  map 41% reduce 0%
2022-04-06 03:26:31,552 INFO mapreduce.Job:  map 42% reduce 0%
2022-04-06 03:26:32,558 INFO mapreduce.Job:  map 43% reduce 0%
2022-04-06 03:26:33,565 INFO mapreduce.Job:  map 45% reduce 0%
2022-04-06 03:26:34,574 INFO mapreduce.Job:  map 52% reduce 0%
2022-04-06 03:26:35,591 INFO mapreduce.Job:  map 59% reduce 0%
2022-04-06 03:26:36,598 INFO mapreduce.Job:  map 66% reduce 0%
2022-04-06 03:26:39,621 INFO mapreduce.Job:  map 68% reduce 0%
2022-04-06 03:26:40,626 INFO mapreduce.Job:  map 73% reduce 0%
2022-04-06 03:26:41,637 INFO mapreduce.Job:  map 76% reduce 0%
2022-04-06 03:26:45,659 INFO mapreduce.Job:  map 77% reduce 0%
2022-04-06 03:26:46,667 INFO mapreduce.Job:  map 79% reduce 0%
2022-04-06 03:26:47,678 INFO mapreduce.Job:  map 81% reduce 0%
2022-04-06 03:26:50,695 INFO mapreduce.Job:  map 83% reduce 0%
2022-04-06 03:26:54,718 INFO mapreduce.Job:  map 95% reduce 0%
2022-04-06 03:26:55,724 INFO mapreduce.Job:  map 100% reduce 0%
2022-04-06 03:26:56,729 INFO mapreduce.Job:  map 100% reduce 35%
2022-04-06 03:27:02,761 INFO mapreduce.Job:  map 100% reduce 70%
2022-04-06 03:27:08,803 INFO mapreduce.Job:  map 100% reduce 85%
2022-04-06 03:27:14,839 INFO mapreduce.Job:  map 100% reduce 100%
2022-04-06 03:27:15,852 INFO mapreduce.Job: Job job_1649215488661_0001 completed successfully
2022-04-06 03:27:15,966 INFO mapreduce.Job: Counters: 56
        File System Counters
                FILE: Number of bytes read=188842462
                FILE: Number of bytes written=381833531
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1884594574
                HDFS: Number of bytes written=660549
                HDFS: Number of read operations=47
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Killed map tasks=1
                Launched map tasks=15
                Launched reduce tasks=1
                Data-local map tasks=13
                Rack-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=969034
                Total time spent by all reduces in occupied slots (ms)=37094
                Total time spent by all map tasks (ms)=969034
                Total time spent by all reduce tasks (ms)=37094
                Total vcore-milliseconds taken by all map tasks=969034
                Total vcore-milliseconds taken by all reduce tasks=37094
                Total megabyte-milliseconds taken by all map tasks=992290816
                Total megabyte-milliseconds taken by all reduce tasks=37984256
        Map-Reduce Framework
                Map input records=9980450
                Map output records=9969859
                Map output bytes=168902738
                Map output materialized bytes=188842540
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
```

```
        Map-Reduce Framework
                Map input records=9980450
                Map output records=9969859
                Map output bytes=168902738
                Map output materialized bytes=188842540
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=39280
                Reduce shuffle bytes=188842540
                Reduce input records=9969859
                Reduce output records=39279
                Spilled Records=19939718
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=4404
                CPU time spent (ms)=159330
                Physical memory (bytes) snapshot=4770906112
                Virtual memory (bytes) snapshot=41959186432
                Total committed heap usage (bytes)=3349151744
                Peak Map Physical memory (bytes)=340594688
                Peak Map Virtual memory (bytes)=2821439488
                Peak Reduce Physical memory (bytes)=345702400
                Peak Reduce Virtual memory (bytes)=2826993664
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1884593202
        File Output Format Counters
                Bytes Written=660549
2022-04-06 03:27:15,967 INFO streaming.StreamJob: Output directory: /Part3/output/
Broadway        119050
3rd Ave 86994
SB MAIN ST @ 37TH AV    59782
5th Ave 54508
EB W 14TH STREET @ 5    52180
2nd Ave 51536
Madison Ave     51055
Lexington Ave   47930
WB W 181ST ST @ AUDU    46282
1st Ave 39152
```

4 . Which color of the vehicle is most likely to get a ticket?

Conclusion: The final output shows that **GY, BK,** and **WH,** are the most likely colors to get a ticket as they are the most .

MapReduce Process:

Mapper : Using regex we have filtering the data where colors names and setting it as key of value value 1 as counter.

```
import sys
import re
rule = re.compile('[A-Za-z]')
for line in sys.stdin:
 line=line.strip(',').split(',')
 line_len = len(line)
 if line_len == 43:
   color= line[33]
   match = rule.search(color)
   if  match:
     print('%s\t%s' % ( color, '1'))
 else:
     Continue
```

Reducer: Passing the output from mapper which has only values for the color as key and counter as value. Sorting all the values with the same key and combining them by adding their counters gives us the final count for each key value.

```
import sys
from operator import itemgetter
dic_color_count = {}
total_count=0
for line in sys.stdin:
   color, count = line.split('\t',1)
   try:
      count = int(count)
      dic_color_count [color] = dic_color_count .get(color, 0) + count
      total_count+=count
```

```
    except ValueError:
        pass
sorted_dict_color_count = sorted(dic_color_count .items(), key=itemgetter(1))[::-1]
for color, count in sorted_dict_color_count:
    print ('%s\t%s' % (color, count/total_count))
```

Test.sh: Only output the key value with maximum count value using head -10 in tesh.sh file on hadoop.

```
on_1649216321944_0001/
2022-04-06 03:39:15,780 INFO mapreduce.Job: Running job: job_1649216321944_0001
2022-04-06 03:39:26,973 INFO mapreduce.Job: Job job_1649216321944_0001 running in uber mode : false
2022-04-06 03:39:26,974 INFO mapreduce.Job:  map 0% reduce 0%
2022-04-06 03:40:04,283 INFO mapreduce.Job:  map 5% reduce 0%
2022-04-06 03:40:05,289 INFO mapreduce.Job:  map 8% reduce 0%
2022-04-06 03:40:10,324 INFO mapreduce.Job:  map 13% reduce 0%
2022-04-06 03:40:11,330 INFO mapreduce.Job:  map 16% reduce 0%
2022-04-06 03:40:12,336 INFO mapreduce.Job:  map 18% reduce 0%
2022-04-06 03:40:13,343 INFO mapreduce.Job:  map 23% reduce 0%
2022-04-06 03:40:16,364 INFO mapreduce.Job:  map 26% reduce 0%
2022-04-06 03:40:17,371 INFO mapreduce.Job:  map 33% reduce 0%
2022-04-06 03:40:18,377 INFO mapreduce.Job:  map 35% reduce 0%
2022-04-06 03:40:19,383 INFO mapreduce.Job:  map 41% reduce 0%
2022-04-06 03:40:22,411 INFO mapreduce.Job:  map 44% reduce 0%
2022-04-06 03:40:23,421 INFO mapreduce.Job:  map 53% reduce 0%
2022-04-06 03:40:24,427 INFO mapreduce.Job:  map 60% reduce 0%
2022-04-06 03:40:25,434 INFO mapreduce.Job:  map 67% reduce 0%
2022-04-06 03:40:30,471 INFO mapreduce.Job:  map 69% reduce 0%
2022-04-06 03:40:31,476 INFO mapreduce.Job:  map 74% reduce 0%
2022-04-06 03:40:32,483 INFO mapreduce.Job:  map 77% reduce 0%
2022-04-06 03:40:37,514 INFO mapreduce.Job:  map 81% reduce 0%
2022-04-06 03:40:38,519 INFO mapreduce.Job:  map 83% reduce 0%
2022-04-06 03:40:39,525 INFO mapreduce.Job:  map 88% reduce 0%
2022-04-06 03:40:40,531 INFO mapreduce.Job:  map 100% reduce 0%
2022-04-06 03:40:44,551 INFO mapreduce.Job:  map 100% reduce 62%
2022-04-06 03:40:50,582 INFO mapreduce.Job:  map 100% reduce 79%
2022-04-06 03:40:56,612 INFO mapreduce.Job:  map 100% reduce 91%
2022-04-06 03:40:58,624 INFO mapreduce.Job:  map 100% reduce 100%
2022-04-06 03:40:59,639 INFO mapreduce.Job: Job job_1649216321944_0001 completed successfully
2022-04-06 03:40:59,767 INFO mapreduce.Job: Counters: 56
        File System Counters
                FILE: Number of bytes read=70948103
                FILE: Number of bytes written=146044813
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1884594574
                HDFS: Number of bytes written=6910
                HDFS: Number of read operations=47
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Killed map tasks=1
                Launched map tasks=15
                Launched reduce tasks=1
                Data-local map tasks=13
                Rack-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=899878
                Total time spent by all reduces in occupied slots (ms)=33442
                Total time spent by all map tasks (ms)=899878
                Total time spent by all reduce tasks (ms)=33442
                Total vcore-milliseconds taken by all map tasks=899878
                Total vcore-milliseconds taken by all reduce tasks=33442
                Total megabyte-milliseconds taken by all map tasks=921475072
                Total megabyte-milliseconds taken by all reduce tasks=34244608
        Map-Reduce Framework
                Map input records=9980450
                Map output records=9287002
                Map output bytes=52374093
                Map output materialized bytes=70948181
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=942
                Reduce shuffle bytes=70948181
                Reduce input records=9287002
                Reduce output records=942
                Spilled Records=18574004
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
```

```
        Map-Reduce Framework
                Map input records=9980450
                Map output records=9287002
                Map output bytes=52374093
                Map output materialized bytes=70948181
                Input split bytes=1372
                Combine input records=0
                Combine output records=0
                Reduce input groups=942
                Reduce shuffle bytes=70948181
                Reduce input records=9287002
                Reduce output records=942
                Spilled Records=18574004
                Shuffled Maps =14
                Failed Shuffles=0
                Merged Map outputs=14
                GC time elapsed (ms)=4507
                CPU time spent (ms)=140400
                Physical memory (bytes) snapshot=4491882496
                Virtual memory (bytes) snapshot=41851625472
                Total committed heap usage (bytes)=3251634176
                Peak Map Physical memory (bytes)=321630208
                Peak Map Virtual memory (bytes)=2819063808
                Peak Reduce Physical memory (bytes)=279449600
                Peak Reduce Virtual memory (bytes)=2820636672
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1884593202
        File Output Format Counters
                Bytes Written=6910
2022-04-06 03:40:59,767 INFO streaming.StreamJob: Output directory: /Part4/output/
GY      1942304
WH      1777470
BK      1700794
BL      653521
WHITE   652620
Deleted /Part4/input
Deleted /Part4/output
Stopping namenodes on [instance-0.c.big-data-339500.internal]
Stopping datanodes
Stopping secondary namenodes [instance-0]
Stopping nodemanagers
10.128.0.4: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
10.128.0.3: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager
WARNING: Use of this script to stop the MR JobHistory daemon is deprecated.
WARNING: Attempting to execute replacement "mapred --daemon stop" instead.
root@instance-0:/BigData_Project/part1/Part4#
```