

Task 1:

```
pi2018_nyu_edu@nyu-dataproc-m:~$ ls
AB_NYC_2019.csv HW1 HW2 HW3 NeighborhoodCount.java NeighborhoodCountMapper.java NeighborhoodCountReducer.java
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath` -d . MaxTemperatureMapper.java
javac: file not found: MaxTemperatureMapper.java
Usage: javac <options> <source files>
use -help for a list of possible options
pi2018_nyu_edu@nyu-dataproc-m:~$ 
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath` -d . NeighborhoodCountMapper.java
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath` -d . NeighborhoodCountReducer.java
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath` -d . NeighborhoodCount.java
Note: NeighborhoodCount.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
pi2018_nyu_edu@nyu-dataproc-m:~$ ls
AB_NYC_2019.csv HW2 NeighborhoodCount.class NeighborhoodCountMapper.class NeighborhoodCountReducer.class
HW1 HW3 NeighborhoodCount.java NeighborhoodCountMapper.java NeighborhoodCountReducer.java
pi2018_nyu_edu@nyu-dataproc-m:~$ jar -cvf maxNeighborhoodCount.jar *.class
added manifest
adding: NeighborhoodCount.class(in = 1449) (out= 824) (deflated 43%)
adding: NeighborhoodCountMapper.class(in = 1678) (out= 669)(deflated 60%)
adding: NeighborhoodCountReducer.class(in = 1618) (out= 676) (deflated 58%)
pi2018_nyu_edu@nyu-dataproc-m:~$ ls
AB_NYC_2019.csv HW2 NeighborhoodCount.class NeighborhoodCountMapper.class NeighborhoodCountReducer.class maxNeighborhoodCount.jar
HW1 HW3 NeighborhoodCount.java NeighborhoodCountMapper.java NeighborhoodCountReducer.java
```

```
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -put AB_NYC_2019.csv input_hw4
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -ls
Found 9 items
drwxr-xr-x  - pi2018_nyu_edu pi2018_nyu_edu          0 2023-09-29 01:05 HW2
-rw-r--r--  1 pi2018_nyu_edu pi2018_nyu_edu 1092 2023-09-21 16:41 MaxTemperatureMapper.java
drwxrwxr-x+ - pi2018_nyu_edu pi2018_nyu_edu          0 2023-09-14 15:42 dirToShareAccess
-rw-r--r--  1 pi2018_nyu_edu pi2018_nyu_edu 571 2023-09-29 01:11 input
-rw-r--r--  1 pi2018_nyu_edu pi2018_nyu_edu 279 2023-09-21 16:40 input.txt
-rw-r--r--  1 pi2018_nyu_edu pi2018_nyu_edu 571 2023-09-29 01:38 input_hw3
-rw-r--r--  1 pi2018_nyu_edu pi2018_nyu_edu 7077973 2023-10-04 18:13 input_hw4
drwxr-xr-x  - pi2018_nyu_edu pi2018_nyu_edu          0 2023-09-22 16:58 new_output
drwxr-xr-x  - pi2018_nyu_edu pi2018_nyu_edu          0 2023-09-29 01:48 output_hw3
pi2018_nyu_edu@nyu-dataproc-m:~$
```

```
pi2018_nyu_edu@nyu-dataproc-m:~$ hadoop jar maxNeighborhoodCount.jar NeighborhoodCount input_hw4 output_hw4
2023-10-04 18:31:42.031 INFO client.RMProxy: Connecting to ResourceManager at nyu-dataproc-m/192.168.1.38:8032
2023-10-04 18:31:42.206 INFO client.AHSProxy: Connecting to Application History server at nyu-dataproc-m/192.168.1.38:10200
2023-10-04 18:31:42.368 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-04 18:31:42.382 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pi2018_nyu_edu/.staging/job_1691775874963_8797
2023-10-04 18:31:42.599 INFO input.FileInputFormat: Total input files to process : 1
2023-10-04 18:31:42.664 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-04 18:31:42.828 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1691775874963_8797
2023-10-04 18:31:42.830 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-04 18:31:42.994 INFO conf.Configuration: resource-types.xml not found
2023-10-04 18:31:42.994 INFO resource.ResourcesUtils: Unable to find 'resource-types.xml'.
2023-10-04 18:31:43.261 INFO impl.YarnClientImpl: Submitted application application_1691775874963_8797
2023-10-04 18:31:43.294 INFO mapreduce.Job: The url to track the job: http://nyu-dataproc-m:8088/proxy/application_1691775874963_8797/
2023-10-04 18:31:43.294 INFO mapreduce.Job: Running job: job_1691775874963_8797
2023-10-04 18:31:50.387 INFO mapreduce.Job: Job job_1691775874963_8797 running in uber mode : false
2023-10-04 18:31:50.388 INFO mapreduce.Job: map 0% reduce 0%
2023-10-04 18:31:56.456 INFO mapreduce.Job: map 100% reduce 0%
2023-10-04 18:32:01.487 INFO mapreduce.Job: map 100% reduce 100%
2023-10-04 18:32:03.507 INFO mapreduce.Job: Job job_1691775874963_8797 completed successfully
2023-10-04 18:32:03.597 INFO mapreduce.Job: Counters:
  File System Counters
    FILE: Number of bytes read=1346423
    FILE: Number of bytes written=3184895
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of Bytes read=1078039
    HDFS: Number of bytes written=751186
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=3
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=1
    Launched reduce tasks=1
    Data-local map tasks=1
    Total time spent by all maps in occupied slots (ms)=13484
    Total time spent by all reduces in occupied slots (ms)=11552
    Total time spent by all map tasks (ms)=3371
    Total time spent by all reduce tasks (ms)=2888
    Total vcore-milliseconds taken by all map tasks=3371
    Total vcore-milliseconds taken by all reduce tasks=2888
    Total megabyte-milliseconds taken by all map tasks=13807616
    Total megabyte-milliseconds taken by all reduce tasks=11829248
Map-Reduce Framework
```

```
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -cat output_hw4/part-r-00000
Bronx Allerton 36
Bronx Baychester 7
Bronx Belmont 22
Bronx Bronxdale 18
Bronx Castle Hill 7
Bronx City Island 16
Bronx Claremont Village 25
Bronx Clason Point 21
Bronx Co-op City 2
Bronx Concourse 47
Bronx Concourse Village 31
Bronx East Morrisania 9
Bronx Eastchester 10
Bronx Edenwald 8
Bronx Fieldston 10
Bronx Fordham 59
Bronx Highbridge 26
Bronx Hunts Point 17
Bronx Kingsbridge 61
Bronx Longwood 57
Bronx Melrose 8
Bronx Morris Heights 16
Bronx Morris Park 14
Bronx Morrisania 18
Bronx Mott Haven 48
Bronx Mount Eden 6
Bronx Mount Hope 18
Bronx North Riverdale 10
Bronx Norwood 26
Bronx Olinville 4
Bronx Parkchester 37
Bronx Pelham Bay 15
Bronx Pelham Gardens 23
Bronx Port Morris 44
Bronx Riverdale 10
Bronx Schuylerville 13
Bronx Soundview 14
Bronx Spuyten Duyvil 2
Bronx Throgs Neck 19
Bronx Tremont 10
Bronx Unionport 5
Bronx University Heights 16
Bronx Van Nest 10
Bronx Wakefield 45
Bronx West Farms 1
Bronx Westchester Square 7
```

Brooklyn Bath Beach	13
Brooklyn Bay Ridge	122
Brooklyn Bedford-Stuyvesant	3200
Brooklyn Bensonhurst	63
Brooklyn Bergen Beach	9
Brooklyn Boerum Hill	142
Brooklyn Borough Park	120
Brooklyn Brighton Beach	58
Brooklyn Brooklyn Heights	133
Brooklyn Brownsville	51
Brooklyn Bushwick	2082
Brooklyn Canarsie	126
Brooklyn Carroll Gardens	188
Brooklyn Clinton Hill	489
Brooklyn Cobble Hill	80
Brooklyn Columbia St	35
Brooklyn Coney Island	15
Brooklyn Crown Heights	1306
Brooklyn Cypress Hills	107
Brooklyn DUMBO	32
Brooklyn Downtown Brooklyn	71
Brooklyn Dyker Heights	11
Brooklyn East Flatbush	437
Brooklyn East New York	183
Brooklyn Flatbush	512
Brooklyn Flatlands	68
Brooklyn Fort Greene	405
Brooklyn Fort Hamilton	43
Brooklyn Gowanus	200
Brooklyn Gravesend	61
Brooklyn Greenpoint	933
Brooklyn Kensington	139
Brooklyn Manhattan Beach	5
Brooklyn Midwood	91
Brooklyn Mill Basin	4
Brooklyn Navy Yard	11
Brooklyn Park Slope	431
Brooklyn Prospect Heights	283
Brooklyn Prospect-Lefferts Gardens	450
Brooklyn Red Hook	68
Brooklyn Sea Gate	6
Brooklyn Sheepshead Bay	144
Brooklyn South Slope	242
Brooklyn Sunset Park	317
Brooklyn Vinegar Hill	30
Brooklyn Williamsburg	3354
Brooklyn Windsor Terrace	142
Manhattan Battery Park City	56
Manhattan Chelsea	944

Brooklyn Sheepshead Bay	144
Brooklyn South Slope	242
Brooklyn Sunset Park	317
Brooklyn Vinegar Hill	30
Brooklyn Williamsburg	3354
Brooklyn Windsor Terrace	142
Manhattan Battery Park City	56
Manhattan Chelsea	944
Manhattan Chinatown	300
Manhattan Civic Center	45
Manhattan East Harlem	978
Manhattan East Village	1594
Manhattan Financial District	688
Manhattan Flatiron District	67
Manhattan Gramercy	291
Manhattan Greenwich Village	348
Manhattan Hell's Kitchen	1746
Manhattan Inwood	210
Manhattan Kips Bay	399
Manhattan Little Italy	106
Manhattan Lower East Side	786
Manhattan Marble Hill	11
Manhattan Midtown	1368
Manhattan Morningside Heights	290
Manhattan Murray Hill	443
Manhattan NoHo	69
Manhattan Nolita	227
Manhattan Roosevelt Island	67
Manhattan SoHo	312
Manhattan Stuyvesant Town	31
Manhattan Theater District	253
Manhattan Tribeca	140
Manhattan Two Bridges	56
Manhattan Upper East Side	1544
Manhattan Upper West Side	1700
Manhattan Washington Heights	757
Manhattan West Village	676
Queens Arverne	58
Queens Astoria	742
Queens Bay Terrace	6
Queens Bayside	36
Queens Bayswater	13
Queens Bellerose	12
Queens Breezy Point	3
Queens Briarwood	43
Queens Cambria Heights	22
Queens College Point	17
Queens Corona	48
Queens Ditmars Steinway	246
Queens Douglaston	5
Queens East Elmhurst	149
Queens Edgemere	10
Queens Elmhurst	187
Queens Far Rockaway	24
Queens Flushing	349
Queens Forest Hills	113
Queens Fresh Meadows	28
Queens Glendale	44
Queens Hollis	9
Queens Holliswood	3
Queens Howard Beach	16
Queens Jackson Heights	150
Queens Jamaica	202
Queens Jamaica Estates	17
Queens Jamaica Hills	6
Queens Kew Gardens	27
Queens Kew Gardens Hills	
Queens Laurelton	18
Queens Little Neck	4
Queens Long Island City	442
Queens Maspeth	96
Queens Middle Village	27
Queens Neponsit	3
Queens Ozone Park	54
Queens Queens Village	54
Queens Rego Park	85
Queens Richmond Hill	79
Queens Ridgewood	351
Queens Rockaway Beach	52
Queens Rosedale	53
Queens South Ozone Park	33

....

Task 2 Part A:

```
pi2018_nyu_edu@nyu-dataprof-m:~$ javac -classpath `yarn classpath`:. -d . CountLines.java
Note: CountLines.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
pi2018_nyu_edu@nyu-dataprof-m:~$ jar -cvf allLines.jar *.class
added manifest
adding: CountLines.class(in = 1375) (out= 790) (deflated 42%)
adding: CountLinesMapper.class(in = 1618) (out= 635) (deflated 60%)
adding: CountLinesReducer.class(in = 1604) (out= 669) (deflated 58%)
pi2018_nyu_edu@nyu-dataprof-m:~$ hadoop jar allLines.jar CountLines input_hw4 output_hw4A
2023-10-04 20:01:05,734 INFO client.RMProxy: Connecting to ResourceManager at nyu-dataprof-m/192.168.1.38:8032
2023-10-04 20:01:05,919 INFO client.AHSProxy: Connecting to Application History server at nyu-dataprof-m/192.168.1.38:10200
2023-10-04 20:01:06,340 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-04 20:01:06,354 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pi2018_nyu_edu/.staging/job_1691775874963_8880
2023-10-04 20:01:06,569 INFO input.FileInputFormat: Total input files to process : 1
2023-10-04 20:01:06,635 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-04 20:01:06,786 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1691775874963_8880
2023-10-04 20:01:06,788 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-04 20:01:06,956 INFO conf.Configuration: resource-types.xml not found
2023-10-04 20:01:06,957 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-04 20:01:07,097 INFO impl.YarnClientImpl: Submitted application application_1691775874963_8880
2023-10-04 20:01:07,131 INFO mapreduce.Job: The url to track the job: http://nyu-dataprof-m:8088/proxy/application_1691775874963_8880/
2023-10-04 20:01:07,132 INFO mapreduce.Job: Running job: job_1691775874963_8880
2023-10-04 20:01:14,220 INFO mapreduce.Job: Job job_1691775874963_8880 running in uber mode : false
2023-10-04 20:01:14,221 INFO mapreduce.Job: map 0% reduce 0%
2023-10-04 20:01:19,279 INFO mapreduce.Job: map 100% reduce 0%
2023-10-04 20:01:25,316 INFO mapreduce.Job: map 100% reduce 19%
2023-10-04 20:01:26,322 INFO mapreduce.Job: map 100% reduce 38%
2023-10-04 20:01:27,328 INFO mapreduce.Job: map 100% reduce 57%
2023-10-04 20:01:28,333 INFO mapreduce.Job: map 100% reduce 81%
2023-10-04 20:01:29,339 INFO mapreduce.Job: map 100% reduce 100%
2023-10-04 20:01:31,357 INFO mapreduce.Job: Job job_1691775874963_8880 completed successfully
2023-10-04 20:01:31,444 INFO mapreduce.Job: Counters: 55
    File System Counters
        FILE: Number of bytes read=2159846
        FILE: Number of bytes written=16129896
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=7078089
File Input Format Counters
    Bytes Read=7077973
File Output Format Counters
    Bytes Written=44
```

```
pi2018_nyu_edu@nyu-dataprof-m:~$ hdfs dfs -cat output_hw4A/part-r-00000
Total number of lines in AirBNB file: 49081
pi2018_nyu_edu@nyu-dataprof-m:~$
```

Task 2 Part B:

```
pi2018_nyu_edu@nyu-dataprof-m:~$ javac -classpath `yarn classpath`:. -d . CountBadShortRecords.java

pi2018_nyu_edu@nyu-dataprof-m:~$ jar -cvf allShortLines.jar *.class
added manifest
adding: CountBadShortRecords.class(in = 1608) (out= 871) (deflated 45%)
adding: CountBadShortRecordsMapper.class(in = 1875) (out= 769) (deflated 58%)
adding: CountBadShortRecordsReducer.class(in = 1624) (out= 679) (deflated 58%)
pi2018_nyu_edu@nyu-dataprof-m:~$ hadoop jar allShortLines.jar CountBadShortRecords input_hw4 output_hw4B
2023-10-04 20:22:28,692 INFO client.RMProxy: Connecting to ResourceManager at nyu-dataprof-m/192.168.1.38:8032
2023-10-04 20:22:28,866 INFO client.AHSProxy: Connecting to Application History server at nyu-dataprof-m/192.168.1.38:10200
2023-10-04 20:22:29,175 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-04 20:22:29,189 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pi2018_nyu_edu/.staging/job_1691775874963_8888
2023-10-04 20:22:29,394 INFO input.FileInputFormat: Total input files to process : 1
2023-10-04 20:22:29,473 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-04 20:22:29,623 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1691775874963_8888
2023-10-04 20:22:29,625 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-04 20:22:29,788 INFO conf.Configuration: resource-types.xml not found
2023-10-04 20:22:29,788 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-04 20:22:29,928 INFO impl.YarnClientImpl: Submitted application application_1691775874963_8888
2023-10-04 20:22:29,962 INFO mapreduce.Job: The url to track the job: http://nyu-dataprof-m:8088/proxy/application_1691775874963_8888
2023-10-04 20:22:29,962 INFO mapreduce.Job: Running job: job_1691775874963_8888
2023-10-04 20:22:37,052 INFO mapreduce.Job: Job job_1691775874963_8888 running in uber mode : false
2023-10-04 20:22:37,054 INFO mapreduce.Job: map 0% reduce 0%
2023-10-04 20:22:42,118 INFO mapreduce.Job: map 100% reduce 0%
2023-10-04 20:22:48,153 INFO mapreduce.Job: map 100% reduce 100%
2023-10-04 20:22:49,166 INFO mapreduce.Job: Job job_1691775874963_8888 completed successfully
2023-10-04 20:22:49,250 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=17056
        FILE: Number of bytes written=526181
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=7078089
        HDFS: Number of bytes written=48
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=3
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
```

```
Peak Reduce Virtual Memory (bytes): 1011500000
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=7077973
File Output Format Counters
    Bytes Written=48
pi2018_nyu_edu@nyu-dataprof-m:~$ hdfs dfs -cat output_hw4B/part-r-00000
Total number of short lines in AirBNB file:      341
pi2018_nyu_edu@nyu-dataprof-m:~$
```

Task 2 Part C:

```
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath`:. -d . CountBadLongRecords.java
pi2018_nyu_edu@nyu-dataproc-m:~$ jar -cvf allLongLines.jar *.class
added manifest
adding: CountBadLongRecords.class(in = 1603) (out= 870) (deflated 45%)
adding: CountBadLongRecordsMapper.class(in = 1872) (out= 766) (deflated 59%)
adding: CountBadLongRecordsReducer.class(in = 1622) (out= 678) (deflated 58%)
pi2018_nyu_edu@nyu-dataproc-m:~$ hadoop jar allLongLines.jar CountBadLongRecords input hw4 output hw4C
2023-10-04 20:30:10,395 INFO client.RMProxy: Connecting to ResourceManager at nyu-dataproc-m/192.168.1.38:8032
2023-10-04 20:30:10,568 INFO client.AHSProxy: Connecting to Application History server at nyu-dataproc-m/192.168.1.38:10200
2023-10-04 20:30:10,839 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-04 20:30:10,852 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pi2018_nyu_edu/.staging/job_1691775874963_8896
2023-10-04 20:30:11,058 INFO input.FileInputFormat: Total input files to process : 1
2023-10-04 20:30:11,120 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-04 20:30:11,274 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1691775874963_8896
2023-10-04 20:30:11,276 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-04 20:30:11,438 INFO conf.Configuration: resource-types.xml not found
2023-10-04 20:30:11,438 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-04 20:30:11,585 INFO impl.YarnClientImpl: Submitted application application_1691775874963_8896
2023-10-04 20:30:11,619 INFO mapreduce.Job: The url to track the job: http://nyu-dataproc-m:8088/proxy/application_1691775874963_8896
2023-10-04 20:30:11,619 INFO mapreduce.Job: Running job: job_1691775874963_8896
2023-10-04 20:30:18,797 INFO mapreduce.Job: Job job_1691775874963_8896 running in uber mode : false
2023-10-04 20:30:18,798 INFO mapreduce.Job: map 0% reduce 0%
2023-10-04 20:30:24,862 INFO mapreduce.Job: map 100% reduce 0%
2023-10-04 20:30:29,903 INFO mapreduce.Job: map 100% reduce 100%
2023-10-04 20:30:31,921 INFO mapreduce.Job: Job job_1691775874963_8896 completed successfully
2023-10-04 20:30:32,004 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=344574
        FILE: Number of bytes written=1181211
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=7078089
        HDFS: Number of bytes written=48
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=3
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=11952
```

```
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
    File Input Format Counters
        Bytes Read=7077973
    File Output Format Counters
        Bytes Written=48
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -cat output_hw4C/part-r-00000
Total number of long lines in AirBNB file:      7032
pi2018_nyu_edu@nyu-dataproc-m:~$
```

Task 2 Part D:

```
pi2018_nyu_edu@nyu-dataproc-m:~$ javac -classpath `yarn classpath`:. -d . CountGoodRecords.java
pi2018_nyu_edu@nyu-dataproc-m:~$ jar -cvf allGoodLines.jar *.class
added manifest
adding: CountGoodRecords.class(in = 1588) (out= 869) (deflated 45%)
adding: CountGoodRecordsMapper.class(in = 1866) (out= 764) (deflated 59%)
adding: CountGoodRecordsReducer.class(in = 1616) (out= 676) (deflated 58%)
pi2018_nyu_edu@nyu-dataproc-m:~$ hadoop jar allGoodLines.jar CountGoodRecords input_hw4 output_hw4D
2023-10-04 20:37:09,843 INFO client.RMProxy: Connecting to ResourceManager at nyu-dataproc-m/192.168.1.38:8032
2023-10-04 20:37:10,011 INFO client.AHSProxy: Connecting to Application History server at nyu-dataproc-m/192.168.1.38:10200
2023-10-04 20:37:10,306 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-04 20:37:10,319 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/pi2018_nyu_edu/.staging/job_1691775874963_8908
2023-10-04 20:37:10,526 INFO input.FileInputFormat: Total input files to process : 1
2023-10-04 20:37:10,588 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-04 20:37:10,732 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1691775874963_8908
2023-10-04 20:37:10,733 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-04 20:37:10,887 INFO conf.Configuration: resource-types.xml not found
2023-10-04 20:37:10,887 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-04 20:37:11,039 INFO impl.YarnClientImpl: Submitted application application_1691775874963_8908
2023-10-04 20:37:11,074 INFO mapreduce.Job: The url to track the job: http://nyu-dataproc-m:8088/proxy/application_1691775874963_8908/
2023-10-04 20:37:11,074 INFO mapreduce.Job: Running job: job_1691775874963_8908
2023-10-04 20:37:18,158 INFO mapreduce.Job: Job job_1691775874963_8908 running in uber mode : false
2023-10-04 20:37:18,159 INFO mapreduce.Job: map 0% reduce 0%
2023-10-04 20:37:23,216 INFO mapreduce.Job: map 100% reduce 0%
2023-10-04 20:37:29,250 INFO mapreduce.Job: map 100% reduce 100%
2023-10-04 20:37:30,262 INFO mapreduce.Job: Job job_1691775874963_8908 completed successfully
2023-10-04 20:37:30,350 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=2043698
        FILE: Number of bytes written=4579441
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=7078089
        HDFS: Number of bytes written=49
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=3
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Rack-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=13048
```

```
Bytes Read=7077973
File Output Format Counters
Bytes Written=49
pi2018_nyu_edu@nyu-dataproc-m:~$ hdfs dfs -cat output_hw4D/part-r-00000
Total number of good lines in AirBNB file: 41708
pi2018_nyu_edu@nyu-dataproc-m:~$
```