Created by Yash Patel

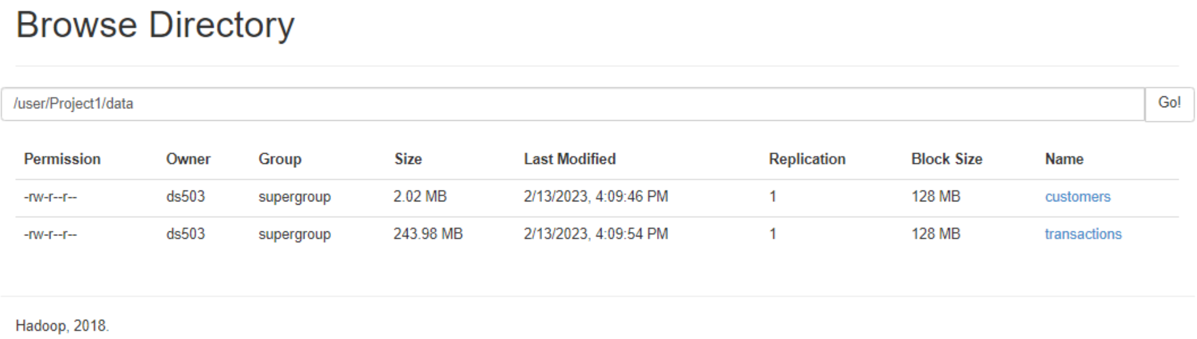
Question	Status	Comment
Q1	Fully Working	Running Datasets.java will generate the two files called customers and transactions
Q2	Fully Working	
Q3.1	Fully Working	Query1.java
Q3.2	Fully Working	Query2.java
Q3.3	Fully Working	Query3.java
Q4.1	Fully Working	query1.pig
Q4.2	Fully Working	query2.pig
Q4.3	Fully Working	query3.pig

Q1.

Run Datasets.java. Doesn't need any arguments and it will generate customers and transactions.

Q2.

Proof the datasets are in the file system



Browse Directory

/user/Project1/data

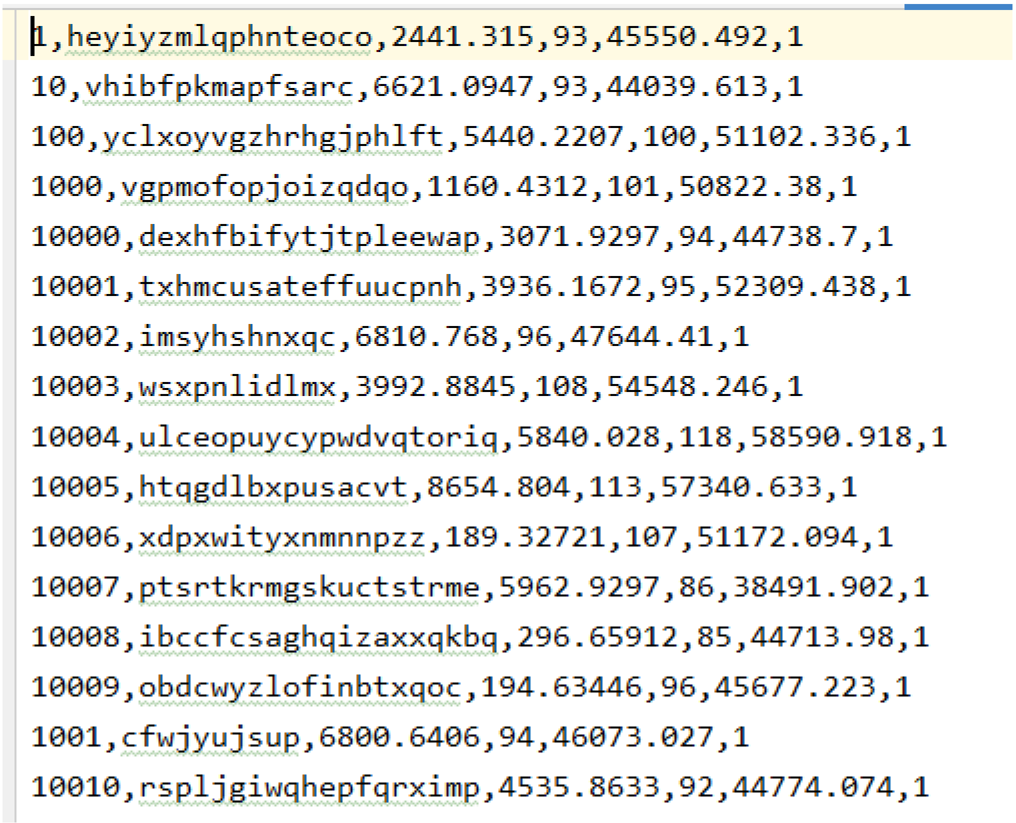
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	ds503	supergroup	2.02 MB	2/13/2023, 4:09:46 PM	1	128 MB	customers
-rw-r--r--	ds503	supergroup	243.98 MB	2/13/2023, 4:09:54 PM	1	128 MB	transactions

Hadoop, 2018.

Q3.1

Running Query1.java with arguments: <pathToCustomers> <pathToTransactions> <outputFileName>

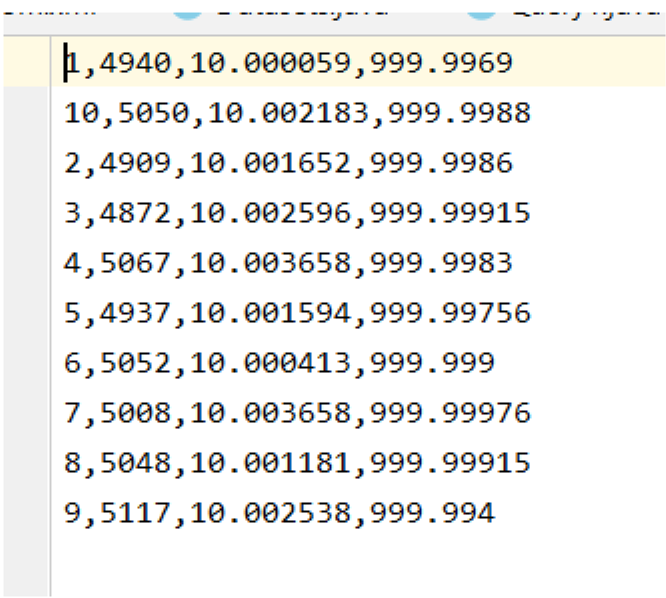
Doing so will generate:



```
1,heyiyzmlqphnteoco,2441.315,93,45550.492,1
10,vhibfpkmapfsarc,6621.0947,93,44039.613,1
100,yclxoyvgzhrhgjphlft,5440.2207,100,51102.336,1
1000,vgpmofopjoizqdqo,1160.4312,101,50822.38,1
10000,dexhfbifytjtpleewap,3071.9297,94,44738.7,1
10001,txhmcusateffuucpnh,3936.1672,95,52309.438,1
10002,imsyhshnxqc,6810.768,96,47644.41,1
10003,wsxpnlidlmx,3992.8845,108,54548.246,1
10004,ulceopuycypwdvqtoriq,5840.028,118,58590.918,1
10005,htqgdlbxpusacvt,8654.804,113,57340.633,1
10006,xdpxwityxnmnnpzz,189.32721,107,51172.094,1
10007,ptsrtkrmgskuctstrme,5962.9297,86,38491.902,1
10008,ibccfcsaghqizaxxqkbq,296.65912,85,44713.98,1
10009,obdcwyzlofinbtxqoc,194.63446,96,45677.223,1
1001,cfwjyujsup,6800.6406,94,46073.027,1
10010,rspljgiwqhepfqrximp,4535.8633,92,44774.074,1
```

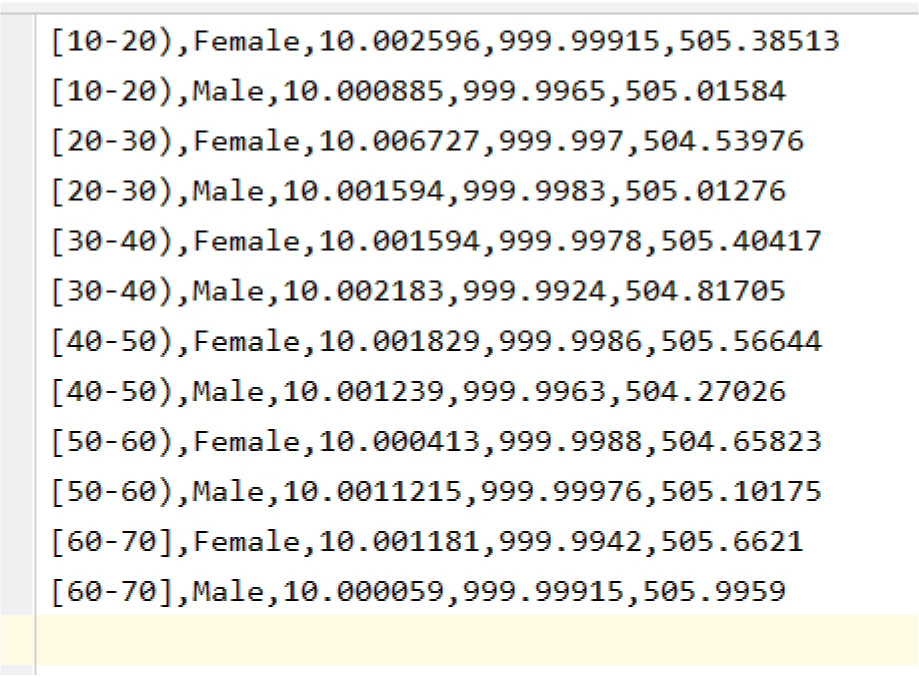
Q3.2

Run Query2 with arguments <pathToCustomers> <pathToTransactions> <outputFileName> This job is done with 2 map-reduce functions. There will be a file called query2FirstOutput that stores the intermediary data between the two map-reduce functions. The final results look like this:



```
1,4940,10.000059,999.9969
10,5050,10.002183,999.9988
2,4909,10.001652,999.9986
3,4872,10.002596,999.99915
4,5067,10.003658,999.9983
5,4937,10.001594,999.99756
6,5052,10.000413,999.999
7,5008,10.003658,999.99976
8,5048,10.001181,999.99915
9,5117,10.002538,999.994
```

Q3.3
Run Query3 with arguments <pathToCustomers> <pathToTransactions> <outputFileName>



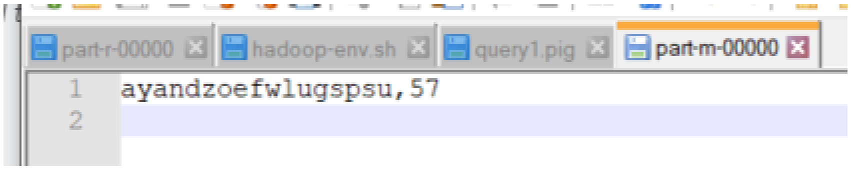
```
[10-20),Female,10.002596,999.99915,505.38513
[10-20),Male,10.000885,999.9965,505.01584
[20-30),Female,10.006727,999.997,504.53976
[20-30),Male,10.001594,999.9983,505.01276
[30-40),Female,10.001594,999.9978,505.40417
[30-40),Male,10.002183,999.9924,504.81705
[40-50),Female,10.001829,999.9986,505.56644
[40-50),Male,10.001239,999.9963,504.27026
[50-60),Female,10.000413,999.9988,504.65823
[50-60),Male,10.0011215,999.99976,505.10175
[60-70],Female,10.001181,999.9942,505.6621
[60-70],Male,10.000059,999.99915,505.9959
```

Q4.1

To run the pig scripts use:

```
pig -x mapreduce query1.pig
```

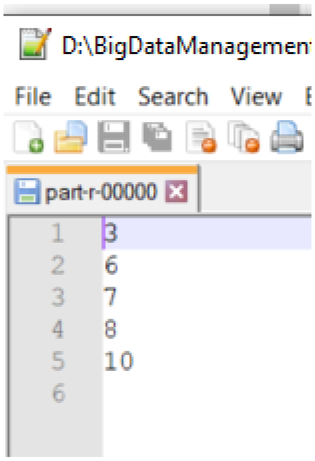
Able to get all of the minimum numbers of transcount. Here is an example output file



```
ayandzoefwlugspsu,57
```

Q4.2

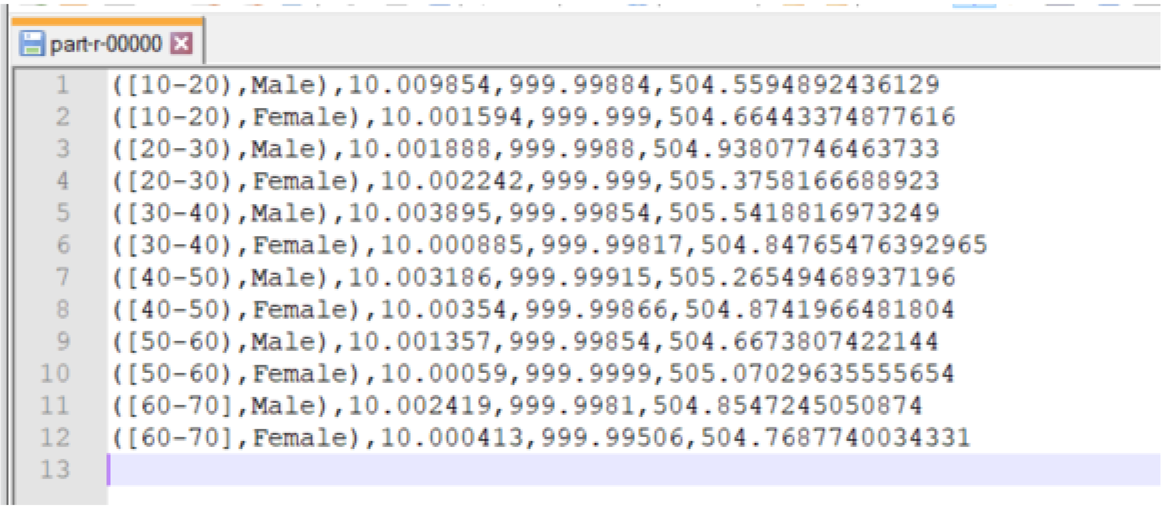
Able to generate just the country codes



```
3
6
7
8
10
```

Q4.3

You need to have getAge.jar to be in the same file as the query3.pig file in order to run. This jar file was made with the pig:0.15.0 jar included. Also included replicated join if that was needed for this part as well.



```
([10-20),Male),10.009854,999.99884,504.5594892436129
([10-20),Female),10.001594,999.999,504.66443374877616
([20-30),Male),10.001888,999.9988,504.93807746463733
([20-30),Female),10.002242,999.999,505.3758166688923
([30-40),Male),10.003895,999.99854,505.5418816973249
([30-40),Female),10.000885,999.99817,504.84765476392965
([40-50),Male),10.003186,999.99915,505.26549468937196
([40-50),Female),10.00354,999.99866,504.8741966481804
([50-60),Male),10.001357,999.99854,504.6673807422144
([50-60),Female),10.00059,999.9999,505.07029635555654
([60-70],Male),10.002419,999.9981,504.8547245050874
([60-70],Female),10.000413,999.99506,504.7687740034331
```