Garv Ghai
CSE B-32
180905232
Lab 5

q1)

mapper.py

```
import sys
import pandas as pd
df=pd.read_csv('heart_disease_data.csv')
age=df['age']
for word in age:
	print(word,'\t',1)
```

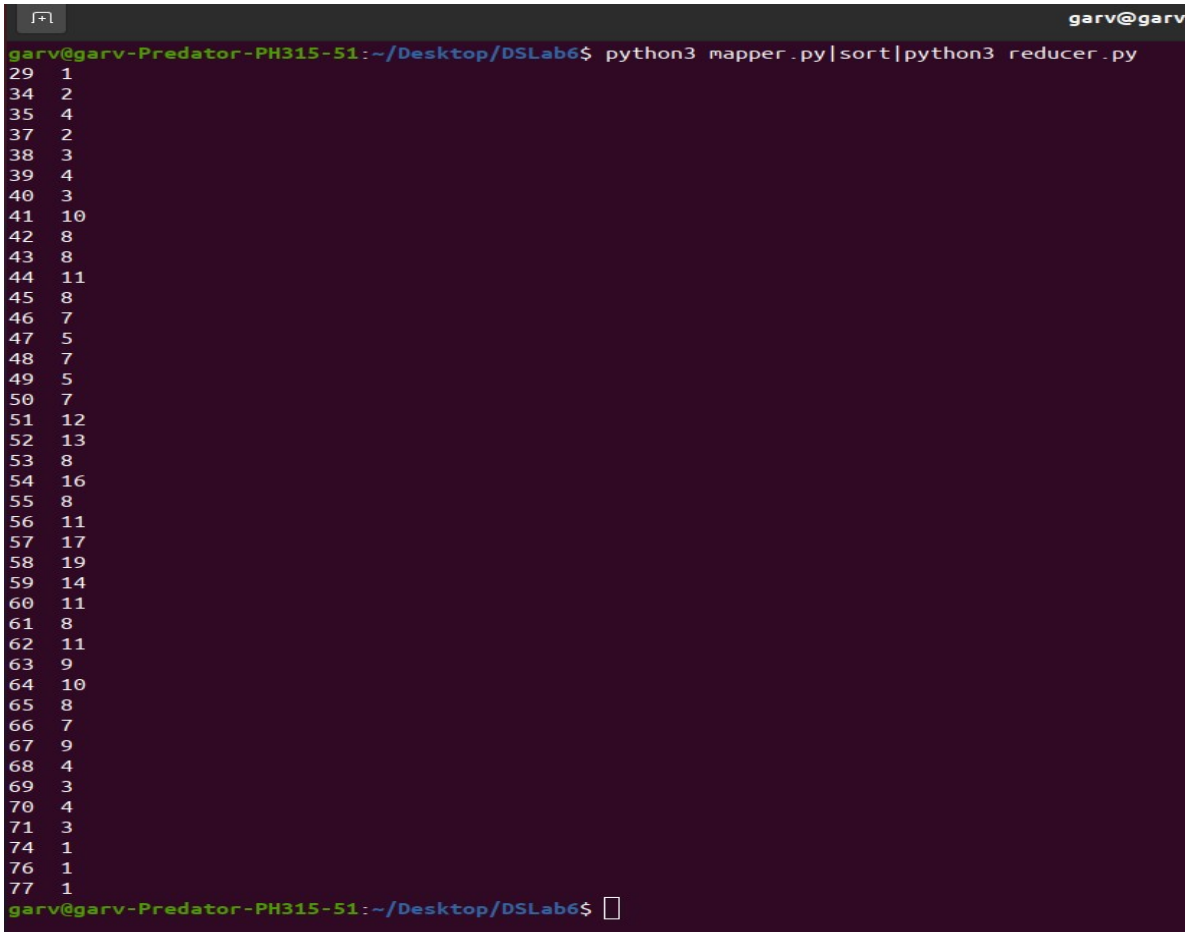
reducer.py

```
from operator import itemgetter
import sys
current_word = None
current_count = 0
word = None
# input comes from STDIN
for line in sys.stdin:
	line = line.strip()
	word,count = line.split('\t', 1)
	try:
		count = int(count)
	except ValueError:
		continue
	if current_word == word:
		current_count += count
	else:
		if current_word:
			print (current_word, current_count)
		current_count = count
		current_word = word
if current_word == word:
	print (current_word, current_count)
```

I/P:

Taking the age column in the heart disease file.

O/P:



q2)

freqmap1.py

```
from __future__ import print_function
import sys
import pandas as pd
df=pd.read_csv('covid_19_data.csv')
country=df['Deaths']
for word in country:
        print(word,'\t',1)
```

freqread1.py

```
from __future__ import print_function
import sys
lastWord = None
sum = 0
for line in sys.stdin:
        word, count = line.strip().split('\t', 1)
        count = int(count)
        if lastWord==None:
                lastWord = word
```

```
                sum = count
                continue
        if word==lastWord:
                sum += count
        else:
                print(lastWord,'\t',sum)
                sum = count
                lastWord = word
# output last word
if lastWord == word:
        print(lastWord,'\t',sum)
```

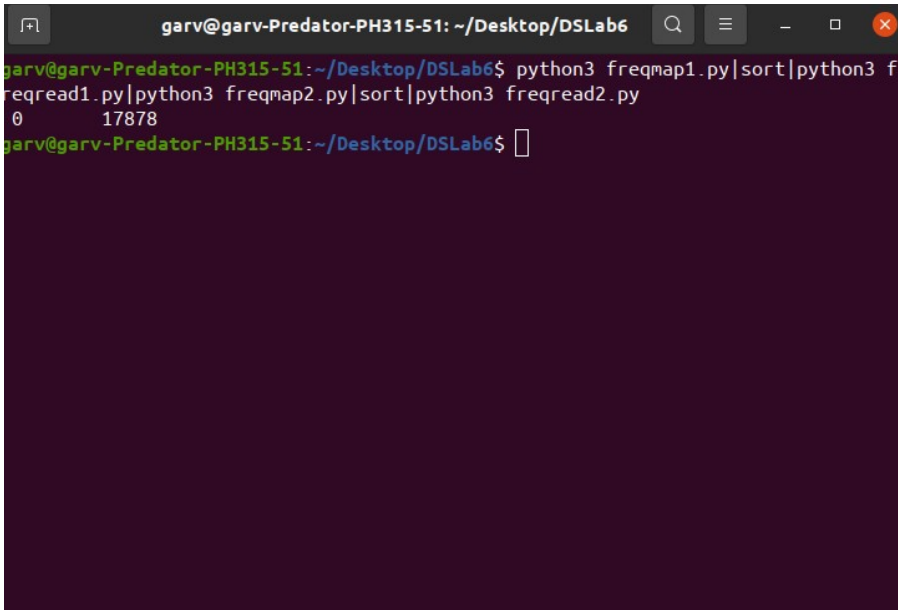
freqmap2.py

```
from __future__ import print_function
import sys
# input comes from STDIN (standard input)
for line in sys.stdin:
        word, count = line.strip().split('\t', 1)
        count = int(count)
        print(count,'\t',word);
```

freqread2.py

```
from __future__ import print_function
import sys
mostFreq = []
currentMax = -1
for line in sys.stdin:
        count, word = line.strip().split('\t', 1)
        count = int(count)
        if count > currentMax:
                currentMax = count
                mostFreq = [ word ]
        elif count == currentMax:
                mostFreq.append( word )
# output mostFreq word(s)
for word in mostFreq:
        print(word,'\t',currentMax)
```

O/P:



q3)

itemmap.py

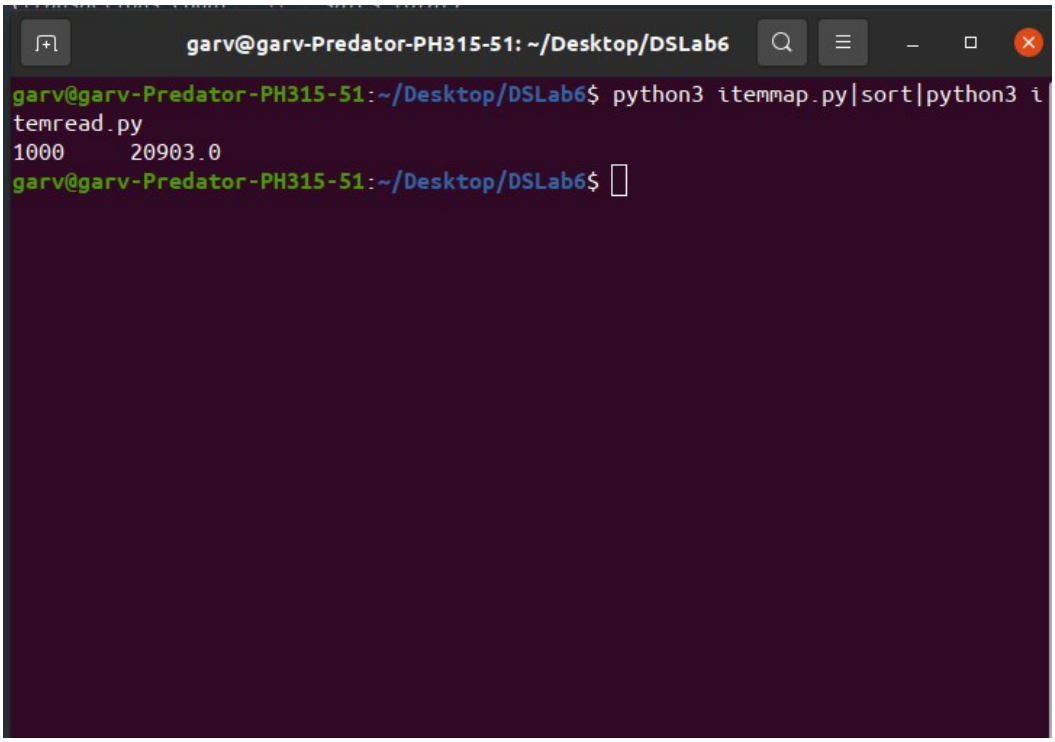
```
import pandas as pd
import numpy as np
dataframe = pd.read_excel("German_Credit.xlsx")
x = dataframe["CreditAmount"]
y = dataframe["DurationOfCreditInMonths"]
for i in range(len(x)):
        print("{0}\t{1}".format(x[i],y[i]))
```

itemread.py

```
import fileinput
transactions_count = 0
sales_total = 0
for line in fileinput.input():
        data = line.strip().split("\t")
        if len(data) != 2:
# Something has gone wrong. Skip this line.
                continue
        current_key, current_value = data
        transactions_count += 1
        sales_total += float(current_value)
print (transactions_count, "\t", sales_total)
```

O/P:

```
garv@garv-Predator-PH315-51:~/Desktop/DSLab6$ python3 itemmap.py
1049    18
2799    9
841     12
2122    12
2171    12
2241    10
3398    8
1361    6
1098    18
3758    24
3905    11
6187    30
1957    6
7582    48
1936    18
2647    6
3939    11
3213    18
2337    36
7228    11
3676    6
3124    12
2384    36
1424    12
4716    6
4771    11
652     12
1154    9
3556    15
4796    42
3017    30
3535    36
6614    36
1376    24
1721    15
860     6
1495    12
```



q4)

sepmap.py

```
import pandas as pd
import numpy as np
def read_input(dataframe):
	for x in dataframe["age"]:
		yield x
def main(separator = "\t"):
	filename = input("Enter a filename: ")
	dataframe = pd.read_csv(filename)
	data = read_input(dataframe)
	for x in data:
		print("%s%s%d" %(x,separator,1))
if __name__ == "__main__":
	main()
```

sepread.py

```
from itertools import groupby
from operator import itemgetter
import sys
def read_mapper_output(file, separator='\t'):
	for line in file:
		yield line.rstrip().split(separator, 1)
def main(separator='\t'):
	data = read_mapper_output(sys.stdin, separator=separator)
	for current_word, group in groupby(data, itemgetter(0)):
		try:
			total_count = sum(int(count) for current_word, count in group)
			print ("%s%s%d" % (current_word, separator, total_count))
		except ValueError:
			pass
if __name__ == "__main__":
	main()
```

O/P:

```
arv@garv-Predator-PH315-51:~/Desktop/DSLab6$ python3 sepmap.py|sort|python3 sepread.py
eart_disease_data.csv
9       1
4       2
5       4
7       2
8       3
9       4
0       3
1       10
2       8
3       8
4       11
5       8
6       7
7       5
8       7
9       5
0       7
1       12
2       13
3       8
4       16
5       8
6       11
7       17
8       19
9       14
0       11
1       8
2       11
3       8
4       10
5       8
6       7
7       9
8       4
9       3
0       4
1       3
4       1
6       1
7       1
nter a filename: 63    1
arv@garv-Predator-PH315-51:~/Desktop/DSLab6$
```

q5)

map.py

```
import numpy as np
import pandas as pd
dataframe = pd.read_csv("covid_19_data.csv")
country_name = dataframe["Country/Region"]
peak_cases = dataframe["Confirmed"]
for x in range(len(country_name)):
        print("%s\t%d" %(country_name[x],peak_cases[x]))
```

read.py

```
import fileinput
max_value = 0
old_key = None
for line in fileinput.input():
        data = line.strip().split("\t")
        if len(data) != 2:
# Something has gone wrong. Skip this line.
                continue
        current_key, current_value = data
```

```
# Refresh for new keys (i.e. locations in the example context)
        if old_key and old_key != current_key:
                print (old_key, "\t", max_value)
                old_key = current_key
                max_value = 0
                old_key = current_key
        if float(current_value) > float(max_value):
                max_value = float(current_value)
if old_key != None:
        print (old_key, "\t", max_value)
```

O/P:

q6)

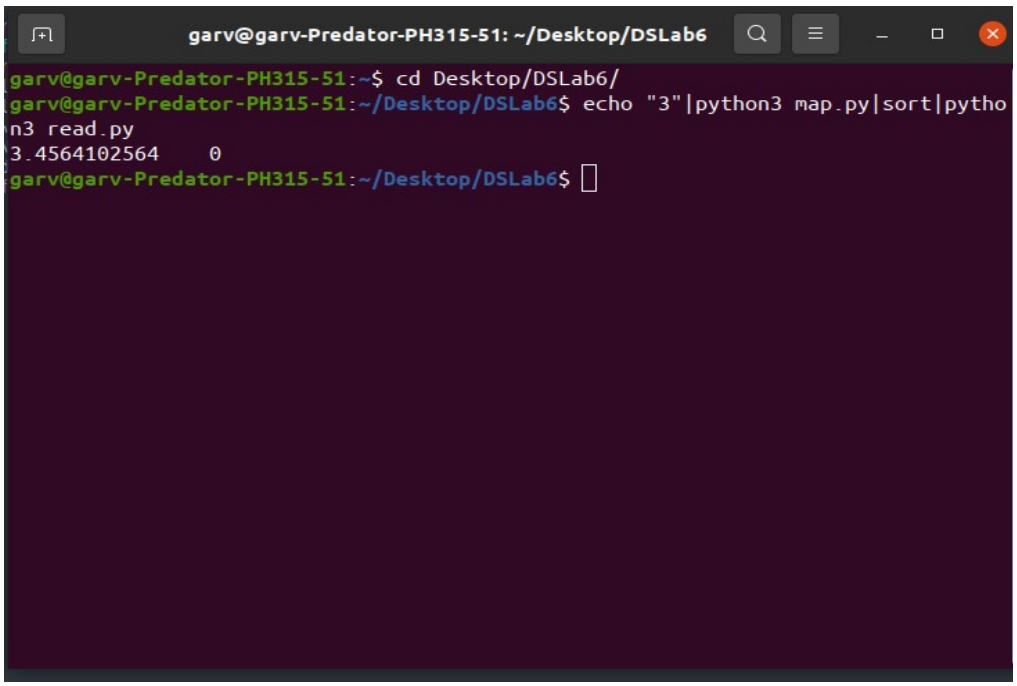
Mapper :

```
import sys
def f( x ):
   return 4.0 / ( 1.0 + x*x )
for line in sys.stdin:
   line = line.strip()
   words = line.split()
   N = int( words[0] )
   deltaX = 1.0 / N
   for i in range( 0, N ):
      print( "1\t%1.10f" % ( f( i * deltaX )*deltaX ) )
```

Reducer:

```
from __future__ import print_function
from operator import itemgetter
import sys
sum = 0
for line in sys.stdin:
     line = line.strip()
     word, count = line.split('\t', 1)
     try:
        count = float(count)
     except ValueError:
        # count was not a number, so silently
        # ignore/discard this line
        #print( "--skipping (%s, %s)" % ( str(word), str(count) ) )
        continue
     sum += count
print( '%1.10f\t0' % sum )
```

Output:



q7)

```
from operator import itemgetter
import sys
current_year = None
year_tot = [0]
year_male = [0]
year_fem = [0]
month_tot = [0]
month_male = [0]
month_fem = [0]
specyear = 2001
i = 1
while i<40:
        year_tot.append(0)
        year_male.append(0)
        year_fem.append(0)
        i += 1
i = 0
while i<12:
        month_tot.append(0)
        month_male.append(0)
        month_fem.append(0)
        i += 1
year = None
for line in sys.stdin:
   line = line.strip()
   sex, month, year = line.split(' ')
   sex = int(sex)
   month = int(month)
   year = int(year)
   year_tot[year-1980] += 1
   if sex == 0:
        year_male[year-1980] += 1
   else:
        year_fem[year-1980] += 1
   if year == specyear:
        month_tot[month-1] += 1
        if sex == 0:
                month_male[month-1] += 1
        else:
                month_fem[month-1] += 1
i = 0
while i<40:
        if year_tot[i] == 0:
                i += 1
                continue
        print('Year %d Total: %d' %(i+1980, year_tot[i]))
        print('Males: %d' %(year_male[i]))
        print('Females: %d' %(year_fem[i]))
```

```
        print('\n')
        i += 1
print('Year %d'% (specyear))
i = 0
while i<12:
        if month_tot[i] == 0:
                i += 1
                continue
        print('Month %d Total: %d' %(i+1, month_tot[i]))
        print('Males: %d' %(month_male[i]))
        print('Females: %d' %(month_fem[i]))
        print('\n')
        i += 1
```

O/P:

```
Year 2007 Total: 5
Males: 1
Females: 4


Year 2010 Total: 10
Males: 5
Females: 5


Year 2013 Total: 6
Males: 2
Females: 4


Year 2016 Total: 8
Males: 5
Females: 3


Year 2019 Total: 6
Males: 2
Females: 4


Year 2001
Month 2 Total: 1
Males: 0
Females: 1


Month 5 Total: 1
Males: 0
Females: 1


Month 8 Total: 1
Males: 0
Females: 1


Month 10 Total: 1
Males: 0
Females: 1


Month 12 Total: 1
Males: 1
Females: 0
```

q8)

```
from operator import itemgetter
import sys
odd_count = 0
even_count = 0
for line in sys.stdin:
    line = line.strip()
    num = line.split()
    for currnum in num:
        try:
            odd_count = int(odd_count)
            even_count = int(even_count)
            currnum = int(currnum)
        except ValueError:
            continue
        if currnum%2 == 0:
            even_count += 1
        else:
            odd_count += 1
print ('%s odd and %s even' % (odd_count, even_count))
```

O/P: