Name: Ankita Ghosh

Reg no: 180905354

Sec: A Roll no: 41

DISTRIBUTED SYSTEMS LAB 5

**Question 1.**

Program:

**Mapper:**

import sys

import pandas as pd

df=pd.read\_csv('heart\_disease\_data.csv')

#print("xyz")

age=df['oldpeak']

for word in age:

print(word,'\t',1)

**Reducer:**

from operator import itemgetter

import sys

#print("abc")

current\_word = None

current\_count = 0

word = None

# input comes from STDIN for line in sys.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)

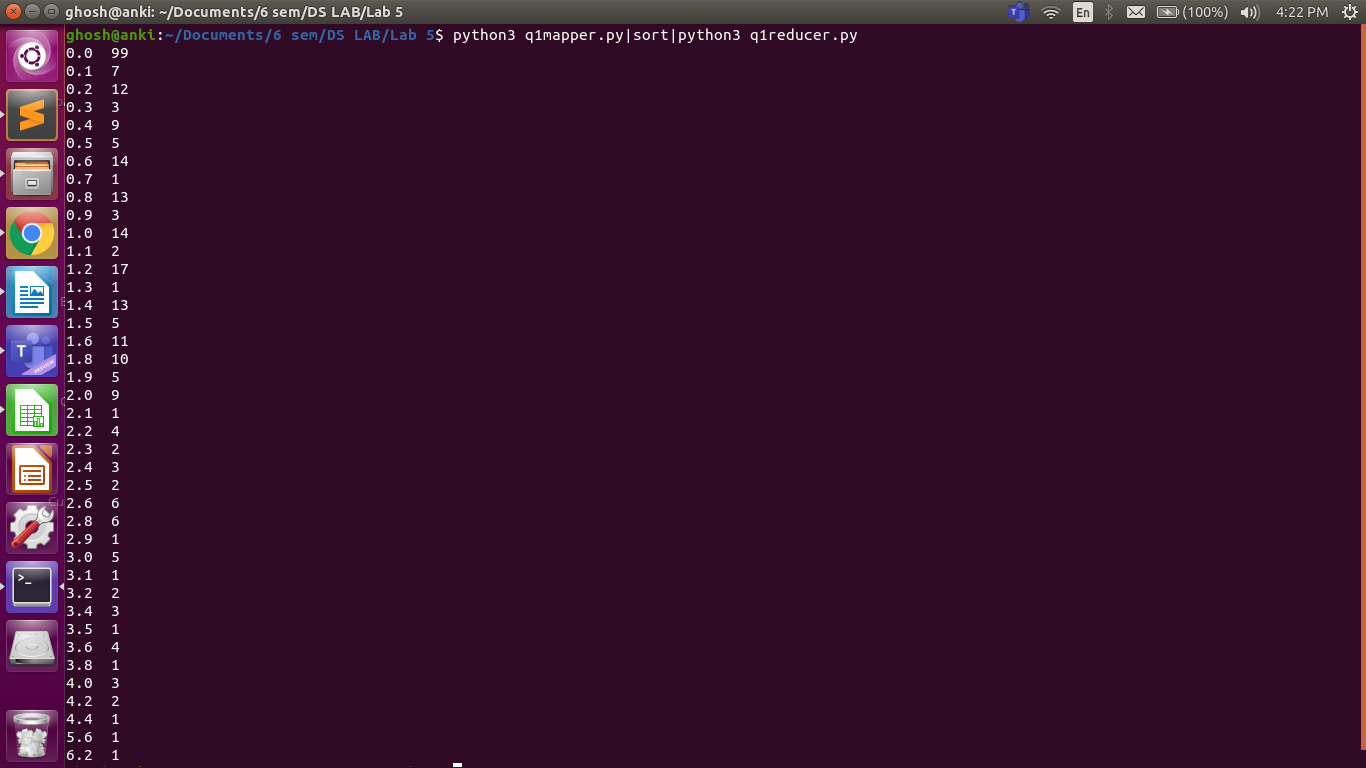
current\_count = count

current\_word = word

if current\_word == word:

print (current\_word, current\_count)

Output:



**Question 2.**

Program:

**Mapper 1:**

import fileinput

import pandas as pd

from pandas.core import frame

Cov = pd.read\_csv("heart\_disease\_data.csv")

Frame=pd.DataFrame(Cov.values, columns = ["age", "sex", "cp", "trestbps", "chol", "fbs","restecg", "thalach", "exang", "oldpeak", "slope", "ca", "thal", "target"])

Frame.style.hide\_index()

arr=Frame['trestbps'].to\_string(index=False).split('\n')

print(arr)

arr1=[]

for ele in arr:

arr1.append(ele.strip())

arr1.sort()

for ele in arr1:

print(ele,1)

**Reducer 1:**

from operator import itemgetter

import sys

curr\_word=None

curr\_count=0

word=None

for line in sys.stdin:

word,count=line.split(' ',1)

count=1

try:

count = int(count)

except ValueError:

continue

if curr\_word == word:

curr\_count += count

else:

if curr\_word:

print (curr\_word, curr\_count)

curr\_count = count

curr\_word = word

if curr\_word == word:

print(curr\_word,curr\_count)

**Mapper 2:**

from \_\_future\_\_ import print\_function

import sys

for line in sys.stdin:

word,count = line.strip().split(' ', 1)

print(count,word)

count = int(count)

print(count,word)

**Reducer 2:**

from \_\_future\_\_ import print\_function

from operator import itemgetter

import sys

mostFreq = []

currentMax = -1

for line in sys.stdin:

count, word = line.strip().split(' ', 1)

count = int(count)

if count >= currentMax:

currentMax = count

mostFreq = [ word ]

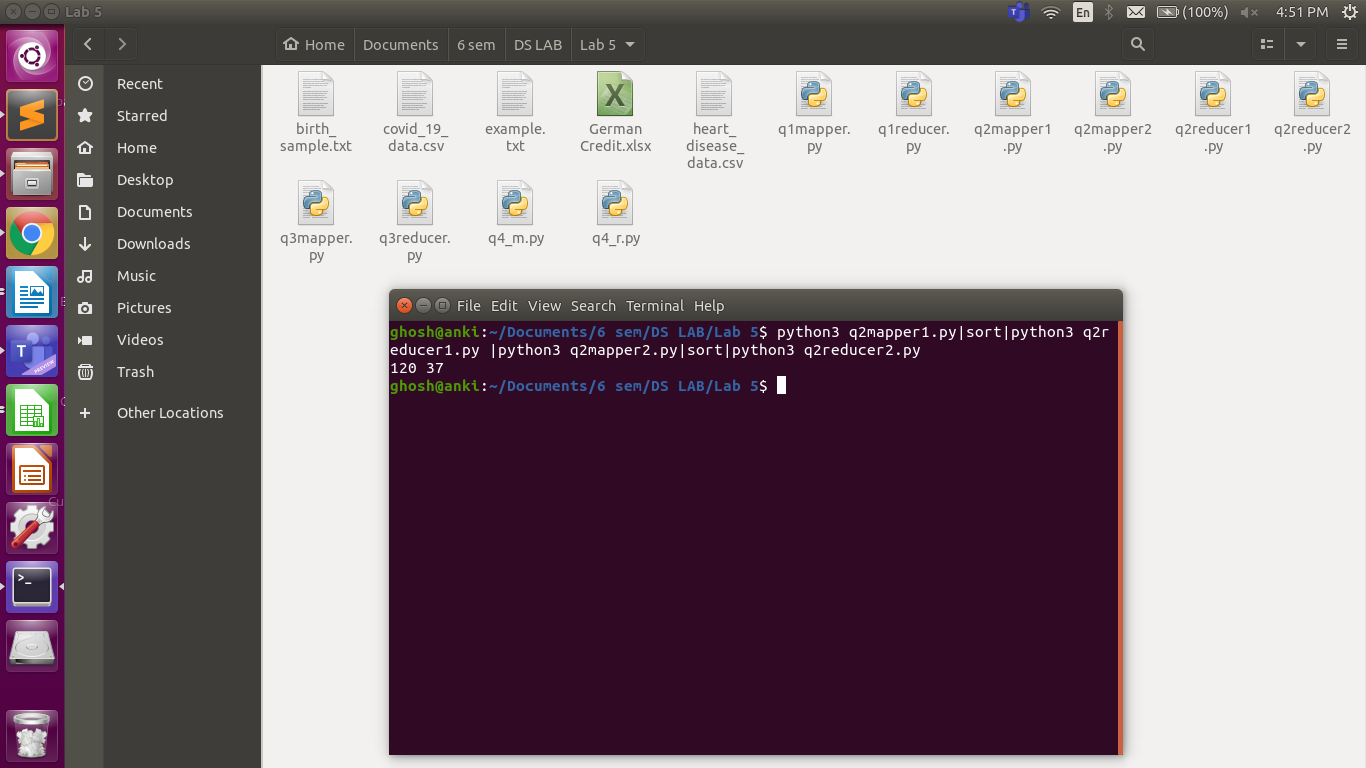
elif count == currentMax:

mostFreq.append( word )

for word in mostFreq:

print(word, currentMax )

Output:



**Question 3.**

Program:

**Mapper:**

import fileinput

for line in fileinput.input():

data = line.strip().split("\t")

if len(data) == 6:

date, time, location, item, cost, payment = data

print ("{0}\t{1}".format(location,cost))

**Reducer:**

import fileinput

transactions\_count = 0

sales\_total = 0

for line in fileinput.input():

data = line.strip().split("\t")

if len(data) != 2:

continue

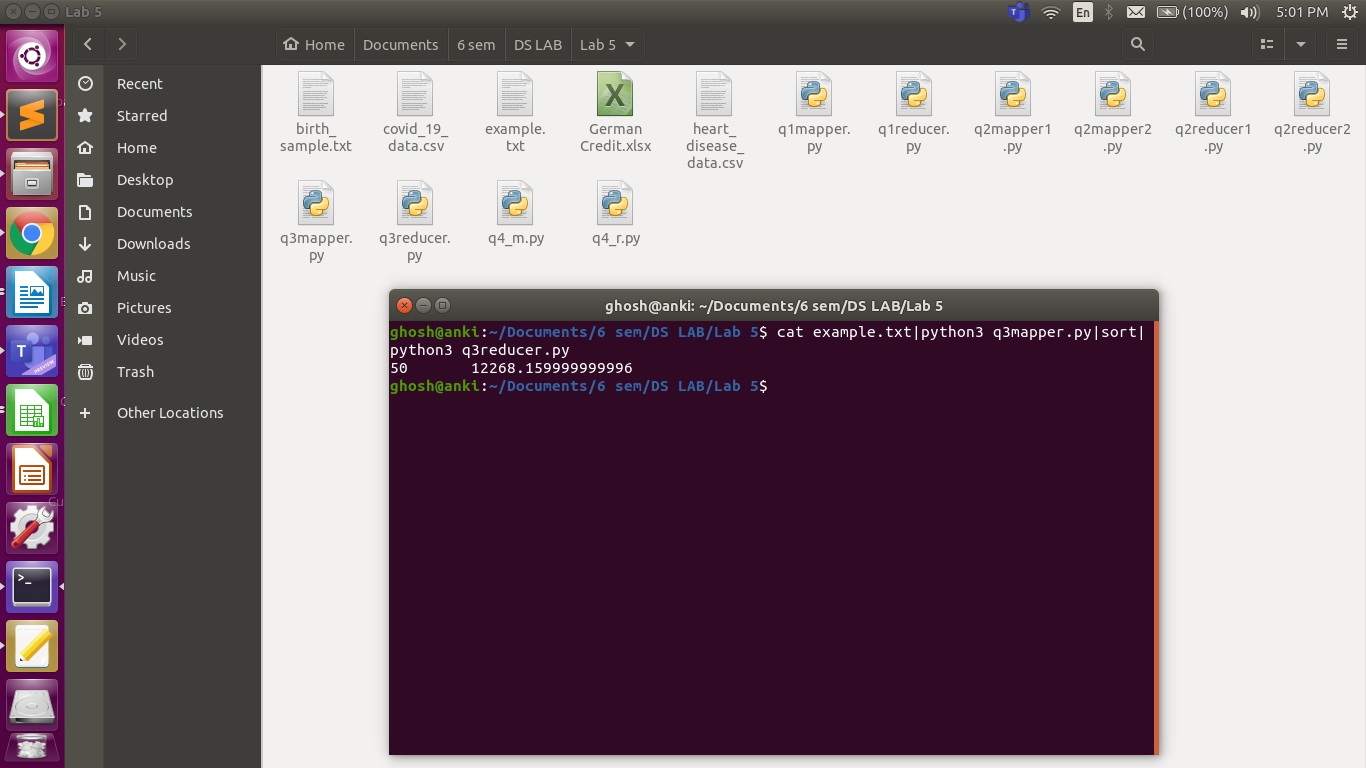
current\_key, current\_value = data

transactions\_count += 1

sales\_total += float(current\_value)

print (transactions\_count, "\t", sales\_total)

Output:



**Question 4.**

Program:

**Mapper:**

from os import sep

import pandas as pd

import sys

df = pd.read\_excel('German Credit.xlsx',engine='openpyxl')

x=df['CreditAmount'][:100]

print(x)

def read\_input(x):

for line in x:

line=str(line)

yield line.split()

def main(separator='-->'):

data = read\_input(x)

for words in data:

for word in words:

print(word, separator, 1)

if \_\_name\_\_=='\_\_main\_\_':

main()

**Reducer:**

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='-->'):

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))

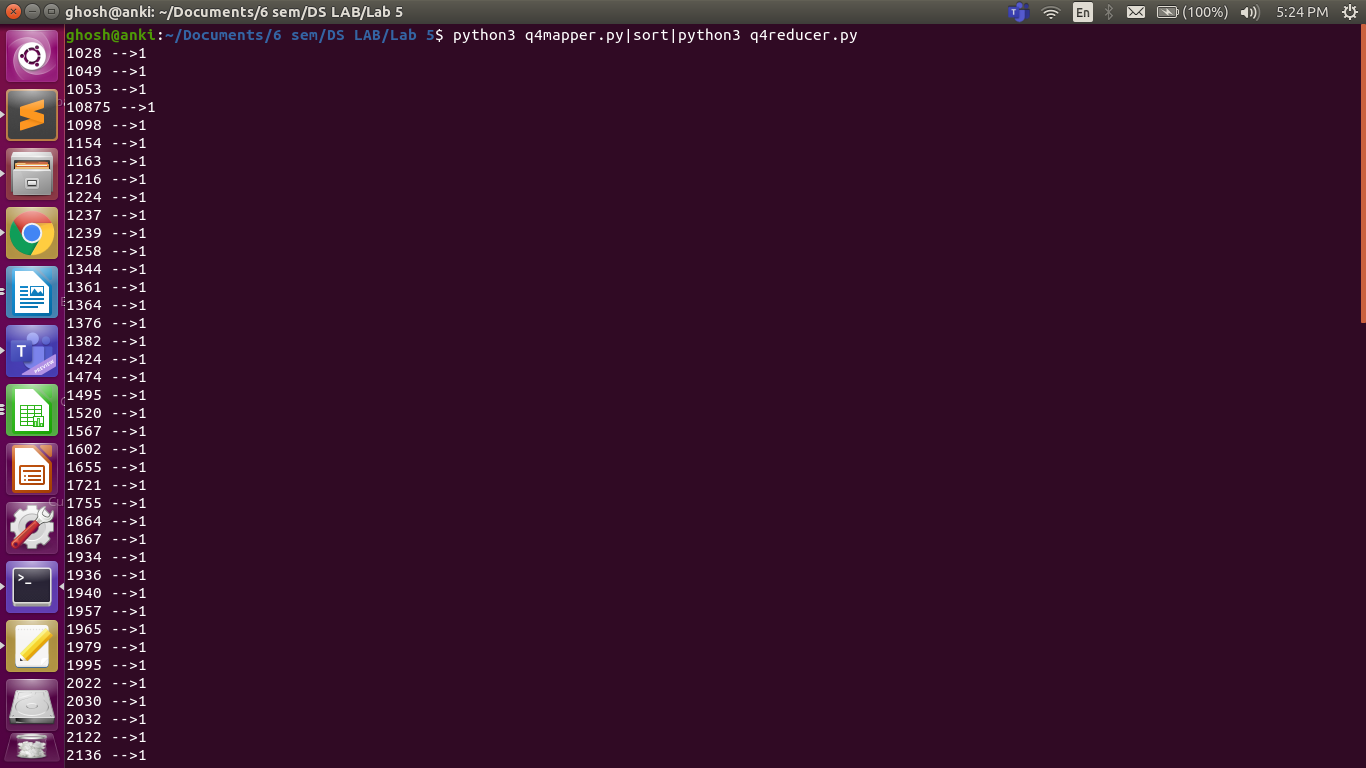
except ValueError:

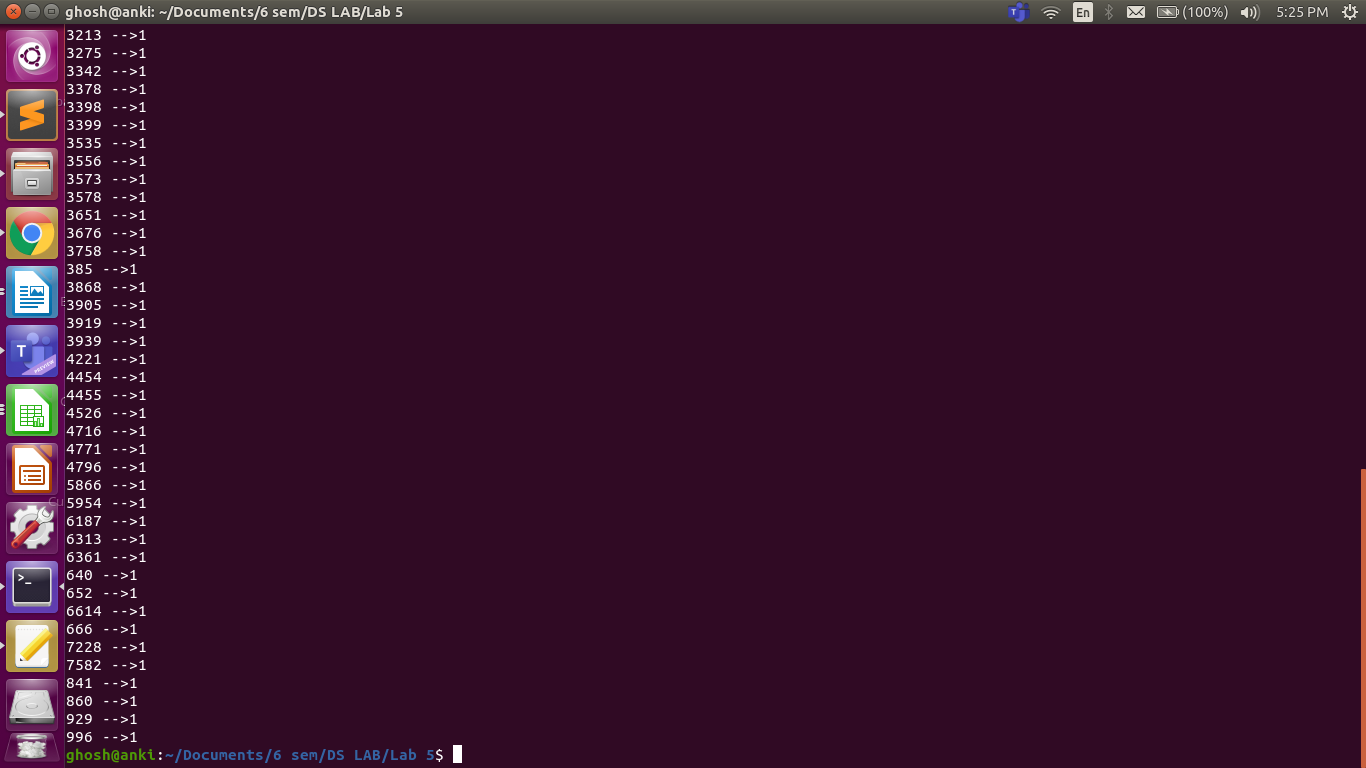
pass

if \_\_name\_\_ == "\_\_main\_\_":

main()

Output:





**Question 5.**

Program:

**Mapper:**

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]))  
  
**Reducer:**

import fileinput

max\_value = 0

old\_key = None

for line in fileinput.input():

data = line.strip().split("\t")

if len(data) != 2:

continue

current\_key, current\_value = data

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

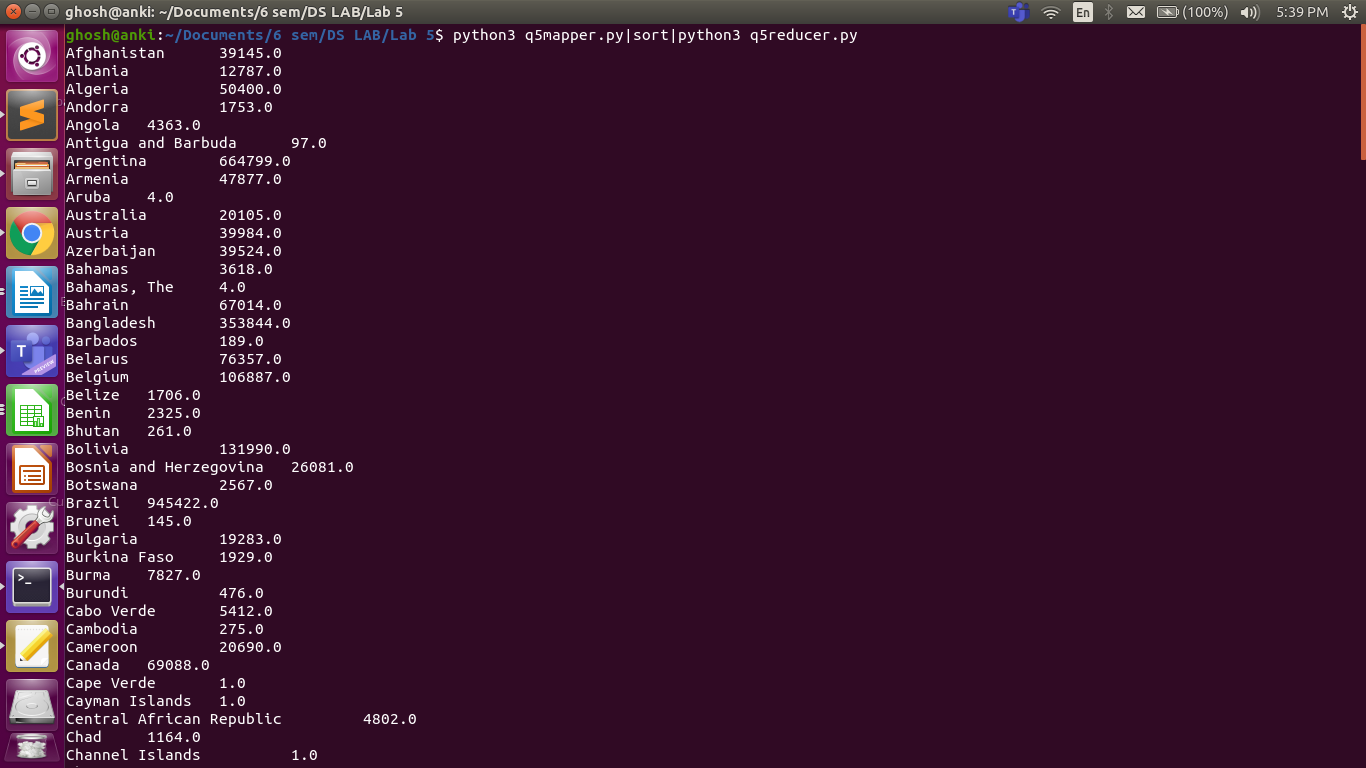
if float(current\_value) > float(max\_value):

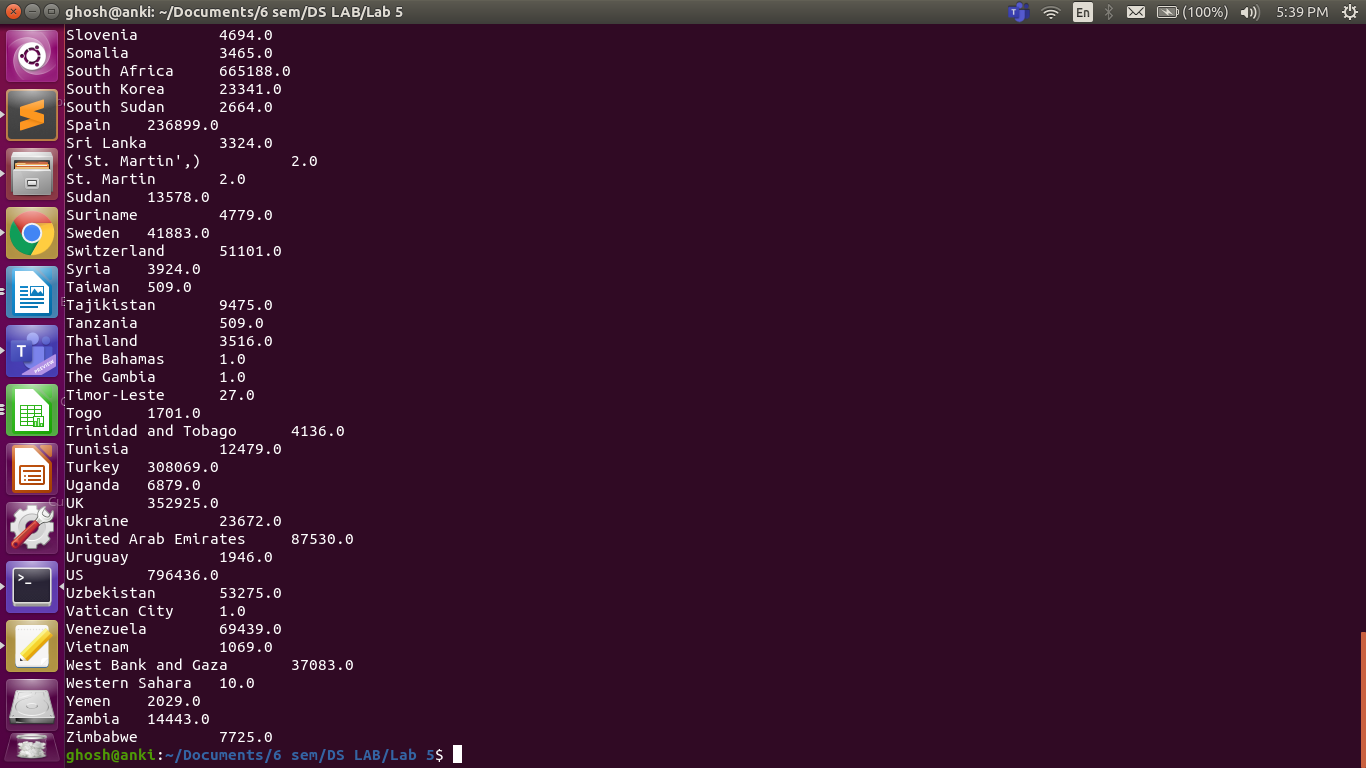
max\_value = float(current\_value)

if old\_key != None:

print (old\_key, "\t", max\_value)

Output:





**Question 6.**

Program:

**Mapper:**

import fileinput

import sys

def main():

for line in sys.stdin:

line = line.strip()

data = line.split()

if len(data) >= 64:

print(f'{data[1]}\t{data[31]}')

if \_\_name\_\_ == '\_\_main\_\_':

main()

**Reducer:**

import fileinput

yearMcount = 0

yearFcount = 0

monMcount = 0

monFcount = 0

prevyear = '201001'

for line in fileinput.input():

data = line.strip().split()

if len(data) != 2:

continue

curryear, gender = data

if prevyear[:4] == curryear[:4]:

if prevyear[4:] == curryear[4:]:

if gender == 'M':

yearMcount += 1

monMcount += 1

else:

yearFcount += 1

monFcount += 1

else:

print(

f"Month {prevyear[4:]}\t Male Births: {monMcount}, Female Births: {monFcount}, Total Births: {monFcount + monMcount}"

)

prevyear = curryear

monMcount = monFcount = 0

if gender == 'M':

yearMcount += 1

monMcount += 1

else:

yearFcount += 1

monFcount += 1

else:

print(

f"Year: {prevyear[:4]}\t Male Births: {yearMcount}\t Female Births:{yearFcount}\t Total Births: {yearFcount+yearMcount}\n\n"

)

yearMcount = 0

yearFcount = 0

prevyear = curryear

if gender == 'M':

yearMcount += 1

monMcount += 1

else:

yearFcount += 1

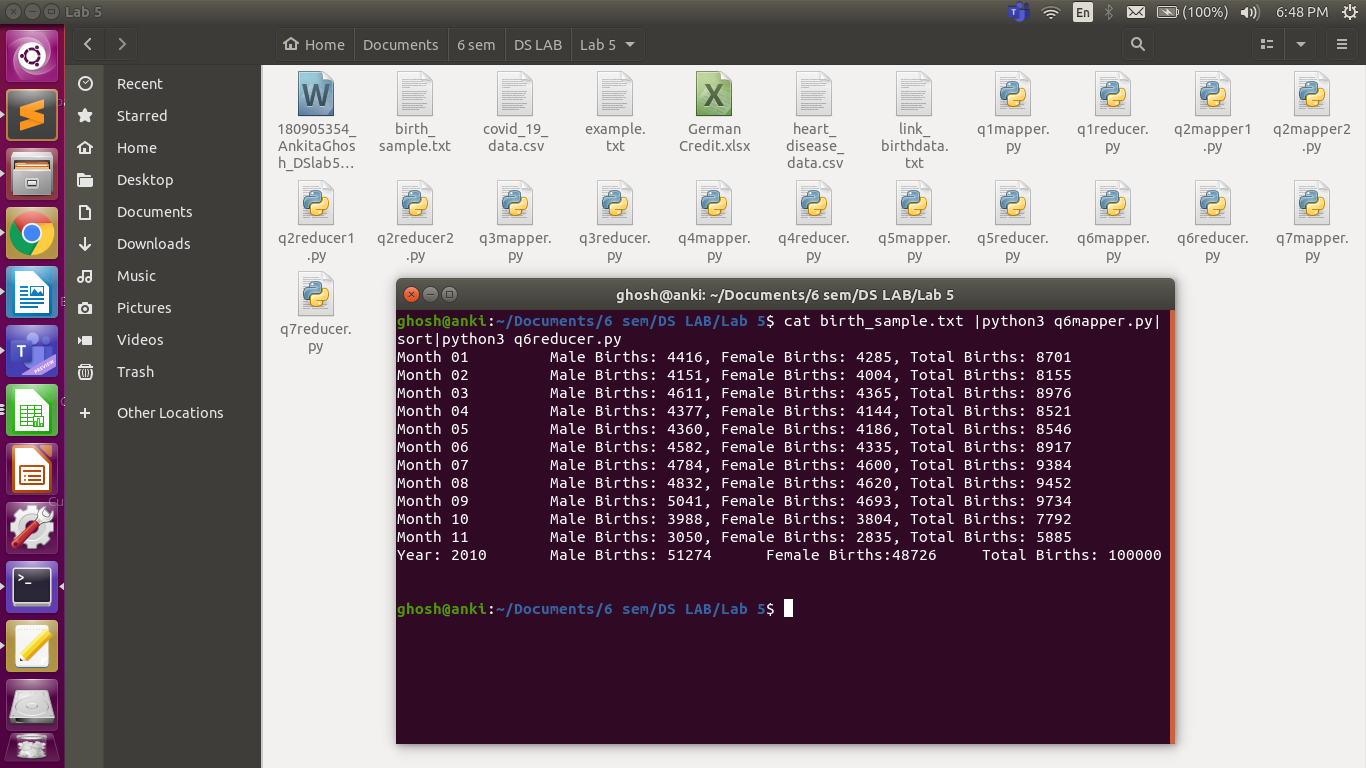
monFcount += 1

print(

f"Year: {prevyear[:4]}\t Male Births: {yearMcount}\t Female Births:{yearFcount}\t Total Births: {yearFcount+yearMcount}\n\n"

)

Output:



**Question 7.**

Program:

**Mapper:**

#!/usr/bin/env python

import sys

for line in sys.stdin:

line = line.strip()

words = line.split()

for word in words:

print ('%s\t%s\t' % (word, 1))

**Reducer:**

#!/usr/bin/env python

from operator import itemgetter

import sys

current\_word = None

current\_count = 0

word = None

even=0

odd=0

for line in sys.stdin:

line = line.strip()

word, count = line.split('\t', 1)

count = int(count)

if current\_word == word:

current\_count += count

else:

if current\_word:

#print ('%s\t%s' % (current\_word, current\_count))

if (int(current\_word) %2==0):

even +=current\_count

else:

odd+=current\_count

current\_count = count

current\_word = word

if current\_word == word:

#print ('%s\t%s\t' % (current\_word, current\_count))

if (int(current\_word) %2==0):

even +=current\_count

else:

odd+=current\_count

print("EVEN COUNT = ",even)

print("ODD COUNT = ",odd)

Output:

