Programming Assignment 2

```
from pyspark import SparkConf
from pyspark import SparkContext
import pandas as pd
import xlrd
sc = SparkContext.getOrCreate()

months=['ANNUAL','JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV','DEC']
df=pd.read_excel("data.xlsx",header=None,skiprows=1,
names=['CANADA','ANNUAL','JAN','FEB','MAR','APR','MAY','JUN','JUL','AUG','SEP','OCT','NOV'
,'DEC','YEARS','# CITIES'])

for i in months:
  demo= df[[i , '# CITIES']]
  demo = demo.apply (pd.to_numeric, errors='coerce')
  final_df=demo.dropna()

df_par = sc.parallelize(final_df[i])
df_par2 = sc.parallelize(final_df['# CITIES']) columnData=df_par.collect()
citiesData=df_par2.collect()
wgt=[]
for j in range(0, len(columnData)):
   wgt.append(columnData[j] * citiesData[j])

def returnNth(lst, n1, n):
  return lst[n1::n]

weightOfAverageTemp = []
at_cities = []
totalWeightSum = 0
totalCities=0

for j in range(0, len(wgt)):
    weightOfAverageTemp=returnNth(wgt, 0, 4)
    for k in range(0, len(weightOfAverageTemp)):
        totalWeightSum = totalWeightSum + weightOfAverageTemp[k]

for k in range(0, len(citiesData)):
        at_cities=returnNth(citiesData, 0, 4)
        for j in range(0, len(at_cities)):
            totalCities = totalCities + at_cities[j]
```

```
average_temperature = totalWeightSum/totalCities
weightPrecipation = []
pt_cities = []
totalSumPrecipitation = 0
totalSumPrecipitationOfCities = 0

for k in range(0, len(wgt)):
      weightPrecipation=returnNth(wgt, 3, 4)
      for j in range(0, len(weightPrecipation)):
          totalSumPrecipitation = totalSumPrecipitation + weightPrecipation[j]

for k in range(0, len(citiesData)):
       pretp_cities=returnNth(citiesData, 3, 4)
       for j in range(0, len(pretp_cities)):
           totalSumPrecipitationOfCities = totalSumPrecipitationOfCities + pretp_cities[j]

prep_temp = totalSumPrecipitation/totalSumPrecipitationOfCities
print("-------------------------------------------------------------")
print(" | " + "The Average Temperature for" + " | " +i+"\t"+ " | " +
str(round(average_temperature, 2)), "\tF" , " | ")
print("-------------------------------------------------------------")
print(" | " + "The Average Precipitation " + " | " +i+"\t" + " | " + str(round(prep_temp, 2)),"\tIN",
" | ")
```

OUTPUT

```
-------------------------------------------------------------------------------

| The Average Temperature for | ANNUAL | 37.95 F |

-------------------------------------------------------------------------------

| The Average Precipitation | ANNUAL | 34.47 IN |

-------------------------------------------------------------------------------

| The Average Temperature for | JAN | 12.11 F |

-------------------------------------------------------------------------------

| The Average Precipitation | JAN | 3.21 IN |

-------------------------------------------------------------------------------
```

| The Average Temperature for | FEB | 15.52 F |

| The Average Precipitation | FEB | 2.29 IN |

| The Average Temperature for | MAR | 24.63 F |

| The Average Precipitation | MAR | 2.42 IN |

| The Average Temperature for | APR | 37.36 F |

| The Average Precipitation | APR | 2.35 IN |

| The Average Temperature for | MAY | 48.42 F |

| The Average Precipitation | MAY | 2.74 IN |

| The Average Temperature for | JUN | 57.05 F |

| The Average Precipitation | JUN | 3.15 IN |

| The Average Temperature for | JUL | 62.26 F |

| The Average Precipitation | JUL | 3.05 IN |

| The Average Temperature for | AUG | 60.89 F |

| The Average Precipitation | AUG | 2.88 IN |

| The Average Temperature for | SEP | 52.42 F |

| The Average Precipitation | SEP | 2.93 IN |

| The Average Temperature for | OCT | 41.21 F |

| The Average Precipitation | OCT | 3.16 IN |

| The Average Temperature for | NOV | 28.09 F |

| The Average Precipitation | NOV | 3.44 IN |

| The Average Temperature for | DEC | 16.98 F |

| The Average Precipitation | DEC | 3.15 IN |

PS C:\PythonAssignment\databricksdemo> SUCCESS: The process with PID 4236 (child process of PID 14212) has been terminated.

SUCCESS: The process with PID 14212 (child process of PID 9768) has been terminated.
SUCCESS: The process with PID 9768 (child process of PID 19036) has been terminated.