

The problem with the `pd.read_excel()` is that pandas no longer supports anything other than xls files in their latest version.

To work around this do the following:

- 1) Install the openpyxl engine using the following command

```
python3 -m pip install openpyxl
```

- 2) In the `pd.read_excel` function, pass an additional engine parameter and set the engine to openpyxl

```
df = pd.read_excel('german.xlsx', engine='openpyxl')
```

You will now be able to read excel files to pandas

Map Reduce program example

testmapper.py

```
testmapper.py > ...
1  import pandas as pd
2  df = pd.read_excel('german.xlsx', engine='openpyxl')
3  for country in df['CreditAmount']:
4
5      print('%d\t%d' %(country, 1))
```

testreducer.py

```
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('%s\t%d' %(lastWord, sum))
        sum = count
        lastWord = word

if lastWord == word:
    print('%s\t%s' %(lastWord, sum))
```

OUTPUT

```
886      1
888      1
894      1
8947     1
8978     1
900      1
902      1
9034     1
9055     1
907      1
909      1
915      1
9157     1
918      1
926      1
9271     1
9277     1
9283     1
929      1
930      1
931      1
932      1
935      1
936      1
937      1
939      1
9398     1
9436     1
947      1
950      1
951      1
9566     1
9572     1
958      1
959      1
960      1
9629     1
975      1
976      2
983      1
9857     1
9960     1
996      1
999      1
200905124@chick-for-updates:~/200905124/week5$
```

Program demonstrating plt function on excel file

```
testplot.py > ...  
1  import pandas as pd  
2  import matplotlib.pyplot as plt  
3  df = pd.read_excel('german.xlsx',engine='openpyxl')  
4  
5  plt.scatter(df['CreditAmount'],df['DurationOfCreditInMonths'])  
6  print(plt.show())
```

Output

