

Name: Muhammad Sheroz

SID: 9852

Big Data Quiz 1

```
In [ ]: from pyspark.sql import SparkSession
```

```
In [ ]: spark = SparkSession.builder.appName("pySPWordCOUNTCSV").getOrCreate()
```

```
In [ ]: df = spark.read.csv('E:\Semesters\Summer 2022\Bid_data\100_sales_records.csv', header =
```

```
In [ ]: #df.show()
```

```
In [ ]: dfrdd = df.rdd
```

```
In [ ]: jobMap = dfrdd.map(lambda x: [x[0], float(x[8])* float(x[9])])

#smSal = jobMap.reduceByKey(lambda a,b: a+b)
```

```
In [ ]: jbMp = jobMap.collect()
jbMp
```

```
Out[ ]: [['Australia and Oceania', 2533654.0],
['Central America and the Caribbean', 576782.7999999999],
['Europe', 1158502.59],
['Sub-Saharan Africa', 75591.66],
['Sub-Saharan Africa', 3296425.02],
['Australia and Oceania', 759202.72],
['Sub-Saharan Africa', 2798046.4899999998],
['Sub-Saharan Africa', 1245112.92],
['Sub-Saharan Africa', 496101.10000000003],
['Sub-Saharan Africa', 1356180.0999999999],
['Asia', 19103.44],
['Sub-Saharan Africa', 455479.04],
['Asia', 902980.64],
['Central America and the Caribbean', 5997054.9799999995],
['Asia', 400558.73000000004],
['Europe', 182825.44],
['Asia', 3039414.4],
['Sub-Saharan Africa', 257653.50000000003],
['Asia', 2559474.1],
['Australia and Oceania', 2492526.12],
['Europe', 1901836.0],
['Europe', 324971.44],
['Central America and the Caribbean', 339490.5],
['Australia and Oceania', 20404.71],
['Europe', 414371.10000000003],
['Europe', 793518.0],
['Australia and Oceania', 50363.340000000004],
['Sub-Saharan Africa', 54319.26],
['Europe', 243133.80000000002],
['Sub-Saharan Africa', 1583799.9],
['Europe', 3786589.1999999997],
['Sub-Saharan Africa', 173676.25],
['Australia and Oceania', 1904138.04],
['Asia', 5513227.5],
['Sub-Saharan Africa', 1117953.6600000001],
```

```

['Central America and the Caribbean', 523807.57],
['Middle East and North Africa', 35304.72],
['Sub-Saharan Africa', 2011149.63],
['Asia', 4368316.680000001],
['Europe', 2596374.27],
['Sub-Saharan Africa', 246415.95],
['Middle East and North Africa', 3162704.8],
['Sub-Saharan Africa', 435466.89999999997],
['Europe', 26344.260000000002],
['Asia', 648030.4],
['Sub-Saharan Africa', 197883.4],
['Europe', 3876652.4],
['Europe', 22312.29],
['Europe', 856973.76],
['Sub-Saharan Africa', 824431.8600000001],
['Europe', 247956.32],
['Sub-Saharan Africa', 89623.98],
['Middle East and North Africa', 835759.1],
['Sub-Saharan Africa', 2251232.97],
['Sub-Saharan Africa', 71253.21],
['Sub-Saharan Africa', 380512.96],
['Australia and Oceania', 1082418.4],
['Europe', 1244708.4],
['Europe', 188452.13999999998],
['Sub-Saharan Africa', 3154398.0],
['Australia and Oceania', 140287.4],
['Europe', 1212580.0],
['Sub-Saharan Africa', 3593376.7800000003],
['Middle East and North Africa', 668356.48],
['Central America and the Caribbean', 745426.0],
['Sub-Saharan Africa', 1957344.4],
['Sub-Saharan Africa', 707454.88],
['Central America and the Caribbean', 600821.4400000001],
['Europe', 5396577.2700000005],
['Sub-Saharan Africa', 802333.76],
['Asia', 3262562.1],
['Middle East and North Africa', 6279.09],
['Sub-Saharan Africa', 272410.45],
['Sub-Saharan Africa', 1780539.2],
['Middle East and North Africa', 4324782.399999999],
['North America', 4647149.58],
['Australia and Oceania', 445033.55000000005],
['Asia', 574951.92],
['Europe', 2198981.92],
['Australia and Oceania', 4220728.8],
['Europe', 3015902.51],
['Middle East and North Africa', 861563.52],
['Middle East and North Africa', 2836990.8],
['Sub-Saharan Africa', 623289.3],
['Sub-Saharan Africa', 994765.42],
['North America', 524870.06],
['Sub-Saharan Africa', 418936.05000000005],
['Sub-Saharan Africa', 1419101.52],
['Middle East and North Africa', 4870.26],
['Europe', 221117.0],
['Sub-Saharan Africa', 617347.0800000001],
['Australia and Oceania', 445508.05000000005],
['Middle East and North Africa', 1316095.4100000001],
['Europe', 3458252.0],
['Central America and the Caribbean', 387002.2],
['Sub-Saharan Africa', 97040.64],
['Asia', 58471.11],
['Sub-Saharan Africa', 228779.1],
['North America', 471336.91000000003],
['Sub-Saharan Africa', 3586605.09]]

```

```
In [ ]: #maxsale = sorted(jobMap.collect())  
        #print(len,maxsale)  
        #print(maxsale)
```

```
In [ ]: from wordcloud import WordCloud  
        wordcloud = WordCloud(background_color = 'white', width =1000, height=500).generate(str
```

```
In [ ]: import matplotlib.pyplot as plt
```

```
In [ ]: plt.imshow(wordcloud, interpolation= 'bilinear')  
        plt.axis('off')  
        plt.rcParams['figure.figsize']= [15,0]
```

