from pyspark import SparkContext, SparkConf

conf = SparkConf().setAppName("product-analysis").setMaster("local")

sc = SparkContext(conf=conf)

input\_file = "mindtree.csv"

products = sc.textFile(input\_file)

header = products.first()

products = products.filter(lambda x: x != header) # Filter out the header row

products = products.map(lambda x: x.split(","))

expensive\_products = products.filter(lambda x: float(x[3]) > 1000)

category\_products = expensive\_products.map(lambda x: (x[2], (int(x[4]), float(x[3]))))

category\_totals = category\_products.reduceByKey(lambda x, y: (x[0] + y[0], x[1] + y[1]))

category\_revenue = category\_totals.mapValues(lambda x: (x[0], round(x[1] \* x[0], 2)))

output\_file = "category\_totals.csv"

category\_revenue.map(lambda x: ",".join([str(e) for e in x])).saveAsTextFile(output\_file)

sc.stop()