**W205 Section 3 Lab 4**

**Sanjay Dorairaj**

# Create an RDD with tuples were the there is a key and a value. But in contrast to the example above the key is removed from the value portion of the key-value tuple. Submit the code and a print out of the first tuple.

|  |
| --- |
| >>> narcoticsCrimeTuples=narcoticsCrimes.map(lambda x:(x.split(",")[0],x.split(",")[1:]))  >>> print narcoticsCrimeTuples.first()  (u'10184515', [u'HY372204', u'08/06/2015 11:55:00 PM', u'033XX W DIVERSEY AVE', u'2027', u'NARCOTICS', u'POSS: CRACK', u'STREET', u'true', u'false', u'1412', u'014', u'35', u'22', u'18', u'1153440', u'1918377', u'2015', u'08/13/2015 12:57:42 PM', u'41.931870591', u'-87.711546895', u'"(41.931870591', u' -87.711546895)"'])  >>> |

# **SUBMISSION 2:** Submit the the code for executing the above query as a Spark SQL python call. Also submit the number of rows in the result

|  |
| --- |
| >>> query='select count(\*) from Web\_Session\_Log where REFERERURL="http://www.ebay.com"'  >>> results=sqlContext.sql(query)  >>> results.show()  +----+  | \_c0|  +----+  |3943|  +----+ |