SUBMISSION 1:

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.

I went about this a little different:

I defined a function to create tuples from RDD:

>>> def tuple_maker(item): return item[0], item[1:]

Then I applied it to first entry to test:

>>> test = narcoticsCrimeRecords.first()

>>> first_tuple = tuple_maker(test)

>>> first_tuple[0]
u'10184515'

>>> first_tuple[1]
[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.

Spark SQL python call:

>>> results = sqlContext.sql('SELECT count(*) from Web_Session_Log where REFERERURL = "http://www.ebay.com"')

>>> results.show()

Number of results:

```
+----+
| _c0|
+----+
|3943|
+——+
```

Screenshot of additional query: