Project 2 - Hadoop and Spark

In [1]:

```
###NOTE### This is running on Anaconda3 - Python 3.5.4 and version 2.3.0 Spark Apache
```

In [2]:

```
# Must adjust memory management properties accordingly
```

In [3]:

```
import findspark
findspark.init()

from pyspark import SparkContext
from pyspark import SparkConf
```

In [4]:

```
conf = SparkConf().setMaster("local").setAppName("Project2SPARKH").set('spark.executor.mem
ory', '4G').set('NotebookApp.iopub_data_rate_limit','99999999999').set('NotebookApp.rate_l
imit_window', '5.0').set('spark.driver.memory', '45g').set('spark.driver.maxReultSize', '1
0g')
```

In [5]:

```
sc1 = SparkContext(conf=conf)
```

In [6]:

```
sc1
```

Out[6]:

SparkContext

Spark UI

Version

v2.3.0

Master

local

AppName

Project2SPARKHADOOP

In [7]:

```
## Import necessary modules
```

In [8]:

```
import findspark
findspark.init()
import pyspark
from pyspark.sql import SparkSession
from pyspark.sql import SQLContext
from pyspark.sql.types import StructType, StructField, IntegerType, StringType
from pyspark.sql.types import *
import pandas as pd
from math import sqrt
from numpy import array
from pyspark.ml.clustering import KMeans
from pyspark import SparkContext
sc = SparkContext.getOrCreate()
sqlContext = SQLContext(sc)
from pyspark.ml.feature import OneHotEncoder, StringIndexer, VectorAssembler
from pyspark.ml import Pipeline
from pyspark.mllib.clustering import KMeansModel
import numpy as np
import pandas as pd
from pyspark import SparkContext
from pyspark.ml.clustering import KMeans
from pyspark.ml.feature import VectorAssembler
from pyspark.sql import SQLContext
```

In [9]:

```
sqlContext = SQLContext(sc)
```

In [10]:

```
## Import dataset through Pandas then convert to a Pyspark DataFrame
```

In [11]:

```
jack = pd.read_csv("Dataset2.csv")
```

In [12]:

```
jacket = sqlContext.createDataFrame(jack, ["_c0","Timestamp","Name","Code","Latitude","Lon
gitude"])
```

In [13]:

```
## Choosing which columns will be incorporated for the conversion to create the clusters
## Latitude and Longitude
## Naming on variable to another variable (df = jacket)
```

In [14]:

```
FEATURES_COL = ['Latitude', 'Longitude']
```

In [15]:

```
df = jacket
```

In [16]:

```
+---+-------------------+--------+--------------------+-----------------+---------------
---+
|_c0|          Timestamp|    Name|                Code|         Latitude|          Longit
ude|
+---+-------------------+--------+--------------------+-----------------+---------------
---+
|  0|2014-03-15:10:10:20|Sorrento|8cc3b47e-bd01-448...|     33.6894754264|-117.54330825299
999|
|  1|2014-03-15:10:10:20|  MeeToo|ef8c7564-0a1a-465...|     37.4321088904|     -121.485029
632|
|  2|2014-03-15:10:10:20|  MeeToo|23eba027-b95a-472...|39.437890834899996|-120.93897848600
001|
|  3|2014-03-15:10:10:20|Sorrento|707daba1-5640-4d6...|39.363518676700004|-119.40033470799
999|
|  4|2014-03-15:10:10:20|   Ronin|db66fe81-aa55-43b...|     33.1913581092|-116.44824264299
999|
|  5|2014-03-15:10:10:20|Sorrento|ffa18088-69a0-433...|     33.8343543748|     -117.330000
857|
|  6|2014-03-15:10:10:20|Sorrento|66d678e6-9c87-48d...|     37.3803954321|     -121.840756
755|
|  7|2014-03-15:10:10:20|  MeeToo|673f7e4b-d52b-44f...|     34.1841062345|       -117.9435
329|
|  8|2014-03-15:10:10:20|   Ronin|a678ccc3-b0d2-452...|     32.2850556785|-111.81958373399
999|
|  9|2014-03-15:10:10:20|Sorrento|86bef6ae-2f1c-42e...|     45.2400522984|     -122.377467
861|
| 10|2014-03-15:10:10:20|  iFruit|27178d24-3a61-42f...|     37.9248961741|-122.20686816700
001|
| 11|2014-03-15:10:10:20| Titanic|b4a15931-9a69-469...|     38.1653163975|-122.15160837799
999|
| 12|2014-03-15:10:10:20|   Ronin|e75dc777-b531-4db...|      33.323126641|     -116.472234
745|
| 13|2014-03-15:10:10:20|   Ronin|d4ebd9ae-4dad-4fb...|     33.1774985363|     -116.889226
299|
| 14|2014-03-15:10:10:20|   Ronin|b954db08-1f97-431...|     32.2083493316|-111.43410271299
999|
| 15|2014-03-15:10:10:20|  MeeToo|16085fbf-cda5-448...|     34.0487620041|     -111.928871
717|
| 16|2014-03-15:10:10:20|  iFruit|6474caf1-7bbf-459...|     37.9031053656|-121.56145134200
001|
| 17|2014-03-15:10:10:20|  MeeToo|668e6f06-a8aa-4be...|      36.032967794|     -118.970108
886|
| 18|2014-03-15:10:10:20|   Ronin|6d195272-8dba-42d...|     45.0400810371|     -117.858004
521|
| 19|2014-03-15:10:10:20|Sorrento|d228cdab-8b35-473...|     35.2338863976|       -114.3057
523|
+---+-------------------+--------+--------------------+-----------------+---------------
---+
only showing top 20 rows
```

In [17]:

```
## First step before generating clusters based on features is to convert Latitude and Long
itude to float data types
```

In [18]:

```
df_feat = df.select(*(df[c].cast("float").alias(c) for c in df.columns[1:]))
```

In [19]:

```
df_feat.show()
```

```
+---------+----+----+---------+-----------+
|Timestamp|Name|Code| Latitude|  Longitude|
+---------+----+----+---------+-----------+
|     null|null|null|33.689476|-117.543304|
|     null|null|null| 37.43211| -121.48503|
|     null|null|null| 39.43789| -120.93898|
|     null|null|null|39.363518| -119.40034|
|     null|null|null|33.191357| -116.44824|
|     null|null|null|33.834354|    -117.33|
|     null|null|null|37.380394| -121.84076|
|     null|null|null|34.184105|-117.943535|
|     null|null|null|32.285057| -111.81958|
|     null|null|null| 45.24005|-122.377464|
|     null|null|null|37.924896| -122.20687|
|     null|null|null|38.165318| -122.15161|
|     null|null|null|33.323128| -116.47224|
|     null|null|null|33.177498| -116.88923|
|     null|null|null| 32.20835|-111.434105|
|     null|null|null|34.048763| -111.92887|
|     null|null|null|37.903107|-121.561455|
|     null|null|null|36.032967| -118.97011|
|     null|null|null| 45.04008|   -117.858|
|     null|null|null|35.233887|-114.305756|
+---------+----+----+---------+-----------+
only showing top 20 rows
```

In [20]:

```
for col in df.columns:
    if col in FEATURES_COL:
        df = df.withColumn(col,df[col].cast('float'))
```

In [21]:

```
df.show()
```

```
+---+-------------------+--------+-------------------+---------+-----------+
|_c0|          Timestamp|    Name|               Code| Latitude|  Longitude|
+---+-------------------+--------+-------------------+---------+-----------+
|  0|2014-03-15:10:10:20|Sorrento|8cc3b47e-bd01-448...|33.689476|-117.543304|
|  1|2014-03-15:10:10:20|  MeeToo|ef8c7564-0a1a-465...| 37.43211| -121.48503|
|  2|2014-03-15:10:10:20|  MeeToo|23eba027-b95a-472...| 39.43789| -120.93898|
|  3|2014-03-15:10:10:20|Sorrento|707daba1-5640-4d6...|39.363518| -119.40034|
|  4|2014-03-15:10:10:20|   Ronin|db66fe81-aa55-43b...|33.191357| -116.44824|
|  5|2014-03-15:10:10:20|Sorrento|ffa18088-69a0-433...|33.834354|    -117.33|
|  6|2014-03-15:10:10:20|Sorrento|66d678e6-9c87-48d...|37.380394| -121.84076|
|  7|2014-03-15:10:10:20|  MeeToo|673f7e4b-d52b-44f...|34.184105|-117.943535|
|  8|2014-03-15:10:10:20|   Ronin|a678ccc3-b0d2-452...|32.285057| -111.81958|
|  9|2014-03-15:10:10:20|Sorrento|86bef6ae-2f1c-42e...| 45.24005|-122.377464|
| 10|2014-03-15:10:10:20|  iFruit|27178d24-3a61-42f...|37.924896| -122.20687|
| 11|2014-03-15:10:10:20| Titanic|b4a15931-9a69-469...|38.165318| -122.15161|
| 12|2014-03-15:10:10:20|   Ronin|e75dc777-b531-4db...|33.323128| -116.47224|
| 13|2014-03-15:10:10:20|   Ronin|d4ebd9ae-4dad-4fb...|33.177498| -116.88923|
| 14|2014-03-15:10:10:20|   Ronin|b954db08-1f97-431...| 32.20835|-111.434105|
| 15|2014-03-15:10:10:20|  MeeToo|16085fbf-cda5-448...|34.048763| -111.92887|
| 16|2014-03-15:10:10:20|  iFruit|6474caf1-7bbf-459...|37.903107|-121.561455|
| 17|2014-03-15:10:10:20|  MeeToo|668e6f06-a8aa-4be...|36.032967| -118.97011|
| 18|2014-03-15:10:10:20|   Ronin|6d195272-8dba-42d...| 45.04008|   -117.858|
| 19|2014-03-15:10:10:20|Sorrento|d228cdab-8b35-473...|35.233887|-114.305756|
+---+-------------------+--------+-------------------+---------+-----------+
only showing top 20 rows
```

In [22]:

```
df = df.na.drop()
```

In [23]:

```
+---+-------------------+--------+--------------------+---------+-----------+
|_c0|          Timestamp|    Name|                Code| Latitude|  Longitude|
+---+-------------------+--------+--------------------+---------+-----------+
|  0|2014-03-15:10:10:20|Sorrento|8cc3b47e-bd01-448...|33.689476|-117.543304|
|  1|2014-03-15:10:10:20|  MeeToo|ef8c7564-0a1a-465...| 37.43211| -121.48503|
|  2|2014-03-15:10:10:20|  MeeToo|23eba027-b95a-472...| 39.43789| -120.93898|
|  3|2014-03-15:10:10:20|Sorrento|707daba1-5640-4d6...|39.363518| -119.40034|
|  4|2014-03-15:10:10:20|   Ronin|db66fe81-aa55-43b...|33.191357| -116.44824|
|  5|2014-03-15:10:10:20|Sorrento|ffa18088-69a0-433...|33.834354|    -117.33|
|  6|2014-03-15:10:10:20|Sorrento|66d678e6-9c87-48d...|37.380394| -121.84076|
|  7|2014-03-15:10:10:20|  MeeToo|673f7e4b-d52b-44f...|34.184105|-117.943535|
|  8|2014-03-15:10:10:20|   Ronin|a678ccc3-b0d2-452...|32.285057| -111.81958|
|  9|2014-03-15:10:10:20|Sorrento|86bef6ae-2f1c-42e...| 45.24005|-122.377464|
| 10|2014-03-15:10:10:20|  iFruit|27178d24-3a61-42f...|37.924896| -122.20687|
| 11|2014-03-15:10:10:20| Titanic|b4a15931-9a69-469...|38.165318| -122.15161|
| 12|2014-03-15:10:10:20|   Ronin|e75dc777-b531-4db...|33.323128| -116.47224|
| 13|2014-03-15:10:10:20|   Ronin|d4ebd9ae-4dad-4fb...|33.177498| -116.88923|
| 14|2014-03-15:10:10:20|   Ronin|b954db08-1f97-431...| 32.20835|-111.434105|
| 15|2014-03-15:10:10:20|  MeeToo|16085fbf-cda5-448...|34.048763| -111.92887|
| 16|2014-03-15:10:10:20|  iFruit|6474caf1-7bbf-459...|37.903107|-121.561455|
| 17|2014-03-15:10:10:20|  MeeToo|668e6f06-a8aa-4be...|36.032967| -118.97011|
| 18|2014-03-15:10:10:20|   Ronin|6d195272-8dba-42d...| 45.04008|   -117.858|
| 19|2014-03-15:10:10:20|Sorrento|d228cdab-8b35-473...|35.233887|-114.305756|
+---+-------------------+--------+--------------------+---------+-----------+
only showing top 20 rows
```

In [24]:

```
## Vectorizing the "FEATURES_COL" variable
## selecting all columns to be displayed
```

In [25]:

```
vecAssembler = VectorAssembler(inputCols=FEATURES_COL, outputCol="Features")
df_kmeans = vecAssembler.transform(df).select('_c0',  'Timestamp', 'Name', 'Code', 'Latitu
de', 'Longitude','Features')
```

In [26]:

```
+---+-------------------+--------+--------------------+---------+-----------+-------------
-------+
|_c0|          Timestamp|    Name|                Code| Latitude|  Longitude|            F
eatures|
+---+-------------------+--------+--------------------+---------+-----------+-------------
-------+
|  0|2014-03-15:10:10:20|Sorrento|8cc3b47e-bd01-448...|33.689476|-117.543304|[33.689476013
1835...|
|  1|2014-03-15:10:10:20|  MeeToo|ef8c7564-0a1a-465...| 37.43211| -121.48503|[37.432109832
7636...|
|  2|2014-03-15:10:10:20|  MeeToo|23eba027-b95a-472...| 39.43789| -120.93898|[39.437889099
1210...|
|  3|2014-03-15:10:10:20|Sorrento|707daba1-5640-4d6...|39.363518| -119.40034|[39.363517761
2304...|
|  4|2014-03-15:10:10:20|   Ronin|db66fe81-aa55-43b...|33.191357| -116.44824|[33.191356658
9355...|
|  5|2014-03-15:10:10:20|Sorrento|ffa18088-69a0-433...|33.834354|    -117.33|[33.834354400
6347...|
|  6|2014-03-15:10:10:20|Sorrento|66d678e6-9c87-48d...|37.380394| -121.84076|[37.380393981
9335...|
|  7|2014-03-15:10:10:20|  MeeToo|673f7e4b-d52b-44f...|34.184105|-117.943535|[34.184104919
4335...|
|  8|2014-03-15:10:10:20|   Ronin|a678ccc3-b0d2-452...|32.285057| -111.81958|[32.285057067
8710...|
|  9|2014-03-15:10:10:20|Sorrento|86bef6ae-2f1c-42e...| 45.24005|-122.377464|[45.240051269
5312...|
| 10|2014-03-15:10:10:20|  iFruit|27178d24-3a61-42f...|37.924896| -122.20687|[37.924896240
2343...|
| 11|2014-03-15:10:10:20| Titanic|b4a15931-9a69-469...|38.165318| -122.15161|[38.165317535
4003...|
| 12|2014-03-15:10:10:20|   Ronin|e75dc777-b531-4db...|33.323128| -116.47224|[33.323127746
5820...|
| 13|2014-03-15:10:10:20|   Ronin|d4ebd9ae-4dad-4fb...|33.177498| -116.88923|[33.177497863
7695...|
| 14|2014-03-15:10:10:20|   Ronin|b954db08-1f97-431...| 32.20835|-111.434105|[32.208351135
2539...|
| 15|2014-03-15:10:10:20|  MeeToo|16085fbf-cda5-448...|34.048763| -111.92887|[34.048763275
1464...|
| 16|2014-03-15:10:10:20|  iFruit|6474caf1-7bbf-459...|37.903107|-121.561455|[37.903106689
4531...|
| 17|2014-03-15:10:10:20|  MeeToo|668e6f06-a8aa-4be...|36.032967| -118.97011|[36.032966613
7695...|
| 18|2014-03-15:10:10:20|   Ronin|6d195272-8dba-42d...| 45.04008|   -117.858|[45.040081024
1699...|
| 19|2014-03-15:10:10:20|Sorrento|d228cdab-8b35-473...|35.233887|-114.305756|[35.233886718
75,-...|
+---+-------------------+--------+--------------------+---------+-----------+-------------
-------+
only showing top 20 rows
```

In [27]:

```
## k in the KMeans function is set to 3 as the project required
### There are three groups = 0, 1, and 2
## Seed is set to 300
## the Prediction Column is named as "Group_Number"

## Cluster Centers are displayed
```

In [28]:

```
k = 3
kmeans = KMeans().setK(k).setSeed(100).setFeaturesCol("Features").setPredictionCol("Group_
Number")
model = kmeans.fit(df_kmeans)

centers = model.clusterCenters()

print("Cluster Centers: ")
for center in centers:
    print(center)
```

```
Cluster Centers:
[  34.52887063 -116.34533272]
[ 0.  0.]
[  39.57394629 -121.24864998]
```

In [29]:

```
## Displaying certain columns in the PySpark DataFrame for proper display - one of which w
ill be the cluster number
## otherwise known as "Group_Number"
## three rows are displayed
```

In [43]:

```
transformed = model.transform(df_kmeans).select('_c0', 'Name', 'Code', 'Latitude', 'Longit
ude','Features', 'Group_Number')
rows = transformed.collect()
```

In [44]:

```
[Row(_c0=0, Name='Sorrento', Code='8cc3b47e-bd01-4482-b500-28f2342679af', Latitude=33.6894
76013183594, Longitude=-117.54330444335938, Features=DenseVector([33.6895, -117.5433]), Gr
oup_Number=0), Row(_c0=1, Name='MeeToo', Code='ef8c7564-0a1a-4650-a655-c8bbd5f8f943', Lati
tude=37.43210983276367, Longitude=-121.48503112792969, Features=DenseVector([37.4321, -12
1.485]), Group_Number=2), Row(_c0=2, Name='MeeToo', Code='23eba027-b95a-4729-9a4b-a3cca51c
5548', Latitude=39.437889099121094, Longitude=-120.93898010253906, Features=DenseVector([3
9.4379, -120.939]), Group_Number=2)]
```

In [45]:

```
df_pred = sqlContext.createDataFrame(rows)
```

In [46]:

```
+---+--------+--------------------+-----------------+------------------+---------------
----+------------+
|_c0|    Name|                Code|         Latitude|         Longitude|            Feat
ures|Group_Number|
+---+--------+--------------------+-----------------+------------------+---------------
----+------------+
|  0|Sorrento|8cc3b47e-bd01-448...|33.689476013183594|-117.54330444335938|[33.689476013183
5...|           0|
|  1|  MeeToo|ef8c7564-0a1a-465...| 37.43210983276367|-121.48503112792969|[37.432109832763
6...|           2|
|  2|  MeeToo|23eba027-b95a-472...|39.437889099121094|-120.93898010253906|[39.437889099121
0...|           2|
|  3|Sorrento|707daba1-5640-4d6...| 39.36351776123047|-119.40033721923828|[39.363517761230
4...|           2|
|  4|   Ronin|db66fe81-aa55-43b...| 33.19135665893555|    -116.4482421875|[33.191356658935
5...|           0|
|  5|Sorrento|ffa18088-69a0-433...|33.834354400634766|-117.33000183105469|[33.834354400634
7...|           0|
|  6|Sorrento|66d678e6-9c87-48d...|37.380393981933594|-121.84075927734375|[37.380393981933
5...|           2|
|  7|  MeeToo|673f7e4b-d52b-44f...|34.184104919433594|-117.94353485107422|[34.184104919433
5...|           0|
|  8|   Ronin|a678ccc3-b0d2-452...|32.285057067871094|  -111.819580078125|[32.285057067871
0...|           0|
|  9|Sorrento|86bef6ae-2f1c-42e...| 45.24005126953125| -122.3774642944336|[45.240051269531
2...|           2|
| 10|  iFruit|27178d24-3a61-42f...|37.924896240234375|-122.20687103271484|[37.924896240234
3...|           2|
| 11| Titanic|b4a15931-9a69-469...| 38.16531753540039|  -122.151611328125|[38.165317535400
3...|           2|
| 12|   Ronin|e75dc777-b531-4db...| 33.32312774658203|-116.47223663330078|[33.323127746582
0...|           0|
| 13|   Ronin|d4ebd9ae-4dad-4fb...| 33.17749786376953|-116.88922882080078|[33.177497863769
5...|           0|
| 14|   Ronin|b954db08-1f97-431...|32.208351135253906| -111.4341049194336|[32.208351135253
9...|           0|
| 15|  MeeToo|16085fbf-cda5-448...|34.048763275146484|-111.92887115478516|[34.048763275146
4...|           0|
| 16|  iFruit|6474caf1-7bbf-459...|37.903106689453125|-121.56145477294922|[37.903106689453
1...|           2|
| 17|  MeeToo|668e6f06-a8aa-4be...| 36.03296661376953|-118.97010803222656|[36.032966613769
5...|           0|
| 18|   Ronin|6d195272-8dba-42d...| 45.04008102416992|-117.85800170898438|[45.040081024169
9...|           2|
| 19|Sorrento|d228cdab-8b35-473...|    35.23388671875|-114.30575561523438|[35.2338867187
5,-...|           0|
+---+--------+--------------------+-----------------+------------------+---------------
----+------------+
only showing top 20 rows
```

In [47]:

```
df_pred.groupBy("Group_Number").count().show()
```

```
+------------+------+
|Group_Number| count|
+------------+------+
|           0|233916|
|           1| 27683|
|           2|197941|
+------------+------+
```

In [48]:

```
## Converted the PySpark DataFrame back to a Pandas DataFrame
## showing the first nine rows
```

In [49]:

```
PandasDF=df_pred.toPandas()
```

In [50]:

Out[50]:

	_c0	Name	Code	Latitude	Longitude	Features	Group_Number
0	0	Sorrento	8cc3b47e-bd01-4482-b500-28f2342679af	33.689476	-117.543304	[33.6894760132, -117.543304443]	0
1	1	MeeToo	ef8c7564-0a1a-4650-a655-c8bbd5f8f943	37.432110	-121.485031	[37.4321098328, -121.485031128]	2
2	2	MeeToo	23eba027-b95a-4729-9a4b-a3cca51c5548	39.437889	-120.938980	[39.4378890991, -120.938980103]	2
3	3	Sorrento	707daba1-5640-4d60-a6d9-1d6fa0645be0	39.363518	-119.400337	[39.3635177612, -119.400337219]	2
4	4	Ronin	db66fe81-aa55-43b4-9418-fc6e7a00f891	33.191357	-116.448242	[33.1913566589, -116.448242188]	0
5	5	Sorrento	ffa18088-69a0-433e-84b8-006b2b9cc1d0	33.834354	-117.330002	[33.8343544006, -117.330001831]	0
6	6	Sorrento	66d678e6-9c87-48d2-a415-8d5035e54a23	37.380394	-121.840759	[37.3803939819, -121.840759277]	2
7	7	MeeToo	673f7e4b-d52b-44fc-8826-aea460c3481a	34.184105	-117.943535	[34.1841049194, -117.943534851]	0
8	8	Ronin	a678ccc3-b0d2-452d-bf89-85bd095e28ee	32.285057	-111.819580	[32.2850570679, -111.819580078]	0

In [51]:

```
PandasDF.to_csv("Project2_file.csv")
```

In [52]:

```
## converting Pandas DataFrame to .csv file ##
```

In [53]:

```
jam=pd.read_csv("Project2_file.csv")
```

In [54]:

Out[54]:

	Unnamed: 0	_c0	Name	Code	Latitude	Longitude	Features	C
0	0	0	Sorrento	8cc3b47e-bd01-4482-b500-28f2342679af	33.689476	-117.543304	[33.6894760132,-117.543304443]	0
1	1	1	MeeToo	ef8c7564-0a1a-4650-a655-c8bbd5f8f943	37.432110	-121.485031	[37.4321098328,-121.485031128]	2
2	2	2	MeeToo	23eba027-b95a-4729-9a4b-a3cca51c5548	39.437889	-120.938980	[39.4378890991,-120.938980103]	2
3	3	3	Sorrento	707daba1-5640-4d60-a6d9-1d6fa0645be0	39.363518	-119.400337	[39.3635177612,-119.400337219]	2
4	4	4	Ronin	db66fe81-aa55-43b4-9418-fc6e7a00f891	33.191357	-116.448242	[33.1913566589,-116.448242188]	0
5	5	5	Sorrento	ffa18088-69a0-433e-84b8-006b2b9cc1d0	33.834354	-117.330002	[33.8343544006,-117.330001831]	0
6	6	6	Sorrento	66d678e6-9c87-48d2-a415-8d5035e54a23	37.380394	-121.840759	[37.3803939819,-121.840759277]	2
7	7	7	MeeToo	673f7e4b-d52b-44fc-8826-aea460c3481a	34.184105	-117.943535	[34.1841049194,-117.943534851]	0
8	8	8	Ronin	a678ccc3-b0d2-452d-bf89-85bd095e28ee	32.285057	-111.819580	[32.2850570679,-111.819580078]	0
9				86bef6ae-				

In [55]:

```
#source:https://rsandstroem.github.io/sparkkmeans.html#
```