# [https://avatars2.githubusercontent.com/u/4156894?v=3&s=100](http://www.calstatela.edu/centers/hipic) CIS5200 Term Project Tutorial

#### Authors: Neel Savla, Neel Patel, Constance Vance

#### Instructor: [Jongwook Woo](https://www.linkedin.com/in/jongwook-woo-7081a85)

#### Date: 12/10/2017

**Lab Tutorial**

Neel Savla ([nsavla@calstatela.edu](mailto:nsavla@calstatela.edu))

Neel Patel ([npatel84@calstatela.edu](mailto:npatel84@calstatela.edu))

Constance Vance ([cvance5@calstatela.edu](mailto:cvance5@calstatela.edu))

12/10/2017

**Chicago Crime From 2001 – Present**

**Objectives**

**List what your objectives are.** In this hands-on lab, you will learn how to:

* Top Crimes Each Year from 2012-2016
* Domestic and Non-Domestic Crimes from 2012-2016
* Likely Areas of Crime
* Police Districts with the Most Arrests.
* Finding the highest crime type in Chicago from 2001 – present date.
* Finding the safest places in Chicago which have the least crimes from year 2001 till present days.
* Finding The Location Where The Police Are Most Active In Chicago From 2001 Till Present Date.
* Number of different crime committed per year in Chicago from 2001 to 2017
* Different types of crime committed from 2001 to 2017 in Chicago.
* Number of the crimes committed in the respective year in Chicago

**Platform Spec**

* IBM Bluemix BigInsights
* Version: IOP 4.2
* Number of CPU : 4VCPU
* Number of nodes: 1 Data Node
* Total Memory Size: 24GB RAM

Step 1: Upload the data

wget <https://data.cityofchicago.org/api/views/ijzp-q8t2/rows.csv?accessType=DOWNLOAD>

Bash$ hdfs dfs -mkdir /tmp/chicagocrimes/

Bash$ hdfs dfs -put Crimes.csv /tmp/chicagocrimes/

Bash$ hive

Step 2: Creating Primary Table

Hive> CREATE EXTERNAL TABLE `crimes\_staging` (

`id` int,

`case\_number` string,

`crime\_date` string,

`block` string,

`iucr` string,

`primary\_type` string,

`description` string,

`location\_description` string,

`arrest` boolean,

`domestic` boolean,

`beat` int,

`district` int,

`ward` int,

`community\_area` int,

`fbi\_code` string,

`x\_coordinate` int,

`y\_coordinate` int,

`year` int,

`updated\_on` string,

`latitude` double,

`longitude` double,

`location` string)

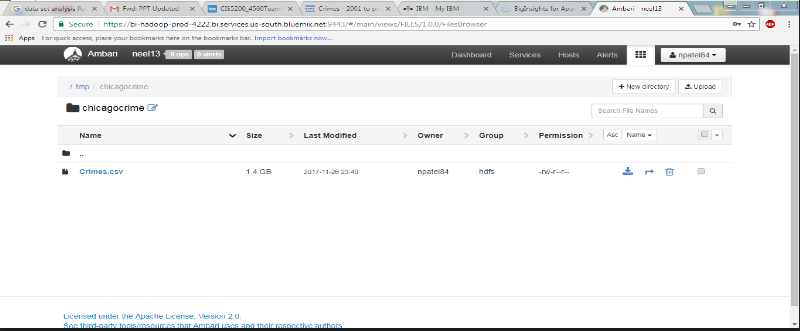
ROW FORMAT DELIMITED

FIELDS TERMINATED BY ","

LOCATION "/tmp/chicagocrime"

TBLPROPERTIES (

'skip.header.line.count' = '1');



Step 3: Visualization

**1.For Maps :**

**Finding the highest crime type in Chicago from 2001 – present date :**

drop table if exists crime\_location;

CREATE TABLE IF NOT EXISTS crime\_location

AS SELECT

id ,case\_number ,crime\_date ,primary\_type ,description ,location\_description ,arrest ,latitude ,longitude

from crimes\_staging;

**1.1For Making 00000-00 file ,for tablue :**

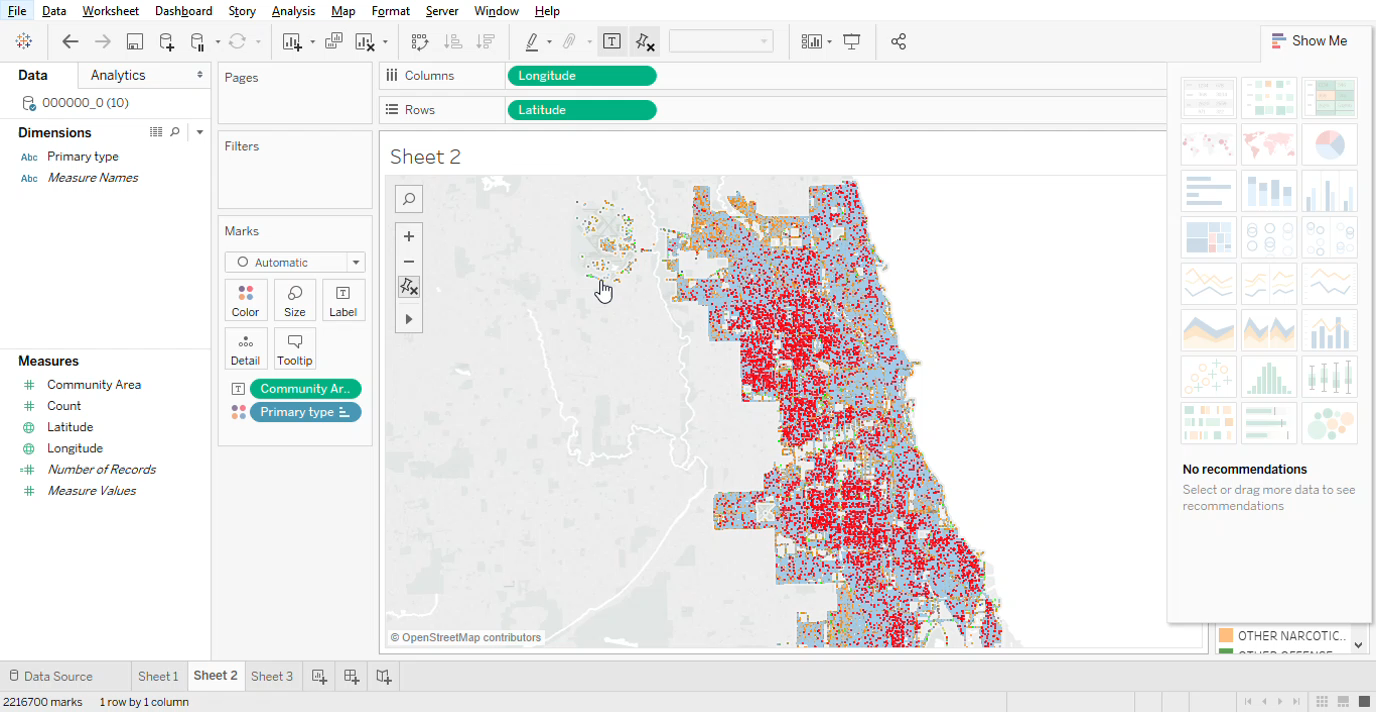
INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime\_location

WHERE latitude is NOT NULL and longitude is NOT NULL and primary\_type is NOT NULL;

**1.2Now Import the 0000-00 file to tablue:**



**2.Finding the safest places in Chicago which have the least crimes from year 2001 till present days.**

drop table if exists crime\_location\_community\_area;

CREATE TABLE IF NOT EXISTS crime\_location\_community\_area

AS SELECT

community\_area,latitude ,longitude, primary\_type, count(community\_area) AS cnt

from crimes\_staging

group by primary\_type,community\_area, latitude, longitude

order by cnt DESC;

**2.1For Making 00000-00 file ,for tablue**

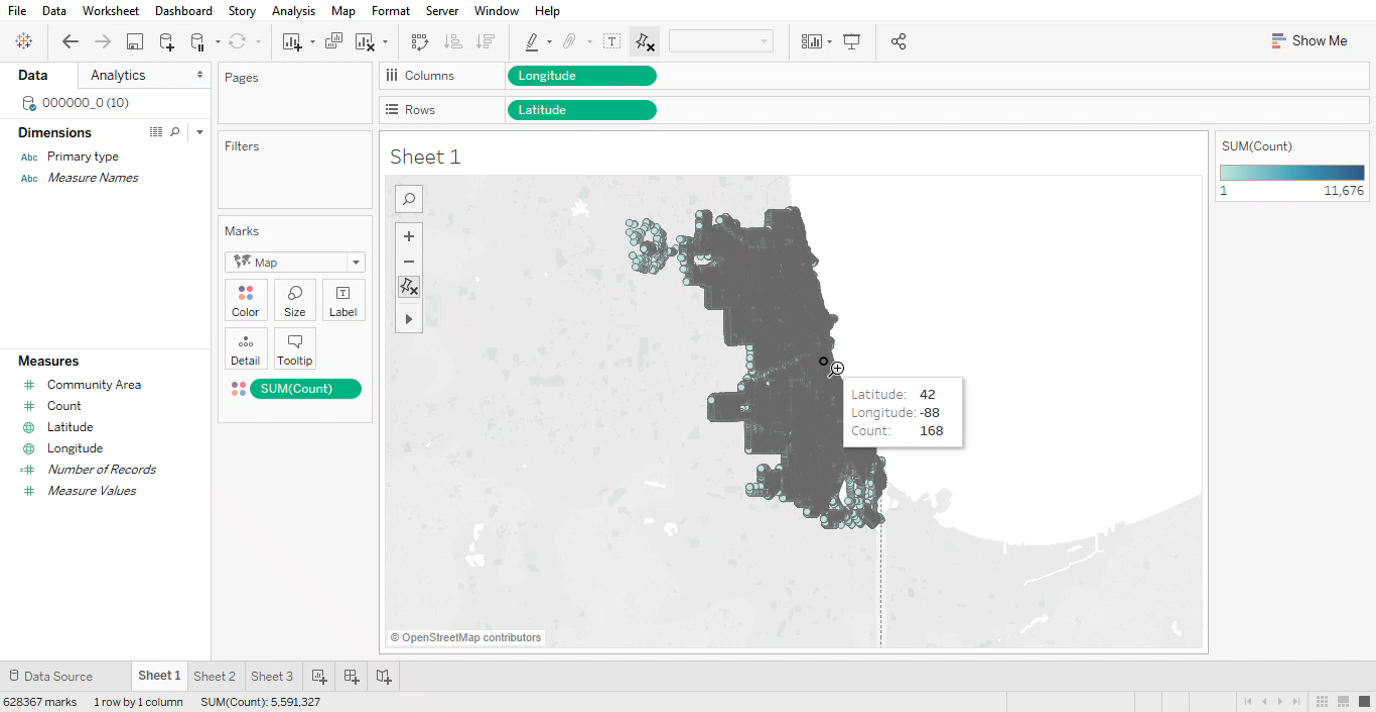
INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime\_location\_community\_area

WHERE latitude is NOT NULL and longitude is NOT NULL and community\_area is NOT NULL;

**2.2Now Import the 0000-00 file to tablue:**



**3.Finding The Location Where The Police Are Most Active In Chicago From 2001 Till Present Date.**

drop table if exists crime\_arrest;

CREATE TABLE IF NOT EXISTS crime\_arrest

AS SELECT

primary\_type ,arrest ,beat,latitude, longitude

from crimes\_staging

group by primary\_type,arrest ,beat,latitude, longitude;

**3.1For Making 00000-00 file ,for tablue**

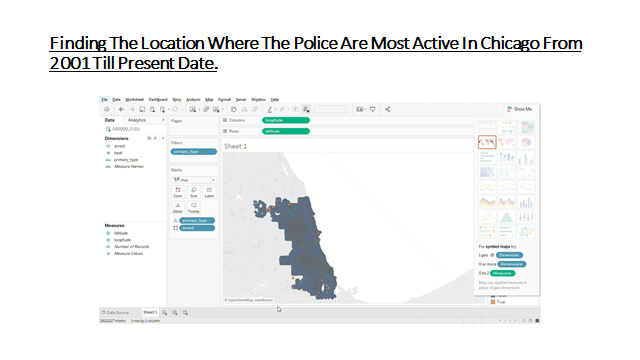
INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime\_arrest

WHERE latitude is NOT NULL and longitude is NOT NULL and arrest is NOT NULL;

**3.2Now Import the 0000-00 file to tablue:**



Step 3.1 Visualizations for Graph :

**1.Number of different crime committed per year in Chicago:**

drop table if exists crime;

CREATE TABLE IF NOT EXISTS crime

AS SELECT

id ,case\_number ,crime\_date ,primary\_type ,description ,arrest

from crimes\_staging;

**1.1For Making 00000-00 file ,for tablue**

INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime

WHERE primary\_type is NOT NULL and year is NOT NULL;

**1.2Now Import the 0000-00 file to tablue:**



**2.The pie chart Illustrate percentage of crime committed from 2001 to 2017 in Chicago :**

drop table if exists crime\_perc;

CREATE TABLE IF NOT EXISTS crime\_ perc

AS SELECT

primary\_type,year, count(arrest) AS cnt

from crimes\_staging

group by primary\_type,year

order by cnt DESC;

**2.1For Making 00000-00 file ,for tablue**

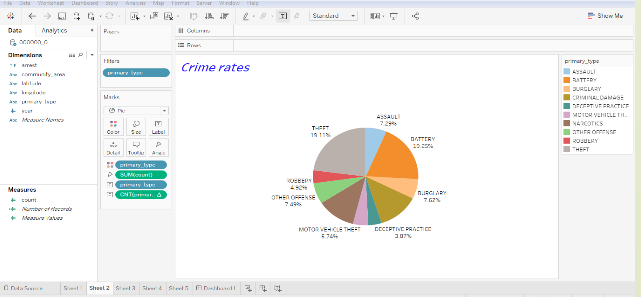
INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime\_perc

WHERE primary\_type is NOT NULL;

**2.2Now Import the 0000-00 file to tablue:**



**3.Different crimes committed in chicago with respect to years from 2001 to present.**

drop table if exists crime\_type\_year;

CREATE TABLE IF NOT EXISTS crime\_type\_year

AS SELECT

primary\_type,year, count(arrest) AS cnt

from crimes\_staging

group by primary\_type,year

order by cnt DESC;

**3.1For Making 00000-00 file ,for tablue**

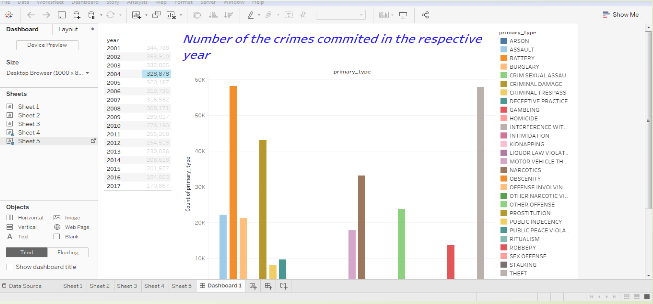
INSERT OVERWRITE DIRECTORY '/tmp/'

ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

SELECT \* FROM crime\_type\_year

WHERE primary\_type is NOT NULL and year is NOT NULL;

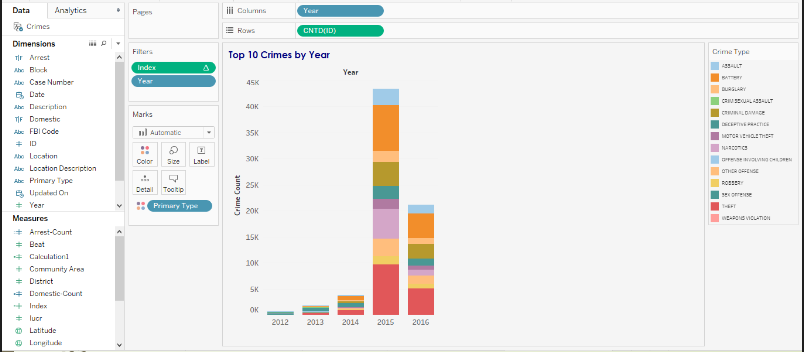
**3.2Now Import the 0000-00 file to tablue:**



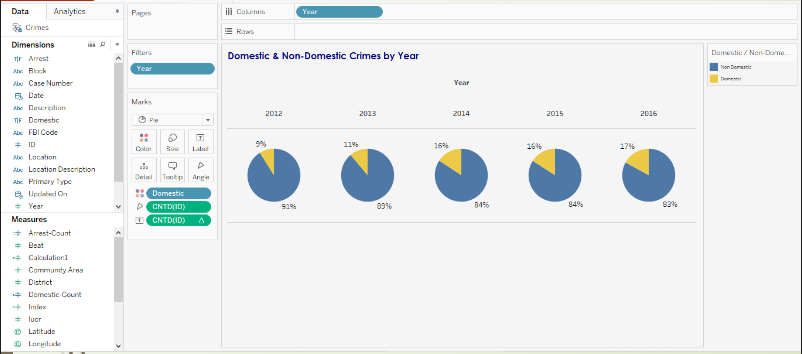
**Top Crimes Each Year from 2012-2016 :**

**The following code was used to create the last two tables in Tableau**

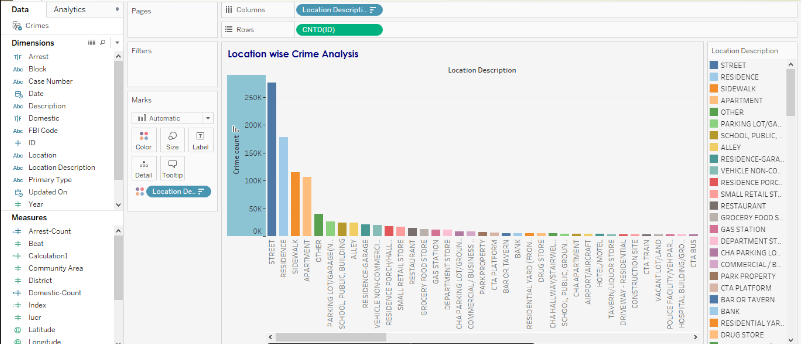
4.1. *create table crimesdata.crimes stored as orc as select \* from crimesdata.crimes\_staging;*



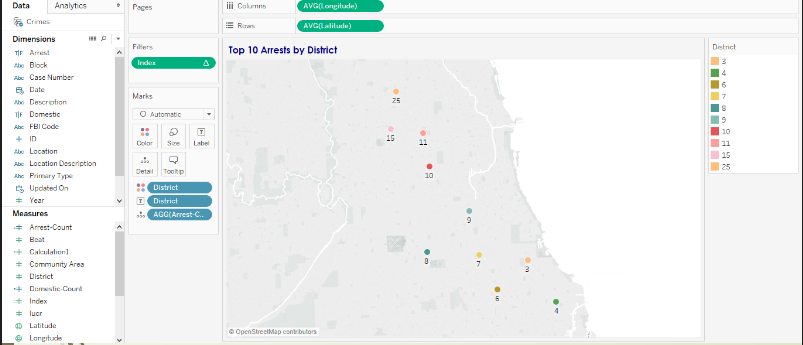
**Domestic and Non-Domestic Crimes from 2012-2016 :**



**Likely Areas of Crime :**



**Police Districts with the Most Arrests:**



References

* 1. URL of Data Source, https://data.cityofchicago.org/api/views/ijzp-q8t2/rows.csv?accessType=DOWNLOAD
  2. https://data.cityofchicago.org
  3. git hub link