CIS5200 Term Project Tutorial

Authors: Neel Savla, Neel Patel, Constance Vance

Instructor: Jongwook Woo

Date: 12/10/2017

Lab Tutorial

Neel Savla (nsavla@calstatela.edu)

Neel Patel (npatel84@calstatela.edu)

Constance Vance (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

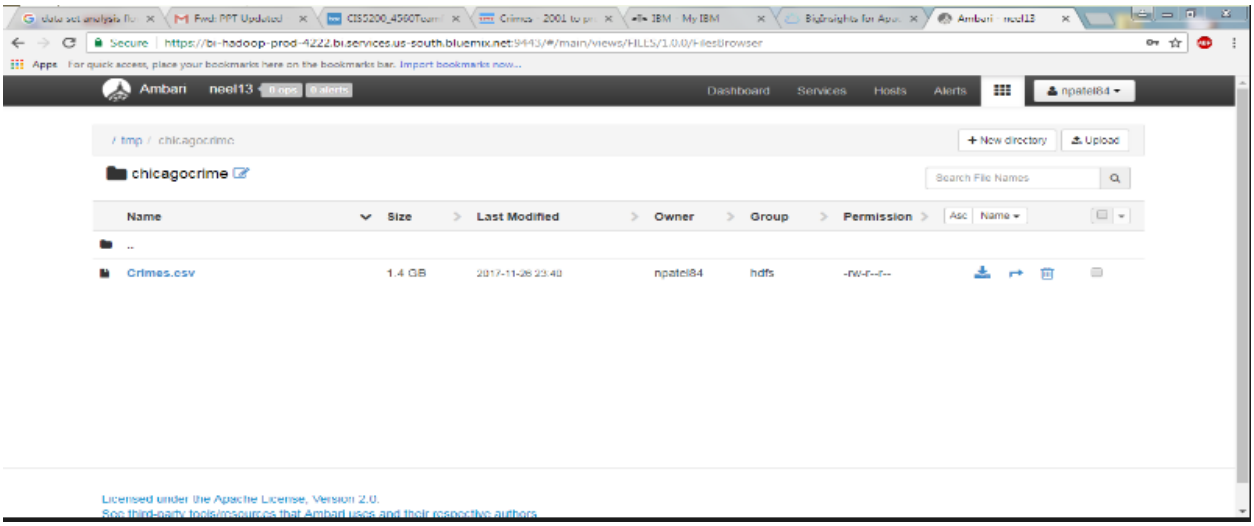
Step 1: Upload the data

```
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 :

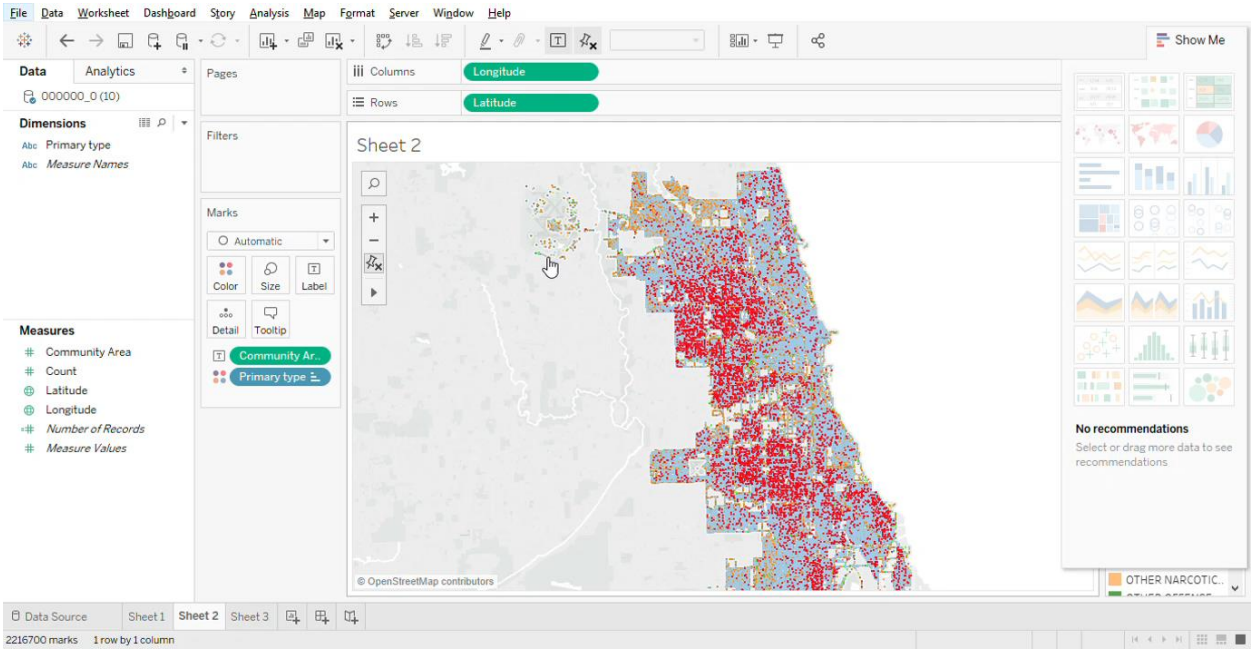
Finding the highest crime type in Chicago from 2001 – present date :

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:



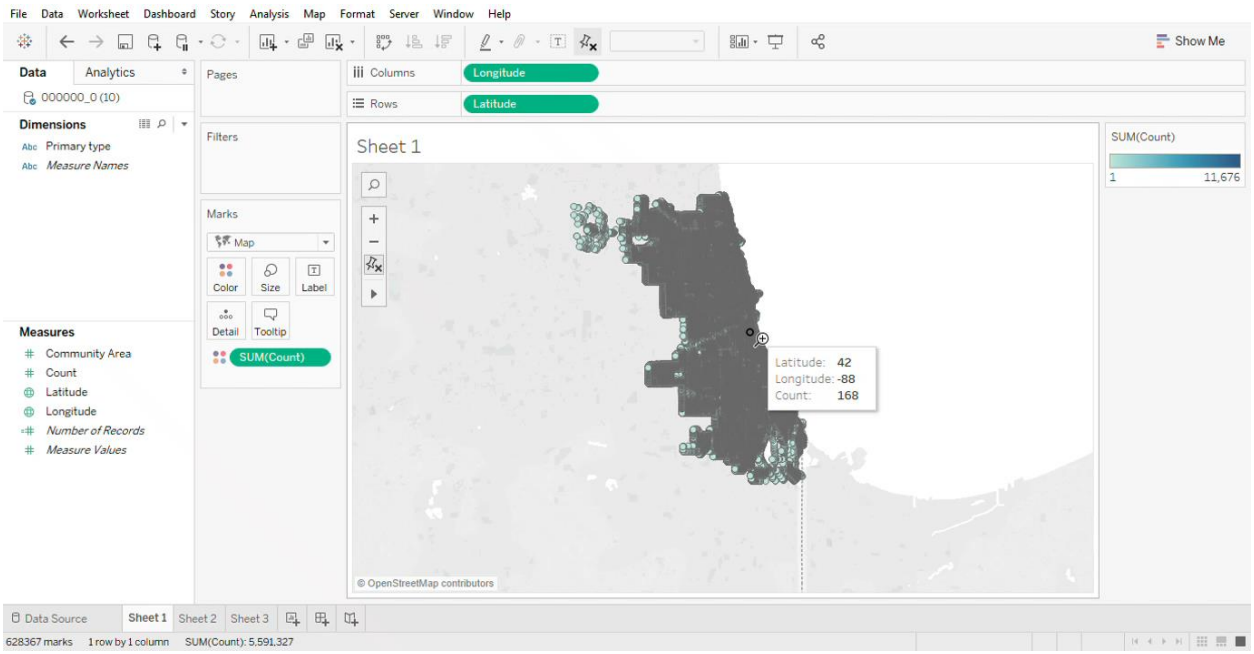
2.Finding the safest places in Chicago which have the least crimes from year 2001 till present days.

```
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.

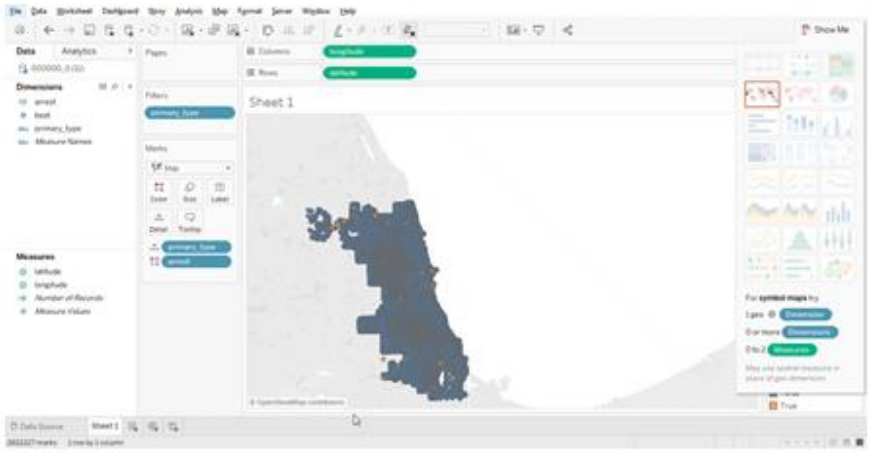
```
drop table if exists crime_arrest;
CREATE TABLE IF NOT EXISTS crime_arrest
AS SELECT
primary_type ,arrest ,beat,latitude, longitude
from crimes_staging
group by primary_type,arrest ,beat,latitude, longitude;
```

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:

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



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:

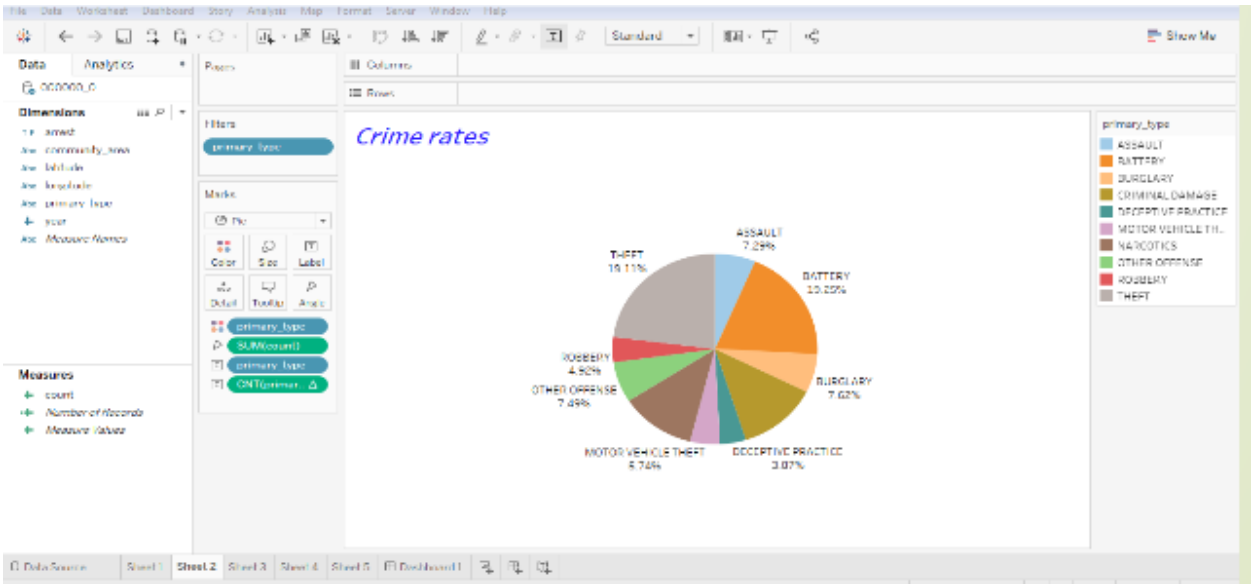
2.The pie chart Illustrate percentage of crime committed from 2001 to 2017 in Chicago :

```
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

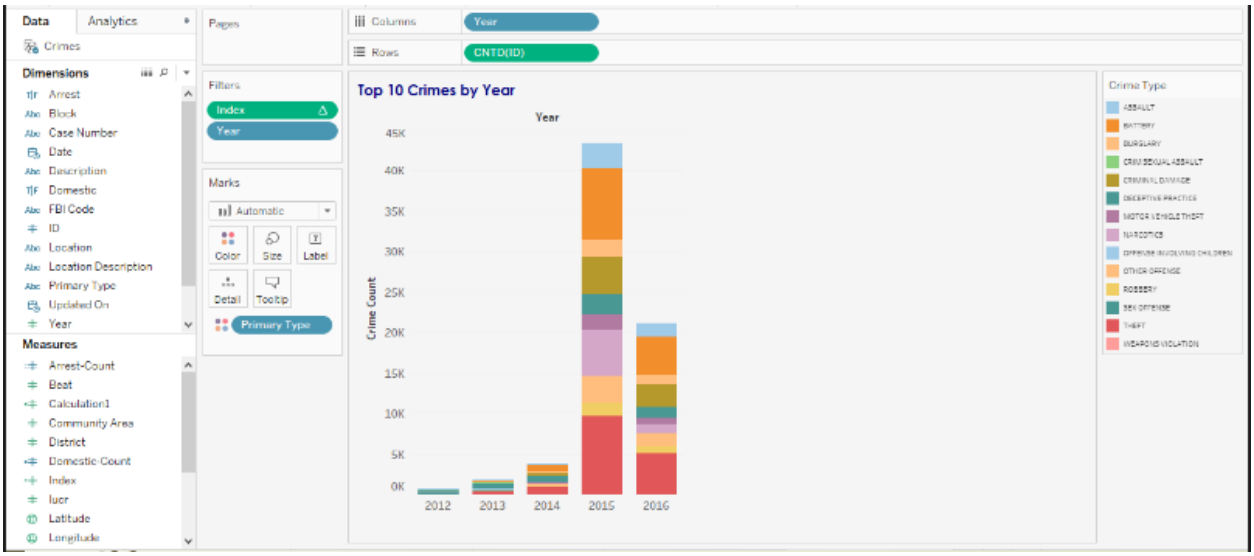
```
INSERT OVERWRITE DIRECTORY '/tmp/'
 ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
SELECT * FROM crime_type_year
WHERE primary_type is NOT NULL and year is NOT NULL;
```

3.2Now Import the 0000-00 file to tablue:

Top Crimes Each Year from 2012-2016 :

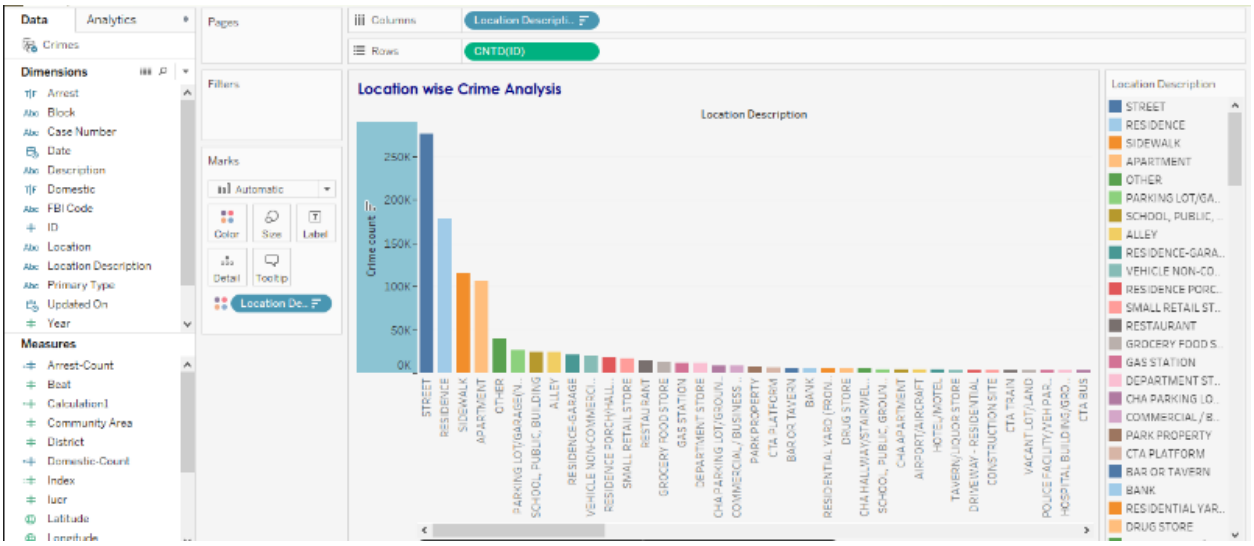
The following code was used to create the last two tables in Tableau

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

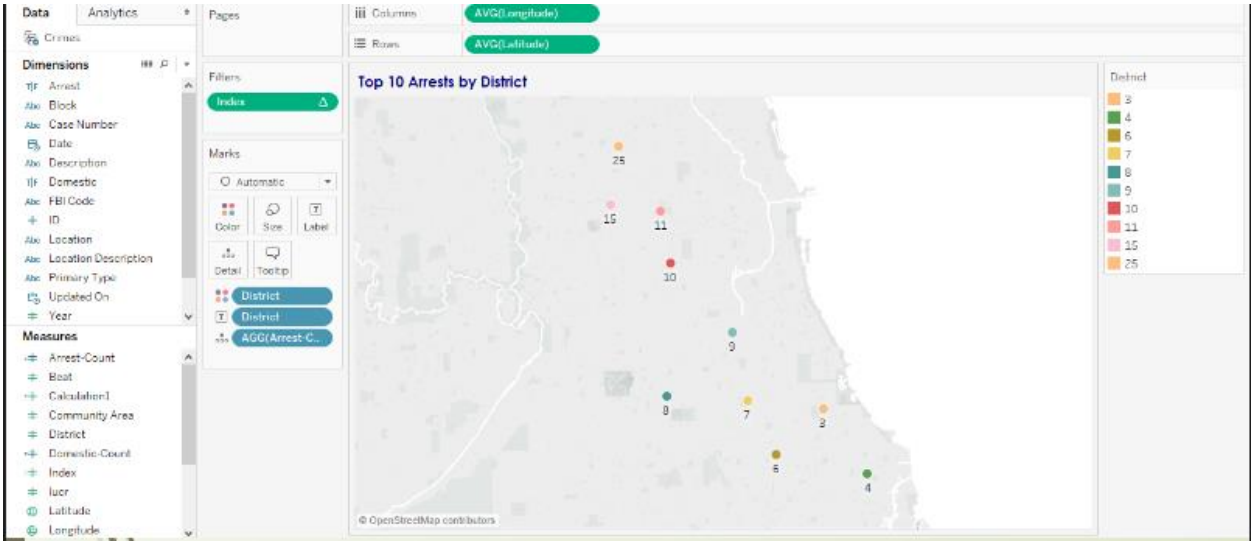


Domestic and Non-Domestic Crimes from 2012-2016 :

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