

RASHTREEYA SIKSHANA SAMITHI TRUST
RV COLLEGE OF ENGINEERING®
(Autonomous Institution Affiliated To VTU, BELAGAVI)
RV Vidyaniketan Post, Mysuru Road,
BENGALURU - 560 059

MASTER OF COMPUTER APPLICATIONS



Assignment - Fifth Semester

BIG DATA ANALYTICS

18MCA52

Faculty In-charge

Dr.Vijayalakshmi.M.N

&

Prof. Shaila.H.Koppad

USN	1RZ18MCA08
NAME	Chetan Pavate
Academic Year	2020 – 2021

RASHTREEYA SIKSHANA SAMITHI TRUST (RSST)

The last decade of pre-independence India was marked with several initiatives and entrepreneurial ventures. One such unique venture was the founding of Rashtreeya Sikshana Samithi Trust (RSST), by **Sri. M.C. Sivananda Sarma**, a freedom fighter and a scholar in the year 1940. The organization started with a noble mission to impart quality education to all sections of the society, without any favour or bias towards anyone. Today, the charitable trust provides avenue for quality education, catering to a wide sector of educational needs, starting from kindergarten to post graduate education as well as research in advanced engineering, medical and architecture domains. The sustained growth and success is due to the dedicated efforts of the management and their continued commitment to the founder's vision, mission, quality, continuous improvement and concern towards social responsibility. Today the trust is managed by a very distinguished board of trustees led by **Dr. M.K. Panduranga Setty, President & Chairman Governing body, Sri. C.V. Hayagriv and Shri Panditharadhy, Vice Presidents, Shri K.G. Subbarama Setty, Hon. Treasurer, Shri A.V.S. Murthy, Hon. Secretary and Shri D.P. Nagaraj, Hon. Joint Secretary**. The board of trustees recognizes the importance of holistic education and a need for a learning environment that nurtures healthy competition and innovation. Rashtreeya Sikshana Samithi Trust **manages over twenty educational institutions including schools, colleges offering degree, post graduate programs and doctoral programs in different specialties**. RSST continuously strives to create state of art infrastructure, recruit excellent faculty and facilitate efficient administration in all its institutions to provide congenial ambience for learning. Today, RSST is regarded as one of the finest and the best managements for education in the country

RV COLLEGE OF ENGINEERING®

Marching towards Excellence in Education, Research and Innovation

Rashtreeya Vidyalaya College of Engineering (RVCE) established in 1963 is one of the earliest self-financing engineering colleges in the country. RV College of Engineering® is the flagship institution of RSST. The institution provides opportunities to all sections, including the under privileged, differently abled and socially marginalized people to gain engineering skills through various programs like WEST-Women Empowerment and Skill Training etc.

RVCE is rated amongst the top five self-financing Engineering colleges in the country. Some magazines have rated it as the best among private institutions in the country, including in terms of **Best Return on Investment for a student**. RVCE is a preferred destination for top ranking aspirants, both for UG, PG and Doctoral programs. RVCE is an **Autonomous college**, affiliated to **Visvesvaraya Technological University (VTU)**, Belagavi. The institution has its own Academic Council which is empowered to approve the academic curriculum as suggested by Board of Studies of various programs and as per the **VTU guidelines for autonomous institutions**. RVCE currently offers **12 Bachelor, 17 Master programmes** and 16 centres of research to carry out research and consultancy activities in the departments. All UG programs have been accredited multiple times. Some of the P.G. programs have been accredited and other eligible PG programs have applied for accreditation. The Institution currently has student **strength of about 5600, faculty strength of around 400 and around 300 Support Staff**. The institution has set itself a Vision "Leadership in

Technical Education, Interdisciplinary Research & Innovation, with a focus on Sustainable and Inclusive Technologies". All the departments are aligned to the vision of sustainable and inclusive technology development with a focus to contribute to both technological leadership of the nation and welfare of all sections of the society. RVCE is rapidly expanding its R&D activity and Industry academic collaborations.

VISION

***Leadership in Technical Education, Interdisciplinary Research & Innovation, with
a Focus on Sustainable and Inclusive Technologies***

MISSION

- *To deliver Outcome Based Quality Education, emphasizing on experiential learning with state of the art infrastructure*
- *To create a conducive environment for interdisciplinary research and innovation*
- *To develop professionals through holistic education focusing on individual growth, discipline, integrity, ethics and social sensitivity*
- *To nurture industry-institution collaboration leading to competency enhancement and entrepreneurship.*
- *To focus on technologies that are sustainable and inclusive, benefiting all sections of the society*

PROFILE OF THE DEPARTMENT

The Department of Master of Computer Applications was established in year 1997 and is the first PG program started in RVCE. The programs offered by the department include, Masters of Computer Applications, M.Sc. by Research and Ph.D. Degree. These programs are affiliated to Visvesvaraya Technological University, Belagavi. The program obtained academic autonomy in the year 2016. The sanctioned intake of students for first year of MCA is 120 students. The MCA program is accredited for third time by National Board of Accreditation, New Delhi. Our graduates have the distinction of obtaining high positions in reputed IT industry. The faculties are from diverse background, committed, and highly qualified with Doctorates in various specializations. They deliver quality education to students through their rich research experience. The faculties are engaged in active research works and projects funded by AICTE, NRB, DRDO agencies and industries to the tune of Rs.83.21 Lakhs have been completed in the last three years and ongoing research projects worth Rs. 25 Lakhs. Faculties have also taken up consultancy works and completed 28.48 Lakhs worth projects and have ongoing of 11.36 Lakh worth. The department has state-of-the-art infrastructure and computing facilities supported by high speed Ethernet and wireless access.

Vision

Pioneering in ICT Enabled Quality Education and Research with a focus on Sustainable and Inclusive Applications

Mission

- M1 To adapt novel methodologies for quality education through experiential learning
- M2 To empower students with continuous, holistic education, emphasizing on discipline, ethics and social commitment
- M3 To become a vibrant knowledge center for research and software development
- M4 To continuously build capacity steering towards industry - institute collaborative research and entrepreneurial competencies
- M5 To utilize and develop free and open source software tools for sustainable and inclusive growth

Program Educational Objectives (PEO)

MCA graduates will be able to

- PEO1** Practice software engineering principles and standards to develop software to meet

customer requirements across verticals

- PEO2** Contribute to build sustainable and inclusive applications using mathematical, simulation and meta-heuristic models
- PEO3** Demonstrate entrepreneurial qualities through individual competence and team work
- PEO4** Achieve successful professional career with integrity and societal commitments leading to lifelong learning

Program Outcomes (PO)

MCA graduates will be able to

PO1 Computational Knowledge:

Acquire in-depth computational knowledge and mathematics with an ability to abstract and conceptualize models from defined problems and requirements

PO2 Problem Analysis:

Identify, formulate, conduct literature survey and solve complex computing problems through analysis as well as provide optimal solutions

PO3 Design / Development of Solutions:

Design and evaluate solutions for complex problems, components or processes that meet specified needs after considering public health and safety, cultural, societal, and environmental factors

PO4 Conduct investigations of complex Computing problems:

Conduct literature survey to analyze and extract information relevant to unfamiliar problems and synthesize information to provide valid conclusions and interpret data by applying appropriate research methods, tools and design experiments

PO5 Use of Modern Tool:

Create, select, adapt and apply appropriate techniques, resources, and modern IT tools to complex computing system activities, with an understanding of the limitations

PO6 Professional Ethics:

Understand and commit to professional ethics and cyber regulations, responsibilities, and norms of professional computing practices

PO7 Life-long Learning:

Engage in lifelong learning independently for continual development to improve knowledge and competence as a computing professional

PO8 Project management and finance:

Demonstrate knowledge and understanding of management principles and apply these to multidisciplinary software development as a team member and manage projects efficiently as a leader considering economical and financial factors

PO9 Communication Efficacy:

Understand and communicate effectively with the computing community and with society at large, regarding complex computing systems activities confidently and effectively by writing effective

reports and design documentations by adhering to appropriate standards, make effective presentations and give / receive clear instructions

PO10 Societal and Environmental Concern:

Understand responsibilities and consequences based on societal, environmental, health, safety, legal and cultural issues within local and global contexts relevant to professional computing practices

PO11 Individual and Team Work:

Function effectively as an individual, as a member or leader in diverse teams in multidisciplinary environments

PO12 Innovation and Entrepreneurship:

Identify a timely opportunity for entrepreneurship and use innovation to pursue and create value addition for the betterment of the individual and society at large

Program Specific Criteria (PSC)

The MCA program will enable the students, by the time they graduate to:

- PSC1:** Explain the principles of mathematics, computing and business foundations
- PSC2:** Demonstrate the use of software tools and technologies relevant to various verticals
- PSC3:** Design and develop software products, processes and systems for real world situations

Program Specific Outcomes (PSO)

MCA graduates will be able to

- PSO1:** Solve real world computing system problems of various industries by understanding and applying the principles of mathematics, computing techniques and business concepts
- PSO2:** Design, test, develop and maintain desktop, web, mobile and cross platform software applications using modern tools and technologies

RV COLLEGE OF ENGINEERING[®], BENGALURU – 560059
(Autonomous Institution Affiliated to VTU, Belagavi)
Department of Master of Computer Applications

C E R T I F I C A T E

This is to certify that Mr./Ms. Chetan Pavate USN 1RZ18MCA08 of **5th Semester Master of Computer Applications** has satisfactorily completed the **Assignment of Big Data Analytics** prescribed for the academic year **2020 – 2021**.

LAB MARKS	
Max	Obtained
30	

**Signature of the
Student**

**Signature of the
Faculty In - Charge**

**Signature of Director
MCA**

Big Data Analytics

18MCA52

WPS Office | BPD_CrimeData_Original.xlsx | BPD_CrimeData_Inside Or Outside | BPD_CrimeData_District.xlsx | +

Home Insert Page Layout Formulas Data Review View Tools

C14 fx 04:26:00

ID	CrimeDate	CrimeTime	Description
1	18-06-2016	00:33:00	2700 CHESLEY AV I
2	18-06-2016	00:39:00	2700 FAIT AVN O
3	18-06-2016	01:00:05	2400 CYLBURN AVO
4	18-06-2016	01:53:00	2300 ORLEANS ST O
5	18-06-2016	02:05:00	800 N WOLFE ST I
6	18-06-2016	02:35:00	ST & IMLA ST O
7	18-06-2016	02:35:00	ST & IMLA ST O
8	18-06-2016	02:45:00	2400 W BELVEDER I
9	18-06-2016	03:45:00	1100 RUSSELL ST I
10	18-06-2016	04:00:00	ST & N PATTERSON I
11	18-06-2016	04:00:00	200 N LIBERTY ST O
12	18-06-2016	04:00:00	609 S SANN ST I
13	18-06-2016	04:26:00	2600 PENNSYLVIA I
14	18-06-2016	02:25:01	2300 DENISON ST O
15	18-06-2016	02:25:30	4200 NICHOLAS A O
16	18-06-2016	23:00:00	2500 W BALTIMORE O
17	18-06-2016	23:15:00	700 RAMSAY ST O
18	18-06-2016	00:00:00	300 S CASTLE ST I
19	18-06-2016	00:00:00	4000 CEDARDALE O
20	18-06-2016	00:30:00	1300 N CAROLINE O
21	18-06-2016	00:30:00	2400 F FEDERAL S I
22	18-06-2016	04:45:00	3500 E BALTIMOR O
23	18-06-2016	05:03:00	3000 RAYNER AVE I
24	18-06-2016	05:15:00	1000 FELL ST O
25	18-06-2016	05:30:00	1500 RUSSELL ST I
26	18-06-2016	06:40:00	RD & GWYNNS #1
27	18-06-2016	07:05:00	1800 ORLEANS ST I
28	18-06-2016	08:00:00	3200 LABYRINTH II
29	18-06-2016	08:00:00	200 GUILFORD AVO
30	18-06-2016	08:14:00	3700 W BELVEDER I
31	18-06-2016	09:00:00	6800 FABDEL AVI
32	18-06-2016	09:30:00	400 PURROW ST I
33	18-06-2016	09:40:00	5700 CHINQUAPIN I
34	18-06-2016	09:40:00	2400 ETTING ST I
35	18-06-2016	10:00:00	3900 GROVELAND I
36	18-06-2016	10:10:00	400 W FRANKLIN I
37	18-06-2016	10:10:00	400 W FRANKLIN I

HP 1148 PM

U: 0.25 KB/s D: 0.44 KB/s ENG

BPD_CrimeData Inside Or Outside

WPS Office | BPD_CrimeData_Original.xlsx | BPD_CrimeData_Inside Or Outside | BPD_CrimeData_District.xlsx | +

Home Insert Page Layout Formulas Data Review View Tools

G7 fx

ID	CrimeDate	CrimeTime	Description	Weapon	District
1	18-06-2016	00:33:00	I	HANDS	NORTHEASTERN
2	18-06-2016	00:39:00	O	KNIFE	SOUTHEASTERN
3	18-06-2016	01:00:05	O	FIREARM	NORTHERN
4	18-06-2016	01:53:00	O	FIREARM	SOUTHEASTERN
5	18-06-2016	02:05:00	I	OTHER	EASTERN
6	18-06-2016	02:35:00	O	OTHER	SOUTHEASTERN
7	18-06-2016	02:35:00	O	OTHER	SOUTHEASTERN
8	18-06-2016	02:45:00	I	HANDS	NORTHERN
9	18-06-2016	03:45:00	I	OTHER	SOUTHERN
10	18-06-2016	04:00:00	I	HANDS	EASTERN
11	18-06-2016	04:00:00	O	FIREARM	CENTRAL
12	18-06-2016	04:00:00	I	OTHER	SOUTHEASTERN
13	18-06-2016	04:26:00	I	HANDS	WESTERN
14	18-06-2016	02:25:01	O	FIREARM	SOUTHWESTERN
15	18-06-2016	02:25:30	O	FIREARM	NORTHEASTERN
16	18-06-2016	23:00:00	O	HANDS	SOUTHWESTERN
17	18-06-2016	23:15:00	O	OTHER	SOUTHERN
18	18-06-2016	00:00:00	I	OTHER	SOUTHEASTERN
19	18-06-2016	00:00:00	O	OTHER	NORTHWESTERN
20	18-06-2016	00:30:00	O	HANDS	EASTERN
21	18-06-2016	00:30:00	I	HANDS	EASTERN
22	18-06-2016	04:45:00	O	OTHER	SOUTHEASTERN
23	18-06-2016	05:03:00	I	OTHER	SOUTHWESTERN
24	18-06-2016	05:15:00	O	OTHER	SOUTHEASTERN
25	18-06-2016	05:30:00	I	OTHER	SOUTHERN
26	18-06-2016	06:40:00	I	OTHER	WESTERN
27	18-06-2016	07:05:00	I	OTHER	EASTERN
28	18-06-2016	08:00:00	I	OTHER	NORTHWESTERN
29	18-06-2016	08:00:00	O	OTHER	CENTRAL
30	18-06-2016	08:14:00	I	OTHER	NORTHWESTERN
31	18-06-2016	09:00:00	I	HANDS	NORTHEASTERN
32	18-06-2016	09:30:00	I	OTHER	SOUTHWESTERN
33	18-06-2016	09:40:00	I	HANDS	NORTHERN
34	18-06-2016	09:40:00	I	HANDS	WESTERN
35	18-06-2016	10:00:00	I	HANDS	NORTHWESTERN
36	18-06-2016	10:10:00	I	OTHER	CENTRAL
37	18-06-2016	10:10:00	I	OTHER	CENTRAL

HP 1149 PM

U: 0.25 KB/s D: 0.61 KB/s ENG 11:49 PM

BPD_CrimeData District

Big Data Analytics

18MCA52

ID	CrimeCode	Description	Weapon	Total Incidents
1	4E	I	HANDS	2
2	4B	O	KNIFE	3
3	9S	O	FIREARM	1
4	3AF	O	FIREARM	1
5	6C	I	OTHER	1
6	4C	O	OTHER	1
7	74C	O	OTHER	1
8	84E	I	HANDS	1
9	95D	I	OTHER	1
10	104E	I	HANDS	1
11	113AF	O	FIREARM	1
12	126G	I	OTHER	1
13	134E	I	HANDS	1
14	149S	O	FIREARM	1
15	159S	O	FIREARM	1
16	164E	O	HANDS	1
17	174C	O	OTHER	1
18	185A	I	OTHER	1
19	196H	O	OTHER	1
20	204E	O	HANDS	1
21	214E	I	HANDS	1
22	224F	O	OTHER	1
23	235A	I	OTHER	1
24	246D	O	OTHER	1
25	254C	I	OTHER	1
26	266G	I	OTHER	1
27	274F	I	OTHER	1
28	285A	I	OTHER	1
29	296D	O	OTHER	1
30	306G	I	OTHER	1
31	314E	I	HANDS	1
32	325A	I	OTHER	1
33	334E	I	HANDS	1
34	344E	I	HANDS	1
35	354E	I	HANDS	1
36	363GO	I	OTHER	1

BPD_CrimeData Crime Code

ID	CrimeDate	CrimeTime	Total Incidents
1	18-06-2016	00:33:00	2
2	18-06-2016	00:39:00	3
3	18-06-2016	01:00:00	1
4	18-06-2016	01:33:00	1
5	18-06-2016	02:00:00	1
6	18-06-2016	02:35:00	1
7	18-06-2016	02:35:00	1
8	18-06-2016	02:45:00	1
9	18-06-2016	03:45:00	1
10	18-06-2016	04:00:00	1
11	18-06-2016	04:00:00	1
12	18-06-2016	04:00:00	1
13	18-06-2016	04:26:00	1
14	18-06-2016	02:25:01	1
15	18-06-2016	02:25:30	1
16	18-06-2016	23:00:00	1
17	18-06-2016	23:15:00	1
18	18-06-2016	00:00:00	1
19	18-06-2016	00:00:00	1
20	18-06-2016	00:30:00	1
21	18-06-2016	00:30:00	1
22	18-06-2016	04:45:00	1
23	18-06-2016	05:03:00	1
24	18-06-2016	05:15:00	1
25	18-06-2016	05:30:00	1
26	18-06-2016	06:40:00	1
27	18-06-2016	07:05:00	1
28	18-06-2016	08:00:00	1
29	18-06-2016	08:00:00	1
30	18-06-2016	08:14:00	1
31	18-06-2016	09:00:00	1
32	18-06-2016	09:30:00	1
33	18-06-2016	09:40:00	1
34	18-06-2016	09:40:00	1
35	18-06-2016	10:00:00	1
36	18-06-2016	10:10:00	1

BPD_CrimeData Co-ordinates

1) 1.Mapper and Reducer function for the chosen Area/Domain

The below code has the mapper and reducer functions where, mapper is used to group the location post number and the reducer will find total number of crimes for each key value of post.

```

package chetan.crime;
import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.DoubleWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class Crime{
public static class CrimeMapper extends Mapper<Object, Text, Text, IntWritable> {
@Override
public void map(Object key, Text value, Context context) throws IOException,
InterruptedException {
String line[] = value.toString().split(",");
if(line.length==12)
{
String lpost = line[8];
int val = Integer.parseInt(line[8].trim());
context.write(new Text(lpost), new IntWritable(val));
}
}
}

public static class CrimeReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
int sum = 0;
for (IntWritable value : values) {
sum += value.get();
}
context.write(key, new IntWritable(sum));
}
}

public static void main(String[] args) throws Exception {
if (args.length != 2) {
System.err.println("Usage: Crime <input path> <output path>");
System.exit(-1);
}
Job job = new Job();
job.setJarByClass(Crime.class);
job.setJobName("Crimes");
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
job.setMapperClass(CrimeMapper.class);
job.setReducerClass(CrimeReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

OUTPUT Map Reducer 1

```
cpavate148@cluster-ac0b-m:~/MR/mra$ ls
_SUCCESS part-r-00000 part-r-00001 part-r-00002 part-r-00003 part-r-00004
cpavate148@cluster-ac0b-m:~/MR/mra$ cat part-r-00000
EASTERN 44
NORTHERN 85
cpavate148@cluster-ac0b-m:~/MR/mra$ cat part-r-00001
SOUTHERN 60
WESTERN 36
cpavate148@cluster-ac0b-m:~/MR/mra$ cat part-r-00002
cpavate148@cluster-ac0b-m:~/MR/mra$ cat part-r-00003
NORTHEASTERN 74
SOUTHEASTERN 50
cpavate148@cluster-ac0b-m:~/MR/mra$ cat part-r-00004
CENTRAL 67
NORTHWESTERN 69
SOUTHWESTERN 40
cpavate148@cluster-ac0b-m:~/MR/mra$ []
```

2. Mapper and Reducer function for finding the total incidents using weapons.

The below code has the mapper and reducer functions where, mapper is used to group the Weapons and the reducer will find total number of crimes for each Weapon.

```
package chetan.crime;

import java.io.IOException;
import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class CovidMR2 {
    public static class CMapper extends Mapper<Object, Text, Text, IntWritable> {

        @Override
        public void map(Object key, Text value, Context context) throws IOException,
        InterruptedException {
            String line[] = value.toString().split(",");
            if(line.length==12)
            {
                String district = line[8];
                int incidents = line[11];
                context.write(new Text(district), new IntWritable(incidents));
            }
        }
    }

    public static class CReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
        public void reduce(Text key, Iterable<IntWritable> values, Context context) throws
        IOException, InterruptedException {
            int sum = 0;
            for (IntWritable value : values) {
                sum += value.get();
            }
        }
    }
}
```

```

        }
        context.write(key, new IntWritable(sum));
    }
}

public static void main(String[] args) throws Exception {
    if (args.length != 2) {
        System.err.println("Usage: Covid <input path> <output path>");
        System.exit(-1);
    }
    Job job = new Job();
    job.setJarByClass(CovidMR2.class);
    job.setJobName("Covid19");
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    job.setMapperClass(CMapper.class);
    job.setReducerClass(CReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

OUTPUT Map Reducer 2

```

cpavate148@cluster-ac0b-m:~/MR/mr02$ ls
_SUCCESS part-r-00000 part-r-00001 part-r-00002 part-r-00003 part-r-00004
cpavate148@cluster-ac0b-m:~/MR/mr02$ cat part-r-00000
cpavate148@cluster-ac0b-m:~/MR/mr02$ cat part-r-00001
KNIFE    23
cpavate148@cluster-ac0b-m:~/MR/mr02$ cat part-r-00002
HANDS    315
OTHER    128
cpavate148@cluster-ac0b-m:~/MR/mr02$ cat part-r-00003
FIREARM  59
cpavate148@cluster-ac0b-m:~/MR/mr02$ cat part-r-00004
cpavate148@cluster-ac0b-m:~/MR/mr02$ 

```

2) Pig Querries for the give Domain and Dataset

1.Total known no of crimes that have happened in the City.

```

q1 = LOAD '/que/ccordinates.csv' using PigStorage(',') AS
(ID:int,CrimeDate:chararray,CrimeTime:chararray,Incidents:int,Location1:chararray);
a = group q1 all;
b = foreach a generate SUM(q1.Incidents);
dump b;

```

```

cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q1/part-r-00000
110999
cpavate148@cluster-ac0b-m:~/Queries$ 

```

2. Parameter substitutions to find the total no of crimes by a particular weapon (Firearm).

```
=> load '/query/ccode.csv' USING PigStorage(',') AS (ID:int, CrimeCode:chararray, Description:chararray, Weapon:chararray, Incidents:int);
A = FILTER @ BY Weapon=='FIREARM';
group_A = GROUP A ALL;
A_count = FOREACH group_A GENERATE SUM(A.Incidents);
dump A_count;
```

`pig -param weapon=FIREARM q2a.pig`

```
cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q2/part-r-00000
19046
cpavate148@cluster-ac0b-m:~/Queries$
```

3. Total crimes committed inside house(join).

```
cdic = LOAD '/query/cdistrict.csv' USING PigStorage(',') AS (ID:int, CrimeDate:chararray, CrimeTime:chararray, Description:chararray, Weapon:chararray, District:chararray);
ccd = LOAD '/query/ccode.csv' USING PigStorage(',') AS (ID:int, CrimeCode:chararray, Description:chararray, Weapon:chararray, Incidents:int);
filter_cdic = FILTER cdic BY Description == '$desc';
filter_ccd= FILTER ccd BY Description == '$desc';
join_query= JOIN filter_ccd by ID, filter_cdic by ID;
join_res = group join_query ALL;
res = FOREACH join_res GENERATE SUM(join_query.filter_ccd::Incidents);
dump res;
```

```
cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q3/part-r-00000
52203
cpavate148@cluster-ac0b-m:~/Queries$
```

4. Total crimes in a particualar location using Join.

```
cdi = LOAD '/query/cio.csv' USING PigStorage(',') AS (ID:int, CrimeDate:chararray, CrimeTime:chararray, Location:chararray, Description:chararray);
cc = LOAD '/query/clweapon.csv' USING PigStorage(',') AS (ID:int, Location:chararray, Description:chararray, Weapon:chararray, Post:int);
filter_cdic = FILTER cdi BY Location == '$desc';
filter_ccd= FILTER cc BY Location == '$desc';
join_query= JOIN filter_ccd by ID, filter_cdic by ID;
join_res = group join_query ALL;
res = FOREACH join_res GENERATE COUNT(join_query.filter_ccd::Post);
dump res;
```

```
cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q4/part-r-00000
203
cpavate148@cluster-ac0b-m:~/Queries$
```

5. Checking no of incidents occurred and it belongs to which district using Join.

```

cdic = LOAD '/query/cdistrict.csv' USING PigStorage(',') AS (ID:int, CrimeDate:chararray,
CrimeTime:chararray, Description:chararray, Weapon:chararray, District:chararray);
ccd = LOAD '/query/ccordinates.csv' using PigStorage(',') AS
(ID:int,CrimeDate:chararray,CrimeTime:chararray,Incidents:int);
q1 = filter cdic by ID is not null;
q2 = filter ccd by ID is not null;
grp1 = group cdic by ID;
grp2 = group ccd by ID;
qu1 = foreach grp1 generate group as ID, cdic.District;
qu2 = foreach grp2 generate group as ID, ccd.Incidents;
j = join qu1 by ID, qu2 by ID;
query1 = FOREACH j GENERATE $0,$1,$3;
dump query1;

```

```

192009, { (SOUTHEASTERN) }, { (1) }
192014, { (SOUTHWESTERN) }, { (1) }
192022, { (SOUTHWESTERN) }, { (1) }
192026, { (NORTHERN) }, { (1) }
192034, { (SOUTHERN) }, { (1) }
192040, { (NORTHWESTERN) }, { (1) }
192041, { (NORTHWESTERN) }, { (1) }
192043, { (CENTRAL) }, { (1) }
192046, { (NORTHEASTERN) }, { (1) }
192047, { (EASTERN) }, { (1) }
192048, { (SOUTHERN) }, { (1) }
192049, { (NORTHWESTERN) }, { (1) }
192052, { (SOUTHEASTERN) }, { (1) }
192056, { (NORTHEASTERN) }, { (1) }
192057, { (NORTHEASTERN) }, { (1) }
192058, { (EASTERN) }, { (1) }
192059, { (NORTHERN) }, { (1) }
192062, { (SOUTHERN) }, { (1) }
192064, { (EASTERN) }, { (1) }
192070, { (NORTHEASTERN) }, { (1) }
192073, { (WESTERN) }, { (1) }
192076, { (NORTHEASTERN) }, { (cpavate148@cluster-ac0b-m:~/Queries$)
cpavate148@cluster-ac0b-m:~/Queries$ █

```

6. List Different weapons used for committing crimes

```

ccd = LOAD '/query/ccode.csv' USING PigStorage(',') AS (ID:int, CrimeCode:chararray,
Description:chararray, Weapon:chararray, Incidents:int);
q3a = filter ccd by (Weapon IS NOT NULL);
q3b = FOREACH q3a GENERATE(Weapon) as wp;
q3c = DISTINCT q3b;
dump q3c;

```

```

cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q6/part-r-00000
HANDS
KNIFE
OTHER
Weapon
FIREARM
cpavate148@cluster-ac0b-m:~/Queries$ █

```

7. Total no of incidents by every weapon

```

org = LOAD '/query/original.csv' USING PigStorage(',') AS (ID:int, Weapon:chararray, District:chararray,
Neighbourhood:chararray, Incidents:int);
ftr = filter org by (Weapon IS NOT NULL) and (Incidents IS NOT NULL);
grp = group ftr by (Weapon, Incidents);
res = FOREACH grp generate FLATTEN(group) as (Weapon, Incidents), COUNT($1) as c;
dump res;

```

```

cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q7/part-r-00000
HANDS,1,48251
HANDS,2,1
KNIFE,1,9201
KNIFE,3,1
OTHER,1,34498
FIREARM,1,19046
cpavate148@cluster-ac0b-m:~/Queries$ 

```

8. Neighbourhood and weapons using which more than 1 incidents happened on a particular day

```

org = LOAD '/query/original.csv' USING PigStorage(',') AS (ID:int, Weapon:chararray, District:chararray,
Neighbourhood:chararray, Incidents:int);
query4 = filter org by Incidents>1;
query4a = group query4 all;
query4b = foreach query4a generate (query4.District, query4.Weapon);
Dump query4b;

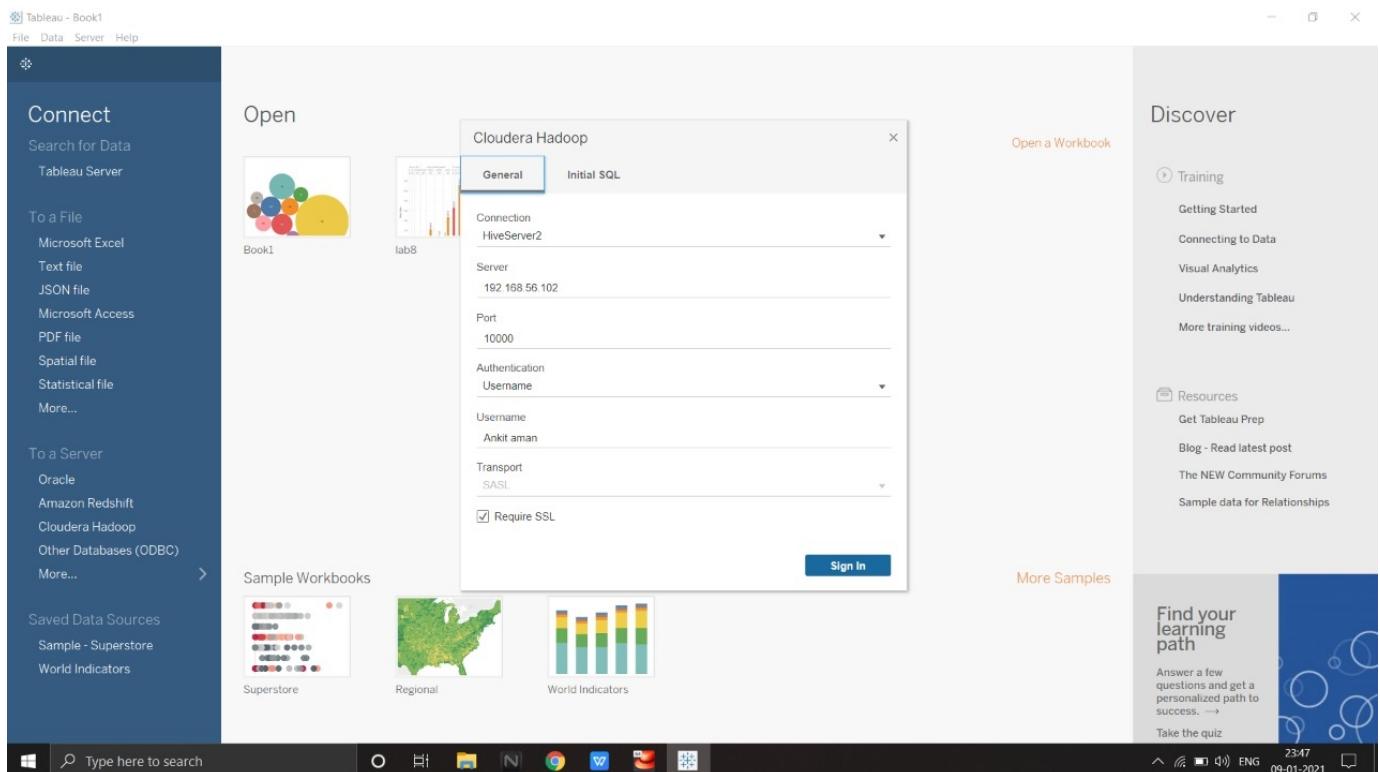
```

```

cpavate148@cluster-ac0b-m:~/Queries$ hdfs dfs -cat hdfs://cluster-ac0b-m/user/cpavate148/q8/part-r-00000
({(SOUTHEASTERN),(NORTHEASTERN)}, {(KNIFE),(HANDS)})
cpavate148@cluster-ac0b-m:~/Queries$ 

```

3) Connectivity of Hortonworks Hadoop Hive with Tableau Desktop to show Visualization of Queries.



Big Data Analytics

18MCA52

Tableau - Book1

File Data Server Window Help

Connections Add
192.168.56.102 Cloudera Hadoop

Schema crime

Table
Enter table name P +
Exact Contains Starts with

New Custom SQL
New Union

org (crime.org) (crime)

Connection Live Extract Filters 0 | Add

Need more data?
Drag tables here to relate them. Learn more

Sort fields Data source order ▾ Show aliases Show hidden fields 1,000 rows

ID	Crimedate	Crimestime	Crimecode	Location	Description	Weapon	Post	District	Neighbourhood	Incidents
1	18-06-2016	0:33:00	4E	2700 CHESLEY AVE	I	HANDS	424	NORTHEASTERN	North Harford R...	2
2	18-06-2016	0:39:00	4B	2700 FAIT AVE	O	KNIFE	232	SOUTHEASTERN	Canton	3
3	18-06-2016	1:00:05	95	2400 CYLBURN AV	O	FIREARM	532	NORTHERN	Levindale	1
4	18-06-2016	1:53:00	3AF	2300 ORLEANS ST	O	FIREARM	221	SOUTHEASTERN	McElderry Park	1
5	18-06-2016	2:05:00	6C	800 N WOLFE ST	I	OTHER	321	EASTERN	Middle East	1
6	18-06-2016	2:35:00	4C	ST & IMLA ST	O	OTHER	241	SOUTHEASTERN	Bayview	1
7	18-06-2016	2:35:00	4C	ST & IMLA ST	O	OTHER	241	SOUTHEASTERN	Bayview	1
8	18-06-2016	2:45:00	4E	2400 W BELVEDER...	I	HANDS	532	NORTHERN	Levindale	1
9	18-06-2016	3:45:00	5D	1100 RUSSELL ST	I	OTHER	941	SOUTHERN	Carroll - Camden ...	1
10	18-06-2016	4:00:00	4E	ST & N PATTER...	I	HANDS	321	EASTERN	McElderry Park	1
11	18-06-2016	4:00:00	3AF	200 N LIBERTY ST	O	FIREARM	111	CENTRAL	Downtown	1

Data Source Sheet 1 Sheet 2 Sheet 3 Sheet 4 Sheet 5 ⏪ ⏩

Type here to search 23:59 09-01-2021

1. Total Incidents that happened per Quarter-Month.

Tableau - assignment - Tableau license expires in 13 days

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Data Analytics Pages

BFD_CrimeData (BPD_Cr...

Search Filters

Tables

- Crime Code
- Crime Date
- Crime Time
- Description
- District
- Location
- Location1
- Neighborhood
- Weapon
- Measure Names
- Post
- Total Incidents
- BFD_CrimeData (Count)
- Measure Values

Columns QUARTER(Crime ... MONTH(Crime Da..)

Rows SUM(Total incidents)

incidents happen in month

Crime Date

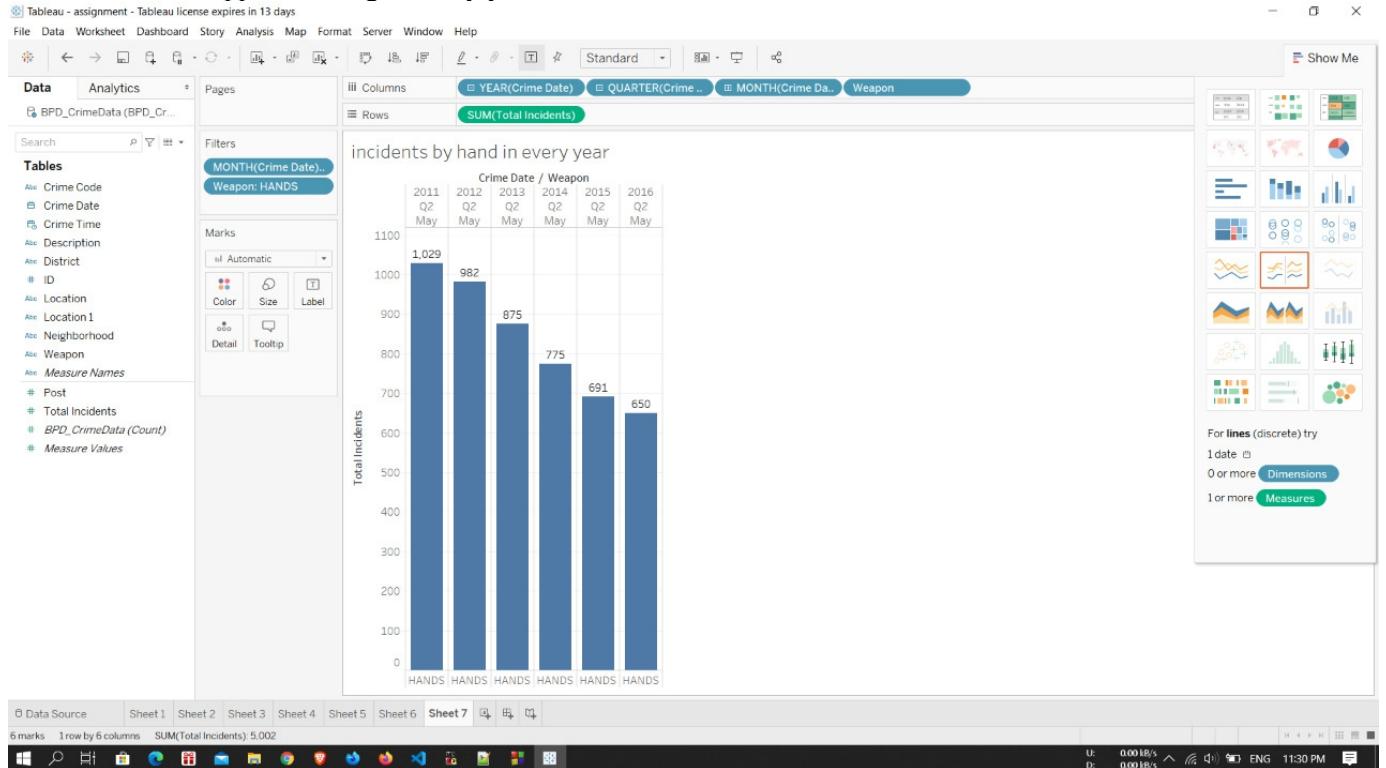
Month	Q1	Q2	Q3	Q4
January	1,229	1,372	1,349	1,180
February	1,126	1,228	1,159	1,112
March	1,584	1,636	1,271	963
April	1,628	1,402	1,400	1,237
May	1,708	1,679	1,531	1,286
June	1,442	1,527	1,460	1,274
July	1,633	1,451	1,409	1,543
August	1,568	1,465	1,500	1,460
September	1,532	1,541	1,399	1,304
October	1,570	1,367	1,336	1,307
November	1,446	1,301	1,217	1,238
December	1,463	1,374	1,312	1,181

For lines (discrete) try
1 date 0 or more Dimensions 1 or more Measures

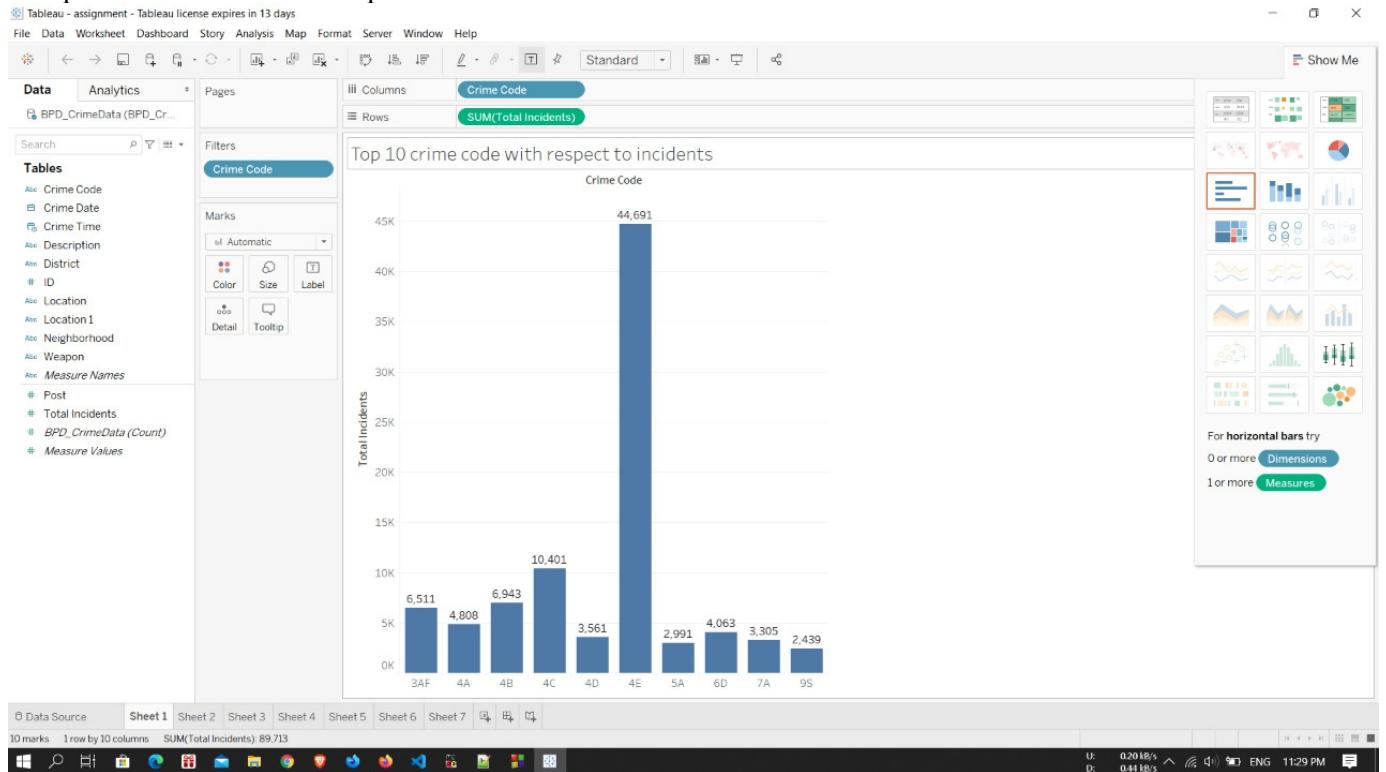
Data Source Sheet 1 Sheet 2 Sheet 3 Sheet 4 Sheet 5 Sheet 6 Sheet 7 ⏪ ⏩

66 marks 1 row by 12 columns SUM(Total Incidents) 10,999 U: 0.19 kB/s D: 0.12 kB/s 11:30 PM ENG

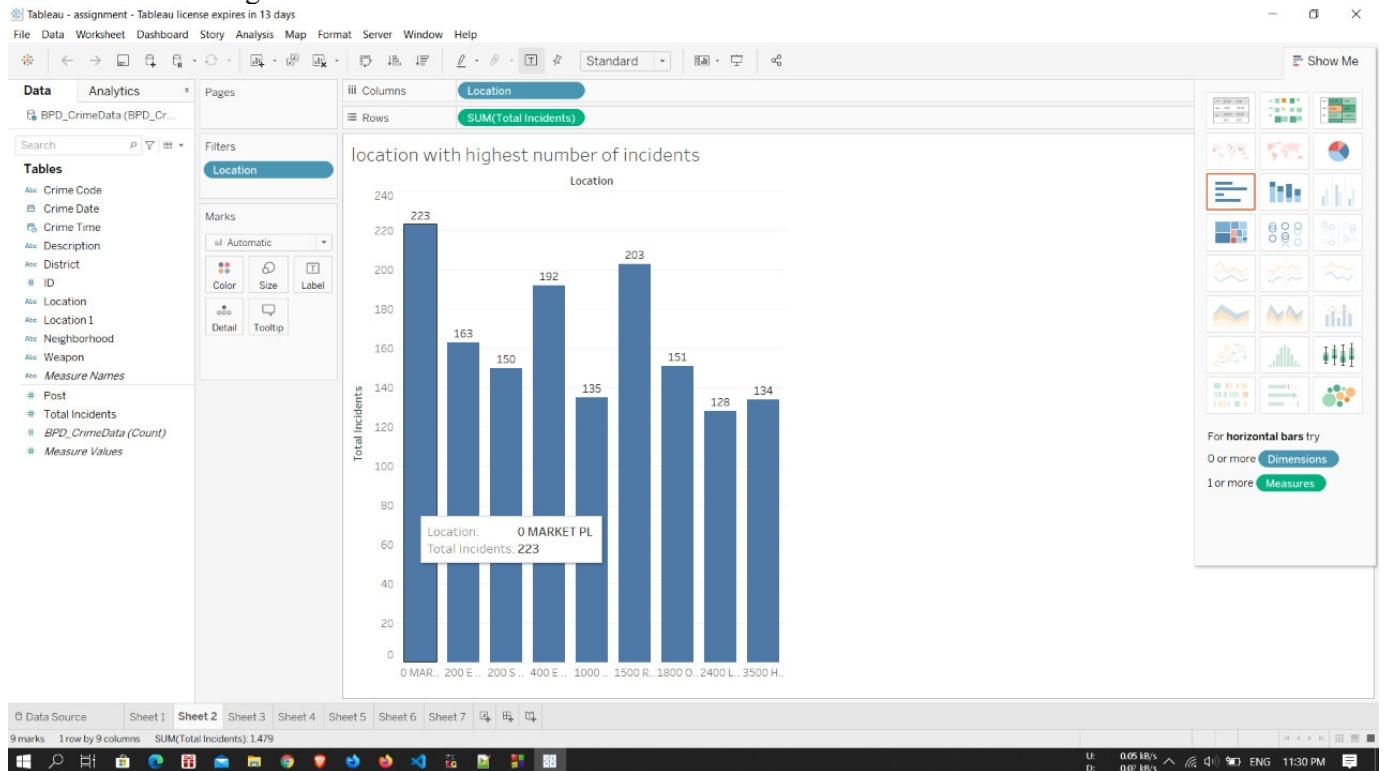
2. Incidents that happened using hand by year.



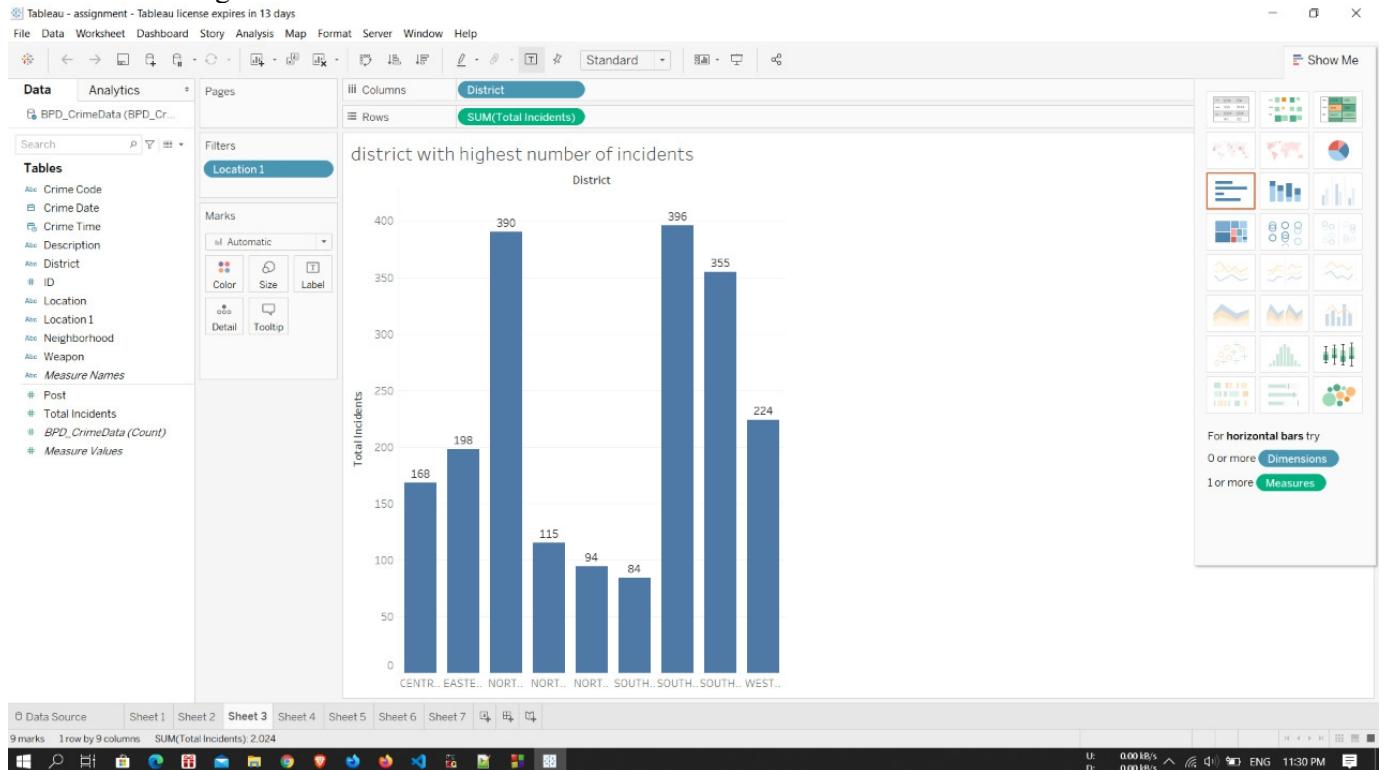
3. Top 10 Crime-Code with respect to incidents.



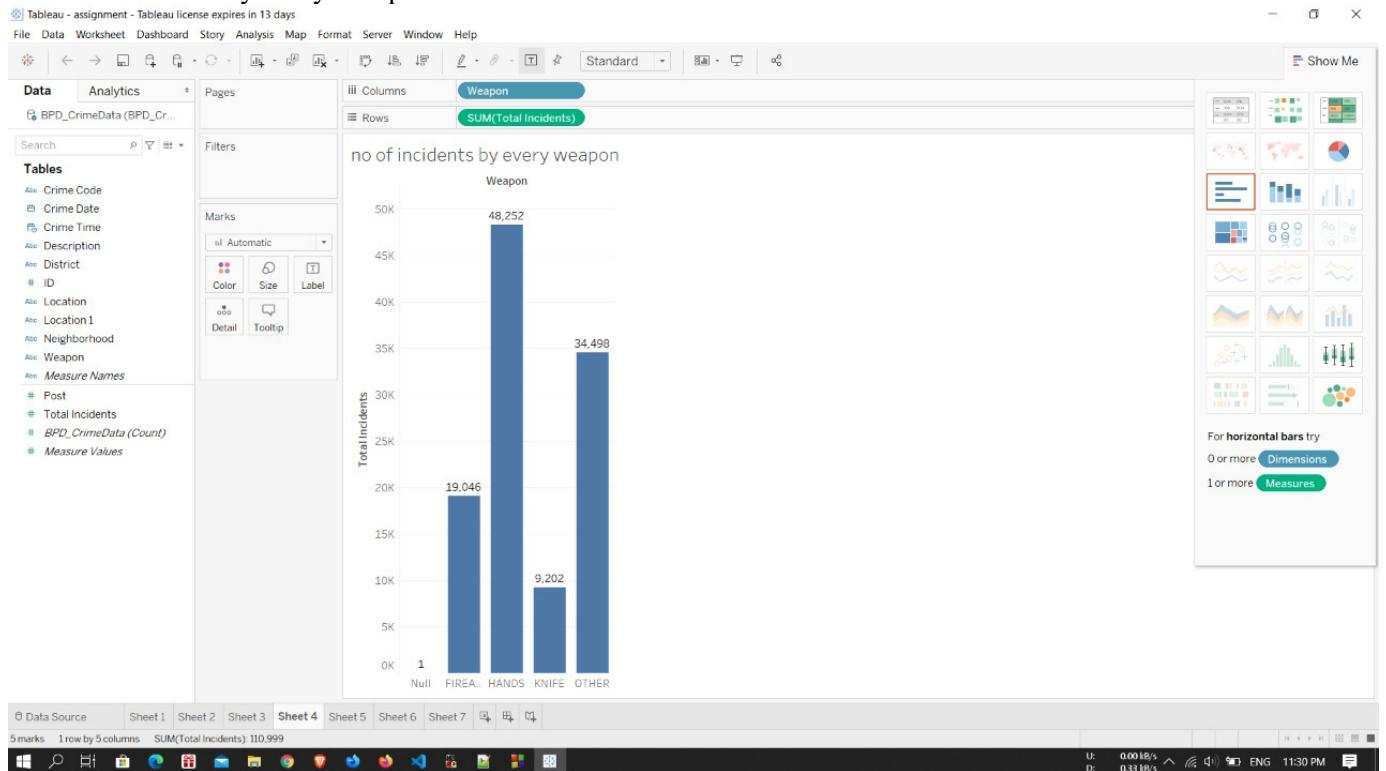
4. Location with Highest no of incidents.



5. District with highest no of incidents.



6. No of incidents by every Weapon.



7. District with Highest no of incidents using Hands.

