# *Proof of Concept*

***On***

**Title: Analysis of Global Terrorism Data**

**Submitted for the requirement of**

**Big Data Engineering course**

**BACHELOR OF ENGINEERING**
**COMPUTER SCIENCE & ENGINEERING**
**(Big Data and Analytics)**
**CSP-329**

**Semester-5**
**POC Group no 12**


**Submitted to:**
Ms. Gurpreet Kaur
Project Supervisor

**Submitted by:**
Kaustav Bora
(17BCS3775),
Aman Shah
(17BCS3784)

## ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. A special gratitude we give to our 5th semester B.D.E project supervisor, Ms. Gurpreet Kaur, whose contribution in stimulating suggestions and encouragement, helped us to coordinate our project and especially in writing this report. Furthermore, we would also like to acknowledge with much appreciation her crucial role, in giving the permission to use all required equipment and the necessary materials to complete the task 'Analysis of Global Terrorism data' using Apache Hadoop (Hive), IBM Cognos, MS Excel and Python.

## OVERVIEW

Terrorism is, in the broadest sense, the use of intentional violence, generally against civilians, for political purposes. It is used in this regard primarily to refer to violence during peacetime or in context of war against non-combatants (mostly civilians and neutral military personnel). With the global terrorism dataset, it mostly signifies about the attack types including various information's like effected areas, casualties, targeted areas etc.

## RATIONALE

Terrorism has always been a greatest threat to a country and its civilians so, every nation is much more interested to get insights of that data. With the proper analysis of such a large data we can find ways to counter attack the cases and prevent them. With proper analysis we can attain the main reason of such crime and find much more about the context of such events.

## OBJECTIVES

1. To provide information about the causes of such incidents
2. To optimize the ways of preventing such incidents in the near future
3. To located most target areas.
4. To make an analysis of why terrorism occurs

## COLUMNS AND DATA TYPES:

| | |
|---|---|
| eventid | string |
| iyear | int |
| imonth | int |
| iday | int |
| extendedd | int |
| country | string |
| country_txt | string |
| region | string |
| region_txt | string |
| provstate | string |
| city | string |
| latitude | double |
| longitude | double |
| specificity | int |
| vicinity | int |
| locationn | string |
| summary | string |
| crit1 | int |
| crit2 | int |
| crit3 | int |
| doubter | int |

| | |
|---|---|
| alternative | int |
| alternative_txt | string |
| multiple | int |
| success | int |
| suicide | int |
| attacktype1 | int |
| attacktype1_txt | string |
| targtype1 | int |
| targtype1_txt | string |
| targsubtype1 | int |
| targsubtype1_txt | string |
| corp1 | string |
| target1 | string |
| natlty1 | int |
| natlty1_txt | string |
| gname | string |
| motive | string |
| guncertain1 | int |
| individual | int |
| nperps | int |
| nperpcap | int |
| claimed | int |
| weaptype1 | int |
| weaptype1_txt | string |
| weapsubtype1 | int |
| weapsubtype1_txt | string |
| nkill | int |
| nkillter | int |
| nwound | int |
| nwoundte | int |
| property | int |
| propextent | int |
| propextent_txt | string |
| ishostkid | int |
| nhostkid | int |
| dbsource | string |

## PROBLEM STATEMENTS:

**Dataset Download link:**
https://www.kaggle.com/northon/globalterrorismdatabase-compact/download

1. Which country is mostly affected by terrorism?
2. Compare the number of terror attack between India and US.
3. Name the region highest affected by terrorism?
4. Display the name of the group that has carried most terror attack.
5. Count the no of success and failure of the terrorist attack.
6. List the general method of attack used.
7. List the highest number of fatalities by terrorism by Year
8. List the highest number of fatalities by month due to terrorism.
9. Compare the no of terror attacks in India by Year.

## CREATING A DATABASE AND A TABLE

```
hive> create table crime(eventid string,iyear int,imonth int,iday int,extendedd int,country stri
gion_txt string,provstate string ,city string,latitude double,longitude double,specificity int,v
tring,crit1 int,crit2 int,crit3 int,doubter int,alternative int,alternative_txt string,multiple
l int,attacktype1_txt string,targtype1 int,targtype1_txt string,targsubtype1 int,targsubtype1_tx
atlty1 int,natlty1_txt string,gname string,motive string,guncertain1 int,individual int,nperps i
int,weaptype1_txt string,weapsubtype1 int,weapsubtype1_txt string,nkill int,nkillter int,nwound
nt int,propextent_txt string,ishostkid int,nhostkid int,dbsource string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
```

## LOADING DATA INTO THE TABLE

```
hive> load data local inpath '/home/cloudera/Desktop/b/a.csv' into table crime;
Loading data to table projectbde.crime
Table projectbde.crime stats: [numFiles=1, totalSize=97220904]
OK
Time taken: 1.018 seconds
hive>
```

## PROBLEM 1:

**Which country is mostly affected by terrorism?**

## CREATING A SPECIFIC TABLE

```
hive> create external table details_country(eventid string,iyear int,imonth int,iday int,country_txt string,region_txt string,summary string,attacktype1_txt
string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/home/cloudera/Desktop/store/';
OK
Time taken: 0.05 seconds
hive>
```

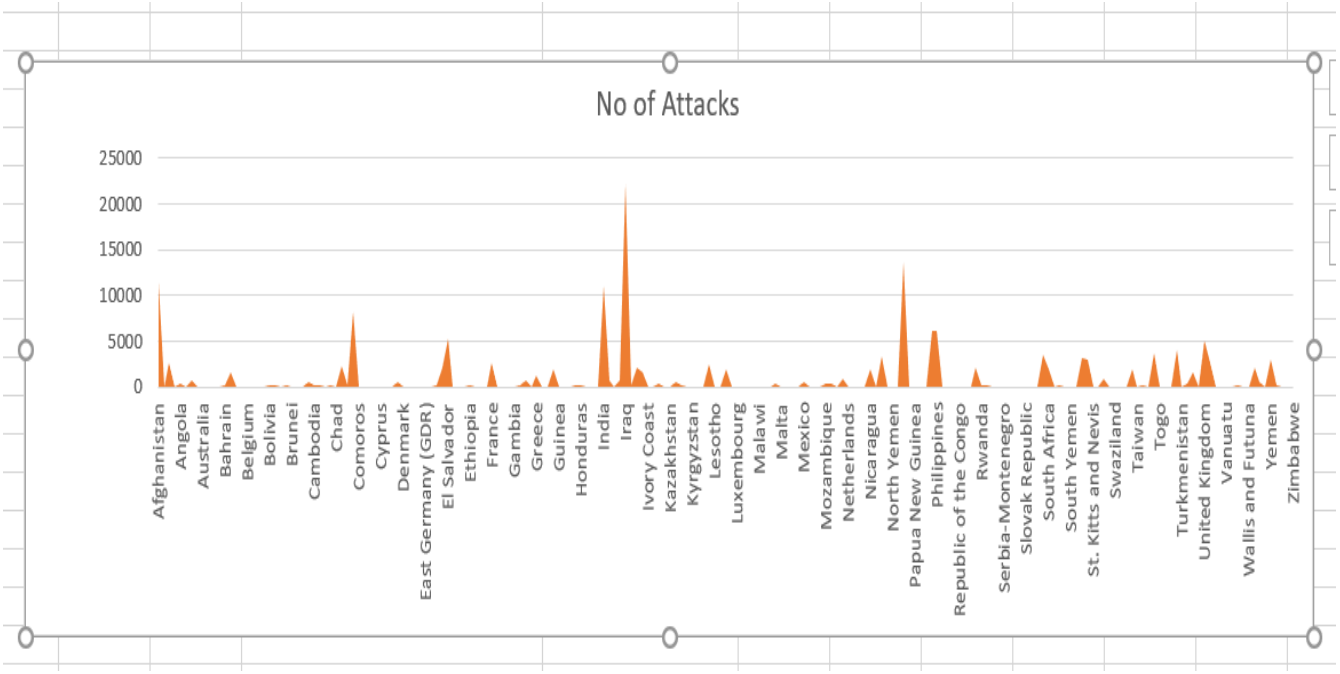
## INSERTING DATA TO IT:

```
hive> insert overwrite table details_country
    > select eventid,iyear,imonth,iday,country_txt,region_txt,summary,attacktype1_txt from crime;
Query ID = cloudera_20191031230303_f24e83a5-701e-4bb6-8409-cca604f1bf67
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572584136784_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572584136784_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2019-10-31 23:03:30,916 Stage-1 map = 0%,  reduce = 0%
2019-10-31 23:03:38,280 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.25 sec
MapReduce Total cumulative CPU time: 3 seconds 250 msec
Ended Job = job_1572584136784_0011
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/home/cloudera/Desktop/store/.hive-staging_hive_2019-10-31_23-03-24_021_3203756477131205155-1/-ext-10000
Loading data to table projectbde.details_country
Table projectbde.details_country stats: [numFiles=1, numRows=170350, totalSize=18176705, rawDataSize=18006355]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.25 sec   HDFS Read: 97228698 HDFS Write: 18176797 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 250 msec
OK
Time taken: 16.596 seconds
hive>
```

## FINAL QUERY:

```
hive> select * from details_country1 order by noofattack desc limit 1;
Query ID = cloudera_20191031231818_3989038a-399a-4493-a710-0fe9d147751b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572584136784_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572584136784_0013/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-10-31 23:18:35,959 Stage-1 map = 0%,  reduce = 0%
2019-10-31 23:18:41,180 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.06 sec
2019-10-31 23:18:48,480 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.46 sec
MapReduce Total cumulative CPU time: 2 seconds 460 msec
Ended Job = job_1572584136784_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.46 sec   HDFS Read: 9636 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 460 msec
OK
Iraq    22130
Time taken: 20.14 seconds, Fetched: 1 row(s)
hive>
```

## VISUALIZATION:



## PROBLEM 2:

**Compare the number of terror attack between India and US.**

## QUERY:

```
hive> create external table Us(Year int,country string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/lastq/';
OK
Time taken: 0.092 seconds
hive> insert overwrite table Us
    > select iyear,country_txt from crime where country_txt='United States';
Query ID = cloudera_20191103221919_b2f63f1b-a210-4d12-9c4c-a07ef9a73edd
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572842835005_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572842835005_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2019-11-03 22:20:06,599 Stage-1 map = 0%,  reduce = 0%
```

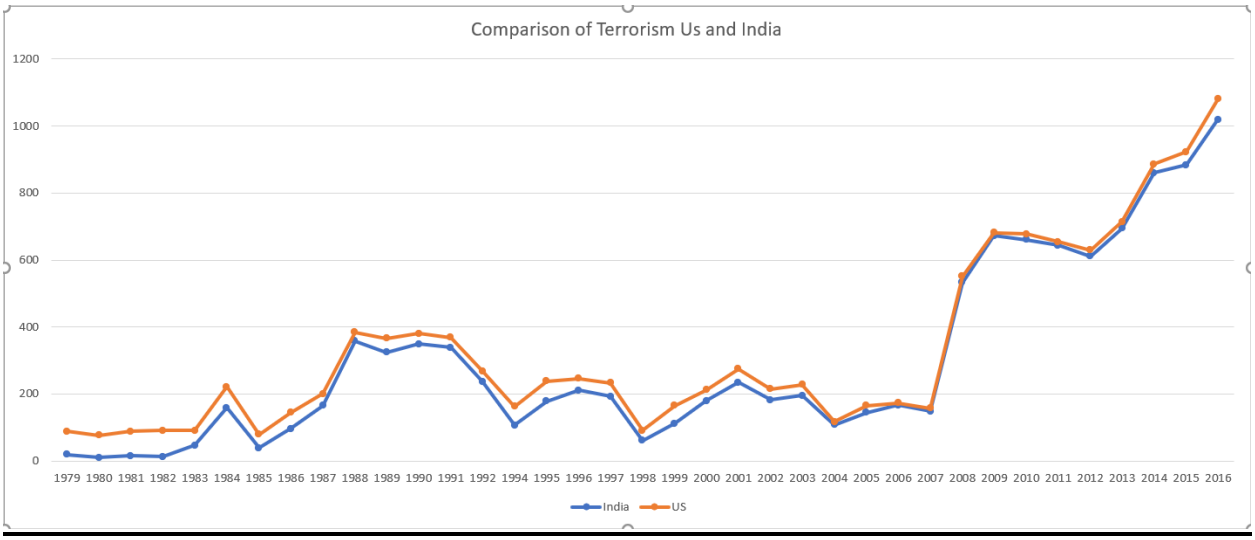
```
hive> create external table Usattack(Year int,noofattacksUS int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/lastq/';
OK
Time taken: 0.109 seconds
hive> insert overwrite table Usattack
    > select year,count(*) from Usattack group by Year;
Query ID = cloudera_20191103222323_3f60bb31-a0e3-42f8-b776-c122fe77a9e0
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572842835005_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572842835005_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0009
```

```
hive> create table UsIndia(Year int,noofattacksIndia int,noofattacksUs int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/lastq/finalout/';
OK
Time taken: 0.105 seconds
```

```
hive> insert overwrite table UsIndia
    > select i.year,i.noofattacks,u.noofattacksus from Indiaattack i full outer join Usattack u on (i.year=u.year);
Query ID = cloudera_20191103222929_fb53f13a-d8f2-48f0-8068-92a4b837c79d
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572842835005_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572842835005_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0010
```

```
hive> select * from USIndia;
OK
NULL    NULL    468
NULL    NULL    247
1972    1       64
NULL    NULL    58
NULL    NULL    94
1975    1       149
1976    1       105
1977    1       130
NULL    NULL    87
1979    20      69
1980    10      67
1981    16      73
1982    13      78
1983    47      44
1984    159     63
1985    39      40
1986    96      49
1987    166     34
1988    358     27
1989    324     42
1990    349     32
1991    339     30
1992    237     32
1994    107     56
1995    179     60
1996    211     35
1997    193     40
1998    61      30
1999    112     53
2000    180     32
2001    234     41
2002    182     33
2003    196     32
```

## VISUALIZATION:



## PROBLEM 3:

**Name the region highest affected by terrorism?**

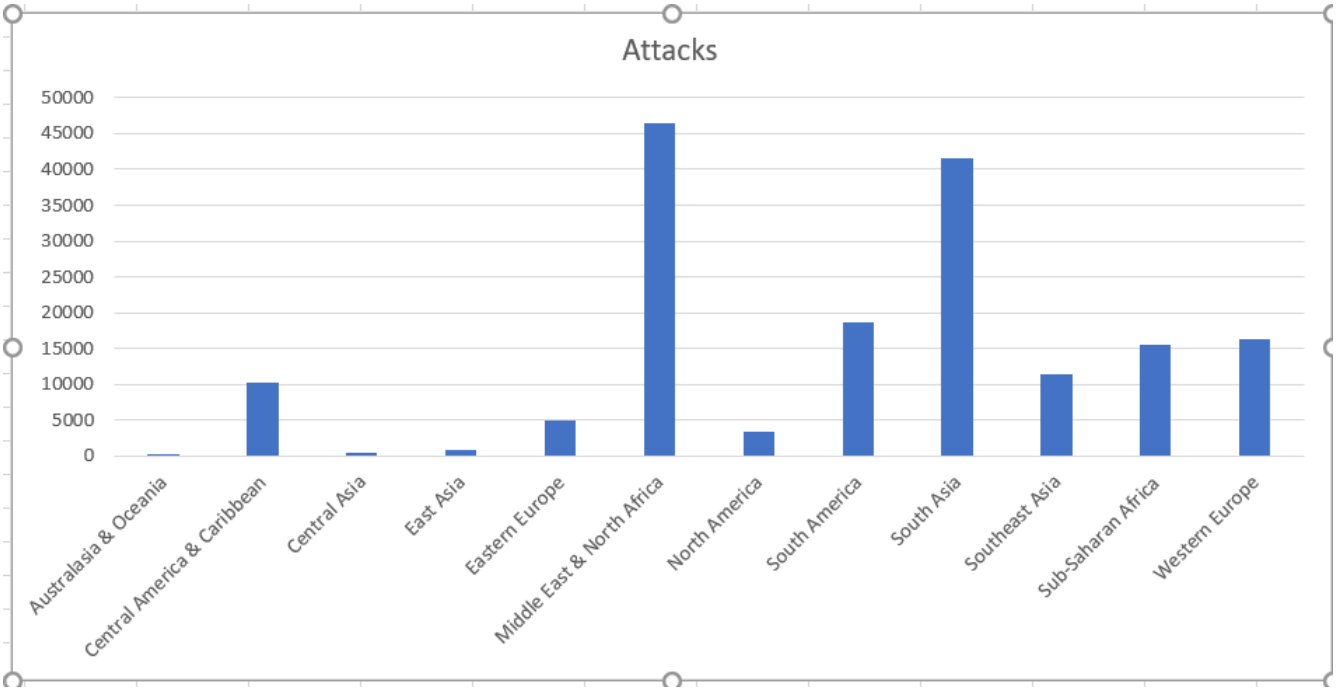
## CREATE AN EXTERNAL TABLE

```
hive> create external table region(region string,Noofattacks int)
    > row format delimited
    > fields terminated by ','
    > location '/user/cloudera/project/region/';
OK
Time taken: 0.056 seconds
hive> insert overwrite table region
    > select region_txt,count(*) from crime group by region_txt;
Query ID = cloudera_20191031235555_a78176c0-e78c-4b0d-89d1-21757e11694b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572584136784_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/applica
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-10-31 23:55:49,418 Stage-1 map = 0%,  reduce = 0%
2019-10-31 23:55:55,632 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.24 sec
```

## FINAL QUERY:

```
hive> select * from region order by noofattacks desc limit 1;
Query ID = cloudera_20191031235757_6446f3c1-453c-4dbc-a05f-21f8d1ca5864
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572584136784_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572584136784_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-10-31 23:57:54,596 Stage-1 map = 0%,  reduce = 0%
2019-10-31 23:58:00,903 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.91 sec
2019-10-31 23:58:07,186 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.34 sec
MapReduce Total cumulative CPU time: 2 seconds 340 msec
Ended Job = job_1572584136784_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.34 sec   HDFS Read: 7092 HDFS Write: 33 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 340 msec
OK
Middle East & North Africa      46511
Time taken: 19.402 seconds, Fetched: 1 row(s)
hive>
```

## VISUALIZATION:



## PROBLEM 4:

**Display the name of the group that has carried most terror attack.**

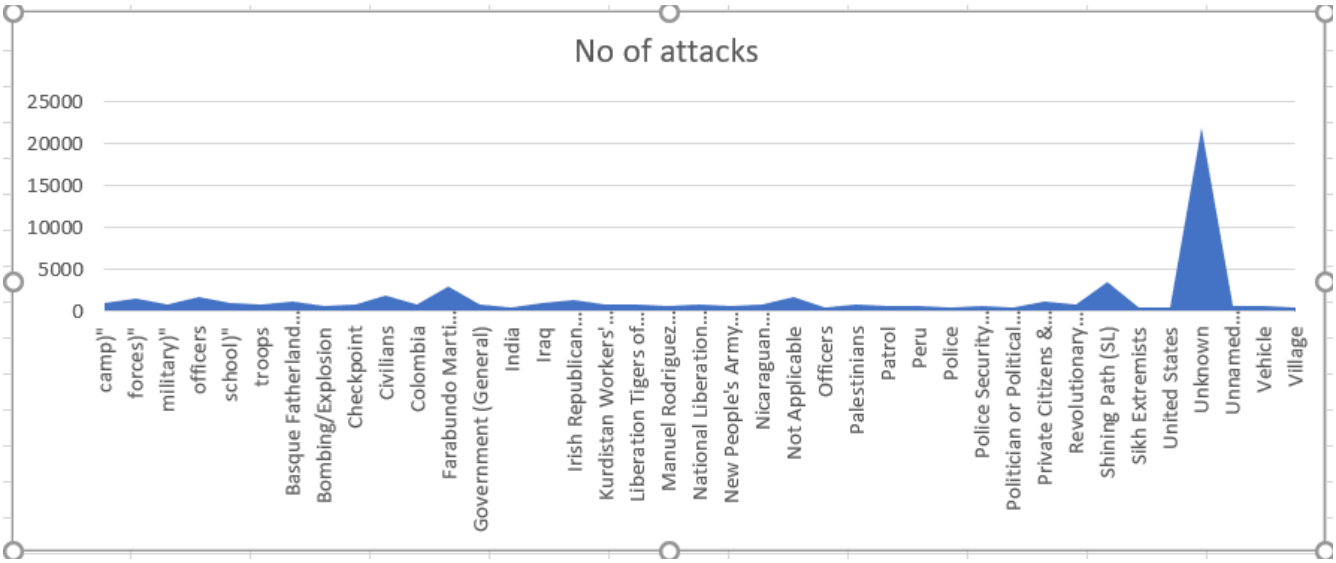
## CREATING A TABLE:

```
hive> create table about(eventid string,gname string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > ;
OK
Time taken: 0.104 seconds
hive> insert overwrite table about
    > select eventid,gname from crime;
Query ID = cloudera_20191101010404_1d03802b-a6e4-438b-82dc-4bb361b8f222
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572584136784_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572584136784_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2019-11-01 01:04:32,071 Stage-1 map = 0%,  reduce = 0%
2019-11-01 01:04:46,316 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.91 sec
MapReduce Total cumulative CPU time: 6 seconds 910 msec
Ended Job = job_1572584136784_0018
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/projectbde.db/about/.hive-staging_hive_2019-11-01_01-04-19_520_9144932137091898354-1/-ext
-10000
Loading data to table projectbde.about
Table projectbde.about stats: [numFiles=1, numRows=170350, totalSize=4695960, rawDataSize=4525610]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 6.91 sec   HDFS Read: 97228202 HDFS Write: 4696041 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 910 msec
OK
Time taken: 29.539 seconds
```

## FINAL QUERY:

```
hive> select * from terrname order by noofattack desc limit 1;
Query ID = cloudera_20191101011313_b5659520-0e91-4377-96ac-0473fccd12cc
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572584136784_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572584136784_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572584136784_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-01 01:13:51,169 Stage-1 map = 0%,  reduce = 0%
2019-11-01 01:14:03,326 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.15 sec
2019-11-01 01:14:15,583 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 9.27 sec
MapReduce Total cumulative CPU time: 9 seconds 270 msec
Ended Job = job_1572584136784_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 9.27 sec   HDFS Read: 518211 HDFS Write: 14 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 270 msec
OK
Unknown 21998
Time taken: 37.91 seconds, Fetched: 1 row(s)
hive>
```

## VISULAIZATION



**PROBLEM 5:**

**Count the no of success and failure of the terrorist attack.**

**COMMANDS:**

**QUERY FOR FINDING THE NO OF SUCCESS:**

```
hive> select count(*) from Op where success=1;
Query ID = cloudera_20191101032121_a48440a8-9680-4f10-837e-9cc5d36d4c0b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
```

**RESULT:**

```
Total MapReduce CPU Time Spent: 3 seconds 780 msec
OK
89210
Time taken: 21.61 seconds, Fetched: 1 row(s)
```

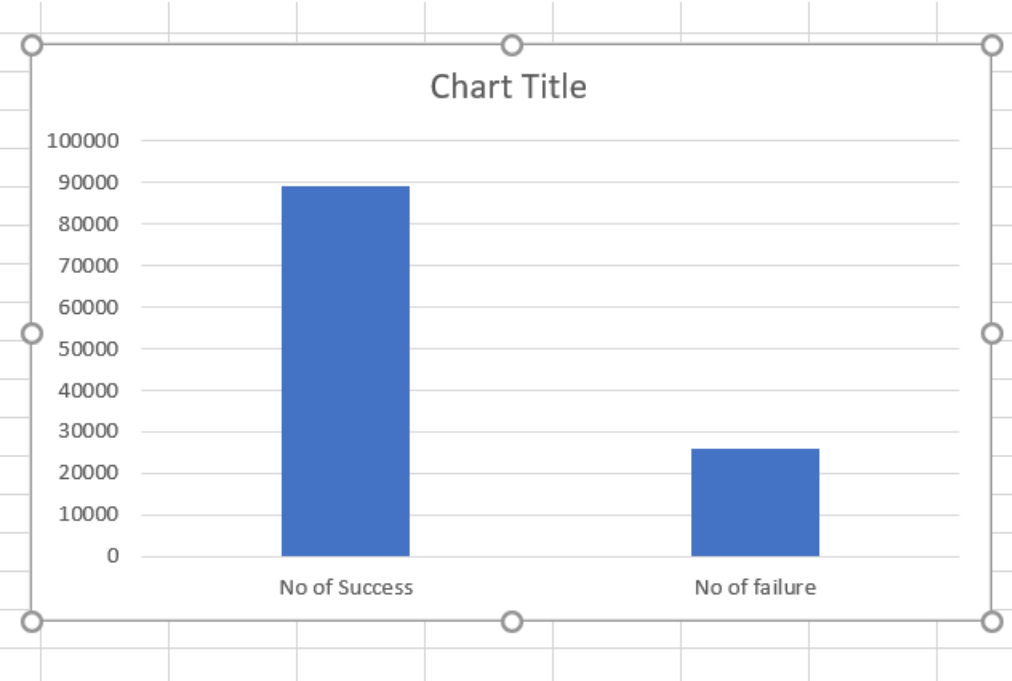
**QUERY FOR FINDING THE NO OF FAILURE :**

```
hive> select count(*) from Op where success=0;
Query ID = cloudera_20191101032525_e56ee764-fba4-40c7-946e-4fc48740eb68
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```

**RESULT:**

```
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
26125
Time taken: 21.064 seconds, Fetched: 1 row(s)
```

## VISUALIZATION



**PROBLEM 6:**

**List the general method of attack used.**

**CREATING EXTERNAL TABLE**

```
hive> create external table categorytype(category string)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/attack/'
    > ;
OK
Time taken: 0.061 seconds
```

**INSERTING DATA TO THE EXTERNAL TABLE:**

```
hive> insert overwrite table categorytype
   > select attacktype1_txt from crime where attacktype1_txt IN ('Assassination','Hostage Taking (Kidnapping)','Bombing/Explosion','Facility/Infrastructure Attack','Armed Assault','Hijacking','Unknown');
Query ID = cloudera_20191102124949_c4a14f77-0e3b-49db-b57b-d15dc3b520de
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572720843485_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572720843485_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572720843485_0006
```

**CREATING ANOTHER EXTERNAL TABLE FOR STORING THE CATEGORY AND GROUP BY COUNT INFORMATION :**

```
hive> create external table categorycount(category string,noofoccurance int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/cloudera/attacktype/count/';
OK
Time taken: 0.062 seconds
hive>
```

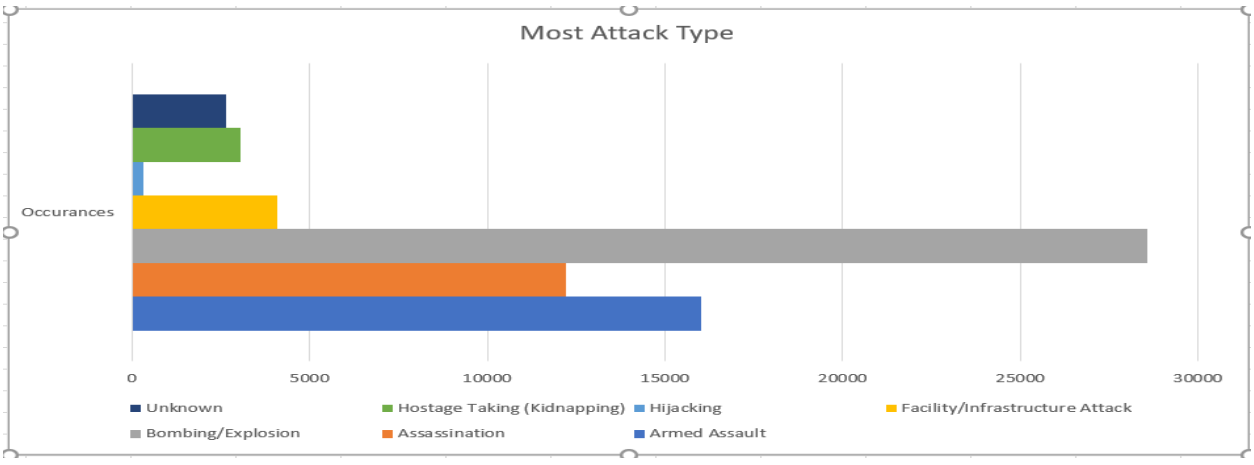
## INSERTING THE DATA FROM THE CRIME TABLE:

```
hive> insert overwrite table categorycount
    > select category,count(*) from categorytype group by category;
Query ID = cloudera_20191102125353_6ff711d8-6040-46f5-ac75-63c8074dc5ad
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
```

## FINAL QUERY :

```
hive> select * from categorycount order by  noofoccurance desc limit 1;
Query ID = cloudera_20191102125555_e7da8ed3-20b1-4814-8477-9dcd7a0f9392
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572720843485_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572720843485_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572720843485_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-02 12:55:41,419 Stage-1 map = 0%,  reduce = 0%
2019-11-02 12:55:46,695 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.02 sec
2019-11-02 12:55:52,922 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.37 sec
MapReduce Total cumulative CPU time: 2 seconds 370 msec
Ended Job = job_1572720843485_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.37 sec   HDFS Read: 7158 HDFS Write: 24 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 370 msec
OK
Bombing/Explosion       28569
Time taken: 19.332 seconds, Fetched: 1 row(s)
hive>
```

## VISULAIZATION:



## PROBLEM 7:

**List the highest number of fatalities by terrorism by Year**

## QUERIES:

```
hive> create external table kills_year(iyear int,attacks int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/newquestion/';
OK
Time taken: 1.304 seconds
hive>
```

```
hive> insert overwrite table kills_year
    > select iyear,nkill from crime;
Query ID = cloudera_20191103202828_e0ca37d2-ebf3-4743-a09c-705416b3ac88
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572841308881_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572841308881_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572841308881_0001
```

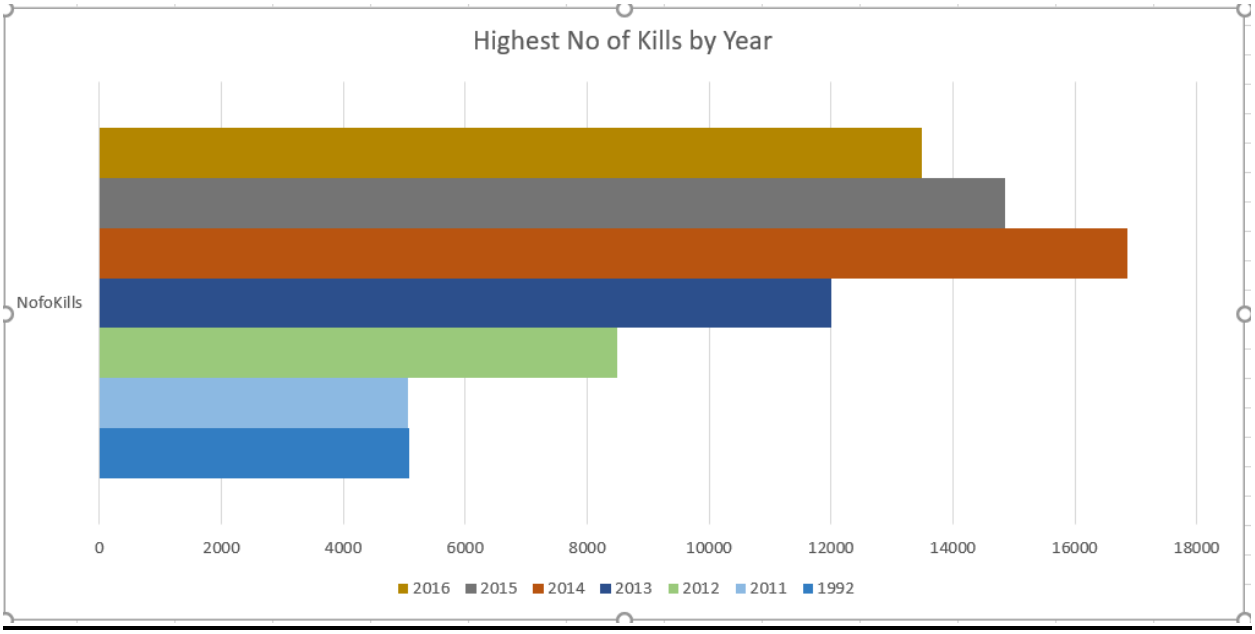
```
hive> select sum(attacks) from kills_year;
Query ID = cloudera_20191103203030_da1bc8fe-7e88-471b-990a-dcc5a350ec43
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572841308881_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572841308881_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572841308881_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-03 20:30:14,543 Stage-1 map = 0%,  reduce = 0%
2019-11-03 20:30:22,847 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.86 sec
2019-11-03 20:30:29,073 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.16 sec
MapReduce Total cumulative CPU time: 4 seconds 160 msec
Ended Job = job_1572841308881_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.16 sec   HDFS Read: 1278451 HDFS Write: 6 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 160 msec
OK
83953
Time taken: 21.429 seconds, Fetched: 1 row(s)
hive>
```

```
hive> insert overwrite table kills1
    > select iyear,count(*) from kills_year group by iyear;
Query ID = cloudera_20191103203232_75bf3e04-77ec-4ff6-81a6-8c8b5655bdd9
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572841308881_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572841308881_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572841308881_0003
```

## FINAL QUERY :

```
hive> select * from kills1 order by desc limit 1;
FAILED: SemanticException Line 0:-1 Invalid table alias or column reference 'desc': (possible column names are: year, noofkills)
hive> select * from kills1 order by noofkills desc limit 1;
Query ID = cloudera_20191103203333_576ecff4-a270-41ca-9a0b-62fba8682184
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572841308881_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572841308881_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572841308881_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-03 20:33:51,361 Stage-1 map = 0%,  reduce = 0%
2019-11-03 20:33:56,553 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.97 sec
2019-11-03 20:34:02,780 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.06 sec
MapReduce Total cumulative CPU time: 2 seconds 60 msec
Ended Job = job_1572841308881_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.06 sec   HDFS Read: 7266 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 60 msec
OK
2014    16860
Time taken: 17.293 seconds, Fetched: 1 row(s)
hive>
```

## VISUALIZATION:



**PROBLEM 8:**

**List the highest number of fatalities by month due to terrorism.**

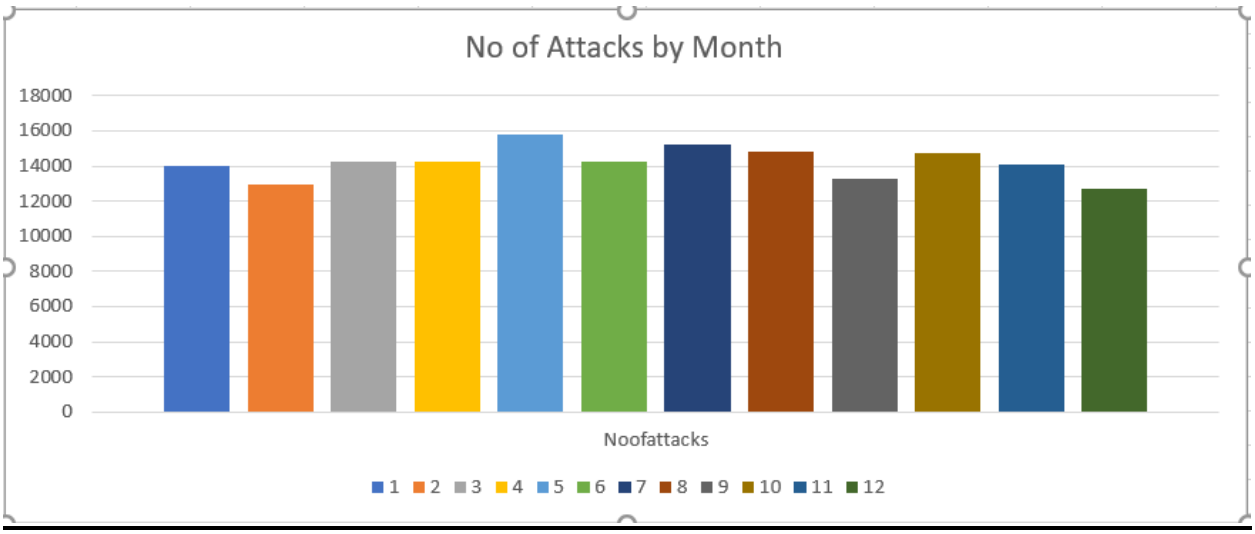
**QUERIES:**

```
hive> create external table month(imonth int,noofattacks int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/newquestion/2/';
OK
Time taken: 1.132 seconds
hive>
```

```
hive> insert overwrite table month
    > select imonth,count(*) from crime group by imonth;
Query ID = cloudera_20191103205252_a0d3f77a-d864-4430-944f-04cb0cd61bd4
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
```

```
hive> select * from month order by noofattacks desc limit 1;
Query ID = cloudera_20191103205353_5884fb61-54f2-4ac0-b524-9af539b6bf2c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572842835005_0002, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1572842835005_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-03 20:53:39,920 Stage-1 map = 0%,  reduce = 0%
2019-11-03 20:53:47,226 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.07 sec
2019-11-03 20:53:54,517 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.29 sec
MapReduce Total cumulative CPU time: 3 seconds 290 msec
Ended Job = job_1572842835005_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.29 sec   HDFS Read: 6935 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 290 msec
OK
5       15771
Time taken: 22.309 seconds, Fetched: 1 row(s)
```

**VISUALIZATION:**



**PROBLEM 9:**

**Compare the no of terror attacks in India by Year.**

**QUERY:**

```
hive> create external table Country(country string,iyear int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/newquestion/India/';
OK
Time taken: 0.042 seconds
```

```
hive> insert overwrite table Country
   > select country_txt,iyear from crime where country_txt='India';
Query ID = cloudera_20191103212323_d5089b24-926f-45bc-b961-41e7458ef8dd
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1572842835005_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572842835005_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0004
```

```
hive> select * from country limit 5;
OK
India   1972
India   1975
India   1976
India   1977
India   1979
Time taken: 0.043 seconds, Fetched: 5 row(s)
hive> create external table Indiaattack(Year int,noofattacks int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile
    > location '/user/cloudera/project/kaustav/';
OK
Time taken: 0.046 seconds
hive>
```

```
hive> insert overwrite table Indiaattack
    > select iyear,count(*) from country group by iyear;
Query ID = cloudera_20191103212626_9abf0ee7-aa75-4958-9a3d-67f26d7636ae
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
```

## FINAL QUERY:

```
hive> select * from Indiaattack order by noofattacks desc limit 1;
Query ID = cloudera_20191103212727_29997989-4df8-4470-a708-df45e988351c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1572842835005_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1572842835005_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1572842835005_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2019-11-03 21:27:31,662 Stage-1 map = 0%,  reduce = 0%
2019-11-03 21:27:36,878 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 0.91 sec
2019-11-03 21:27:42,096 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.1 sec
MapReduce Total cumulative CPU time: 2 seconds 100 msec
Ended Job = job_1572842835005_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.1 sec   HDFS Read: 7281 HDFS Write: 10 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 100 msec
OK
2016    1019
Time taken: 18.819 seconds, Fetched: 1 row(s)
```

## VISUALIZATION