

Crime Dataset:

The columns in Crime.csv are Area_Name,Year,Group_Name,Sub_Group_Name,Cases_Property_Recovered,Cases_Property_Stolen,Value_of_Property_Recovered,Value_of_Property_Stolen

1.Remove the headers

2.Copy the local file in hdfs

```
#hadoop dfs -put 'file' /
```

LOAD Dataset:

```
crime_record = LOAD '/Crime.csv' using PigStorage(',') as  
(area:chararray,year:int,groups:chararray,subgroups:chararray,casesrecovered:int,casesstolen:in  
t,valuerecovered:int,valuestolen:int);
```

```
grunt> crime_record = LOAD '/Crime.csv' using PigStorage(',') as (area:chararray  
,year:int,groups:chararray,subgroups:chararray,casesrecovered:int,casesstolen:in  
t,valuerecovered:int,valuestolen:int);
```

View Dataset:

```
dump crime_record;
```

```
(Rajasthan,2010,Total Property,7. Total Property Stolen & Recovered,8551,28152,8  
54388626,1395764020)  
(Sikkim,2010,Total Property,7. Total Property Stolen & Recovered,38,134,1444190,  
9445146)  
(Tamil Nadu,2010,Total Property,7. Total Property Stolen & Recovered,16125,21509  
,660311804,1317919190)  
(Tripura,2010,Total Property,7. Total Property Stolen & Recovered,192,879,566610  
2,33032746)  
(Uttar Pradesh,2010,Total Property,7. Total Property Stolen & Recovered,9130,350  
68,577591772,1442670414)
```

Analysing Dataset:

1)Crimes accured in TamilNadu:

```
filter_data = FILTER crime_record by area == 'Tamil Nadu';  
dump filter_data;
```

```
(Tamil Nadu,2002,Total Property,7. Total Property Stolen & Recovered,19263,25825  
,391752230,695528915)  
(Tamil Nadu,2003,Total Property,7. Total Property Stolen & Recovered,18758,24082  
,368499696,539833082)  
(Tamil Nadu,2004,Total Property,7. Total Property Stolen & Recovered,18546,22489  
,356073796,554827854)  
(Tamil Nadu,2005,Total Property,7. Total Property Stolen & Recovered,17227,20352  
,313875556,434945098)  
(Tamil Nadu,2006,Total Property,7. Total Property Stolen & Recovered,15620,17787  
,357485712,488594368)  
(Tamil Nadu,2007,Total Property,7. Total Property Stolen & Recovered,15374,17875  
,495527011,680515631)
```

2)Crimes in TamilNadu at 2010:

```
filter_data = FILTER crime_record by area == 'Tamil Nadu' and year==2010;  
dump filter_data;
```

```
(Tamil Nadu,2010,Burglary - Property,3. Burglary,2953,4715,120836243,247787824)  
(Tamil Nadu,2010,Criminal Breach of Trust - Property,5. Criminal Breach of Trust  
,49,156,10621862,375413547)  
(Tamil Nadu,2010,Dacoity -Property,1. Dacoity,74,85,62272356,80347154)  
(Tamil Nadu,2010,Other heads of Property,6. Other Property,139,153,31084802,3279  
7329)  
(Tamil Nadu,2010,Robbery - Property,2. Robbery,1326,1817,70981709,117163496)  
(Tamil Nadu,2010,Theft - Property,4. Theft,11584,14583,364514832,464409840)  
(Tamil Nadu,2010,Total Property,7. Total Property Stolen & Recovered,16125,21509  
,660311804,1317919190)
```

3)Crimes Average Value Stolen by Area:

```
GroupByArea = GROUP crime_record by area;  
dump GroupByArea;  
area_avg_value = foreach GroupByArea GENERATE  
group,AVG(crime_record.valuestolen);  
dump area_avg_value;
```

```
(Tamil Nadu,2.0060000668571427E8)  
(Daman & Diu,5360775.0)  
(Lakshadweep,142943.89855072464)  
(Maharashtra,8.03073665468085E8)  
(Uttarakhand,2.336641665714286E7)  
(West Bengal,2.105838181029412E8)  
(Chhattisgarh,8.243038345714286E7)  
(Uttar Pradesh,3.408068046857143E8)
```

4)Maximum Value Recovered by year:

```
GroupByyear = GROUP crime_record by year;  
dump GroupByyear;  
max_val_rec = FOREACH GroupByyear GENERATE  
group,MAX(crime_record.valuerecovered);  
dump max_val_rec;
```

```
(2001,856697065)  
(2002,1300723470)  
(2003,749453074)  
(2004,810624001)  
(2005,885513709)  
(2006,1038628130)  
(2007,1462054517)
```

5)Average ValueStolen of cases by area and year:

```
GroupByareayear = GROUP crime_record by (area,year);  
dump GroupByareayear;  
avg_area_year = FOREACH GroupByareayear GENERATE group,  
AVG(crime_record.valuestolen);  
dump avg_area_year;
```

```
((Puducherry,2008),8576757.714285715)  
((Puducherry,2009),8311051.142857143)  
((Puducherry,2010),8642709.714285715)  
((Tamil Nadu,2001),2.1000393657142857E8)  
((Tamil Nadu,2002),1.9872254714285713E8)  
((Tamil Nadu,2003),1.5423802342857143E8)  
((Tamil Nadu,2004),1.58522244E8)  
((Tamil Nadu,2005),1.24270028E8)
```