ABSTRACT

Crime Prediction is mainly concerned with predicting where and when and type of crime which will happen. This project is based on the research paper Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots published by International Journal of Data Mining & Knowledge Management Process (IJDKP).

In this project we will be using datasets available in the internet. There are multiple datasets available online. In the research paper, the datasets which are used are Denver Crimes Dataset and Los Angeles Crimes Dataset. The three algorithms which we are using after data pre-processing are all well-known classification methods of machine learning algorithms.

Decision Tree Algorithm, Naive Bayes Algorithm Apriori Algorithm and Random Forest are all supervised learning algorithm in the scope of project. This project not only predicts crime but also predicts crime-hotspots and subsequent crimes which follows one crime after another.

Due to large variety of crimes, crimes in this project are mainly categorized into Burglary, Robbery, Vandalism, Aggravated and Assault.

INTRODUCTION

Platform Details:

RStudio is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics.

RStudio is available in open source and commercial editions and runs on the desktop (Windows, macOS, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro

RStudio provides a mechanism for executing R functions interactively from within the IDE through the Addins menu. This enables packages to include Graphical User Interfaces (GUIs) for increased accessibility. Popular packages that use Addins based GUIs include:

- bookdown - a knitr extension to create books
  - colourpicker - a graphical tool to pick colours for plots
  - datasets.load - a graphical tool to search and load datasets
- googleAuthR - Authenticate with Google APIs System

algorithm 1 description :

Apriori Algorithm

Apriori is one of the basic algorithms for mining frequent patterns. It scans the dataset to collect all itemsets that satisfy a predefined minimum support. We implemented this model using an open source tool.

We implemented the algorithm on location and time features and excluded the crime type feature.

Pseudocode

C_k: Candidate item set of size k

L_k : Frequent item set of size k

L_1= {frequent items};

For (k=1; L_k!=Φ;k++) do begin

C_{k+1} = candidates generated from L_k;

For each transaction t in database do

Increment the count of all candidates in C_{k+1}

Those are contained in t

L_{k+1} = candidates in C_{k+1} with min_support

End

Return U_k L_k;

Algorithm 2 description

A naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers assume strong, or naive, independence between attributes of data points. Popular uses of naive Bayes classifiers include spam filters, text analysis and medical diagnosis. These classifiers are widely used for machine learning because they are simple to implement.

Naive Bayes is also known as simple Bayes or independence Bayes. Pseudocode

Input:

Training dataset T,

F= (f_1, f_2, f_3,.., f_n) // value of the predictor variable in testing dataset.

Output:

A class of testing dataset.

Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of f_i using the gauss density equation in each class;

   Until the probability of all predictor variables (f_1, f_2, f_3,.., f_n) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

Algorithm 3 description

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too.

The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data (training data).

The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each **internal node** of the tree corresponds to an attribute, and each **leaf node** corresponds to a class label.

Pseudocode

1. Place the best attribute of the dataset at the **root** of the tree.
2. Split the training set into **subsets**. Subsets should be made in such a way that each subset contains data with the same value for an attribute.

3. Repeat step 1 and step 2 on each subset until you find **leaf nodes** in all the branches of the tree.

Algorithm 4 description

Random forest algorithm

Random forests or **random decision forests** are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

Pseudocode

1. Randomly select **“k”** features from total **“m”** features.
   1. Where **k << m**
2. Among the **“k”** features, calculate the node **“d”** using the best split point.
3. Split the node into **daughter nodes** using the **best split**.
4. Repeat **1 to 3** steps until “l” number of nodes has been reached.
5. Build forest by repeating steps **1 to 4** for “n” number times to create **“n” number of trees**.

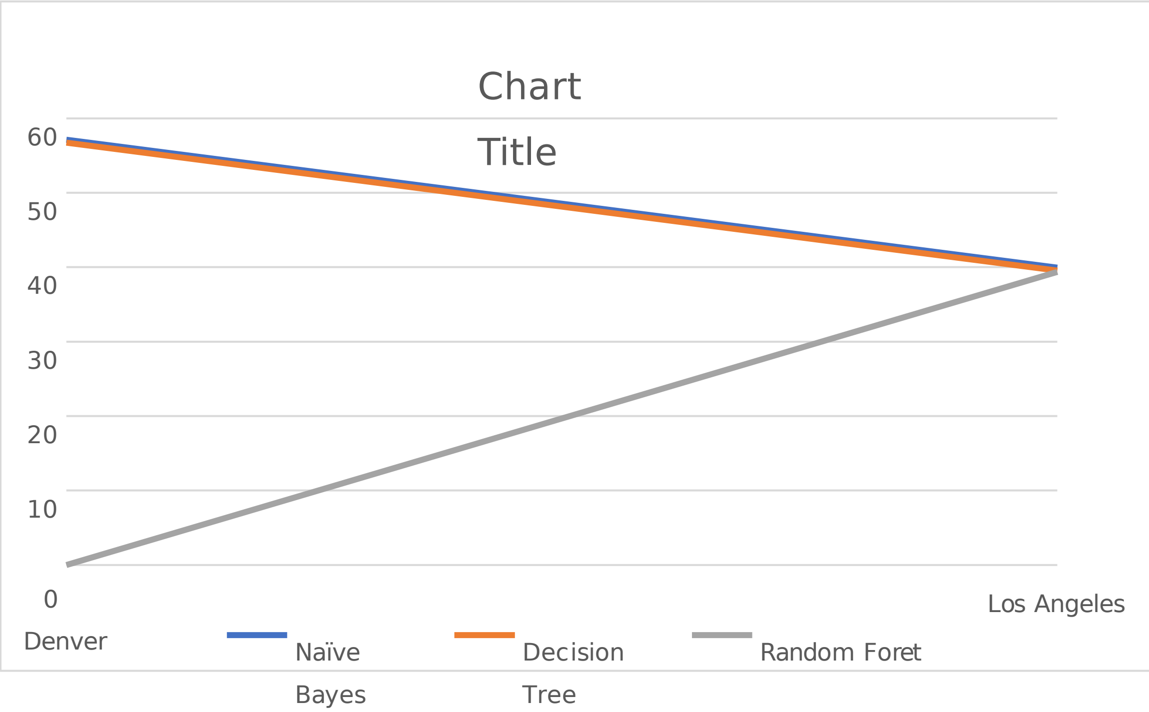
Proposed system

Our proposed system is that we not use crime prediction for predicting the crime but we also use crime prediction to find areas with more crimes. Our system also provides provisions on which crimes associate together. Getting knowledge of what types of crimes, times, day and location where they occur or will be occurring can be major boost in surveillance in that areas which can further enhance the security.

Result analysis

Dataset And Algorithm	**Accuracy in %**
Denver Naive Bayes	57.1 %

Los Angeles Naive Bayes	39.91%
Denver Decision Tree	56.7%
Los Angeles Decision Tree	39.5%
Los Angeles Random Forest	39.38%



Conclusion

We applied Apriori algorithm to find frequent crime patterns in both cities. After that, we applied Decision Tree and Naïve Bayesian classifiers to help predicting future crimes in a specific location within a particular time. We achieved 57% of prediction accuracy in Denver and 4O% prediction accuracy in Los Angeles. We achieved more accuracy than the research paper in Denver Dataset.

Source Code

1. Datasets

1.1) Denver Crimes Dataset

denverdata<-read.csv("C:/Users/Srix/Desktop/crime.csv")

A1 | INCIDENT_ID

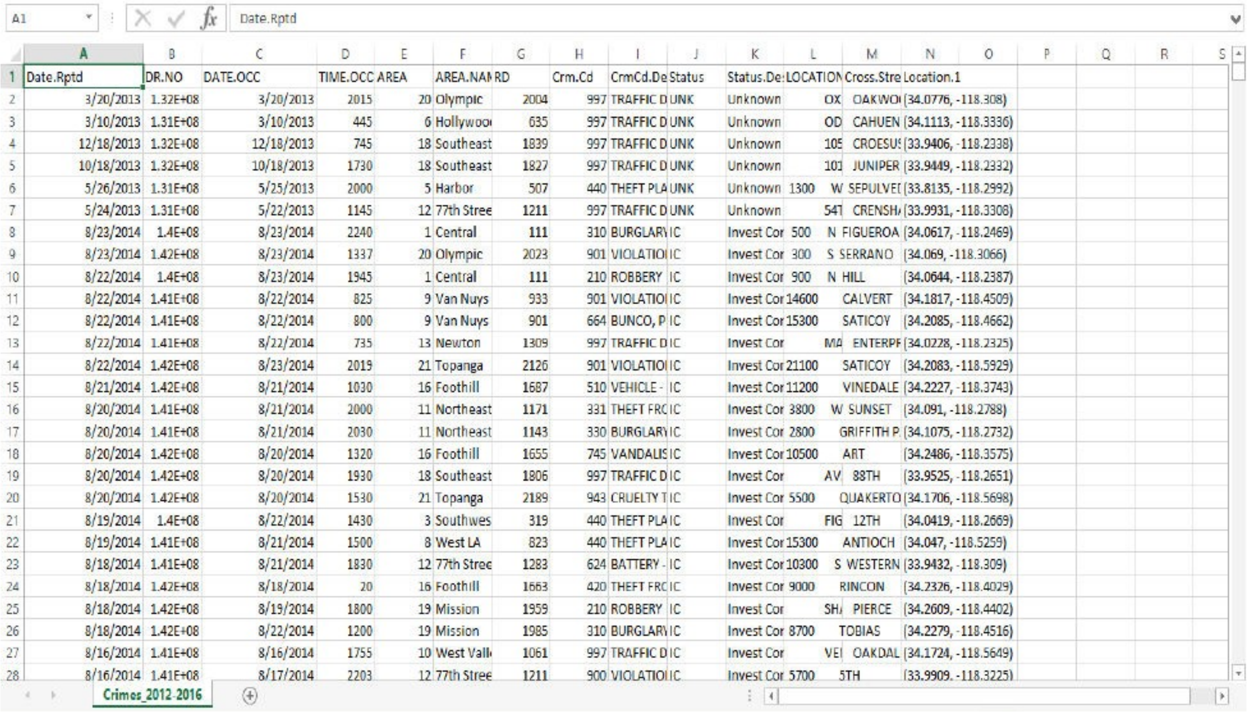
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	INCIDENT	OFFENSE_	OFFENSE_	OFFENSE_	OFFENSE_	OFFENSE_	FIRST_OCCURRENCE_I	LAST_OCCURREN	REPORTED_DAT	INCIDENT_	GEO_X	GEO_Y	GEO_LON	GEO_LAT	DISTRICT_	PRECINCT	NEIGHBOI	IS_CRIME	IS
2	2.02E+09	2.02E+15	5213	0	weapon-u	all-other-	6/15/2016 23:31		6/15/2016 23:31		3193983	1707251	-104.81	39.77319	5	521	montbell	1	
3	2.02E+10	2.02E+16	2399	0	theft-othe	larceny	10/11/2017 12:30	10/11/2017 16:55	1/29/2018 17:53		3201943	1711852	-104.781	39.78565	5	522	gateway-g	1	
4	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	3/4/2016 20:00	4/25/2016 8:00	4/26/2016 21:02	2932 S JOS	3152762	1667011	-104.957	39.66349	3	314	wellshire	1	
5	2.02E+08	2.02E+14	2399	0	theft-othe	larceny	1/30/2018 19:20		1/30/2018 22:29	705 S COL	3157162	1681320	-104.941	39.7027	3	312	belcaro	1	
6	2.02E+09	2.02E+15	2303	0	theft-shop	larceny	6/22/2017 20:53		6/23/2017 16:09	2810 E 1ST	3153211	1686545	-104.955	39.71711	3	311	cherry-cre	1	
7	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/31/2018 0:44		1/31/2018 1:29	2100 BLOC	3151310	1696020	-104.962	39.74315	6	622	city-park-	1	
8	2.02E+10	2.02E+16	2304	0	theft-part	theft-fron	6/1/2017 12:15	1/26/2018 12:15	1/26/2018 12:24	995 N FED	3133441	1692147	-105.026	39.73279	1	122	villa-park	1	
9	2.02E+08	2.02E+14	5707	0	criminal-t	all-other-	1/30/2018 7:40		1/30/2018 10:33	E SPEER BI	3145202	1688799	-104.984	39.72342	3	311	speer	1	
10	2.02E+08	2.02E+14	5401	0	traffic-acc	traffic-acc	1/30/2018 9:10		1/30/2018 9:17	W 13TH AV	3142965	1693682	-104.992	39.73686	6	611	civic-cent	0	
11	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	1/31/2018 0:55	1/31/2018 6:55	1/31/2018 7:07	2828 N ZU	3136231	1701209	-105.015	39.75763	1	113	highland	1	
12	2.02E+08	2.02E+14	3572	0	drug-metl	drug-alcol	1/30/2018 20:04		1/30/2018 22:12	E EVANS A	3161788	1672521	-104.925	39.67846	3	323	goldsmith	1	
13	2.02E+10	2.02E+16	2304	0	theft-part	theft-fron	1/28/2018 5:25	1/30/2018 17:25	1/30/2018 17:42	1220 S INC	3141343	1678270	-104.998	39.69458	4	422	ruby-hill	1	
14	2.02E+09	2.02E+15	2303	0	theft-shop	larceny	5/31/2016 16:54		5/31/2016 18:30	1505 S COI	3156992	1676337	-104.942	39.68902	3	312	cory-merr	1	
15	2.02E+09	2.02E+15	2404	0	theft-of-n	auto-theft	2/24/2017 16:55	5/24/2017 8:01	6/12/2017 15:24	2850 W 26	3134022	1700050	-105.023	39.75448	1	121	jefferson-	1	
16	2.02E+09	2.02E+15	2303	0	theft-shop	larceny	4/16/2016 15:25		4/16/2016 16:06	950 S QUE	3168039	1680120	-104.903	39.69921	3	321	windsor	1	
17	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	5/28/2017 2:00	5/28/2017 7:00	5/28/2017 19:58	3400 BLK N	3136548	1703738	-105.014	39.76456	1	113	highland	1	
18	2.02E+09	2.02E+15	2404	0	theft-of-n	auto-theft	6/16/2017 21:30	6/17/2017 7:12	6/17/2017 7:12	6825 E ILIF	3165912	1671324	-104.911	39.6751	3	323	goldsmith	1	
19	2.02E+08	2.02E+14	5707	0	criminal-t	all-other-	1/29/2018 14:50		1/29/2018 14:50	10TH / CO	3140932	1694856	-104.999	39.74012	1	123	lincoln-pa	1	
20	2.02E+08	2.02E+14	2609	0	fraud-by-	white-coll	1/4/2018 12:00	1/28/2018 18:00	1/29/2018 14:15	803 S PEAI	3146104	1680940	-104.981	39.70183	3	312	washingtc	1	
21	2.02E+08	2.02E+14	5441	0	traffic-acc	traffic-acc	1/30/2018 12:38		1/30/2018 12:38	N HAVAN	3178274	1708100	-104.866	39.77583	5	511	stapleton	0	
22	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/30/2018 11:27		1/30/2018 12:00	3900 BLK N	3183349	1706918	-104.848	39.77249	5	512	stapleton	1	
23	2.02E+08	2.02E+14	5441	0	traffic-acc	traffic-acc	1/29/2018 14:16		1/29/2018 15:00	E LOWRY E	3173298	1689004	-104.884	39.72351	3	321	lowry-fiel	0	
24	2.02E+08	2.02E+14	5401	0	traffic-acc	traffic-acc	1/30/2018 15:45		1/30/2018 17:13	29TH ST /	3147424	1700922	-104.976	39.75667	2	211	five-point	0	
25	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/30/2018 12:30		1/30/2018 13:09	3200 BLOC	3154698	1710827	-104.95	39.78374	2	212	elyria-swa	1	
26	2.02E+08	2.02E+14	2203	0	burglary-t	burglary	1/29/2018 21:30	1/30/2018 8:30	1/30/2018 8:57	5770 E WA	3163262	1671541	-104.92	39.67575	3	323	goldsmith	1	
27	2.02E+08	2.02E+14	2799	0	theft-emt	white-coll	10/30/2017 12:00	12/21/2017 12:00	1/30/2018 15:19	7400 E HAI	3168440	1663019	-104.902	39.65226	3	324	hampden-	1	
28	2.02E+10	2.02E+16	1316	0	threats-to	public-dis	1/30/2018 14:30		1/30/2018 15:50	8900 PENA	3231979	1738721	-104.674	39.8587	7	759	dia	1	

crime

Attributes	**Data Type**	**Number Of Distinct**	**Value**
Offense_Category_Id	Nominal	14 Categories	AggravatedAssault All-other-crimes Arson Burglary Drug-alcohol Larceny Murder Other- crimesagainst- persons Public-disorder Robbery Sexual-assault Theft- frommotor-vehicle White-collarcrime
First_occurance Date	Date and time	Unlimited	6/13/14 21:30
Neighborhood_Id	Nominal	78 Neighborhoods	
Is_crime	Binary	2 values	0 or 1

1.2) Los Angeles Crimes Dataset

losdata<-read.csv("C:/Users/Srix/Desktop/Crimes_2012-2016.csv")



	Date.Rptd	DR.NO	DATE.OCC	TIME.OCC	AREA	AREA.NAM	RD	Crm.Cd	CrmCd.De	Status	Status.De	LOCATION	Cross.Stre	Location.1
2	3/20/2013	1.32E+08	3/20/2013	2015	20	Olympic	2004	997	TRAFFIC D	UNK	Unknown	OX	OAKWO	(34.0776, -118.308)
3	3/10/2013	1.31E+08	3/10/2013	445	6	Hollywoo	635	997	TRAFFIC D	UNK	Unknown	OD	CAHUEN	(34.1113, -118.3336)
4	12/18/2013	1.32E+08	12/18/2013	745	18	Southeast	1839	997	TRAFFIC D	UNK	Unknown	105	CROESU	(33.9406, -118.2338)
5	10/18/2013	1.32E+08	10/18/2013	1730	18	Southeast	1827	997	TRAFFIC D	UNK	Unknown	101	JUNIPER	(33.9449, -118.2332)
6	5/26/2013	1.31E+08	5/25/2013	2000	5	Harbor	507	440	THEFT PLA	UNK	Unknown	1300 W	SEPULVE	(33.8135, -118.2992)
7	5/24/2013	1.31E+08	5/22/2013	1145	12	77th Stree	1211	997	TRAFFIC D	UNK	Unknown	54T	CRENSH	(33.9931, -118.3308)
8	8/23/2014	1.4E+08	8/23/2014	2240	1	Central	111	310	BURGLARY	IC	Invest Cor	500 N	FIGUEROA	(34.0617, -118.2469)
9	8/23/2014	1.42E+08	8/23/2014	1337	20	Olympic	2023	901	VIOLATIO	IC	Invest Cor	300 S	SERRANO	(34.069, -118.3066)
10	8/22/2014	1.4E+08	8/23/2014	1945	1	Central	111	210	ROBBERY	IC	Invest Cor	900 N	HILL	(34.0644, -118.2387)
11	8/22/2014	1.41E+08	8/22/2014	825	9	Van Nuys	933	901	VIOLATIO	IC	Invest Cor	14600	CALVERT	(34.1817, -118.4509)
12	8/22/2014	1.41E+08	8/22/2014	800	9	Van Nuys	901	664	BUNCO, P	IC	Invest Cor	15300	SATICOY	(34.2085, -118.4662)
13	8/22/2014	1.41E+08	8/22/2014	735	13	Newton	1309	997	TRAFFIC D	IC	Invest Cor	MA	ENTERPF	(34.0228, -118.2325)
14	8/22/2014	1.42E+08	8/23/2014	2019	21	Topanga	2126	901	VIOLATIO	IC	Invest Cor	21100	SATICOY	(34.2083, -118.5929)
15	8/21/2014	1.42E+08	8/21/2014	1030	16	Foothill	1687	510	VEHICLE -	IC	Invest Cor	11200	VINEDALE	(34.2227, -118.3743)
16	8/20/2014	1.41E+08	8/21/2014	2000	11	Northeast	1171	331	THEFT FRC	IC	Invest Cor	3800 W	SUNSET	(34.091, -118.2788)
17	8/20/2014	1.41E+08	8/21/2014	2030	11	Northeast	1143	330	BURGLARY	IC	Invest Cor	2800	GRIFFITH P	(34.1075, -118.2732)
18	8/20/2014	1.42E+08	8/20/2014	1320	16	Foothill	1655	745	VANDALIS	IC	Invest Cor	10500	ART	(34.2486, -118.3575)
19	8/20/2014	1.42E+08	8/20/2014	1930	18	Southeast	1806	997	TRAFFIC D	IC	Invest Cor	AV	88TH	(33.9525, -118.2651)
20	8/20/2014	1.42E+08	8/20/2014	1530	21	Topanga	2189	943	CRUELTY T	IC	Invest Cor	5500	QUAKERTO	(34.1706, -118.5698)
21	8/19/2014	1.4E+08	8/22/2014	1430	3	Southwes	319	440	THEFT PLA	IC	Invest Cor	FIG	12TH	(34.0419, -118.2669)
22	8/19/2014	1.41E+08	8/21/2014	1500	8	West LA	823	440	THEFT PLA	IC	Invest Cor	15300	ANTIOCH	(34.047, -118.5259)
23	8/18/2014	1.41E+08	8/21/2014	1830	12	77th Stree	1283	624	BATTERY -	IC	Invest Cor	10300 S	WESTERN	(33.9432, -118.309)
24	8/18/2014	1.42E+08	8/18/2014	20	16	Foothill	1663	420	THEFT FRC	IC	Invest Cor	9000	RINCON	(34.2326, -118.4029)
25	8/18/2014	1.42E+08	8/19/2014	1800	19	Mission	1959	210	ROBBERY	IC	Invest Cor	SH	PIERCE	(34.2609, -118.4402)
26	8/18/2014	1.42E+08	8/22/2014	1200	19	Mission	1985	310	BURGLARY	IC	Invest Cor	8700	TOBIAS	(34.2279, -118.4516)
27	8/16/2014	1.41E+08	8/16/2014	1755	10	West Vall	1061	997	TRAFFIC D	IC	Invest Cor	VE	OAKDAL	(34.1724, -118.5649)
28	8/16/2014	1.41E+08	8/17/2014	2203	12	77th Stree	1211	900	VIOLATIO	IC	Invest Cor	5700	5TH	(33.9909, -118.3225)

Attributes	**Data Type**	**Number Of Distinct**	**Value**
Crm Cd Desc	Nominal	128 Categories	Burglary Robbery Vandalism Aggravated Assault Etc.
Date Occ	Date	Unlimited	8/23/14
Time Occ	Time	Unlimited	2200
Area Name	Nominal	21 Names	

2. Data Preprocessing

We performed the following preprocessing steps on the two datasets:

2.1 Data Cleaning

There are some missing values in some attributes such as last_occurance_date and incident_address in Denver dataset. However, we found that all attributes containing missing values are not of our key attributes. Therefore, we did not need to clean them. All key attributes in datasets were completed with cleaned values in both datasets. In addition, we did not found any noisy or inconsistent values in these attributes.

2.2 Data Reduction

Among the 19 attributes in Denver crimes dataset we just selected four of them.
The attribute “Is_Crime” indicates whether the instance belongs to a crime or accident. While we concern with crime information, we used the attribute “Is_Crime” to filter the instances and remove all the irrelevant ones.

In Los Angeles Crimes dataset we just selected four of them.

Code

```
losdata1<-subset(losdata,select = c(AREA.NAME,CrmCd.Desc,DATE.OCC,TIME.OCC))

denverdata1<-subset(denverdata,select =

c(OFFENSE_CATEGORY_ID,FIRST_OCCURRENCE_DATE,NEIGHBORHOOD_ID,IS_CR

I

ME)) denverdata2<-subset(denverdata1,IS_CRIME==1,select =

c(OFFENSE_CATEGORY_ID,FIRST_OCCURRENCE_DATE,NEIGHBORHOOD_ID,IS_CRI
ME))
```

2.3 Data Integration

First, to avoid different attribute naming, we unified the key attribute names for both crime datasets as follow: Crime_Type, Crime_Date, and Crime_Location. Crime_Location represents the neighborhood attribute for Denver dataset whereas the Area attribute for Los Angeles dataset.

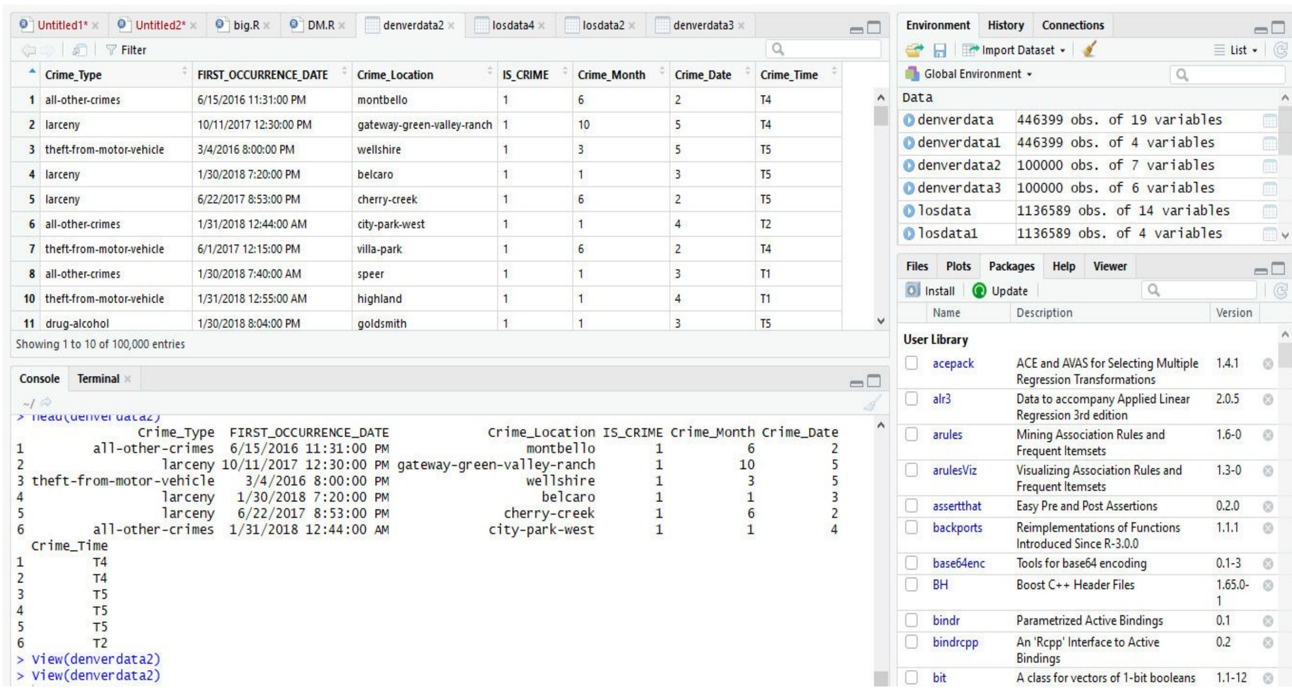
Code

```
colnames(losdata21)[2]<-"Crime_Type"     colnames(losdata21)

[1]<-"Crime_Location"     colnames(denverdata2)[1]<-"Crime_Type"

colnames(denverdata2)[3]<-"Crime_Location"     denverdata2<-
```

```
denverdata2[1:100000,] losdata2<-losdata21[1:100000,] func<-
function(a){

denverdata2["Crime_Month"]="Mon
th"  denverdata2["Crime_Day"]="Day"
denverdata2["Crime_Time"]="Time"
for(i in 1:nrow(denverdata2))
{     c=c()
   str<-denverdata2[["FIRST_OCCURRENCE_DATE"]][i]         str1<-
as.character(str)
     denverdata2[["Crime_Month"]][i]=strsplit(str1,split = '/')[[1]][1]
     denverdata2[["Crime_Day"]][i]=as.character((as.numeric(strsplit(str1,split = '/')[[1]][2])%
%7)+1)

c=strsplit(strsplit(as.character(denverdata2[["FIRST_OCCURRENCE_DATE"]][i]),split='/')[[1]]
[3],split
=" ")[[1]][c(2,3)]  if(c[2]=="PM"){
     d=strsplit(c[1],split=":")[[1]][1]
if(d<1){
         denverdata2[["Crime_Time"]][i]="T3";
       }
       else if(d>=1 && d<5){  denverdata2[["Crime_Time"]]
         [i]="T4";
       }
       else if(d>=5 && d<9){  denverdata2[["Crime_Time"]]
         [i]="T5";
```

```
        }
    else{
        denverdata2[["Crime_Time"]][i]="T6";
        }
      }
    else{
      e=strsplit(c[1],split=":")[[1]][1] if(d<1){
        denverdata2[["Crime_Time"]][i]="T6";
        }
        else if(d>=1 && d<5){ denverdata2[["Crime_Time"]]
        [i]="T1";
        }
        else if(d>=5 && d<9){ denverdata2[["Crime_Time"]]
        [i]="T2";
        }
    else{
        denverdata2[["Crime_Time"]][i]="T3";
        }
      }
    }
  return(denverdata2);
}
denverdata2<-func(denverdata2)
```

	Crime_Type	FIRST_OCCURRENCE_DATE	Crime_Location	IS_CRIME	Crime_Month	Crime_Date	Crime_Time
1	all-other-crimes	6/15/2016 11:31:00 PM	montbello	1	6	2	T4
2	larceny	10/11/2017 12:30:00 PM	gateway-green-valley-ranch	1	10	5	T4
3	theft-from-motor-vehicle	3/4/2016 8:00:00 PM	wellshire	1	3	5	T5
4	larceny	1/30/2018 7:20:00 PM	belcaro	1	1	3	T5
5	larceny	6/22/2017 8:53:00 PM	cherry-creek	1	6	2	T5
6	all-other-crimes	1/31/2018 12:44:00 AM	city-park-west	1	1	4	T2
7	theft-from-motor-vehicle	6/1/2017 12:15:00 PM	villa-park	1	6	2	T4
8	all-other-crimes	1/30/2018 7:40:00 AM	speer	1	1	3	T1
10	theft-from-motor-vehicle	1/31/2018 12:55:00 AM	highland	1	1	4	T1



losdata2[["Crime_Time"]]="Time"

losdata2[["Crime_Day"]]="Day"

losdata2[["Crime_Month"]]="Month"

func5<-function(a){ for(i in

1:nrow(losdata2)){

c<-losdata2[["TIME.OCC"]][i]%

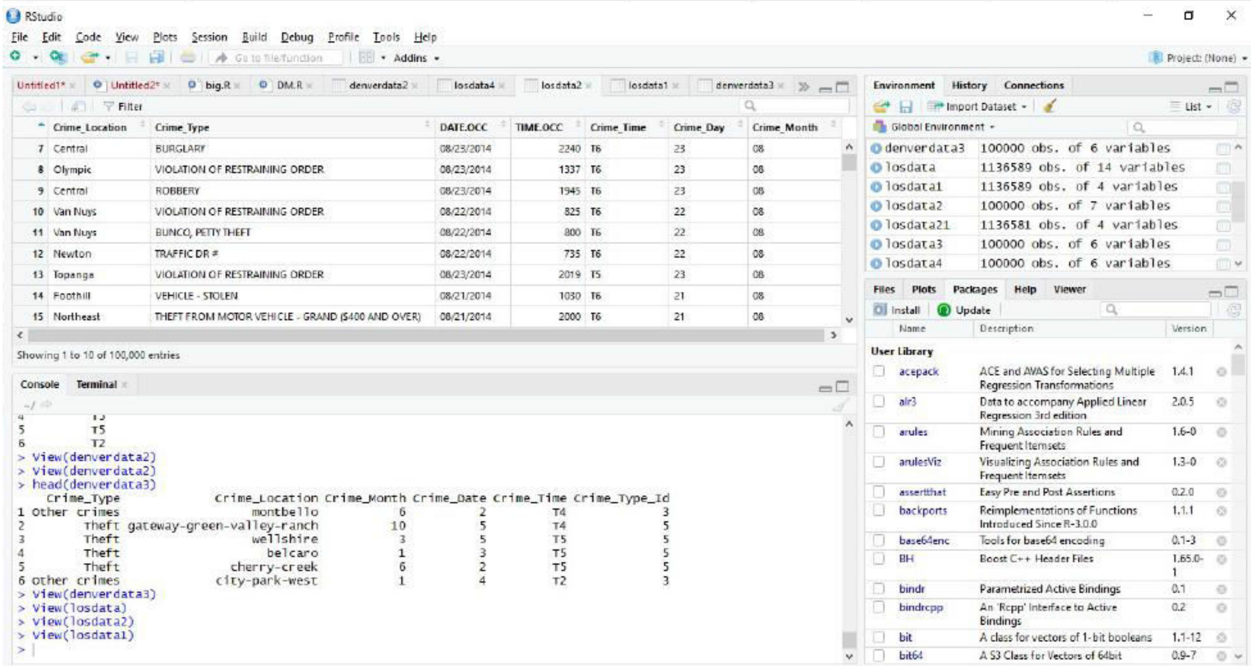
%100 str<-

losdata2[["DATE.OCC"]][i] str1<-

as.character(str)

```
    losdata2[["Crime_Month"]][i]=strsplit(str1,split="/")[[1]][1]
   losdata2[["Crime_Day"]][i]=(as.numeric(strsplit(str1,split="/")[[1]][2])%%7+1)
if(c<1){
     losdata2[["Crime_Time"]][i]="T6";
    }
    else if(c>=1 && c<5){ losdata2[["Crime_Time"]]
     [i]="T1";
    }
    else if(c>=5 && c<9){ losdata2[["Crime_Time"]]
     [i]="T2";
    }
    else if(c>=9 && c<13)
  {      losdata2[["Crime_Time"]][i]="T3";
    }
    else if(c>=13 && c<17)
  {      losdata2[["Crime_Time"]][i]="T4";
    }
    else if(c>=17 && c<21)
  {      losdata2[["Crime_Time"]][i]="T5";
    }
  else{
      losdata2[["Crime_Time"]][i]="T6";
    }
  }
   return(losdata2)
  }
  losdata2<-func5(losdata2)
```

	Crime_Location	Crime_Type	DATE.OCC	TIME.OCC	Crime_Time	Crime_Day	Crime_Month
7	Central	BURGLARY	08/23/2014	2240	T6	23	08
8	Olympic	VIOLATION OF RESTRAINING ORDER	08/23/2014	1337	T6	23	08
9	Central	ROBBERY	08/23/2014	1945	T6	23	08
10	Van Nuys	VIOLATION OF RESTRAINING ORDER	08/22/2014	825	T6	22	08
11	Van Nuys	BUNCO, PETTY THEFT	08/22/2014	800	T6	22	08
12	Newton	TRAFFIC DR #	08/22/2014	735	T6	22	08
13	Topanga	VIOLATION OF RESTRAINING ORDER	08/23/2014	2019	T5	23	08
14	Foothill	VEHICLE - STOLEN	08/21/2014	1030	T6	21	08
15	Northeast	THEFT FROM MOTOR VEHICLE - GRAND ($400 AND OVER)	08/21/2014	2000	T6	21	08



2.4 Data Transformation and Discretization

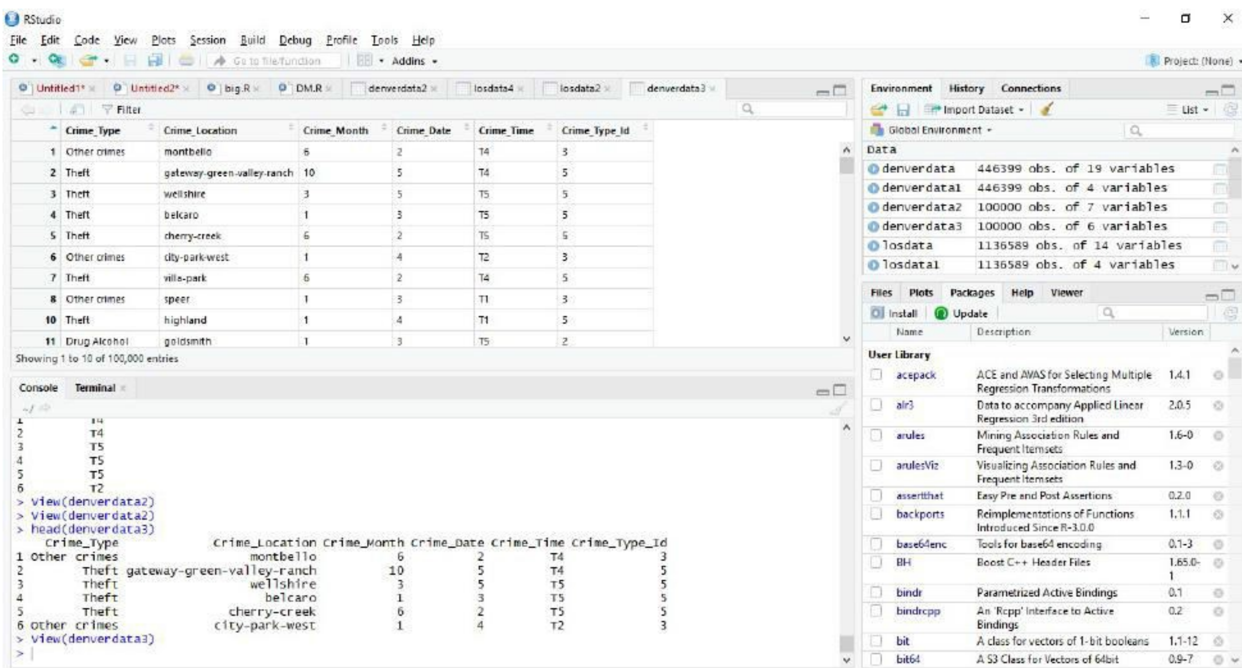
We applied data transformation to map Crime_Type attribute values to fall within the similar groups.

Code

```
denverdata3<-
subset(denverdata2,select=c(Crime_Type,Crime_Location,Crime_Month,Crime_Day,Crim
e_Time)) denverdata3["Crime_Type_Id"]="Crime_Type_Id" func1<-function(a){
  denverdata3$Crime_Type<-
as.character(denverdata3$Crime_Type)   for(i in
1:nrow(denverdata3)){
  if(denverdata3[["Crime_Type"]][i]=="aggravated-assault"     ||
denverdata3[["Crime_Type"]][i]=="sexual-assault")
```

```
{     denverdata3[["Crime_Type"]][i]="Assault";
denverdata3[["Crime_Type_Id"]][i]=1;
      }
                                       else          if(denverdata3[["Crime_Type"]][i]=="drug-alcohol")
  {                                              denverdata3[["Crime_Type"]][i]="Drug          Alcohol";
  denverdata3[["Crime_Type_Id"]][i]=2;

      }
    else if(denverdata3[["Crime_Type"]][i]=="all-other-crimes" ||
denverdata3[["Crime_Type"]][i]=="other-crimes-against-persons"){
denverdata3[["Crime_Type"]][i]="Other crimes";
denverdata3[["Crime_Type_Id"]][i]=3;
      }
      else if(denverdata3[["Crime_Type"]][i]=="public-disorder")
{     denverdata3[["Crime_Type"]][i]="Public Disorder";
denverdata3[["Crime_Type_Id"]][i]=4;
      }
      else if(denverdata3[["Crime_Type"]][i]=="white-collar-crime")
{     denverdata3[["Crime_Type"]][i]="White Collar Crime";
denverdata3[["Crime_Type_Id"]][i]=6;
      }
  else{
      denverdata3[["Crime_Type"]][i]="Theft"; denverdata3[["Crime_Type_Id"]][i]=5;
      }
    }
    return(denverdata3);
  }
  denverdata3<-func1(denverdata3)
```

	Crime_Type	Crime_Location	Crime_Month	Crime_Date	Crime_Time	Crime_Type_Id
1	Other crimes	montbello	6	2	T4	3
2	Theft	gateway-green-valley-ranch	10	5	T4	5
3	Theft	wellshire	3	5	T5	5
4	Theft	belcaro	1	3	T5	5
5	Theft	cherry-creek	6	2	T5	5
6	Other crimes	city-park-west	1	4	T2	3
7	Theft	villa-park	6	2	T4	5
8	Other crimes	speer	1	3	T1	3
10	Theft	highland	1	4	T1	5

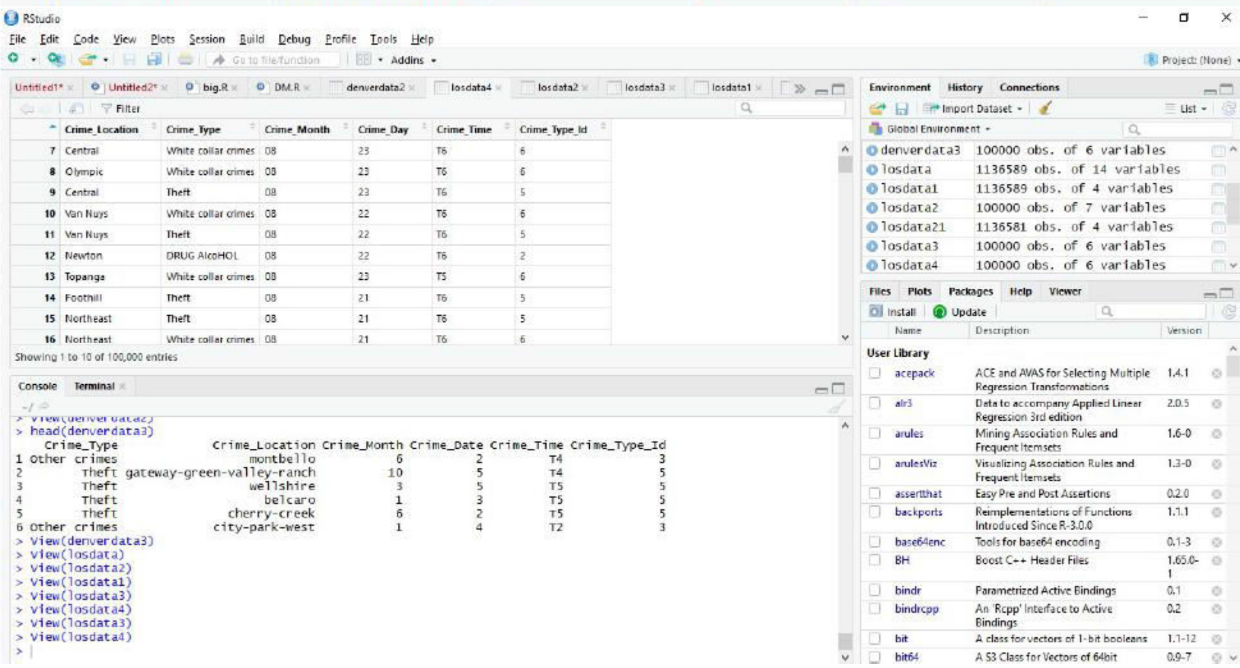



```
losdata3<-
subset(losdata2,select =
c("Crime_Location","Crime_Type","Crime_Month","Crime_Day","Crime_

Time")) losdata3["Crime_Type_Id"]="Type" func6<-function(a){   for(i in

1:nrow(losdata3)){

  losdata3$Crime_Type=as.character(losdata3$Crime_Type)

if((grepl("BURGALORY",losdata3[["Crime_Type"]][i])==TRUE) ||

(grepl("ROBBERY",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("THEFT",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("STOLEN",losdata3[["Crime_Type"]][i])==TRUE)
    || (grepl("SNATCHING",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("STEALING",losdata3[["Crime_Type"]][i])==TRUE) ||
```

```
(grepl("FELONY",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("PICKPOCKET",losdata3[["Crime_Type"]][i])==TRUE)
    || (grepl("BUNCO",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("EMBEZZLEMENT",losdata3[["Crime_Type"]][i])==TRUE) )
    {
  losdata3[["Crime_Type"]][i]="Theft"
losdata3[["Crime_Type_Id"]][i]=5
  }
 else
if((grepl("ASSAULT",losdata3[["Crime_Type"]][i])==TRUE)||
(grepl("BATTERY",losdata3[["Crime_Ty pe"]][i])==TRUE) ||
(grepl("VANDALISM",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("THREATS",losdata3[["Crime_Type"]][i])==TRUE)
        || (grepl("ABUSE",losdata3[["Crime_Type"]][i])==TRUE)
 ||(grepl("SHOTS",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("SCARE",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("SEXUAL",losdata3[["Crime_Type"]][i])==TRUE) ||
        (grepl("THROWING",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("COPULATION",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("EXTORTION",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("STALKING",losdata3[["Crime_Type"]][i])==TRUE)
       || (grepl("KIDNAPPING",losdata3[["Crime_Type"]]
[i])==TRUE) ||
 (grepl("ENDANGERMENT",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("ANNOYING",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("HOMICIDE",losdata3[["Crime_Type"]][i])==TRUE))
{    losdata3[["Crime_Type"]][i]="ASSAULT"
losdata3[["Crime_Type_Id"]][i]=1
  }
  else if((grepl("DR",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("BRANDISH",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("WEAPON",losdata3[["Crime_Type"]][i])==TRUE) ||
 (grepl("LEWD",losdata3[["Crime_Type"]][i])==TRUE))
```

```
{       losdata3[["Crime_Type"]][i]="DRUG AlcoHOL"
losdata3[["Crime_Type_Id"]][i]=2
      }
    else if((grepl("DISTURBING",losdata3[["Crime_Type"]]
[i])==TRUE) ||
  (grepl("ARSON",losdata3[["Crime_Type"]][i])==TRUE))
{       losdata3[["Crime_Type"]][i]="PUBLIC DISORDER"
losdata3[["Crime_Type_Id"]][i]=4
      }
    else if((grepl("TRESPASSING",losdata3[["Crime_Type"]]
[i])==TRUE) ||
(grepl("CRUELTY",losdata3[["Crime_Type"]][i])==TRUE) ||
(grepl("WITHOUT",losdata3[["Crime_Type"]][i])==TRUE))
{       losdata3[["Crime_Type"]][i]="Other crimes"
losdata3[["Crime_Type_Id"]][i]=3
      }
  else{
      losdata3[["Crime_Type"]][i]="White collar crimes"
losdata3[["Crime_Type_Id"]][i]=6
      }
  }
    return(losdata3);
  }
  losdata4<-func6(losdata3)
```

	Crime_Location	Crime_Type	Crime_Month	Crime_Day	Crime_Time	Crime_Type_Id
7	Central	White collar crimes	08	23	T6	6
8	Olympic	White collar crimes	08	23	T6	6
9	Central	Theft	08	23	T6	5
10	Van Nuys	White collar crimes	08	22	T6	6
11	Van Nuys	Theft	08	22	T6	5
12	Newton	DRUG AlcoHOL	08	22	T6	2
13	Topanga	White collar crimes	08	23	T5	6
14	Foothill	Theft	08	21	T6	5
15	Northeast	Theft	08	21	T6	5
16	Northeast	White collar crimes	08	21	T6	6



denverdata3$Crime_Time<-as.factor(denverdata3$Crime_Time)

denverdata3$Crime_Month<-as.numeric(denverdata3$Crime_Month)

denverdata3$Crime_Day<-as.numeric(denverdata3$Crime_Day)

denverdata3$Crime_Type_Id<- as.numeric(denverdata3$Crime_Type_Id)

denverdata3$Crime_Type<- as.factor(denverdata3$Crime_Type)

```
losdata4$Crime_Type<-as.factor(losdata4$Crime_Type)

losdata4$Crime_Month<-as.numeric(losdata4$Crime_Month)

losdata4$Crime_Day<-as.numeric(losdata4$Crime_Day)

losdata4$Crime_Time<-as.factor(losdata4$Crime_Time)

losdata4$Crime_Type_Id<-as.numeric(losdata4$Crime_Type_Id)
```

Attribute	**Number Of Distinct Values**	**Value**
Crime_Type	6	Assault Drug Alcohol Other crimes Public Disorder Theft White collar crime
Crime_Type_Id	6	1.Assault 2.Drug Alcohol 3.Other crimes 4.Public Disorder 5.Theft 6.White collar crime
Crime_Month	12	Months Name
Crime_Day	7	Days Of the Week
Crime_Time	6	T1: 1am to 4:59am T2: 5am to 8:59am T3: 9am to 12:59pm T4: 13pm to 16:59pm T5: 17pm to 20:59pm T6: 21pm to 0:59am
Crime_Location	Denver:78 LA:21	Same

3. Models Building

3.1 Apriori Algorithm

Apriori is one of the basic algorithms for mining frequent patterns. It scans the dataset to collect all itemsets that satisfy a predefined minimum support. We implemented this model using an open source tool.
We implemented the algorithm on location and time features and excluded the crime type feature.

DenverDataset Code

```
denverdata4<-subset(denverdata3,select=c("Crime_Location","Crime_Date","Crime_Time"))

write.csv(denverdata4,'denverdata4.csv')
```

	Crime_Location	Crime_Date	Crime_Time
1	montbello	2	T4
2	gateway-green-valley-ranch	5	T4
3	wellshire	5	T5
4	belcaro	3	T5
5	cherry-creek	2	T5
6	city-park-west	4	T2
7	villa-park	2	T4
8	speer	3	T1
10	highland	4	T1
11	goldsmith	3	T5
12	ruby-hill	1	T2
13	cory-merrill	4	T4
14	jefferson-park	4	T4
15	windsor	3	T4
16	highland	1	T1
17	goldsmith	3	T6
18	lincoln-park	2	T4
19	washington-park-west	5	T4
21	stapleton	3	T1
24	elyria-swansea	3	T4

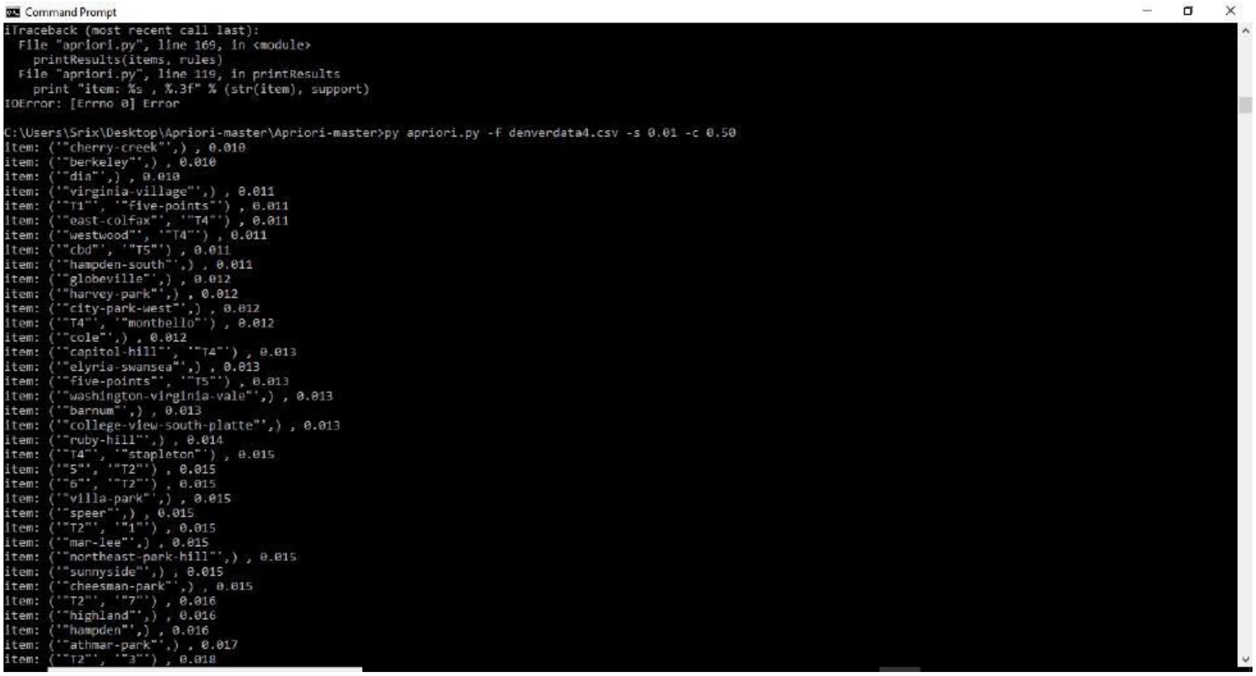
```
C:\Users\Srix\Desktop\Apriori-master\Apriori-master>py apriori.py -f denverdata4.csv -s 0.0012
-c 0.50
```

```
Apriori DenverApriori -s=0.0012 -c=0.50
item: ('harvey-park-south', 'T4', '4') , 0.001
item: ('T4', 'cherry-creek', '4') , 0.001
item: ('five-points', 'T2', '6') , 0.001
item: ('T5', 'athmar-park', 'Theft') , 0.001
item: ('T6', '3', 'cbd') , 0.001
item: ('cory-merrill', 'T4', '5') , 0.001
item: ('2', 'gateway-green-valley-ranch', 'Other crimes') , 0.001
item: ('5', '7', 'hampden-south') , 0.001
item: ('Public Disorder', '2', 'T3') , 0.001
item: ('5', '7', 'virginia-village') , 0.001
item: ('8', 'Drug Alcohol', '6') , 0.001
item: ('1', 'dia', 'Theft') , 0.001
item: ('lowry-field', '3', '5') , 0.001
item: ('elyria-swansea', '3', '2') , 0.001
item: ('villa-park', '4', 'Other crimes') , 0.001
item: ('2', 'Drug Alcohol', '6', '8') , 0.001
item: ('3', '2', 'gateway-green-valley-ranch', 'Other crimes') , 0.001
item: ('1', '4', '6', 'T1') , 0.001
item: ('athmar-park', '5', 'Theft', 'T5') , 0.001
item: ('3', '5', '4', 'T6') , 0.001
item: ('3', '2', 'five-points', 'T1') , 0.001
item: ('Public Disorder', '5', '7', 'T4') , 0.001
item: ('Public Disorder', '2', '4', 'T3') , 0.001
item: ('3', '4', 'villa-park', 'Other crimes') , 0.001
item: ('1', '3', 'westwood', 'T4') , 0.001
item: ('3', '4', 'T4', 'capitol-hill') , 0.001
item: ('1', '5', 'Theft', 'dia') , 0.001
item: ('3', '2', '9', 'T5') , 0.001
item: ('T4', 'Public Disorder', '5', '4', '7') , 0.001
item: ('7', 'regis') , 0.001
item: ('11', 'east-colfax') , 0.001
item: ('cory-merrill', '6') , 0.001
```

```
C:\Users\Srix\Desktop\Apriori-master\Apriori-master>py apriori.py -f denverdata4.csv -s 0.0012 -c 0.50
item: ('"7"', '"west-highland"') , 0.001
item: ('"university"', '"4"') , 0.001
item: ('"6"', '"sloan-lake"') , 0.001
item: ('"4"', '"windsor"') , 0.001
item: ('"1"', '"valverde"') , 0.001
item: ('"5"', '"auraria"') , 0.001
item: ('"college-view-south-platte"', '"T2"') , 0.001
item: ('"clayton"', '"1"') , 0.001
item: ('"T1"', '"marston"') , 0.001
item: ('"5"', '"T4"', '"west-colfax"') , 0.001
item: ('"lincoln-park"', '"T1"', '"2"') , 0.001
item: ('"7"', '"montbello"', '"T5"') , 0.001
item: ('"congress-park"', '"7"') , 0.001
item: ('"5"', '"montclair"') , 0.001
item: ('"5"', '"west-highland"') , 0.001
item: ('"lincoln-park"', '"T4"', '"6"') , 0.001
item: ('"T1"', '"1"', '"cbd"') , 0.001
item: ('"7"', '"T1"', '"cbd"') , 0.001
item: ('"regis"', '"T5"') , 0.001
item: ('"washington-virginia-vale"', '"T2"') , 0.001
item: ('"montclair"', '"2"') , 0.001
item: ('"5"', '"cherry-creek"') , 0.001
item: ('"east-colfax"', '"T6"') , 0.001
item: ('"overland"', '"2"') , 0.001
item: ('"hilltop"', '"T1"') , 0.001
item: ('"5"', '"sloan-lake"') , 0.001
item: ('"westwood"', '"T4"', '"1"') , 0.001
item: ('"T4"', '"kennedy"') , 0.001
item: ('"capitol-hill"', '"T5"', '"4"') , 0.001
item: ('"5"', '"T1"', '"five-points"') , 0.001
item: ('"montbello"', '"3"', '"T5"') , 0.001
item: ('"T4"', '"west-colfax"', '"4"') , 0.001
item: ('"7"', '"harvey-park-south"') , 0.001
```

C:\Users\Srix\Desktop\Apriori-master\Apriori-master>py apriori.py -f denverdata4.csv -s 0.01 -c 0.50

Apriori DenverApriori -s=0.001 -c=0.50
item: ('5', '9', 'T4', 'Other crimes') , 0.001
item: ('T1', '1', '3', '2', 'Other crimes') , 0.001
item: ('Theft', 'T2', '1', '5', '7') , 0.001
item: ('cbd', 'T4', '3', '7', 'Other crimes') , 0.001
item: ('T4', '3', '5', '9', 'Other crimes') , 0.001
item: ('skyland', '4') , 0.002
item: ('cherry-creek', 'T1') , 0.002
item: ('4', 'west-colfax', '6') , 0.002
item: ('6', '7', 'five-points') , 0.002
item: ('3', 'T5', 'mar-lee') , 0.002
item: ('valverde', '3', 'T4') , 0.002
item: ('5', 'mar-lee', '7') , 0.002
item: ('T4', '7', 'gateway-green-valley-ranch') , 0.002
item: ('north-capitol-hill', 'Theft', '6') , 0.002
item: ('9', '3', 'east-colfax') , 0.002
item: ('cheesman-park', '5', '6') , 0.002
item: ('valverde', '3', '5') , 0.002
item: ('1', '8', 'T2') , 0.002
item: ('1', 'civic-center', '5') , 0.002
item: ('10', '5', 'cbd') , 0.002
item: ('10', 'T1', '6') , 0.002
item: ('montbello', '3', 'Public Disorder') , 0.002
item: ('1', 'east-colfax', 'Theft') , 0.002
item: ('2', '7', 'five-points', 'T4') , 0.002
item: ('3', '4', 'five-points', 'T1') , 0.002
item: ('1', '2', 'five-points', 'T4') , 0.002
item: ('10', '3', '5', 'T5') , 0.002
item: ('Public Disorder', '4', 'montbello', '3') , 0.002
item: ('3', '12', 'Theft', 'T4') , 0.002
item: ('north-capitol-hill', '5', 'Theft', '6') , 0.002
item: ('1', '5', 'Theft', 'east-colfax') , 0.002
item: ('3', '4', 'stapleton', 'T4') , 0.002



Los Angeles Dataset

losdata5<-subset(losdata4,select = c("Crime_Location","Crime_Day","Crime_Time"));

write.csv(losdata5,'losdata5.csv')

	Crime_Location	Crime_Day	Crime_Time
7	Central	4	T6
8	Olympic	4	T6
9	Central	4	T6
10	Van Nuys	3	T6
11	Van Nuys	3	T6
12	Newton	3	T6
13	Topanga	3	T5
14	Foothill	1	T6
15	Northeast	1	T6
16	Northeast	1	T6
17	Foothill	7	T5
18	Southeast	7	T6
19	Topanga	7	T6
20	Southwest	2	T6
21	West LA	1	T6
22	77th Street	1	T6
23	Foothill	5	T5
24	Mission	6	T6

C:\Users\Srix\Desktop\Apriori-master\Apriori-master>py apriori.py -f losdata5.csv -s 0.0018 -c 0.50

Apriori LosDataApriori -s=0.0018 -c=0.50
item: ('Northeast', '5', '6') , 0.004
item: ('T6', '12', '5', 'Harbor') , 0.004
item: ('5', 'Theft', '9', 'Hollywood') , 0.004
item: ('Devonshire', '2', '5', 'T6') , 0.004
item: ('9', 'ASSAULT', '77th Street', 'T6') , 0.004
item: ('77th Street', 'T6', '1', 'ASSAULT', '9') , 0.004
item: ('12', '77th Street', 'Theft', 'T6') , 0.004
item: ('2', '77th Street', 'Theft', 'T6') , 0.004
item: ('Southeast', '5', 'T6', '8') , 0.004
item: ('12', 'Theft', 'T6', 'Pacific') , 0.004
item: ('12', 'T6', 'Pacific', '5', 'Theft') , 0.004
item: ('77th Street', 'T6', '2', '5', 'Theft') , 0.004
item: ('12', '77th Street', 'T6', '5', 'Theft') , 0.004
item: ('ASSAULT', '77th Street', '3') , 0.004
item: ('1', 'ASSAULT', '77th Street', '3') , 0.004
item: ('T6', '4', '6', '9') , 0.004
item: ('10', 'Topanga', 'Theft', 'T6') , 0.004
item: ('77th Street', '6', 'T6', '8') , 0.004
item: ('10', 'T6', 'Topanga', '5', 'Theft') , 0.004
item: ('11', '5', 'West Valley') , 0.004
item: ('11', '5', 'DRUG AlcoHOL') , 0.004
item: ('8', 'Newton', '5') , 0.004
item: ('9', 'DRUG AlcoHOL', '7', 'T6') , 0.004
item: ('2', 'Theft', 'T6', 'N Hollywood') , 0.004
item: ('3', '12', 'T6', '2') , 0.004
item: ('11', '2', 'DRUG AlcoHOL', '5') , 0.004
item: ('T6', '2', '7', '9') , 0.004
item: ('T6', 'N Hollywood', '2', '5', 'Theft') , 0.004
item: ('T6', '2', 'DRUG AlcoHOL', '7', '9') , 0.004
item: ('8', '5', 'Foothill') , 0.004
item: ('Rampart', '1', '5') , 0.004
item: ('12', '77th Street', '6') , 0.004

`

C:\Users\Srix\Desktop\Apriori-master\Apriori-master>py apriori.py -f losdata5.csv -s 0.01 -c 0.50

Apriori LosDataApriori -s=0.01 -c=0.50
item: ('11', 'Theft', '7') , 0.010
item: ('11', '5', '7') , 0.010
item: ('11', '5', 'Theft', '7') , 0.010
item: ('T6', 'Rampart', '2') , 0.010
item: ('Van Nuys', 'ASSAULT') , 0.010
item: ('9', 'Southeast') , 0.010
item: ('11', '2', '6') , 0.010
item: ('1', 'Van Nuys', 'ASSAULT') , 0.010
item: ('1', '2', 'T6', '12') , 0.010
item: ('1', '9', '6', 'T6') , 0.010
item: ('N Hollywood', 'DRUG AlcoHOL') , 0.010
item: ('N Hollywood', '2', 'DRUG AlcoHOL') , 0.010
item: ('T6', '8', 'N Hollywood') , 0.010
item: ('9', '5', '4') , 0.010
item: ('DRUG AlcoHOL', 'Pacific') , 0.010
item: ('2', 'DRUG AlcoHOL', 'Pacific') , 0.010
item: ('T6', 'Newton', '6') , 0.010
item: ('1', 'T6', 'West LA') , 0.010
item: ('T6', 'Harbor', '2') , 0.010
item: ('T6', 'N Hollywood', 'ASSAULT') , 0.010
item: ('1', 'ASSAULT', 'T6', 'N Hollywood') , 0.010
item: ('10', '5', '7') , 0.010
item: ('1', '11', 'Theft') , 0.010
item: ('10', 'Theft', '7') , 0.010
item: ('10', '5', 'Theft', '7') , 0.010
item: ('11', '1', '5', 'Theft') , 0.010
item: ('77th Street', '4') , 0.010
item: ('12', 'Mission') , 0.010
item: ('11', 'Pacific') , 0.010
item: ('8', '5', '7') , 0.010
item: ('8', '7', 'Theft') , 0.010
item: ('5', '7', 'Theft', '8') , 0.010

	A	B	C	D
1	**DenverData4 Sup-0.0012 Conf-0.50**	**DenverData4 Sup-0.01 Conf-0.50**	**LosData5 Sup- 0.0018 Conf -0.50**	**LosData5 Sup- 0.01 Conf -0.50**
2	item: ("'7'", "'west-highland'") , 0.001	item: ("'cherry-creek'",) , 0.010	item: ("'Harbor'", "'T5'") , 0.002	item: ("'77th Street'", "'4'") , 0.010
3	item: ("'university'", "'4'") , 0.001	item: ("'berkeley'",) , 0.010	item: ("'T4'", "'Foothill'") , 0.002	item: ("'2'", "'N Hollywood'") , 0.011
4	item: ("'6'", "'sloan-lake'") , 0.001	item: ("'dia'",) , 0.010	item: ("'T5'", "'Southeast'") , 0.002	item: ("'Southwest'", "'2'") , 0.011
5	item: ("'4'", "'windsor'") , 0.001	item: ("'virginia-village'",) , 0.011	item: ("'Van Nuys'", "'T5'") , 0.002	item: ("'77th Street'", "'3'") , 0.011
6	item: ("'1'", "'valverde'") , 0.001	item: ("'T1'", "'five-points'") , 0.011	item: ("'7'", "'T1'") , 0.002	item: ("'77th Street'", "'2'") , 0.012
7	item: ("'5'", "'auraria'") , 0.001	item: ("'east-colfax'", "'T4'") , 0.011	item: ("'Wilshire'", "'T5'") , 0.002	item: ("'T1'",) , 0.016
8	item: ("'college-view-south-platte'", "'T2'") , 0.001	item: ("'westwood'", "'T4'") , 0.011	item: ("'West Valley'", "'T5'") , 0.002	item: ("'T2'",) , 0.023
9	item: ("'clayton'", "'1'") , 0.001	item: ("'cbd'", "'T5'") , 0.011	item: ("'Newton'", "'T5'") , 0.002	item: ("'Hollenbeck'", "'T6'") , 0.030
10	item: ("'T1'", "'marston'") , 0.001	item: ("'hampden-south'",) , 0.011	item: ("'T3'", "'Southeast'") , 0.002	item: ("'Central'", "'T6'") , 0.032
11	item: ("'5'", "'T4'", "'west-colfax'") , 0.001	item: ("'globeville'",) , 0.012	item: ("'Hollywood'", "'T5'") , 0.002	item: ("'Foothill'", "'T6'") , 0.032
12	item: ("'lincoln-park'", "'T1'", "'2'") , 0.001	item: ("'harvey-park'",) , 0.012	item: ("'Pacific'", "'T5'") , 0.002	item: ("'T3'",) , 0.033
13	item: ("'7'", "'montbello'", "'T5'") , 0.001	item: ("'city-park-west'",) , 0.012	item: ("'T3'", "'Southwest'") , 0.002	item: ("'Rampart'", "'T6'") , 0.035
14	item: ("'congress-park'", "'7'") , 0.001	item: ("'T4'", "'montbello'") , 0.012	item: ("'T4'", "'Hollenbeck'") , 0.002	item: ("'Hollenbeck'",) , 0.036
15	item: ("'5'", "'montclair'") , 0.001	item: ("'cole'",) , 0.012	item: ("'T4'", "'Rampart'") , 0.002	item: ("'Topanga'", "'T6'") , 0.036
16	item: ("'5'", "'west-highland'") , 0.001	item: ("'capitol-hill'", "'T4'") , 0.013	item: ("'T1'", "'1'") , 0.002	item: ("'T6'", "'Newton'") , 0.036
17	item: ("'lincoln-park'", "'T4'", "'6'") , 0.001	item: ("'elyria-swansea'",) , 0.013	item: ("'T1'", "'4'") , 0.002	item: ("'Wilshire'", "'T6'") , 0.036
18	item: ("'T1'", "'1'", "'cbd'") , 0.001	item: ("'five-points'", "'T5'") , 0.013	item: ("'Central'", "'T5'") , 0.002	item: ("'Hollywood'", "'T6'") , 0.036
19	item: ("'7'", "'T1'", "'cbd'") , 0.001	item: ("'washington-virginia-vale'",) , 0.013	item: ("'Mission'", "'T5'") , 0.002	item: ("'T6'", "'West Valley'") , 0.037
20	item: ("'regis'", "'T5'") , 0.001	item: ("'barnum'",) , 0.013	item: ("'T1'", "'3'") , 0.002	item: ("'Olympic'", "'T6'") , 0.037
21	item: ("'washington-virginia-vale'", "'T2'") , 0.001	item: ("'college-view-south-platte'",) , 0.013	item: ("'Rampart'", "'T5'") , 0.002	item: ("'Harbor'", "'T6'") , 0.037
22	item: ("'montclair'", "'2'") , 0.001	item: ("'ruby-hill'",) , 0.014	item: ("'T4'", "'Harbor'") , 0.002	item: ("'West LA'", "'T6'") , 0.038
23	item: ("'5'", "'cherry-creek'") , 0.001	item: ("'T4'", "'stapleton'") , 0.015	item: ("'T4'", "'Newton'") , 0.002	item: ("'Foothill'",) , 0.038
24	item: ("'east-colfax'", "'T6'") , 0.001	item: ("'5'", "'T2'") , 0.015	item: ("'T4'", "'Van Nuys'") , 0.002	item: ("'Devonshire'", "'T6'") , 0.039
25	item: ("'overland'", "'2'") , 0.001	item: ("'6'", "'T2'") , 0.015	item: ("'T4'", "'Topanga'") , 0.002	item: ("'Van Nuys'", "'T6'") , 0.040
26	item: ("'hilltop'", "'T1'") , 0.001	item: ("'villa-park'",) , 0.015	item: ("'77th Street'", "'T3'") , 0.002	item: ("'Central'",) , 0.040
27	item: ("'5'", "'sloan-lake'") , 0.001	item: ("'speer'",) , 0.015	item: ("'N Hollywood'", "'T5'") , 0.002	item: ("'T6'", "'Southeast'") , 0.041
28	item: ("'westwood'", "'T4'", "'1'") , 0.001	item: ("'T2'", "'1'") , 0.015	item: ("'T4'", "'Wilshire'") , 0.003	item: ("'Rampart'",) , 0.043

3.2 NaiveBayes

ind<-sample(2,nrow(denverdata3),replace=TRUE,prob=c(0.8,0.2))

ind1<-sample(2,nrow(losdata4),replace=TRUE,prob=c(0.8,0.2))

denvernaive<subset(denverdata3,select=c(Crime_Type,Crime_Location

,Crime_Month,Crime_Day,Crime_Time))

losnaive<subset(losdata4,select=c(Crime_Type,Crime_Location,Crime_Month,Crime_Day,Crim

e_Time))

train1<-denvernaive[ind==1,]

test1<-denvernaive[ind==2,]

train2<-losnaive[ind1==1,]

test2<-losnaive[ind1==2,]

library(e1071)

model1<naiveBayes(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day),

data=train1)

model2<-naiveBayes(Crime_Type~.,data=train2)

naive1<-confusionMatrix(pred1,test1$Crime_Type)

```
                    Reference
Prediction           Assault Drug Alcohol Other crimes Public Disorder Theft White Collar Crime
  Assault                  0            0            0               0     0                  0
  Drug Alcohol             0            0            0               0     0                  0
  Other crimes            40          177          439              94   357                 10
  Public Disorder          0            0            0               0     0                  0
  Theft                  556         1209         3841            2111 11046                229
  White Collar Crime       0            0            0               0     0                  0

$overall
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull AccuracyPValue  McnemarPValue 
    0.57113730     0.05248087     0.56426215     0.57799188     0.56705953     0.12301471            NaN 

$byClass
                          Sensitivity Specificity Pos Pred Value Neg Pred Value Precision    Recall        F1 Prevalence Detection Rate Detection Prevalence
Class: Assault              0.0000000  1.00000000            NaN      0.9703615        NA 0.0000000        NA 0.02963847     0.00000000           0.00000000
Class: Drug Alcohol         0.0000000  1.00000000            NaN      0.9310756        NA 0.0000000        NA 0.06892436     0.00000000           0.00000000
Class: Other crimes         0.1025701  0.95716722      0.3930170      0.7977570 0.3930170 0.1025701 0.1626830 0.21284002     0.02183102           0.05554727
Class: Public Disorder      0.0000000  1.00000000            NaN      0.8903476        NA 0.0000000        NA 0.10965239     0.00000000           0.00000000
Class: Theft                0.9686924  0.08729612      0.5816133      0.6803939 0.5816133 0.9686924 0.7268301 0.56705953     0.54930628           0.94445273
Class: White Collar Crime   0.0000000  1.00000000            NaN      0.9881148        NA 0.0000000        NA 0.01188523     0.00000000           0.00000000
                          Balanced Accuracy
Class: Assault                    0.5000000
Class: Drug Alcohol               0.5000000
Class: Other crimes               0.5298687
Class: Public Disorder            0.5000000
Class: Theft                      0.5279943
Class: White Collar Crime         0.5000000

$mode
[1] "sens_spec"

$dots
list()
```

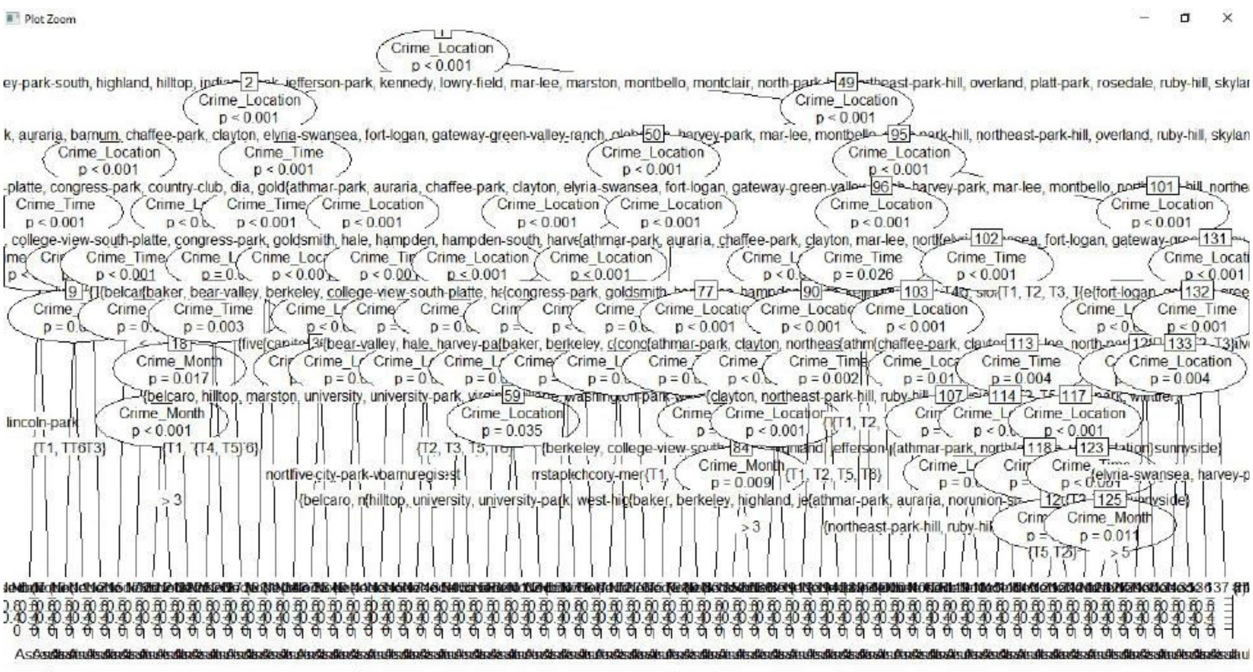
naive2<-confusionMatrix(pred2,test2$Crime_Type)

```
                    Reference
Prediction           ASSAULT DRUG AlcoHOL Other crimes PUBLIC DISORDER Theft White collar crimes
  ASSAULT                782          426           14               7   648                  276
  DRUG AlcoHOL           277          374           12               4   314                  179
  Other crimes             0            0            0               0     0                    0
  PUBLIC DISORDER          0            0            0               0     0                    0
  Theft                 3788         2683          132              35  6807                 3194
  White collar crimes      0            0            0               0     0                    0

$overall
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull AccuracyPValue  McnemarPValue
   0.399107859    0.059501575    0.392302773    0.405942357    0.389384523    0.002509505            NaN

$byClass
                           Sensitivity Specificity Pos Pred Value Neg Pred Value Precision    Recall        F1  Prevalence Detection Rate Detection Prevalence
Class: ASSAULT               0.1613369   0.9092354      0.3632141      0.7716164 0.3632141 0.1613369 0.2234286 0.242933039     0.03919407           0.10790898
Class: DRUG AlcoHOL          0.1073787   0.9522740      0.3224138      0.8345573 0.3224138 0.1073787 0.1611027 0.174568966     0.01874499           0.05813953
Class: Other crimes          0.0000000   1.0000000            NaN      0.9920810        NA 0.0000000        NA 0.007919006     0.00000000           0.00000000
Class: PUBLIC DISORDER       0.0000000   1.0000000            NaN      0.9976945        NA 0.0000000        NA 0.002305533     0.00000000           0.00000000
Class: Theft                 0.8761745   0.1929738      0.4090991      0.7096287 0.4090991 0.8761745 0.5577679 0.389384523     0.34116881           0.83395148
Class: White collar crimes   0.0000000   1.0000000            NaN      0.8171111        NA 0.0000000        NA 0.182888933     0.00000000           0.00000000
                           Balanced Accuracy
Class: ASSAULT                     0.5352861
Class: DRUG AlcoHOL                0.5298263
Class: Other crimes                0.5000000
Class: PUBLIC DISORDER             0.5000000
Class: Theft                       0.5345742
Class: White collar crimes         0.5000000

$mode
[1] "sens_spec"

$dots
list()
```

3.3 Decision Tree DenverDataset

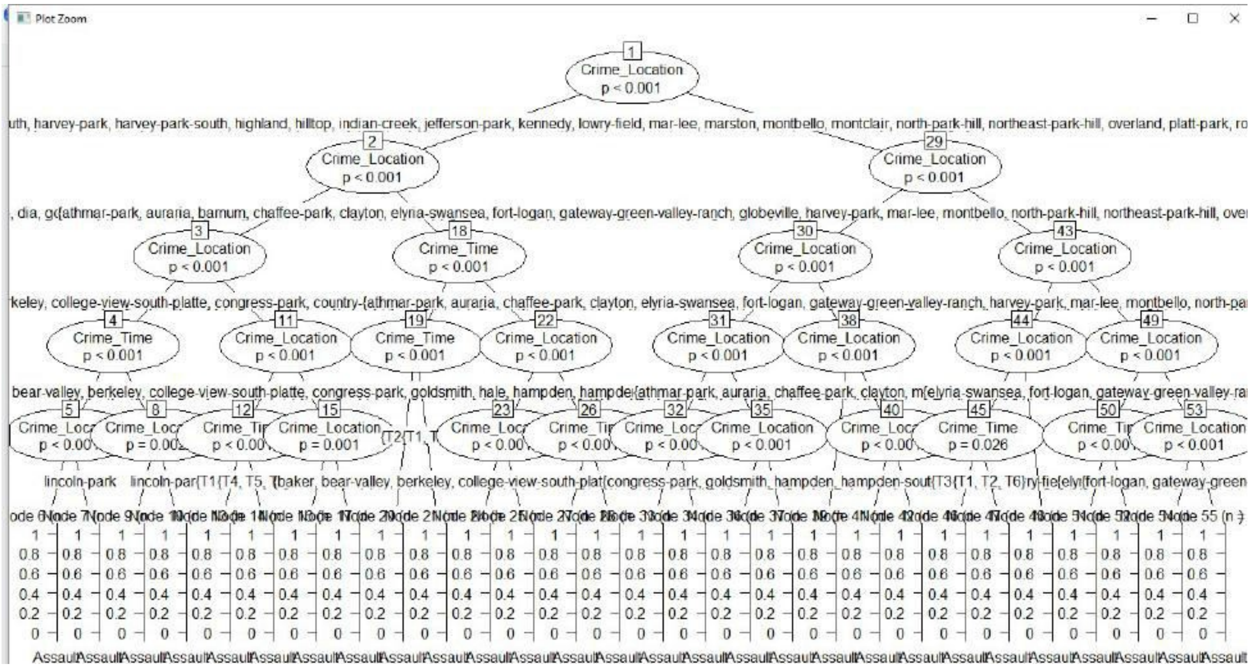
When n=10

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 10))



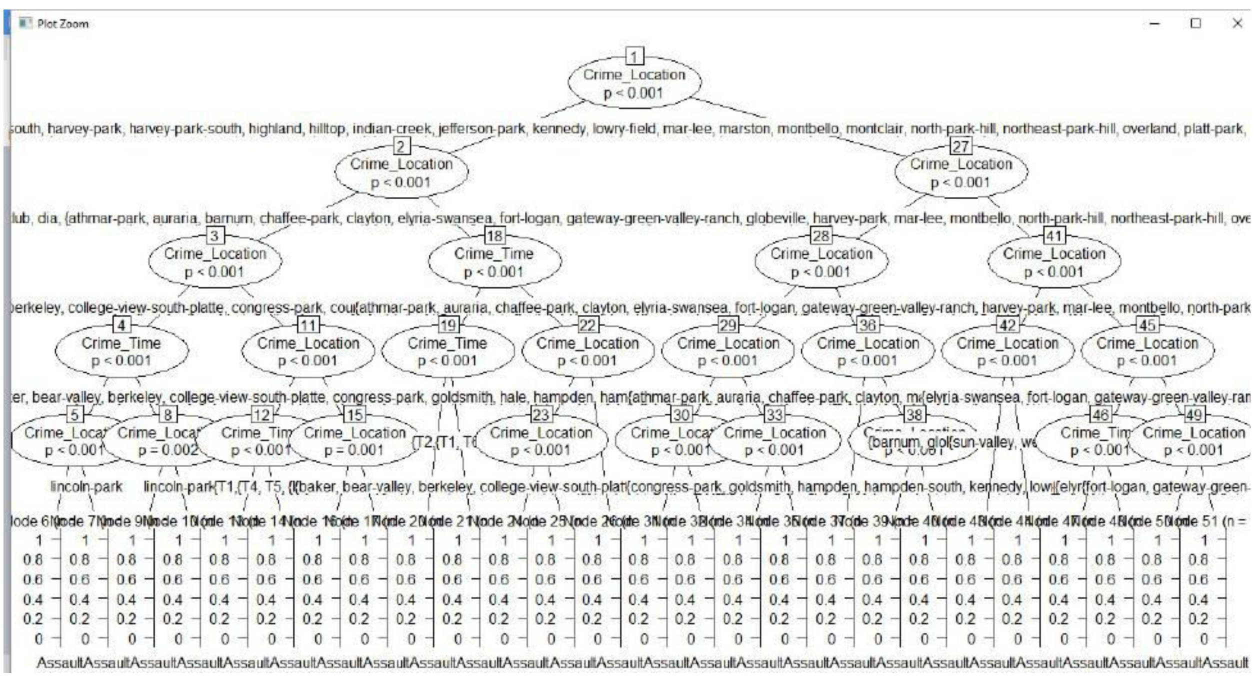
When n=5

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 5))



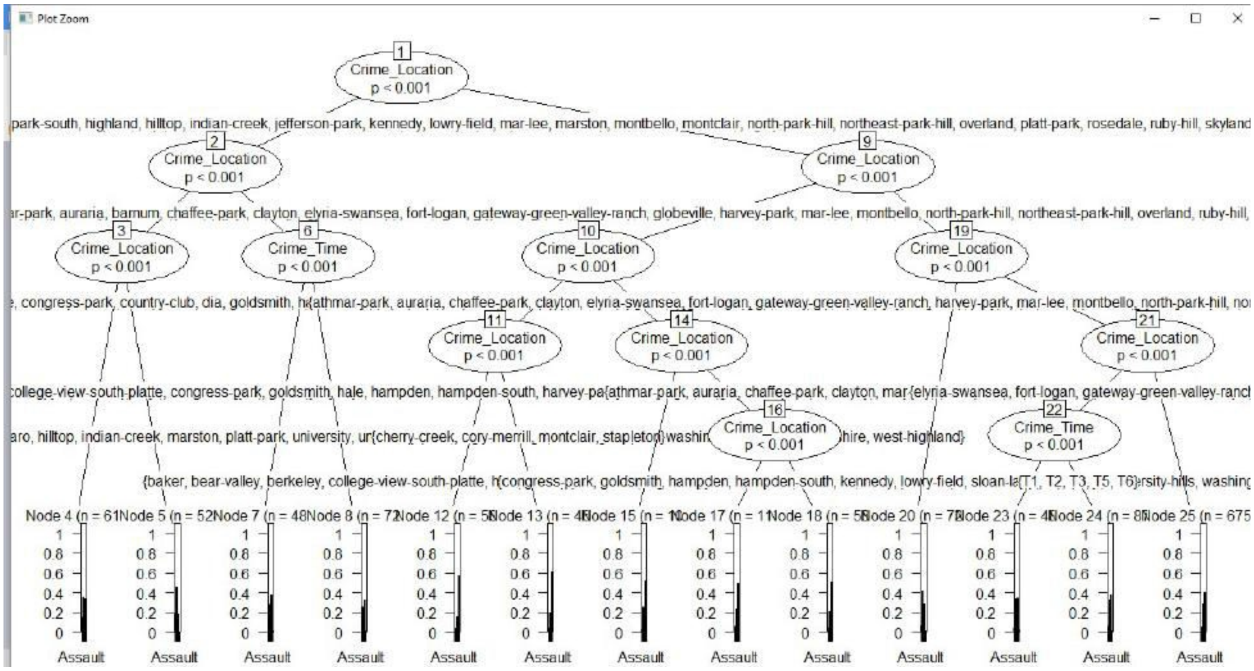
When MinimumSplit = 1000

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 5,mincriterion=0.99,minsplit=1000))



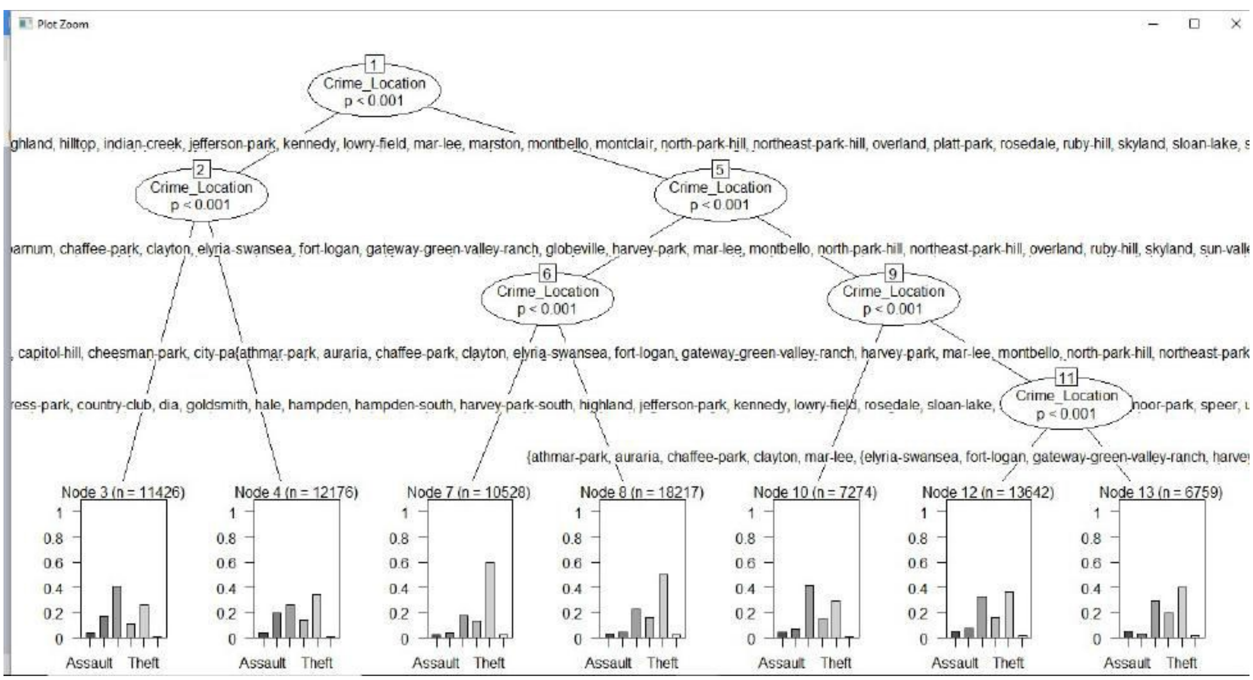
When MinimumSplit = 10000

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 5,mincriterion=0.99,minsplit=10000))

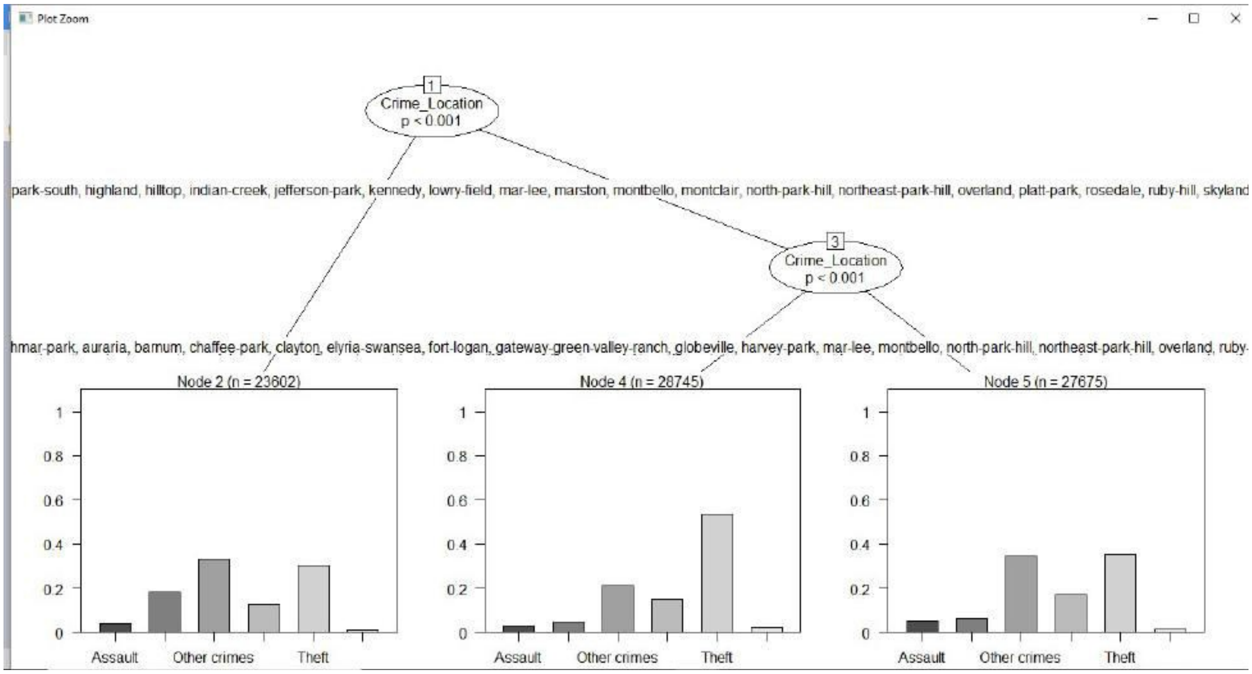


When minimumSplit = 20000

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 5,mincriterion=0.99,minsplit=20000))

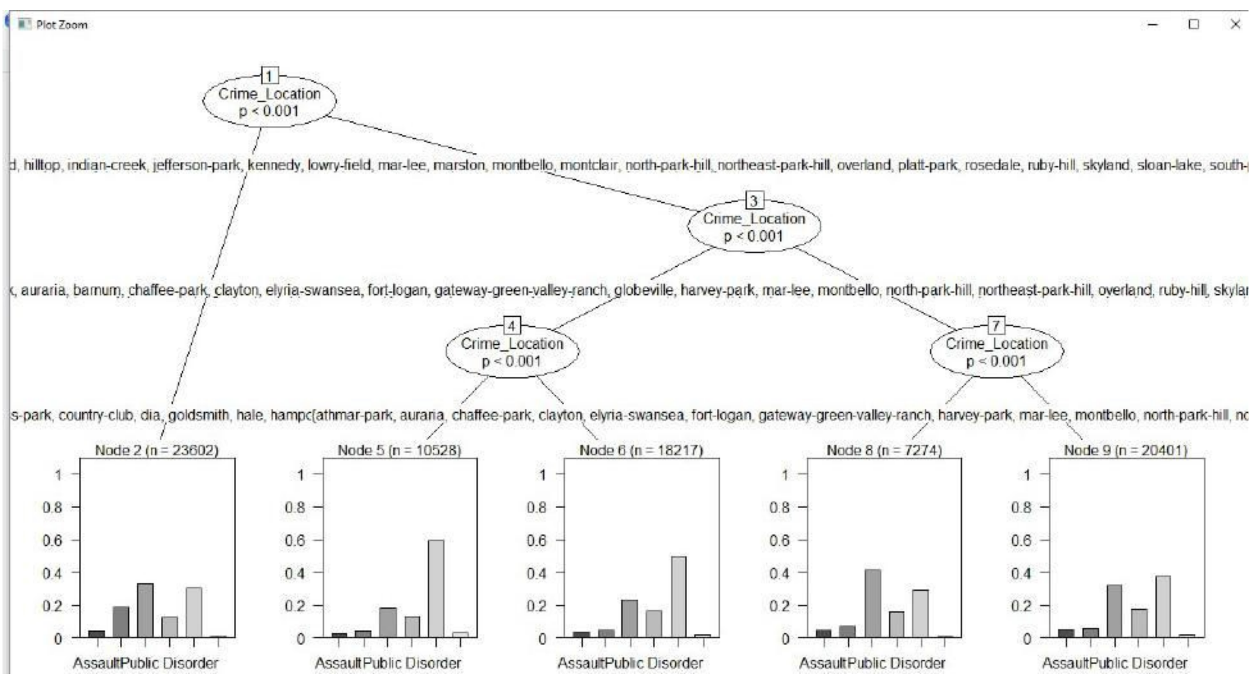


When Minimum Split = 30000 model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 5,mincriterion=0.99,minsplit=30000))



When MinimumSplit = 25000

model3<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day), data=train1,controls = ctree_control(maxdepth = 6,mincriterion=0.99,minsplit=25000))

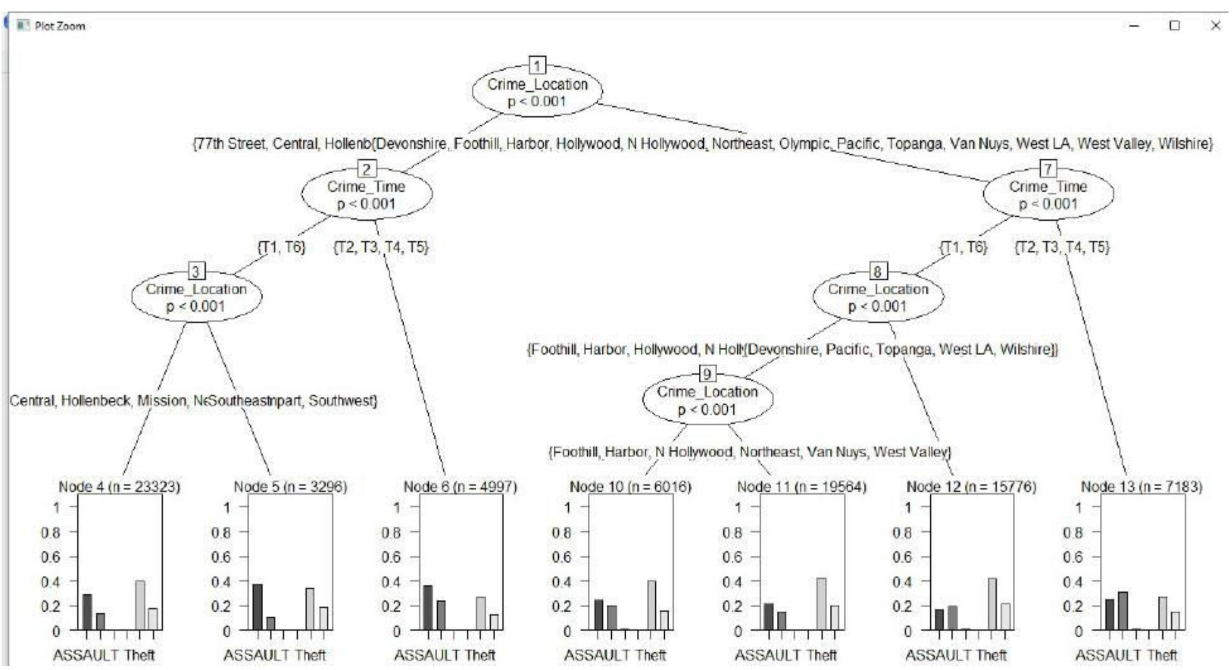


accuracy3<-confusionMatrix(predict(model3,test1[,-1]),test1$Crime_Type)

```
                    Reference
Prediction           Assault Drug Alcohol Other crimes Public Disorder Theft White Collar Crime
  Assault                  0            0            0               0     0                  0
  Drug Alcohol             0            0            0               0     0                  0
  Other crimes             0            0            0               0     0                  0
  Public Disorder          0            0            0               0     0                  0
  Theft                  596         1386         4280            2205 11403                239
  White Collar Crime       0            0            0               0     0                  0

$overall
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull AccuracyPValue  McnemarPValue 
     0.5670595      0.0000000      0.5601771      0.5739226      0.5670595      0.5029658            NaN 

$byClass
                          Sensitivity Specificity Pos Pred Value Neg Pred Value Precision Recall        F1 Prevalence
Class: Assault                      0           1            NaN      0.9703615        NA      0        NA 0.02963847
Class: Drug Alcohol                 0           1            NaN      0.9310756        NA      0        NA 0.06892436
Class: Other crimes                 0           1            NaN      0.7871600        NA      0        NA 0.21284002
Class: Public Disorder              0           1            NaN      0.8903476        NA      0        NA 0.10965239
Class: Theft                        1           0      0.5670595            NaN 0.5670595      1 0.7237243 0.56705953
Class: White Collar Crime           0           1            NaN      0.9881148        NA      0        NA 0.01188523
                          Detection Rate Detection Prevalence Balanced Accuracy
Class: Assault                 0.0000000                    0               0.5
Class: Drug Alcohol            0.0000000                    0               0.5
Class: Other crimes            0.0000000                    0               0.5
Class: Public Disorder         0.0000000                    0               0.5
Class: Theft                   0.5670595                    1               0.5
Class: White Collar Crime      0.0000000                    0               0.5

$mode
[1] "sens_spec"

$dots
list()
```

Los Angeles Dataset

model4<-ctree(Crime_Type~(Crime_Location+Crime_Month+Crime_Time+Crime_Day),data=train2,controls = ctree_control(maxdepth = 6,mincriterion = 0.99,minsplit = 25000))



accuracy4<-confusionMatrix(predict(model4,test2[,-1]),test2$Crime_Type)

```
                    Reference
Prediction            ASSAULT DRUG AlcoHOL Other crimes PUBLIC DISORDER Theft White collar crimes
  ASSAULT                 608          483           14               6   519                 225
  DRUG AlcoHOL            277          398           13               4   359                 197
  Other crimes              0            0            0               0     0                   0
  PUBLIC DISORDER           0            0            0               0     0                   0
  Theft                  3962         2602          131              36  6891                3227
  White collar crimes       0            0            0               0     0                   0

$overall
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull AccuracyPValue  McnemarPValue
    0.39579992     0.05248527     0.38900493     0.40262528     0.38938452     0.03216525            NaN

$byClass
                           Sensitivity Specificity Pos Pred Value Neg Pred Value Precision    Recall        F1  Prevalence Detection Rate Detection Prevalence
Class: ASSAULT               0.1254384   0.9174446      0.3277628      0.7657623 0.3277628 0.1254384 0.1814384 0.242933039     0.03047314           0.09297314
Class: DRUG AlcoHOL          0.1142693   0.9483879      0.3189103      0.8350620 0.3189103 0.1142693 0.1682520 0.174568966     0.01994787           0.06255012
Class: Other crimes          0.0000000   1.0000000            NaN      0.9920810        NA 0.0000000        NA 0.007919006     0.00000000           0.00000000
Class: PUBLIC DISORDER       0.0000000   1.0000000            NaN      0.9976945        NA 0.0000000        NA 0.002305533     0.00000000           0.00000000
Class: Theft                 0.8869867   0.1826315      0.4089857      0.7170480 0.4089857 0.8869867 0.5598343 0.389384523     0.34537891           0.84447674
Class: White collar crimes   0.0000000   1.0000000            NaN      0.8171111        NA 0.0000000        NA 0.182888933     0.00000000           0.00000000
                           Balanced Accuracy
Class: ASSAULT                     0.5214415
Class: DRUG AlcoHOL                0.5313286
Class: Other crimes                0.5000000
Class: PUBLIC DISORDER             0.5000000
Class: Theft                       0.5348091
Class: White collar crimes         0.5000000

$mode
[1] "sens_spec"

$dots
list()
```

#Random Forest

library(randomForest)

denverdata3$Crime_Time<-as.numeric(as.character(denverdata3$Crime_Time))

denverdata3$Crime_Type<-as.numeric(as.character(denverdata3$Crime_Type))

denverdata3$Crime_Location<-as.numeric(as.character(denverdata3$Crime_Location))

losdata4$Crime_Type<-as.numeric(as.character(losdata4$Crime_Type))

losdata4$Crime_Time<-as.numeric(as.character(losdata4$Crime_Time))

losdata4$Crime_Location<-as.numeric(as.character(losdata4$Crime_Location))

ind<-sample(2,nrow(denverdata3),replace=TRUE,prob=c(0.8,0.2))

ind1<-sample(2,nrow(losdata4),replace=TRUE,prob=c(0.8,0.2))

denvernaive1<-
subset(denverdata3,select=c(Crime_Type,Crime_Location,Crime_Month,Crime_Day,Crime_Ti
me))

losnaive1<-
subset(losdata4,select=c(Crime_Type,Crime_Location,Crime_Month,Crime_Day,Crime_Time))

train11<-denvernaive[ind==1,]

test11<-denvernaive[ind==2,]

train21<-losnaive[ind1==1,]

test21<-losnaive[ind1==2,]

modelrandom1<-
randomForest(Crime_Type~.,data=train21,ntree=100,mtype=6,importance=TRUE)

predict2=predict(modelrandom1,test21,type="class")

accuracy5<-confusionMatrix(predict2,test21$Crime_Type)

```
                    Reference
Prediction           ASSAULT DRUG AlcoHOL Other crimes PUBLIC DISORDER Theft White collar crimes
  ASSAULT                731          414           12               4   641                 307
  DRUG AlcoHOL           233          292            7               4   261                 117
  Other crimes             0            1            0               0     0                   0
  PUBLIC DISORDER          0            0            0               0     0                   0
  Theft                 3927         2784          115              36  6850                3142
  White collar crimes     51           64            2               1    71                  51

$overall
      Accuracy          Kappa  AccuracyLower  AccuracyUpper   AccuracyNull AccuracyPValue  McnemarPValue
    0.39387613     0.05042107     0.38711527     0.40066767     0.38885575     0.07312754            NaN

$byClass
                           Sensitivity Specificity Pos Pred Value Neg Pred Value Precision     Recall         F1  Prevalence Detection Rate Detection Prevalence
Class: ASSAULT              0.14791582   0.9091987      0.3466098      0.7661725 0.3466098 0.14791582 0.20734648 0.245650661    0.036335620         1.048315e-01
Class: DRUG AlcoHOL         0.08213783   0.9624464      0.3194748      0.8300875 0.3194748 0.08213783 0.13067800 0.176707426    0.014514365         4.543195e-02
Class: Other crimes         0.00000000   0.9999500      0.0000000      0.9932395 0.0000000 0.00000000        NaN 0.006760115    0.000000000         4.970673e-05
Class: PUBLIC DISORDER      0.00000000   1.0000000            NaN      0.9977632        NA 0.00000000         NA 0.002236803    0.000000000         0.000000e+00
Class: Theft                0.87562316   0.1863359      0.4064317      0.7018995 0.4064317 0.87562316 0.55517283 0.388855751    0.340491102         8.377572e-01
Class: White collar crimes  0.01410008   0.9885461      0.2125000      0.8206057 0.2125000 0.01410008 0.02644542 0.179789243    0.002535043         1.192962e-02
                           Balanced Accuracy
Class: ASSAULT                     0.5285573
Class: DRUG AlcoHOL                0.5222921
Class: Other crimes                0.4999750
Class: PUBLIC DISORDER             0.5000000
Class: Theft                       0.5309795
Class: White collar crimes         0.5013231

$mode
[1] "sens_spec"

$dots
list()
```