BIG DATA ECOSYSTEMS

MSBA - 321

CHICAGO POLICE DEPARTMENT- ILLINOIS UNIFORM CRIME REPORTING (IUCR) CODES

BY

PORRSELVI

CHICAGO POLICE DEPARTMENT-ILLINOIS UNIFORM CRIME REPORTING (IUCR) CODES

INTRODUCTION

Illinois Uniform Crime Reporting (IUCR) codes are four digit codes that law enforcement agencies use to classify criminal incidents when taking individual reports. These codes are also used to aggregate types of cases for statistical purposes. In Illinois, the Illinois State Police establish IUCR codes, but the agencies can add codes to suit their individual needs. The Chicago Police Department currently uses more than 350 IUCR codes to classify criminal offenses, divided into "Index" and "Non-Index" offenses.

USING HIVE TECHNOLOGY, I HAVE ANALYZED CHICAGO POLICE DEPARTMENT DATA OF ILLINOIS UNIFORM CRIME REPORTING (IUCR).

HIVE QUERY LANGUAGE:

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy.

The term 'Big Data' is used for collections of large datasets that include huge volume, high velocity, and a variety of data that is increasing day by day. Using traditional data management systems, it is difficult to process Big Data. Therefore, the Apache Software Foundation

introduced a framework called Hadoop to solve Big Data management and processing challenges.

Hadoop:

Hadoop is an open-source framework to store and process Big Data in a distributed environment. It contains two modules, one is MapReduce and another is Hadoop Distributed File System (HDFS).

- **MapReduce:** It is a parallel programming model for processing large amounts of structured, semi-structured, and unstructured data on large clusters of commodity hardware.

- **HDFS:**Hadoop Distributed File System is a part of Hadoop framework, used to store and process the datasets. It provides a fault-tolerant file system to run on commodity hardware.

The Hadoop ecosystem contains different sub-projects (tools) such as Sqoop, Pig, and Hive that are used to help Hadoop modules.

Hive: It is a platform used to develop SQL type scripts to do MapReduce operations.

DESCRIPTION OF THE DATA:

The data type for all the columns are **STRING.**

COLUMN NAMES

1. crimes_2001_to_present.id

2. crimes_2001_to_present.case number

3. crimes_2001_to_present.date

4. crimes_2001_to_present.block

5. crimes_2001_to_present.iucr

6. crimes_2001_to_present.primary type

7. crimes_2001_to_present.description

8. crimes_2001_to_present.location description

9. crimes_2001_to_present.arrest

10. crimes_2001_to_present.domestic

HIVESQL TO KNOW ABOUT ALL THE COLUMNS:

desc crimes_2001_to_present;

	A	B	C	D	E
1	col_name	col_name	data_type	comment	
2	id	y coordinate	string	from deserializer	
3	case number	year	string	from deserializer	
4	date	updated on	string	from deserializer	
5	block	latitude	string	from deserializer	
6	iucr	longitude	string	from deserializer	
7	primary type	location	string	from deserializer	
8	description	historical wards 2003-201	string	from deserializer	
9	location description	zip codes	string	from deserializer	
10	arrest	community areas	string	from deserializer	
11	domestic	census tracts	string	from deserializer	
12	beat	wards	string	from deserializer	
13	district	boundaries - zip codes	string	from deserializer	
14	ward	police districts	string	from deserializer	
15	community area	police beats	string	from deserializer	
16		fbi code	string	from deserializer	
17					
18					
19					
20					
21					
22					
23					

USING COUNT FUNCTION AND GROUP BY STATEMENT TO RETRIEVE THE DETAILS FOR 'LOCATION DESCRIPTION':
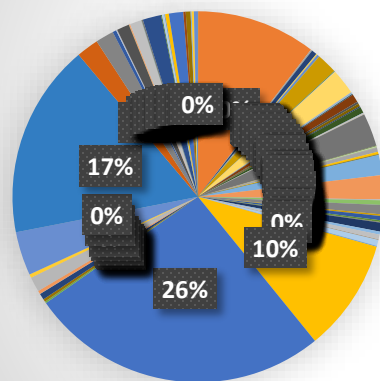
SELECT `location description`, COUNT(*)

FROM crimes_2001_to_present

GROUP BY `location description`;

	location description	_c1
1	location description	_c1
2	AIRCRAFT	652
3	APARTMENT	716655
4	ATHLETIC CLUB	8301
5	BANQUET HALL	1
6	BASEMENT	31
7	BOAT/WATERCRAFT	669
8	CHA PARKING LOT	42
9	CHURCH	6
10	CTA "L" TRAIN	2
11	CTA TRACKS - RIGHT OF WAY	124
12	DRUG STORE	31159
13	FIRE STATION	1024
14	FUNERAL PARLOR	1
15	GOVERNMENT BUILDING	2
16	HIGHWAY/EXPRESSWAY	1046
17	LIBRARY	6059
18	NURSING HOME/RETIREMENT HOME	13952
19	OFFICE	15
20	POOL ROOM	900
21	RETAIL STORE	71
22	RIVER BANK	4
23	SMALL RETAIL STORE	122907

USING COUNT FUNCTION I DETERMINED TOTAL NUMBER OF ARREST BASED ON TRUE AND FALSE

select arrest, count(True)

from crimes_2001_to_present

GROUP BY arrest;

	arrest	_c1
1	false	4983011
2	true	1901263

THE FLLOWING CHART SHOWING TOTAL NUMBER OF ARREST BASED ON TRUE AND FALSE
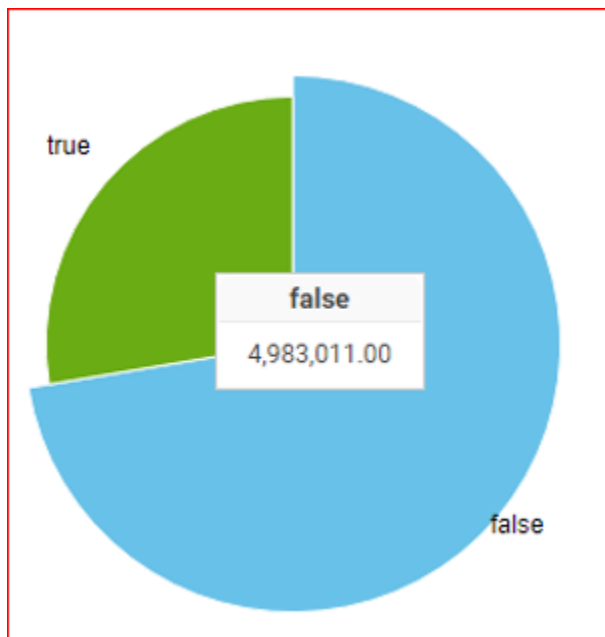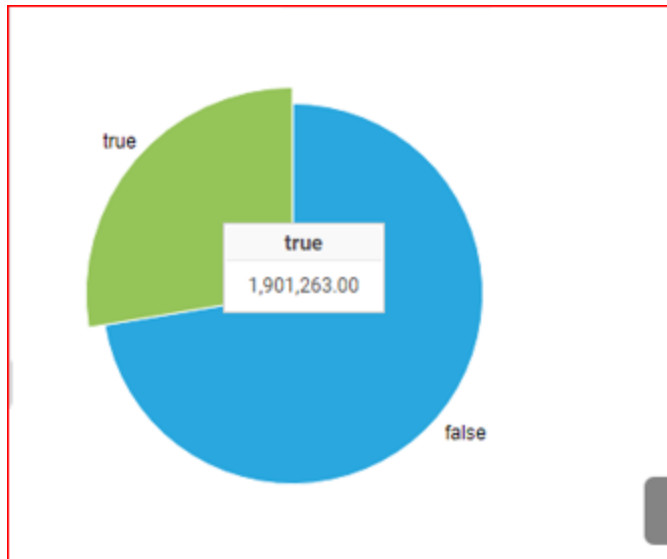
I HAVE SELELCTED ONLY APARTMENT FROM THE LOCATION DESCRIPTION COLUMN BASED ON GAMBLING FROM THE PRIMARY TYPE COLUMN:

select id, `primary type`,`location description` from crimes_2001_to_present

where `primary type`='GAMBLING'

and

`location description`='APARTMENT'

and year= 2018;

crimes_2001	crimes_2001_to_present.primary type	crimes_2001	crimes_2001_to_present.location descriptio	crimes_2001	crimes_2001	crimes_2001_to_present.beat	crimes_2001_to_present.district	crime
1661	GAMBLING	GAME/DICE	APARTMENT	true	false	0312	003	20
1661	GAMBLING	GAME/DICE	APARTMENT	true	false	0331	003	5

HERE I HAVE USED COUNT FUNCTION, ORDER BY, AND GROUP BY STATEMENT AS WELL:

select `location description`, count(*) as counts from

crimes_2001_to_present group by `location description`

order by counts desc

limit 10;

THE FOLLOWING DATA SHOWS ALL THE DISTINCT LOCATIONS FROM THE LOCATION DESCRIPTION COLUMN AND TOTAL NUMBER OF LOCATIONS AS WELL.

	A	B	C
1	location description	counts	
2	STREET	1799841	
3	RESIDENCE	1166173	
4	APARTMENT	716655	
5	SIDEWALK	674886	
6	OTHER	262109	
7	PARKING LOT/GARAGE(NON.RESID.)	197612	
8	ALLEY	153234	
9	SCHOOL, PUBLIC, BUILDING	144468	
10	RESIDENCE-GARAGE	133153	
11	SMALL RETAIL STORE	122907	
12			
13			
14			
15			
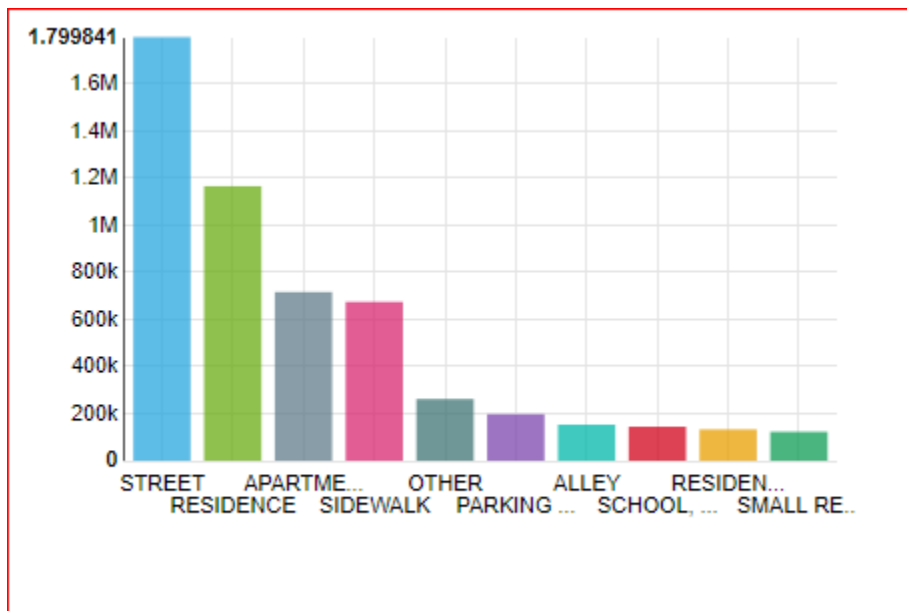
HERE, ALL THE ATRIBUTES ARE DECLARED AS A STRING EVEN INTEGER VALUE ALSO. TO USE MATHEMATICAL FUNCTION WE NEED TO CAST FROM STRING INTO INTEGER. USING SUB QUERY I SELECTED THE DATA FOR ONLY AT WARD NUMBER 10 AND THE DISTRICT NUMBER 4. THE ABOVE DETAILS DESCRIBED IN THE CHART.

select id, description, CAST(ward as INT),CAST(district as INT) from crimes_2001_to_present

where ward IN (select ward from crimes_2001_to_present

where district > 003

group by ward)

limit 20;

	id	description	ward	district
1	id	description	ward	district
2	10526457	$500 AND UNDER	10	4
3	5189037	TO VEHICLE	10	4
4	11504015	DOMESTIC BATTERY SIMPLE	10	4
5	5188127	TO PROPERTY	10	4
6	11385997	TO LAND	10	4
7	5188826	SIMPLE	10	4
8	10582975	ATT: AUTOMOBILE	10	4
9	5192239	POSS: CANNABIS 30GMS OR LESS	10	4
10	10248200	FORCIBLE ENTRY	10	4
11	9801601	$500 AND UNDER	10	4
12	11528821	AGGRAVATED: OTHER DANG WEAPON	10	4
13	9427914	SIMPLE	10	4
14	11472644	FORGERY	10	4
15	9664194	RECKLESS CONDUCT	10	4
16	11336400	AGG: HANDS/FIST/FEET NO/MINOR INJURY	10	4
17	9564349	$500 AND UNDER	10	4
18	10297110	SIMPLE	10	4
19	9419621	ILLEGAL USE CASH CARD	10	4
20	10315414	SIMPLE	10	4
21	10080748	OVER $500	10	4
22				
23				

HERE I HAVE USED MANY AND CONDITIONS AND SUBQUERY TO DETERMIND ONLY AUTOMOBILE THEFT HAPPENED IN THE STREET WITH THE CONDITIONS OF ARREST AND DOMESTIC ARE TRUE AT THE YEAR OF 2018.THE FOLLOWING TABLE SHOWS THE DATA AND THE CHAART DESCRIBES IN DETAILD VISUALIZATION.

select id, description, iucr, arrest,domestic, year, `location description` from crimes_2001_to_present

where year= '2018'

and `location description`="STREET"

and arrest= "true"

and domestic="true"

and description

 IN

(select description from crimes_2001_to_present

where iucr ='0910'

and description ="AUTOMOBILE"

group by description

);

	A	B	C	D	E	F	G	H
1	id	description	iucr	arrest	domestic	year	location description	
2	11376862	AUTOMOBILI	0910	true	true	2018	STREET	
3	11510089	AUTOMOBILI	0910	true	true	2018	STREET	
4	11229146	AUTOMOBILI	0910	true	true	2018	STREET	
5	11477742	AUTOMOBILI	0910	true	true	2018	STREET	
6								
7								
8								
9								

SELECT COUNT(*) as iucrdet,iucr,

  CASE WHEN iucr <910 THEN 'not arrest'

    WHEN iucr >910 THEN 'arrest'

    ELSE 'case'

  END as arrestdet

FROM crimes_2001_to_present

GROUP BY iucr

LIMIT 10;

	A	B	C
1	iucrdet	iucr	arrestdet
2	1	0130	not arrest
3	1712	0261	not arrest
4	1056	0263	not arrest
5	9655	0110	not arrest
6	2818	0266	not arrest
7	34	0262	not arrest
8	307	0264	not arrest
9	44	0142	not arrest
10	3	0141	not arrest
11	4567	0265	not arrest
12			
13			
14			
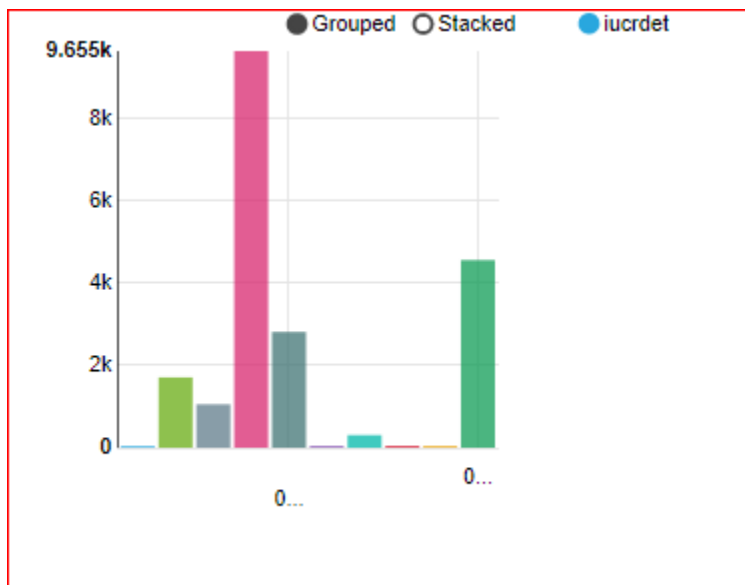
USING PIE CHART I HAVE DESCRIBED THE NUMBER OF BEAT. IN 2018 AT THE STREET SOME THEFT HAPPENED ESPECIALY I HAVE DRILLED TO TAKE OUT THE DETAILS OF CHICAGO POLICE DEPARTMENT DATA.

SELECT description,arrest,`location description`,year ,CAST(beat AS INT) as INTBEAT FROM crimes_2001_to_present

where description ="AUTOMOBILE"

and arrest="true"

and `location description`="STREET"

and year = '2018';

	description	arrest	location description	year	intbeat	
1	description	arrest	location description	year	intbeat	
2	AUTOMOBILE	true	STREET	2018	1234	
3	AUTOMOBILE	true	STREET	2018	1234	
4	AUTOMOBILE	true	STREET	2018	123	
5	AUTOMOBILE	true	STREET	2018	1834	
6	AUTOMOBILE	true	STREET	2018	1014	
7	AUTOMOBILE	true	STREET	2018	511	
8	AUTOMOBILE	true	STREET	2018	724	
9	AUTOMOBILE	true	STREET	2018	324	
10	AUTOMOBILE	true	STREET	2018	825	
11	AUTOMOBILE	true	STREET	2018	413	
12	AUTOMOBILE	true	STREET	2018	2412	
13	AUTOMOBILE	true	STREET	2018	1031	
14	AUTOMOBILE	true	STREET	2018	2412	
15	AUTOMOBILE	true	STREET	2018	124	
16	AUTOMOBILE	true	STREET	2018	1421	
17	AUTOMOBILE	true	STREET	2018	1632	
18	AUTOMOBILE	true	STREET	2018	623	
19	AUTOMOBILE	true	STREET	2018	824	
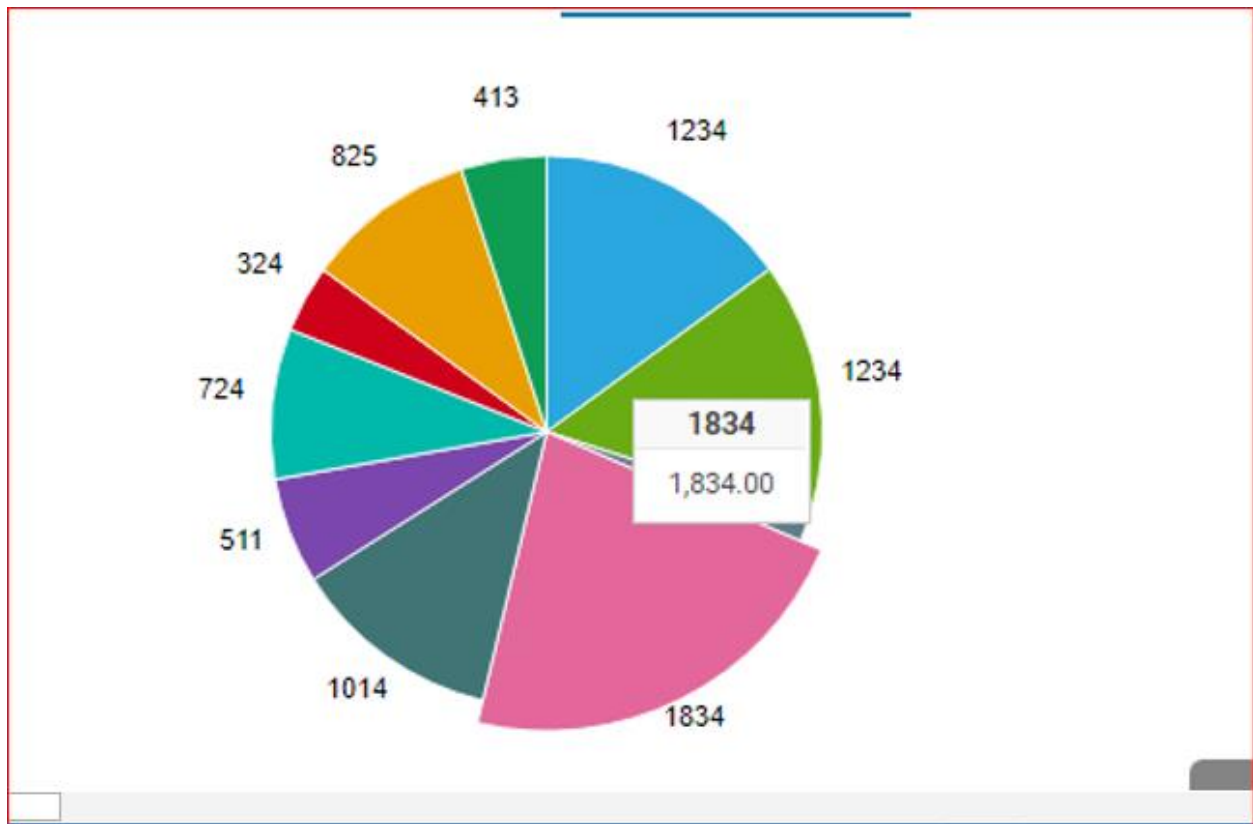
TO SKIP NULL VALUES I USED COUNT FUNCTION TO DETERMIN THE PARTICULAR ID ARRESTED DETAILS. ID NUMBER '11704541' ARRESTED ONLY ONE TIME IN HIS LIFE.

select count(block) as COUNTBLOCK,

count(arrest) as COUNTARREST

from crimes_2001_to_present

where id = '11704541';

Results (1)

	countblock	countarrest
1	1	1

I HAVE DRILLED THE DATA SET TO IDENTIFY THE NUMBER OF DISTRICTS AT THE PARTICULAR WARD USING AVRAGE FUNCTION, GROUP BY, AND ORDER BY STATEMENT.THE FOLLOWING DATA TABLE AND THE DIFFERENT TYPE OF BAR CHARTS DESCRIBE THE DETAILS.

SELECT

  ward,avg_district

FROM

 (SELECT

   ward,

   AVG(district) as avg_district

  FROM

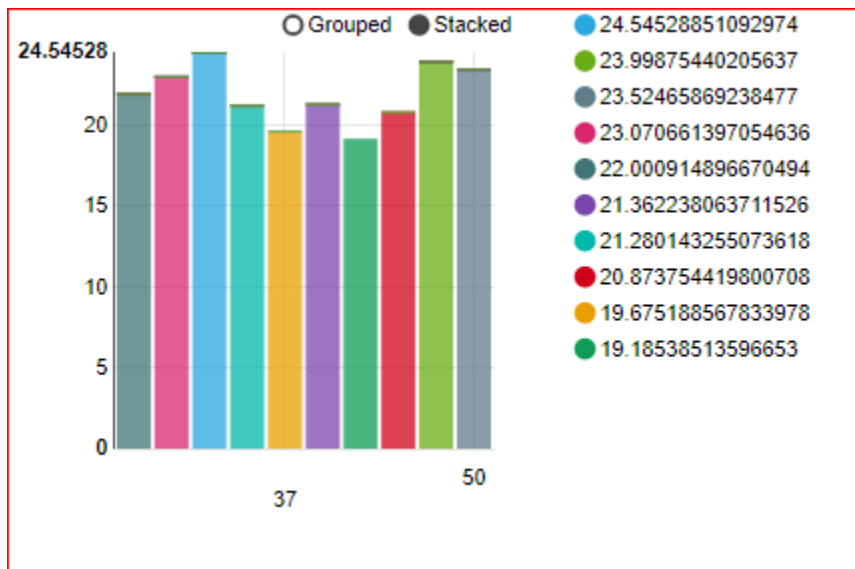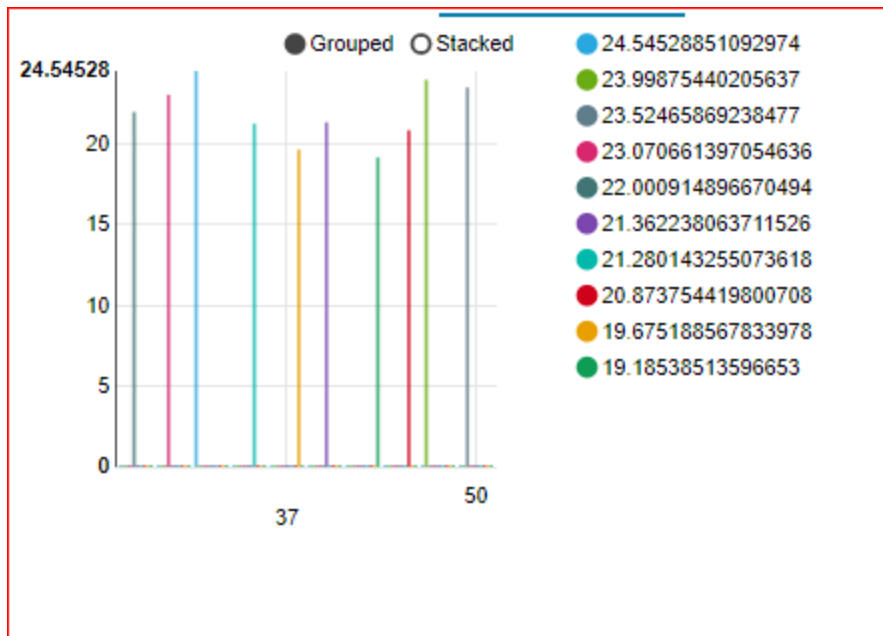 **crimes_2001_to_present**

  GROUP BY ward

  ORDER BY avg_district DESC

  LIMIT 10) as AVGDISRTICTWARD;

	A	B	C
1	ward	avg_district	
2	31	24.5452885	
3	49	23.9987544	
4	50	23.5246587	
5	30	23.0706614	
6	19	22.0009149	
7	40	21.3622381	
8	36	21.2801433	
9	48	20.8737544	
10	37	19.6751886	
11	47	19.1853851	
12			
13			
14			

- 24.54528851092974
- 23.99875440205637
- 23.52465869238477
- 23.070661397054636
- 22.000914896670494
- 21.362238063711526
- 21.280143255073618
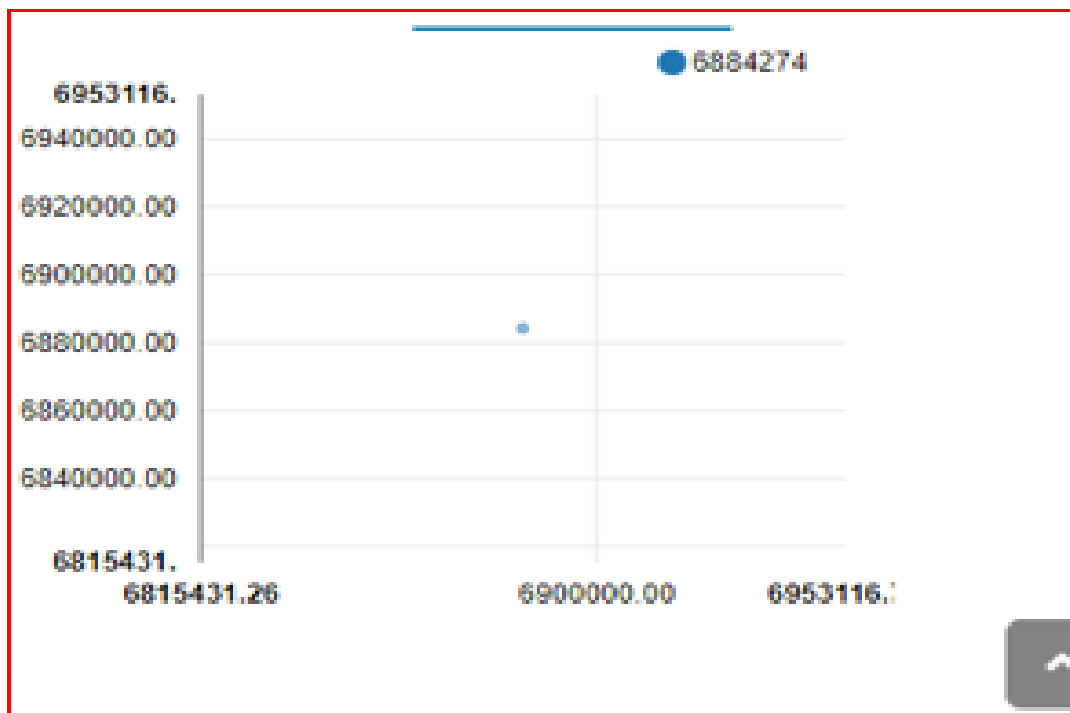- 20.873754419800708
- 19.675188567833978
- 19.18538513596653

USING COUNT FUNCTION I HAVE DETERMINED TOTAL NUMBER OF ROBBERY HAPPENED ONLY AT THE STREET. THE KIND OF ROBBERY IS AUTOMOBILE.

select count(description='AUTOMOBILE')as AUTOMOBILE, count('primary type'='ROBBERY') as ROBBERY,

count('location description'='STREET') as STREET, count(1) as TOTAL

from crimes_2001_to_present;

	A	B	C	D	E	F
1	automobi	robbery	street	total		
2	6884274	6884274	6884274	6884274		
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

A1 : automobile

● 6884274

6953116.
6940000.00
6920000.00
6900000.00
6880000.00
6860000.00
6840000.00
6815431.
    6815431.26   6900000.00   6953116.

THE FOLLOWING TABLE PROVIDES THE DETAILS OF MOTOR VEHICLE THEFT AHPPENED AT THE STREET, AND THE ARREST IS TRUE WITH NUMBER OF BEAT. FOR VISULAIZATION I HAVE USED BAR CHARTS TO PROVIDE MORE DETAILS.

SELECT 'primary type', `description`,beat ,CAST(beat AS INT) as INTBEAT FROM crimes_2001_to_present

where description ="AUTOMOBILE"

and arrest="true"

and 'primary type'="MOTOR VEHICLE THEFT"

and `location description`="STREET";

	crimes_2001_to_present.primary type	crimes_2001_to_present.description	crimes_2001	crimes_2001	crimes_2001	crimes_2001_to_present.bea
1						
2	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET	true	true	2213
3	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET	true	true	0434
4	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET	true	true	0713
5	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET	true	true	0934
6						
7						
8						
9						
10						

I HAVE DILLED THE DAT SET TO TAKE OUT THE DETAILS FOR THE PARTICULAR ID, USING ARITHMATIC OPERATOR, SUB QUERY,AND OPERATOR, CAST FUNCTION IDENTIFIED FOR THAT PARTICULE THEIF ATH STEET WHAT HE DID.IS HE ARRESTED OR NOT?
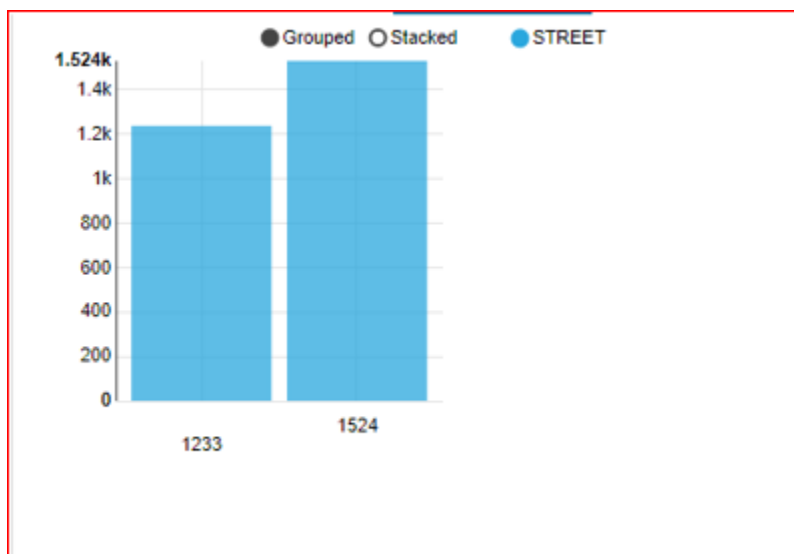
```
select id, description, iucr, arrest,domestic, year, `location description`, cAST(beat as INT)
from crimes_2001_to_present

where year= '2018'

and `location description`="STREET"

and arrest= "true"

and domestic="true"

and beat >='1000'

and description

 IN

(select description from crimes_2001_to_present

where iucr ='0910'

and description ="AUTOMOBILE"

group by description

);
```

	A	B	C	D	E	F	G	H	I
1	id	description	iucr	arrest	domestic	year	location desc	_c7	
2	11510089	AUTOMOBILI	0910	true	true	2018	STREET	2213	
3									
4									
5									
6									
7									
8									
9									
10									

USING SUBQUERY, BETWEEN OPERATOR, AND CAST FUNCTION IDENTIFIED THE DETAILS ID, DESCRIPTION,DISTRICT, WARD NUMBER BETWEEN 2011 TO 2018 AND THE DISTRICT BETWEEN 003 TO 005. THE BAR CHART PROVIDES THE VISUALIZATION OF THE DATA TO UNDERSTAND EASLY.

select id, description, district, CAST(ward as INT) from crimes_2001_to_present

where year between 2011 and 2018

and ward IN (select ward from crimes_2001_to_present

where district between 003 and 005

);

USING SUBQUERY, BETWEEN OPERATOR, AND CAST FUNCTION IDENTIFIED THE DETAILS ID, DESCRIPTION, DISTRICT, WARD NUMBER BETWEEN 2011 AND 2018. HERE, **I TOOK OUT THE DATA BETWEEN 2011 AND 2104.**

select id, description, district, CAST(ward as INT) from crimes_2001_to_present

where year between 2011 and 2014

and ward IN (select ward from crimes_2001_to_present

where district between 003 and 005

);

	A	B	C	D	E	F	G	H
1	year	location description	_c7	id	description	iucr	arrest	domestic
2	2012	STREET	1121	8694631	AUTOMOBILI	0910	true	true
3	2012	STREET	1331	8684174	AUTOMOBILI	0910	true	true
4	2012	STREET	1123	8552898	AUTOMOBILI	0910	true	true
5	2012	STREET	1512	8529667	AUTOMOBILI	0910	true	true
6	2012	STREET	1011	8473901	AUTOMOBILI	0910	true	true
7	2018	STREET	2213	11510089	AUTOMOBILI	0910	true	true
8	2014	STREET	1233	9814062	AUTOMOBILI	0910	true	true
9	2014	STREET	1524	9727359	AUTOMOBILI	0910	true	true
10								
11								
12								
13								
14								

TO VIEW ALL THE DATA IN 2019, THE WARD NUMBER IS 7, COMMUNITY AREA BETWEEN 1 AND 50 AND THE DESCRIPTION IS SIMPLE.

SELECT * FROM crimes_2001_to_present

WHERE description = 'SIMPLE'

and year= '2019'

and ward='7'

and `community area` between 1 and 50

and beat= '0421'

and arrest='true';
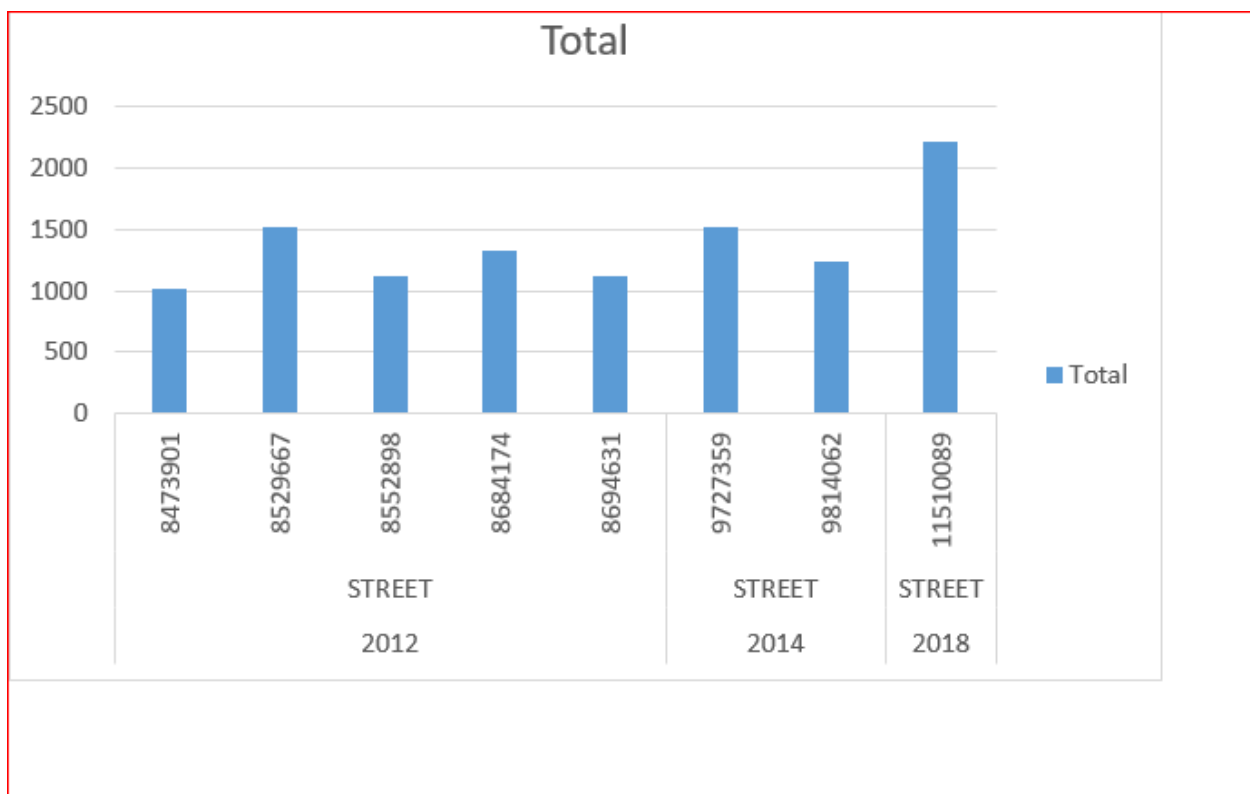
WHEN THE BEET EXCEEDS GREATER THAN 1000 IN AUTOMOBILE THEFT AT THE STREET IN 2012 WITH IUCR NUMBER 0910- TO SEE HOW MANY AUTOMOBILE THEFT HAPPENED. THE BAR CHART VISUALIZED IN THE DIFFERENT COLOR TO THE USER.

```
select id, description, iucr, arrest,domestic, year, `location description`, CAST(beat as INT)
from crimes_2001_to_present

where year= '2012'

and `location description`="STREET"

and arrest= "true"

and domestic="true"

and beat >='1000'

and description

 IN

(select description from crimes_2001_to_present

where iucr ='0910'

and description ="AUTOMOBILE"

group by description

);
```
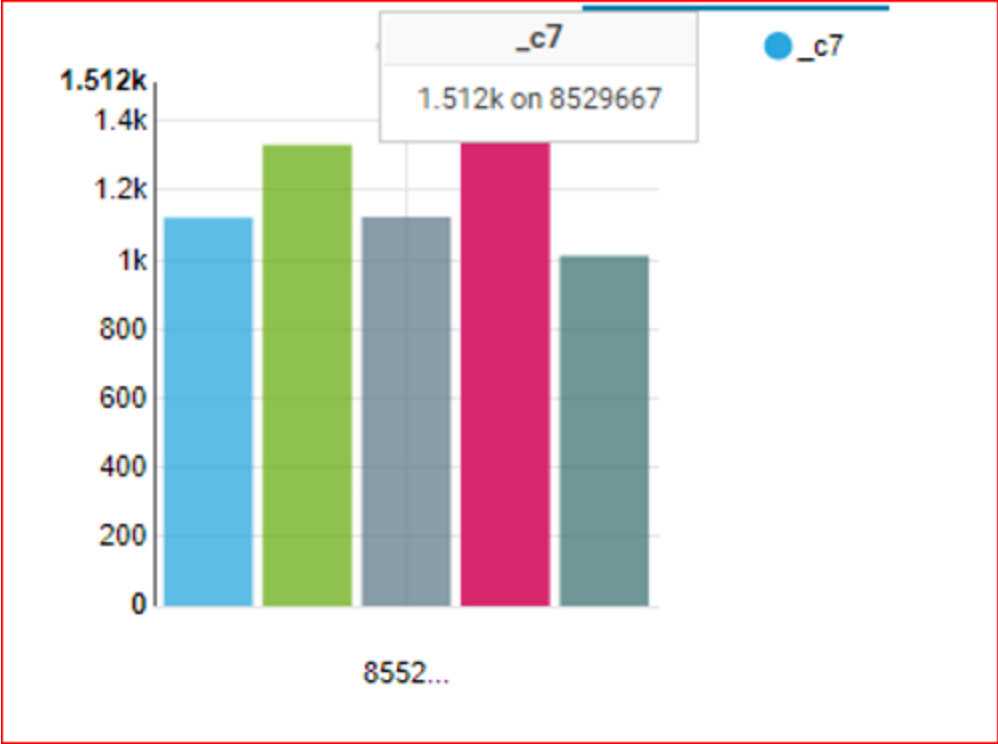
	description	iucr	arrest	domestic	year	location de
1	AUTOMOBILE	0910	true	true	2012	STREET
2	AUTOMOBILE	0910	true	true	2012	STREET
3	AUTOMOBILE	0910	true	true	2012	STREET
4	AUTOMOBILE	0910	true	true	2012	STREET
5	AUTOMOBILE	0910	true	true	2012	STREET

TO SEE IN WARD NUMBER 37 AT DISTRICT NUMBER 011WHAT HAPPENED

select id, arrest, description, ward,district

from crimes_2001_to_present

where iucr='2014'

and

district='011'

and

ward = '37';

HERE I USED WHERE CLAUSE, COUNT FUNCTION, TWO SUBQUERIES, GROUP BY, AVERAGE FUNCTIONAND ORDER BY STATEMENT WITH LIMIT NUMBER 10. I IDENTIFIED THE AVERAGE OF BEAT WITH PARTICULAR WARD.

SELECT COUNT (ward),

description

FROM crimes_2001_to_present

WHERE ward IN

(SELECT ward

FROM

(SELECT ward,

AVG(beat) as avg_beat

FROM crimes_2001_to_present

GROUP BY ward

ORDER BY avg_beat DESC

LIMIT 10) as data_frmcrime)

GROUP BY description

ORDER BY description DESC

LIMIT 10;

	A	B
1	_c0	description
2	8	WIREROOM/SPORTS
3	5	WIREROOM/HORSES
4	7	VIOLENT OFFENDER: DUTY TO REGISTER
5	12	VIOLENT OFFENDER: ANNUAL REGISTRATION
6	22	VIOLATION OF STALKING NO CONTACT ORDER
7	9	VIOLATION OF SMOKING BAN
8	56	VIOLATION OF CIVIL NO CONTACT ORDER
9	5	VIOLATION GPS MONITORING DEVICE
10	4090	VIOLATE ORDER OF PROTECTION
11	1	VIOL CHARITABLE GAME ACT
12		
13		
14		
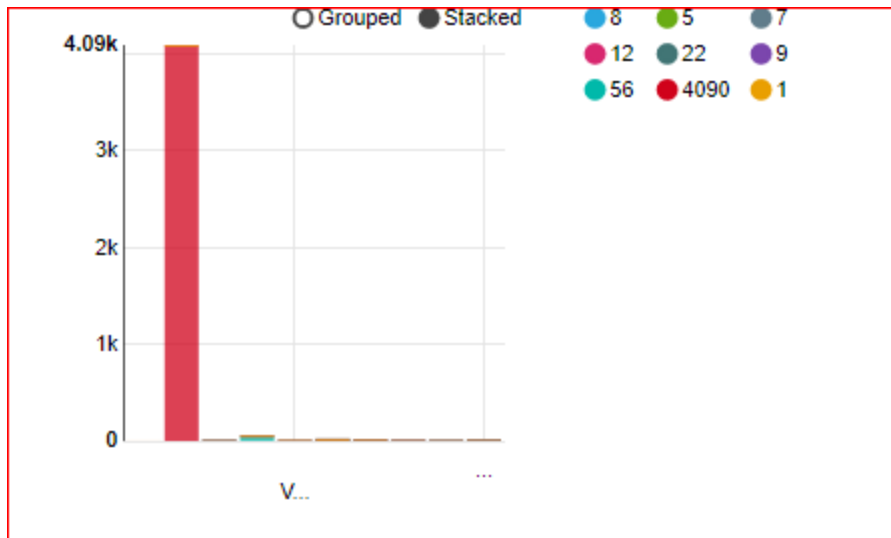
TO SELECT AUTOMOBILE FROM THE DESCRIPTION COLUMN WITH CONDITION OF ARREST IS TRUE AND THE DOMESTIC COLUMN IS EQUAL TO TRUE AT THE PARTICULAR YEAR OF 2012. I USED THE FOLLOWING QUERY TO TAKE OUT THE PARTICULAR DATA.

select id, description, iucr, arrest, domestic, year, `location description`, CAST(beat as INT) from crimes_2001_to_present

where year= '2012'

and `location description`="STREET"

and arrest= "true"

and domestic="true"

and beat >='1000'

and description

 IN

(select description from crimes_2001_to_present

where iucr ='0910'

and description ="AUTOMOBILE"

group by description

);

DATA:

	id	description	iucr	arrest	domestic	year	location desc	_c7
1	id	description	iucr	arrest	domestic	year	location desc	_c7
2	8694631	AUTOMOBILI	0910	true	true	2012	STREET	1121
3	8684174	AUTOMOBILI	0910	true	true	2012	STREET	1331
4	8552898	AUTOMOBILI	0910	true	true	2012	STREET	1123
5	8529667	AUTOMOBILI	0910	true	true	2012	STREET	1512
6	8473901	AUTOMOBILI	0910	true	true	2012	STREET	1011
7								
8								
9								
10								
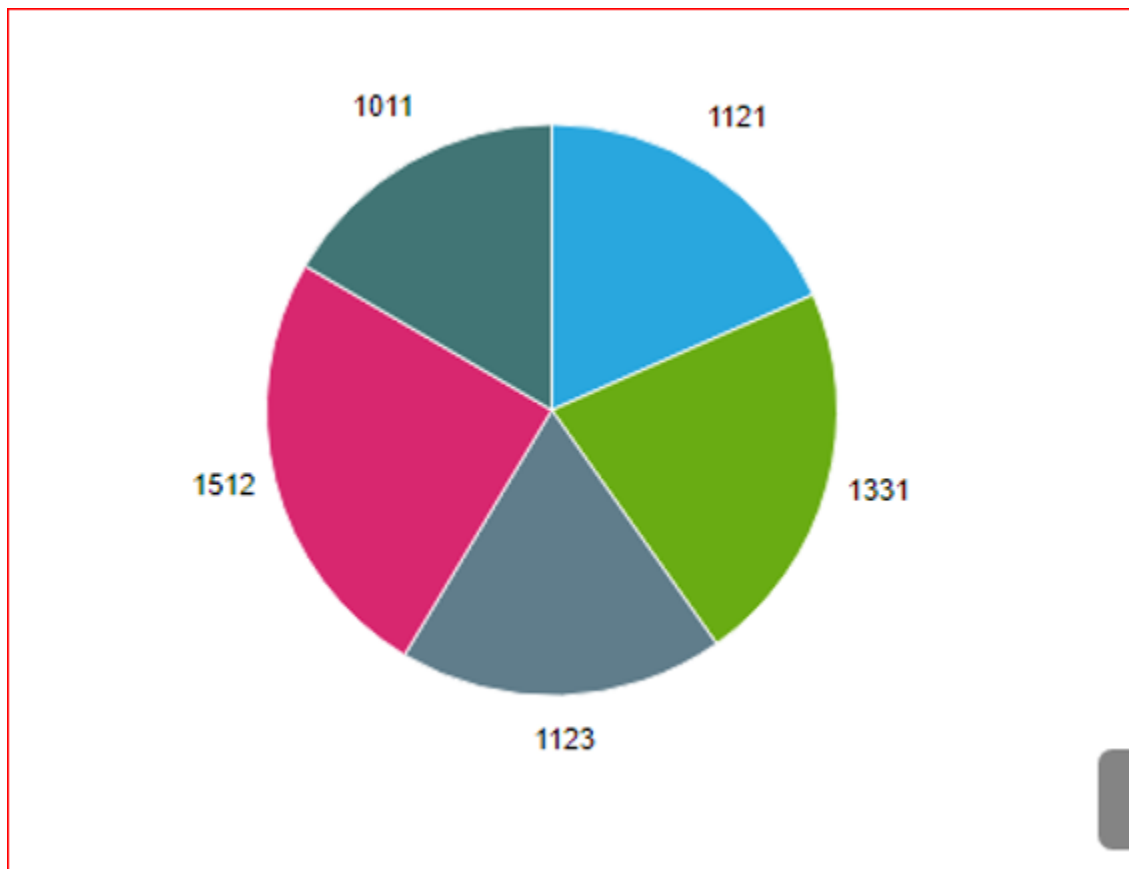
THE ABOVE THE CHART SHOWS WHERE BEAT IS GREATER THAN 1000, SO IT SHOWS FROM 1011 TO 1021.

CONCLUSION:

In conclusion, I have used the HIVE environment to analyze Chicago police department data set. Using HIVESQL I have drilled the data set to take out the theft information because the data set contains unorganized data. To get information from the unstructured data can use the SQL language.

I have collected all the data based on the description, particular id details, total number of beat, an average of the ward in the particular district, and based on the years. This Chicago police Department report provides all the information. In my opinion, the HIVE environment tool is very convenient to use the SQL language.