Volume 9, No.5, September - October 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse313952020.pdf https://doi.org/10.30534/ijatcse/2020/313952020



# An Optimal Warehouse Design for Crime Dataset

Haider Alsharqi<sup>1</sup>, Kadhim B. S. Aljanabi<sup>2</sup>

<sup>1</sup>Information Technology Research and Development Centre, University of Kufa, Najaf, Iraq haider.alsharqi@uokufa.edu.iq <sup>2</sup>Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq kadhim.aljanabi@uokufa.edu.iq

## ABSTRACT

Data science and analytics represent one of the most emerging fields nowadays. Collecting, storing and analyzing the data are challenging issues in the field since they require the most advanced techniques and technologies. Data Warehouse and Data Marts represent some solutions for collecting, storing and accessing the data. Good Warehouse design leads to better analysis results.

Among different application fields of the data, crime data is an important and complex discipline that contains a number of complex relationships between its contents, a wide range of applications and its crucial importance.

The aim of the work in this paper is building an optimal Data warehouse for crime dataset using real crime data collected from the internet. Among the different DW modules available in this field galaxy module is used in this work. The data warehouse will support the decision-making process for lawmaker and police departments by understanding crime subjects, and statistics that allow them to track actions, foretell the probability of occurring crimes and efficiently use supplies which are inverted in this paper.

The proposed design of the DW shows more reliability, better storing and accessing capabilities and lower anomalies among the other designs. The proposed design was supported with a crime database design to remove heterogonous of the data and to apply some preprocessing issues from which they require data is extracted, transformed and loaded (ETL) into the warehouse.

Finally, more than six million high quality, clean, and preprocessed of crime records data are available for the researchers.

Key words: Data warehouse, Database, Data Reduction, Preprocessing.

## **1. INTRODUCTION**

The fast growth of cloud computing and data acquisition and storage technologies, from business and research centres of governments and different organizations, has led to a vast number of unprecedented, complex from data that has been gathered and produced publicly available [1]. It has become more critical to extort meaningful information and provide new insights for understanding patterns from such data stores. Data mining can efficiently address the difficulties of data that are too enormous, unstructured, and fast-moving that not handle by traditional approaches [2]. Data mining is an innovative, interdisciplinary, and growing research field, which can build models and techniques across several areas for inferring useful information and hidden patterns from data [3],[4].The data mining techniques, like clustering, classification and association, can be used for data analysis and prediction by beneficial extract information from raw data.

1- Clustering: is unsupervised learning of data mining technique and is an automatic manner in which the data divided into groups whose segments are similar to each other these are called clusters. For similarity between elements in each cluster, different measures can reflect, such as the distance measure, which uses in the K-Means algorithm, this applies for clustering and samples that are closest to each other can be held as one cluster. So, in image and video databases, clustering can use to discover exciting patterns also characteristics and support content-based retrievals of images and videos using low-level features such as colour, shape, histogram descriptions, texture [5]. There are several algorithms for clustering can divide into two groups:

**A. Traditional Clustering Algorithms:** algorithms based on one of the features like Partition, Hierarchy, Fuzzy Theory, Distribution, Density, Graph Theory, Grid, Fractal Theory and Model.

B. Modern Clustering Algorithms: algorithms based on one of the features like Kernel, Ensemble, Swarm Intelligence, Quantum Theory, Spectral Graph Theory, Affinity Propagation and Density and Distance. Also, the algorithm for Clustering Spatial Data, Clustering Algorithm for Data Streams and Clustering Algorithm for Large-Scale Data [6].
2- Classification: is very strictly associated with the clustering and assigns to as supervised learning. A classification is an approach in data mining which the algorithm learning from the input data and then applies this

information to classify new results. The classification has been examined widely by the database and Artificial Intelligence areas. Some well-known methods of classification are Decision Tree, k-Nearest Neighbor, DNF Rules, Neural Networks and Bayesian Classifiers [7],[8].

**3- Association:** is a process of finding relationships between different attributes in large data sets in various types of databases. These attributes maybe 0 or 1, or they may be quantitative. The concept in association rule is finding the kind of causalities between the values of the various attributes. Association rule mining includes the use of machine learning models to analyze data for patterns, or co-occurrence, in a database. It identifies many if-then relationships, which are called association rules. Association rule generation has considerable importance in data mining because of the ability of its use as an essential mechanism for knowledge finding. For example, a supermarket which the data that records for the different events is the sets of items bought by each customer. For that, it may be helpful to find how the shopping behaviour of one element influences the shopping behaviour of another, Association Rules support for detecting such relationships correctly. Before-mentioned data may use to make target marketing choices. It can also be generalized to classify on big data [9].

There are various types of mining algorithms using. Despite such diversity, some methods are more frequently use [10]. In addition to these methods, a data warehouse, invented by Bill Inmon, is a procedure used in data analysis, database format to analyzing and to report which recognize as OLAP (Online Analytical Processing). From this process, data analysis and doing statistics can be quickly and accurately with refreshed data. Data warehouse structure is a database form that focused on the utilization of statistics. With a data warehouse, architecture reporting process becomes faster, interactive and also can be in real-time [11]. Data warehouse evolved from a normalized database by using ETL (Extraction Transformation and Loading). So, the applications of data mining used to extract knowledge from data like medicine, marketing, weather, crime and various complex data.

There are many models in data warehouse design which represented star, snowflake and galaxy schema. These schemas are created based on reporting needed. in the following sentences, we list the standard schemas with a brief description for each schema:

**1-Star schema:** is the most straightforward and widely used form of data warehouse schemas, include one or more fact tables referencing any dimension tables. Star schema is more efficient with working on simpler queries,

**2-Snowflake schema:** is similar to the star schema. However, in snowflake dimensions are arrangement in various related tables with multiple levels of table relations,

**3-Galaxy schema:** is more complicated of star and snowflake schemas, it contains multiple fact tables connected with various dimension tables that sorted in one level or more, seem like groups of stars [12].

Due to continuous urbanization and growing populations,

especially in enormous societies, increasing Violent crimes and accidents recently led to evidence of crucial and different databases that resulted in needing for getting valuable information to analyze crimes by using diverse technologies [13]. Data mining methods are applied to crime data to identify the complexity of the relationship between the criminals and crime pattern by using clustering and classification algorithms with a suitable data structure.

## 2. RELATED WORKS

The academic literature on data warehouse design has revealed the emergence of several contrasting themes.

Agapito et al. (2020) [14] design COVID-WAREHOUSE where they integrate and save the COVID-19 data, several pollutions and weather data made available in Italy. Also, the data warehouse supports Public Health to understand how the pandemic is spreading in time and geographic area and to associate the pandemic to pollution and climate data in a particular region.

ANUSHA and Jyothi (2020) [15] design a data warehouse for a medical information system to examine the process of data, take decisions, foresee diseases and find cures with the help of data warehouse.

Quitaleg and Ortiz (2020) [16] design and Development of Data Warehouse of Highland Vegetable Crops, this warehouse offers an efficient method for analysis and statistics to the big data in agriculture. It takes data about vegetables, farmers, consumers and other factors and requirements concerning agricultural production. A data warehouse introduces as a solution to agricultural data issues. A.Abdo et al. (2019) [17] design a system for crime prediction by collecting data from Egyptian forensic medicine and create a data warehouse for this data to apply data mining techniques on it, Where the system achieved acceptable results about 98%.

Ari Setyawan et al. (2019) [11] design a data warehouse for an insurance company will help to perform data analysis and to report better and more efficient.it is helps for analyzing the results of insurance sales and used as a reference to the administration to make decisions.

Prabawa et al. (2019) [18] building data warehouse for e-travel company by using the fourth steps dimensional modelling methodology, the design using snowflake scheme become the solution to predict business trends, maintain quality, enhance competitiveness, and exist in the long run for companies.

Sudarmojo et al. (2018) [19] design a data warehouse for a library to analyze data about the transaction process for getting smart decisions in similar service feature evaluation in libraries.

Farooqui et al. (2018) [20] introduce a methodology for the construction of a data warehouse for a medical information system; this data warehouse will help to improve the data analysis and assisting clinical managers to identify decisive patterns, diseases, and their support by enhancing the decision making.

Choo and Chua (2018) [21] implement and design a data warehouse for the semi-structured research literature mining, this study shows data warehouse able to support for exposing hidden pattern or trend of certain aspects of the research literature data such as the keyword locus of a journal. Also discover information about the authors who had cooperated before in writing research literature.

Sutedja et al. (2018) [22] design data warehouse to support active student management by using four stages used by Ralph Kimball in designing a data warehouse. This data warehouse will help the university to analyze active student and make decisions in the student area.

## 3. METHODOLOGY

The proposed approach to design and implement a data warehouse for crime dataset contains the following six stages:

1. Data collection

2. Convert CSV file to SQL table

3. Data Preprocessing

4. Database Design (an intermediate stage to overcome data heterogeneous)

5. Data warehouse design and the second stage of preprocessing

6. Evaluation and analysis

A Proposed Warehouse Design Phases are shown in Figure (1).

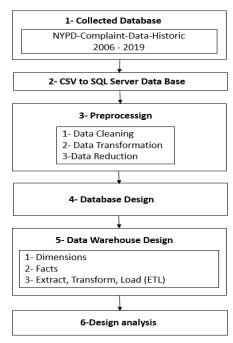


Figure 1: Proposed Warehouse Design Phases

## 3.1 Collected Data

The Crime dataset used in this paper is real data on New York City (NYPD Complaint Data Historic) [23]. Which contain 35 columns with 6,500,871 rows. In CSV file, the data describe the type of crime "(OFNS\_DESC, PD\_DESC, CRM\_ATPT\_CPTD\_CD, LAW\_CAT\_CD )", suspect "(SUSP\_AGE\_GROUP,SUSP\_RACE,SUSP\_SEX )", victims "(VIC\_AGE\_GROUP,VIC\_RACE,VIC\_SEX)" and the location "(BORO\_NM, Latitude, Longitude)" for each crime reported on it in NYPD from 2006 to the end of 2019. Sample of collected raw data is shown in Table (1), Table (2) shows the detailed description of the data attributes.

Table 1: Sample of Collected Raw Data from CSV File

	A	В	с	D	E	F	G	н	1	1	K	L			N	0	F		Q
	IMPLNT_N UM	CMPLNT_F R_DT	CMPLNT_ FR_TM	CMPLNT_T O_DT	CMPLNT_ TO_TM		RPT_DT	KY_CD	OFNS_DES	PD_CD	PD_DESC	CRM_ATPI CPTD_CD	- LAW_C	AT_CD I	IORO_NM	LOC_OF_ CCUR_DI C	O PREM	_TYP JU	RIS_DESC
	522575447	8/29/2005	13:00:00			43	8/30/2006	578	HARRASSN ENT 2	638	HARASSMI NT,SUBD 3,4,5	COMPLETE	D VIOLA	TION	BRONX	INSIDE	RESID	T. N.Y.	POLICE DEI
	403507361	11/5/2006	11:00:00	11/5/2006	17:40:00	66	11/5/2006	107	BURGLAR	221	BURGLARY RESIDENCE DAY	COMPLETE	D FEU		ROOKLYN	INSIDE	RESID	T. N.Y.	POLICE DE
	531420068	9/8/2006	23:30:00	9/9/2006	0:01:00	106	9/9/2006	347	INTOXICAT ED & IMPAIRED DRIVING		INTOXICAT ED DRIVING,A COHOL	COMPLETE	D MISDEN	IEANOR	QUEENS	FRONT O	F STR	EET N.Y.	POLICE DE
	995609899	12/13/2011	18:40:00	12/13/2011	18:49:00	79	12/13/2011	341	PETIT	333	LARCENY, ETIT FROM STORE- SHOPL		D MISDEN	IEANOR	IROOKLYN	INSIDE	CHU STC		POLICE DE
	480667624	8/14/2009	4:20:00			30	8/14/2009	113	FORGERY	729	FORGERY,I TC.,UNCLA SIFIED-FELO	COMPLETE	D FEU	DNY M	ANHATTAN		STR	ET N.Y.	POLICE DEI
l	R	s	т	U	v	W	×		Y	z	AA	AB	AC	AD	AE	AF	AG	AH	AI
	URISDICTI DN_CODE	PARKS_NM	HADEVELO PT	HOUSING_ PSA	X_COORD CD	Y_COOI CD	RD_ SUSP_A _GROU	GE P SUSF	P_RACE SU	P_SEX 1	RANSIT_DI STRICT	Latitude	Longitude	Lat_Lon	PATROL_B ORO	STATION _NAME	VIC_AGE _GROUP	VIC_RACE	VIC_SEX
Ī	0	NA		NA	1018029	24074	17	UNK	NOWN	м		40.827414	73.877946	4051, -73.8	PATROL BORO BRONK		25-44	BLACK HISPANIC	F
	0	NA		NA	982556	17138	IS					40.637097	74.006105	16864, -74.0	PATROL BORO BKLYN SOUTH		45-64	ASIAN / PACIFIC ISLANDER	F
	0	NA		NA	1028213	18678	16					40.67926	-73.8415	i0229, -73.8	PATROL BORO QUEENS SOUTH			UNKNOWN	E
	0	NA		NA	1000788	18971	.8					40.687402	73.940369	)1619, -73.9	PATROL BORO BKLYN NORTH			UNKNOWN	D
	0	NA		NA	1000029	24224	IS					40.831576	73.942983	6128, -73.9	PATROL BORO MAN NORTH			UNKNOWN	E

Table 2: Attribute Description of The Collected Raw Data

	Column	•
	Name	Column Description
1	CMPLNT	Randomly generated persistent ID for
	NUM	each complaint
2	CMDI NIT	Exact date of occurrence for the reported
	CMPLNT_	event (or starting date of occurrence, if
	FR_DT	CMPLNT_TO_DT exists)
3	CMPLNT	Exact time of occurrence for the reported
	FR_TM	event (or starting time of occurrence, if
	FK_IM	CMPLNT_TO_TM exists)
4	CMPLNT	Ending date of occurrence for the
	TO DT	reported event, if exact time of
	10_D1	occurrence is unknown
5	CMPLNT	Ending time of occurrence for the
	TO_TM	reported event, if exact time of
	_	occurrence is unknown
6	ADDR_PC	The precinct in which the incident
	T CD	occurred
7	RPT_DT	Date event was reported to police
8	KY_CD	Three digit offense classification code
9	OFNS_DE	Description of offense corresponding
	SC	with key code
10	PD CD	Three digit internal classification code
	FD_CD	(more granular than Key Code)
11		Description of internal classification
	PD_DESC	corresponding with PD code (more
		granular than Offense Description)
12	CRM_ATP	Indicator of whether crime was
	T_CPTD_	successfully completed or attempted, but
	CD	failed or was interrupted prematurely
13	LAW_CA	Level of offense: felony, misdemeanor,
	T CD	violation
14	BORO_N	The name of the borough in which the
	М	incident occurred

15	LOC_OF_ OCCUR D ESC	Specific location of occurrence in or around the premises; inside, opposite of, front of, rear of
16	PREM_TY P DESC	Specific description of premises; grocery store, residence, street, etc.
17	JURIS_DE SC	Description of the jurisdiction code
18	JURISDIC TION CO DE	Jurisdiction responsible for incident. Either internal, like Police(0), Transit(1), and Housing(2); or external(3), like Correction, Port Authority, etc.
19	PARKS_N M	Name of NYC park, playground or greenspace of occurrence, if applicable (state parks are not included)
20	HADEVEL OPT	Name of NYCHA housing development of occurrence, if applicable
21	HOUSING PSA	Development Level Code
22	X_COORD CD	X-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
23	Y COORD CD	Y-coordinate for New York State Plane Coordinate System, Long Island Zone, NAD 83, units feet (FIPS 3104)
24	SUSP_AG E GROUP	Suspect's Age Group
25	SUSP_RA CE	Suspect's Race Description
26	SUSP_SE X	Suspect's Sex Description
27	TRANSIT_ DISTRICT	Transit district in which the offense occurred.
28	Latitude	Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
29	Longitude	Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
30	Lat Lon	Geospatial Location Point (Latitude and Longitude combined)
31	PATROL_ BORO	The name of the patrol borough in which the incident occurred
32	STATION NAME	Transit station name
33	VIC_AGE GROUP	Victim's Age Group
34	VIC_RAC E	Victim's Race Description
35	VIC SEX	Victim's Sex Description

## 3.2 Converting CSV file to SQL table

Converted the CSV file to SQL Server database by implemented python 3.6 code with (pyodbc and pandas) packages, this language was used because it is open-source language, can be used for any platform operating system and also suitable for data mining and big data [24]. The SQL table is shown in Figure (2).

	/***** Scr	ipt for Selec	tTopNRows com	mand from SSM	s *****/						
	SELECT [CMP	UNT_NUM], [CMP	UNT_FR_DT], [C	MPLNT_FR_TH],	CMPLNT_TO_DT	CMPLNT_TO	TM], [ADDR	PCT_CC	], [RPT_DT], [KY_CD], [OFNS_DB	SC], [P	D_CD], [PD_DESC] ADEVELOPT], [HOUSING_PSA], [X_COORD_C
											ION_NAME],[VIC_AGE_GROUP],[VIC_RACE
					ore prprocessi						
	6 • 6										
	esults 🗄 Messag	~									
	CMPLNT NUM	CMPLNT_FR_DT	CMPLNT FR TM	CMPINT TO DT	CMPLNT TO TM	ADDR PCT_CD	RPT DT	KY CD	OFNS DESC	PD CD	PD DESC
	522575447	08/29/2006	13:00:00	nan	040	43	08/30/2005	578	HARRASSMENT 2	638	HARASSMENT SUBD 3.4.5
	403507361	11/05/2006	11:00:00	11/05/2008	17.40.00	66	1105/2006	107	BURGLARY	221	BURGLARY RESIDENCE DAY
	631420068	09/08/2006	23.30.00	09/09/2005	00.01.00	105	09/09/2006	347	INTOXICATED & IMPAIRED DRIVING	905	INTOXICATED DRIVING ALCOHOL
	995609899	12/13/2011	18:40:00	12/13/2011	18.49.00	79	12/13/2011	341	PETIT LARCENY	333	LARCENY, PETIT FROM STORE-SHOPL
	480667624	08/14/2009	04/20:00	nan	030	30	08/14/2009	113	FORGERY	729	FORGERY ETC. UNCLASSIFIED-FELO
	605915885	10/27/2009	21:10:00	nan	nan	69	10/29/2009	105	ROBBERY	397	ROBBERY OPEN AREA UNCLASSIFIED
	820084484	12/16/2006	20.00.00	12/16/2005	20.05:00	40	12/16/2006	235	DANGEROUS DRUGS	567	MARUUANA, POSSESSION 4 & 5
	799044409	07/21/2007	23.15.00	07/21/2007	23.45.00	75	07/22/2007	109	GRAND LARCENY	421	LARCENY GRAND FROM VEHICLE MOTORCYCL
	211477358	07/07/2008	21:30:00	07:05/2008	11.45.00	34	07:08/2008	341	PETIT LARCENY	321	LARCENY, PETIT FROM AUTO
0	986887005	09/17/2007	16:00:00	nan	nan	48	09/19/2007	578	HARRASSMENT 2	637	HARASSMENT, SUBD 1, CIVILIAN
1	207970722	01/06/2007	16:00:00	02/26/2007	12:00:00	68	02/26/2007	340	FRAUDS	718	FRAUD.UNCLASSIFIED-MISDEMEANOR
2	537893224	04/26/2009	23.00.00	04/27/2009	11:30:00	122	04/27/2009	351	CRIMINAL MISCHIEF & RELATED OF	254	MISCHIEF, CRIMINAL 4, OF NOTOR
3	924188035	01/08/2010	22:00:00	nan	nan	75	01/11/2010	344	ASSAULT 3 & RELATED OFFENSES	101	ASSAULT 3
4	321158154	11/02/2007	11:55:00	11/02/2007	11:56:00	71	1102/2007	235	DANGEROUS DRUGS	567	MARUUANA, POSSESSION 4 & 5
5	103912654	08/30/2009	01:00:00	08/30/2009	01.01.00	109	08/30/2009		FELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED
6	591753685	12/09/2008	15:10:00	12/09/2008	15:14:00	110	12/09/2008	105	ROBBERY	384	ROBBERY POCKETBOOK/CARRIED BAG
7	289318739	07/18/2007	07.00.00	11/30/2007	12:00:00	23	11/30/2007	341	PETIT LARCENY	313	LARCENY, PETIT BY FALSE PROMISE
8	574443172	04/27/2006	17:00:00	04/28/2008	22.40.00	44	04/28/2006	361	OFF. AGNST PUB ORD SENSBLTY &	639	AGGRAVATED HARASSMENT 2
9	935932450	04/11/2007	17:00:00	04/14/2007	18.00.00	10	04/14/2007	109	GRAND LARCENY	438	LARCENY, GRAND FROM BUILDING (NON-RESID
1	800480404	11/05/03/06	11-16-00			77	11/05/2008	161	ORE MONGTOND ODD CENCERTY F	655	ACCORDANIATED MADACEMENT 5
	any executed suc										VO Crime NYPD OLD 00:12:39 6500870

Figure 2: Sample of Data from SQL Table

## 3.3 Data Preprocessing

Data preprocessing is a significant stage of data analysis because real-world data is impure, contain a lot of missing values or duplicating data, and high-performance mining methods expect quality data [25]. So, we will apply three steps of data preprocessing to enhancement the quality of the dataset.

## 3.3.1 Data Cleaning

• Deleting(CMPLNT\_TO\_DT,CMPLNT\_TO\_TM,

JURIS\_DESC, URISDICTION\_CODE, PARKS\_NM) colums because these columns not important in future analysis. In the other hand, deleted the columns that have Too many missing values(HADEVELOPT, HOUSING PSA, TRANSIT DISTR ICT, STATION\_NAME) also deleted the columns that have duplicated data(X\_COORD\_CD, Y\_COORD\_CD, Lat\_Lon). • Deleting every row with missing value in columns (CMPLNT\_FR\_DT, CMPLNT\_FR\_TM, OFNS\_DESC, CRM ATPT CPTD CD, BORO NM, Latitude, Longitude, PATROL BORO), and convert every missing value in these columns (LOC\_OF\_OCCUR\_DESC, PREM\_TYP\_DESC, SUSP\_AGE\_GROUP, SUSP\_RACE, SUSP\_SEX, VIC AGE GROUP, VIC RACE, VIC SEX) to (UNKNOWN) value. And also add same (OFNS DESC) with (PD DESC) to rows that have missing value in (PD DESC) column.

## 3.3.2 Data Transformation

In this step, we convert "(ADDR\_PCT\_CD, KY\_CD, PD\_CD, Latitude, Longitude)" columns to numerical data from string format after export it from CSV file to SQL Server Database. Also converting "( CMPLNT\_FR\_DT, CMPLNT\_FR\_TM, RPT\_DT)" columns into date time format.

## 3.3.3 Data Reduction

Rounding the Latitude and Longitude columns into two decimals numbers to reduce the distinct values in these columns. So, it can be used with Patrol name "(P\_name)" to find the Approximate location for crime.

After the preprocessing stage, the SQL table contains 23 columns and 6,435,235 rows. The column names have been

changed into more meaningful names. Sample of data after preprocessing phase is shown in Figure (3).

	,[Pres_ty	pe_desc],[SU	SP_AGE_GROU	JP], [SUSP_R/	ACE], [SUSP	P_SEX],[patrol_plac	e],[VIC_AS	me],[C_DES_Code],[C_DESC E_GROUP],[VIC_RACE],[VIC	SEX],[Latitud	e],[Longit	ude]			
		ta_NYPD_CRIM	E].[dbo].[0	Crimes_after	r_Process]	1								
00 1														
1 8	suls 🔓 Mess													
	ID C_Dab	C_Time 8-29 13:00:00.00	Preci.	Reported_date 2006-08-30		Name ARRASSMENT 2	C_D 638	. C_DESC HARASSMENT,SUBD 3,4,5		C_status COMPLETED	C_Law_DESC VIOLATION	P_name REONX	P_around_p INSIDE	pre
	2 2005-0			2006-08-30		URGLARY	221	RURGLARY RESIDENCE DAY		COMPLETED	FELONY	ERONX EROCKLYN	INSIDE	
	3 2005-0			2006-09-09		TOXICATED & INPAIRED D		INTOXICATED DRMING ALCOH	10	COMPLETED	MISDEMEANOR		FRONT OF	
	4 2011-1			2011-12-13		ETITLARCENY	333	LARCENY PETIT FROM STORE		COMPLETED	MISDEMEANOR		INSIDE	
	5 2009-0	8-14 04 20 00.00	000000 30	2009-08-14		ORGERY	729	FORGERY ETC. UNCLASSIFIED		COMPLETED	FELONY	MANHATTAN	UNKNOWN	N
	6 2009-1	0-27 21:10:00.00	49 00000	2009-10-29	106 Rt	OBBERY	397	ROBBERY OPEN AREA UNCLA	SSIFIED	COMPLETED	FELONY	BROOKLYN	FRONT OF	F
	7 2008-1	2-16 20:00:00.00	000000 40	2008-12-16	235 Dr	ANGEROUS DRUGS	567	MARUUANA, POSSESSION 4.8	5	COMPLETED	MISDEMEANOR	BRONX	INSIDE	
	8 2007-0	7-21 23:15:00.00	000000 75	2007-07-22	100 GI	RAND LARCENY	421	LARCENY, GRAND FROM VEHI	LEMOTORCYCLE	COMPLETED	FELONY	BROOKLYN	FRONT OF	
		7-07 21:30:00.00		2008-07-08		ETIT LARCENY	321	LARCENY, PETIT FROM AUTO		COMPLETED	MISDEMEANOR		UNKNOWN	
0		9-17 16:00:00.00		2007-09-19		ARRASSMENT 2	637	HARASSMENT, SUBD 1, CIVILIA		COMPLETED	VIOLATION	BRONX	UNKNOWN	N
1	11 2007-0			2007-02-26		RAUDS	718	FRAUD, UNCLASSIFIED-MISDE		COMPLETED	MISDEMEANOR		INSIDE	
2	12 2009.0		000000 122	2009-04-27		RIMINAL MISCHIEF & RELA		MISCHIEF, CRIMINAL 4, OF NO	TOR	COMPLETED	MISDEMEANOR		FRONT OF	
3	13 2010-0			2010-01-11		SSAULT 3 & RELATED OFF		ASSAULT 3		COMPLETED	MISDEMEANOR		UNKNOWN	
4		102 1155:00.00		2007-11-02		ANGEROUS DRUGS	567	MARIJUANA, POSSESSION 4.8	5	COMPLETED	MISDEMEANOR		UNKNOWN	
5	15 2009-0			2009-08-30		ELONY ASSAULT	109	ASSAULT 2,1,UNCLASSIFIED		COMPLETED	FELONY	QUEENS	UNKNOWN	
3	16 2008-1			2008-12-09		OBBERY	384	ROBBERY, POCKETBOOK/CARP		ATTEMPTED	FELONY	QUEENS	UNKNOWN	N
7	17 2007-0			2007-11-30		ETIT LARCENY	313	LARCENY, PETIT BY FALSE PRI		COMPLETED	MISDEMEANOR		INSIDE	
3		4-11 17.00.00.00		2007-04-14		RAND LARCENY	438	LARCENY, GRAND FROM BUILT		COMPLETED	FELONY	MANHATTAN	INSIDE	
9		1-25 11:15:00.00		2006-11-25		FF. AGNST PUB ORD SENS		AGGRAVATED HARASSMENT		COMPLETED	MISDEMEANOR		INSIDE	
)	20 2010-0	13-23 01:00:00:00 16-30 17:00:00:00		2010-03-24 2009-07-01		FF. AGNST PUB ORD SENS ARRASSMENT 2	BLTY & 639 637	AGGRAVATED HARASSMENT HARASSMENT SUBD 1 CIVILIA		COMPLETED COMPLETED	MISDEMEANOR	BROOKLYN	INSIDE	
V	/***** 5	SKR4C2\CLEVO( tript for Sel	LectTopNRov				viel (C Nar	DESKTOP-43KR4C2 (12.0 SP			Data_NYPD_CR			rc
20	SELECT [I]	SKR4C2/CLEVO ( tript for Se) D_Crime],[C_I	LectTopNRov Date],[C_Ti SP_AGE_GROL	lme],[Precin JP],[SUSP_RA	<pre>hct_no],[R vCE],[SUSP</pre>	Reported_date],[C_Co _SEX],[patrol_place		DESKTOP-43KR4C2 (12.0 SP me], [C_DES_Code], [C_DESC] s_GROUP], [VIC_R4CE], [VIC	[,[C_status],[I	C_Law_DESC]	, [P_name], [P			rc
50	SELECT [II , [Prem_ty] FROM [Dat	SKR4C2\(LEVO ( tript for Se) D_Crime],[C_I De_desc],[SU	LectTopNRov Date],[C_Ti SP_AGE_GROL	lme],[Precin JP],[SUSP_RA	<pre>hct_no],[R vCE],[SUSP</pre>	Reported_date],[C_Co SEX],[patrol_place		we], [C_DES_Code], [C_DESC	[,[C_status],[I	C_Law_DESC]	, [P_name], [P			ra
0.	SELECT [II , [Prem_ty] FROM [Dar	SKR4C2/(LEVO ( cript for Se D_Crime],[C_ De_desc],[SU ta_NYPD_CRIM	LectTopNRov Date],[C_Ti SP_AGE_GROL	lme],[Precin JP],[SUSP_RA	<pre>hct_no],[R vCE],[SUSP</pre>	Reported_date],[C_Co SEX],[patrol_place		we], [C_DES_Code], [C_DESC	[,[C_status],[I	C_Law_DESC]	, [P_name], [P			
0.	SELECT [II , [Prem_ty] FROM [Dar south 12 Masse	SKR4C2(CLEVO ( cript for Se D_Crime],[C_ Se_desc],[SU ta_NYPD_CRIMI ages	LectTopNRov Date],[C_Ti SP_AGE_GROL E].[dbo].[C	ime],[Precir JP],[SUSP_RJ Trimes_after	nct_no],[R NCE],[SUSP N_Process]	Reported_date],[C_Cc P_SEX],[patrol_place	e],[VIC_AG	we],[C_DES_Code],[C_DESC e_GROUP],[VIC_RACE],[VIC_	,[C_status],[i SEX],[Latitud	C_Law_DESC] e],[Longitu	, [P_name], [P ide]	_around_pre	n)	>
0.	SELECT [I] SELECT [I] FROM [Dar ROM [Dar Muts ]] Massi P_name	SKR4C2/CLEVO ( cript for Se )_Crime],[C_ se_desc],[SU ta_NYPD_CRIM eps P_around_prem	LectTopNRov Date],[C_Ti SP_AGE_GROL E].[dbo].[C Prem_hpe_de	ime],[Precir JP],[SUSP_R# Inimes_after %0	susp_AGE_0	Reported_date),[C_Cc P_SEX],[patrol_place   GROUP_SUSP_RACE	e],[VIC_AG	we], [C_DES_Code], [C_DESC] _GROUP], [VIC_RACE], [VIC_ patrol_phon	,[C_status],[I SEX],[Latitud	C_Law_DESC] e],[Longitu VIC_RACE	, [P_name], [P ide]	_around_pro	m] de Longhude	>
0.	SELECT [II , [Prem_ty] FROM [Dar south 12 Masse	SKR4C2(CLEVO ( cript for Se D_Crime],[C_ Se_desc],[SU ta_NYPD_CRIMI ages	LectTopNRov Date],[C_Ti SP_AGE_GROL E].[dbo].[C Prom_type_dr RESIDENCE	Lme], [Precir JP], [SUSP_RA Inimes_after SSC - APT HOUSE	nct_no],[R NCE],[SUSP N_Process]	Reported_date],[C_Cc P_SEX],[patrol_place	SUSP_SEX	we],[C_DES_Code],[C_DESC e_GROUP],[VIC_RACE],[VIC_	,[C_status],[i SEX],[Latitud	C_Law_DESC e],[Longitu VIC_RACE BLACK HISP)	, [P_name], [P ide]	_around_pre	n)	>
	SELECT [I] , [Prem_ty] FROM [Dar 6 * < mults [] Massa P_name BRONX	SKR4C2(CLEVO ( cript for Sel )crime], [C_] pe_desc], [SU ta_NYPD_CRIM ages P_around_prem INSIDE	LectTopNRov Date],[C_Ti SP_AGE_GROL E].[dbo].[C Prom_type_dr RESIDENCE	Lme], [Precir JP], [SUSP_RA Inimes_after SSC - APT HOUSE - APT HOUSE	susp_age_o	Reported_date],[C_Cc P_SEX],[patrol_place GROUP_SUSP_RACE UNENCOWN	<pre>b), [VIC_AG SUSP_SEX</pre>	we], [C_DES_Code], [C_DESC _GROUP], [VIC_RACE], [VIC_ patrol_place PATROL_BORD BRONK	(,[C_status],[I SEX],[Latitude VIC_AGE_GROUP 25-44	C_Law_DESC e],[Longitu VIC_RACE BLACK HISP)	, [P_name], [P de]	_around_pro	m] de Longhude -73.85	,
0.	SELECT [II , [Pres_ty] FROM [Dar 6 * < P_name BROOK BROOKLYN	SKR4C2(CLEVO) cript for Sel D_Crime], [C_1 De_desc], [SUI ta_WYPD_CRIMI ages P_around_prem mesiDE mesiDE	LectTopNRov Date],[C_Ti SP_AGE_GROL E].[dbo].[C Prem_hyps_dk RESIDENCE RESIDENCE	Lme], [Precir JP], [SUSP_R# Inimes_after SSO - APT HOUSE - APT HOUSE	susp_Age_ unknown	Reported_date),[C_Cc P_SEX],[patrol_place GROUP_SUSP_RACE UNENCOWN UNENCOWN	SUSP_SEX M UNKNOWN	w]; [C_DES_Code]; [C_DESC 	([C_status])[ SEX],[Latitud VIC_AGE_GROUP 25-44 45-64	C_Law_DESC] e],[Longitu VIC_RACE BLACK HISPI ASIAN/PACI	, [P_name], [P de]	"_around_pro	m] 5e Longitude -73.85 -74.01	>
0.	SELECT [II , [Pres_ty] FROM [Dar 6 * < P_name BROOKX BROOKLYN QUEENS	SKR4C2/CLEVO( cript for Se D_Crime],[C_] se_desc],[SU ta_NYPD_CRIM ages P_around_prom RESIDE FRONT OF INSIDE	LectTopNRov Date] [C_T3 SP_AGE_GROL E] . [dbo] . [C Prem_hype_dc RESIDENCE RESIDENCE STREET	Lme], [Precir JP], [SUSP_R4 Crimes_after SSO - APT. HOUSE - APT. HOUSE E	susp_Age_o	Reported_date),[C_Cc _SEX],[patrol_place GROUP_SUSP_RACE UNENCOWN UNENCOWN UNENCOWN	SUSP_SEX M UNKNOWN UNKNOWN	(C_DES_Code], [C_DESC _GROUP], [VIC_RACE], [VIC patrol_pace PATROL BORO BRONK PATROL BORO BRONK PATROL BORO BRONK	VIC_AGE_GROUP 25-44 45-64 UNRNOWN	C_Law_DESC] e],[Longitu VIC_RACE BLACK HISP) ASIAN (PACI UNRXOWN	I, (P_name), (P ide)	_around_pro	m] de Longhude -73.85 -74.01 -73.84 -73.94	>
0.	SELECT [I] SELECT [I] FROM [Dar FROM [Dar FROM [Dar BRONX BROOKLYN BROOKLYN BROOKLYN	SKR4C2(CLEVO ( cript for Se) Crime], [C_] 2e_desc], [SU: ta_WYPD_CRIMI ResiDE FRONT OF	LectTopNRov Date], [C_Ti SP_AGE_GROL E].[dbo].[dbo].[dbo].[dbo]. Prom_hype_dc RESIDENCE RESIDENCE STREET CHAIN STOR	ime], [Precir JP], [SUSP_RJ Inimes_after -APT. HOUSE - APT. HOUSE E	NCE], [SUSP Process] SUSP_AGE_C UNKNOWN UNKNOWN UNKNOWN UNKNOWN	Reported_date), [C_CC _SEX), [patrol_place   GROUP SUSP_RACE UNENCOWN UNENCOWN UNENCOWN UNENCOWN	SUSP_SEX N UNKNOWN UNKNOWN	b) [C_DES_Code] [C_DESC _GROUP] [VIC_RACE] [VIC_ patrol_pace PATROL BORD BROKK PATROL BORD BRYN SOUTH PATROL BORD BLYN SOUTH PATROL BORD BLYN SOUTH	VIC_AGE_GROUP 25-44 45-64 UNRXXVIN UNRXXVIN	C_Law_DESC] e],[Longitu VIC_RADE BLACK HISP ASIAN/PACE UNRXXVWN UNRXXVWN	I, [P_name], [P Ide]	"_around_pre AC_SEX_Lattu = 40.82 = 40.84 = 40.85 = 40.85 = 40.85	m] de Longhude -73.85 -74.01 -73.84 -73.94	>
0.	Select [I] , [Prem_ty] FROM [Dar 6 * < P_name BRONX BROOKLYN BROOKLYN BROOKLYN BROOKLYN BROOKLYN BROOKLYN	SCR4C2(CLEVO) pript for Se Crime], [C_] pe_desc], [SU taVVPD_CRIM ages P_around_prem mesiDE RESIDE UNENDE UNENDE UNENDE UNENDE	lectTopNRov Date], [C_T3 P_AGE_GROL ].[dbo].[C Prom_type_de RESIDENCE STREET CHAIN STOR STREET STREET RESIDENCE	ime], [Precir IP], [SUSP_RA Primes_after -APT. HOUSE - APT. HOUSE E	Ct_no], [R CE], [SUSP Process] SUSP_AGE_C UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	Reported_date], [C_CC _SEX], [patrol_place GROUP SUSP_RACE URKNOWN URKNOWN URKNOWN URKNOWN URKNOWN	SUSP_SEX N UNKNOWN UNKNOWN N UNKNOWN N UNKNOWN	wij, (C_DES_Code), (C_DESC.	I, [C_status], [( SEX], [Latitud 25-44 45-64 UNROXOWN UNROXOWN 18-24 UNROXOWN 18-24 UNROXOWN	C_Law_DESC) e],[Longitu ULC_RACE BLACK HISP ASIAN/PACI UNDACWIN UNDACWIN UNDACWIN	I, [P_name], [P Ide] NIIC F FICISLANDER F E C E	_around_pre	m] 5e Longhude -73.85 -74.01 -73.84 -73.94 -73.94	>
0.	SELECT [I] , [Pres_ty] FROM [Dar FROM [Dar P_name BROOKLYN BROOKLYN BROOKLYN BROOKLYN BROOKLYN BROOKLYN	SCR4C2(CLEVO) cript for Se D_Crime], [C_1 re_desc], [SU ta_WYPD_CRIMI ages P_around_prom NESIDE FROMT OF NESIDE INSIDE FROMT OF NESIDE FROMT OF	LectTopNRov Date], [C_T3 P_AGE_GROL E]. [dbo]. [C Prem_hype_dk RESIDENCE RESIDENCE STREET STREET RESIDENCE STREET RESIDENCE STREET	ine], [Precir IP], [SUSP_RA Crimes_after SSC - APT. HOUSE - APT. HOUSE E - APT. HOUSE	susp_Age_( UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	Reported_date},[C_CC _SEX],[patrol_place   	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	*), (C_DES_Code), (C_DESC 	(,[C_status]).[I SEX],[Latitude VIC_AGE_GROUP 25-44 4544 UNRNOWN UNRNOWN UNRNOWN UNRNOWN 15-24 UNRNOWN 25-44	C_Law_DESC] a], [Longitu UK_RADE BLACK HISP) ASIAN (PACI UNRXOWN UNRXOWN BLACK UNRXOWN BLACK	I, (P_name), (P ide) NIC P FC ISLANDER F E E E E	AC_SEX Lamo AC_SEX Lamo 40.82 40.84 5 40.85 6 40.85 40.83 4 40.64 4 40.64	m] 50 Longhuide -73,86 -74,01 -73,84 -73,94 -73,94	>
R	Annyllagi SELECT [II] FROM [Dark FROM [Dark G ~ C P_Dame BROKX BROXX BR	SCR4C2(CLEVO ( cript for Sel Crime], [C_] pe_desc], [SUU ta_NYPD_CRIM ages P_around_prem mesiDE FRONT OF mesiDE FRONT OF MESIDE FRONT OF MESIDE FRONT OF MESIDE	LectTopNRow Date].[C_T3 SP_AGE_GROL E].[dbo].[C Prom_hype_dk RESDENCE STREET CHAIN STOR STREET RESDENCE STREET STREET STREET STREET	ime], [Precir IP], [SUSP_RF Crimes_efter SSO - APT. HOUSE E - APT. HOUSE E	Ct_no], [R CE], [SUSP _Process] UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	Reported_date],[C_CC P_SEX],[patrol_place UNENDOWN UNENDOWN UNENDOWN UNENDOWN UNENDOWN UNENDOWN UNENDOWN UNENDOWN UNENDOWN	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	wij, (C_DES_Code), (C_DESC.	VIC_AGE_GROUP 2544 4544 UNROXOWN UNROXOWN UNROXOWN UNROXOWN UNROXOWN 2544 2544 2544	C_Law_DESC] e], [Longitu VIC_RADE BLACK HISP- ASIAN FRACE UNROXVMN BLACK UNROXVMN BLACK WHITE	, (P_name), (P ide) NHC F FROISLANDER F E E F	*_around_ord AC_SEX_La900 40.03 40.04 04.065 04.065 04.065 40.05 4	m] 50 Longitude -73.85 -74.01 -73.84 -73.94 -73.91 -73.92 -73.92 -73.94	>
	Anny10agi SELECT [II] FROM [Oan-Ty] FROM [Oan-Ty] FROM [Oan-Ty] BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN	SKR4C2(CLEVO) cript for Sel _Crime],[Clu e_desc],[Clu ta_NVPD_CRIMI RestDE P_around_prom RestDE FRONT OF INSIDE FRONT OF INSIDE FRONT OF INSIDE FRONT OF UNKNOWN UNKNOWN	LectTopNRov Date], [C_T3 P_4GE_GROL E], [dbo], [C Prom_type_dk RESIDENCE STREET RESIDENCE STREET RESIDENCE STREET STREET STREET STREET STREET	ine], [Precir IP], [SUSP_RF Crimes_after SSO - APT. HOUSE E - APT. HOUSE E - APT. HOUSE	htt_no], (R KCE), (SUSP_AGE_( Process) SUSP_AGE_( UNROUWN UNROUWN UNROUWN UNROUWN UNROUWN UNROUWN UNROUWN <18	Seported_date},[C_CC P_SEX],[patrol_place Innonown LINNOWN LINNOWN LINNOWN LINNOWN LINNOWN BLACK LINNOWN LINNOWN LINNOWN LINNOWN LINNOWN BLACK HSYMMC	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN N UNKNOWN UNKNOWN UNKNOWN N	w), (C_DES_Code), (C_DESC 	I,[C_status],[1 SEX],[Latitudy 25-44 45-64 UNROXOWN UNROXOWN UNROXOWN UNROXOWN UNROXOWN 25-44 25-44 25-44 418	C_Law_DESC] J.(Longitu VIC_RACE BLACK HEP BLACK HEP BLACK UNROOVIN BLACK WHITE BLACK WHITE BLACK	I, [P_name], [P de]	AC_SEX Lation ACC_SEX Lation	m] 20 Longitude -73,85 -74,01 -73,84 -73,94 -73,94 -73,94 -73,94 -73,89	,
	Arry10.sql SELECT [I] PROM [Dar FROM [Dar BROKLYN BROKLYN MANHATTAN BROKLYN MANHATTAN BROKLYN MANHATTAN BROKLYN	SKR4C2(CLEVO) cript for Se D_Crime], [CJ] so_desc], [SU so_desc], [SU so_desc], [SU sound_pren NSIDE FRONT OF NSIDE FRONT OF NSIDE FRONT OF NSIDE INSIDE	LectTopNRov Jate]. (C_Ti SP_AGE_GROL []. (dbo]. [C Prom.type_dc RESDENCE STREET CHAIN STOR STREET STREET STREET STREET STREET STREET STREET	ime], [Precir IP], [SUSP_RF Inimes_efter -APT. HOUSE -APT. HOUSE E -APT. HOUSE	Inct_no], (R ICE], (SUSP_AGE_0 SUSP_AGE_0 UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION UNIXION X18 UNIXION	Seported_date);[C_CC _SEX);[partol_place] GROUP SUSP.FACE LNANDOWN LNANDOWN LNANDOWN LNANDOWN BLACK LNANDOWN BLACK HISPAGE UNNOWN	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	(C_DES_Code), (C_DESC 	VC_AGE_GROUP 25-44 UNRXXXVN UNRXXXVN UNRXXXVN UNRXXXVN UNRXXXVN 25-44 UNRXXXVN 25-44 45-64	C_Law_DESC) e),(Longitu VIC_RACE BLACK HISP) ASNAV PACI UNRXXVM UNRXXVM UNRXXVM UNRXXVM UNRXXVM UNRXXVM UNRXXVM BLACK WHITE BLACK	I, [P_name], [P ide]	"_around_pre	m] -73.86 -73.84 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94	>
	Anny10agi SELECT [II] PROM [Dars] FROM [Dars] FROM [Dars] Promains () Mossi BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN	SRA4C2VCLEVO ( cript for 5e) _crime).[c] _gcime).[c] _edes).[s] edes).[s] edes).[s] edes).[s] edes).[s] edes).[s] edes).[s] edes).[s] FROMT of mesiDE FROMT of mesiDE	LectTopNRov Jate]. [Ti prKGE_GROL ]. [dbo]. [ Prom_hypo_dk RESIDENCE STREET CHAIN STOR STREET STREET RESIDENCE STREET STREET RESIDENCE OTHER	ine], [Precir IP], [SUSP_RA rimes_after -APT. HOUSE E -APT. HOUSE HOUSE	htt_no], (R KE), (SUSP_AGE_S SUSP_AGE_S UNROWN UNR	Apported_date)_[C_CC 2_SEX](patrol_place UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN UBNOWN	SUSP_SEX N URKNOWN URKNOWN URKNOWN URKNOWN URKNOWN N URKNOWN N URKNOWN N URKNOWN	b) (C. DES. Code) , (C. DESC. , GROUP) , (VIC., BACE) , [VIC. , GROUP) , (VIC., BACE) , [VIC. , DEVEND, BOOD BRONK PATEND, B	I,[C_status],[1 SEX],[Latitudi 25:44 45:44 UNRNOVIN UNRNOVIN UNRNOVIN UNRNOVIN UNRNOVIN UNRNOVIN UNRNOVIN 25:44 25:44 45:64	C_Law_DESC] VIC_RACE BLACK HISP ASAM (PAC) UNRNOVIM UNRNOVIM BLACK WHITE BLACK WHITE BLACK WHITE BLACK	I, [P_name], [P ide]	"_around_pre- AC_SEX_La%u = 4005 = 40	m] 50 Longitude -73,85 -73,84 -73,84 -73,94 -74,94 -74,	>
	Arry10.sql SELECT [I] PROM_DIA PROM_DIA PROM_DIA BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN BROCKLYN	SKR4C2VCLEVO ( cript for 5e 5c pc_dse_l(c) pc_dsec_l(c_) pc_dsec_l(c_) pc_dsec_l(c_) pc_dsec_l(c_) pc_dsec_l(c_) pc_dsec_l(c_) nestDe nestDe nestDe FRONT OF nestDE FRONT OF nestDE FRONT OF nestDE FRONT OF INSIDE FRONT OF INSIDE FRONT OF INSIDE FRONT OF UNRYOWN	LectTopNRov Jate].[C_Ti SP_AGE_GROL [].[dbo].[C Prom_type_dk RESIDENCE RESIDENCE STREET RESIDENCE STREET RESIDENCE STREET RESIDENCE OTHER OTHER	Ime], [Precir IP], [SUS=_RA IP], [SUS=_RA IP], SUS=_RA IP] IP] IP] IP] IP] IP] IP] IP] IP] IP]	htt_no], (R KE), (SUSP_AGE_S SUSP_AGE_S UNROWN UNROWN UNROWN UNROWN UNROWN UNROWN <18 UNROWN UNROWN UNROWN UNROWN UNROWN UNROWN	Seported_date)_[C_CC Sec[]_SEX],[patrol_place UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP UNENCOMP	SUSP_SEX N URKNOWN URKNOWN URKNOWN URKNOWN URKNOWN N URKNOWN N N N	a) [, C, DES_COG4), [, C, DESC , GROUP], [, VIC_RACE], [, VIC_ , GROUP], [, VIC_RACE], [, VIC_ , DESC ,	VIC_AGE_GROUP SEX],[Latitude VIC_AGE_GROUP 25-44 45-64 UNROXOWN UNROXOWN UNROXOWN 25-44 418 45-64 45-64 45-24	C_LAW_DESC) VIC_RADE BLACK HISPA ASIAN (PACI UNRXXVIM BLACK UNRXXVIM BLACK WHITE BLACK WHITE BLACK WHITE BLACK WHITE BLACK	I, [P_name], [P ide]	around_pro	m] 5 Longitude -73,85 -74,01 -73,34 -73,34 -73,34 -73,34 -73,34 -73,34 -73,34 -73,34 -73,34 -73,39 -73,39 -74,04 -74,04 -74,99 -73,99	>
	SELECT [I] SELECT [I] FROM [Darborner] FROM [	SCH4C2VCLEVO ( cript for Seion Seion Seidesch (SUS) seidesch (SUS) seidesch (SUS) seidesch (SUS) seidesch (SUS) Pianut (SUS) Pianut (SUS) RestDE	LectTopNRov Jate) (Ti y_AGE_GROUE) [ (dbo) . (c Pron_hype_dk RESDENCE STREET CHAIN STOR STREET STREET STREET STREET STREET STREET STREET STREET STREET OTHER STREET OTHER STREET	Ime], [Pracir Ing], [SUSP_RA Primes_after so -APT HOUSE E HOUSE	htt_no], (R KE), (SUSP SUSP.AGE ( UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN	Reported_date)_[[_cc sEK],[patrol_place GROUP SUSP_PACE LUNNOWN LUNNOWN LUNNOWN LUNNOWN LUNNOWN LUNNOWN LUNNOWN BLACK HEFMAG	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN	b) [C 265, Co44], [C 2655; GROUP], [VTC_SKC5], [VTC pHrd_sRee PHTRL, BobD BROKK PHTRL, BOBD BROKK P	VIC_AGE_GROUP 25-44 UNRNOWN UNRNOWN UNRNOWN 15-24 UNRNOWN 15-24 UNRNOWN 25-44 25-44 45-64 45-64 45-64 10-24 UNRNOWN	VIC_RACE BLACK HISP ASAM (PACK UNROXVM UNROXVM UNROXVM UNROXVM BLACK WHITE AMERICAN II BLACK WHITE AMERICAN II UNROXVM	I, [P_name], [P ide]	"_around_pre AC_SEX_La90u = 4084 = 4085 = 4085 = 4085 = 4085 = 4085 = 4085 = 4086 = 4086 = 4086 = 4086 = 4086 = 4086 = 4086 = 4086 = 4086 =	m] b Longitude -73.88 -74.01 -73.84 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -74.09 -73.94	>
	Arry10.sql SELECT [11] FROM [Dar FROM [Dar FROM [Dar FROM [Dar FROM [Dar BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN BROCK:YN	SKR4C2VCLEVO ( cript fon 5e: pc_dse_l(c_) pc_dse_l(c_) pc_dse_l(c_) ps_dse_l(c_) ps_dse_l(c_) ps_dse_l(c_) mside mside mside mside mside from of mside from of mside from of mside from of mside from of mside from of mside from of mside from of mside from of mside from of mside	LectTopNRov Jate], (T1 Pron_type_dk RESOURCE STREET RESOURCE STREET STREET RESOURCE STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET	Ine]; [Precir Ines_rimes_after rimes_after - APT HOUSE - APT HOUSE HOUSE	htt_no], (R CE), (SUSP_AGE, SUSP_AGE, UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN UNNOWN	SEC) (Strong Strong Str	SUSP_SEX M UNKNOWN UNKNOWN UNKNOWN M UNKNOWN M UNKNOWN M UNKNOWN M UNKNOWN M UNKNOWN M UNKNOWN M	(c) (255, Codd); (c) (255; (c) (257, Codd); (c) (255; (c) (c) (c) (c) (c) (c) (c) (c) (c) (c)	I, [C_status],[Latitud SEX],[Latitud 25-44 45-54 10000000 10000000 15-24 418 45-64 418 45-64 45-664 45-664 45-664 45-664 45-664 45-664 45-664 45-664 45-6664 45-6664 45-6666666666	C_Law_DESC] (Longitu VIC_RADE BLACK HIPS ASIAN (PACI UNRXXVIN UNRXXVIN BLACK WHITE BLACK WHITE BLACK WHITE BLACK UNRXXVIN WHITE BLACK	I, (P_name), (P ide) NNIC P PCISIANDER F C NDIANNALAS	*_around_pre	m] 50 Longluide -73.85 -74.01 -73.84 -73.94 -73.94 -73.94 -73.89 -73.89 -74.04 -73.90 -73.90 -73.94 -73.91 -73.94	>
214 23 34 45 55	SELECT [I] SELECT [I] FROM [Dar FROM [Dar FROM [Dar FROM [Dar BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN BROCK1YN STATENIS_ BROCK1YN STATENIS_	SCR4C2VCLEVO ( cript for Seiderson	LectTopNRov Jate], [13 Pron_hype_dk [], [diso], [C [], [diso], [diso	Ime) _ (Precision) JP) , (SUSP_RA Infimes_after sec _APT HOUSE _APT HOUSE _APT HOUSE HOUSE HOUSE REMARKET	htt_no], (R KCE], (SUSP SUSP_AGE_C UNROWN	Seported date ) [[_cc sEK] [patrol_place LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN LUNCOWN	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN M UNKNOWN M UNKNOWN M N N N N N N N N N N	(c) 265 _code), (c) 265C (c) 265 _code), (c) 265C (c) 26	I, [C_status], [Lititud SEX], [Latitud 23-44 45-44 UNONOON UNONOON UNONOON UNONOON 23-44 45-64 45-64 45-64 45-64 45-64 45-64 45-64	C_Law_DESC] VIC_FADE BLACK HISP ASIM/ IPAC UNROXVIN BLACK WHITE BLACK WHITE BLACK WHITE BLACK	I, (P_name), (P de) NNIC I FICISLANDER F E E E E E E E E E E E E E E E E E E E	"_around_pro	m] 3e Longstude -73.86 -73.84 -73.84 -73.94 -73.94 -73.94 -73.94 -73.89 -73.	>
210 60 7	SELECT [II SELECT [II SELECT [II FROM_UN FROM [Dave BROCK] BROCK [P BROCK] BROCK YN BROCK YN	SCH4C2VCLEVD ( tript for 5e) cript for 5e) FRONT OF FRONT OF	LectTopNRov Jate], (T3 Prom_type_dk E]_(dbo). (d E]_(dbo). (d E]_(db	Ime)_(Precision) (SUSP_RA rimes_after are - APT HOUSE E HOUSE RMARKET	htt_no], [8 ct_no], [8 susp_Ade_s unnown u	Seported_stc)_[C_CC	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN M UNKNOWN M UNKNOWN M N N N N N N N N N N	a) (C, DES, Code), (C, DESC, Gendra), (VCC, SeeE), (VCC, Merchan, C, SeeE), (VCC, Merchan, C, SeeE), (VCC, Merchan, Desch Bistone, Martice, Boold Bistone, Martice, Martice, Boold Bistone, Martice, Boold Bistone, Martice, Martice, Boold Bistone, Martice	I, [C_status], [L SEX], [Latitude SEX], [Latitude 23-44 UNNOONN 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44 25-44	C_Law_DESC] C_Law_DESC] (Longitu VIC_RADE BLACK HEPS ASUM/FACI UNROXOM BLACK WHTE BLACK WHTE BLACK WHTE BLACK WHTE HEPSC WHTE HEPSC WHTE HEPSC WHTE HEPSC WHTE HEPSC WHTE HEPSC HEP	I, (P_name), (P ide) NNIC P PCISIANDER F C NDIANNALAS	*_around_pre	m] 52 Longitust -73,85 -73,85 -73,84 -73,84 -73,84 -73,84 -73,84 -73,89 -73,89 -74,04 -73,89 -73,	>
012345573	Setter [1] SELECT [1] SELECT [1] SELECT [1] FROM [Darb 6 ~ C BROKLYN	SKH4C2VCLEVO [ tript for Sei 	LectTopNRov Jate], [C_T3 Pren_hpp_dc absolute], [dbo], [C absolute], [dbo], [C absolute] absolute strater absolute] strater residence strater residence strater residence strater residence strater strater residence absolute absolute food supel other residence absolute absol	Ime) [Precise IP] (SUSP_RA IP] (SUSP_RA I	cct_no], [R (cc), [SUSP_ Process] SUSP_AGE ( UNNOWN UNNO UNNO	SEC)         SEC)         SEC)           SEC)         SEC)         SEC           GROUP         SUMPLANCE         NAME           UNNOWNY         UNNOWNY         UNNOWNY	SUSP_SEX N UNKNOWN UNKNOWN UNKNOWN UNKNOWN UNKNOWN M UNKNOWN M UNKNOWN M N N N N N N N N N N	w), (C, 265, (Cele), (C, 2652, (C, 2652, (Cele), (C, 2652, (C, 2652, (Cele)), (VZC, 2662), (VZC, 2660), (VZC, 2660, (S, 2652, VZC, 2660, (S, 2652, (	), [C_status], [L SEX], [Latitud SEX], [Latitud UNROXON UNROXON 15-24 UNROXON 15-24 UNROXON 25-44 45-4 4 45-4 4 45-4 4 45-4 4 45-4 4 45-4	C_LAW_DESC[] ), [Longitu VIC_RACE BLACK HISP ASIAN (PACI UNR020VM BLACK UNR020VM BLACK UNR020VM BLACK WHITE BLACK WHITE AMERICAN II BLACK WHITE HISP UNR020VM WHITE HISP UNR020VM	I, (P_name), (P de) NNIC I FICISLANDER F E E E E E E E E E E E E E E E E E E E	*_around_prv *_4005 * 4082 * 4084 * 4084 * 4084 * 4084 * 4085 * 4065 * 4065 * 4065 * 4065 * 4065 * 4065 * 4065 * 4055 *	m] 5 Longitude -73.85 -74.01 -73.84 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.94 -73.85 -73.85 -73.85 -73.85 -73.94	>
00,00,00,00,00,00,00,00,00,00,00,00,00,	SECURITY ID and Security ID and Security ID and Security FROM (David) Security FROM (David) FROM	SCH4C2VCLEVD ( cript for 5e) cript for 5e) FRONT OF HEADE FRONT OF HEADE HEA	LectTopNRov Jate], (C_73 Pres, Jype, de Resources) CHAIN STORE STREET CHAIN STORE STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET STREET COTHER STREET STREET COTHER STREET FOOD SUPEI COTHER STREET FOOD SUPEI COTHER STREET FOOD SUPEI COTHER STREET FOOD SUPEI COTHER STREET FOOD SUPEI COTHER STREET	Ine) [Precise Prime_rise Prime_rise Ref APT HOUSE E HOUSE HOUSE APT HOUSE APT HOUSE HOUSE	<pre>st_no], [8 cc], [SUSP_AGE_( UNROGWN UNROM</pre>	SR0, (partopace)           SR1, (parto_pace)           SR1, (parto_pace)           SR0, (parto_pace)           UNROWN	<ul> <li>FVIC_AGE</li> <li>SUSP_SEX</li> <li>N</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>URKNOWN</li> <li>N</li> <li>URKNOWN</li> <li>N</li> <li>URKNOWN</li> <li>N</li> <li>N</li> <li>N</li> <li>N</li> <li>N</li> <li>N</li> <li>N</li> <li>N</li> <li>F</li> </ul>	a) (C, DES, Coole), (C, DESC, (C, DES, Coole), (C, DESC, (C, DESC, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC), (C, DESC),	I, [C_status], [Latitude SEX], [Latitude 2544 UNNAONN UNNAONN UNNAONN UNNAONN 2544 4544 4544 4544 4544 4544 4544 45	C_Law_DESC] C_Law_DESC] (Longitu VIC_RADE BLACK HEB BLACK HEB BLACK WHTE BLACK WHTE HACK WHTE WHTE WHTE	I, (P_name), (P Ide)	Action of the second se	m] 50 Longitude -73.385 -74.011 -73.384 -73.391 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -74.092 -74	>
012345573	Setter [1] SELECT [1] SELECT [1] SELECT [1] FROM [Darb 6 ~ C BROKLYN	SKH4C2VCLEVO [ tript for Sei 	LectTopNRov Jate], (C_T1 Pre, Jpc, Jk RESDENCE E], (dbo), (C E], (dbo), (C E), (dbo),	Ime)_[Precise IP}], [SUSP_RA Prime_after SO APT HOUSE -APT HOUSE HOUSE HOUSE HOUSE HOUSE	cct_no], [R (cc), [SUSP_ Process] SUSP_AGE ( UNNOWN UNNO UNNO	SEC)         SEC)         SEC)           SEC)         SEC)         SEC           GROUP         SUMPLANCE         NAME           UNNOWNY         UNNOWNY         UNNOWNY	IVEC_AGE SUSP_SEX M URKNOWN URKNOWN URKNOWN URKNOWN URKNOWN M URKNOWN M URKNOWN M N N N N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN N CHKNOWN CHCHCONN CHKNOWN CHKNOWN CHCHCONN CHKNOWN CHKNOWN CHKNOWN CHKNOWN CHCHCONN CHKNOWN CHKNOWN CHKNOWN CHKNOWN CHKNOWN CHCHCON	w), (C, 265, (Cele), (C, 2652, (C, 2652, (Cele), (C, 2652, (C, 2652, (Cele)), (VZC, 2662), (VZC, 2660), (VZC, 2660, (S, 2652, VZC, 2660, (S, 2652, (	), [C_status], [L SEX], [Latitud SEX], [Latitud UNROXON UNROXON 15-24 UNROXON 15-24 UNROXON 25-44 45-4 4 45-4 4 45-4 4 45-4 4 45-4 4 45-4	C_LAW_DESC[] ), [Longitu VIC_RACE BLACK HISP ASIAN (PACI UNR020VM BLACK UNR020VM BLACK UNR020VM BLACK WHITE BLACK WHITE AMERICAN II BLACK WHITE HISP UNR020VM WHITE HISP UNR020VM	I, (P_name), (P ide) NIIC I INIC I I INIC I I I I I I I I I I I I I I I I I I I	*_around_prv *_4005 * 4082 * 4084 * 4084 * 4084 * 4085 * 4085 * 4085 * 4085 * 4085 * 4085 * 4085 * 4086 * 4086	m] 50 Longitude -73.385 -74.011 -73.384 -73.391 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -73.394 -74.092 -74	

Figure 3: Sample of Data After Preprocessing Phase

#### 3.4 Database Design

In this stage, data was converted into a query-based format suitable for data Loading from the database to DW. Three different DB schemas have been designed and created to fulfill the ETL requirements. Each ERD meets specific needs for decision-makers, and ERD is a well-known process for executing relational database [26]. An ERD contains several entities and connectors that imagine two crucial information; The main tables within the system and the inter-relationships between these tables, which are very important in designing different fact tables in the proposed warehouse. Figure (4) shows the first design for ER. This DB diagram consists of (Crimes, ListCrimesTypes, List\_Places, Info\_Premises, ListSuspectClasses,ListVictimClasses, Patrol) tables with relations(Crime\_Places,Crime\_Patrol,ListPlaces\_InfoPremis es,Crime\_Suspect,Crime\_Victim,Suspect\_Victim).

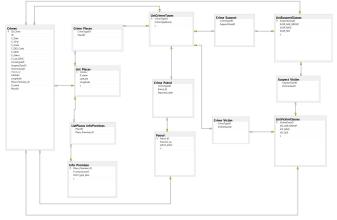


Figure 4: The First Entity Relation Diagram for The Proposed Database

Figure (5) shows the second design of ERD which consist of (Crimes\_Full,List\_Places,ListSuspectClasses,ListVictimClas ses) tables and relations (Crime\_Places, Crime\_Suspect, Crime\_Victim).

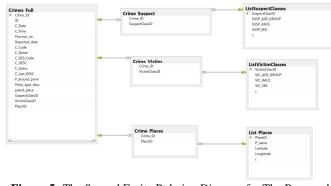


Figure 5: The Second Entity Relation Diagram for The Proposed Database

Figure (6) shows the third design of ERD which consist of ( crimes, ListCrimesTypes, ListPlaces, Info\_Premises, ListSuspectClasses, ListVictimClasses) tables and relations (Crime\_Places,Crime\_Place\_Victim,Crime\_Place\_Victim\_S uspect,ListPlaces\_InfoPremises).

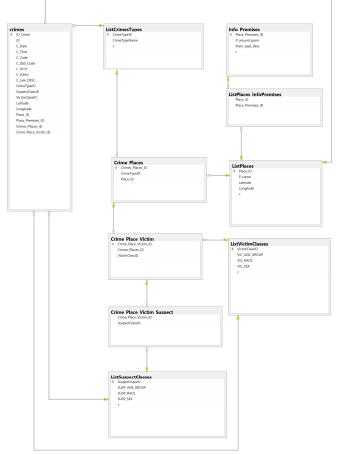


Figure 6: The Third Entity Relation Diagram for The Proposed Database

Haider Alsharqi et al., International Journal of Advanced Trends in Computer Science and Engineering, 9(5), September-October 2020, 9080 - 9088

## 3.5 Data Warehouse Design

In this stage, a galaxy schema data warehouse design which contains various fact connected with multiple dimension that sorted in one level or more (normalized). The different fact tables and dimensions were selected from the database tables and the required analysis.

## 3.5.1 Dimensions

3

4

5

6

7

8

9

10

ListVictimClasses

The dimensions of the DW has been choosing from the data base, the dimensions called the "soul "of DW. Various analysis can be applied on the DW and thus create a number of dimensions such as the type of crime, the time of the crime, the law description for a crime, the crime location name and description it, the patrol that deals with the crime and some specific information about a crime. Table (3) describe each dimension in the proposed data warehouse.

 Table 3: Dimensions Table

 Diminution Table
 Descriptive Attributes

 1
 ListCrimesTypes
 CrimeTypeID, CrimeTypeName

 2
 Crimes
 ID\_Crime,ID,Crime\_Type\_ID,C

 2
 Crimes
 ID\_Crime,ID,Crime\_Type\_ID,C

 2
 Crimes
 ID\_Crime,ID,Crime\_Type\_ID,C

 2
 Example
 ID\_Crime,ID,Crime\_Type\_ID,C

 2
 ID\_Crime,ID,Crime\_Type\_ID,C
 Time,Precinct\_no,Reported\_da

 4
 te,C\_DES\_Code,C\_DESC,Crim
 e\_Status\_ID,Crime\_Law\_Desc\_

 1D,Crime\_date\_ID,Date\_Year\_I
 D,Date\_Month\_ID,Date\_Day\_I
 D

	D,Date_Wonth_ID,Date_Day_I
Crime_Date	Crime_date_ID,Crime_Date,Ye ar,Month_Name,Day_Name,Dat e_Year_ID,Date_Month_ID,Dat e_Day_ID
Crime_Law_Desc	Crime_Law_Desc_ID,Law_Desc ription
Crime_Status	Crime_Status_ID, Crime_Status
ListPlaces	Place_ID, P_name, Latitude, Longitude
ListPlaces_InfoPre mises	Place_ID, Place_Premises_ID
Info_Premises	Place_Premises_ID,P_around_p rem, Prem_type_desc
ListSuspectClasses	SuspectClassID,SUSP_AGE_G ROUP,SUSP_RACE,SUSP_SeX

VictimClassID,VIC\_AGE\_GRO

UP, VIC RACE, VIC SEX

11	Patrol	Patrol_ID,Precinct_no,patrol_pl
		ace
12	Dim_Date_Year	Date_Year_ID, Year
13	Dim_Date_Month	Date_Month_ID, Month
14	Dim_Date_Day	Date_Day_ID, Day_Name

## 3.5.2 Facts

Fact tables represent the place where almost all the analytical processes are applied [27], and hence they contain the dimension key and the measures. In this step of a design data warehouse is to choose carefully, the fact that will appear in the facts table. Can be obtained different reports (analysis). From the fact tables that mentioned in Table (4) such as the crime type with location and suspect, the measure of crime type with location and victims, and the measure of crime type and the Patrol. Each fact table measure (s) is distributed on the timeline (day, month and year).

Table 4: Different DW Fact Tables

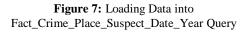
	Fact Table	Descriptive
1	Fact_Crime_Pla ce_Suspect_Dat e_Year	CrimeTypeID, Place_ID, Suspect_ID, Date_Year_ID, Measure
2	Fact_Crime_Pla ce_Victim_Date _Year	CrimeTypeID, Place_ID,Victim_ID, Date_Year_ID, Measure
3	Fact_Crime_Da te_Year_Patrol	CrimeTypeID,Date_Year_ID,Patrol_I D, Measure
4	Fact_Crime_Da te_Month_Patro l	CrimeTypeID,Date_Month_ID,Patrol_ ID, Measure
5	Fact_Crime_Pla ce_Suspect_Dat e_Month	CrimeTypeID, Place_ID, Suspect_ID, Date_Month_ID, Measure
6	Fact_Crime_Pla ce_Victim_Date _Month	CrimeTypeID, Place_ID, Victim_ID, Date_Month_ID, Measure
7	Fact_Crime_Da te_Day_Patrol	CrimeTypeID,Date_Day_ID,Patrol_ID, Measure
8	Fact_Crime_Pla ce_Suspect_Dat	CrimeTypeID, Place_ID, Suspect_ID, Date_Day_ID, Measure

	e_Day	
9	Fact_Crime_Pla	CrimeTypeID, Place_ID, Victim_ID,
	_	Date_Day_ID, Measure
	_Day	

# 3.5.3 Extract, Transform, Load (ETL)

By executing queries in SQL server to extract the data from database and transform the data from tables to in dimensions and facts form and finally loading the data into the data warehouse. Sample of ETL queries is shown in Figure (7).

sef fact\_crime\_p\_3KR4C2ACLEVO[57]) × = INSERT INTO [Fact\_Crime\_Place\_Suspect\_Date\_Year] ([CrimeType\_ID].[Place\_ID].[Suspect\_ID].[Date\_Year\_ID].[Measure]) SELECT [Crime\_Type\_ID].[Place\_ID].[Suspect\_ID].[Date\_Year\_ID]. FROM crimes [GROUP BY [Crime\_Type\_ID].[Place\_ID].[Suspect\_ID].[Date\_Year\_ID]



The galaxy schema of the proposed data warehouse is shown in Figure (8).

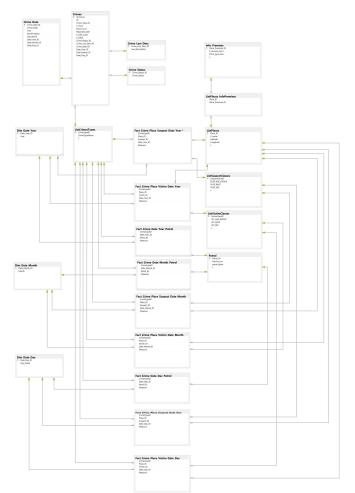


Figure 8: The Proposed Data Warehouse

The proposed warehouse design fulfils the following issues:

## 3.6.1 Data Redundancy

In the proposed data warehouse shown in Figure (8), The data redundancy was eliminated such as (X\_COORD\_CD, Y\_COORD\_CD, Lat\_Lon) columns which were replaced by (Latitude, Longitude) columns. Eliminating the redundancy reduces the data inconsistency, data corruption resulting from errors in writing, reading, storage, or processing data and helping prevent the increase in database size and storage costs.

# 3.6.2 Access flexibility

Ease of access to data and reports needed by a data analyst, through the facts and dimensions in the proposed data warehouse.

## 3.6.3 Enforcing security, ownership and privacy

Data security, privacy and ownership on the proposal data warehouse as shown in Figure (10) can be enforced by dividing the data warehouse into multiple data marts each one has fact with its dimensions according to the required reports. Data privacy and ownership was achieved in different data marts to restrict the access of data and reports for authenticated people.

It is important to notice that the NYPD crime dataset doesn't contain some features like job and education for criminal, which helps get more specific analysis to crime and criminal. This problem cannot be solved even using the most advanced techniques in solving missing values and attributes.

## 4. CONCLUSION AND RECOMMENDATIONS

An optimal galaxy DW for the crime dataset was designed to solve the different features of the analysis including redundancy, access flexibility and enforcing security, ownership and privacy. A structure of data warehouse for crime dataset, as a solution to big data of crime in major cities. We collected real data from New York City (NYPD Complaint Data Historic) available on the internet. The proposed design consists of six stages from collecting data, converting CSV file to SQL table, data preprocessing, design database and design data warehouse finally the analysis. This proposed DW will serve the data analyst and law enforcement forces by getting important analytics, resolve different complex queries as per their need and knowing the complexity of the relationship between crimes with locations, victims and suspects.

It is important to recommend that the proposed data warehouse represent a repository for preprocessed real crime data with huge amount of records (more than six million) available for other research work, and the flexibility of design enables the researcher to add new data and features. Finally, it is of great importance to mention that using a large number of fact tables (nine in this paper) and intern data marts is very important to improve the performance of the analysis algorithms and their complexities.Further, this paper can develop it by using data mining techniques like clustering, classification on the DW to predicted the crime and improve the efforts made by the lawmen and police departments.

# REFERENCES

[1] B. B. Miftachul Huda, Andino Maseleno, Pardimin Atmotiyoso, Maragustam Siregar, Roslee Ahmad, Kamarul Azmi Jasmi, Nasrul Hisyam Nor Muhamad, Mohd Ismail Mustari, "Big Data Emerging Technology: Insights into Innovative Environment for Online Learning Resources," *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 1, pp. 23–36, 2018.

- [2] M. Feng *et al.*, "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," *IEEE Access*, vol. 7. pp. 106111–106123, 2019.
- U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," 2016 2nd Asian Conference on Defence Technology, ACDT 2016. pp. 123–128, 2016.
- [4] H. K. Fatlawi, A. F. H. Alharan, and N. S. Ali, "An efficient hybrid model for reliable classification of high dimensional data using k-means clustering and bagging ensemble classifier," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 24, pp. 8379–8398, 2018.
- [5] S. H. Farhad Soleimanian Gharehchopogh, "An Optimization K-Modes Clustering Algorithm with Elephant Herding Optimization Algorithm for Crime Clustering," *ournal Adv. Comput. Eng. Technol.*, vol. 6, no. 2, pp. 78–87, 2020.
- [6] D. Xu and Y. Tian, "A Comprehensive Survey of Clustering Algorithms," *Annals of Data Science*, vol. 2, no. 2. pp. 165–193, 2015.
- [7] S. Hussain, R. Atallah, A. Kamsin, and J. Hazarika, "Classification, clustering and association rule mining in educational datasets using data mining tools: A case study," in *Advances in Intelligent Systems and Computing*, 2019, vol. 765, pp. 196–211.
- [8] A. F. H. Alharan, H. K. Fatlawi, and N. S. Ali, "A cluster-based feature selection method for image texture classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3. pp. 1433–1442, 2019.
- [9] C. C. Aggarwal and P. S. Yu, "Data mining techniques for associations, clustering and classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1574. pp. 13–23, 1999.
- [10] X. Wu et al., "Top 10 algorithms in data mining," Knowl Inf Syst, vol. 14, pp. 1–37, 2008.
- [11] R. Ari Setyawan, E. Prasetyo, and A. S. Girsang, "Design and Implementation Data Warehouse in Insurance Company," *Journal of Physics: Conference Series*, vol. 1175, no. 1. 2019.
- [12] M. Subekti, J. Junaidi, H. L. H. S. Warnars, and Y. Heryadi, "The 3 steps of best data warehouse model design with leaning implementation for sales

transaction in franchise restaurant," 2017 IEEE International Conference on Cybernetics and Computational Intelligence, CyberneticsCOM 2017 -Proceedings, vol. 2017-Novem. pp. 170–174, 2018.

- [13] C. N. S.Vijayarani, E.Suganya, "A Comprehensive Analysis of Crime Analysis Using Data Mining Techniques," vol. 9, no. 1. pp. 114–123, 2020.
- [14] M. C. Giuseppe Agapito, Chiara Zucco, "COVID-WAREHOUSE A Data Warehouse of Italian COVID-19, Pollution, and Climate Data." 2020.
- [15] D. B. S. J. CHIRRA ANUSHA, "DESIGN OF A DATA WAREHOUSE FOR MEDICAL INFORMATION SYSTEM USING DATA MINING TECHNIQUES," J. Eng. Sci., vol. 11, no. 6, pp. 308–313, 2020.
- [16] A. R. Quitaleg and M. G. Ortiz, "Design and Development of Data Warehouse Framework of Highland Vegetable Crops for Benguet," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 803, no. 1.
- [17] A. Abdo, H. Fahmy, and A. A. Shaker, "Mining Forensic Medicine Data for Crime Prediction," vol. 17, no. 6. pp. 56–62, 2019.
- [18] I. N. A. Prabawa, D. Agung, K. Arimbawa, and I. G. N. Janardana, "Analysis and Design Data Warehouse For E-Travel Business Optimization," *International Journal of Engineering and Emerging Technology*, vol. 4, no. 1. 2019.
- [19] Y. P. Sudarmojo, "Design of Library Data Warehouse Using OLTP Result of Services Analysis," vol. 3, no. 1. pp. 62–65, 2018.
- [20] N. A. Farooqui and R. Mehra, "Design of a data warehouse for medical information system using data mining techniques," *PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing.* pp. 199–203, 2018.
- [21] T. Y. Choo and H. N. Chua, "Data Warehouse Design and Implementation for Research Literature Mining," Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018. 2018.
- [22] I. Sutedja, P. Yudha, N. Khotimah, and C. Vasthi, "Building a Data Warehouse to Support Active Student Management: Analysis and Design," Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018. pp. 460–465, 2018.
- [23] "NYPD Complaint Data Historic | NYC Open Data." [Online]. Available: https://data.cityofnewyork.us/Public-Safety/NYPD-C omplaint-Data-Historic/qgea-i56i. [Accessed: 14-Oct-2020].
- [24] A. Kumar and S. P. Panda, "A Survey: How Python Pitches in IT-World," in *Proceedings of the International Conference on Machine Learning, Big*

Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019, 2019, pp. 248–251.

- [25] C. Z. & Q. Y. Shichao Zhang, "Data preparation for data mining," *Appl. Artif. Intell.*, pp. 375–381, 2003.
- [26] A. Elfaki, A. Aljaedi, and Y. Duan, "Mapping ERD to knowledge graph," in *Proceedings - 2019 IEEE World Congress on Services, SERVICES 2019*, 2019, pp. 110–114.
- [27] M. R. Ralph Kimball, *The Data Warehouse Toolkit: The Defi nitive Guide to Dimensional Modeling*, Third Edit. John Wiley & Sons, Inc., 2013.