Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
May 2024 5(5), 199–220

FORENSIC INVESTIGATION OF CALL DATA RECORD (CDR) USING STATISTICAL AND MAPPING TOOLS

Gouri Rajendra Uplenchwar [1], Rahul Kailas Bharati [1], Dr. Shobha Kamalakar Bawiskar [1]

[1] Government Institute of Forensic Science, Nipat Niranjan Nagar, Near Caves Road, Aurangabad, Maharashtra, India

ABSTRACT

As technology advances, it appears that the undesirable side effects do as well. It also results in increase in crime rates. It has been noted that many significant crimes are occurring with the use of only smartphones because they are convenient and simple to use due to their GUI characteristics. Police departments and many other law enforcement agencies have used CDR since many years till now to collect the evidence including both civil and criminal cases. People are naturally leaving their footprints behind them. Numerous studies on various aspects of CDR have been on-going for a decade. In this research, the researchers utilized the MS EXCEL and Maptive tools to simply and easily gather satellite- based imagery of a site's precise location in order to analyse CDR.

Keywords: Call Data Record (CDR), Clustering, Cellular Network, Fraud Detection, Autofilter, Maptive, User Behaviour Analysis, Telecom, Logistic Regression, Mobility Analysis, Anomaly Detections.

Corresponding Author

Dr. Shobha Kamalakar Bawiskar,
shobha_bawiskar@yahoo.co.in

DOI
10.29121/shodhkosh.v5.i5.2024.1665

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

1. INTRODUCTION

The information used by the investigative agency to look into any case comes from a variety of sources. One method is call detail recording (CDR). For the purpose of solving any case, it is now standard procedure to examine the call records of the suspect, victim, etc.

The detailed data record of every telephone call that passes through a phone exchange or any other piece of telecommunication's equipment is known as a call detail record, or CDR. The relevant telephone exchange keeps track of the record, which includes information on the call's time, duration, source and destination numbers, completion status, and other details. The main data breakdowns in CDR reports are by user or by phone number. Specific metrics, such as call duration and time of the call for that individual, can be displayed in CDR reports that are only

generated for one user. It's critical to review CDR reports for each individual in order to determine who spent the most time on the phone, how long calls lasted, and how much each person spent per call. The analysis of CDR datasets should centre on the characteristics of the underlying social networks that were discovered through the data. Credit risk management is centred on scenarios in which borrowers might default on their loans. Banks determine each prospective customer's credit score (Doyle et al., 2019). Telecom providers continuously a mass a sizable volume of CDR from which further information can be extracted at little cost and in the form of valuable datasets. Analysis of CDR data can be helpful in a variety of disciplines, including network monitoring, analysing the economic situation in a particular area, conducting socioeconomic research geared toward marketing and the creation of social networks, and modifying the services that are being offered. Mobile

location is a crucial source of data for examining the spatial-temporal dynamics of user communities, but there have been relatively few studies on it due to issues with data availability due to privacy and confidentiality (Bianchi et al., 2016).

The use of Spatial Temporal Recurrent Neural Networks (ST-RNN), which capture periodic temporal contexts, has been proposed by the authors as a new way to improve the performance of the spatial-temporal model. This model has time-specific and distance- specific transitions (Liu et al., 2016). Time stamps, encrypted user IDs, event categories, Location Area Codes (LAC), and cell tower IDs are all included in the raw CDR. Using raw CDR data, a special independent cell tower database provides the positions of cell towers and approximates the locations of people (Leng, 2016). The authors' suggested dynamic Bayesian Network (DBN) model takes into account the patterns of user mobility. The optimal model was chosen for each person and circumstance using three separate criteria: entropy, highest probability, and ensemble. The inspiration for the DBN model came from common daily mobility patterns, and the three key features are the day of the week, the time of day, and their combination(Dash et al., 2015). The authors made predictions about future travel patterns based on the recommendations of travellers, using detailed sets of information for each tourist (N. C. Chen et al., 2017)

Types of CDR

1) **IPDR:** IPDR is used to measure IP network traffic.
2) **SDR:** Subscriber Data gives the name of a subscriber, their address, telephone number, and any other information that is used to identify a subscriber of a communications services provider.

CDR is now admissible in court as proof. Electronic records produced in court are admissible as evidence. In accordance with section 65 (B) of the Indian Evidence Act, 1872, a CDR must be produced together with a certificate.

CDR can only be acquired by law enforcement. Police officers with the level of Superintendent of Police (SP) and higher are permitted by the court to request CDR from any telecom company whenever there is a legal matter. They ask the telecom firm for a CDR and provide them with case numbers, FIR numbers, etc. Every month, police officers are required to report to the district magistrate (DM) the CDR they have collected. After receiving a request from a law enforcement agency or court, telecom firms respond with CDR information within an hour. Under Licensing Condition No. 39.20 of the License Agreement for Unified License, operators in India are permitted to maintain customer records, including CDR databases and details of Internet Protocol (IP), for at least the previous year. CDR cannot be accessed or obtained by anybody because it needs judicial approval because it is used to demonstrate one's identity.

CDR Related Offences and legal aspects

1) There are different crimes which require to seek CDR data from Telecom Company to investigate case further.

   For example crimes like Telephonic Harassment, Murder, Stalking, Kidnapping/Abduction, Blackmailing, Cybercri me, Riot, Trafficking, Criminal Trespassing, Bigamy / Infidelity, Rape

2) There are certain crimes which may be perform on CDR data. For example, crimes like Breach of privacy of person whose CDR data was requested, Breach of confidentiality, Breach of integrity, tampering with CDR data i.e., data manipulation, Framing of CDR data as such for not finding culprit

- **IT-2000 ACT**

Anyone who commits a crime using digital evidence (CDR) may be held accountable under the following sections.

1) **Section 43**: If someone is harming a computer, computer system, or computer network in an effort to destroy or wipe any data or digital proof, this section applies.
2) **Section 65:** If somebody tampers with computer source documents.
3) **Section 72:** This section specifies penalties for violating privacy or secrecy.
4) **Section 72(A):** This section imposes penalties for disclosing any material in violation of a valid contract for the purpose of causing someone else harm or unjust gain.
5) **Section 76:** Anyone who confiscates a computer or other electronic device that contains or is in possession of digital evidence is responsible for violating this section.

- **Indian Evidence Act, 1872**

1) **Section 65(B):** Admissibility of any electronic evidence in court.
2) **Section 85(B):** Presumption as to any electronic records and electronic signature.

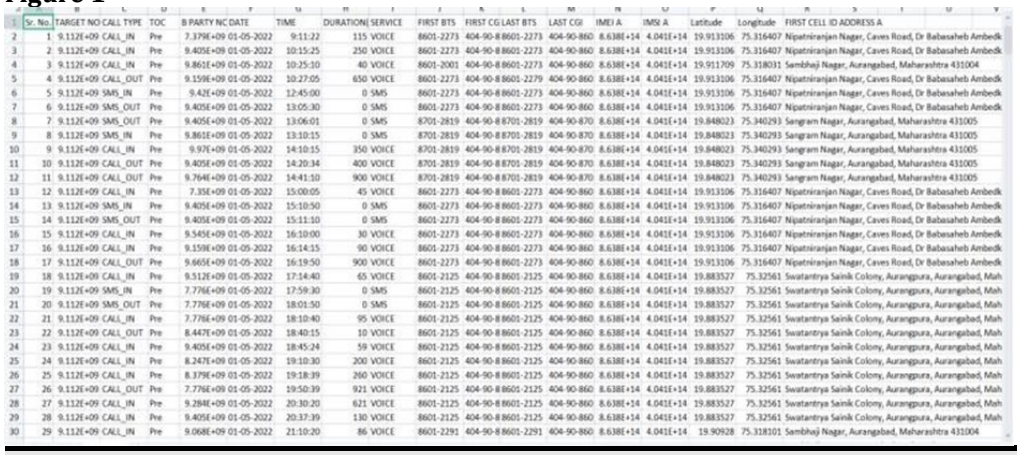
AIM: Analysis of CDR database with various tools

OBJECTIVE:

1) Creation of CDR database.
2) Analysis of CDR using: MS-EXCEL.
3) Prediction of location by using Logistic Regression (Machine Learning Technique).

Example of Raw and processed CDR: Raw CDR is unformatted CDR got from telecom company. Large number from raw CDR is not easily understandable by anyone as shown.

Figure 1



Sr. No.	TARGET NO	CALL TYPE	TOC	B PARTY NC	DATE	TIME	DURATION	SERVICE	FIRST BTS	FIRST CG	LAST BTS	LAST CGI	IMEI A	IMSI A	Latitude	Longitude	FIRST CELL ID ADDRESS A
1	9.112E+09	CALL_IN	Pre	7.379E+09	01-05-2022	9:11:22	115	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
2	9.112E+09	CALL_IN	Pre	9.405E+09	01-05-2022	10:15:25	250	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
3	9.112E+09	CALL_IN	Pre	9.861E+09	01-05-2022	10:25:10	40	VOICE	8601-2001	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.911709	75.318031	Sambhaji Nagar, Aurangabad, Maharashtra 431004
4	9.112E+09	CALL_OUT	Pre	9.159E+09	01-05-2022	10:27:05	650	VOICE	8601-2273	404-90-8	8601-2279	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
5	9.112E+09	SMS_IN	Pre	9.42E+09	01-05-2022	12:45:00	0	SMS	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
6	9.112E+09	SMS_OUT	Pre	9.405E+09	01-05-2022	13:05:30	0	SMS	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
7	9.112E+09	SMS_OUT	Pre	9.405E+09	01-05-2022	13:06:01	0	SMS	8701-2819	404-90-8	8701-2819	404-90-870	8.638E+14	4.041E+14	19.848023	75.340293	Sangram Nagar, Aurangabad, Maharashtra 431005
8	9.112E+09	SMS_IN	Pre	9.861E+09	01-05-2022	13:10:15	0	SMS	8701-2819	404-90-8	8701-2819	404-90-870	8.638E+14	4.041E+14	19.848023	75.340293	Sangram Nagar, Aurangabad, Maharashtra 431005
9	9.112E+09	CALL_IN	Pre	9.97E+09	01-05-2022	14:10:15	350	VOICE	8701-2819	404-90-8	8701-2819	404-90-870	8.638E+14	4.041E+14	19.848023	75.340293	Sangram Nagar, Aurangabad, Maharashtra 431005
10	9.112E+09	CALL_OUT	Pre	9.405E+09	01-05-2022	14:20:34	400	VOICE	8701-2819	404-90-8	8701-2819	404-90-870	8.638E+14	4.041E+14	19.848023	75.340293	Sangram Nagar, Aurangabad, Maharashtra 431005
11	9.112E+09	CALL_OUT	Pre	9.764E+09	01-05-2022	14:41:10	900	VOICE	8701-2819	404-90-8	8701-2819	404-90-870	8.638E+14	4.041E+14	19.848023	75.340293	Sangram Nagar, Aurangabad, Maharashtra 431005
12	9.112E+09	CALL_IN	Pre	7.35E+09	01-05-2022	15:00:05	45	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
13	9.112E+09	SMS_IN	Pre	9.405E+09	01-05-2022	15:10:50	0	SMS	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
14	9.112E+09	SMS_OUT	Pre	9.405E+09	01-05-2022	15:11:10	0	SMS	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
15	9.112E+09	CALL_IN	Pre	9.545E+09	01-05-2022	16:10:00	30	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
16	9.112E+09	CALL_IN	Pre	9.159E+09	01-05-2022	16:14:15	90	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
17	9.112E+09	CALL_OUT	Pre	9.665E+09	01-05-2022	16:19:50	900	VOICE	8601-2273	404-90-8	8601-2273	404-90-860	8.638E+14	4.041E+14	19.913106	75.316407	Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedk
18	9.112E+09	CALL_IN	Pre	9.512E+09	01-05-2022	17:14:40	65	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
19	9.112E+09	SMS_IN	Pre	7.776E+09	01-05-2022	17:59:30	0	SMS	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
20	9.112E+09	SMS_OUT	Pre	7.776E+09	01-05-2022	18:01:50	0	SMS	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
21	9.112E+09	CALL_IN	Pre	7.776E+09	01-05-2022	18:10:40	95	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
22	9.112E+09	CALL_OUT	Pre	8.447E+09	01-05-2022	18:40:15	10	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
23	9.112E+09	CALL_IN	Pre	9.405E+09	01-05-2022	18:45:24	59	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
24	9.112E+09	CALL_IN	Pre	8.247E+09	01-05-2022	19:10:30	200	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
25	9.112E+09	CALL_IN	Pre	8.379E+09	01-05-2022	19:18:39	260	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
26	9.112E+09	CALL_OUT	Pre	7.776E+09	01-05-2022	19:50:39	921	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
27	9.112E+09	CALL_IN	Pre	9.284E+09	01-05-2022	20:30:20	621	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
28	9.112E+09	CALL_IN	Pre	9.405E+09	01-05-2022	20:37:39	130	VOICE	8601-2125	404-90-8	8601-2125	404-90-860	8.638E+14	4.041E+14	19.883527	75.32561	Swatantrya Sainik Colony, Aurangpura, Aurangabad, Mah
29	9.112E+09	CALL_IN	Pre	9.068E+09	01-05-2022	21:10:20	86	VOICE	8601-2291	404-90-8	8601-2291	404-90-860	8.638E+14	4.041E+14	19.90928	75.318101	Sambhaji Nagar, Aurangabad, Maharashtra 431004

Figure 1 Raw CDR

2. LTERATURE REVIEW

SR. NO.	Title Of Paper	Year Of Publication	Authors	Methodology	Advantage	Disadvantage
1	Crime Investigation and Criminal Network Analysis Using Archive Call Detail Records (Kumar Et Al., 2017)	2016	Dr. Manish Kumar, Dr. M. Hanuma nthappa, Dr. T. V. Suresh Kumar	Author Proposed Graph Based Analytical Model Using CDR Data Where Pictorial Graph Presentation Can Show How Phone or IMEI Numbers Are Connected by Other Phone Numbers.	By Using Graph Analysis Tool, It Is Easy to Discover Insights in Complex Data.	There Are No Such Techniques to Analyse CDR Along With Social Network Profile.
2	Estimation and monitoring of city-to-city travel times using call detail records (Kujala et al., 2016)	2016	Rainer Kujala, Talayeh Aledavod , Jari Saramäki	CDR data was taken. Determine its typical travel times by computing all inter-observation time distribution from mobility trajectories of different users and smooth the distribution. Estimation of biases to correct it and also estimation error decreases with number of data points.	Ease to extract typical travel time between cities from mobility data of CDR. Gives better spatial and temporal accuracy of data.	Amount of data increased to estimate correct travel time but it also to increase high level of noise Errors in data can give corrupted results.
3	Predicting complex user behaviour from CDR based social networks (Doyle et al., 2019)	2019	Casey Doyle, Zala Herga, Stephen Dipple, Boleslaw K. Szymans ki, Gyorgy Korniss, Dunja Mladenic	Generation of full feature set. Point biserial correlation is used to define correlation between each feature. Use of logistic regression and Principal Component Analysis to build model. Then simple method used for	This model used to investigate different aspects of user behaviour.	To predict realistic result, require large data from longer time.

				feature space reduction.		
4	Identifying hidden visits from sparse call detail record data (Zhan Zhao et al., 2021)	2021	Zhan Zhao, Haris N. Koutsop oulos,Jin hua Zhao	Extract trip from CDR which include stages-localization, movement state inference, hidden visit inference. Hidden visit inference on basis of data fusion. Preprocessing should be done on data. Spatial, temporal, personal attributes categorized in model specification. Logistic regression model was used and feature was analysed, comparison of accuracy different model. Model deployment for trip extraction.	Authors developed data fusion approach to conclude the existence of hidden visit in CDR data.	It is challenging to conclude if such hidden visits exist, when and where it occurs i.e. unable to conclude spatiotemporal patterns of hidden visits for one user.
5	An Efficient Approach of Identifying Tourist by Call Detail Record Analysis (Sikder et al., 2017)	2016	Ratul Sikder, Md. Jamal Uddin, Sajal Halder	Efficient scan for CDR database which is time interval scan for cellular phones location. Tourists' identification by using CDR location data - collect network latitude and longitude with unique identity. Efficient scan of CDR database after time interval and fetch location data with cellular identity and tourists' identification by using CDR location data.	Given methodology is used to identify tourists by using CDR data. This methodology also helps tourist on basis of SMS system. This framework was used to detect long travelled tourists.	This framework is not supported for tourists' detection history which can be large source of data.
6	Relevance of Context for the Temporal Completion of Call Detail Record (G. Chen et al., 2016)	2016	Guangsh uo Chen, Sahar Hoteit, Aline Carneiro Viana, Marco Fiore, Carlos Sarraute	Fine-grained and coarse -grained datasets collected from cellular network and apply recursive look- ahead filter to it. Categorise users into rare CDR and frequent CDR. To increase accuracy authors did comparative study between coarse-grained and ground-truth datasets.	Proposed framework shows capability by using location boundaries of identifying user's locations. Identification of key challenge of CDR based human mobility analysis regards its accuracy. Identification of fixed period location	This framework is not adaptive approach to estimates location and particular time period.

					boundary and particular time period regarding spatial error.	
7	Understanding the bias of call detail records in human mobility research (Ziliang Zhao et al., 2016)	2016	Ziliang Zhao, Shih-Lung Shaw, Yang Xu, Feng Lu, Jie Chen, Ling Yin	Datasets were collected and authors analysed individual human mobility patterns based on total travel distance, radius of gyration and movement entropy. Total travel distance is calculated using sum of Euclidian distance between each pair of consecutive footprints. Radius of gyration is root mean squared distance between set of visited location up to time t and centre of mass. Movement of entropy measures heterogeneity of visited patterns.	Assessing the validity using CDR to estimate human mobility patterns. Helpful to investigate distance decay effect and analyses how CDR can lead to biased understanding of urban dynamics.	Aggregate urban interaction pattern from CDR can be biased and misleading.
8	A Hierarchical Approach for Identifying User Activity Patterns from Mobile Phone Call Detail Records (Khan et al., 2015)	2015	Fahim Hasan Khan, Moham med Eunus Ali, Himel Dev	Authors have proposed hierarchical model with multiple layers for processing CDR for identifying mobility patterns and user activities. For each and every layer, authors have analysed CDR data and calculated distance and working patterns. By behavioural and lifestyle analysis, authors have analysed their mobility patterns, travelling patterns.	Used spatial-temporal data extracted from CDR to identify and analyse user activity and mobility patterns. Urban analysis can be done based on model. This model useful for investigation of densely populated city.	By using large CDR data from longer period of time can be used to enhance efficiency of this model.

After reviewing many papers, there less work on GUI technique tool in last 5-6 years that is reason to choose this topic for my research work. Most of above paper are based on either classification or regression. In classification, authors checked clustering pattern and mobility patterns and user activity patterns of tourists. In regression, authors predicted next location of tourists to visit either new city or country or hidden visit's locations of users.

The uniqueness of this pilot investigation is that we used two distinct tools. 1. Excel, a statistical tool 2. Maptive – It has a GUI and provides a satellite view. There are many programmes available, both paid and free, to analyse and explore the CDR database. The purpose of this research is to provide background information about digital and cybercriminals. I made a totally fake CDR database, and I used a lot of options to evaluate it.

2.1. TOOL ANALYSIS

1) There are many licenced pieces of software available for CDR data analysis. Excel was used to analyse it.

2) EXCEL

Microsoft Excel has the capacity to manage massive amounts of data, and it has a number of features that are useful for analysis. With comparison to any licenced software, CDR data analysis in Excel takes time. To analyse CDR in this situation, one should use several formulas and tables. Some analysis should be performed manually and then entered into Excel.

3) CDR Analyst

CDR BTS and Forensic software are the providers of the licenced software (ISO-9001:2015 certified). This software does all analysis using the tools that are included with it. Additionally, it offers users video tutorials on how to utilise the software. The year 2022 (CDR BTS and Forensic Software).

4) Purple Radiance CDR Software

Purple Radiance Technology Beyond Solution is the provider of the licenced software in question. The manual analysis of CDR data is a laborious operation. In response, Purple Radiance offered CDR software, which is useful for analysing cell numbers, IMEI, tower dumps, SDR, IPDR, and social networks. Graphical results are also provided by this software. (PurpleRadiance, CDR Analysis Software, 2022)

5) CDR software provided by Avenging Security

It is a licenced product offered by AVENGING SECURITY used by numerous law enforcement agencies to work with and analyse CDR. It produces results in under a minute. This programme has fantastic features for analysing each and every column, including filtering columns, tying relationships between numerous CDRs or columns, identifying shared values between sets, tower dump, etc. Avenging Security, 2022 (LIS CDR Analysis Tool | | Cell Id Finder |)

6) TRAP

3rd Eye Techno Solutions CDR is the provider of the licenced software. The raw database file supplied by telecom operators is used by TRAP to do analysis on the CDR database and Tower Data Record (TDR)(3rd Eye | Home, n.d.).

Due to their size, CDR and TDR provided by telecom companies are particularly challenging to manually analyse. They therefore created the TRAP tool, which can transform enormous amounts of data into relevant information. The software includes many capabilities for producing results, such as filtering out, maximum, minimum, count, and relationships between two or more CDR.

3. MATERIAL AND METHODOLOGY

3.1. MATERIALS

3.1.1. CDR DATA

Call Detail Records (CDRs) are transcripts of phone calls that are created by police officers after making a legally required legal request to the relevant telecoms firm. CDR provides a wealth of personal data about individuals that can be used to solve cases. However, we have constructed a mock/demo CDR database for this study project.

3.1.2. MICROSOFT EXCEL

Microsoft Excel is free software that can be downloaded and used on virtually any computer system. It is used to format, organise, and calculate many types of data in spreadsheets. Tasks like max, min, sum, count, comparing two different values or columns, adding a pivot table, inserting various pie charts, histograms, bars, X-Y scatters, etc., are simple to do.

3.1.3. MAPTIVE

Maptive is a licenced piece of software used to create straightforward maps for any project by entering coordinates (latitude, longitude), which are useful for analysis and visualisation. Maptive offers a number of features, including a filter, multiple variable grouping, markers grouped by colour, a heat map, tools for calculating distance, etc.

3.1.4. JUPYTER NOTEBOOK

The IDE to mix Python with all of the tools we'll be using in our solution is Jupyter Notebook. Although some difficult calculations take a while to finish, it is interactive. Instantaneous plot and image display. It may serve as a one-stop shop for all of our needs, and the majority of libraries, including Dlib, OpenCV, NumPy, Pandas, and Scikit-learn (Sklearn), are simple to integrate.

3.2. METHODOLOGY

3.2.1. DATABASE CREATION

STEP 1: Decided-upon Target Number

STEP 2: studied actual CDRs to become familiar with their data fields and excel sheet layouts.

STEP 3: determined timeframe for which CDR was constructed

STEP 4: All fields were used to generate the database. STEP 5- used telecom company name and request id there. STEP 6- CSV-formatted Excel sheet saved.

3.2.2. EXCEL ANALYSIS

STEP 1: formatting the CDR database and clearing it of superfluous information and empty cells.

STEP 2: Determined MAX, MIN, and TOTAL COUNT by choosing specified columns.

STEP 3: The result was organised in ascending order or displayed as a count when the pivot table was inserted and the columns were selected.

STEP 4: Incorporated graphics for a better understanding of the findings

3.2.3. MAPTIVE ANALYSIS

STEP 1: Log in to the licenced Maptive tool.

STEP 2: Added all addresses with the correct location information. STEP 3- Several map formats were plotted onto maptive.

STEP 4: Once more, after inserting geographical location, a map with the routing direction was displayed in a different colour scheme.

3.2.4. LOGISTIC REGRESSION – PLAN OF WORK

pre-processing of cdr

imported packages and libraries

using Python, read a csv file

using the appropriate columns, build new datasets, and read new files

set test and training data

used logistic regression to analyse test and training data

used regression to forecast new data

4. EXPERIMENTAL WORK

4.1. CREATION OF DATABASE

The Call Detail Record (CDR) database has been practically generated. The database's columns and names all match those found in the actual CDR database exactly. The mobile phone number used to produce the CDR belongs to a real person, and all of the data entered into the records is also authentic. There are 446 target records and a total of 18 columns in the CDR database. The CDR database spans the dates May 1 and May 10, 2022. Always mention the telecom provider that provided the raw CDR; in the case study, the provider is Bharti Airtel Limited, and the request identifier is 0000234123.

Figure 2

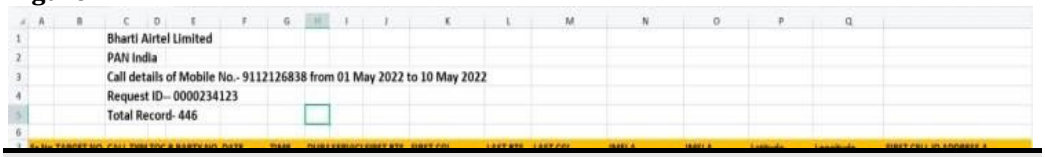


Figure 2 Information Related with CDR

There are several column CDRs included here, as follows:

1) Serial No. Each record in the excel spreadsheet is given a sequential number.
2) Target No. The target number is the mobile phone number for which the CDR database was requested.
3) Call Type—This refers to the call or SMS type, such as CALL-IN, CALL-OUT, SMS-IN, and SMS-OUT.
4) TOC refers to the type of call, such as post-paid or prepaid.
5) B Party Number—This is the number that the Target No. calls or SMSs.
6) Date: This is the day, hour, minute, and year that the target no. was contacted by another party.
7) Time- At what specific time were they contacted by one another.
8) Duration: A call has a duration, however an SMS has no duration.First BTS-BTS is Base Transceiver Station . First BTS gives idea about tower location from where call has started or SMS has been sent.

9) The first CGI is known as global identity. First CGI provides an indication of the tower location from which the call or SMS was initiated. CGI is a network-wide unique identification for BTS. It comprises of a Location Area Code (LAC) of 4 digits, a Cell Identifier (CI) of 4 digits, a Mobile Country Code (MCC) of 3 digits, and a Mobile Network Code (MNC) of 1-2 digits.
10) Last BTS - The last BTS indicates where the call terminated.
11) The last CGI and last BTS provide information about where the conversation ended.
12) The target number is an IMEI (International Mobile Equipment Identity). The mobile device's IMEI provides information about its model number and manufacturing location. International Mobile Subscriber Identity (IMSI A). IMSI provides information about the SIM profile that is kept on the SIM card.
13) Latitude - It provides information about the location from whence the call was initiated.
14) Longitude - It provides information about the place from whence the call was initiated.
15) First Cell ID Address A: This is the address of the location where the call originated.

4.2. EXCEL ANALYSIS

Raw CDR must be transformed into CDR that can be easily read and understood by humans before analysis. After cleaning the CDR database by deleting empty cells, columns, and rows, as well as additional information relating to case specifics or request ID, processed CDR was used for excel analysis.

4.2.1. FILTRATION OF B PARTY COLUMN

Go to the Home tab in MS-Excel. Choose a column from the sheet, apply AutoFilter, then double-click the selected column to return to the AutoFilter menu. Then, the dropdown icon will appear in the first cell of the chosen column. Simply click the icon to bring up a box containing a list of objects, their counts, and total percentages. Items can be arranged in either ascending or descending order, as well as by custom or color sorting. One can order the top 10 products with a maximum or minimum count, or with above- or below-average items, depending on what they need. Following the aforementioned steps, it appears in the filter by value box that the user has called or texted the 940464**** number 98 times in the specified time frame. In 446 records of the CDR database, there are a total of 36 mobile numbers for the B-party.

In Fig. 4.2, a pie-chart representation of the call number filtering according to its count in descending order is displayed together with the overall percentage. A simple and clear representation of the maximum call is provided by the graphical format.

Figure 3

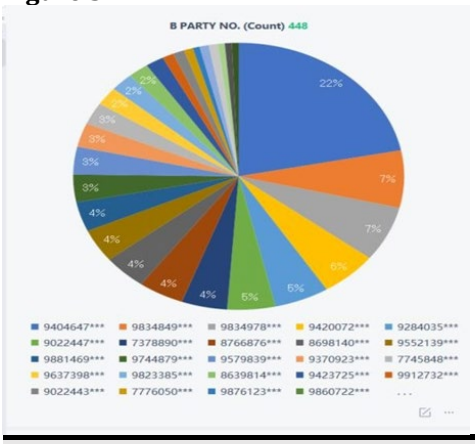


Figure 3 Graphical Representation of Call Filtration with Count

4.2.2. FILTRATION OF DATE

The specified date column is filtered after selecting the Autofilter option on the Home tab. The first year that was displayed was 2022, which is the same for all records. All records also show the same May month. As stated in the database, the CDR's call records cover the period from May 1 to May 10. Therefore, the Autofilter option indicates that on May 8th, 2022, the most calls were made, totaling 81. On May 5th, the minimum call was reached, which is 16. Each date has a total call or message count. The following pie chart provides a visual depiction of the daily call and message volume along with its overall proportion. The graphs below show that 18% of calls and messages took place on May 8, 2022. Only 4% of calls and messages were received on May 5.

Figure 4

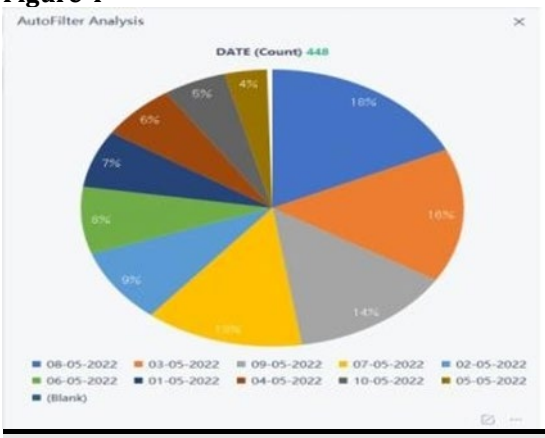


Figure 4 Graphical Representation of No. Of Calls/Message Per Day Filtration of Call Duration

4.2.3. FILTRATION OF CALL DURATION

The highest duration is 0 seconds, which is typically for text messages, according to the Autofilter option of the Home tab. The next duration is 234 and 53 seconds, which is 4 percentage of the total duration.

Figure 5

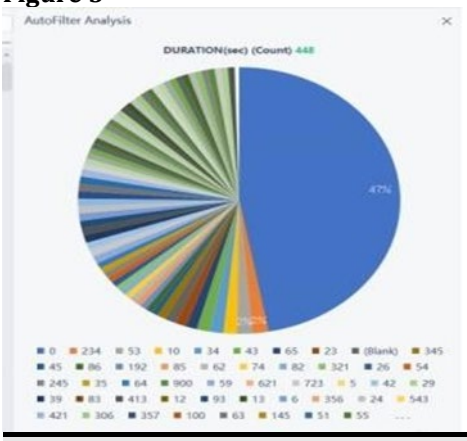


Figure 5 Graphical Representation of Call Duration

Figure 4.5 displays the maximum call duration, which is 1800 seconds, or 5% of the total duration. 966 seconds, or 3% of the total duration, are next.

Figure 6

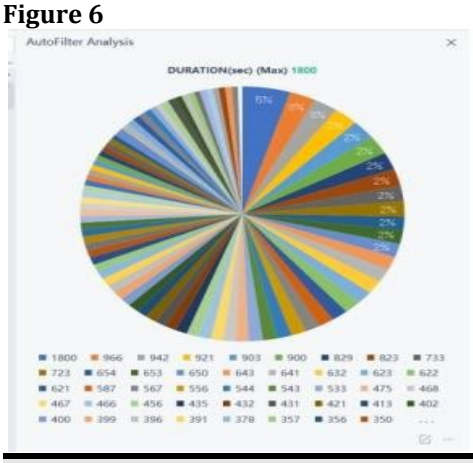


Figure 6 Filtration of Call Duration

4.2.4. FILTRATION OF SERVICE OF CALL

Voice and SMS are filtered using the Autofilter feature of the Home tab. As seen in figure 4.6, 234 of the 446 records had phone calls, whereas 212 have SMS.

Figure 7

Figure 7 Filtration of Services

4.2.5. FILTRATION OF FIRST BTS

It displays the most first BTS location after using the autofilter function. The location that can be reached a maximum of 106 times is 8601-2291. The CDR database of 446 records contains 33 of the first BTS's elements. Figure displays a pie-chart-style illustration of the First BTS. Location 8601-2291 makes up nearly 24 percent of all recordings.

Figure 8

Figure 8 Graphical Representation of First BTS

4.2.6. FILTRATION OF IMEI NUMBER

Filtering of the user's mobile phone's IMEI numbers using the Autofilter option found on the Home tab. There are two IMEI numbers linked to the cellphone number, according to the results of the filtering. However, roughly 289 calls and SMS were sent and received using the IMEI 8637820712*****. Additionally, another IMEI was used 157 times for calls and SMS. Figure 4.8 demonstrates that the IMEI number 8637820712***** utilised 65% of the total IMEI while another was used for 35%.

Figure 9

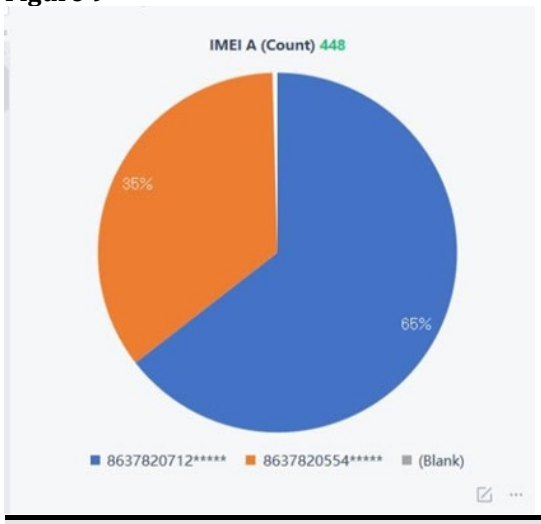


Figure 9 Graphical Representation of Filtered IMEI of user Mobile Phone

4.2.7. FILTRATION OF ADDRESS

The First Cell ID Address will be filtered. Following filtration, filtering by value reveals that 32 different address entries are present in 446 rows. But out of 32, Sambhaji Nagar in Aurangabad received 141 requests, while Nipatniranjan Nagar on Caves Road was the second most requested destination. Sambhaji Nagar Aurangabad appeared in the database for 31% of all addresses as represented graphically in Figure 4.9's first cell ID Address.

Figure 10

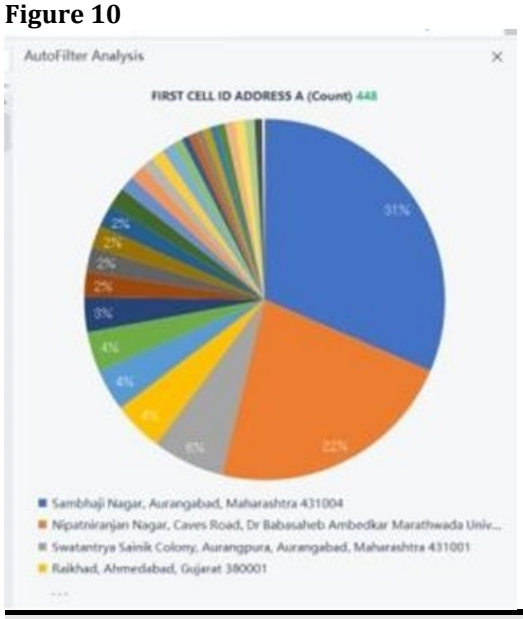


Figure 10 Graphical Representation of Filtered Address

4.2.8. SUM OF CALL DURATION FOR EACH AND EVERY MOBILE NUMBER OF B PARTY IN PIVOT TABLE

First, choose the Excel sheet containing the CDR data. Then, click the Insert tab and then the Insert Pivot Table option. Simply click the "Go into new worksheet" button after selecting your selection. When a new worksheet appears, choose the B party mobile number that is listed in the fields of the pivot table by simply right-clicking on it. This will present you with the following 4 alternatives; choose function add to row labels from the list. Next, simply right- click on the duration field found in the pivot table and choose "append to value label." It will automatically insert the B party mobile number into a specific column from the existing datasheet as Row Labels with a dropdown symbol after completing the aforementioned two steps. The duration column from the existing datasheet is added. After being chosen, the duration column from the old datasheet is added to the new one as the sum of duration. The total time is indicated as 50058 seconds at the end of the records. This pivot table also reveals that during the course of 10 days, the user had over 10463 seconds of conversation with mobile number 9404647***. Figure 4.11 also displays in graphical form the people with whom the user spoke the most and the least.

Figure 11

B PARTY NO.	Sum of DURATION(sec)
7350037***	45
7378890***	1698
7745848***	276
7776050***	1016
8247313***	200
8379052***	260
8446856***	10
8639814***	891
8698140***	2673
8766876***	1652
9022443***	855
9022447***	1823
9067723***	86
9158999***	740
9284035***	5082
9370923***	491
9404647***	10463
9420072***	2567
9423725***	1042
9511771***	65
9545221***	30
9552139***	894
9579839***	2150
9637398***	672
9665350***	900
9744879***	1234
9764420***	900
9823385***	960
9834849***	3319
9834978***	3394
9860722***	40
9876123***	61
9881469***	1889
9912732***	364
9970053***	1316
(blank)	
Grand Total	50058

Figure 11 Pivot Table for Showing Sum of Duration for Every B Party Mobile Number

Figure 12

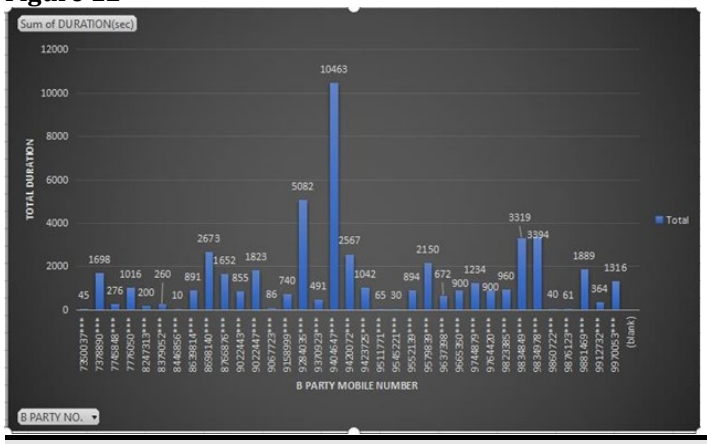


Figure 12 Graphical Representation of Sum of Duration with Respect to Every Mobile Number

- **Maximum duration of call**

Select the Max of Duration option under the values portion of the pivot table fields. Every mobile number's maximum call time will be provided. Figure 4.12 depicts the maximum call time for the mobile number 9284035***.

Figure 13

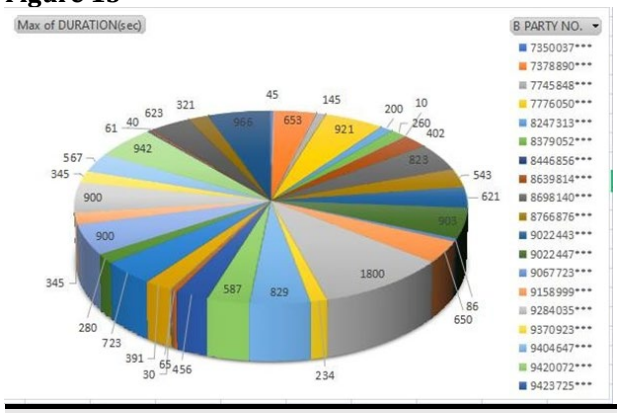


Figure 13 Graphical Representation of Max Duration

4.2.9. TOTAL COUNT OF TYPE OF CALL BY USING PIVOT TABLE

After choosing the Excel sheet, go to the Insert tab, pick the pivot table, and then select the new datasheet. Right-click the B party number in the pivot table fields and then add it to the Rows label. Click the duration option with the right mouse button, add the value label, and call type to the column labels. Choose the duration count option under values. Doing all procedures reveals that there are a total of 114 SMS-IN and 100 SMS-OUT, 124 CALL-IN and 105 CALL-OUT. The graph below clearly demonstrates that users have called the 9404647*** cellphone number the majority of the time using various call types.

Figure 14

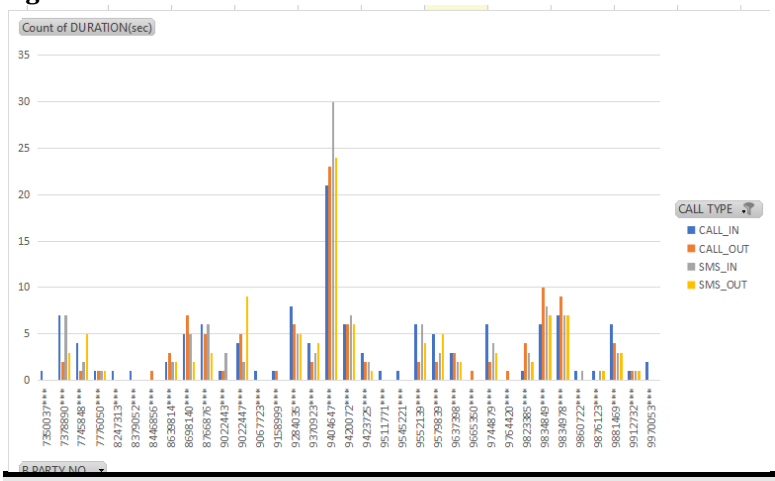


Figure 14 Graphical Representation of Total Count of Type of Call

4.2.10. COUNT OF DURATION OF EVERY MOBILE NUMBER FROM RESPECTIVE ADDRESS

Right-click on the B party number in the pivot table fields to add it to the Rows label, the First Cell ID Address to the Row label, and the Duration to the Values label. A specific number has received calls from the user from several locations. When contacting 9834978, 983489, 9744897, 9404647, 9420072, and 7378990, the user used numerous locations. Right-click on the B party number in the pivot table fields to add it to the Rows label, the First Cell ID Address to the Row label, and the Duration to the Values label. It demonstrates that a user has contacted a specific number from several locations.

- **Count of calls/SMS from respective location**

It displays each location's total calls after inserting the first cell ID address in the rows label and the duration in the values label. The user has been contacted 141 times from Sambhaji Nagar in Aurangabad, according to Fig. 4.14.

Figure 15

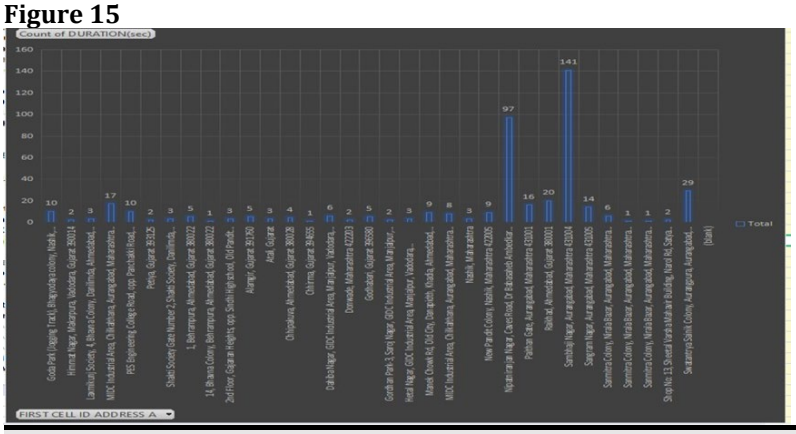
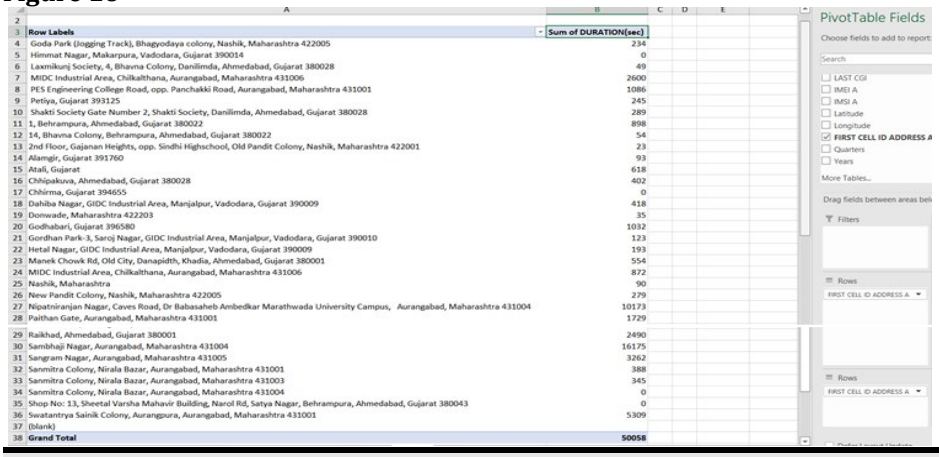


Figure 15 Total Calls/SMS from Respective Location

- **Total duration of contacts from respective locations**

As illustrated in figure 4.15, by placing the address in the rows label and the duration in the values label. It demonstrates that users from Sambhaji Nagar in Aurangabad talked the longest. Next location from which user have talked more is Nipatniranjan Nagar Aurangabad.

Figure 16



Row Labels	Sum of DURATION(sec)
Goda Park (Jogging Track), Bhagyodaya colony, Nashik, Maharashtra 422005	234
Himmat Nagar, Makarpura, Vadodara, Gujarat 390014	0
Laxmikunj Society, 4, Bhavna Colony, Danilimda, Ahmedabad, Gujarat 380028	49
MIDC Industrial Area, Chilkalthana, Aurangabad, Maharashtra 431006	2600
PES Engineering College Road, opp. Panchakki Road, Aurangabad, Maharashtra 431001	1086
Petiya, Gujarat 393125	245
Shakti Society Gate Number 2, Shakti Society, Danilimda, Ahmedabad, Gujarat 380028	289
1, Behrampura, Ahmedabad, Gujarat 380022	898
14, Bhavna Colony, Behrampura, Ahmedabad, Gujarat 380022	54
2nd Floor, Gajanan Heights, opp. Sindhi Highschool, Old Pandit Colony, Nashik, Maharashtra 422001	23
Alamgir, Gujarat 391760	93
Atali, Gujarat	618
Chhipakuva, Ahmedabad, Gujarat 380028	402
Chhirma, Gujarat 394655	0
Dahiba Nagar, GIDC Industrial Area, Manjalpur, Vadodara, Gujarat 390009	418
Donwade, Maharashtra 422203	35
Godhabari, Gujarat 396580	1032
Gordhan Park-3, Saroj Nagar, GIDC Industrial Area, Manjalpur, Vadodara, Gujarat 390010	123
Hetal Nagar, GIDC Industrial Area, Manjalpur, Vadodara, Gujarat 390009	193
Manek Chowk Rd, Old City, Danapidth, Khadia, Ahmedabad, Gujarat 380001	554
MIDC Industrial Area, Chilkalthana, Aurangabad, Maharashtra 431006	872
Nashik, Maharashtra	90
New Pandit Colony, Nashik, Maharashtra 422005	279
Nipatniranjan Nagar, Caves Road, Dr Babasaheb Ambedkar Marathwada University Campus, Aurangabad, Maharashtra 431004	10173
Paithan Gate, Aurangabad, Maharashtra 431001	1729
Raikhad, Ahmedabad, Gujarat 380001	2490
Sambhaji Nagar, Aurangabad, Maharashtra 431004	16175
Sangram Nagar, Aurangabad, Maharashtra 431005	3262
Sanmitra Colony, Nirala Bazar, Aurangabad, Maharashtra 431001	388
Sanmitra Colony, Nirala Bazar, Aurangabad, Maharashtra 431003	345
Sanmitra Colony, Nirala Bazar, Aurangabad, Maharashtra 431004	0
Shop No: 13, Sheetal Varsha Mahavir Building, Narol Rd, Satya Nagar, Behrampura, Ahmedabad, Gujarat 380043	0
Swatantrya Sainik Colony, Aurangpura, Aurangabad, Maharashtra 431001	5309
(blank)	
Grand Total	50058

Figure 16 Total Duration from Every Location

4.3. MAPTIVE TOOL

Maptive is a tool that assists in mapping locations by providing address or coordinate (Latitude, Longitude) information.

Before plotting a map, maptive tool asks for which purpose user is using tool and accoarding to it, user have to give details. In this case study, User inserted Government Institute of Forensic Science Aurangabad and CDR. Next to it, user given all details about address in specific columns like Location name, Address, City, Postal Code, State/province, Country which are as following as shown in figure 4.16.

Figure 17

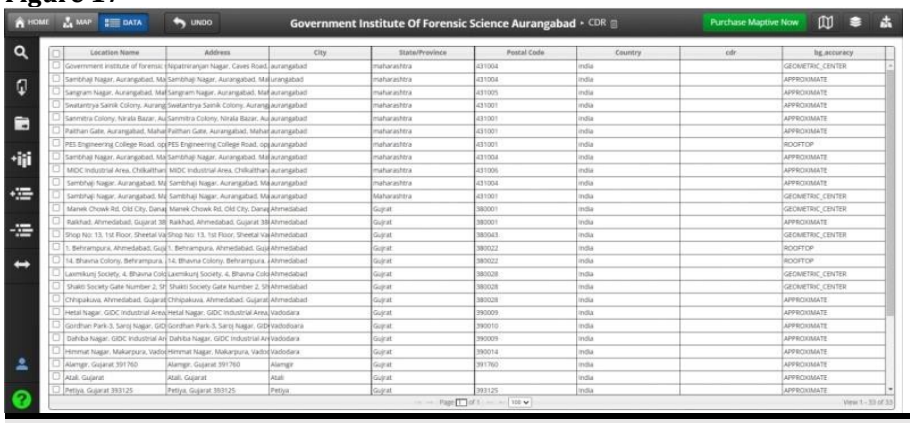


Figure 17 Insertion of All Address Details on Maptive Tool

Following the insertion of all location information, an overall map of India is displayed, which primarily illustrates location concepts. After the first map is zoomed in, a second map is created to show the general location of the cities.

Figure 18

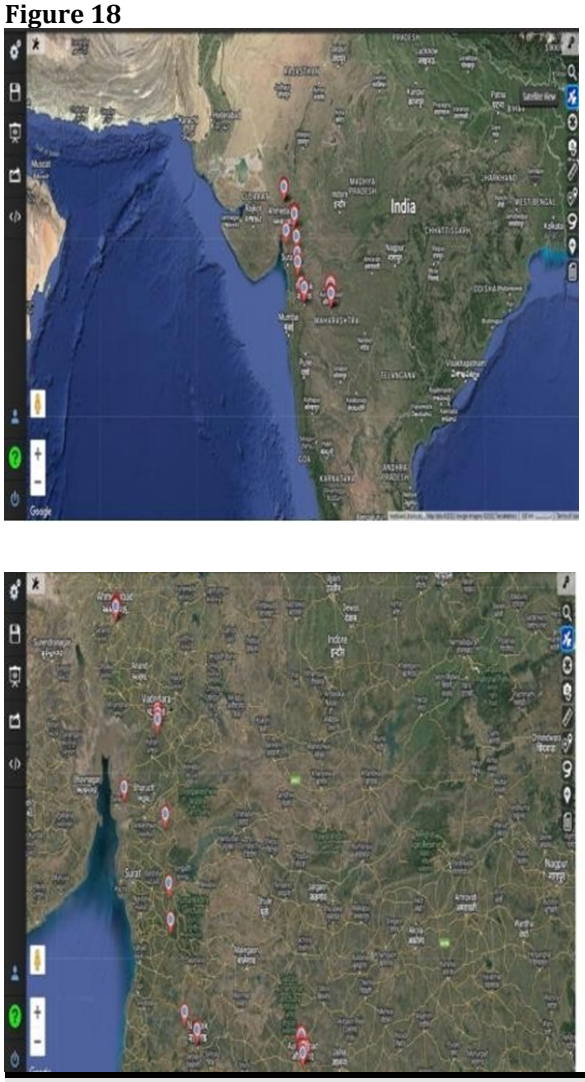


Figure 18 Satellite View of all Location

To pinpoint the precise location, a map of all the cities that the target hasn't been to recently is also shown.

Maptive provides navigation direction with different colours from a highly exact location after receiving location information. The map that follows shows the general area where the user has travelled, and the map that follows shows Aurangabad, where the user has left traces.

Figure 19

Figure 19 Overall Connected Location

4.4. LOGISTIC REGRESSION

When the dependent variable is dichotomous or binary, logistic regression is a statistical method used to create machine learning models. Logistic regression is used to describe data and the connection between a dependent variable and one or more independent variables. The independent variables can be of the nominal, ordinal, or interval types.

Figure 20

```
In [243]: log_reg = LogisticRegression()

In [244]: log_reg.fit(X_train, Y_train)
Out[244]: LogisticRegression()

In [247]: Y_pred = log_reg.predict(X_test)

In [248]: df = pd.DataFrame({'Actual': Y_test, 'Predicted': Y_pred})
          df

In [249]: print("The accuracy of the predicted values is: ", metrics.accuracy_score(Y_test, Y_pred))
          The accuracy of the predicted values is:  0.38392857142857145
```

Figure 20 Logistic Regression

5. OBSERVATION

According to the experimental work, the user had called and talked on the phone more with 9404647*** throughout the course of 10 days, which is the maximum amount of time for a call. The user's longest call lasted almost 1800 seconds with the mobile number 9284035***. The user's longest call was 16175

seconds long from Sambhaji Nagar in Aurangabad, and the next longest call was 10173 seconds long from Nipatniranjan Nagar on Caves Road in Aurangabad. It demonstrates that the user spent more time at those two places. If the CDR reveals two IMEI numbers, the user has changed phones. CALL IN and CALL OUT data show that the user had received more calls than he or she had made. Customer used a pre-paid plan. User travelled to Maharashtra and Gujarat throughout those ten days, primarily stopping in Ahmednagar, Aurangabad, Surat, and Nasik Vadodara. However, user only made one trip to another district as opposed to Aurangabad city because, according to analysis, user was contacted virtually exclusively from Aurangabad city. According to Maptive research, the user visited the city of Aurangabad numerous times over the course of ten days. It is simple to understand the distance between two locations because Maptive displays satellite views of various locations. According to maptive, there are two spots far from the most popular tourist attractions in Aurangabad itself that are in opposing directions from the city.

With its precision, logistic regression forecasts potential locations rather than actual locations. It foresees the potential that the user's location may be Sambhaji Nagar rather than Nipatniranjan Nagar, Shakti Colony, or the actual city of New Pandit Colony.

6. CONCLUSION

It is impossible for the typical person to obtain a user's CDR. To obtain CDR, a copy of the FIR must be enclosed in a suitable channel letter from a judicial body to the telecom company. Even while Excel analysis is free to use, it takes a little bit longer than other CDR analysis programmes. An expert can swiftly analyse CDR using Excel if it is only necessary in a small number of circumstances.

Although Excel is a useful tool for statistical analysis, more complex analyses can be completed more quickly and easily using more expensive tools with third-party licences and other capabilities. When there are fewer CDRs, using Excel to analyse them is easier. Futher The physical position of a site can be determined using satellite pictures thanks to a method called Maptive, which also makes it easier to link to other sites. The use of both statistical and graphical location-based tools helps speed up the investigation process.

DISCLAIMER

Researcher only conducts a pilot study on various available applications. In this research, researcher never encourages / do not promote / do not advertise any specific applications.

CONFLICT OF INTEREST STATEMENT

- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- The authors declare that no financial interests/personal relationships which may be considered as potential competing interests.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

3rd Eye | Home. (n.d.). http://www.3ets.in/cdr-analysis-tool-trap.php

Bianchi, F. M., Rizzi, A., Sadeghian, A., & Moiso, C. (2016). Identifying user habits through data mining on call data records. Engineering Applications of Artificial Intelligence, 54, 49–61.

CDR Analysis Software - PurpleRadiance. (2022).

CDR BTS and Forensic Software. (2022). CDR Training | CDR Software | CDR Analysis | CDR Analysis Software | CDR Analysis Investigation | Mobile CDR Analysis.

Chen, G., Hoteit, S., Viana, A., Fiore, M., & Sarraute, C. (2016). Relevance of Context for the Temporal Completion of Call Detail Record.

Chen, N. C., Xie, W., Welsch, R. E., Larson, K., & Xie, J. (2017). Comprehensive Predictions of Tourists' Next Visit Location Based on Call Detail Records Using Machine Learning and Deep Learning Methods. Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017, 1–6.

Dash, M., Koo, K. K., Gomes, J. B., Krishnaswamy, S. P., Rugeles, D., & Shi-Nash, A. (2015). Next place prediction by understanding mobility patterns. 2015 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2015, 469–474.

Doyle, C., Herga, Z., Dipple, S., Szymanski, B. K., Korniss, G., & Mladenić, D. (2019). Predicting complex user behavior from CDR based social networks. Information Sciences, 500, 217–228.

Khan, F. H., Ali, M. E., & Dev, H. (2015). A hierarchical approach for identifying user activity patterns from mobile phone call detail records. Proceedings of 2015 International Conference on Networking Systems and Security, NSysS 2015, February.

Kujala, R., Aledavood, T., & Saramäki, J. (2016). Estimation and monitoring of city-ity travel times using call detail records. EPJ Data Science, 5(1).

Kumar, M., Hanumanthappa, M., & Kumar, T. V. S. (2017). Crime investigation and criminal network analysis using archive call detail records. 2016 8th International Conference on Advanced Computing, ICoAC 2016, 46–50.

Leng, Y. (2016). Urban computing using call detail records : mobility pattern mining, next-location prediction and location recommendation.

LIS CDR Analysis Tool | | Cell Id Finder | Avenging Security. (2022).

Liu, Q., Wu, S., Wang, L., & Tan, T. (2016). Predicting the next location: A recurrent model with spatial and temporal contexts. 30th AAAI Conference on Artificial Intelligence, AAAI 2016, 194–200.

Sikder, R., Uddin, M. J., & Halder, S. (2017). An efficient approach of identifying tourist by call detail record analysis. IWCI 2016 - 2016 International Workshop on Computational Intelligence, May 2017, 136–141.

Zhao, Zhan, Koutsopoulos, H. N., & Zhao, J. (2021). Identifying Hidden Visits from Sparse Call Detail Record Data. 1–17.

Zhao, Ziliang, Shaw, S. L., Xu, Y., Lu, F., Chen, J., & Yin, L. (2016). Understanding the bias of call detail records in human mobility research. International Journal of Geographical Information Science, 30(9), 1738–

Zhang, Daqiang, Athanasios V. Vasilakos, and Haoyi Xiong. "Predicting location using mobile phone calls." ACM SIGCOMM Computer Communication Review 42, no. 4 (2012): 295-296.

Palchykov V, Mitrovic M, Jo H-H, Saramaki J, Pan RK (2014) Inferring human mobility using communication patterns. Sci Rep 4:6174