CHAPTER - 1

INTRODUCTION

1.1 Introduction

Increased crime rates play a decisive role in hindering the growth of a nation as a unit, especially in the case of developing countries like India. There is a strong need to monitor and analyse the crimes occurring across the country in addition to regulating them. Though the task of crime regulation is solely under the hands of the government, analysis of the crimes can be carried out in order to derive conclusions and suggest necessary actions and measures extracted from analytical analysis of the data sets available. This method can play a significant role in helping the government in taking better actions for maintaining law and order in the nation.

Results in the form of visualizations and story are boards easily understandable to our not so literate politicians who handle the government affairs. Studies like this can play a major role in helping the law and order department to have a clear analysis of crimes and hence take reasoning based actions and measures. Here I provide some of the analysis results of the crime data for Indian sub-continent, a developing nation where crime regulation plays an important role in developing trust among its citizens. A special focus is laid in the study towards CRIMES AGAINST WOMEN to enhance need for better security and dignity of the women and girls in the country.

1.2 Rationale

The rationale of the study includes three important aspects such as the relevance of the topic, interest of the researcher and the researchability of the topic

1.3 Objectives

1. To study the nature of domestic violence on respondents.
2. To study the causative factors of domestic violence.
3. To study the effects of domestic violence on the respondents.
4. To measure the level of stress among the respondents.
5. To understand the coping methods used by the respondents.
6. To understand the help seeking behaviour of the respondents.
7. To understand the awareness of victims regarding the services existing for protection of women from domestic violence

1.4 Existing System

This system named "Case Study on 'Crime on Women in India' " is a data analytical application that aims to predict crime rate. This project is intended to solve the crime rate.

CHAPTER – 2
REQUIREMENT ANALYSIS AND SYSTEM SPECIFICATION

2.1 Feasibility Study

Feasibility is the determination of whether or not a project is worth doing. The process followed in making this determination is called feasibility study. Feasibility Study concentrates on the following areas:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

1.Technical Feasibility

Simply put, Crime rates cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The crime rate is a place where criminal and civilians converge. When there are more criminals than civilians, the crime rate increases. When there are more civilians than criminals, the crime rate decreases. It has more to do with emotion than logic. Because emotion is unpredictable, crime ratesmovements will be unpredictable. It's futile to try to predict where markets are going. They are designed to be unpredictable.

2. Operational Feasibility

The proposed system will not always produce accurate results since it does not account for the human behaviours. Factors like change in unemployement, internal matters, strikes, protests can be taken into account for relating it to the change in Crime rates by the machine.The objective of the system is to give a approximate idea of where the crime rates might be headed. It does not give a long term forecasting of a crime rate value. There are way too many reasons to acknowledge for the long term output of a current crime ratte. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

2.2 Software Requirement Specification Document

Software Requirement Specification (SRS) is a detailed description of a software system to be developed with its functional and non-functional requirements. SRS on QA lead, managers creates test plan and follows the IEEE/ANSI-830-1998 standard. It is based upon the agreement between customer and contractor. It may include some cases when user interact with system software. It is highly recommended to test all cases while making planning.

2.2.1 Purpose

The proposed system is Crime rates prediction and analysis is the act of trying to determine the future value of a state crime. Crimes against women and girl children are on the rise and the place, which we think is safer, is no longer secure. The changing time in todays world have encouraged many women to seek their own identity, the value and the worth of their lives, through the contribution they make to the family and society at large. She now wants to be regarded as an individual with feelings, aspirations, desires, as a 'human' person. The women are demanding the male dominated society, their due respect, dignity and status especially at the work place, where some of them face sexual harassment and gender discrimination. Unfortunately, many women today fall victims in the hands of those men who fail to respect and recognize a woman in her own identity for a variety of reasons. The outcome is domestic violence, battering, torturing both mental and physical.

2.2.2 Functional Requirements

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

• The system should be able to generate an approximate share price.
• The system should collect accurate data in consistent manner.

It requires following tools for implementation:-

R Language (**R** is a programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing).

*R Studio (***RStudio** is a free and open-source integrated development environment (IDE) for R, a programming language for statistical computing and graphics. R Studio was founded by JJ Allaire, creator of the programming language ColdFusion. Hadley Wickham is the Chief Scientist at RStudio)

2.2.4 Non Functional Requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non- functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

- The system should provide better accuracy.
- The system should have simple interface for users to use.
- To perform efficiently in short amount of time.

User requires an internet connection for accessing this application. Knowledge to basic internet usage is required by the user.

Software Requirements:

- Windows 7 or higher
- R Studio
- R Language

Operating System	Linux
Software	Hadoop, Hive, Tableau Java
Database	SQL, CSV
Languages	Bigdata SQL
Linux Version	16.04 and above

Table 1: Software Requirements

2.2 HARDWARE REQUIREMENTS

Main Memory	2GB RAM minimum 8GB Recommended
Hard Disk Space	2GB RAM minimum 4GB Recommended
Processor	1.6GHz

Table 2: Hardware Requirements

2.3 Analysis Model

The model that is basically being followed is the WATER FALL MODEL, which states that the phases are organized in a linear order. First of all, the feasibility study is done. Once that part is over the requirement analysis and project planning begins. If system exists one and modification and addition of new module is needed, analysis of present system can be used as basic model.

The design starts after the requirement analysis is complete and the coding begins after the design is complete. Once the programming is completed, the testing is done. In this model the sequence of activities performed in a software development project are: -

- Requirement Analysis
- Project Planning
- System design
- Detail design
- Coding
- Unit testing
- System integration & testing

Here the linear ordering of these activities is critical. End of the phase and the output of one phase is the input of other phase. The output of each phase is to be consistent with the overall requirement of the system. Some of the qualities of spiral model are also incorporated like after the people concerned with the project review completion of each of the phase the work done.

WATER FALL MODEL was being chosen because all requirements were known beforehand and the objective of our software development is the computerization/automation of an already existing manual working system.

2.3.1 Existing System:

This system named "Crime ratesAnalysis and Prediction using Artificial Neural Networks" is a web application that aims to predict crime ratesvalue using Artificial Neural Network. This project is intended to solve the economic dilemma created in individuals that wants to invest in Stock Market.

2.3.2 Problem Statement:

Crime rate is very vast and difficult to understand. It is considered too uncertain to be predictable due to huge fluctuation of the state. Crime rates prediction task is interesting as well as divides researchers and academics into two groups, those who believe that we can devise

mechanisms to predict the society and those who believe that the society is efficient and whenever new information comes up the society absorbs it by correcting itself, thus there is no space for prediction.

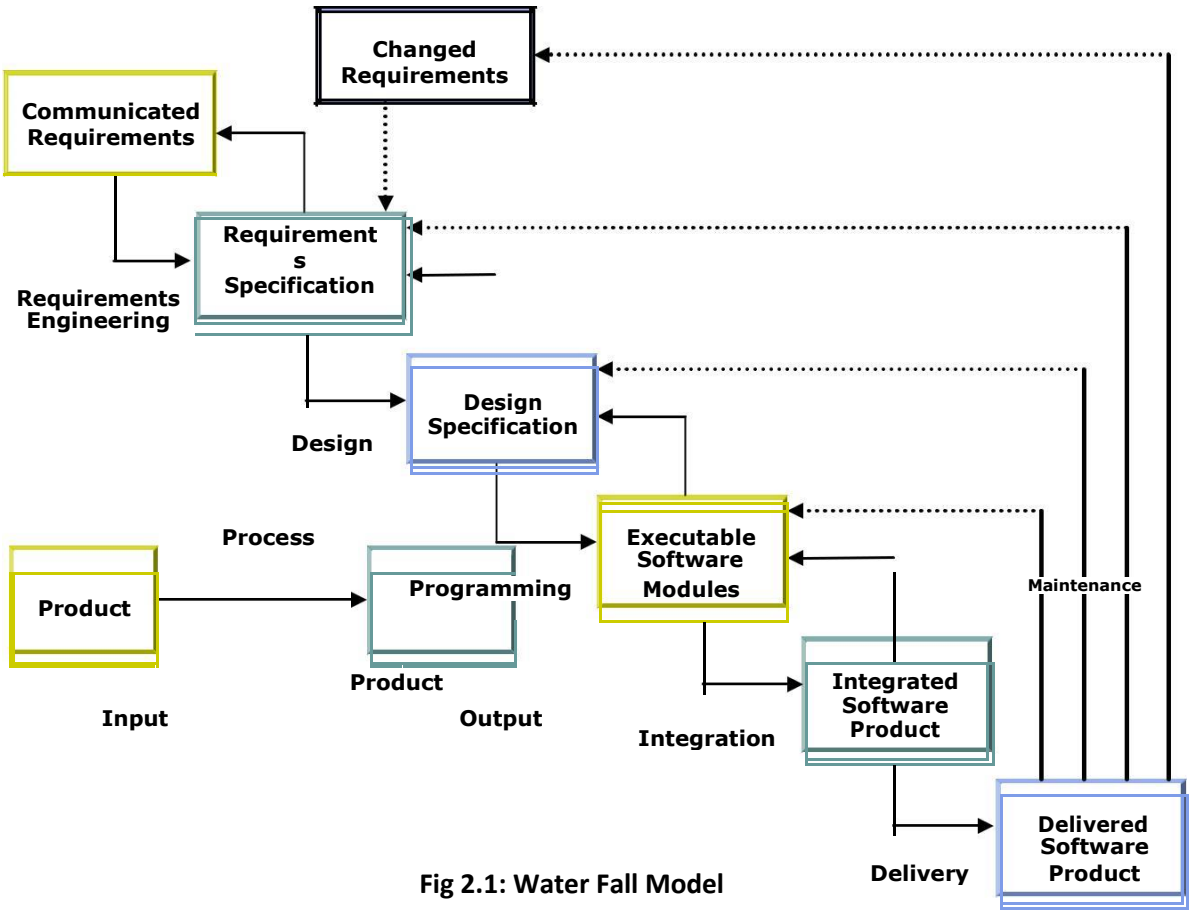


Fig 2.1: Water Fall Model

CHAPTER – 3

SYSTEM DESIGN

3.1 Data Flow Diagram

3.1.1 DFD Model

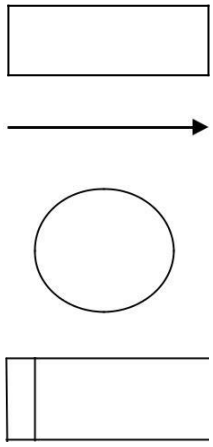
DFDs models the system by depicting

1) External entities from which the data flows and where results terminate.
2) Processes which transform data flows.
3) Data stores from which the data are read or into which data are written by the processes.

Symbols used in DFD

1) A circle represents a process.
2) Straight lines within coming arrows are input data flows.
3) Straight lines without going arrows are output data flows.
4) Processes are given serial numbers for easy reference.

Labels are assigned to Data flow. These aid documentations.



Source or Destination of data

Data Flow

Process

Storage

Fig 3.1: Symbols of the DFD

3.1.2 Steps to Construct Data Flow Diagrams

Four steps are commonly used to construct a DFD:

1) Process should be named and numbered for easy reference. Each name should be representative of the process.
2) The direction of flow is from top to bottom and from left to right.
3) When a process is exploded into lower level details they are numbered.
4) The names of data stores, sources and destinations are written in capital letters.

3.1.3 Rules for constructing a Data Flow Diagram

1) Arrows should not cross each other.
2) Squares, Circles and files must bear names.

3) Decomposed data flow squares and circles can have same names.
4) Choose meaningful names for dataflow.
5) Draw all data flows around the outside of the diagram.
6) An entire system is represented by one DFD which gives the system overview.
7) It is called a context diagram. It gives little detail &is also known as the top level DFD.
8) Context diagram is shown in the next diagram.

3.1.4 DFD Diagrams:

Context level Diagram:

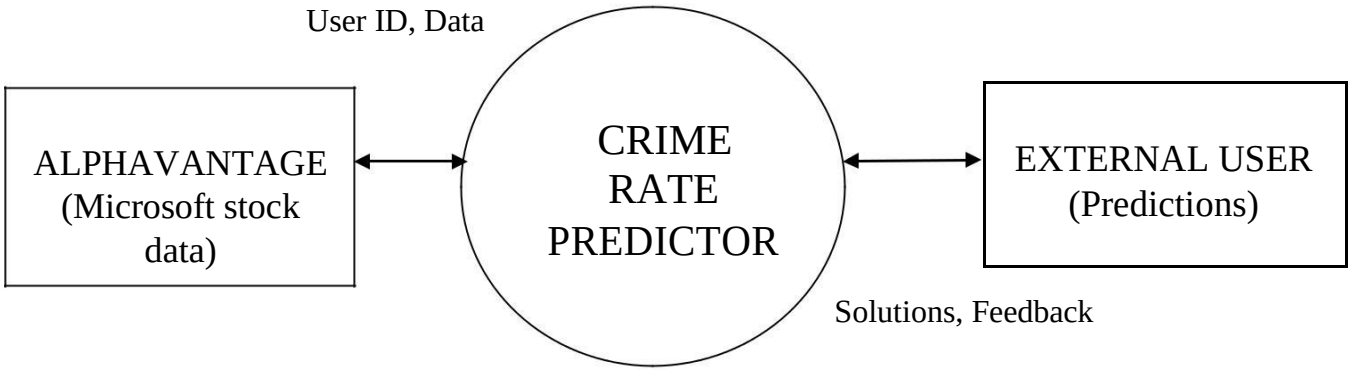


Fig 3.2: Context Level DFD (Overall System)

3.2 ER Diagrams

An ER diagram is a diagram that helps to design databases in an efficient way. Attributes in ER diagrams are usually modeled as an oval with the name of the attribute, linked to the entity or relationship that contains the attribute. Within the relational model the final step can generally be broken down into two further steps, that of determining the grouping of information within the system, generally determining what are the basic objects about which information is being stored, and then determining the relationships between these groups of information, or objects. An Entity Relationship Diagram (ERD) is a visual representation of different data using conventions that describe how these data are related to each other. ER diagrams are most often associated with complex databases that are used in software engineering and IT networks. In particular, ER diagrams are frequently used during the design stage of a development process in order to identify different system elements and their relationships with each other. For example, an inventory software used in a retail shop will have a database that monitors elements such as purchases, item, item type, item source and item price.

Three main components of an ERD are the entities, which are objects or concepts that can have data stored about them, the relationship between those entities, and the cardinality, which defines that relationship in terms of numbers. The three main cardinal relationships are: One-to-one (1:1), One-to-many (1:M), Many-to-many (M: N).

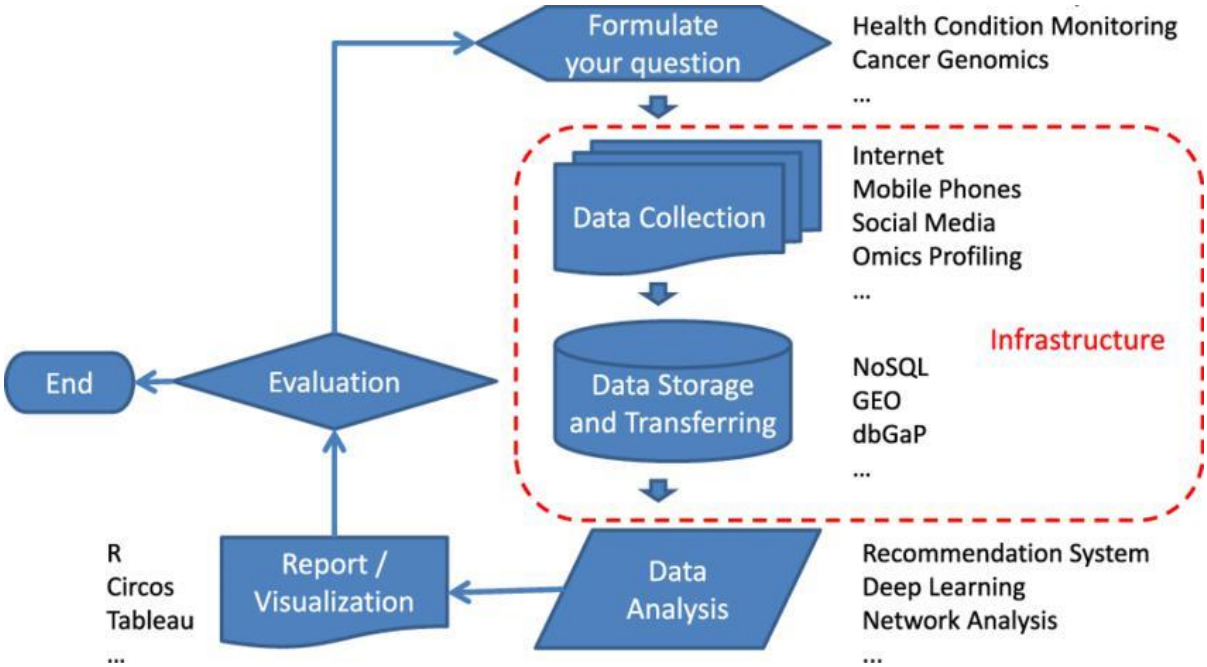


Fig 3.4: E-R diagram

CHAPTER 4
IMPLEMENTATION, TESTING, AND MAINTENANCE

4.1 Introduction to Languages, IDE's.

I. INTRODUCTION

In digital world, data are generated from various sources and the fast transition from digital technologies has led to growth of big data. It provides evolutionary breakthroughs in many fields with collection of large datasets. In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management tools or data processing applications. These are available in structured, semi-structured, and unstructured format in petabytes and beyond. Formally, it is defined from 3Vs to 4Vs. 3Vs refers to volume, velocity, and variety. Volume refers to the huge amount of data that are being generated everyday whereas velocity is the rate of growth and how fast the data are gathered for being analysis. Variety provides information about the types of data such as structured, unstructured, semistructured etc. The fourth V refers to veracity that includes availability and accountability. The prime objective of big data analysis is to process data of high volume, velocity, variety, and veracity using various traditional and computational intelligent techniques . Some of these extraction methods for obtaining helpful information was discussed by Gandomi and Haider .

The following Figure 1 refers to the definition of big data. However exact definition for big data is not defined and there is a believe that it is problem specific. This will help us in obtaining enhanced decision making, insight discovery and optimization while being innovative and cost-effective.

2.Big Data Analytics

The term "Big Data" has recently been applied to datasets that grow so large that they become awkward to work with using traditional database management systems. They are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process the data within a tolerable elapsed time. Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before. Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics based on large data samples reveals and leverages business change. However, the larger the set of data, the more difficult it becomes to manage . In this section, we will start by discussing the characteristics of big data, as well as its importance. Naturally, business benefit can commonly be derived from analyzing larger and more complex data sets that require real time or near-real time capabilities; however, this leads to a need for new data architectures, analytical methods, and tools. Therefore the successive section will elaborate the big data analytics tools and methods, in particular, starting with the big data storage and management, then moving on to the big data analytic processing. It then concludes with some of the various big data analyses which have grown in usage with big data.

3. Characteristics of Big Data

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to enable insights that unlock new sources of business value. Three main features characterize big data: volume, variety, and velocity, or the three V's. The volume of the data is its size, and how enormous it is. Velocity refers to the rate with which data is changing, or how often it is created. Finally, variety includes the different formats and types of data, as well as the different kinds of uses and ways of analyzing the data .

Data volume is the primary attribute of big data. Big data can be quantified by size in TBs or PBs, as well as even the number of records, transactions, tables, or files. Additionally, one of the things that make big data really big is that it's coming from a greater variety of sources than ever before, including logs, clickstreams, and social media. Using these sources for analytics means that common structured data is now joined by unstructured data, such as text and human language, and semi-structured data, such as eXtensible Markup Language (XML) or Rich Site Summary (RSS) feeds. There's also data, which is hard to categorize since it comes from audio, video, and other devices. Furthermore, multi-dimensional data can be drawn from a data warehouse to add historic context to big data. Thus, with big data, variety is just as big as volume. Moreover, big data can be described by its velocity or speed. This is basically the frequency of data generation or the frequency of data delivery. The leading edge of big data is streaming data, which is collected in real-time from the websites . Some researchers and organizations have discussed the addition of a fourth V, or veracity. Veracity focuses on the quality of the data.

4. CHALLENGES IN BIG DATA ANALYTICS

Recent years big data has been accumulated in several domains like health care, public administration, retail, biochemistry, and other interdisciplinary scientific researches. Web-based applications encounter big data frequently, such as social computing, internet text and documents, and internet search indexing. Social computing includes social network analysis, online communities, recommender systems, reputation systems, and prediction markets where as internet search indexing includes ISI, IEEE Xplorer, Scopus, Thomson To handle the challenges we need to know various computational complexities, information security, and computational method, to analyze big data. For example, many statistical methods that perform well for small data size do not scale to voluminous data. Similarly, many computational techniques that perform well for small data face significant challenges in analyzing big data. Various challenges that the health sector face was being researched by much researchers. Here the challenges of big data analytics are classified into four broad categories namely data storage and analysis; knowledge discovery and computational complexities; scalability and visualization of data; and information security. We discuss these issues briefly in the following subsections.

About the Tableau

Tableau is a Business Intelligence tool for visually analyzing the data. Users can create and distribute an interactive and shareable dashboard, which depict the trends, variations, and density of the data in the form of graphs and charts. Tableau can connect to files, relational and Big Data sources to acquire and process data. The software allows data blending and realtime collaboration, which makes it very unique. It is used by businesses, academic researchers, and many government organizations for visual data analysis. It is also positioned as a leader Business Intelligence and Analytics Platform in Gartner Magic Quadrant.

Tableau Features-Tableau provides solutions for all kinds of industries, departments, and data environments. Following are some unique features which enable Tableau to handle diverse scenarios.

Self-Reliant − Tableau does not need a complex software setup. The desktop version which is used by most users is easily installed and contains all the features needed to start and complete data analysis.

Visual Discovery − The user explores and analyzes the data by using visual tools like colors, trend lines, charts, and graphs. There is very little script to be written as nearly everything is done by drag and drop.

Blend Diverse Data Sets − Tableau allows you to blend different relational, semistructured and raw data sources in real time, without expensive up-front integration costs. The users don't need to know the details of how data is stored.

Architecture Agnostic − Tableau works in all kinds of devices where data flows. Hence, the user need not worry about specific hardware or software requirements to use Tablea

Real-Time Collaboration − Tableau can filter, sort, and discuss data on the fly and embed a live dashboard in portals like SharePoint site or Salesforce. You can save your view of data and allow colleagues to subscribe to your interactive dashboards so they see the very latest data just by refreshing their web browser.

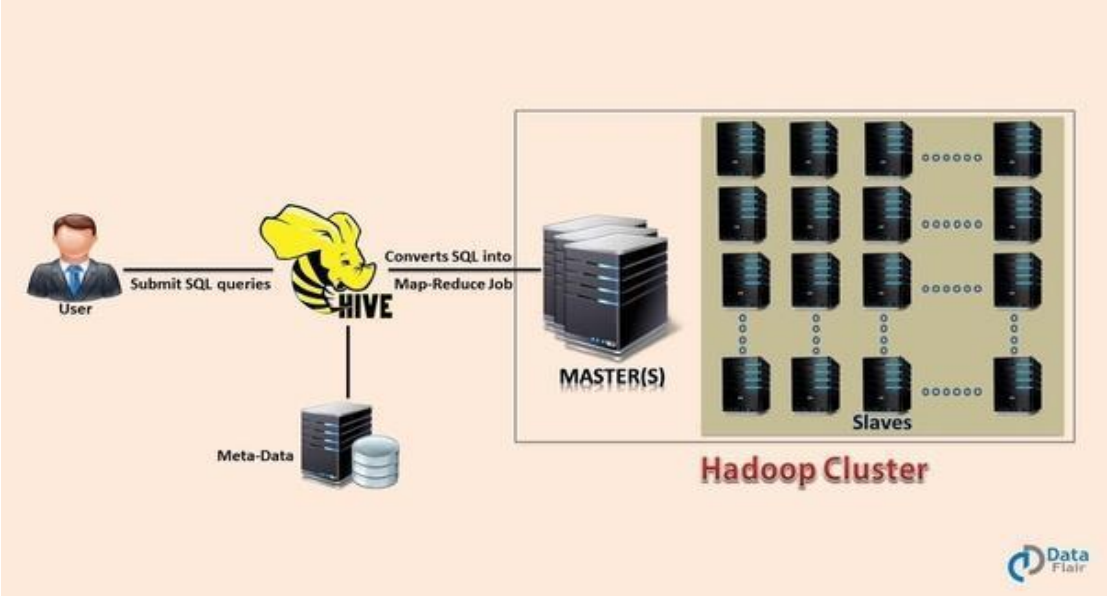
Centralized Data − Tableau server provides a centralized location to manage all of the organization's published data sources. You can delete, change permissions, add tags, and manage schedules in one convenient location.

About Hive

Apache Hive is a data warehousing solution for Hadoop which provides data summarization, query, and ad-hoc analysis. It is used to process structured and semi-structured data in Hadoop. Analysis of large datasets stored in Hadoop's HDFS and also in Amazon S3 filesystem is

supported by Hive. Like SQL hive also provides query language named HiveQL. Ad-hoc queries can be run using Hive for the data analystics.

Hive Aechitecture:The diagram below shows basic architecture of Hive. We have Hadoop cluster, masters, and slaves where HDFS and MapReduce runs. Hive is a high-level component and sits on the top of the Hadoop. Hive is usually installed on the Master (or edge node in case of a big cluster). The hive needs a Metastore which is an RDBMS (it can be any RDBMS like MySQL or Oracle). Hive stores its Metadata inside the RDBMS, which can be deployed on master or on the third machine. User submits a SQL query in the Hive like creating a table then its metadata like a number of rows, columns, etc. gets stored inside the RDBMS. Hive converts this SQL queries into MapReduce jobs and submits it to the cluster. And hence data is processed on the Slaves. Every time user submits DDL SQL queries Hive updates its Metastore.



R Language

R is a language and environment for statistical computing and graphics. It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R can be considered as a different implementation of S. There are some important differences, but much code written for S runs unaltered under R.

R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.

R Studio

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, RedHat/CentOS, and SUSE Linux).

4.3 Project Scheduling

4.3.1 Pert chart

A PERT chart is a project management tool that provides a graphical representation of a project's timeline. The Program Evaluation Review Technique (PERT) breaks down the individual tasks of a project for analysis. PERT charts are considered preferable to Gantt charts because they identify task dependencies, but they're often more difficult to interpret These nodes are linked by vectors or lines that represent various tasks. Dependent tasks are items that must be performed in a specific manner. For example, if an arrow is drawn from Task No. 1 to Task No. 2 on a PERT chart, Task No. 1 must be completed before work on Task No. 2 begins.

	Name	Duration	Start	Finish
1	**Stock Market Price Prediction**	58 days	**6/2/19 8:00 AM**	**26/4/19 5:00 PM**
2	**Problem Analysis**	5 days	**6/2/19 8:00 AM**	**12/2/19 5:00 PM**
3	Feasibility Study	3 days	6/2/19 8:00 AM	8/2/19 5:00 PM
4	Prepapration of SRS	2 days	11/2/19 8:00 AM	12/2/19 5:00 PM
5	Literature Review	8 days	13/2/19 8:00 AM	22/2/19 5:00 PM
6	Dataset Collection	7 days	25/2/19 8:00 AM	5/3/19 5:00 PM
7	Feature Extraction	5 days	6/3/19 8:00 AM	12/3/19 5:00 PM
8	**Architecture Design and Implementation**	15 days	**13/3/19 8:00 AM**	**2/4/19 5:00 PM**
9	Designing the Stock Price Prediction Architecture	7 days	13/3/19 8:00 AM	21/3/19 5:00 PM
10	Coding the price prediction model	5 days	22/3/19 8:00 AM	28/3/19 5:00 PM
11	Training the price prediction model	3 days	29/3/19 8:00 AM	2/4/19 5:00 PM
12	Testing the price prediction network	7 days	3/4/19 8:00 AM	11/4/19 5:00 PM
13	Analysis of the Project	7 days	12/4/19 8:00 AM	22/4/19 5:00 PM
14	Documentation	4 days	23/4/19 8:00 AM	26/4/19 5:00 PM

Fig 4.2 Pert chart

4.3.2 Gantt chart

A Gantt chart is a type of bar chart that illustrates a project schedule, named after its inventor, Henry Gantt (1861{1919), who designed such a chart around the years 1910{1915. Modern Gantt charts also show the dependency relationships between activities and current schedule status. A Gantt chart is a type of bar chart that illustrates a project schedule. This chart lists the tasks to be performed on the vertical axis, and time intervals on the horizontal axis. The width of the horizontal bars in the graph shows the duration of each activity. Gantt charts illustrate the start and dates of the terminal elements and summary elements of a project. Terminal elements and summary elements constitute the structure.

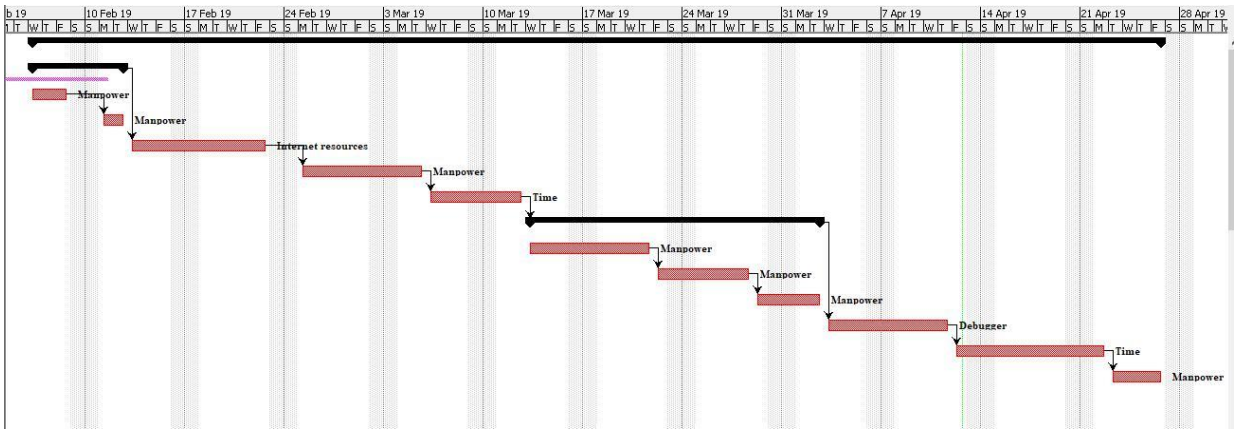


Fig 4.3: Gantt chart

4.3 Testing Techniques and Test Plans

Unit Test: Each independent piece of code works correctly. The primary goal of unit testing is to take the smallest piece of testable software in the application, isolate it from the remainder of the code, and determine whether it behaves exactly as you expect. Each unit is tested separately before integrating them into modules to test the interfaces between modules.

Unit testing has proven its value in that a large percentage of defects are identified during its use. Unit tests are typically written and run by software developers to ensure that code meets its design and behaves as intended. Its implementation can vary from being very manual (pencil and paper) to being formalized as part of build automation.

Integration Test: All units work together without errors.

Interface Test: Usually done at integration stage when modules or sub-systems are combined. Objective is to detect errors or invalid assumptions about interfaces between modules. Reason these are not shown up in unit testing is that test case may perpetuate same incorrect assumption made by module designer. Particularly important when Object Oriented development has been used. Global variable). One places data there and the other retrieves it.

Regression Test: Newly Added features do not introduce errors to other features that are already working.

Load Test (also called Stress Test): The product continues to work under extreme usage. Test system's ability to cope with a specified load (e.g. transactions per second). Plan tests to increase load incrementally. Go beyond design limit until system fails (this test particularly important for distributed systems).

Platform Test: The product works on all the target hardware and software platforms.

Top Down Test: This approach tests high levels of system before detailed components. This is appropriate when developing the system top-down as it is likely to show up structural design errors early. Validation (as distinct from verification) can begin early. Its disadvantage is that stubs needs to be generated (extra effort) and might be impractical if component is complex (e.g. converting an array into a linked list; unrealistic to generate random list. Therefore, end up implementing unit anyway). Test output may be difficult to observe (needs creation of artificial environment). This is not appropriate for OO systems (except within a class).

Bottom Up Test: This is opposite of top-down testing. This testing test low-level unit then works up hierarchy. Its advantages and disadvantages of bottom-up mirror those of top-down. In this testing there is need to write test drivers for each unit. These are as reusable as the unit itself. Combining top-down development with bottom-up testing means that all parts of system

must be implemented before testing can begin, therefore does not accord with incremental approach discussed above.

Back to Back Test: Comparison of test results from different versions of the system (e.g. comparing the prototype with previous version or different configuration). The process involves running the first system, saving test case results. Then running the second system, also saving its results. Finally comparing the results files. It is important to note that no difference doesn't imply no bugs. Both systems may have made the same mistake.

Testing Done in my System: The best testing is to test each subsystem separately as we have done in our project. It is best to test a system during the implementation stage in form of small sub steps rather than large chunks. We have tested each module separately i.e. have completed unit testing first and system testing was done after combining /linking all different Modules with different menus and thorough testing was done. Once each lowest level unit has been tested, units are combined with related units and retested in combination. This proceeds hierarchically bottom-up until the entire system is tested as a whole. Hence we have used the Top Up approach for testing our system.

CHAPTER – 5

OUTPUT SCREENS

The following graph represent the different crimes against women like Rape, Kidnapping, Insult to modesty, Dowry Deaths, Cruelty by Husband.

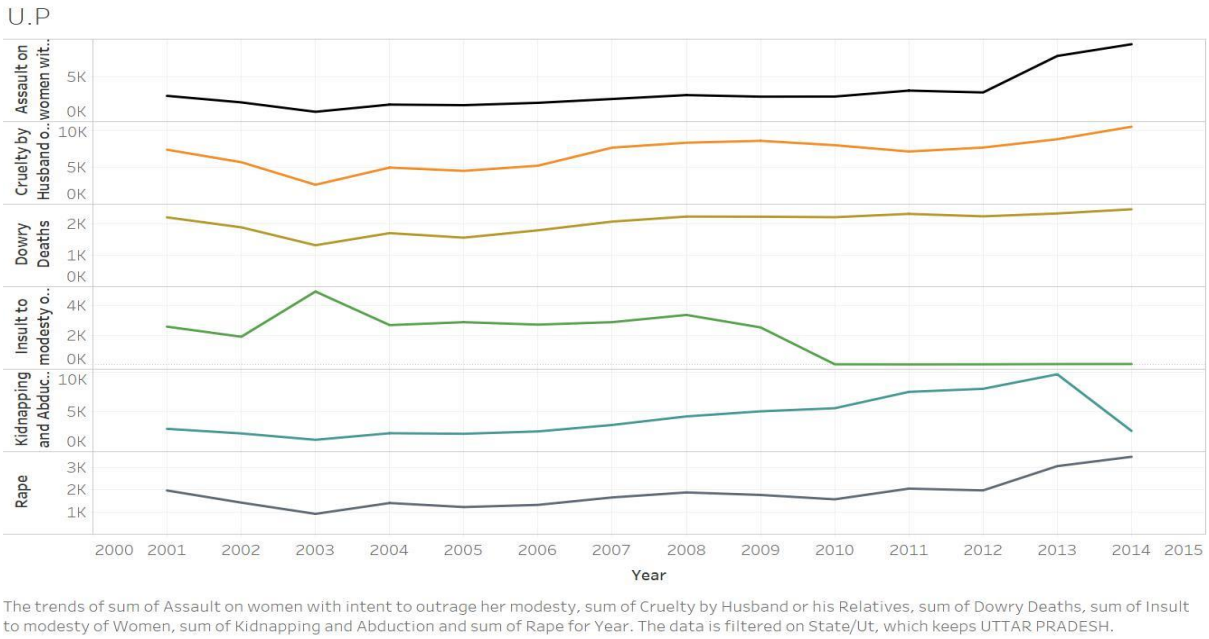
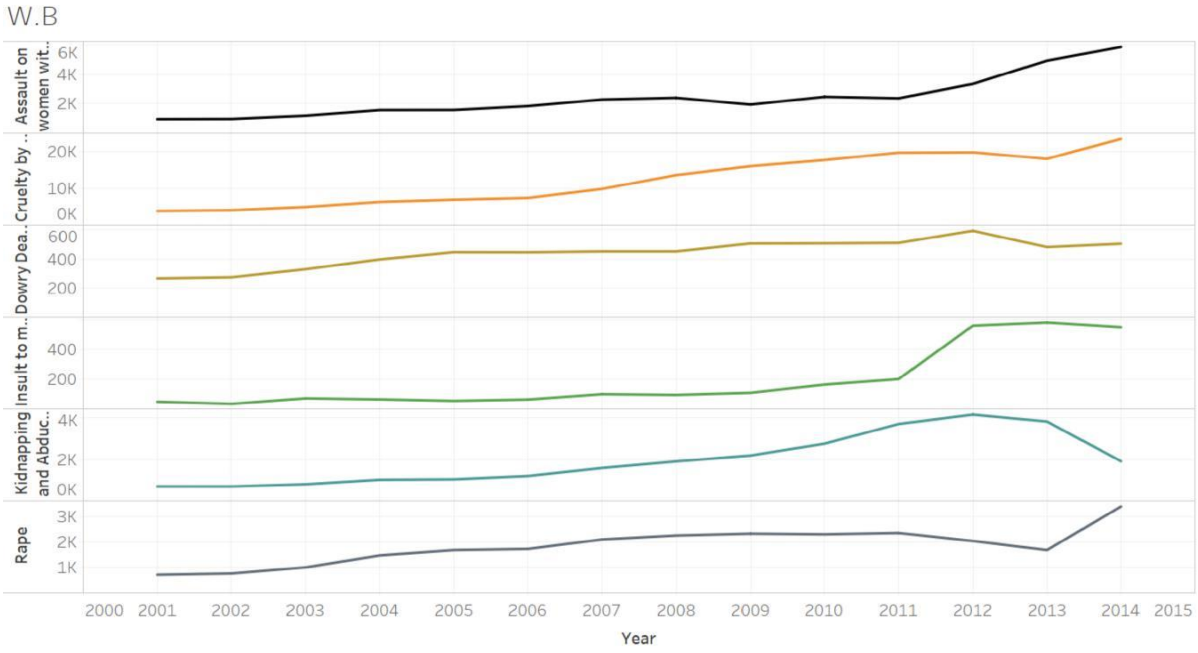


Fig 5.1 Crime on women in U.P

The above image shows the different graphs of crime against women in the state Uttar Pradesh in the time spam of 15 years. Starting from the top the graph is of assault on women vesus years. Second is cruelty by husband on women versus year. Third chart is cases of dowry deaths versus years. Such as the last graph is of rape cases versus the year.



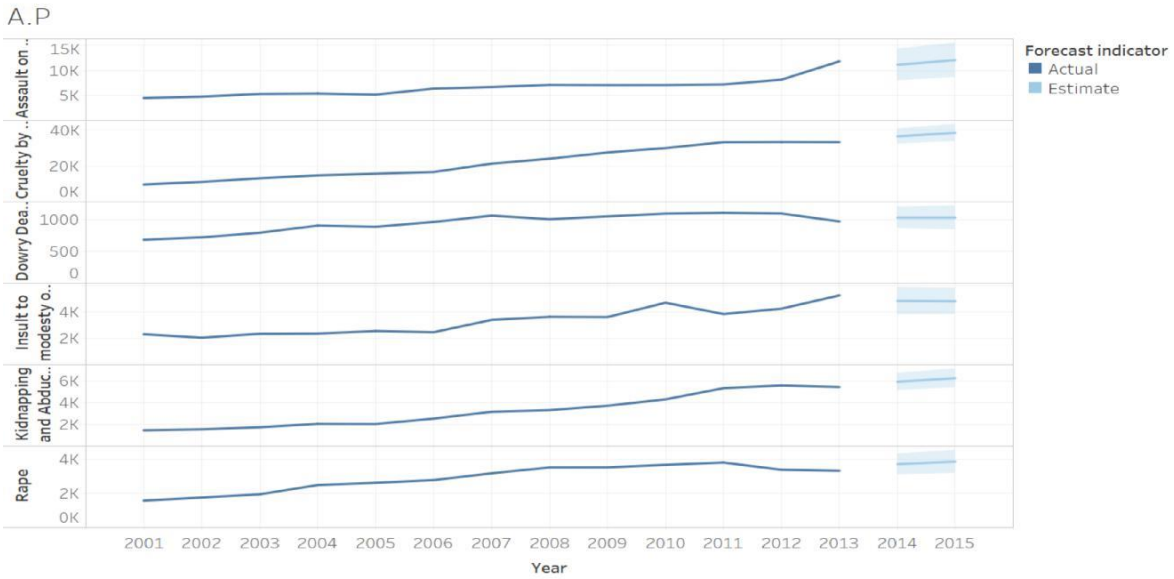
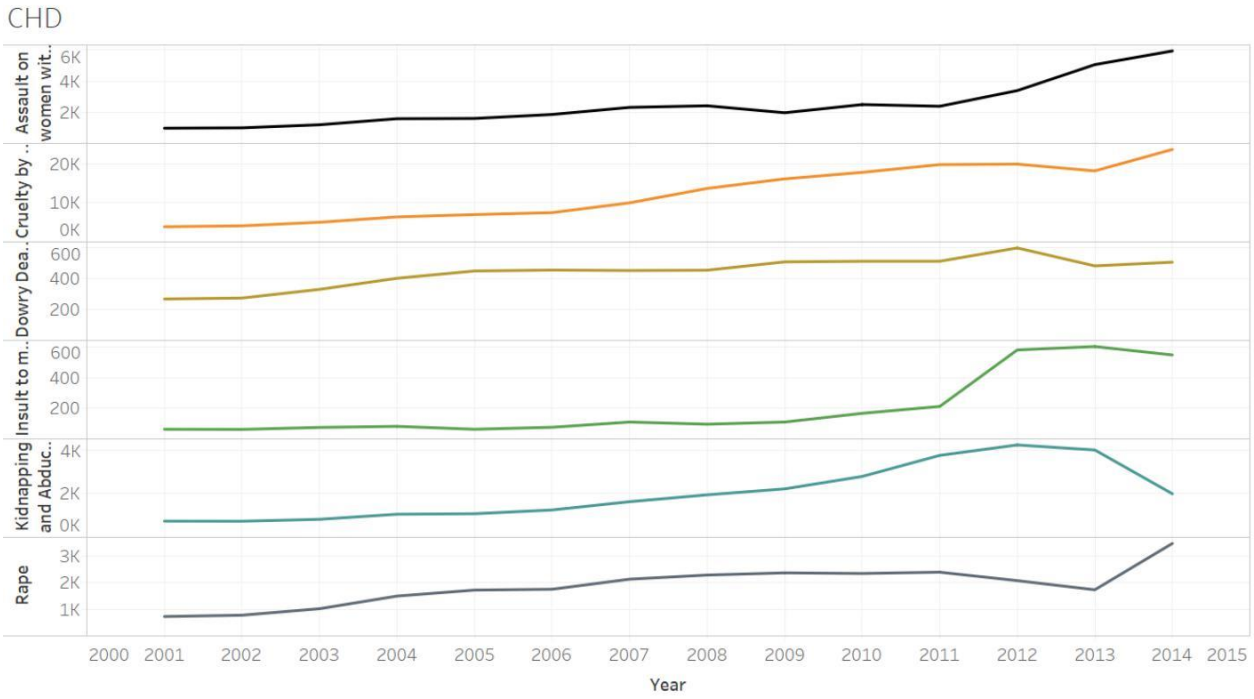
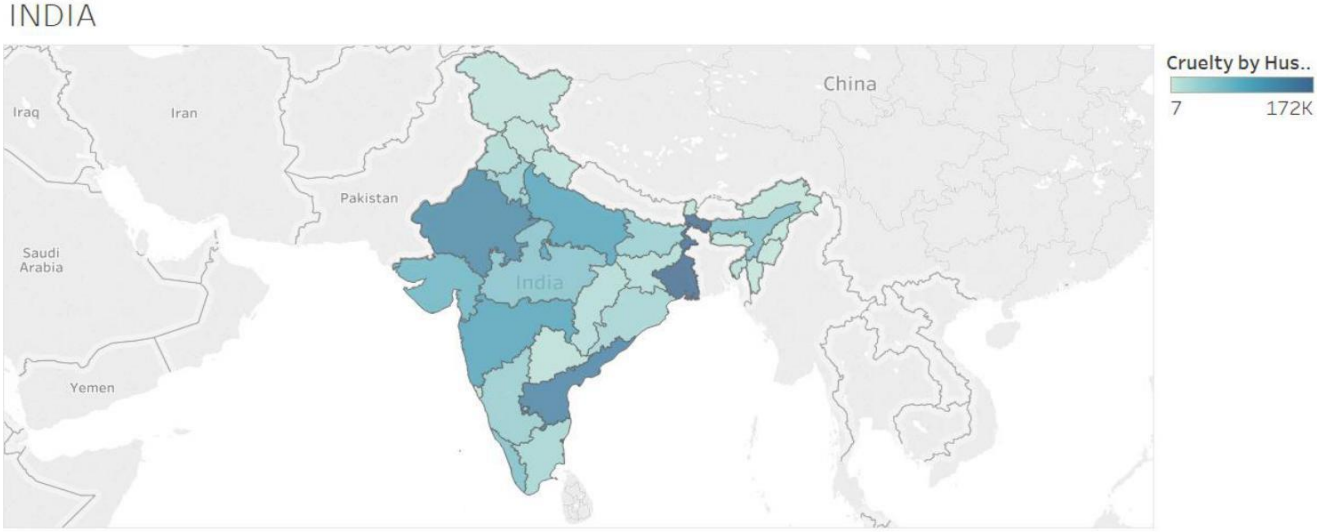


Fig 5.2 Crime on women in W.B and A.P with next year Forecast

The above image shows the different graphs of crime against women in the state West Bengal and Andhra Pradesh in the time spam of 15 years. Starting from the top the graph is of assault on women vesus years. Second is cruelty by husband on women versus year. Third chart is cases of dowry deaths versus years. Such as the last graph is of rape cases versus the year. The graph of Andhra Pradesh shows a prediction of crime rate in the next year.



The trends of sum of Assault on women with intent to outrage her modesty, sum of Cruelty by Husband or his Relatives, sum of Dowry Deaths, sum of Insult to modesty of Women, sum of Kidnapping and Abduction and sum of Rape for Year. The data is filtered on State/Ut, which keeps CHANDIGARH and WEST BENGAL.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Cruelty by Husband or his Relatives. Details are shown for State/Ut.

Fig 5.3 Cruelty by husband in India using colors to diplay no. of crimes

The above graph shows a map of India where different states are filled with different colors. The darkness of the color of the state describes the number of crimes in a year in a particular state. The color scale is mentioned in the image.

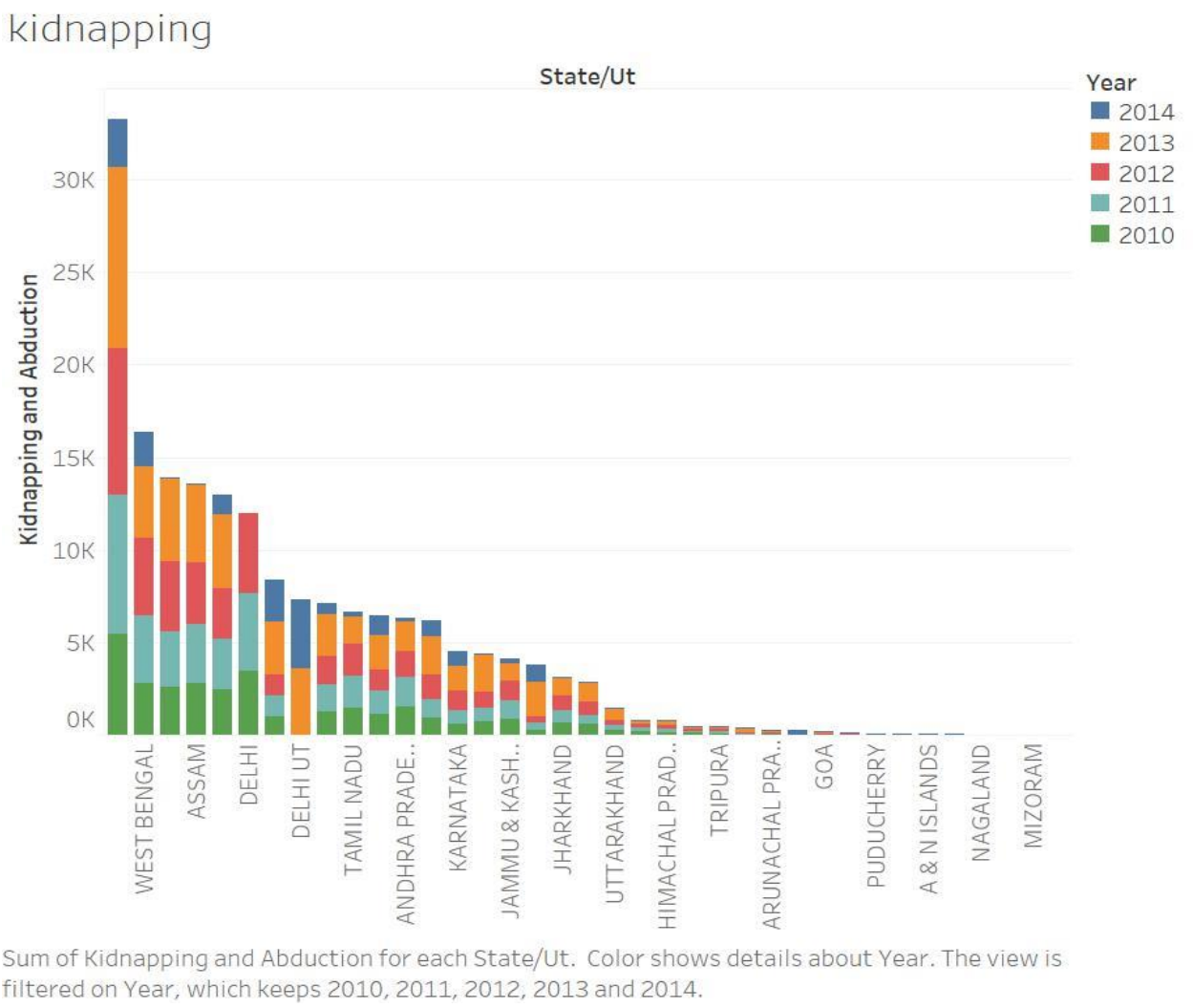


Fig 5.4 Summation of Kidnapping and Abduction of womenin 5 yrs v/s state

Kidnapping is a form of crime associated not just with women or girls but in general with the entire society. Kidnapping can be mainly with an aim of holding people as hostages and demanding cash or grants.A striking feature that can be noticed from the analysis of Kidnap of women data set using box plot shows that the rate is on a constant rise every year and the state of Uttar Pradesh has higher kidnap rates.

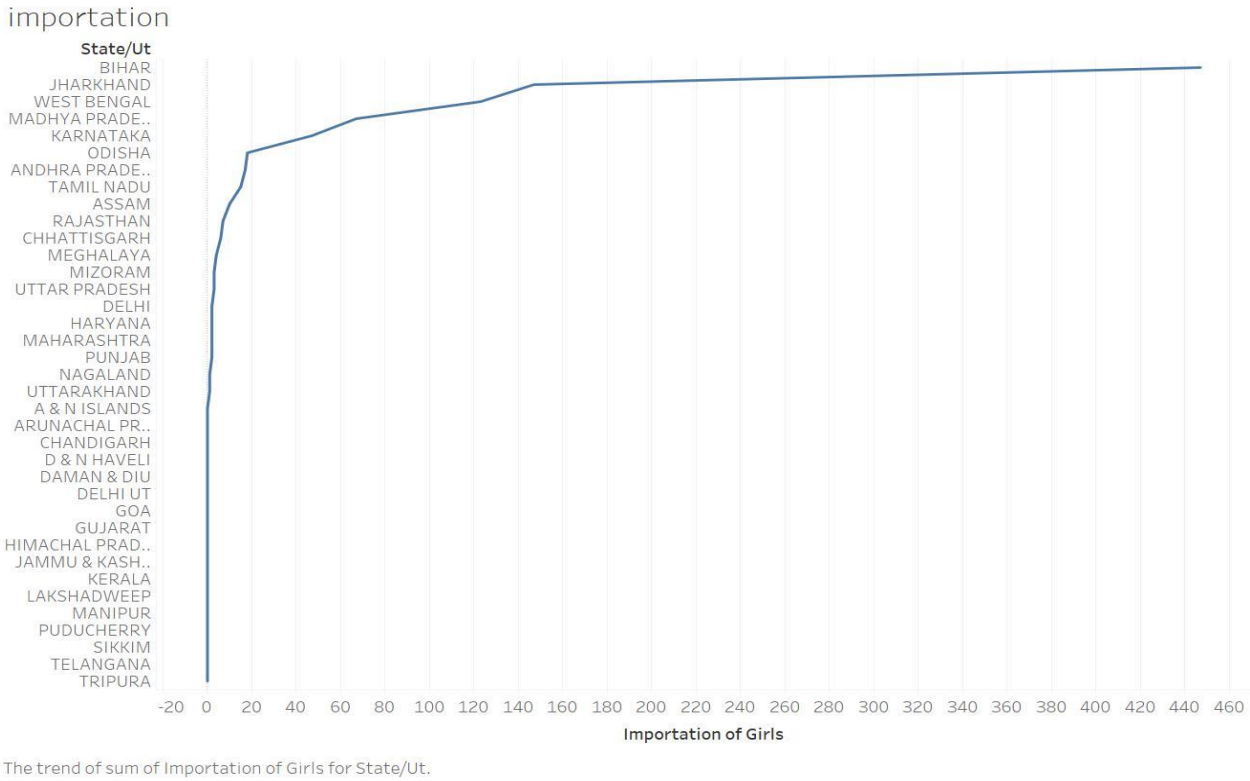


Fig 5.5 The sum of Importation of girl for states

The other more concerning crime against women is the Immoral trafficking and import of girls and women and their use as anti-social elements. As seen from the line graph the states in the southern part of the country see a lot of Immoral trafficking happening while the eastern regions are prone to more of Illegal Import of girls. It would, therefore, be meaningful if rehabilitation programs are launched and implementation machinery is set not only to eradicate the fertile source of such crimes but also for successful rehabilitation of the fallen women who are the victims of circumstances to regain their lost respect to the dignity of person to sustain equality of status, economic and their social empowerment.

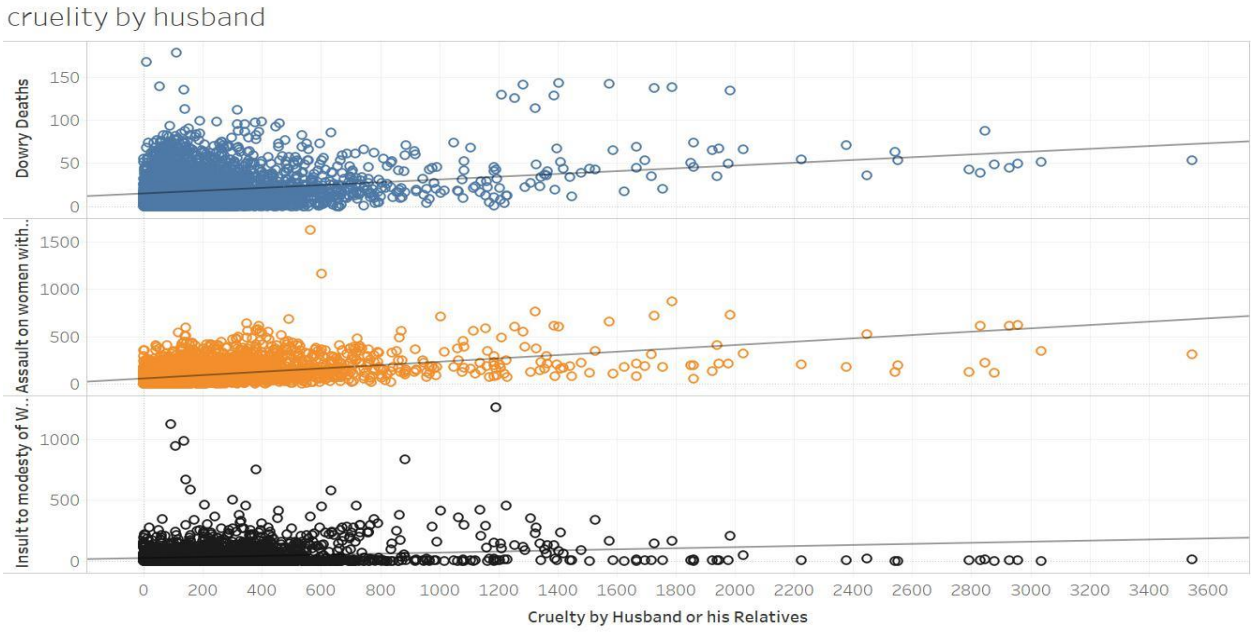


Fig 5.6 Cruelity by Husband

This slide throws light over an important but often overlooked statistic. The slide can turn out to be an effective proof to argue that a women faces more of a danger from her husband and his relatives than what she faces from the outside world. The clustering indicate that, though over the years the number of Insult to modest cases reported has stood almost constant, the cases under cruelty by husband and his relatives for dowry is on a constant rise and is quite large in number compared to insult to modest cases, proving once again that there is more of danger for an Indian women from within than outside.

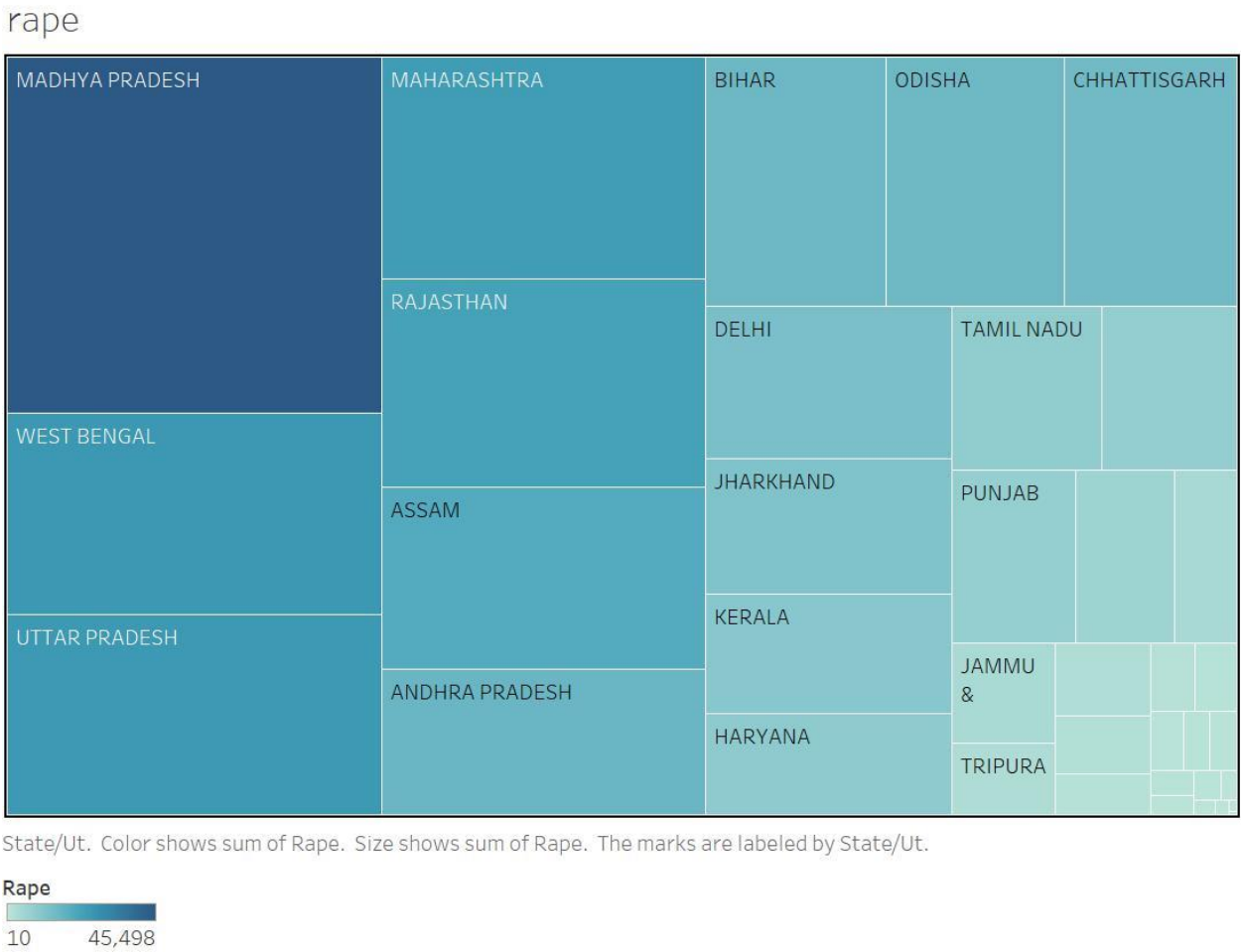


Fig 5.7 Color and Size shows number of Raps in diff States

In a country like India progressing constantly towards development, safety and rights of women play a major role in developing the trust of citizens and the way country is projected in front of the entire world. However, in the resent past there is an acute rise in the crimes against women, particularly RAPE cases. Analysis using Symbol map charts indicates that the minimum number of cases were reported in Daman and Diu . Analysis of the same data using a Heat map shows that rape cases are widely reported in the state of Madhya Pradesh and also has been seeing a constant rise from 2001, sounding an alarm to the state government for appropriate actions

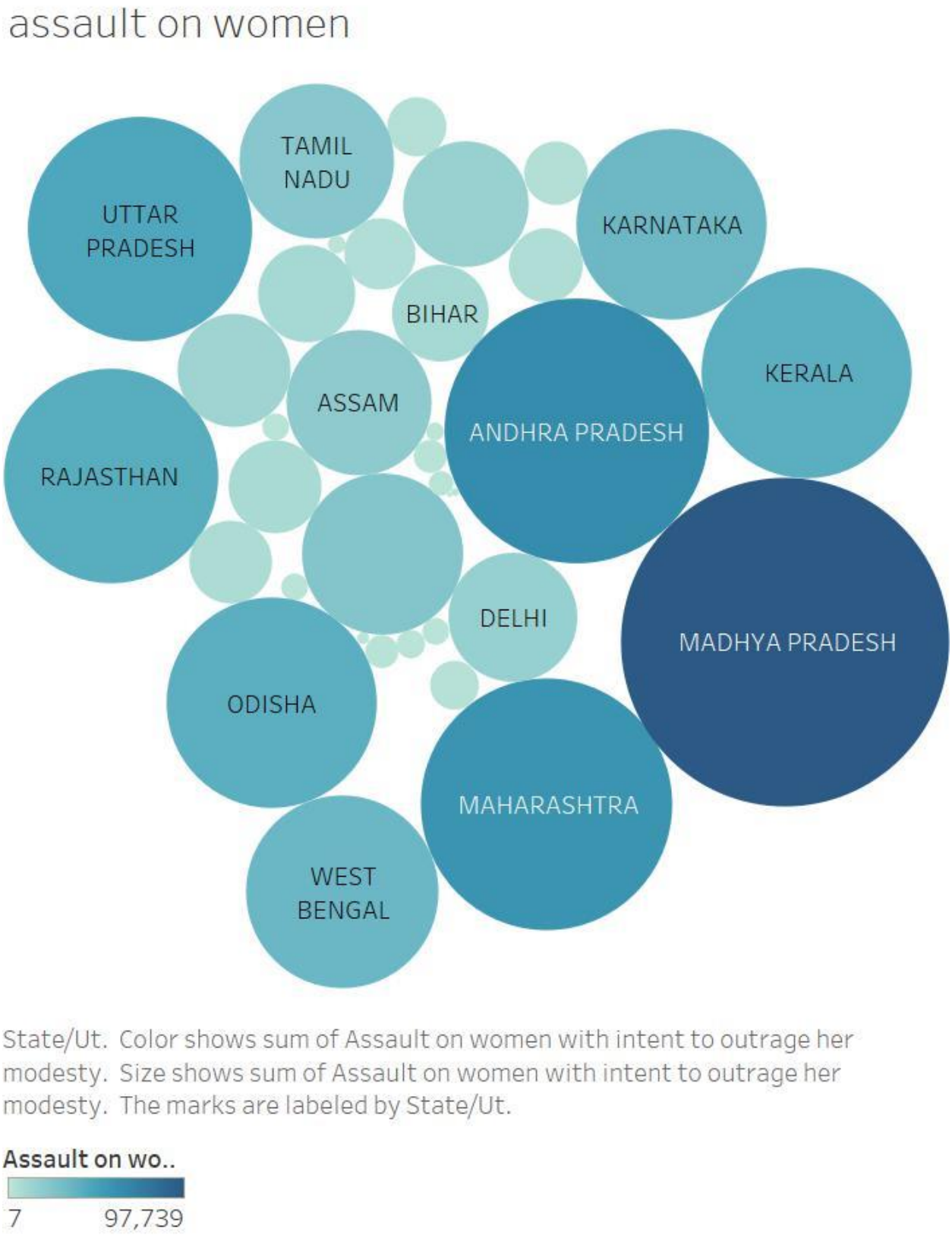


Fig 5.8 Assault on women

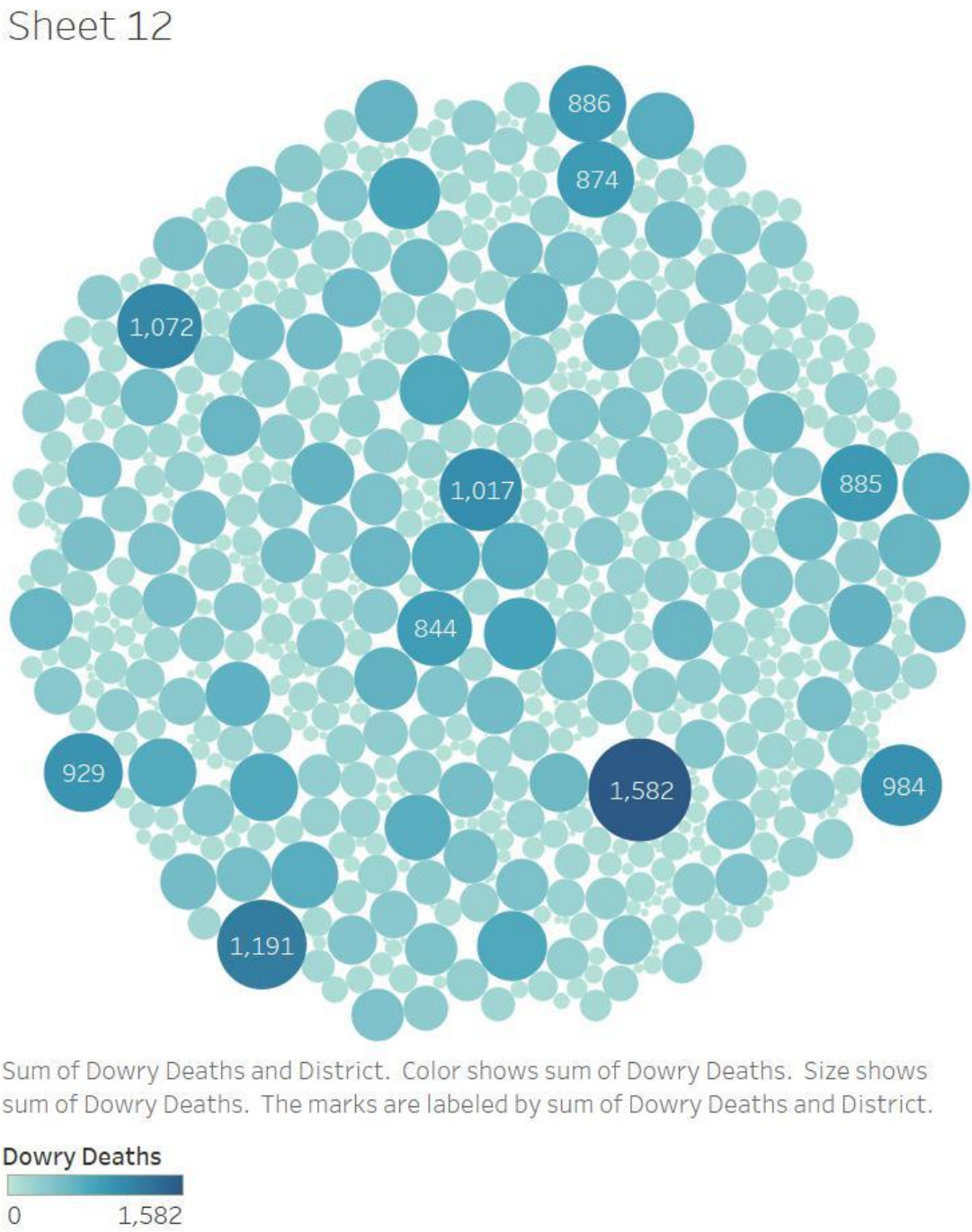


Fig 5.9 Dowry Deaths

DOWRY- best described as D-Donkeys O-Of the first order W-Who can't stand on their feet R-Rely on their wives riches Y-Yet Shameless, is a practice followed in the nation over a long time. Both giving and receiving of dowry is punishable under law. Various laws like Dowry Prohibition Act are passed in the Indian Republic. Analysis using a plot shows that the Delhi UT has higher kidnap rates and the number of dowry cases are increasing every year.

CHAPTER – 6

CONCLUSION & FUTURE SCOPE

6.1 CONCLUSION

This study based on the above mentioned methodology attempts to contribute new insights into the status of women and the effects on them due to domestic violence. This study is also expected to contribute to the realm of knowledge in the field of social work and psychology.

6.2 FUTURE SCOPE

- Potentian improvement can be made to our data collection and analysis method.
- Future research can be done with possible improvement such as more refined data and more accurate algorithm.
- Implementation of discussion forums and economic news portal including other sector apart from hydropower and going in national level.

REFERENCES

[1] https://www.ejmanager.com/mnstemps/67/67-1426406765.pdf?t=1556605149

[2] https://www.sciencedirect.com/science/article/pii/S221457961830296X

[3] K. Tsai and J. Wang, "External technology sourcing and innovation performance in LMT sectors", Research Policy, vol. 38, no. 3, pp. 518-526, 2018.

[4] https://www.academia.edu/38453831/Using_Big_Data_Analytics_in_Supply_Chain_Proposed_Framework_and_Review

[5] https://bigdata-madesimple.com/research-papers-that-changed-the-world-of-big-data/

[6] https://www.researchgate.net/publication/264555968_Big_Data_Analytics_A_Literature_Review_Paper

[7] https://thesai.org/Downloads/Volume7No2/Paper_67A_Survey_on_Big_Data_Analytics_Challenges.pdf

[8] https://thewire.in/women/rising-number-of-crimes-against-women-reflects-decay-in-indias-institutions.

[9] https://www.thehindu.com/news/cities/Delhi/crime-up-by-6-in-2018-slight-dip-in-crime-against-women/article25954302.ece