RAINFALL ANALYSIS IN INDIA USING TABLEAU

A MAJOR PROJECT REPORT

Submitted by

B.HAMSINI
17RH1A0574

Under the Esteemed Guidance of

MS.L.C USHA MAHESWARI

Assistant Professor

in partial fulfillment of the Academic Requirements for the Degree of

BACHELOR OF TECHNOLOGY
Computer Science & Engineering

MALLA REDDY ENGINEERING COLLEGE FOR WOMEN
Accredited by NBA & NAAC with A-Grade
(Autonomous Institution, UGC, Govt. of India)
Accredited by NBA & NAAC with 'A' Grade, UGC, Govt. of India
NIRF Indian Ranking, Accepted by MHRD, Govt. of India,
Band A (6th to 25th) National Ranking by ARIIA, MHRD, Govt. of India
Approved by AICTE, ISO 9001:2015 Certified Institution
AAAA+ Rated by Digital Learning Magazine, AAA+ Rated by Careers 360
3rd Rank CSR, Platinum Rated by AICTE-CII Survey, 141 NationalRanking by India Today Magazine
National Ranking-Top 100 Rank band by Outlook Magazine, National Ranking-Top 100 Rank band by Times News Magazine
2020 - 2021

MALLA REDDY ENGINEERING COLLEGE FOR WOMEN

Accredited by NBA & NAAC with A-Grade

(Autonomous Institution, UGC, Govt. of India)
Accredited by NBA & NAAC with 'A' Grade, UGC, Govt. of India
NIRF Indian Ranking, Accepted by MHRD, Govt. of India,
Band A (6th to 25th) National Ranking by ARIIA, MHRD, Govt. of India
Approved by AICTE, ISO 9001:2015 Certified Institution
AAAA+ Rated by Digital Learning Magazine, AAA+ Rated by Careers 360
3rd Rank CSR, Platinum Rated by AICTE-CII Survey, 141 NationalRanking by India Today Magazine
National Ranking-Top 100 Rank band by Outlook Magazine, National Ranking-Top 100 Rank band by Times News

---

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project work entitled "**RAINFALL ANALYSIS IN INDIA USING TABLEAU**" is carried out by B.HAMSINI (17RH1A0574), in partial fulfillment for the award of degree of BACHELOR OF TECHNOLOGY in Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad during the academic year 2020-2021.

Project Guide	**Head of the Department**
MS.L.C. Usha Maheswari	**Dr. C.V.P.R.Prasad**
Assistant Professor	**Professor**

External Examiner

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby declare that the major project entitled "**RAINFALL ANALYSIS IN INDIA USING TABLEAU**" submitted to Malla Reddy Engineering College for Women, affiliated to Jawaharlal Nehru Technological University, Hyderabad for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a result of original research work done by us.

It is declared that the technical major project report has not been previously submitted to any University or Institute for the award of Degree.

Submitted by

B.HAMSINI (17RH1A0574)

ACKNOWLEDGEMENT

We feel ourselves honored and privileged to place our warm salutation to our college **Malla Reddy Engineering College for Women** and Department of **Computer Science and Engineering** which gave us the opportunity to have expertise in engineering and profound technical knowledge.

We would like to deeply thank our Honorable Minister of Telangana State **Sri.Ch. Malla Reddy Garu,** founder chairman MRGI, the largest cluster of institutions in the state of Telangana for providing us with all the resources in the college to make our project success.

We wish to convey gratitude to our **Principal Dr. Y. Madhavee Latha,** for providing us with the environment and mean to enrich our skills and motivating us in our endeavor and helping us to realize our full potential.

We express our sincere gratitude to **Dr.C.V.P.R.Prasad, Head of the Department** of Computer Science and Engineering for inspiring us to take up a project on this subject and successfully guiding us towards its completion.

We would also like to thank our project coordinator **Mrs.V.Narmada,** for her kind encouragement and overall guidance in viewing this program a good asset with profound gratitude.

We would like to thank our internal guide **M S L C Usha Maheswari** and all the faculty members for their valuable guidance and encouragement towards the completion of our project work.

With Regards and Gratitude

B.HAMSINI(17RH1A0574)

ABSTRACT

In India several attempts have already been done in past to detect rainfall trends. our Indian farmers are highly dependent on rain, because the amount of soil moisture for crop production is totally determined by amount of rainfall. Here we analyse and forecast the long term changes in rainfall across India in divisional level. mapped the rainfall trend pattern for whole country by using visualization technique. most metrological divisions exhibited significant negative trend of rainfall in annual and seasonal scales. Data visualization is a process which aims to communicate data effectively and clearly to the user through graphical representation. rainfall are highly non-linear and complex phenomena, which require advanced computer modelling and recreation for their accurate prediction. forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way.

Tableau can connect to files, relational and Big Data sources to acquire and process data. The software allows data blending and real-time collaboration, which makes it very unique. As it does not require high level of programming expertise, any user with access to data can start using it to derive value from the data. rainfalls in India visualization in Tableau environment. The aim is to show how to extract meaningful data from the raw data and visualize it. The rainfall pattern in the States/Union Territories of India was successfully visualised.

Smart Internz

PROJECT COMPLETION CERTIFICATE

THIS CERTIFIES THAT

B.Hamsini

has successfully completed the project titled

"Analysis Of Rainfall In India Using Tableau"

with expected outcome.

APPROVEDBY

TEAM SMARTINTERNZ

June 22, 2021

INDEX

LIST OF FIGURES

CHAPTER - 1

INTRODUCTION

1.1 INTRODUCTION TO THE PROJECT

Agriculture in India has a giant history. Today, India is ranks second global in farm output. Agriculture and allied sectors like forestry and fisheries accounted for 16.6% of the GDP (Gross Domestic Product) 2009, about 50% of the complete workforce. The financial contribution of agriculture to India's GDP is gradually declining with the country's huge - based monetary growth. Agriculture exercise is a type of commercial enterprise with risk. The manufacturing of Crops depends more than a few elements like on climatic, geological, organic, political and economic factors.

Accurate information about the nature of historic yield of crop is vital modelling input, which useful to farmers & Government Corporation for decision making process in organizing acceptable policies related to next production. The advances in computing and statistics storage have supplied widespread almost of data. The challenge has been to extract expertise from this raw data, Data mining that can bridge the knowledge of the statistics to the crop yield estimation. This lookup aimed to statistics mining techniques and applies them to the a number of variables consisting in the database to set up if significant relationships can be located and the usage of fuzzy logic to discover the circumstance of crops on a number stipulations of rainfall.

In this project, a scan is made to discover the achievable of Principal aspect Analysis for broad vary forecasts of Indian summer season monsoon rainfall. Earlier this approach has drawn interest from lookup employees as it can control the elaborate and deviating options higher than the conservative statistical strategies and has effectively been utilized to a a variety of problems.

Indian summer monsoon rainfall has a share of long interval common is used in this experiment. The goal for the usage of these method for following reasons; to minimize the size of the data set, correspond to the information the usage of a minimal variety of factors that are capable to describe a vital proportion of the whole difference, and two to enhance the grasp of the modifications of the underlying system, by way of rendering these elements as correspond to fundamental 'modes' of local weather variability Summer precipitation in South India is related with monsoon circulated. Connected activity which experiences extensive variability over the spatial and temporal scale. In order to check out the inter-annual variability of summer time rainfall.

Forecasting frost formation on bridge-ways in Iowa is a vital yet tough problem. Frost varieties when water vapour in the air sublimates onto a floor and the floor temperature is beneath freezing. Only small quantities of moisture are wanted to cover surfaces with frost and create hazardous travel conditions. Prediction of rainfall is important in a range of contexts. The chance of rainfall is necessary for many choice makers who are touchy to the occurrence of precipitation. An accurate quantitative rainfall prediction can become aware of the manageable for heavy precipitation and possible associated flash flooding, as properly as providing facts for hydrologic interests.

1.2 Data Visualization

Data visualization is a process which aims to communicate data effectively and clearly to the user through graphical representation. Effective and efficient data visualization is the key part of the discovery process. It is the intermediate between the human intuition and quantitative context of the data, thus an essential component of the scientific path from data into knowledge and understanding. It is a powerful new technology having a great potential to help researchers as well as companies for building revenue decision.

Extracting relevant information and useful knowledge from large mixed-mode data spaces is complex by various challenging mark such as the limitations of data storage formats, a deficit of expert prior knowledge for real-world databases, the difficulty of visualizing the data using inefficient data mining tools, etc. Data mining is a series of steps in the knowledge discovery process, consisting of the use of particular algorithms for generating pattern, as required by the real world.

Huge amount of data becomes important not for its quantity but for the quality of information extracted from it. For a relatively complex real problem with a large data space, all knowledge generating and data mining tools would become obviously inefficient, even unassisted sometimes.

One of the key steps in Business Intelligence process where data is extracted and correlated from various data sources. In today's globalized market most organizations have multiple information repositories. Human Resources, Sales, Customer Management and Marketing all have information systems for their needs. Often each of these departments has multiple databases and applications and with the adoption of SAAS recently, more and more data is kept in different cloud offerings along with some databases in premise.

1.3 Visualization tool kits

Visualization Techniques:

The aim of the project is to forecast the passenger's inflow rate on a particular date by using prophet time series model i.e, when we select a particular date it will show the number of passengers using the airline on that particular date.

Pixel-Oriented Visualization Techniques: Using pixel is an easy way to visualize the value of the data which depends on the dimensions in which the value of dimension represents the colour of the pixel. Given a data set of n dimensions, pixel-oriented techniques create n windows on the screen, one for each dimension. The n dimension values of a record are mapped to n pixels at the corresponding positions in the windows. The corresponding values are reflected by the colours of the pixels. The data values are arranged inside a window, in a global order which is shared by all windows.

Geometric Projection Visualization Techniques: The pixel-oriented visualization techniques fail to help us in understanding the distribution of data in a multidimensional collocation. For example, compact domain in multidimensional collocation cannot be shown by pixel oriented visualization techniques. Geometric projection techniques help users in discovering zestful hurling of multidimensional data sets. Visualizing high dimensional space on a 2-D display is the main challenge that the geometric projection techniques try to address. Using Cartesian coordinates scatter plot displays 2-D source points. Using different colours or shapes a third dimension can be added to represent different data points.

Icon-Based Visualization Techniques: In Icon-based visualization techniques multidimensional data values are represented by using small icons. Two popular icon-based techniques are Stick Figures and Chernoff faces. Chernoff faces were introduced by statistician Herman Chernoff in 1973. As a cartoon human face they display multidimensional data of up to 18 variables (or dimensions). Chernoff faces help reveal trends in the data. Values of the dimensions are represented by the shapes, size, placement, and orientation of the eyes, ears, mouth, and nose, which are the components of the face.

For example, dimensions can be mapped to the following facial characteristics: head eccentricity, nose length, eye eccentricity, eye size, mouth curvature, eye spacing, nose width, mouth width, mouth openness, pupil size and eyebrow slant.

Hierarchical Visualization Techniques: The techniques for visualization discussed so far focus on visualizing various dimensions cumulatively. In spite of it would be stiff to visualize all dimensions at the analogical time for a large data set of high dimensionality. In Hierarchical visualization techniques all dimensions are partitioned into subsets (i.e., subspaces). In hierarchical manner the subspaces are visualized.

Visualizing Complex Data and Relations:

In early days, mainly for numeric data visualization techniques were used. Recently, huge amount of nonnumeric data, such as social networks came into existence. Analysing and visualizing such nonnumeric data attracts a lot of interest. There are many modern visualization techniques devoted to these type of data. For example, multiple people on the social network tag various item such as product review, blog entries and pictures. Statistics of user generated tags is visualized by a tag cloud. Often, in a tag cloud, tags are arranged in a user preferred order or listed alphabetically. The importance of a tag is indicated by colour or font size.

CHAPTER - 2

LITERATURE SURVEY

2.1 LITERATURE REVIEW

Rainfall prediction is one of the most essential and tricky job in the modern world. In general, weather and rainfall are highly non-linear and complex phenomena, which require advanced computer modelling and recreation for their accurate prediction. An Artificial Neural Network can be used to foretell the behaviour of such nonlinear systems. Soft computing deals with approximate models where an approximation answer or result is achieved. Soft computing has three basic components, namely, Artificial Neural Network, Fuzzy logic and Genetic Algorithm. It is commonly used by researchers in the field of rainfall prediction. Human brain is a highly complex, nonlinear, and parallel computer (information processing system). Neural Networks are simplified models of biological neuron system. A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. The fundamental processing element of it is an artificial neuron. Just like the natural neuron in human brain, it can receive inputs, process them and produce the relevant output. Neural Networks are successful of modelling a weather forecast system. Statistical indications chosen are capable of extracting the trends, which can be regarded as elements for creating the models. Statistical warning signs except coefficients of and kurtosis are observed appropriate to extract the hidden patterns existing in weather data. The neural community signal processing strategy for weather forecasting is successful of yielding excellent outcomes and can be considered as a choice to usual meteorological approaches. Accurate climate forecasting performs an integral role for planning day to day activities. Neural community has been use in numerous meteorological purposes together with climate forecasting.

2.2 EXISTING SYSTEM

Despite the increasing demands for high quality rainfall datasets, accurate quantitative documentation of regional rainfall analysis (gridded data) remains as one of the challenging tasks. This is primarily because of the large spatial and temporal variability of rainfall and lack of a comprehensive observing system. The major existing source of large scale rainfall data includes gauge observations, outputs from numerical weather prediction (NWP) models and estimates inferred from satellite observations, and each of these datasets has advantages and disadvantages.

Rain gauge gives relatively accurate point measurement of rainfall but observations are not available over most remote land areas and over oceanic areas. Another problem with the land rain gauge observations is the sampling error if the network is not adequately dense. In the case of numerical weather prediction, though there has been rapid progress in recent times, the rainfall prediction skill over tropics is still not up to the satisfactory standard. Satellite observations of infrared (IR) and microwave radiance have been used successfully to retrieve rainfall information over many parts of the globe. But the accuracy of the product is limited due to the indirect relationship of cloud top temperature and precipitation

Disadvantages of Existing System:

1. Most weighing rain gauges have to be emptied manually by staff and in winter climates are susceptible to the weighing pan freezing. In hot climates, they are susceptible to large evaporation errors, thus oil is often added to create an thin oil layer on the accumulated water's surface to limit evaporation. Extensive post processing of data from weighing rain gauges is required to achieve usable rain data and causes high power consumption and is a source of potential errors.

2. Siphoning rain gauges tend to have a rising error with the increase in rain intensity, much like most tipping bucket rain gauges. Since I do not have experience with siphoning rain gauges.

3. Optical rain sensor like the RG-11 Hydrogen Optical Rain Sensor is not a rain gauge and for this purpose it is the least reliable form of accumulated rain measurement with errors of up to 37 % recorded by our testing.

4. Winds over 25 knots produce oscillations of the equilibrium creating errors in the recording.

5. Rain gauge does not estimate or designate the intensity of the air or winds.

2.3 PROPOSED SYSTEM

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way. This study aims to determine trends in annual and seasonal rainfall and rainy days over different river basins across India. The project reviews, rainfalls in India visualization in Tableau environment. The aim is to show how to extract meaningful data from the raw data and visualize it. The rainfall pattern in the States/Union Territories of India was successfully visualised.

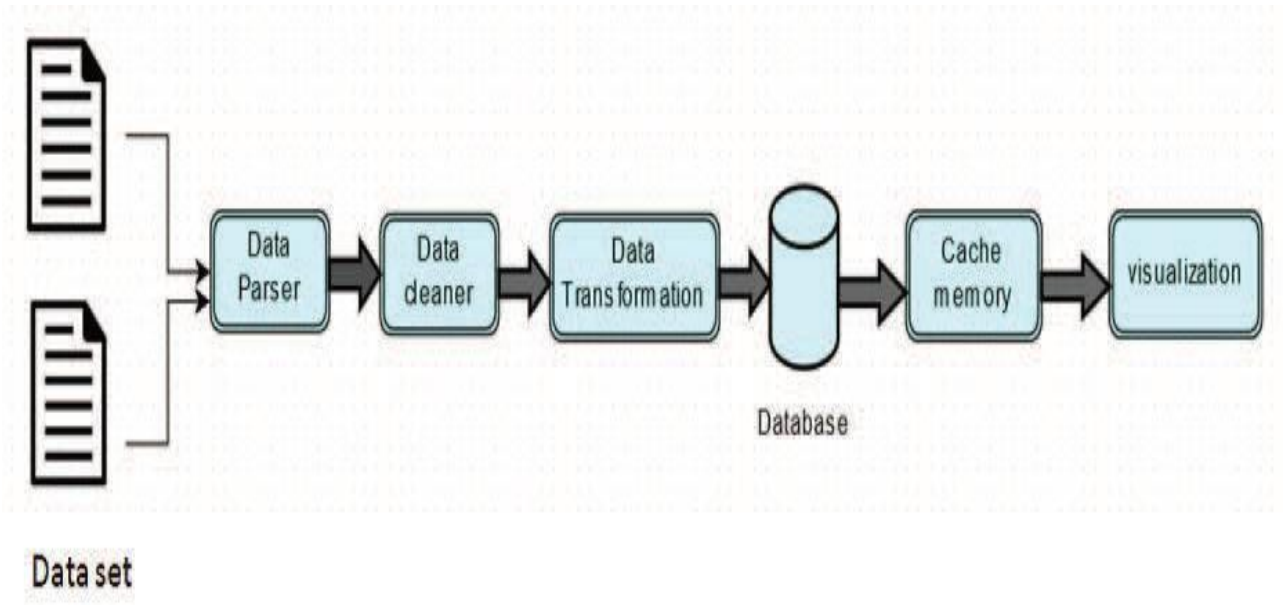


Figure 2: Flow diagram for data processing carried out by the system

Advantages of Proposed System:

1. Trend analysis is very useful for comparative analysis of date in order to measure the performances of patterns over a period of time.
2. Trend analysis (in terms of percentage) is found to be more effective in comparison with the absolutes figures/data.
3. It can be used for failure analysis and as an early warning indicator of impending problems. Where accurate historical information exists and valid relationships between variables can be established, trend analysis is a precise tool for anticipating events.
4. The project reviews, rainfalls in India visualization in Tableau environment.

2.4 FEASIBILITY STUDY

Investigation examines project feasibility, the likelihood the system will be useful to the organization. The main objective of the feasibility study is to test the Technical, Operational and economic feasibility for adding new modules and debugging old running system. All system is feasible if they are unlimited resources and infinite time.

2.4.1 Technical Feasibility

The role of the data assimilation processes is neutralized; the correction is done by forcing correct continuous fields within the target area. These experiments prove that it is necessary to correct the projection of the initial errors on the first unstable plane (the first two leading singular vectors) in order to significantly improve the forecast. These results also clearly demonstrate that the quality of the initial conditions on a limited, but quite large, area could be a major factor influencing the forecast quality.

2.4.2 Operational Feasibility

In a second stage, the focus is on operational aspects. The correction is done through the assimilation of a discrete set of simulated profiles using a 3DVAR analysis system. This leads to studying the impact of the assimilation scheme and to testing different sampling strategies. These experiments suggest that the concept of adaptive observations shows great promise in situations comparable to the one studied here. But the current assimilation systems, such as 3DVAR, require that all the structure of the target has to be well sampled to have a significant beneficial effect; sampling only the extremes does not suffice.

CHAPTER - 3

SOFTWARE REQUIREMENTS ANALYSIS

3.1 DEFINITION OF PROBLEM

Climate is one of the key components in the earth system. There are many variables such as temperature, rainfall, atmospheric pressure, humidity that constitute weather and climate. Climate is usually defined as the average weather. In broad sense, it is the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The analysis of long-term changes in climatic variables is a fundamental task in studies on climate change detection. The understanding of past and recent climate change has received considerable attention through improvements and extensions of numerous datasets and more sophisticated data analyses across the globe (Kumar et al., 2010). Global climate changes may influence long-term rainfall patterns impacting the availability of water, along with the danger of increasing occurrences of droughts and floods

INPUT DESIGN:

Input design is a part of overall system design. The main objective during the input design is as given below:

- To produce a cost-effective method of input.
- To achieve the highest possible level of accuracy.
- To ensure that the input is acceptable and understood by the user.

ERROR AVOIDANCE:

At this stage care is to be taken to ensure that input data remains accurate form the stage at which it is recorded up to the stage in which the data is accepted by the system. This can be achieved only by means of careful control each time the data is handled.

ERROR DETECTION :

Even though every effort is made to avoid the occurrence of errors, still a small proportion of errors is always likely to occur, these types of errors can be discovered by using validations to check the input data.

DATA VALIDATION :

Procedures are designed to detect errors in data at a lower level of detail. Data validations have been included in the system in almost every area where there is a possibility for the user to commit errors. The system will not accept invalid data. Whenever an invalid data is keyed in, the system immediately prompts the user and the user has to again key in the data and the system will accept the data only if the data is correct. Validations have been included where necessary. The system is designed to be a user friendly one. In other words, the system has been designed to communicate effectively with the user. The system has been designed with popup menus.

ERROR MESSAGE DESIGN:

The design of error messages is an important part of the user interface design. As user is bound to commit some errors or other while designing a system the system should be designed to be helpful by providing the user with information regarding the error, he/she has committed. This application must be able to produce output at different modules for different inputs.

OUTPUT DESIGN:

Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provides a permanent copy of the results for later consultation. The various types of outputs in general are:

OUTPUT DEFINITION

The outputs should be defined in terms of the following points:

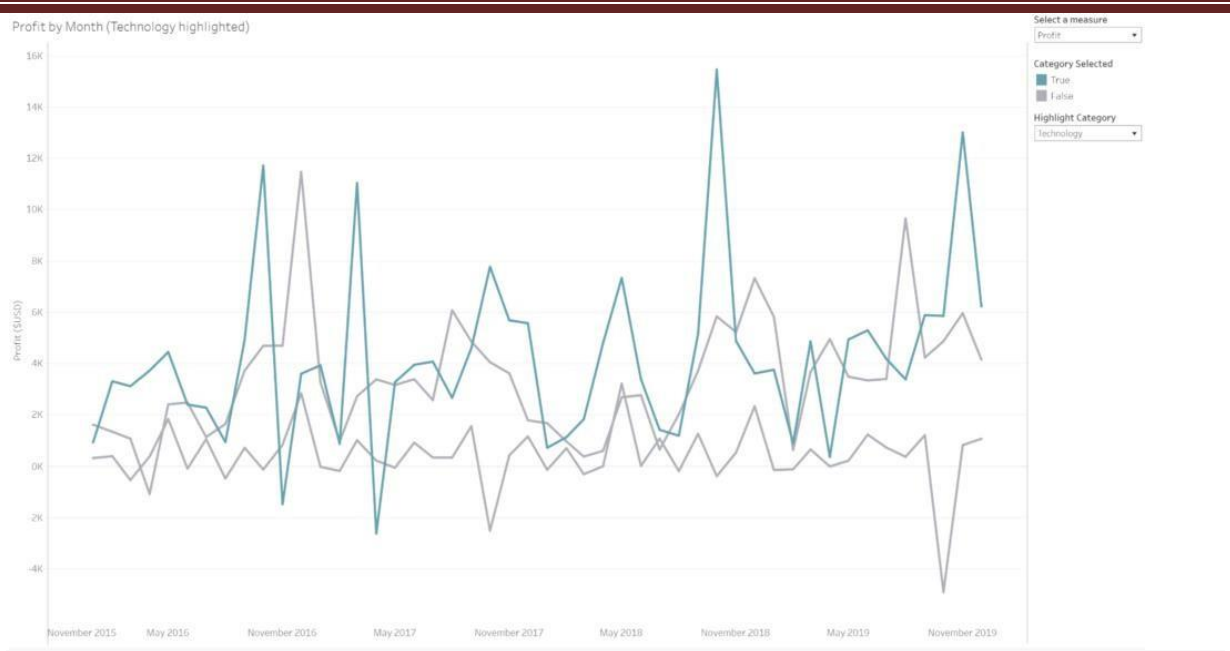
- Type of the output
- Content of the output
- Format of the output

3.2 GRAPHS

Tableau is a world leader in creating beautiful, professional visuals. But equally exciting is its customizability, which allows users to create a wide range of complex visuals in just a few steps.

3.2.1 LINES

Line graphs are helpful for displaying changes over a set period. Use them to view and predict various trends and patterns for your organization. Some common values and data points include sales, profits, customer numbers, subscribers, and rates. You can plot a single value or several in the same line chart.

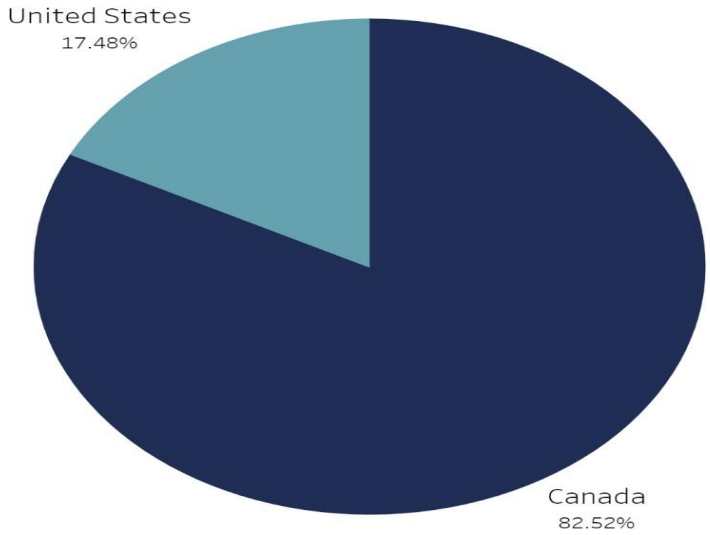

Profit by Month (Technology highlighted)
Profit ($USD)
16K
14K
12K
10K
8K
6K
4K
2K
0K
-2K
-4K
November 2015
May 2016
November 2016
May 2017
November 2017
May 2018
November 2018
May 2019
November 2019
Select a measure
Profit
Category Selected
True
False
Highlight Category
Technology

3.2.2 BARS

Bar charts are great because they're simple. You simply compare data in different categories with vertical or horizontal bars. Bar charts create an easy-to-comprehend visual representation of key data. The size of the bar "columns" represents the data values

3.2.3 PIE CHARTS

Ratios, proportions, and percentages are valuable data points. While not always the best option, pie charts are suitable for displaying this type of data when you only have a few values. Note that it's difficult to compare angles with pie charts. We suggest using them with two to three categories at most in a single chart. You just divide the pie chart into "slices" that represent each value as part of a whole.

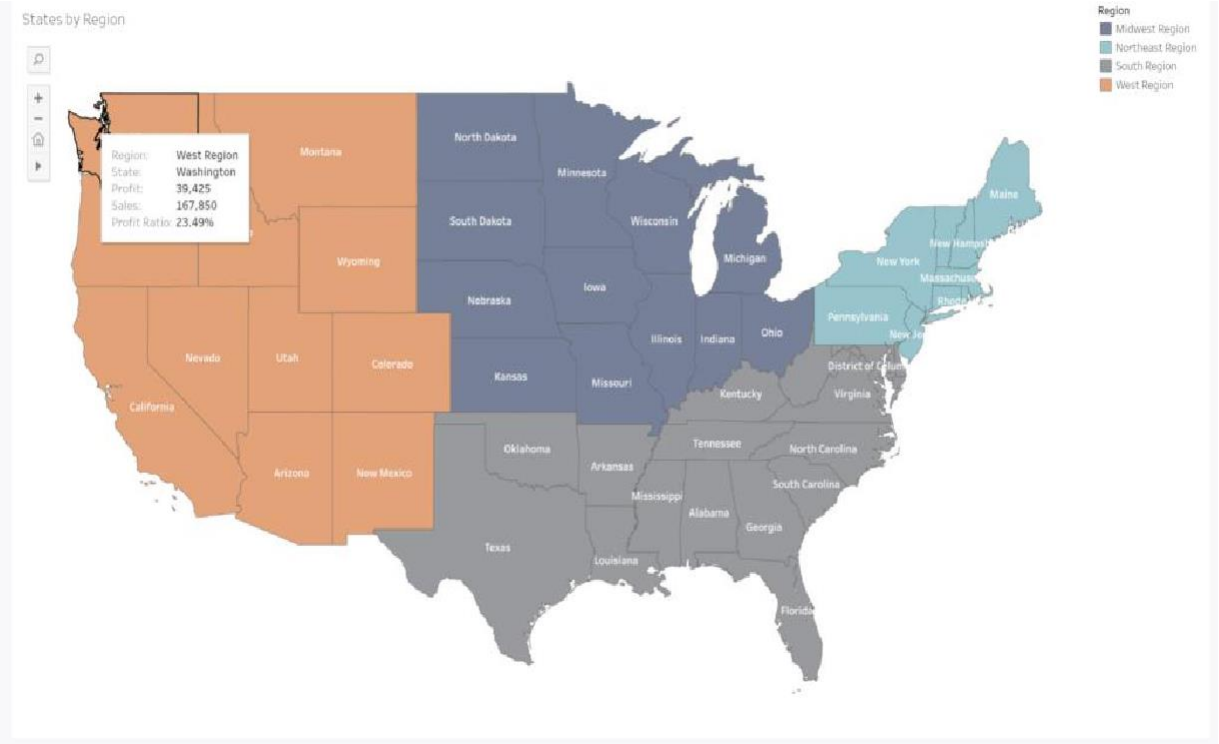


3.2.4 MAPS

Maps can be used for two purposes:

1. To give users a visual way to see divisions in company structure or regions.
2. To display measures like a heat map, showing areas of high or low performance.

Tableau comes with several built-in mapping capabilities to let you design simple maps. You'll need location data to use the tools.

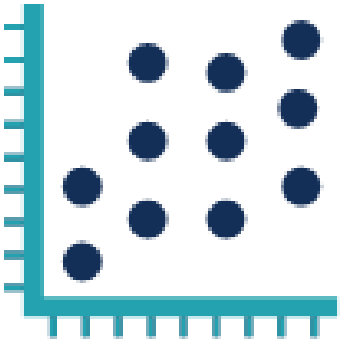


3.2.5 TREE MAPS

Tree maps are a great way to visualize the relative importance of categories within a whole. For this reason, they can be useful as interactive slicers but are less good at comparing values between categories

3.2.6 SCATTER PLOTS

Scatter charts are good at comparing data points across two variables. Additional colour or size can be added to each point to help identify further patterns.



CHAPTER - 4

SOFTWARE DESIGN

4.1 UML DIAGRAM

The Systems Development Life Cycle (SDLC), or Software Development Life Cycle in systems engineering, information systems and software engineering, is the process of creating or altering systems, and the models and methodologies use to develop these systems. Analysis gathers the requirements for the system. This stage includes a detailed study of the business needs of the organization. Options for changing the business process may be considered. Design focuses on high level design like, what programs are needed and how are they going to interact, low-level design (how the individual programs are going to work), interface design (what are the interfaces going to look like) and data design (what data will be required). During these phases, the software's overall structure is defined. Analysis and Design are very crucial in the whole development cycle. Any glitch in the design phase could be very expensive to solve in the later stage of the software development. Much care is taken during this phase. The logical system of the product is developed in this phase. Implementation In this phase the designs are translated into code. Computer programs are written using a conventional programming language or an application generator. Testing In this phase the system is tested. Normally programs are written as a series of individual modules, this subject to separate and detailed test. The system is then tested as a whole. The separate modules are brought together and tested as a complete system. The system is tested to ensure that interfaces between modules work (integration testing), the system works on the intended platform and with the expected volume of data (volume testing) and that the system does what the user requires (beta testing). Maintenance Inevitably the system will need maintenance. Software will definitely undergo change once it is delivered to the customer. There are many reasons for the change. Change could happen because of some unexpected input values into the system.

4.2 USECASE DIAGRAM

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well. The use cases are represented by either circles or ellipses.

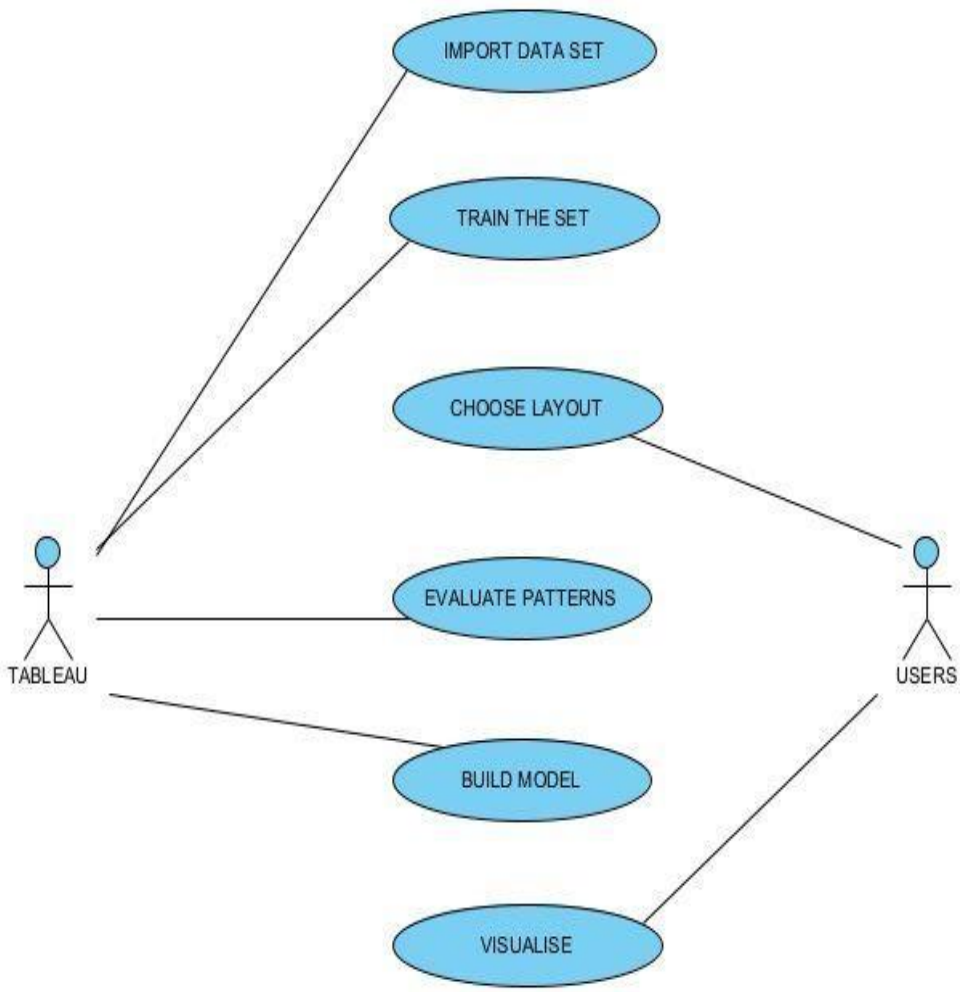


Fig 4.2: Use Case Diagram

4.3 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another.

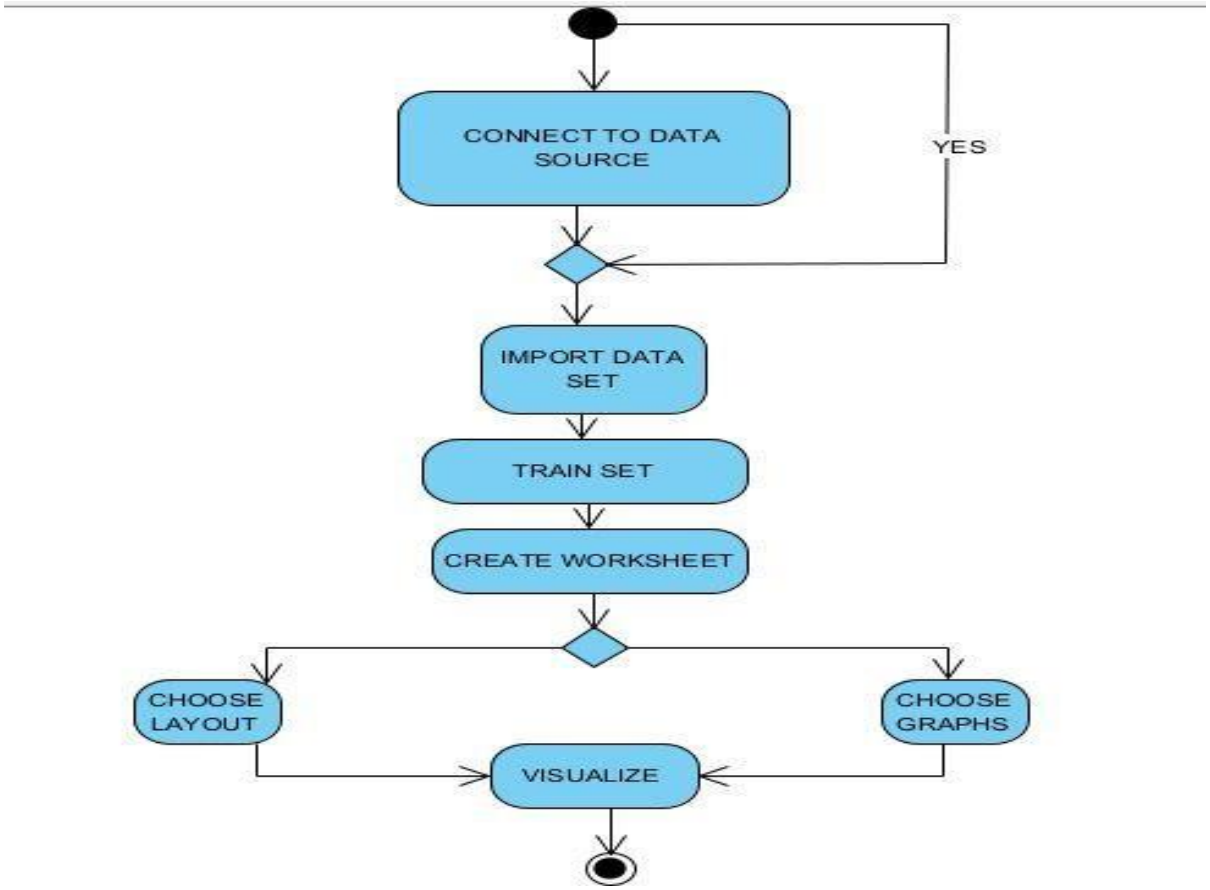


Fig 4.3: Activity model

4.4 SEQUENCE DIAGRAM

Sequence diagrams, commonly used by developers, model the interactions between objects in a single use case. They illustrate how the different parts of a system interact with each other to carry out a function, and the order in which the interactions occur when a particular use case is executed. A sequence diagram is structured in such a way that it represents a timeline which begins at the top and descends gradually to mark the sequence of interactions. Each object has a column and the messages exchanged between them are represented by arrows.

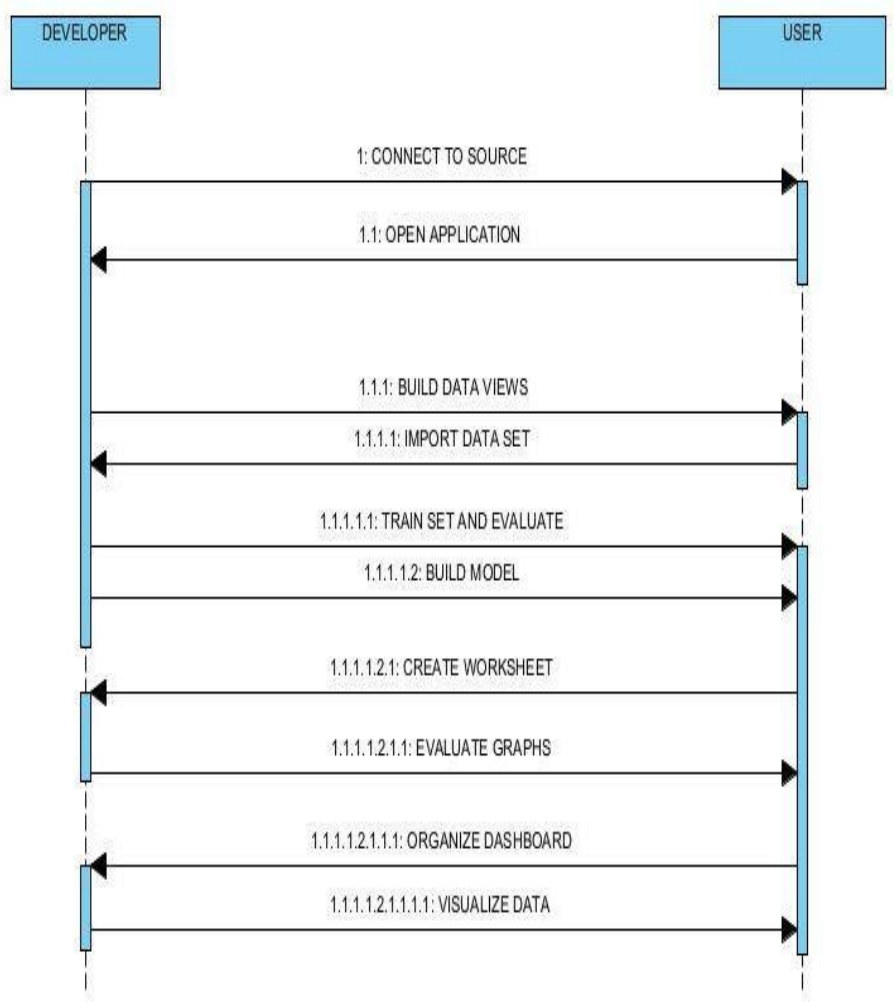


Fig 4.4: Sequence Diagram

4.5 CLASS DIAGRAM

Class diagram is a static diagram. It represents the static view of an application. Class diagram is not only used for visualizing, describing, and documenting different aspects of a system but also for constructing executable code of the software application. Class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modelling of object oriented systems because they are the only UML diagrams, which can be mapped directly with object-oriented languages. Class diagram shows a collection of classes, interfaces, associations, collaborations, and constraints. It is also known as a structural diagram.

Purpose of Class Diagrams:

The purpose of class diagram is to model the static view of an application. Class diagrams are the only diagrams which can be directly mapped with object-oriented languages and thus widely used at the time of construction. UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the applications, however class diagram is a bit different. It is the most popular UML diagram in the coder community.

The purpose of the class diagram can be summarized as:

- Analysis and design of the static view of an application.
- Describe responsibilities of a system.
- Base for component and deployment diagrams.
- Forward and reverse engineering.

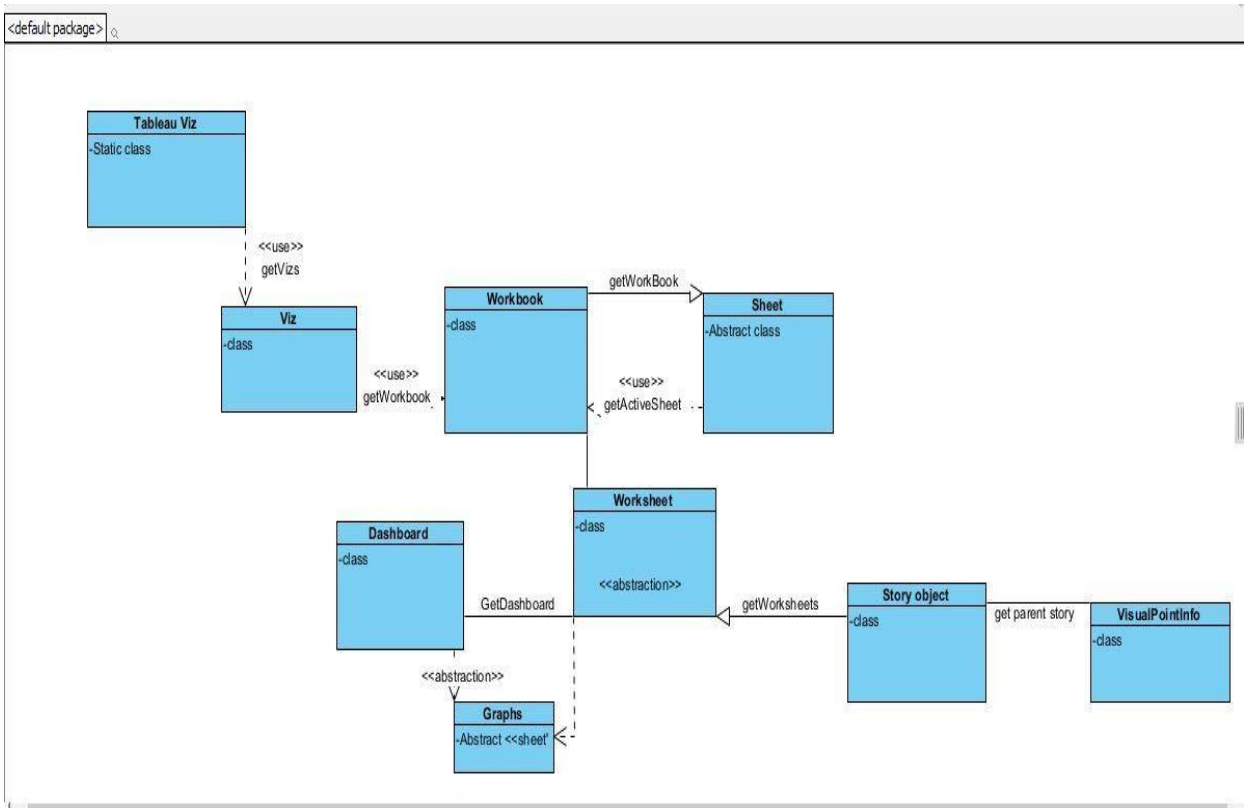


Fig 4.5 Class Diagram

CHAPTER – 5

SOFTWARE AND HARDWARE REQUIREMENTS

5.1 REQUIREMENT SPECIFICATION

Software Requirements

Tableau

Operating System : Windows

Microsoft Server : R2 (2008)

Hardware Requirements

Memory : 2GB and Higher

Screen resolution : 1366 * 768

Intel Pentium : 4 (or) higher

Tableau

Tableau is a Business Intelligence tool for visually analysing 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 real-time 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.

As a leading data visualization tool, Tableau has many desirable and unique features. Its powerful data discovery and exploration application allows you to answer important questions in seconds. You can use Tableau's drag and drop interface to visualize any data, explore different views, and even combine multiple databases easily. It does not require any complex scripting. Anyone who understands the business problems can address it with a visualization of the relevant data. After analysis, sharing with others is as easy as publishing to Tableau Server.

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.

- **Speed of Analysis** − As it does not require high level of programming expertise, any user with access to data can start using it to derive value from the data.

- **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 analyse the data by using visual tools like colours, 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, semi-structured 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 Tableau.

- **Real-Time Collaboration** − Tableau can filter, sort, and discuss data on the fly and embed a live dashboard in portals like SharePoint site or Sales force. 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. It's easy to schedule extract refreshes and manage them in the data server. Administrators can centrally define a schedule for extracts on the server for both incremental and full refreshers.

There are three basic steps involved in creating any Tableau data analysis report.
These three steps are –

- **Connect to a data source** − It involves locating the data and using an appropriate type of connection to read the data.

- **Choose dimensions and measures** − This involves selecting the required columns from the source data for analysis.

- **Apply visualization technique** − This involves applying required visualization methods, such as a specific chart or graph type to the data being analysed.

As Tableau helps in analysing lots of data over diverse time periods, dimensions, and measures, it needs a very meticulous planning to create a good dashboard or story. Hence, it is important to know the approach to design a good dashboard. Like any other field of human endeavour, there are many best practices to be followed to create good worksheets and dashboards.

Though the final outcome expected from a Tableau project is ideally a dashboard with story, there are many intermediate steps which needs to be completed to reach this goal. Following is a flow diagram of design steps that should be ideally followed to create effective dashboards.

OVERVIEW

Tableau Software is a software company headquartered in Seattle, Washington that produces interactive data visualization products focused on business intelligence. Tableau was established at Stanford University's Department of Computer Science between 1997 and 2002.

The main products offered by tableau are:

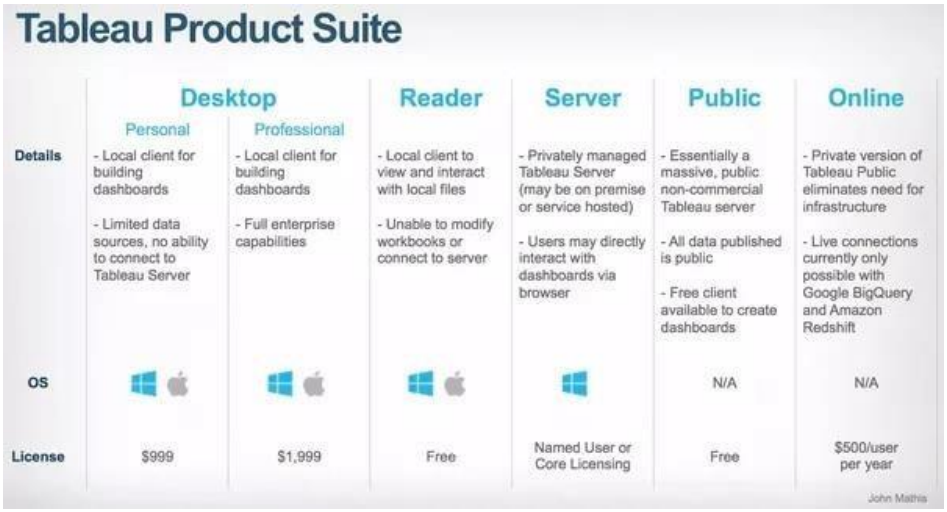
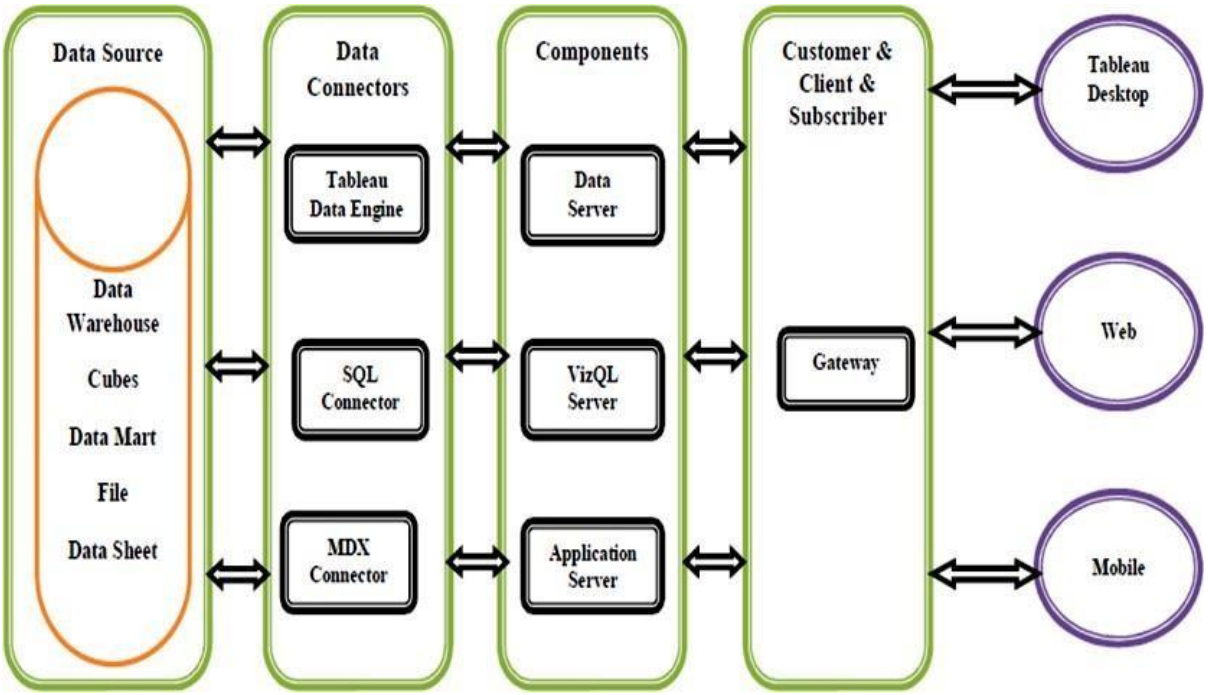


Tableau Desktop, Tableau Public, and Tableau Online, all offer Data Visual Creation and choice depends upon the type of work.

Products	Tableau Desktop	Tableau Public	Tableau Online
Formats Available	Free trial for 14 days	Absolutely Free	Workbooks are created on the Cloud and users can access them from anywhere.
	Free acess to Tableau Desktop for students and academicians for a year		
	Paid Version		
Limitation		User's data and workbooks are made public to all Tableau users.	

TABLEAU ARCHITECTURE:

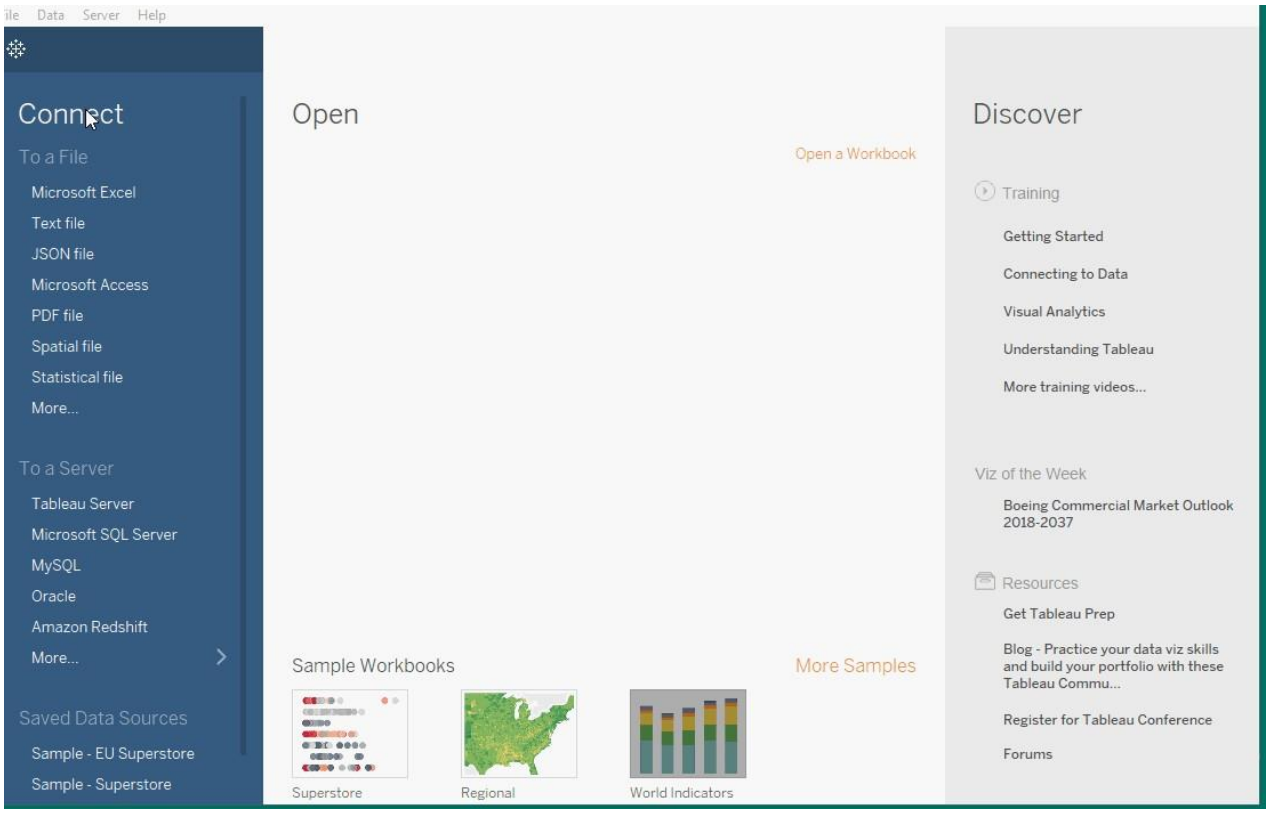
Tableau can connect to multiple data sources. These data sources can be on-premise or remotely located. It can connect to a data base, excel file, and a web application all at the same time. Tableau can connect data from heterogeneous environments. Tableau can work with all these at the simultaneously. Tableau gives simple alternatives to update your data to be quick and responsive with our quick in- memory data engine [29]. It can blend the data from multiple data sources. It can also make the relationship between various types of data sources.



CONNECTION TO DATA SOURCE

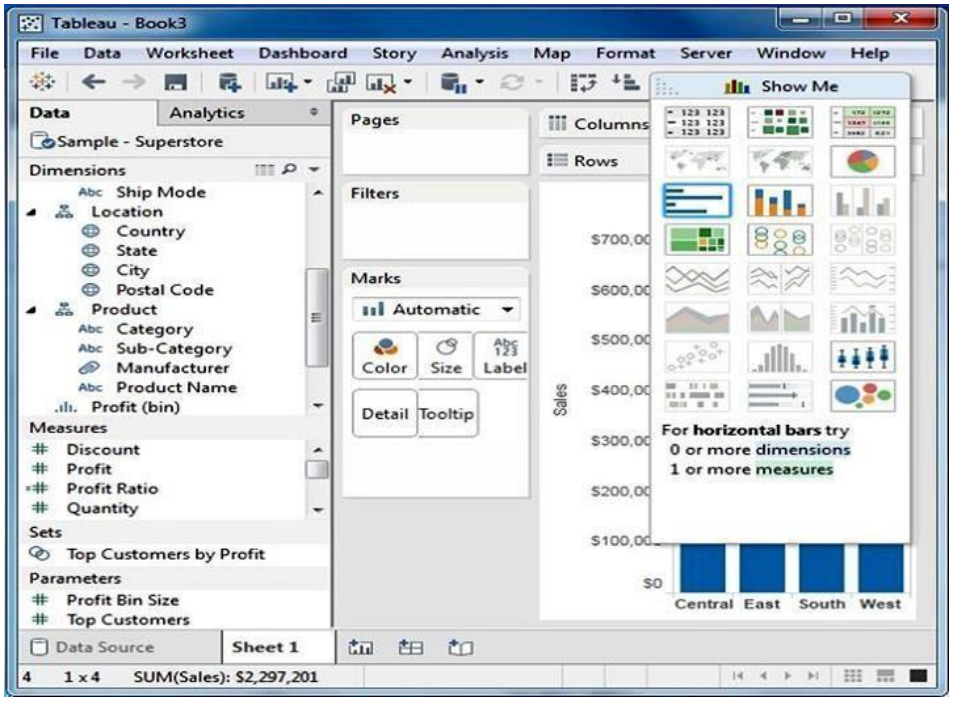
To begin working with Tableau, we need to connect Tableau to the data source. Tableau is compatible with a lot of data sources. The data sources supported by Tableau appear on the left side of the opening screen. Some commonly used data sources are excel, text file, relational database or even on a server. One can also connect to a cloud database source such as Google Analytics, Amazon Redshift, etc.

The launch screen of Tableau Desktop shows the available data sources that one can connect too. It is also dependent on the version of Tableau since the paid version offers more possibilities. On the left side of the screen, there is a connect pane which highlights the available sources. File types are listed first, followed by common server types, or the servers that have been recently connected. You can open previously created workbooks under open tab. Tableau Desktop also provides some sample workbooks under sample workbooks.



Visualization in Tableau

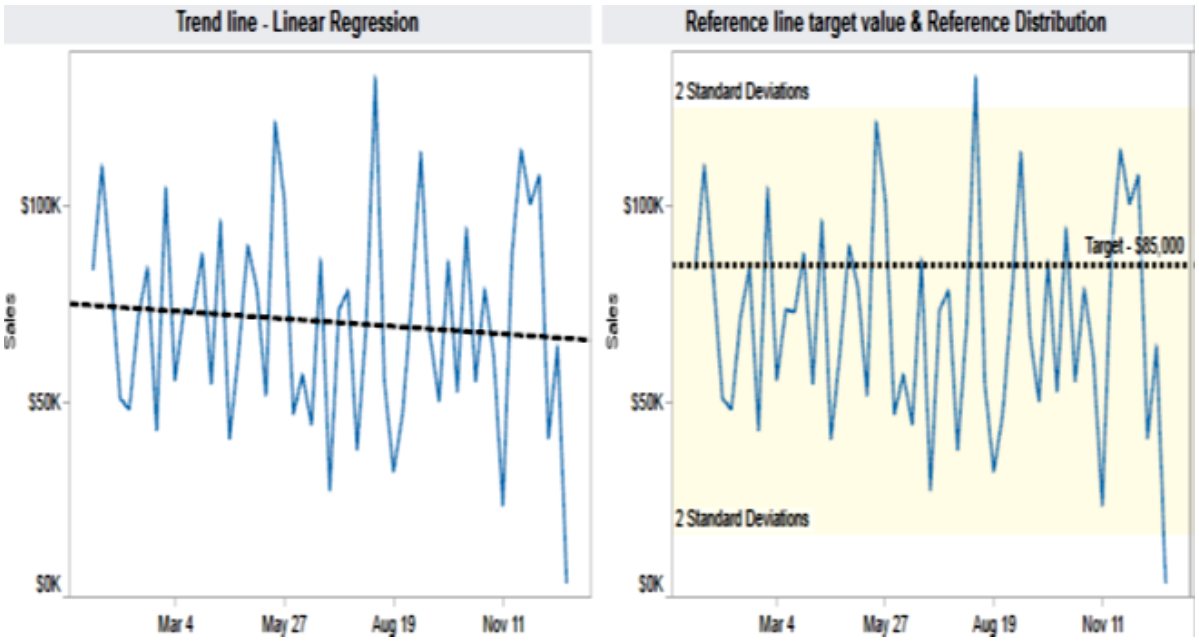
In this section, we are discussing add trend lines, reference lines, and control the way your data is sorted and filtered. Tableau's mission statement is to help you see and understand your data by enabling self-service visual analytics. The software is designed to facilitate analysis for non-technical information consumers.



This is the concept behind Tableau's show me button shown in figure. Show me to be your expert helper. Show me tells you what chart to use and why. It will also help we create complicated visualizations faster and with less effort. Show me looks at the combination of measures and dimensions we are selected and interprets what chart types display the data most effectively. At the bottom of the show me area you also see additional details regarding requirements needed for building any available chart. The time series chart requires one date, one measure, and zero or more dimensions. Pointing at other chart options in the Show Me menu changes the text at the bottom of the menu. This text provides guidance on the combination of data elements required for the chart being considered. Clicking on any of the highlighted show me icons alters the visualization in the worksheet.

Trend Lines and Reference Lines in Tableau:

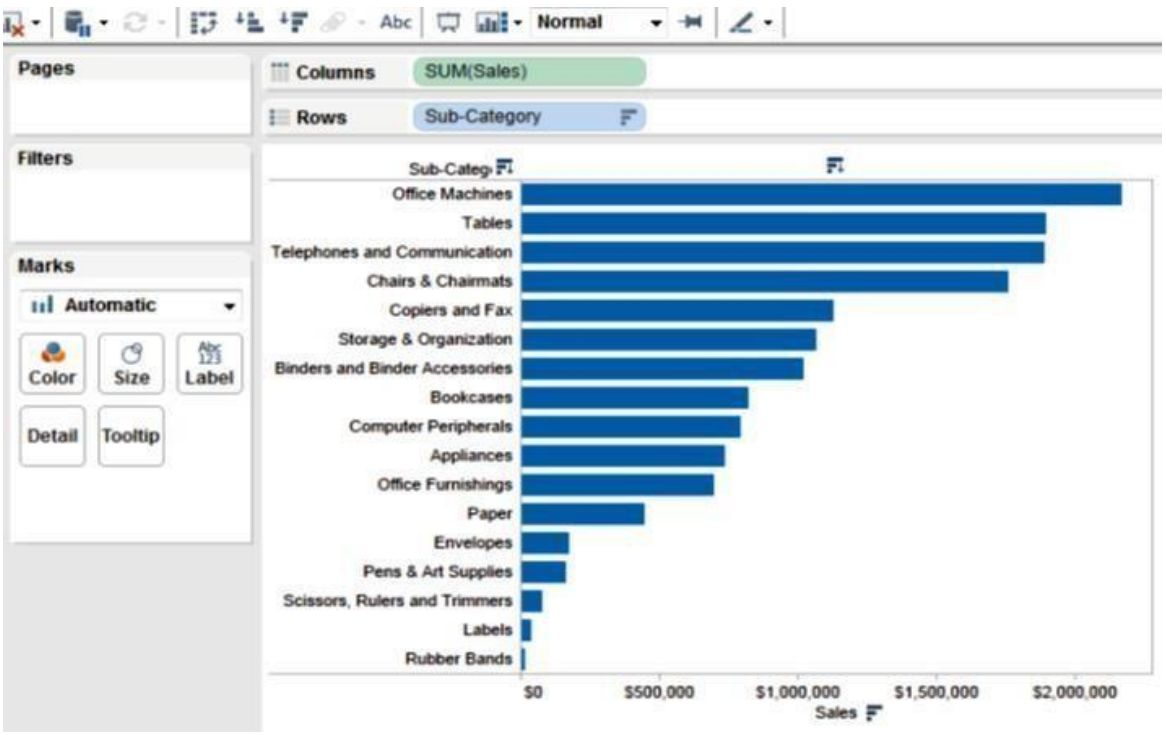
Trend lines are considered to be of great use in data visualization and analysis. In Tableau, we get a lot of options to create different types of trend lines which helps in inferring. Important patterns and trends in our data.



In visualizing granular data sometimes results in random-looking plots shown in figure 6. Trend lines help you interpret the data by fitting a straight or curved line that best represents the pattern contained within detailed data plots. Reference lines provide visual comparisons to benchmark figures, constants, or calculated values that provide insight into marks that don't conform to expected or desired values [28]. Trend lines help you see patterns in data that are not apparent when looking at your chart of the source data by drawing a line that best fits the values in view. Reference lines allow you to compare the actual plot against targets or to create statistical analyses of the deviation contained in the plot; or the range of values based on fixed or calculated numbers.

Sorting Data in Tableau

Tableau provides basic and advanced sorting methods that are easily accessed through icons or menus. The most basic way to sort is via the icons that appear in the toolbar menu. The toolbar menu sort icons provide ascending and descending sorts. In figure shows a bar chart in which a manual sort was applied from the toolbar icon. Tableau also provides sorting icons near the headings and mark axis. If we don't see an icon, hover your mouse near the area and it will appear. The clicking on the sort icon floating over the right side of the sub-category heading provides ascending and descending sorts using the text of the product category headings. The sort icons that appear over and under the mark (bar) axis provide ascending and descending sorts based on the values displayed by the marks, and also add data source order sorting.



CHAPTER – 6

IMPLEMENTATION

6.1 CREATING VIEW

We will start by generating a simple chart. In this section, we will get to know our data and will begin to ask questions about the data to gain insights. There are some important terms that we will encounter in this section.

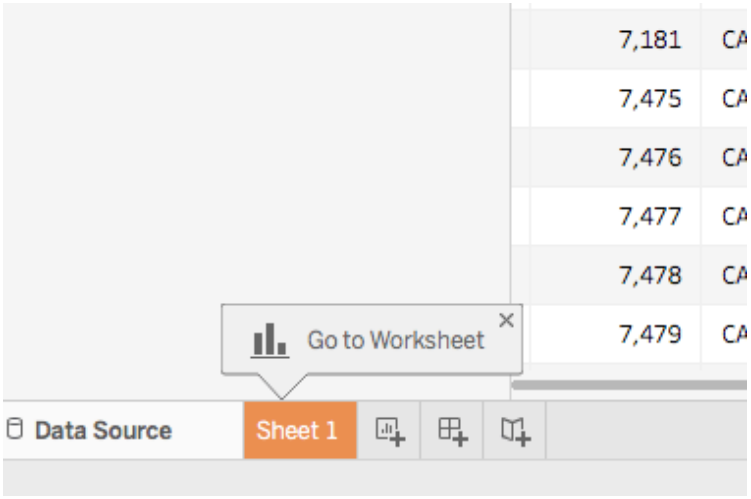
Dimensions are qualitative data, such as a countries, states, year and months. By default, Tableau automatically classifies data that contains qualitative or categorical information as a dimension, for example, any field with text or date values. These fields generally appear as column headers for rows of data, such as Customer Name or Order Date, and also define the level of granularity that shows in the view.

Measures are quantitative numerical data. By default, Tableau treats any field containing this kind of data as a measure, for example, sales transactions or profit. Data that is classified as a measure can be aggregated based on a given dimension, for example, total sales (Measure) by region (Dimension).

Aggregation is the row-level data rolled up to a higher category, such as the sum of sales or total profit.

STEPS:

1. Go to the worksheet. Click on the tab `Sheet 1` at the bottom left of the tableau workspace.



Activity 1

CREATING A MAP VIEW

Map views are beneficial when we are looking at geographic data (the Region field). In the current example, Tableau automatically recognizes that the Country, State, City, and Postal Code fields contain geographical information.

Steps:

1. Visualize the data set loaded into the sheet.

2. All the dimensions are visible in the left side of the sheet.

3. Now, drag and drop the countries dimensions into the sheet.

4. After the map loads select the area map which has only India country as per our data set.

5. Apply map operations in the sheet like map layers to show the data of the states in the India map.

6. Drag and drop the shapes into the detail.

7. Select suitable weather shapes and place it in the map.

8. Now the map shows the states with their annual rainfall.

We observe that Karnataka is having highest rainfall compared to other states r with the dark colour in the map, and the light colour indicates the lowest rainfall st0ates .

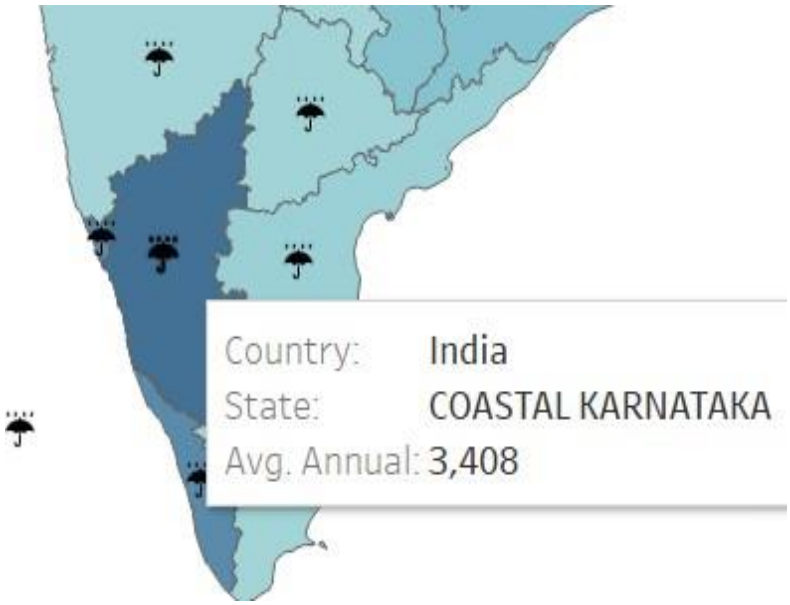
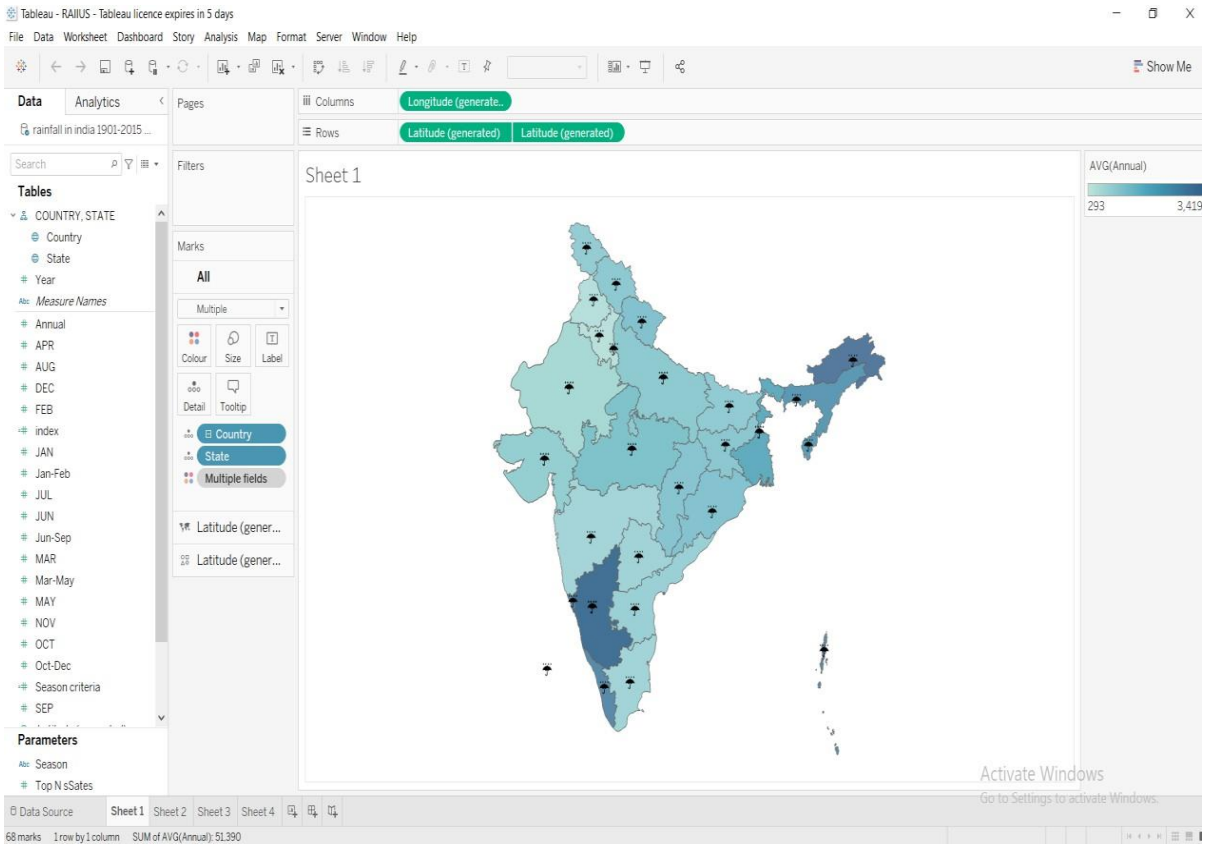


Fig. The map view shows the visualized data representation of our data set in a pictorial format.

It is very easy to understand the data looking the plotted graph representing with the average rainfall in the states throughout the year.

This gives the map with different shades of colours, where the dark colour representing the highest rainfall state and the light colour representing lowest rainfall state for our better understanding.



Getting into the details:

Maps empower us to visualize the data broadly. In the last step, we discovered that Arunachal Pradesh and Karnataka have the highest rainfall. In this section let us draw a Bar chart.

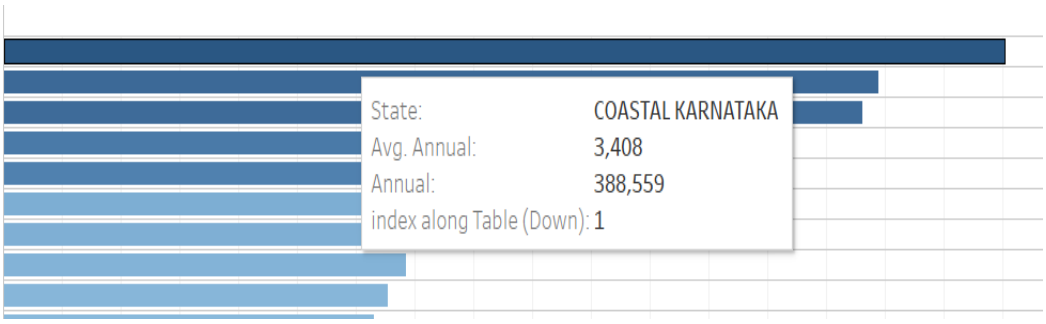
Activity 2

BAR CHARTS

In this activity we see which state is having the highest rainfall using a bar graph.

Steps:

1. First create an index before starting the activity, index is used to find the unique values.



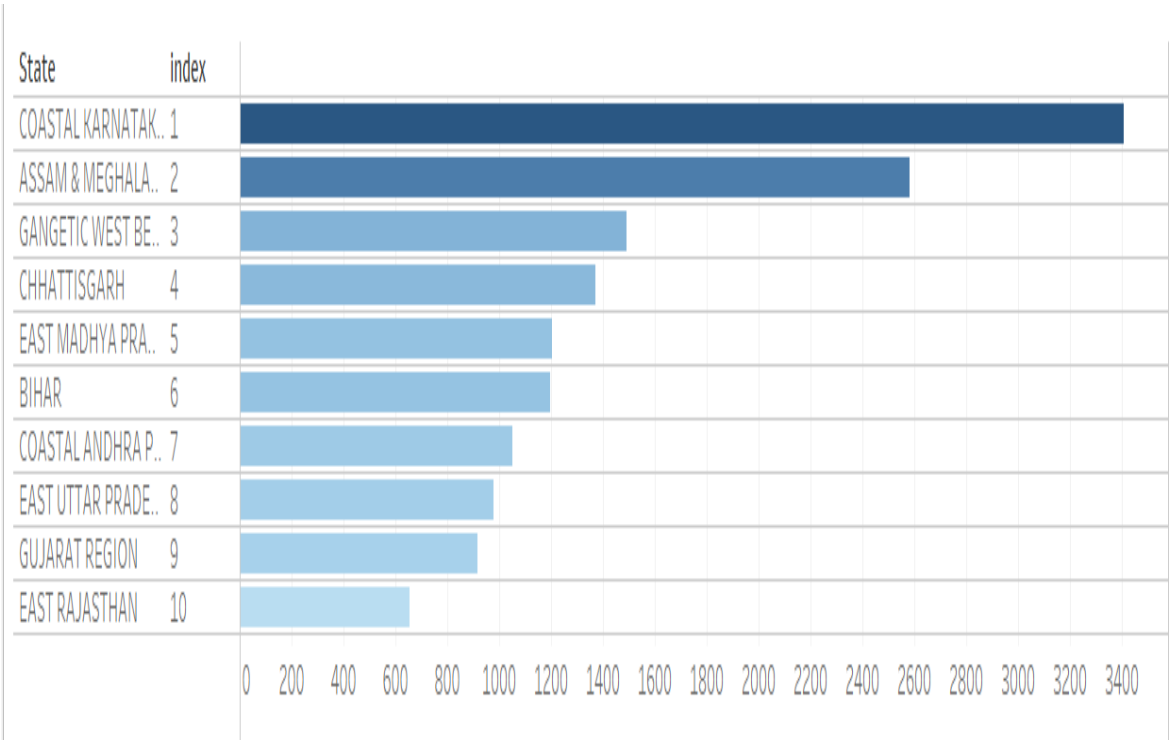
For the graph to be plotted with the states displaying in the bar chart the following dimensions should be considered.

2. Now Drag the stare dimension into the rows, all the states in the data set gets loaded into the worksheet.

3. Next drag the annual into the columns bar, and the change the sum attributes of the annual into average to get the average of all the states.

4. Now drag the index which we have created into the rows, now you the see the index values of the states according to average of rainfall.

5. Now create a parameter to display top N states.

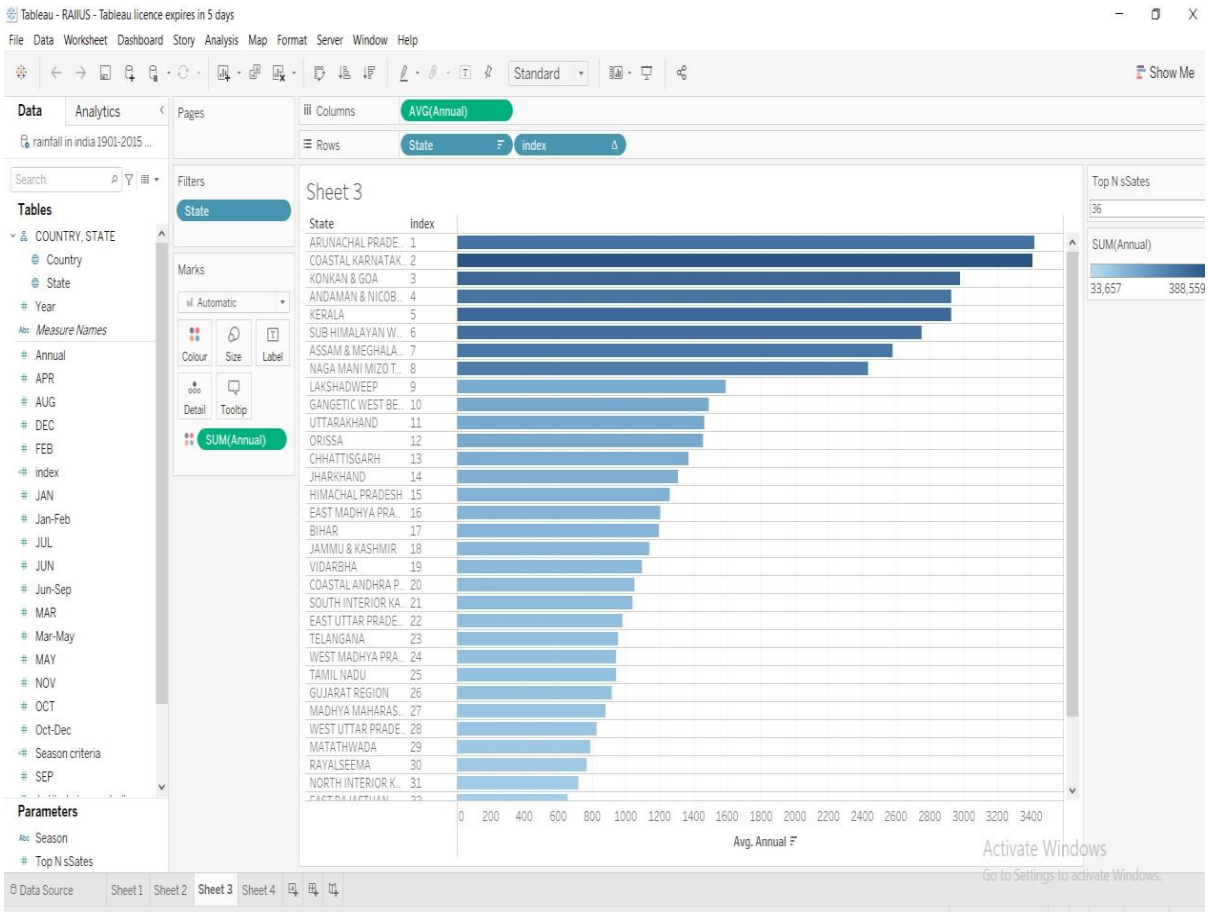


Top 10 States

6. Drag the state into filter parameter and set the field to top, top N states, states, count and apply the functions in the filter parameter so that it will be easy to display the top N states according to the given parameters.

7. Now we can see the states of the country in the descending order according to their average rainfall calculated over a period of time.

8. The Bar graph is displays with the states in descending order as shown:



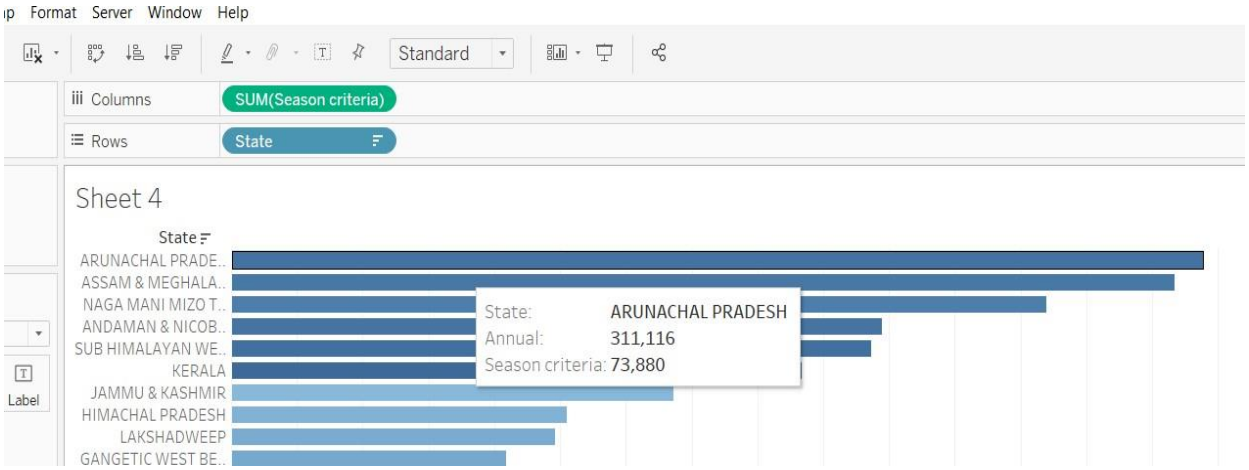
Activity 3

In this activity we will see the seasonal wise rainfall occurred in states. It is the very important aspect because analysing the rainfall based on the season helps the farmers to grow their crops according to the analysis.

To display the seasonal wise rainfall in the state we use bar graphs which gives us the proper pictorial representation of it.

Bar graphs are simple to implement. We can simply compare data in different categories with vertical or horizontal bars. The size of the bar “columns’ represent the data value.

Bar graph create an easy-to-comprehend visual representation of the key data.



In the above figure we can see the top states based on the seasonal wise rainfall analysis.

The following steps are required to obtain the graph.

Steps:

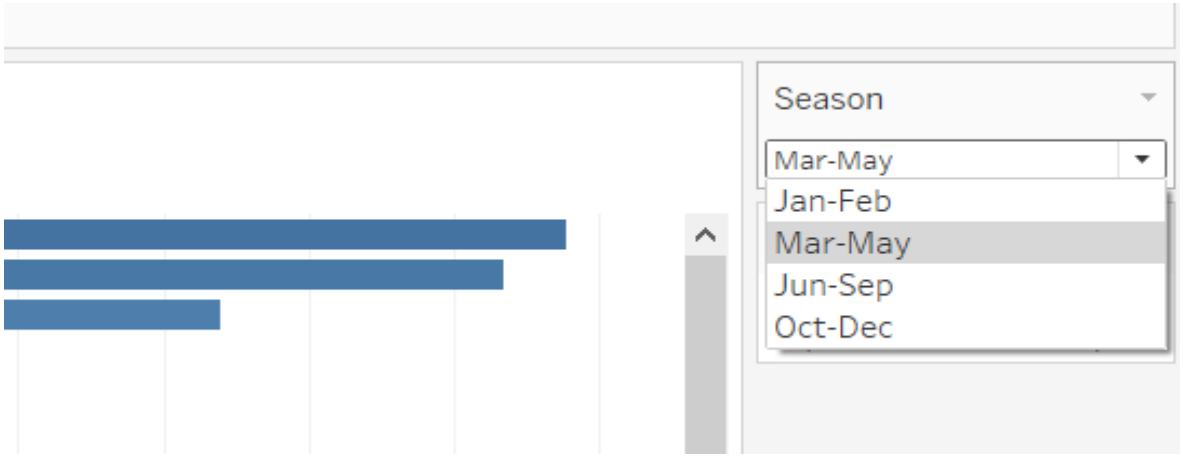
1. In the data set which we have collected has the seasonal wise data of the states over a period of time.

2. In the first step drag the states parameter into the columns, we can see all the states displayed in the sheet without any order.

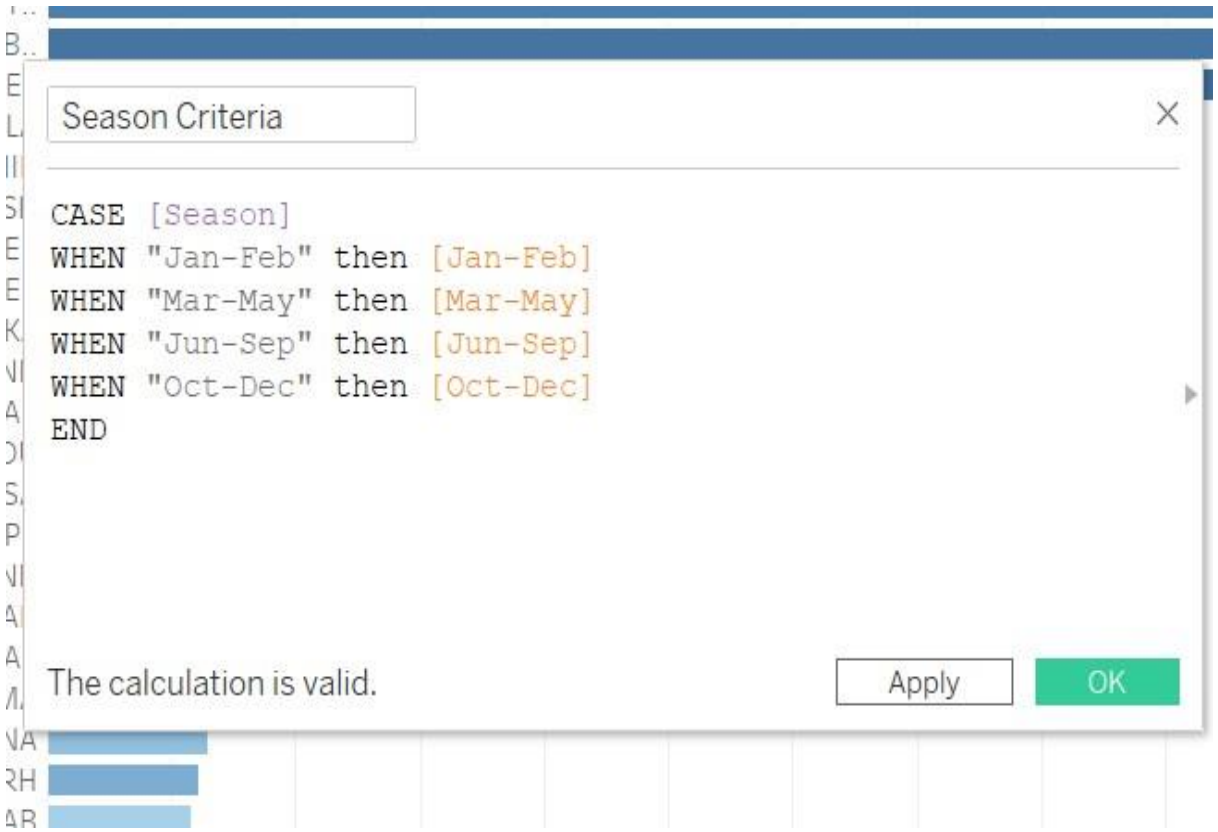
3. Next step, drag the Annual parameter into rows all the annual rainfall of the states will be displayed as bars.

4. Change the annual parameters as the average so that the average of the seasonal rainfall in the states will be displayed.

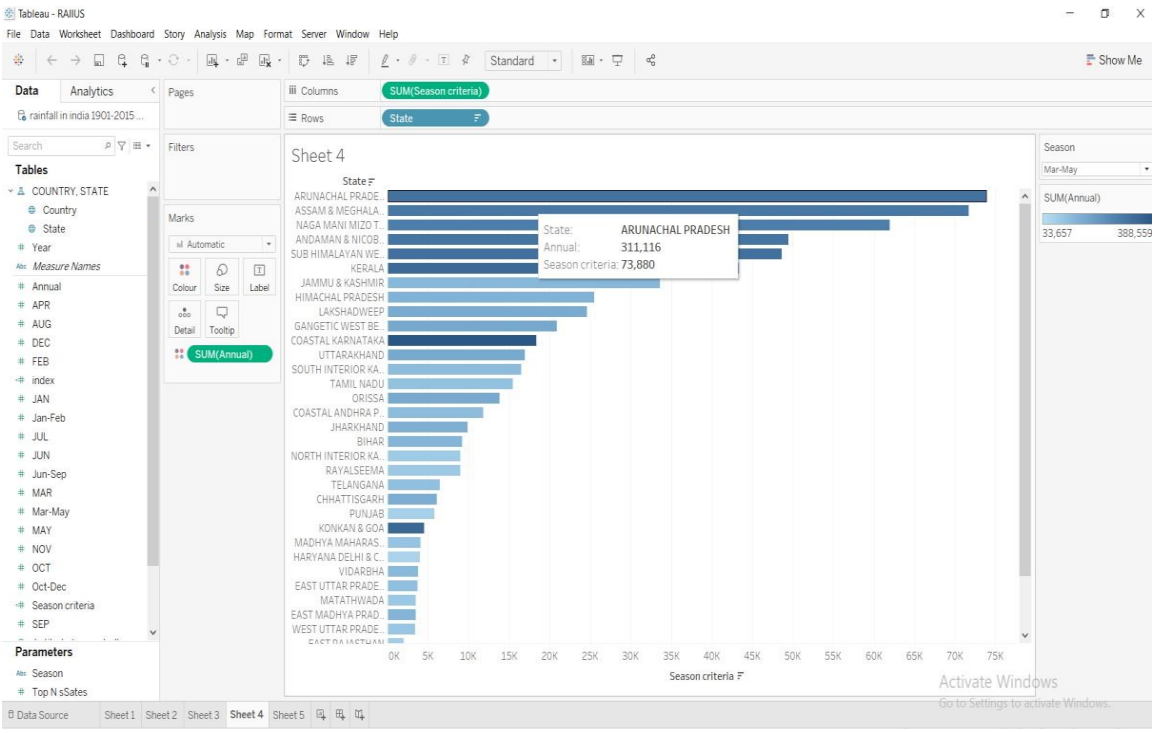
5. Create a new parameter season, set the type as string and enter all the seasonal wise data which we have in the data set Jan-Feb, March-May, June-Sep, Oct-Dec, now click on the show parameter so that it will display on the top right of the sheet as shown



6. Now create a new criteria known as season, In this criteria write a query to display the seasons mention in the dataset, the calculation should be a valid calculation.



7. Now remove the Annual criteria from columns and drag the season criteria into columns and change the sum of the season criteria into the average to get the average of all the season.



Activity 4

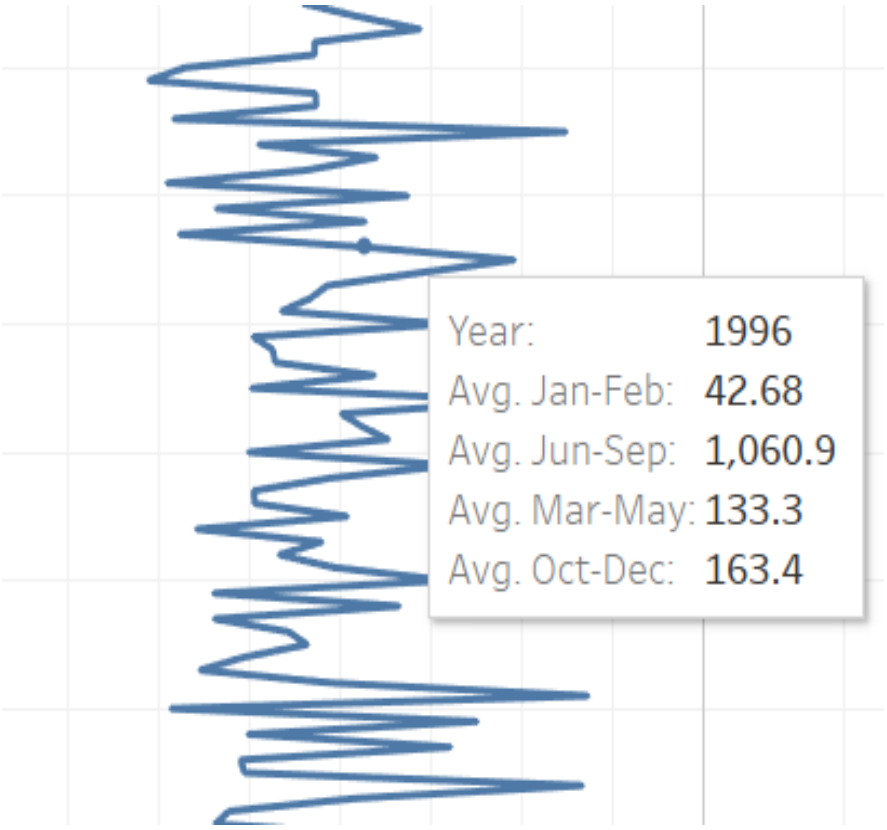
In this activity we will find the rainfall Season Average Seasonal Rainfall. Based on the season wise rainfall analysis will plot the graph.

For this we should divide the months into different sections based on the seasons we have 5 seasons in India based on that we divide the seasons into 5 parts.

Steps:

1. In the first step drag the Year parameter into the rows, the years in the dataset will be displayed in the sheet on the left side of the sheet.

2. Now drag the season parameters into the columns.



3. Here we can see the average rainfall of the states based on the seasonal criteria.

4. Change all the sum criteria of parameter into the average to display the average of all the states.

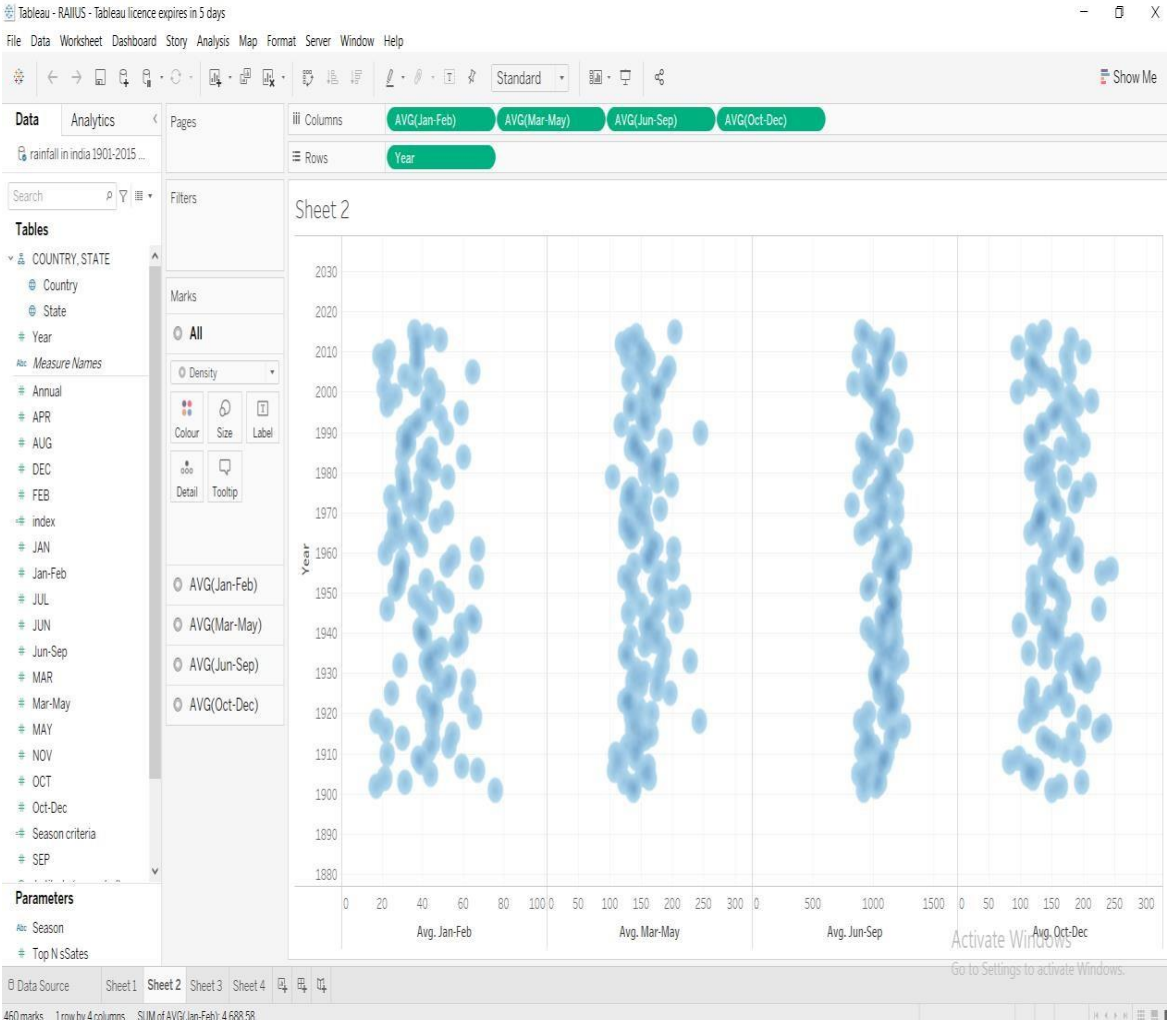
5. Next change the automate into density shape, By this density chart we can easily identify seasonal rainfall in the state based on the dense area

6. The rainfall will be highest in the region where the density shapes are high, the thick area represents the highest rainfall season.

As shown in the figure the dense shape is used to represent the highest rainfall in the season.

7. Now increase the size of the density shape so that it is very clearly visible in which season the rainfall is very high.

8. At the last step the sheet is ready with all the Seasons average of the States displays one after the other, it makes the visualization easy to know in which season or in which tie period we have the highest rainfall.



.

SAVING THE WORK:

Tableau Desktop

To save a Tableau workbook locally, Select File > Save. Specify the workbook file name in the Save As dialog box. Tableau by default saves the file with the .twb extension.

Tableau Public

With Tableau Public all the views and data is made public and anybody on the internet has access to it. Select Server > Tableau Public > Save to Tableau Public and enter the credentials.

Tableau Server

In case the data is confidential, and the story needs to be shared with the entire team, Tableau Server comes in handy. To publish a story to Tableau Server, Select Server > Publish Workbook or click Share on the toolbar. But make sure to create an account first.

CHAPTER 7

RESULT

Dashboard

A dashboard is a collection of several views, enabling one to compare a variety of data simultaneously.

Creating a Dashboard

Steps:

1. Click the new dashboard button.
2. Drag sheet1 to the empty dashboard
3. Drag sheet2 to the dashboard, and drop it on top of the sheet1 view. Both views can be seen at once. To be able to present data in a manner so that others can understand it we can arrange the dashboard to our liking.
4. On the sheet3 worksheet in the dashboard view, click under the region and clear off the show header Repeat the same process for all the other headers. This helps to emphasize only what is needed and hides away the not so important information.
5. Drag sheet4 into the sheet and arrange it into the dashboard.

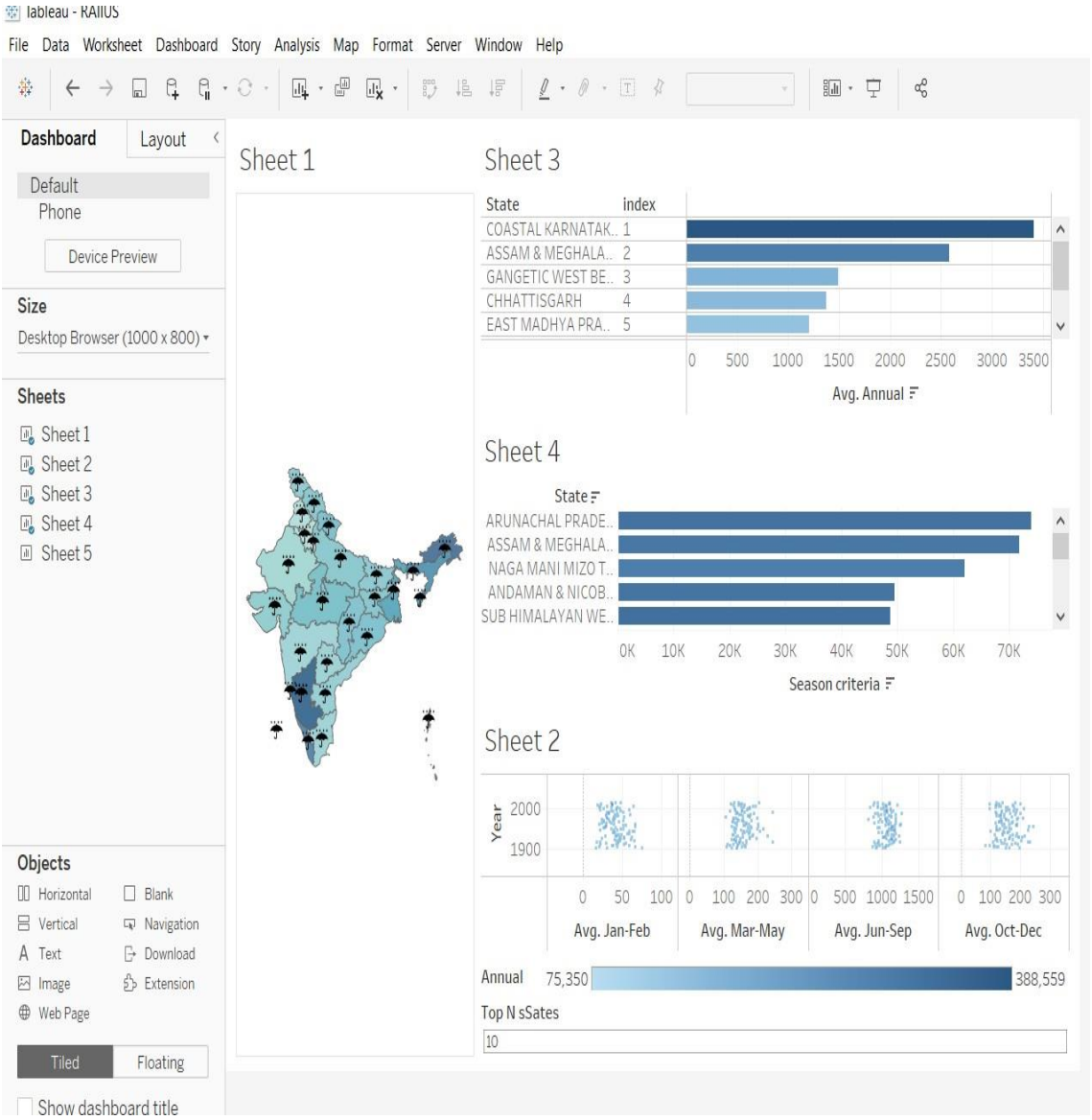
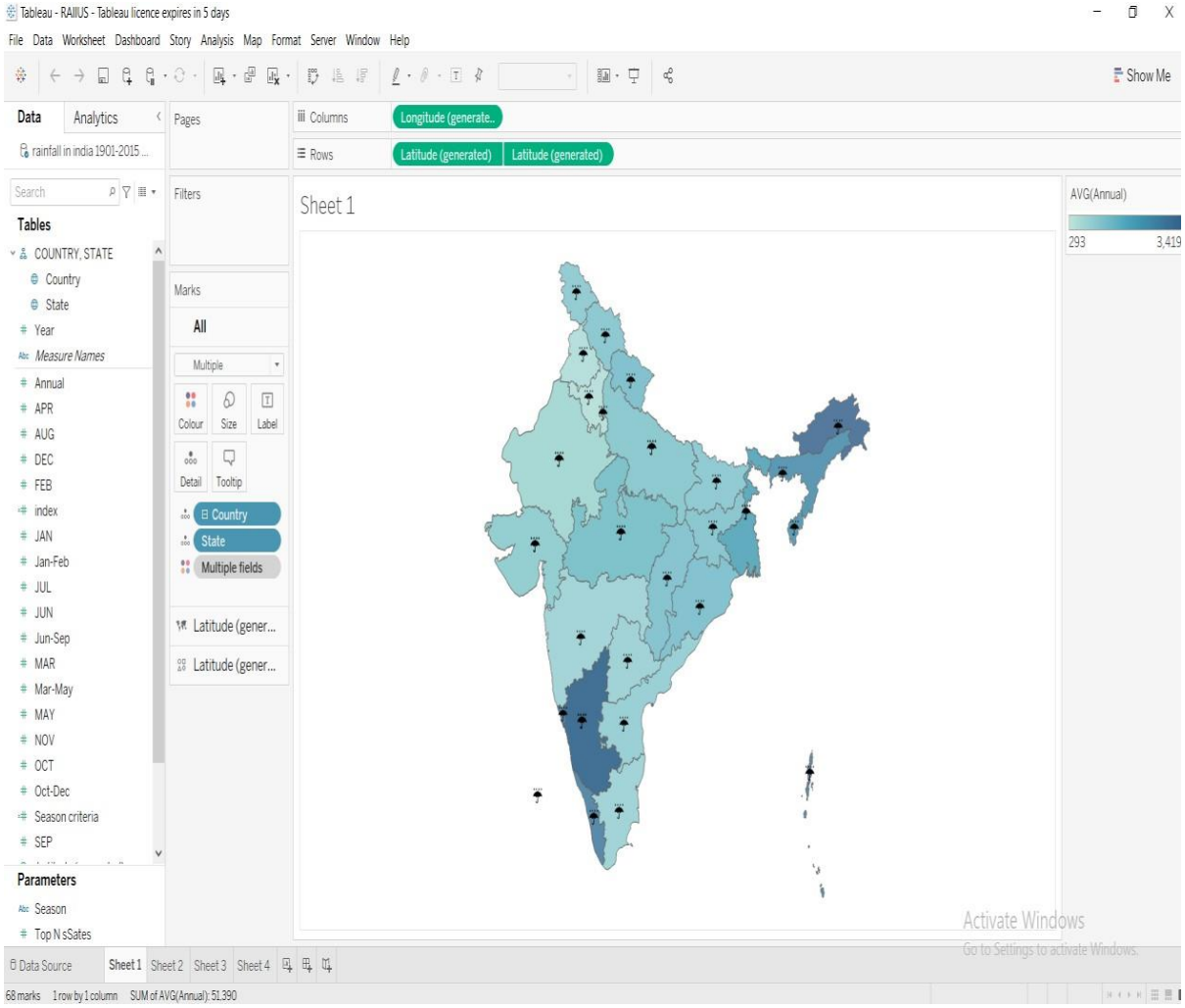
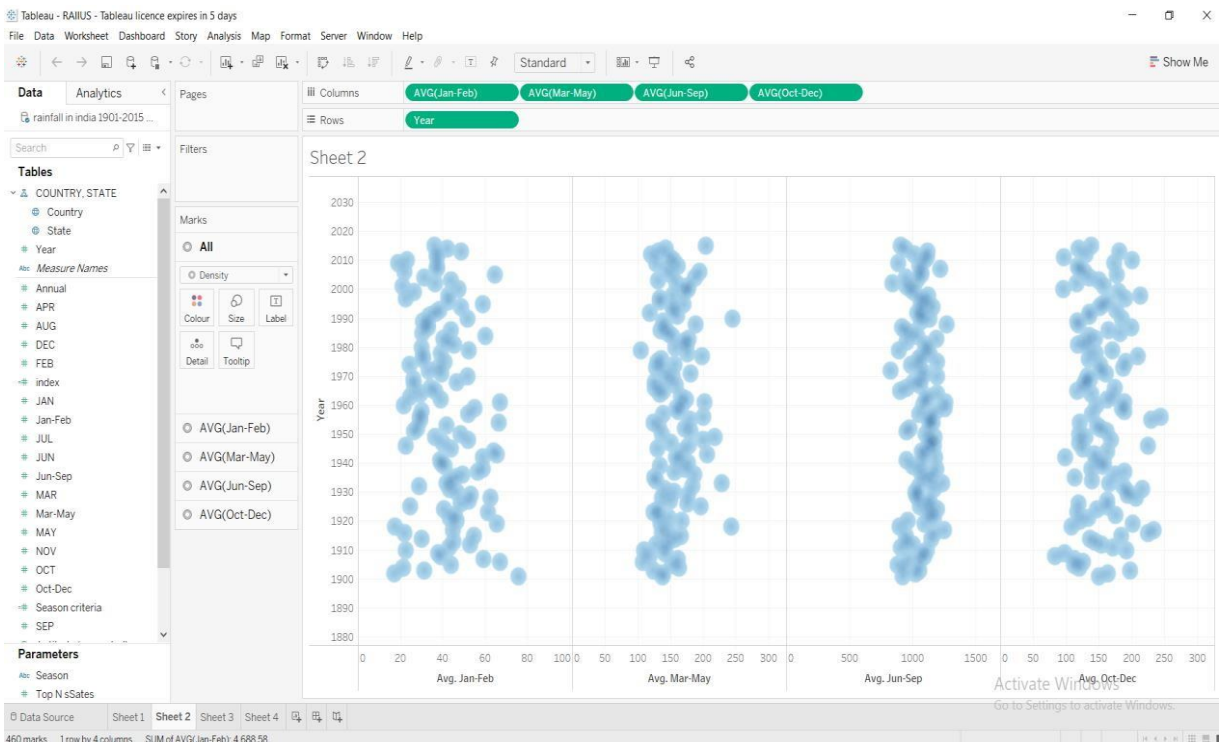


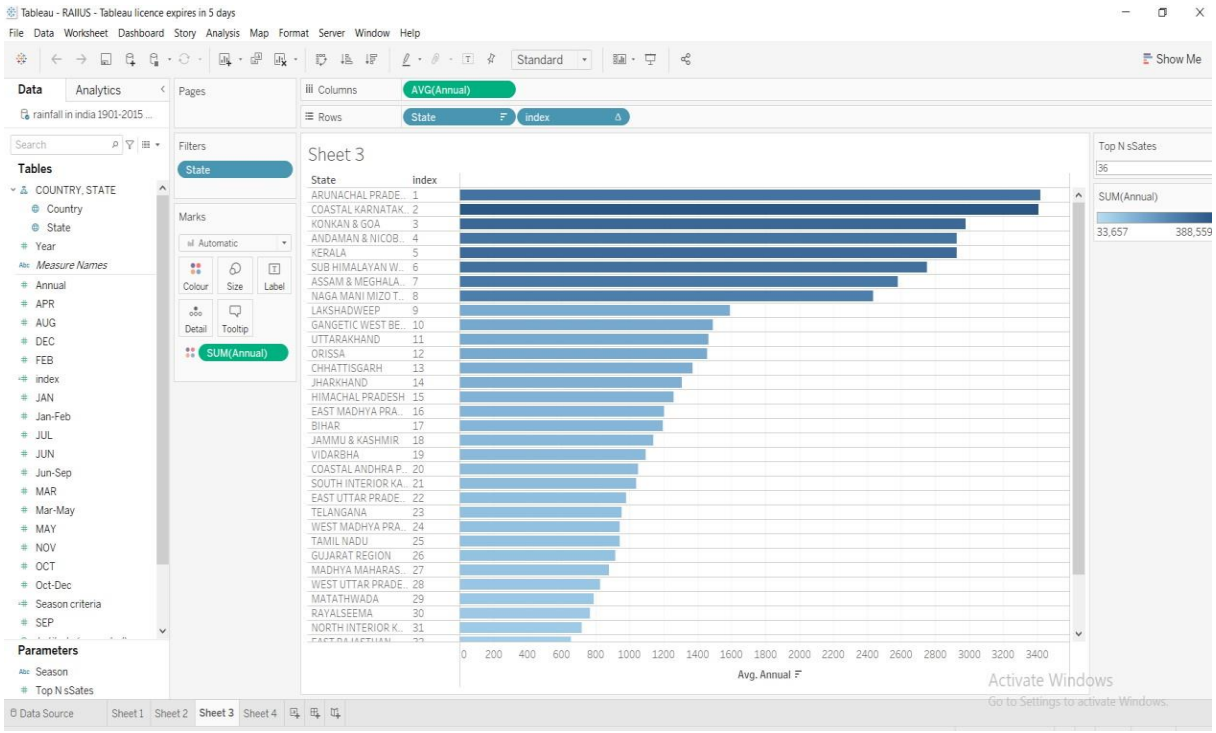
Tableau - RAIIUS
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Dashboard
Layout
Default
Phone
Device Preview
Size
Desktop Browser (1000 x 800)
Sheets
Sheet 1
Sheet 2
Sheet 3
Sheet 4
Sheet 5
Objects
Horizontal
Blank
Vertical
Navigation
Text
Download
Image
Extension
Web Page
Tiled
Floating
Show dashboard title
Sheet 1
Sheet 3
State
index
COASTAL KARNATAK.. 1
ASSAM & MEGHALA.. 2
GANGETIC WEST BE.. 3
CHHATTISGARH 4
EAST MADHYA PRA.. 5
0 500 1000 1500 2000 2500 3000 3500
Avg. Annual
Sheet 4
State
ARUNACHAL PRADE..
ASSAM & MEGHALA..
NAGA MANI MIZO T..
ANDAMAN & NICOB..
SUB HIMALAYAN WE..
0K 10K 20K 30K 40K 50K 60K 70K
Season criteria
Sheet 2
Year
2000
1900
0 50 100
0 100 200 300
0 500 1000 1500
0 100 200 300
Avg. Jan-Feb
Avg. Mar-May
Avg. Jun-Sep
Avg. Oct-Dec
Annual
75,350
388,559
Top N sSates
10

CHAPTER 8

OUTPUTS







Activity 7: Splitting the Dataset into Dependent and Independent variable

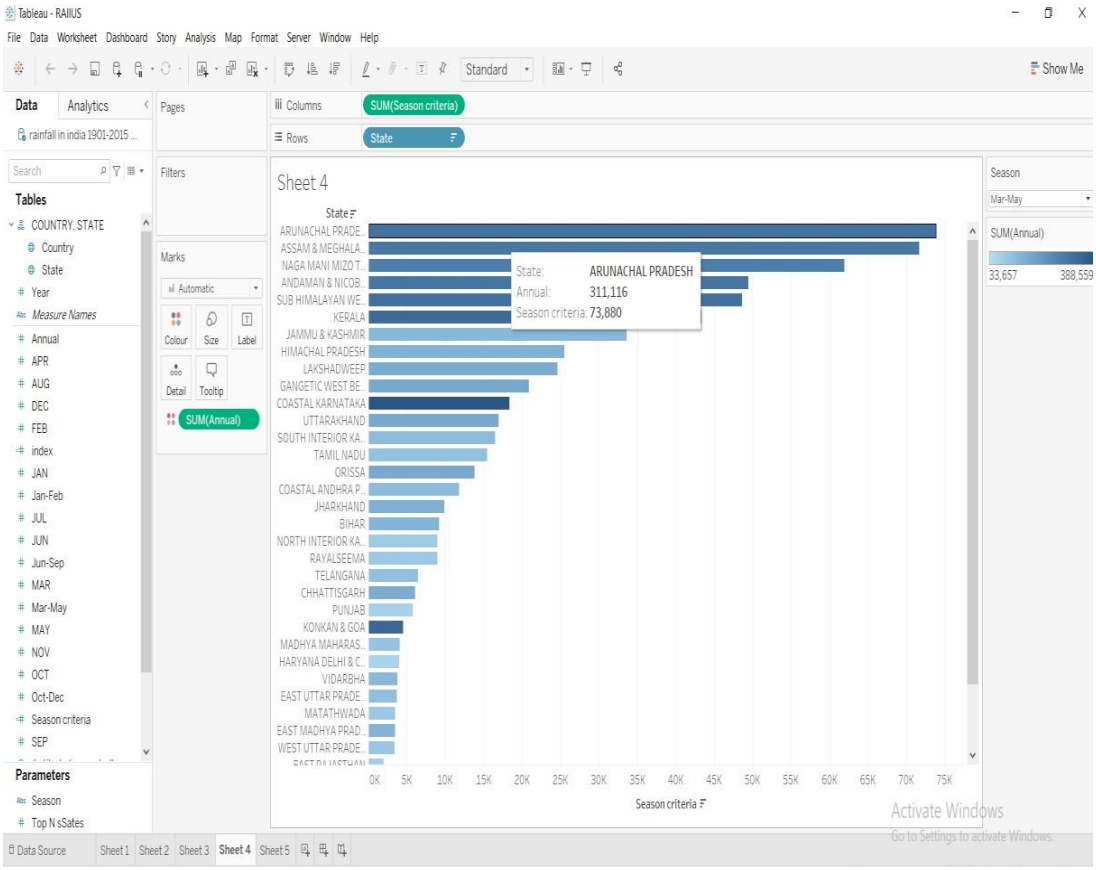


Tableau - RAIIUS
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Data Analytics
rainfall in india 1901-2015 ...
Tables
COUNTRY, STATE
Country
State
Year
Measure Names
Annual
APR
AUG
DEC
FEB
index
JAN
Jan-Feb
JUL
JUN
Jun-Sep
MAR
Mar-May
MAY
NOV
OCT
Oct-Dec
Season criteria
SEP
Parameters
Season
Top N sSates
Pages
Filters
Marks
Automatic
Colour
Size
Label
Detail
Tooltip
SUM(Annual)
Columns
SUM(Season criteria)
Rows
State
Sheet 4
State
ARUNACHAL PRADE..
ASSAM & MEGHALA..
NAGA MANI MIZO T..
ANDAMAN & NICOB..
SUB HIMALAYAN WE..
KERALA
JAMMU & KASHMIR
HIMACHAL PRADESH
LAKSHADWEEP
GANGETIC WEST BE..
COASTAL KARNATAKA
UTTARAKHAND
SOUTH INTERIOR KA..
TAMIL NADU
ORISSA
COASTAL ANDHRA P..
JHARKHAND
BIHAR
NORTH INTERIOR KA..
RAYALSEEMA
TELANGANA
CHHATTISGARH
PUNJAB
KONKAN & GOA
MADHYA MAHARAS..
HARYANA DELHI & C..
VIDARBHA
EAST UTTAR PRADE..
MATATHWADA
EAST MADHYA PRAD..
WEST UTTAR PRADE..
State: ARUNACHAL PRADESH
Annual: 311,116
Season criteria: 73,880
0K 5K 10K 15K 20K 25K 30K 35K 40K 45K 50K 55K 60K 65K 70K 75K
Season criteria
Season
Mar-May
SUM(Annual)
33,657 388,559
Activate Windows
Go to Settings to activate Windows.
Data Source
Sheet 1 Sheet 2 Sheet 3 Sheet 4 Sheet 5

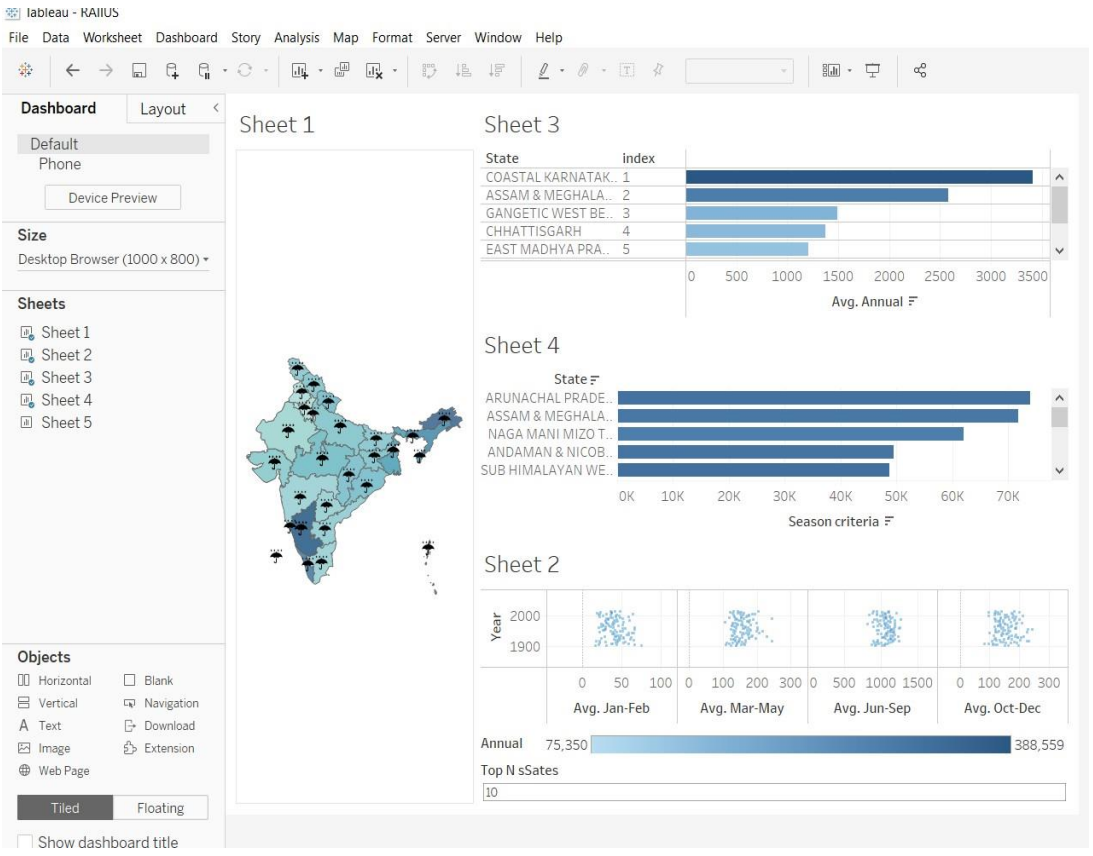


Tableau - RAIIUS
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Dashboard
Layout
Default
Phone
Device Preview
Size
Desktop Browser (1000 x 800)
Sheets
Sheet 1
Sheet 2
Sheet 3
Sheet 4
Sheet 5
Objects
Horizontal
Blank
Vertical
Navigation
Text
Download
Image
Extension
Web Page
Tiled
Floating
Show dashboard title
Sheet 1
Sheet 3
State
index
COASTAL KARNATAK.. 1
ASSAM & MEGHALA.. 2
GANGETIC WEST BE.. 3
CHHATTISGARH 4
EAST MADHYA PRA.. 5
0 500 1000 1500 2000 2500 3000 3500
Avg. Annual
Sheet 4
State
ARUNACHAL PRADE..
ASSAM & MEGHALA..
NAGA MANI MIZO T..
ANDAMAN & NICOB..
SUB HIMALAYAN WE..
0K 10K 20K 30K 40K 50K 60K 70K
Season criteria
Sheet 2
Year
2000
1900
0 50 100
0 100 200 300
0 500 1000 1500
0 100 200 300
Avg. Jan-Feb
Avg. Mar-May
Avg. Jun-Sep
Avg. Oct-Dec
Annual 75,350 388,559
Top N sSates
10

CHAPTER 9

CONCLUSION

Rainfall forecasting is very important because heavy and irregular rainfall can have many impacts like destruction of crops and farms, damage of property so a better forecasting model is essential for an early warning that can minimize risks to life and property and also managing the agricultural farms in better way. This study aims to determine trends in annual and seasonal rainfall and rainy days over different river basins across India. The project reviews, rainfalls in India visualization in Tableau environment. The aim is to show how to extract meaningful data from the raw data and visualize it. The rainfall pattern in the States/Union Territories of India was successfully visualised.

CHAPTER 10

FUTURE ENHANCEMENT

1. Instant detection of rainfall by predetermined datasets.

2. Easy prediction.

3. Time to time indications and early warnings.

4. Hazardous events can be predicted with more efficiency.

5. With evolving trend analysis, this becomes more reliable.

6. It takes rainfall average data from previous years and predicts further probability of rainfall.

7. Highly accurate and adaptable for long run rainfall prediction which can replace older prediction systems.

8. Intensity of rainfall in several areas is analysed.

9. It provides better forecasting model which gives early warning.

10. Beforehand prediction of floods can help safeguard the lands.

11. Efficiency can be improved even more in future forecasts.

12. The economic growth depends on the amount of monsoon rain which could be evaluated accurately.

13. Being secondary agro-based market can be steady with a good monsoon it needs trend analysis.

CHAPTER 11

BIBLIOGRAPHY

For Tableau Installation:

https://www.tableau.com/products/desktop/download
https://help.tableau.com/current/desktopdeploy
https://www.tableau.com/learn/tutorials/on-demand/installation-setup

User Tier:

- VizPortal
- VizQL Server
- Data Server
- Metadata Services

File Types:

- Tableau Work Book
- Tableau Map Source
- Tableau Book Mark
- Tableau Data Source

CHAPTER 12

REFERENCES

[1] Guhathakurta, P. and Rajeevan, M., 2007 "Trends in the rainfall pattern over India", Int. J. Climatol., 28, 1453-1469.

[2] Roy Bhowmik S K, Joardar D and Das Ananda K 2005.Radius of rainfall influence over Indian monsoon region;

[3] Helsel, D. R. and Hirsch, R. M., 1992, "Statistical Methods in Water Resources", Elsevier, Amsterdams, Elsevier Publishers, 529.

[4] Xu, Z. X., Li, J. Y. and Liu, C. M., 2007, "Long-term trend analysis for major climate variables in the Yellow River basin", Hydrological Processes, 21, 1935-1948.

[5] Janowiak J E 1992 Tropical rainfall: A comparison of derived rainfall estimates with model forecast, climatologies and observations; Mon. Wea. Rev.120 448–462.

[6] Janowiak J E and Xie P 1999 CMAPS-OPI: A rain gauge merge product for real timeprecipitation monitoring application; J. Climate 123335–3342.

[7] Roy Bhowmik S K and Sen Roy Soma 2006 Princi-pal Component Analysis to study spatial variability of errors in the INSAT
derived quantitative precipitation estimates over Indian monsoon region; Atmosfera 19255–265.

[8] Roy Bhowmik S K, Joardar D and Das Ananda K 2005 Radius of rainfall influence over Indian monsoon region;Geofizika 22 131–141.