MINOR-1 PROJECT

SYNOPSIS

Statistical Data Visualizer Library

Department of Systemics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007, Uttarakhand

Dr. Ajay Prasad   Dr. Neelu Jyoti Ahuja
Project Guide   **Cluster Head**

Submitted By:

S.no	Specialization	SAP ID	Name
1.	CSF	500084172	Manika Rajpal
2.	CSF	500082387	Sanskar
3.	AI-ML	500082509	Aryan Kumar
4.	AI-ML	500084403	Palak Singh

Index

Contents   Page Number

Chapter 1: Abstract

This project is intended to create an exploration environment, which focuses on the statistical operations and creates a smooth learning for students as the graphical and pictorial representation will help in analysing and understanding large quantities of numerical data and the relationship between them. The goal of this library is that it will make the information easy to understand and through this user can gain insight into their huge amounts of data, also it will help the users to pay attention to areas that represent changes or processes and the deteriorating areas which need progress.

 Keyword: Statistical, Visualizer, Real life Implementation

Chapter 2: Introduction

In the globalized world, there has been the need for displaying massive amounts of data, in a way that it is easily accessible and understandable not only to the data analyser, but also to every user who goes through it and because the use of analytics is no longer limited to big companies with deep pockets also since data is the fuel for many industries as a result of which the amount of data available on the Web has increased drastically so it is difficult for many users to visualize, explore, and use this enormous amount of data.

Today, computers can be used to process large amounts of data. Data visualization is concerned with the design, development, and application of computer-generated graphical representation of the data. It provides effective data representation of data originating from different sources. This enables decision makers to see analytics in visual form and makes it easy for them to make sense of the data. It helps them discover patterns, comprehend information, and form an opinion. A variety of data visualization software packages are important to those who plan to work in the current business environment and also the students who are enrolled in the programming field and are often exposed to some form of statistical and data analysis, due to the growing number of employment opportunities which require these specialized skills. These software packages are also highly researched and maintain a high amount of relevance in academia, as well as corporate worldwide.

Chapter 3: Problem Statement

To create a data representation library which can be used to show and analyse huge amounts of data in the form of graphs and charts according to the user's choice.

Chapter 4: Literature

Each library is designed to provide a solution to a specific feature, also libraries allow us to dictate the flow of an application and add a costume code and quickly give all statistical results. The existing libraries are quite complex for beginners to understand therefore we will be developing a statistical library less complex than others that will help beginners to easily implement and use our library so that they can have a better understanding of any type of data in visual representation.

Existing System:

There are various visualization libraries provided in python like GGplot, matplotlib, plotly, etc. But in java there is no specified library for statistical data visualization.

Need For System:

Libraries available are hard for beginner programmer to operate. In java we don't have any such package or tool which provides us the graphical output of the data. So, our aim is to build and provide a less complex library for beginner programmers to have an overview of the visualization world. We would be building a library which would help user to get an insight about statistical visualization and obtain graphical outputs.

Chapter 5: Objectives

1. To develop a single Java library which minimizes the span of time and visualizes different statistical operations at once.
2. To give a glimpse of a visualization tool for beginner programmers.
3. To get our youth more excited about the Java programming language.

Chapter 6: Methodology

We will be using the Agile Model for our project's development. A statistical visualizer library provides the user the leverage of having the statistical operation that provides the visualized data in a graphical and pictorial format under a single library. So, all this needs to be done in small parts and checked phase wise and a constant interaction with the user is needed. Agile Model deems that the current methods should be tailored with time and according to project's need also it divides the tasks to small frames to deliver particular features for a release. This model also emphasizes user interaction as the students, developers, and testers collaborate throughout the project. However, because this approach is strongly reliant on client engagement, the project may continue in the incorrect direction if the developer is unsure of where he or she wants to go.

ADVANTAGES OF AGILE MODEL

• It is easy to manage.

• Provides flexibility to the developers.

• Delivers a partial working solution before complete making.

• Suitable for changing requirements.

 VISUALIZATION TECHNIQUES

Visualization is the use of computer-supported, visual representation of data. Unlike static data

visualization, interactive data visualization allows users to specify the format used in displaying data.

Visualization techniques we are using are:

- Line graph: It shows the relationship between various different types of items in the data. It can be used to compare changes over a period of time.
- Bar chart: It is used to compare quantities of different categories.
- Scatter plot: It is a two-dimensional plot showing variation of two items.
- Pie chart: It is used to compare the parts of a whole.
- Bubble charts: are used to determine if at least three numerical variables are related or share some kind of pattern.
- Pictograph: is a chart or graph which uses pictures/symbols to represent data.
- Box Plot: Box plots visually show the distribution of numerical data and skewness through displaying the data quartiles (or percentiles) and averages.
- Gantt chart: It maintains a bird's eye view of any projects. It depicts the relationship between the start and end dates of tasks, milestones, and dependent tasks.
- Histogram: A histogram is a graphical display of data using bars of different heights.
- Map Chart: To compare values and show categories across geographical regions.
- Heat Map: A heat map is a data visualization technique that shows the magnitude of a phenomenon as colour in two dimensions.
- Area Chart: It is a graph that combines a line chart and a bar chart to show changes in quantities over time.

Thus, the format of graphs and charts can take the form of bar chart, pie chart, line graph, etc. It is important to understand that every data requires a different chart or graph.

Data visualization uses computer graphics to show patterns, trends, and relationship among elements

of the data. It can generate pie charts, bar charts, scatter plots, and other types of data graphs with simple

pull-down menus and mouse clicks. Colours are carefully selected for certain types of visualization. When

colour is used to represent data; we must choose effective colours to differentiate between data elements.

In data visualization, data is abstracted and summarized. Spatial variables such as position, size, and

shape represents key elements in the data. A visualization system should perform a data reduction,

transform and project the original dataset on a screen.

It should visualize results in the form of charts and graphs and present results in user friendly way.

Chapter 7: System Requirement

1. Software Requirements

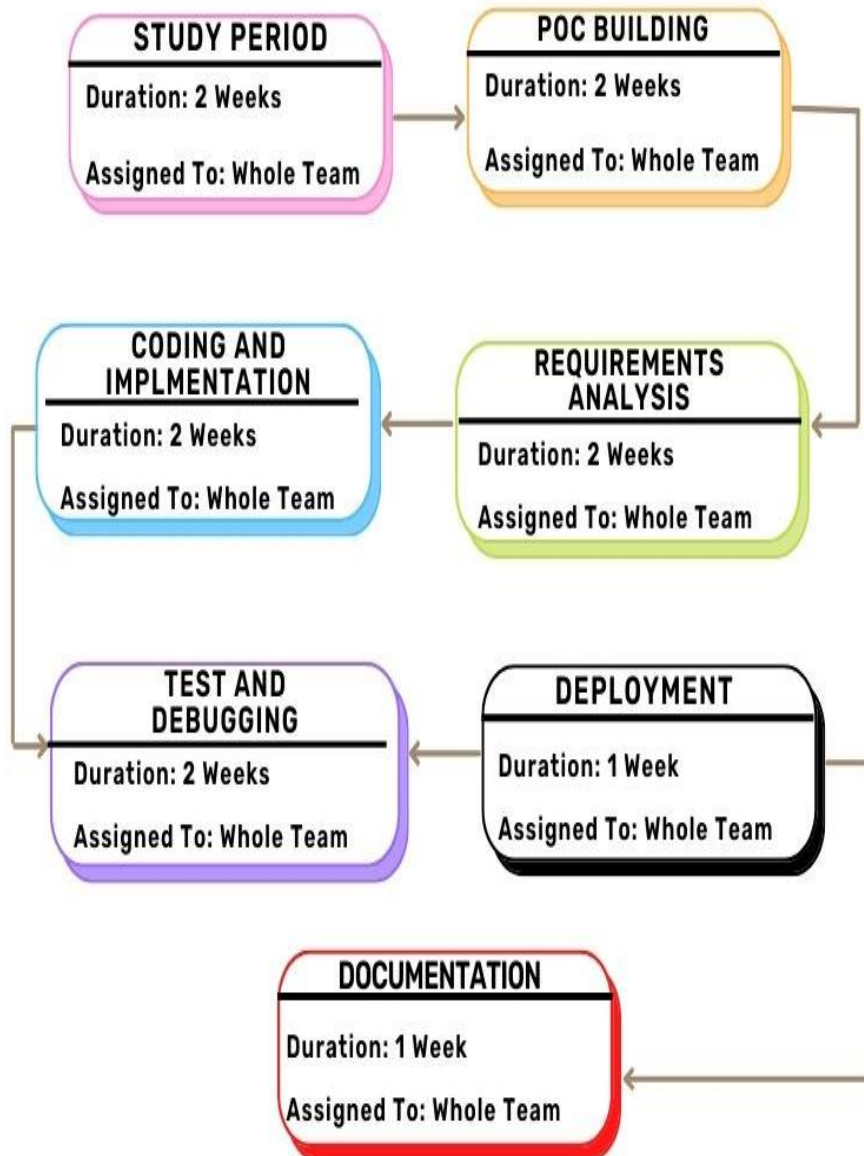Operating System: Windows 10/8/7 (32-bit or 64-bit)

Software: Eclipse/ VS Code

2. Hardware Requirements

Processor: Dual Core 2.7 GHz or better

RAM: 512 MB or higher

Disk Space: 512 MB

Chapter 8: PERT Chart

STUDY PERIOD

Duration: 2 Weeks

Assigned To: Whole Team

POC BUILDING

Duration: 2 Weeks

Assigned To: Whole Team

CODING AND IMPLMENTATION

Duration: 2 Weeks

Assigned To: Whole Team

REQUIREMENTS ANALYSIS

Duration: 2 Weeks

Assigned To: Whole Team

TEST AND DEBUGGING

Duration: 2 Weeks

Assigned To: Whole Team

DEPLOYMENT

Duration: 1 Week

Assigned To: Whole Team

DOCUMENTATION

Duration: 1 Week

Assigned To: Whole Team

Chapter 9: References

- https://www.researchgate.net/publication/311597028_DATA_VISUALIZATION
- https://www.researchgate.net/publication/315870481_A_survey_of_data_visualization_tools_for_analyzing_large_volume_of_data_in_big_data_platform
- https://www.statisticshowto.com/operational-statistics-definition/
- https://www.investopedia.com/terms/p/pert-chart.asp
- https://acadpubl.eu/hub/2018-119-12/articles/5/1164.pdf
- https://reference.wolfram.com/language/guide/StatisticalVisualization.html
- https://www.tableau.com/learn/articles/data-visualization
- https://www.cs.uic.edu/~kzhao/Papers/00_course_Data_visualization.pdf
- https://www.idrc.ca/sites/default/files/sp/Documents%20EN/data-visualization-in-review-summary-report.pdf