Generic Web Console for Visualization and Time Series Forecasting

Koushikraj Rajendran

10506917

Dissertation submitted in part fulfilment of the requirements for the degree of

M.Sc. in Information Systems with Computing

At

Dublin Business School

Supervisor: Obinna Izima

January 2020

Declaration page

I declare that this dissertation that I have submitted to Dublin Business School for the award of M.Sc. in Information Systems with Computing is the result of my own investigations, except where otherwise stated, where it is clearly acknowledged by references. Furthermore, this work has not been submitted for any other degree.

Signed: KOUSHIKRAJ RAJENDRAN

Student Number: 10506917

Date: January 2020

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor and advisor Prof. Obinna Izima for his leadership, encouragement and continuous support for this research. I am also thankful to all the lecturers and academic staffs in Dublin Business School, who helped me to progress with my dissertations. I would also like to extend my gratitude towards the entire TCS-mKrishi team for every insight and experience I gained, which immensely helped me in completing my master's dissertation. I would like to thank all the participant who agreed to take part in the final user evaluation of the prototype.

Finally, I would like to thank my parents and friends for everything they have done for me.

Abstract

Data has become the new fuel of this era. Big Data are being scaled and processed to give useful information's and various visualization techniques are used to representation of data or information in a graph, chart, or other visual format. Visualization plays a vital role in bringing out hidden details like trends and pattern and representation of the information easily. All these data and information can be used to forecast into the future, which can influence major decisions and strategies in any field. Thus, automating two major fields in data industry will help save time and effort, which in-turns enable better understanding and quicker decision making. This dissertation aims to design and implement a framework using Google Charts.js for generating client-side visualization and Long short-term memory (LSTM) an RNN technique for forecasting time series data. Spring boot framework is used for developing this standalone Spring web application, with Java as its background language. With the help of scheduler and runtime library in Java, python script for LSTM was executed at specific interval. The main idea is to automate visualization and prediction process of data as much as possible in order to reduce the time and work done.

Contents

List of Figures

List of Tables

1. Introduction

Visualization changes the information into a visual form and takes advantages of people's natural ability that can quickly identify visual patterns to observe, browse, discriminate and understand information (Y. Qi, G. Shi, X. Yu and Y. Li, 2015). Psychologists have had the chance to communicate effectively of the data rich stories through the human perception studies. Interactive visualizations usually result in discovery and give better performance than the static data tools (Kirk, 2016). As discussed visualization plays a vital role in literally all major fields in which data are involved. But as the data increase the work

required to visualize data becomes complex, as the data needs to be cleansed, prepared and then fit into a model for analysis or visualization repeatedly.

In today's world there is always a requirement for fast, reliable and un-conditioned access to data, this applies to the insights and visualization of these data also. Thus not only making these visualization available on the internet but also providing them with the tool to generate new visualization on the go, is the only way to meet the current requirements of today's fast changing world, where new data's come in every time and the generated insights and trends changes drastically. A generic visualization tool was created which is for flexible data visualization over the web, the output can be used like a plugin in other websites. It had prediction model which will predict the upcoming data from the data present and give a future insight in the visualization. But unfortunately, it only supports three types of chart namely line, bar and radar, which limits the visualization and might not be able to produce a suitable visualization for every data (Roshdy, A. et al,2018).The web-based techniques used by the researchers help in production of interactive data visualizations that can fully support and supplement both the current and future publications. This is quite beneficial not only to the researchers but also to the students, practitioners, the academic community and other interested people from the public (Zhu,2017).

Time Series analysis is the study of an ordered sequence of values collected at equal intervals. Main purpose of this analysis is to determine the key factors and forecast future values which help in taking better decision and reduce the risk of uncertainty. Time Series Analysis is a key player in many sectors, some are stock market and economic forecasting, inventory forecasting, Rainfall Analysis, Yield sales forecasting and many more.

Thus the proposed research is about producing application by combining both these major players of data industry ,which save work and time by automating the data cleansing, formatting and fitting the data in order to visualize and forecast along with the luxury of control and access to it via internet.

1.1 Objectives of Research

Objective of the research is to build a lightweight web module which can provide real time dynamic visualization using client-side rendering and automate time series forecasting.

1.2 Research question

1. Design a web-based system to automate visualizations and forecasting of time series data?
2. Does automating visualization and forecasting improve the experience for users with limited programming background?

1.3 Scope of Research

- To enable users to create visualization of data easily without the hassle of coding.
- To forecast time series data with a suitable algorithm.

1.4 Explanation of Research Question.

To select a suitable web module framework to automate both analysis and dynamic visualization of data with interactivity and also predict data using deep learning techniques. The developed system should allow a user with none or limited programming background to

generate visualizations and forecasting of data present in the database. The research follows the below steps.

- Finding a suitable framework for web project, it should be able to handle the load created by both analysis and generation of visualization.
- Finding a Generic interface for the database, so that the module will be free of database system restrictions.
- Define a predicting machine learning model.
- Select a suitable client -side rendering Visualization Techniques.
- Build the artefact
- Run various Test on the artefact
- Adapt to changes based on the test results

2. Literature Review

2.1 Data visualization

Data Visualization is the process of mapping information to visuals, it makes the data more understandable and brings out hidden insights.IT

2.1.1 Importance of Data Visualization

In this information era, where there is enormous amount of data and each and every single data is important, Visualization has a major role in understanding the data. Visualization

tends to have impart more understanding than mere words and it is the fastest way to convey the meaning of data to others, even a child can interpret a chart (Scott Murray ,2017)

Visualization helps in finding data which are normally hidden in facts. At a certain point, visualization removes the abstract fact which mostly everyone might be clinging on to as important information and help us evolve our understanding of data from “this is the data” to “this is what the data is”.(Hinderman, Bill,2016)

The main role of the Big Data analytics is to minimize the complexity and size of data in the Big Data applications. On the other hand, visualizations help in realizing the data values and acquiring a complete view of all the data (Zhu, 2017).

Psychologist have implemented the use of data visualization and graphical illustrations in presenting numbers or as a means helping other people discover and understand well results or the information presented as well. Interactive visualizations usually result in discovery and give better performance than the static data tools. Interactive visualizations assist well in the gaining of insight from some big data. The interactive visualization may involve scientific processes like the linking and brushing the approaches or the web-based tools (Kirk, 2016).

2.1.2 Web Based Data Visualization

A web based application was developed to impose real-time ocean data in to heat map images using matplot lib , map images are generated on the server on demand and then displayed .The time taken for the image generation depends on the data points to be shown in the map.

(Mwalongo, F. et al.,2016), performed analysis on various web based applications in his paper. And grouped these applications into three major categories

1. Single abstract visualization pipeline. (restricts the user flexibility)
2. Multiple Visualization pipeline (fully exposed to user, has more flexibility but inefficient data management)
3. Combination of Multiple Visualization Pipeline into single web service (more flexibility and more efficient management of data)

Spiros Koulouzis(2012), conducted a research for data transportation between visualization web services for medical image analysis using a visualization workflow engine to generate the visualization pipelines. The experiment was done between centralized and distributed communication model. The centralized model, the data is fetched from central storage every time, whereas for distributed communication data is exchanged between the web services. Then the research concluded that the distributed service had better performance than the centralized service model. Also, instead of transmission of data, URI of data was used in transmission, which resulted in increasing the efficiency of the communication.

Developers use API's to create various charts on their web page. But still they must spend some time to read documentation and understand the API. There are a few commercial tools allowing charts creation (even web based), but none of them provides ability to create just only chart templates which are independent from charts data. (M. Ostruszka, B. Sakowicz and A. Napieralski, 2007)

Client side rendering has some advantages over server side rendering of visualization, With advancement in client side rendering technologies, interactivity for visualization over

web that can be increased ,reduction in latency due round trip time taken in server side rendering(Mwalongo, F. et al.,2016)

2.1.3 Visualization Tools:

Feng Xia (2018), did a survey on available visualization tools, and discussed detailly about that. Some of the notable open source tools related to this proposal are

- D3.js which takes JSON, CSV & XML as input and creates output using combination of CANVAS, SVG and XML. It has a wide variety of visualizations which can be used and provides user the option to create their own visualization
- Chart.js uses JS API to get input of a specific format and generates visualization using canvas alone. It provides 8 chart type and 23 varieties of them.

Apart from this, analysis on various commercial and non-programable tools, data source in the journal can be found.

Zhu(2012),a Data visualization scholar, used a test group of students to compare google charts and maps with various libraries for visualization. It was found that google charts and maps which has client-side rendering enabling interactivity and SVG format output increasing the scalability, was easier to implement than other existing libraries and there was a lot of materials available to support the chart. But the drawback was only 2D charts are available and drawing graphs are still in development stage.

Web-based Time-tunnel is a multi-dimensional 3D visualization tool for assisting exploratory data analysis. The system overlaps charts of different data to find similarities and differences, also a generic algorithm was developed as a functionality to group charts with strong relationship among them. Data-wings are the basic units which are optimized based

on the coefficient of correlation values. The information and order of these units are shared and discussed using gene information (R. Akase and Y. Okada,2015).

Web-based visualization helps in acquiring dynamic data on time and keeping the visualizations updated. The extension of some conventional visualization approaches to handling big data is far from enough in functions. There is a presentation of the advancements of the Big Data visualization and the SWOT analysis of the current software tools used for the Big Data visualization. The SWOT analysis will help in developing new tools and techniques for the big data visualization. This will help develop new methods and tools for big data visualization. However, there is a method that can work well for the big data visualization. It is a new method known as Virtual reality with the ability to handle high dimensions and abstracts (Kirk, 2016).

2.2 Time Series Analysis

Data analysis refers to the sorting, organizing, processing, analysing and studying on the basis of collecting and occupying the data, the whole process of discovering new knowledge (Huang Yihua, 2014).

Time series data are a set of data observed sequentially in time, and the study these time variant data are known as time series analysis. There are four main purpose of time series analysis are Forecasting of future data based on previous data, calculation of transfer function of a system to determine the dependency of output on input, to assess the intervention of unusual events, designing of control schemes. In forecasting the statistical

properties of the time series data plays a vital role, an time series can be either stationary or non- stationary depending on the constant mean and variance level of the data (Tunnicliffe Wilson, Granville, 2016).

Stationary time series have statistical properties which are constant over time, meaning all the properties like variance, mean, joint distribution, third or higher order is same at any point of the data. In general, real world data cannot be 100% stationary, thus a 2nd order or weak stationary is considered stationary. Weak stationary implies that the variance of the data is independent of time and auto correlation depends on lag between the data points rather than the time. The auto correlation plays a major role in creating a prediction model for the data and it is usually normalized to values between -1 to 1 (Nason, GP 2006).

Ratnadip Adhikari, R. K. Agrawal (2013) gave a brief explanation on various stochastic models for stationary time series. The most popular linear models where AR model which predicts the future value of a variable based on a linear combination of past observations and a random error along with a constant and MA model which uses the mean of the past values along with a past errors. Apart from these models AIMA, ARMA , SARIMA for seasonal data which are basically created by combining AR and MA models with slight modifications and Box-Jenkins models were also discussed briefly.

Juang, W. C (2017) successfully developed and published on ARIMA time series model for predicting overcrowding of emergency department with the criteria of minimum Akaike information and Schwartz Bayesian as base to select the p,q,d value range and finally ended up with ARIMA(0,0,1) model with a mean absolute percentage error of 8.91 %.They also used extended autocorrelation function and smallest canonical correlation in addition to plot pattern to evaluate the model composition.

Order difference of a non-stationary time series makes the sequence smooth and appropriate for models like ARMA and ARIMA, but still seasonality might be present in some series. Thus for time series with seasonality SARIMA model was developed from ARIMA. SARIMA includes both the trend and seasonality of a sequence into the equation formulating a model. SARIMA had higher fitting degree for agricultural irrigation where there are abnormal and sudden changes in the data, when compared with the regression and Box-Jenkins method (Shengwei Wang,Juan Feng,Gang Liua, August 2013)

Recommends SARIMA were sensitive to periodic terms and has better results than ARIMA when used in evaluation of long-term runoff forecasting for US state. Another notable information was that the performance of SARIMA declined sharply when the periodic term was not correctly estimated. Some of the reliable models for sequence with drastic changes are SARIMA 55, ASRIMA 22 and SARIMA 11 (Valipour, M. ,2015).

Apart from linear models, machine learning and deep learning techniques can also be used for predicting time series data. Tianhong Pan and Shaoyuan Li(2016),discussed about using lazy learning machine learning technique , a statistical memory based learning strategy to perform time series analysis. They used a similar criterion and locally weighted learning to achieve one-step-ahead best predictive value based on leave-one-out cross validation.

LSTM-DBN is a multi-step time series prediction model which was developed by combining LSTM neural network model with a Bayesian GUI and increasing the throughput accuracy by implementing optimal estimation principle and recursion algorithm on the data. The reclusion calculation structure is based on Bayes and probability theories, which on experimenting showed better results than few models (Q. Xiao and Y. Si,2017).

Auto-regressive filters can also be used for time series analysis, one such filter is Kalman filter which uses a recursive algorithm to estimate dynamic system state via a process of incomplete and noise inclusive measurements. Kalman filters are created for the process of estimation of multidimensional stochastic data as it converts state space theories into mathematical models of physical system. The process uses previous estimate and most recent data to estimate the current value by two steps: estimation and correction of state variables. Even though it was mainly used as linear estimator, there are variants of non-linear estimators also like particle filtering, integrated and unidirectional Kalman filter, as most of the data are non-linear (X. Song, J. Huang and D. Song,2019).

The long-short-term memory has many variants, but none of them differ in the architecture, thus combining a standard LSTM with another methodology seems to be a way of improving the results. One such proposed model was LSTM-Kalman Model, which uses the LSTM output as input for the Kalman filter producing a better result combined than the individual.AS the LSTM produce a dynamic summary of the variations of the data, the Kalman model dynamically adjust he the observed value based on the predicted value of LSTM, thus dependency of dynamic model is eliminated. The researcher used this model on air pollutant data and concluded that the result was more efficient than the individual models (X. Song, J. Huang and D. Song,2019).

Data filtering and analysis are critical tasks in the process of identifying and visualizing the knowledge contained in large datasets, which is needed for informed decision-making (S. Holtz, G. Valle, J. Howard and P. Morreale, 2011).

Data are often in the wrong format and needs to be transformed and integrated to support further analysis. Which requires programming, a complex data integration

infrastructure, and it takes days to produce even simple reports. Clearly, this introduces a significant gap between raw data and insight (E. Kandogan, M. Roth, C. Kieliszewski, F. Özcan, B. Schloss and M. Schmidt,2013).

2.3 Framework

One of the most used web application frameworks is structs framework for J2EE platform, it provides separation of logic and business layer. Interaction between these layers is done by data transmission whenever required. MVC structure in J2EE can be explained as

- View: It is responsible for the output page which is constructed using html
- Model: The logic of the application but structs doesn't support this layer.
- Controller: This layer focus on mapping the inputs to the business logic and the output to the View

Even though Struts is popularly used framework, it has become old and newer technologies with more efficient and advancement (Wojciechowski, J. et al. ,2004).

Developed a lightweight framework by combining Spring and Mybatis. Spring is based on method which made the operation efficiency of Spring MVC much higher and annotation made configuration file more concise than Struts which is class based. Mybatis is a semi-automated ORM framework which supports all types of query and database operations with operating efficiency higher than Hibernate. Thus, by combining Spring MVC annotations and injection of only requested data of methods and Mybatis to improve efficiency of database. More information about the performance and implementation of this lightweight model can be found in the paper (Zhang, D., Wei, Z. and Yang, Y. ,2016).

2.4 Literature Conclusion

Based on the above findings, it is clear that integration of time series analytics and visualization will give good results in the various applications, Due to the large complexity and increased size of the data sets, it is clear that there are upcoming opportunities and challenges experienced by the analysts.

Data Analysis and Visualization both depends on given data, thus various data cleansing algorithms should be imposed to make the data free from irregularities, duplicates and other unwanted data if any. Also, instead of using centralized storage it is advised to use distributed system communicating URI of data instead of data will increase the response time and increase the efficiency of the system.

Visualization plays an important role in communicating the information mined from the data, Interactive and dynamic visualization always captures the user attention ,thus making it interactive and faster loading is very important and also rendering it in the browser reduce latency which can be achieved by client-side rendering. Therefore, there should be a thorough research on the different approaches in specific different areas that will help for easier data exploration.

While developing an artefact to combine both data analysis and visualization, a suitable framework should be chosen in order to handle the load and provide good user experience.

3. Methodology

3.1 Methodology Introduction

Research methodology is a scientific way of solving a research problem systematically. The scope of the methodology is to ensure that the researcher understand the logic behind the methods involved in the research and be able to explain the reason behind using a specific process instead of other available methods(Kothari, 2004).It helps in conceptualization from idea to results. The methodology has to be selected based on the scope and nature of the research, there is no common or standardized methodology for all research(Bell, 2005).The main objective of the research is to develop a web framework for making visualization and forecasting of data easier for the user irrespective of the data.

Firstly, this section describes the research design were research philosophy followed will be discussed, research approach and research strategy. framework of the application which gives details of the theoretical concepts that were used for the development of

application. Secondly, about the identification and selection of data for testing. Then finally, ethics and limitations followed while doing this research.

3.2 Research Design

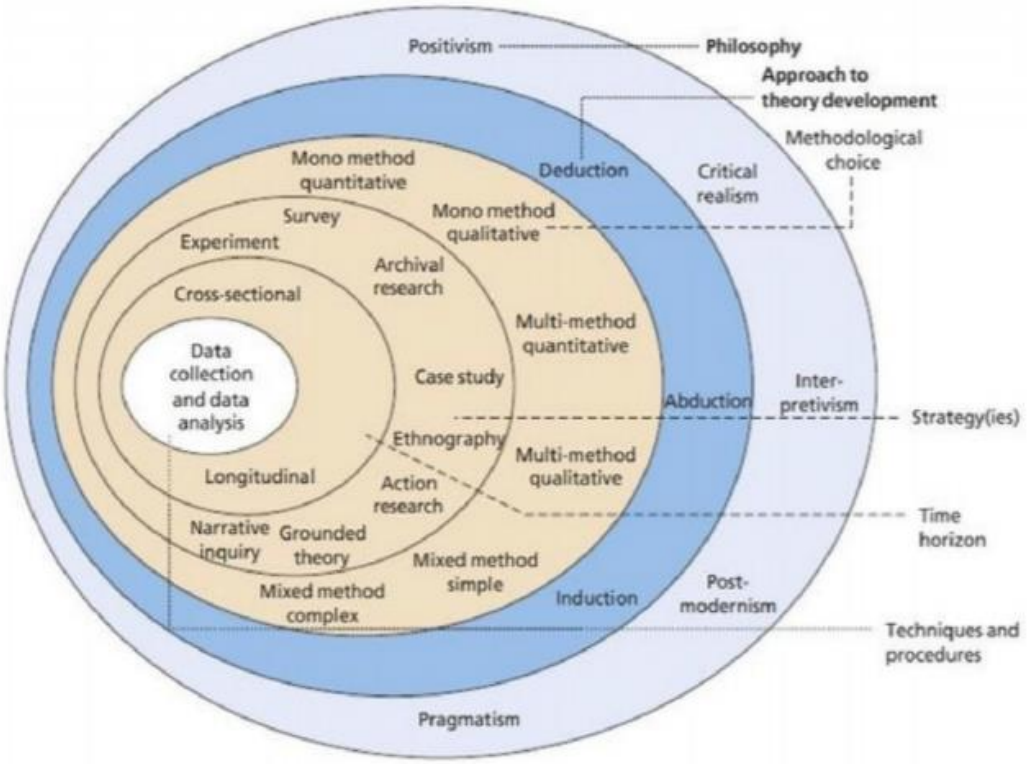


Figure 1: The Research Onion (Source: Collins, 2018)

3.2.1 Research Philosophy

Zheng and Zhou (2011) states that research philosophy is the concept of analysing and interpreting data, which is represented by the first layer of the research onion .Thus, the decision of research philosophy will be influenced by practical considerations, mainly the researcher's perceptiveness on the knowledge and the process used for development (Mark Saunders, Philip Lewis, Adrian Thornhill,2015).

This proposal comprises of a web module which produce visualization and insights to the user. It aims to remove the restriction of accessibility by hosting it in the web, restriction on visualization due to lack of coding knowledge and time, predict the future values and then increase the efficiency and speed by integrating state of the art technologies. Altogether the purpose is to ease the user burden to generate visualization and enhance user experience. Thus, user evaluation is employed here to test and gather user feedback

3.2.2 Research Approach

There are two types of research approach namely, deductive and inductive. Deductive is a top down approach were a hypothesis is formed and then narrow it down to observations and confirmation by testing the observation with the hypothesis. Where else inductive is a bottom-up approach with qualitative research, it begins with the research question and then explore various research papers to solve the research question. In this research inductive approach is used by collecting data about various artefact, interface and frameworks discussed in research papers and then try to develop our artefact.

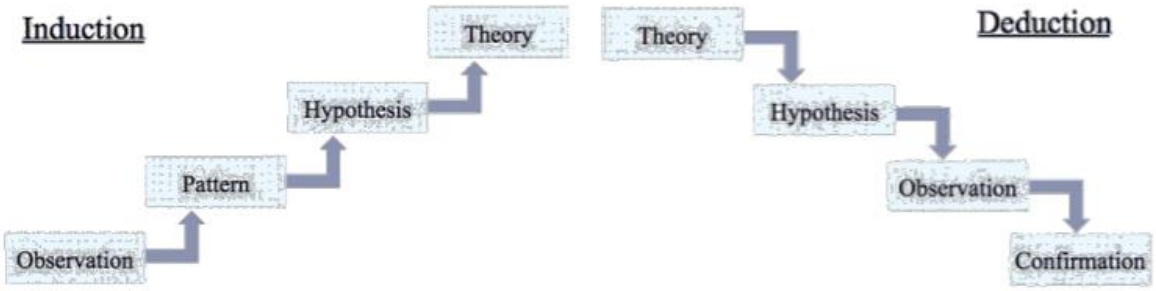


Figure 2:Inductive and Deductive Approach (Source Ragab, M., Arisha, A,2017)

3.2.3 Research Strategy

(Saunders, et al., 2009) explain that the research approach is conducted by the research questions and objectives. It is known as a crucial step in the research in which researcher chooses for collecting the data. There are many research strategies available. No research strategy is better than others. It is credible only when if it assists the research worker in finding the answer to the required research query.

Inductive approach is a qualitative method thus, qualitative method will be used as our research strategy. With the help of User evaluation, the prototype of the artefact will be tested by the user. With the help of google forms the user opinions are captured in a questionnaire and which is then evaluated to give recommendations. De Vaus (2002) identifies questionnaires as methods of getting response from a single or group of people in a predetermined order.

3.2.4 Ethics Applied to the Research.

All the ethical rules and guideline was followed while conducting this research. The following things were maintained in order limit the research to the boundaries of ethics:

- All the literature referenced for this study is referenced and put into the reference list

- No personal data of the user was recorded during the user evaluation. Personal details like
- names or email ids or any kind is asked from the user before participation.
- Integrity of the data from the feedback is maintained without any distortion.
- No data has been stored in the participants devices or no data about the participant device has been collected.

4. Artefacts Design and Development

4.1 Introduction

This section of the study will describe the comprehensive design strategy and development of the artefact. The artefact generates various visualization based on data fetched from the database and implement forecasting for time series data. User need to either select the required data's or write a SQL query to fetch the it, which is stored in a database for future use. Spring Boot framework were used to create a production ready spring MVC module. Spring boot belongs to the no code generation where most the configurations are done automatically with embedded server to run the module making it standalone, at the same time easy to integrate with other modules as a micro service. Java is used at background, where business logic and interaction between database and webpage are streamlined. Google charts.js is a JavaScript file which is used to generate visualization. MySQL DB is used as database used for storing the data for which visualizations are created.

Python is used to implement the deep learning algorithm for forecasting time series data. Eclipse is the IDE used for development of this artefact and MySQL workbench is used to interact the database

Firstly, the researcher will provide a brief overview of the software development life cycle (SDLC) model. The sections of the dissertation are structured in a way that it first discussed in breath about the phases of the software development model. Following development phases of the software development model, it describes different phases of the development of the SDLC model. This section is sub categorized into four sections.

4.2 Software Developmental Model Selection

Software is a collection of instruction written in a programming language in order to execute a function or requirement.

A software life cycle model provides a basic characterization of how a software needs to be developed. How a new software system should be developed are defined usually by prescriptive models, they are used as guidelines on how software development activities should be performed, and in which sequence (Scacchi and Walt,2002).These software life cycle models have the below phases:

- Planning
- Requirement Analysis
- Design and Coding
- Testing
- Deployment and Maintenance

After a series of consideration, the researcher choose to move along with the Iterative and incremental life cycle model as an working functionality is developed and deployed early and other functions can be incremental on further stages on top of the existing functionality ,it also has better risk factor and supports changing requirement.

4.3 Iterative Incremental Model

Adel Alshamrani(2015),stated that in iterative incremental model the first increment usually addresses the core functionality and both known and unknown supplementary features are delivered at the later stages slowly, a partial implementation of the entire system is constructed slowly in each and every increment.

4.3.1 Advantages of the Iterative Incremental model

1. High-risk or prioritized functions are developed first.
2. In High-risk projects, the risk is spread across multiple small increments than as a whole
3. Main functionality reaches the market early
4. Different operational product is delivered at each release.
5. Initial product delivery is faster.
6. Risk of failure and sudden change requests are reduced

4.3.2 Disadvantages of the Iterative Incremental model are −

1. Iterations should be planned effectively.
2. Inclusion of the both known and unknown functionality should be ensured while designing.
3. Precise and early definition of an entire and completely system to allow the definition of increments

4. Module interfaces should be well defined.

4.3.3 When to utilize Iterative and incremental Model

When the project as a whole has high risk or involves new technologies or when the deployment schedule is lengthy, thus developing in smaller incremental steps reduces risk, gets the basic functionality to market and allows the developer to gain better understanding of the technology.

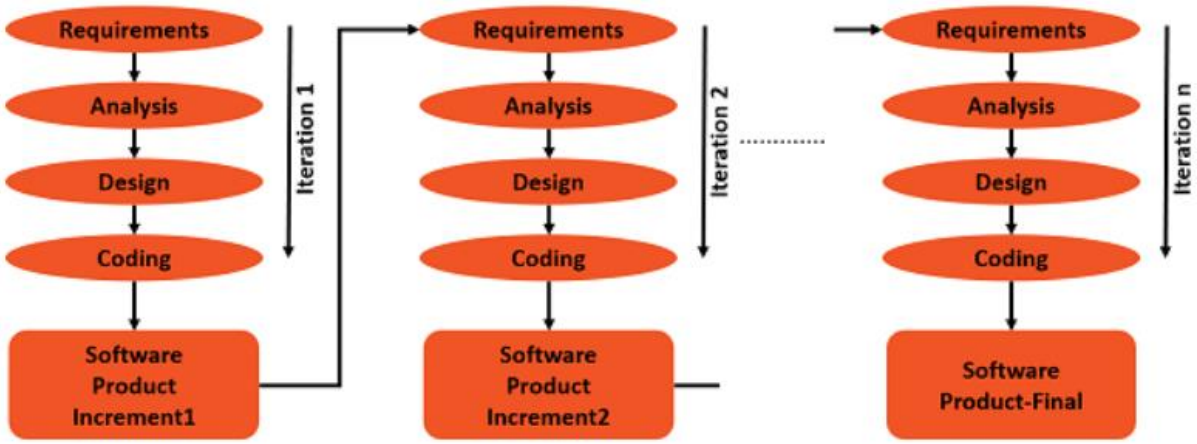


Figure 3: Iterative implementation (source: Suman Babu Chapram,2018)

4.4 Requirement Analysis of Software Developmental Model

Software requirements are nothing but descriptions of the services and the constraints on operation of the software. It doesn't not state what the software does rather reflect on the needs of customers for a system that serves a certain purpose." The process of

finding out, analysing, documenting and checking these services and constraints is called requirements engineering." (Sommerville, p.83,2010).

4.4.1 Functional Requirement

Functional requirements are abstract way of describing the functionality of a system, accepted inputs, outputs and exceptions in way for the users to understand, it varies. It depends on the system type developed and the target group for which it is developed for.

Table 1: Creating Query (For Normal User)

Title	**Creating Query (For Normal User)**
Requirement	User can select the fields from the database which needs to be visualized.
Dependencies	Internet
Restrictions and Risk	Fields from different entity can be joined only if foreign key is present in between them.
Priority	High

Table 2: Creating Query with advance SQL operations (For Developer)

Title	**Creating Query with advance SQL operations (For Developer)**
Requirement	User should be able to enter advance SQL queries instead of selecting the fields for visualization.
Dependencies	Internet, Knowledge on SQL
Restrictions and Risk	Can access sensitive database information's.
Priority	Medium

Table 3: Saving Query (For Developer)

Title	**Saving Query (For Developer)**
Requirement	All query for Visualization should be saved for future use
Dependencies	Internet
Restrictions and Risk	None
Priority	High

Table 4: Selection of visualizations

Title	**Selection of visualizations**

Requirement	System should provide user with the options to select the visualization of their choice
Dependencies	Internet, Knowledge on visualization
Restrictions and Risk	None
Priority	High

Table 5: Display the list of queries

Title	**Display the list of queries**
Requirement	Display list of all saved queries
Dependencies	Internet
Restrictions and Risk	None
Priority	High

Table 6: Model for forecasting data

Title	**Model for forecasting data**
Requirement	Individual deep learning model should be created for every time series data queries
Dependencies	None
Restrictions and Risk	None
Priority	High

Table 7: Saving Model

Title	Saving Model
Requirement	All model must be saved
Dependencies	None
Restrictions and Risk	None
Priority	High

Table 8: Replacing model

Title	Replacing model
Requirement	Models should be replaced in a specific time interval
Dependencies	New data
Restrictions and Risk	None
Priority	Low

Table 9: Forecasting data

Title	Forecasting data
Requirement	Data needs to be forecasted and saved in the database
Dependencies	None
Restrictions and Risk	None
Priority	High

4.4.2 Non-Functional Requirement

As the name suggest these requirements doesn't directly concerned about the system services, rather to the properties of the system and constraints on the implementations of the entire systems. They are more critical than the functional requirements as failing will render the system unusable. There are namely three type of non-functional requirement

- Product Requirement
- Organizational Requirement
- External Requirement

Table 10: Non-Functional Requirement

Title	Description
Scalability	The system should be able to expand and handle large number of queries, model execution simultaneously without or with an acceptable loss in reliability and availability. Also adapting to varying load

Reliability	As an online service, the basic need for the system is it should be available 24*7 to all users with a 99% availability, excluding maintenance downtime
Usability	Main purpose of the system is to reduce the user workload of formatting data and generating visualization, thus usability plays a major role in this project. 80-90% of the input are provided using generated values thus errors are reduced. Appropriate legends, titles, examples, help notes, and alerts are provided to make it easier for the user to understand.
Supportability	As a web application it should run on both desktop and handheld device with internet and support major browsers like chrome, Firefox, Edge.
Performance	Performance here denotes the response time of the webpage. Response time should be not more than 60 seconds for generation of visualization provided with good internet strength. Response time to any action like

	form submission, various internal events should not be more than 10 seconds.

4.5 Design of artefact

In this section the artefact's architectural design will be discussed in detail. First phase of any software design is architectural design, which is nothing but the blueprint of the system. It helps us in allocating the required resources, identifying the backbone of the system and finding connection between different components of the system. (Sommerville, 2000). The reliability, performance and stability of the system greatly depends on the architecture design phase.

This section gives a brief description of all the system design and diagrams related to the design. Following are the main diagrams used to define a system

- System Architecture
- Flowchart
- Sequence Diagram
- Data Flow Diagram
- System Context Diagram

Following are the key components of the system:

1. Creation of Query: First the user needs to select the fields from the available options which is populated from the database and then a suitable title for the query, which will be saved in the database for future usage.

2. Fetching & Formatting Data Based on Query: On selecting the required query for viewing, the data should be fetched based on the query from the database and then to formatted for visualization.
3. Rendering Client-side Visualization: The formatted data should be populated in the JavaScript file based on the type of visualization and the visualization should be rendered in the client browser, irrespective of the resolution, system properties and browser.
4. Forecasting of Data: Deep learning model for each query with time series data need to be created and the reference should be stored in database, which is replaced on intervals. Every model should be executed and forecasted data needs to be stored in database.

4.6 System Architecture Diagram

Figure 2 represents the system architecture of the artefact. The artefact is built on spring MV C framework, Input from the user is sent to the controller which maps it to the respective service, which in turns pass the control to DAO layer, then the DAO layer interacts with the database using ORM layer, the requested data is then formatted and passed to view for rendering the webpage in the browser. The scheduler runs the python script at fixed time and stores the model in file system and reference in the database.

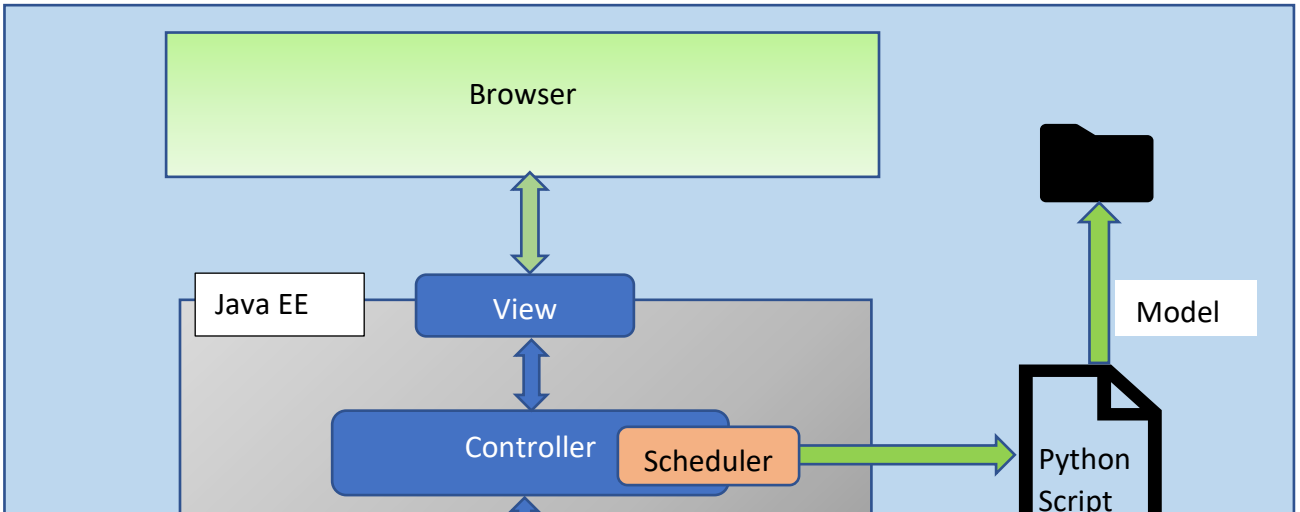


Figure 4: System Architecture

4.7 Sequence Diagrams

Sequence diagrams represents the flow of events in the chronological order of execution to achieve a specific function in a system. Thus, a system can be represented as a collection of sequence diagram. It is a useful tool, when the project is moving from documentation to implementation as it shows the step by step flow of functions. Below is sequence diagram of the main functions of the proposed system(Gerhard Greeff and Ranjan Ghoshal,2004).

In this sequence diagram, the function of adding a query for visualization is pictured.

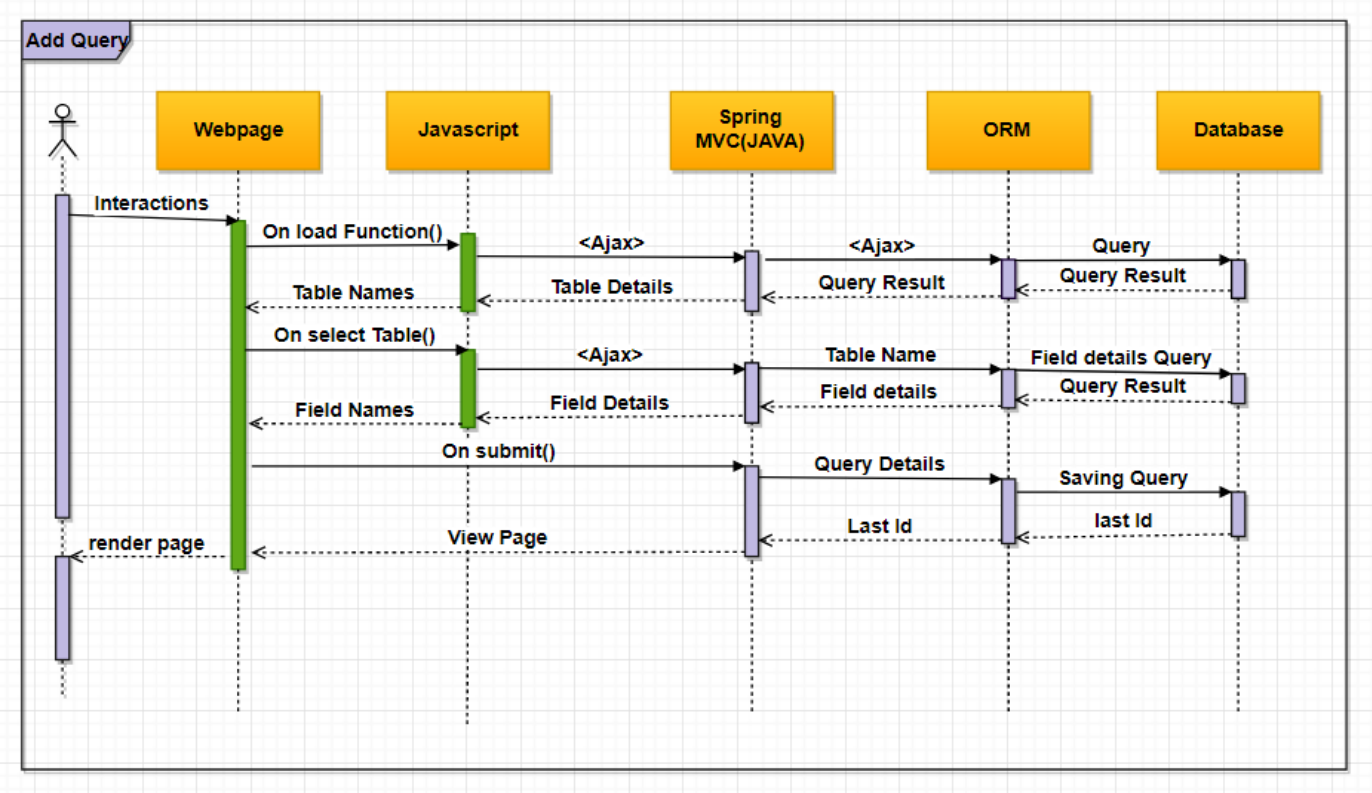


Figure 5 :Sequence Diagram for Adding a Visualization Query

In this sequence diagram, the flow of generation of the visualization of a selected query is described.

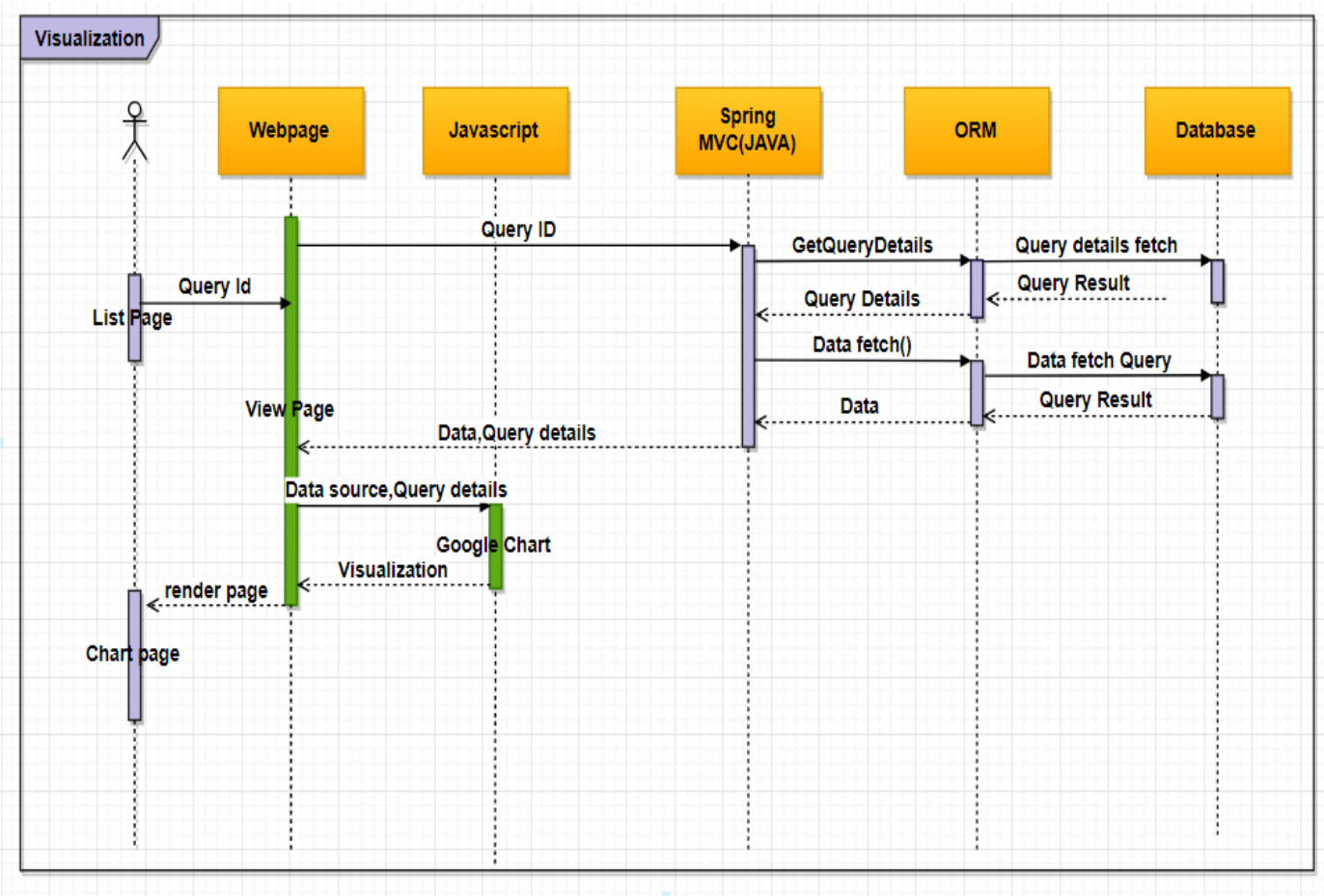


Figure 6:Sequence Diagram for Visualization of data

In this sequence diagram, the forecasting functionality of the system is described.

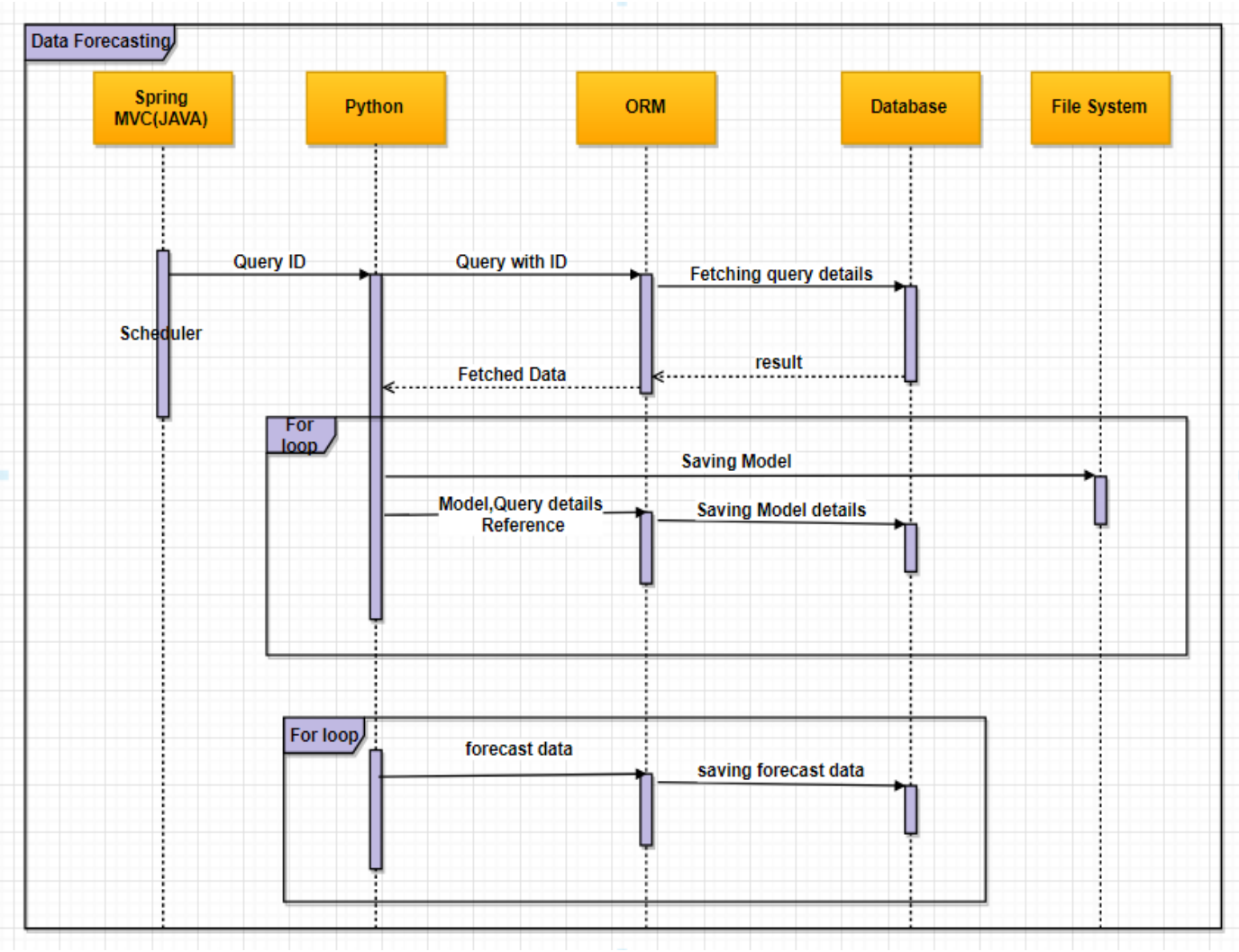


Figure 7:Sequence Diagram for Forecasting

4.6. Technology Overview

Java: The main programming language used in this application is Java. It is an object-oriented, multithreaded language which is both a language and a platform, thus making it easier to run programs created in java to run in any environment. A language which was built specifically for the internet age. It is simple, robust, secure, faster and provides high performance. Java is used for server-side implementation of our web module

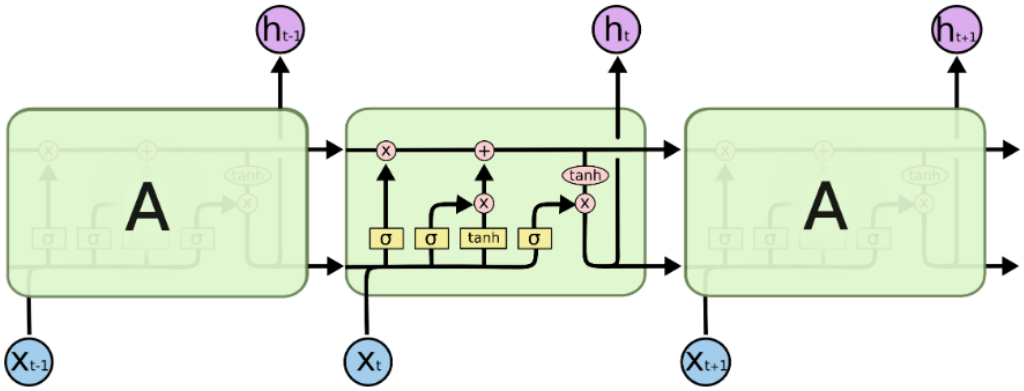
MySQL: MySQL is an open-source relational database management system developed for next generation web, mobile and other applications. Data is stored in form of table and SQL query is used to perform ACID operations. Whereas MySQL workbench is used as a graphical user interface for the database

JavaScript: It is a lightweight, cross platform programming language with object-oriented capabilities used along with HTML. With the help of this, webpages are made more dynamic, client-side rendering became possible and validation can be done on the client side instead of server side. It also became the source code for jQuery. JavaScript is used for rendering visualization

Google Chart JS: It is an open source JavaScript based library for generating charts and graph, developed by google. Our client-side visualization is generated using this JavaScript.

Python: Python is object oriented, interpreted scripting language in which the scripts are interpreted instead of compiled. It is extensively used in machine learning, deep learning, neural networks, etc. long short-term memory (LSTM) deep learning methodology for forecasting data is executed using python.

LSTM: “Long short-term memory (LSTM) is a recurrent neural network with a state memory and multilayer cell structure.” (Kamilya Smagulova,2019). Mainly developed to avoid the problem created in other RNN due to retention of long-term data. It is composed of cell to remember the arbitrary value and input, output and forget gate to change the value of the cell. It has an edge over other algorithms for time series data analysis.



The repeating module in an LSTM contains four interacting layers.

Figure 8:Long Short-Term Memory Internal Structure

4.5. Implementation

In this section, the researcher will discuss about the approaches taken to implement the Generic web-console for dynamic visualization and Time series forecasting application. This study elaborates on a step by step manner of implementing the requirements one by one until the application was completed.

4.5.1 Generating Query Details

This is the first and only input step present in the application, where the user generates the query for visualization by selecting the fields available. The user might enter a incorrect input or a field that might not exist or a field that has multiple occurrence in the database causing ambiguity, in order to eliminate these type of user error, most of the inputs in the applications have been configured as options which are populated by getting the data from the database.

Figure 9 shows the UI of the Add Query page where the user creates the query for visualization, as observed except title, column name and condition field other fields are populated from the database using Ajax and Json, already selected columns and conditions are shown at the right. Also, a column can be selected only once and a condition can be entered only once, if the user tries to select or enter an existing column or condition an alert is shown to the user with the help of JavaScript.

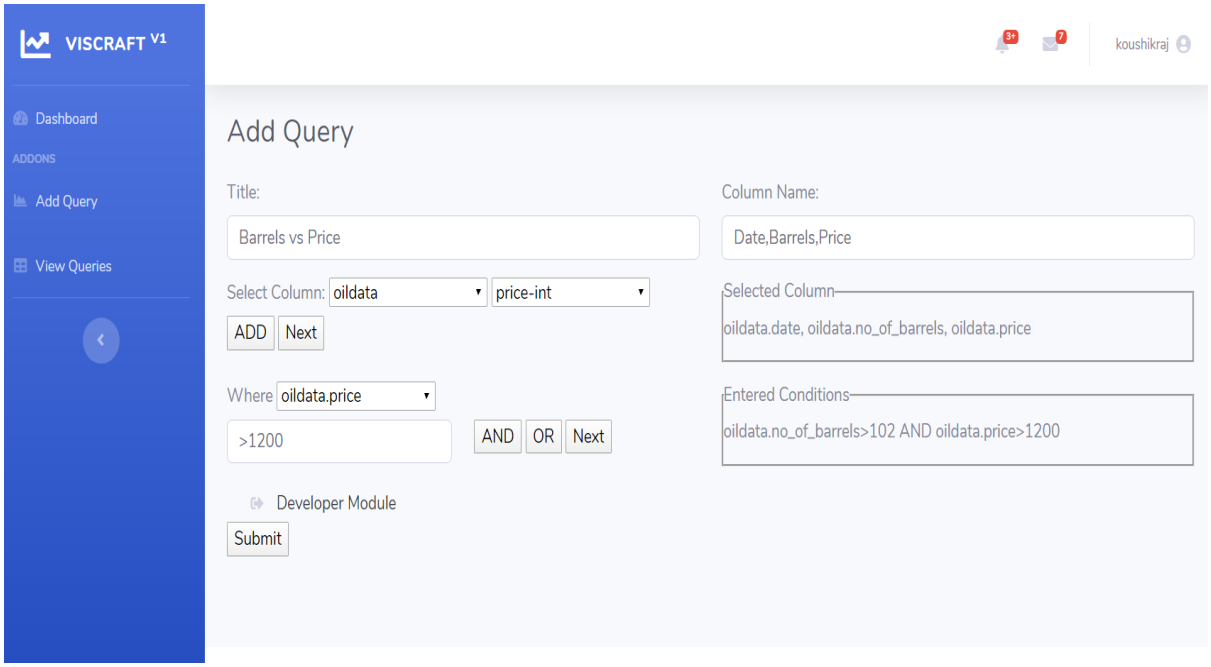


Figure 9:Add Query Page UI

Figure 10 shows the developer section in the add query page, which shows columns needed for visualization but values are populated automatically with the help of JavaScript as the user selects the column and conditions, further the user can modify the generated SQL Query if any advance SQL functions like concat , SUM, etc need to be used in the query or any other column values if needed.

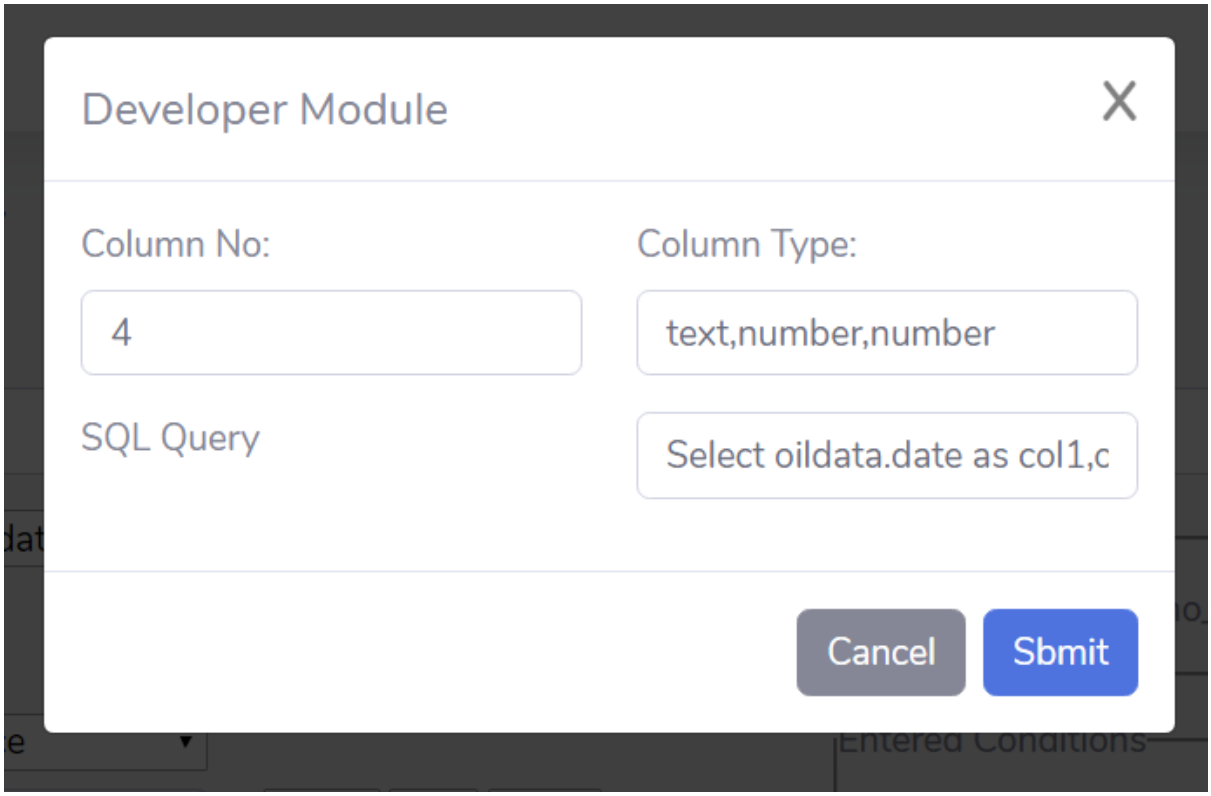


Figure 10:UI of Developer Module Pop-Up

Following is the Ajax code for populating the fields on selecting the table:

```
$('#table_name').on('change', function() {
        $.ajax({
              type: "POST",
              contentType: "application/json",
              url: "/search",
              data: JSON.stringify(this.value),
              dataType: 'json',
              cache: false,
              timeout: 600000,
              success: function (data) {
                  console.log("SUCCESS : ", data);
                  $('#column_name').find('option').remove().ends
                  $('#column_name').append($("  <option  disabled  selected
value> -- select an option -- </option>"));

                  $.each(data, function(key, value) {

                      $('#column_name')
                          .append($("<option></option>")
                          .attr("value",value)
                          .text(value));
                      console.log(value);
                  });
              },
              error: function (e) {

                  var json = "<h4>Ajax Response</h4><pre>"
                      + e.responseText + "</pre>";
                  $('#feedback').html(json);

                  console.log("ERROR : ", e);
                  $("#btn-search").prop("disabled", false);

              }
          });

      });
```

Following is the JavaScript for populating the where once the fields are selected and generating data for hidden fields:

```
$("#list1").click(function() {

	if(no_of_tables==1){
		slected_tables.forEach(fire_ajax_submit);
	 $.ajax({
		type: "POST",
		contentType: "application/json",
		url: "/testing",
		data: JSON.stringify(slected_tables),
		dataType: 'json',
		cache: false,
		timeout: 600000,
		success: function (data) {
		    console.log("SUCCESS : ", data);
		},
		error: function (e) {

		    var json = "<h4>Ajax Response</h4><pre>"
		        + e.responseText + "</pre>";
		    $('#feedback').html(json);

		    console.log("ERROR : ", e);
		    $("#btn-search").prop("disabled", false);

		}
	    });



	 }

	else if( no_of_tables==2){
		for (index = 0; index < slected_tables.length; index++) {
		    console.log(slected_tables[index]);



		}
		slected_tables.forEach(fire_ajax_submit);

	}
	});

$( "#selected_column" ).click(function() {
	table=$('#table_name').val();
	var temp=$('#column_name').val().split("-");
```

```
        field=temp[0];
        var type;

        if(temp[1]==("varchar")||temp[1]==("text"))
        {
        type="col"
        col_type.push("text")
        }
        else if(temp[1]==("int"))
        {
        type="i"
              col_type.push("number")
        }
        else if(temp[1]==("float"))
        {
        type="d"
              col_type.push("number")
        }


    if(list==null){
                list="Select "+table+"."+field+" as "+type+no_of_fields ;
                console.log(list);
                slected_tables.push(table);
                selected_column.push(table+"."+field);
                $('#fields')
          .append(table+"."+field);
                console.log(slected_tables);
                no_of_fields=no_of_fields+1;
        }
        else  if (!list.includes(table+"."+field)) {
                list=list+","+table+"."+field+" as "+type+no_of_fields ;
                console.log(list);
                selected_column.push(table+"."+field);
                $('#fields')
          .append(", "+table+"."+field);
                if (slected_tables.indexOf(table) < 0) {
                      slected_tables.push(table);
                      console.log(slected_tables);
                      no_of_tables=no_of_tables+1;
                      }
                no_of_fields=no_of_fields+1;
        }
                });
```

4.5.2 Fetching and Fitting Data into Graph Data Source

On selecting the Query to be visualized and type of visualization, the details of the query is first fetched from the database , then the SQL Query in the query details is executed to fetch the desired data and other details are used as description for the data to be visualized. The main problem faced by the researcher is categorizing and storing the data

based on the data type, then populating the data source of the visualization. Thus, when the user selects the field the data type of the field is detected and based on that and the position of the field an alias name is given to the field in the SQL query automatically.

Example: SQL Query=" **Select oildata.date as col_1,oildata.no_of_barrels as i2,oildata.price as i3 from oildata where oildata.no_of_barrels>102 AND oildata.price>1200"**

In the above example," oildata.date" is a string type data thus the first part of the alias name is "col_" and it is in first position thus the second part is "1".Simiarly for "**oildata.no_of_barrels**" is an integer thus the alias name is "**i2**".

Figure 11 shows the bean structure called "queryData" designed to store a row of the dataset, thus the entire dataset is stored in a list of queryData, which is then passed to view with the help as model value. The view is nothing but the webpage where the visualization is going to be created, using JSP scriptlet technology the data is populated in JavaScript data source for the visualization.

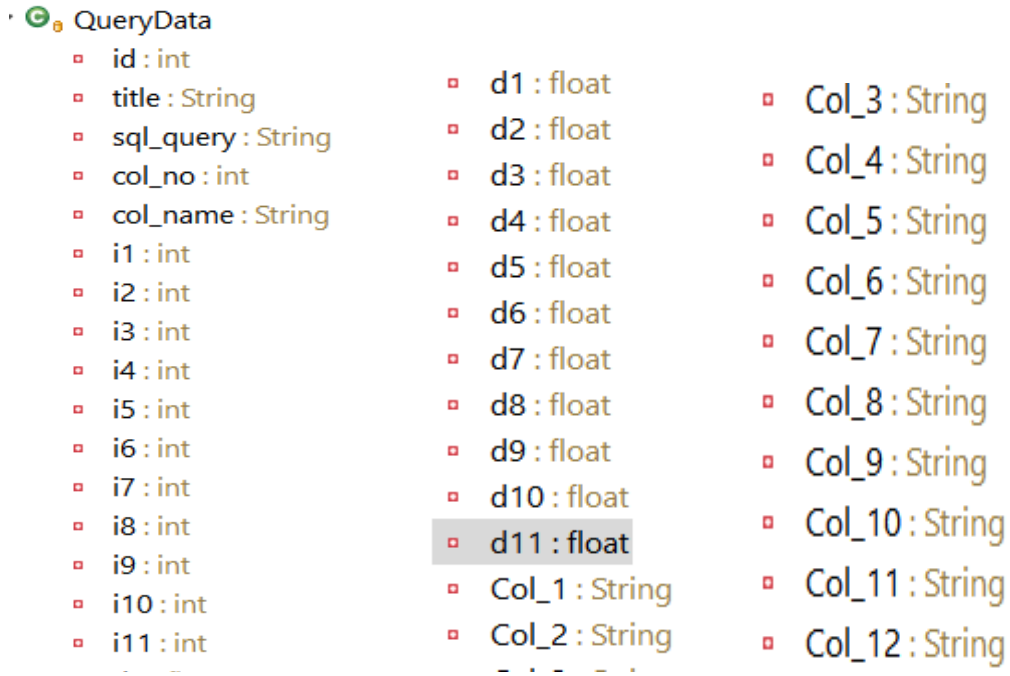


Figure 11 Bean Structure for Storing Data

Following is the code for population of Data and its details in JavaScript using JSP scriptlet:

```
        <% QueryPair queryPair = (QueryPair)request.getAttribute("data");
        List<QueryData> dataList=queryPair.getQueryData();
        QueryDetails queryDetails=queryPair.getQueryDetails();

        int col_size=queryDetails.getCol_no();
        List<String>legends=new
ArrayList(Arrays.asList(queryDetails.getCol_name().split(",")));
        %>
        [
        <% for(int j=0;j<legends.size();j++){%>
        '<%=legends.get(j)%>'      <%if(j<legends.size()-1){%>
        ,
        <%}%>
        <%}%>],
        // ['Year', 'Sales', 'Expenses'],
    <% for(int i=0;i<dataList.size();i++){%>
            [

             <% if(col_size>0){%>'<%if(dataList.get(i).getCol_1()!=null){%>
<%=dataList.get(i).getCol_1()%> <%} else if(dataList.get(i).getD1()!=0){%>
<%=dataList.get(i).getD1()%><%} else{%> <%=dataList.get(i).getI1()%><% }%>'
             <%} if(col_size>1){%>,<%if(dataList.get(i).getCol_2()!=null){%>
<%=dataList.get(i).getCol_2()%> <%} else if(dataList.get(i).getD2()!=0){%>
<%=dataList.get(i).getD2()%><%} else{%> <%=dataList.get(i).getI2()%><% }%>
             <%} if(col_size>2){%>,<%if(dataList.get(i).getCol_3()!=null){%>
<%=dataList.get(i).getCol_3()%> <%} else if(dataList.get(i).getD3()!=0){%>
<%=dataList.get(i).getD3()%><%} else{%> <%=dataList.get(i).getI3()%><% }%>
             <%} if(col_size>3){%>,<%if(dataList.get(i).getCol_4()!=null){%>
<%=dataList.get(i).getCol_4()%> <%} else if(dataList.get(i).getD4()!=0){%>
<%=dataList.get(i).getD4()%><%} else{%> <%=dataList.get(i).getI4()%><% }%>
             <%} if(col_size>4){%>,<%if(dataList.get(i).getCol_5()!=null){%>
<%=dataList.get(i).getCol_5()%> <%} else if(dataList.get(i).getD5()!=0){%>
<%=dataList.get(i).getD5()%><%} else{%> <%=dataList.get(i).getI5()%><% }%>
             <%} if(col_size>5){%>,<%if(dataList.get(i).getCol_6()!=null){%>
<%=dataList.get(i).getCol_6()%> <%} else if(dataList.get(i).getD6()!=0){%>
<%=dataList.get(i).getD6()%><%} else{%> <%=dataList.get(i).getI6()%><% }%>
             <%} if(col_size>6){%>,<%if(dataList.get(i).getCol_7()!=null){%>
<%=dataList.get(i).getCol_7()%> <%} else if(dataList.get(i).getD7()!=0){%>
<%=dataList.get(i).getD7()%><%} else{%> <%=dataList.get(i).getI7()%><% }%>
             <%} if(col_size>6){%>,<%if(dataList.get(i).getCol_8()!=null){%>
<%=dataList.get(i).getCol_8()%> <%} else if(dataList.get(i).getD8()!=0){%>
<%=dataList.get(i).getD8()%><%} else{%> <%=dataList.get(i).getI8()%><% }%>
             <%} if(col_size>6){%>,<%if(dataList.get(i).getCol_9()!=null){%>
<%=dataList.get(i).getCol_9()%> <%} else if(dataList.get(i).getD9()!=0){%>
<%=dataList.get(i).getD9()%><%} else{%> <%=dataList.get(i).getI9()%><% }%>
             <%} if(col_size>6){%>,<%if(dataList.get(i).getCol_10()!=null){%>
<%=dataList.get(i).getCol_10()%> <%} else if(dataList.get(i).getD10()!=0){%>
<%=dataList.get(i).getD10()%><%} else{%> <%=dataList.get(i).getI10()%><% }%>
             <%} if(col_size>6){%>,<%if(dataList.get(i).getCol_11()!=null){%>
<%=dataList.get(i).getCol_11()%> <%} else if(dataList.get(i).getD11()!=0){%>
<%=dataList.get(i).getD11()%><%} else{%> <%=dataList.get(i).getI11()%><% }%>
```

```
<%}%>
]
```

4.5.2 Implementing of LSTM

This is the second segment of the applications where values had to be forecasted based on queries. Python was used as programming language for implementing LSTM. But as the researcher was not sure which model is best fit for time series analysis, so the researcher had to try out a number of Machine learning and deep learning techniques in order to find the best suitable method. As discussed in literature review, time series two main component trend and seasonality. Most of the models like Arima, SArima, FbPhtophet was either not able to capture the seasonality and trends or not able to produce a reliable model with less root mean square error without the intervention of a user. Only LSTM was able to give the lowest root mean square error, as well as reproduce the trends and seasonality to a better than other models. Next comes the number of fields in query, thus each field had to be iterated and model for each field in the query has been generated.

Following is the code for converting dataset to required format.

```
train_size = int(len(df1) * 0.80)
test_size = len(df1) - train_size
train = data[0:train_size]
valid = data[train_size:]
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data)

x_train, y_train = [], []
for i in range(6, len(train)):
        x_train.append(scaled_data[i - 6:i, 0])
        y_train.append(scaled_data[i, 0])
x_train, y_train = np.array(x_train), np.array(y_train)

x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
```

Following is the code for fitting data into the model.

```
model = Sequential()
    model.add(LSTM(units=300, return_sequences=True,
input_shape=(x_train.shape[1], 1)))
    model.add(LSTM(units=25))
    model.add(Dropout(0.15))
    model.add(Dense(1))

    model.compile(loss='mean_squared_error', optimizer='adam')
    history_callback = model.fit(x_train, y_train, epochs=80, batch_size=12,
validation_data=(valX, valY), verbose=1)
```

4.5.3 Saving LSTM Model and Reference in Database

Every model created is saved in the file system on the server using a python library and the name of the model is saved in the database as a reference along with the query it belongs too.

Following is the code for saving the model and reference.

```
import pickle
    dateTimeObj = datetime.now()
    date_time = dateTimeObj.strftime("%m-%d-%Y_%H-%M-%S")
    filename = "Query_"+str(id)+"_ p_value_"+str(y)+"_"+date_time
    pickle.dump(model, open(filename, 'wb'))
    print(valid[[listOfColumnNames[y], 'Predictions']])
    mySql_insert_query = "INSERT INTO prediction_model (sql_id,
p_value_"+str(y)+") VALUES ("+str(id)+",'"+filename+"')ON DUPLICATE KEY UPDATE
p_value_"+str(y)+"='"+filename+"';"
    cursor.execute(mySql_insert_query)
    connection.commit()
```

4.5.6 Scheduling Forecasting of Data

As new data come in the existing model needs to be updated with new model, thus scheduler library of java is used to run the python scripts on specific interval. Once the model is generated then it runs the second script which forecast data based from the model. Then the forecasted data is stored in database with reference to the field and query in string format.

Following is the code for scheduling python scripts from java environment.

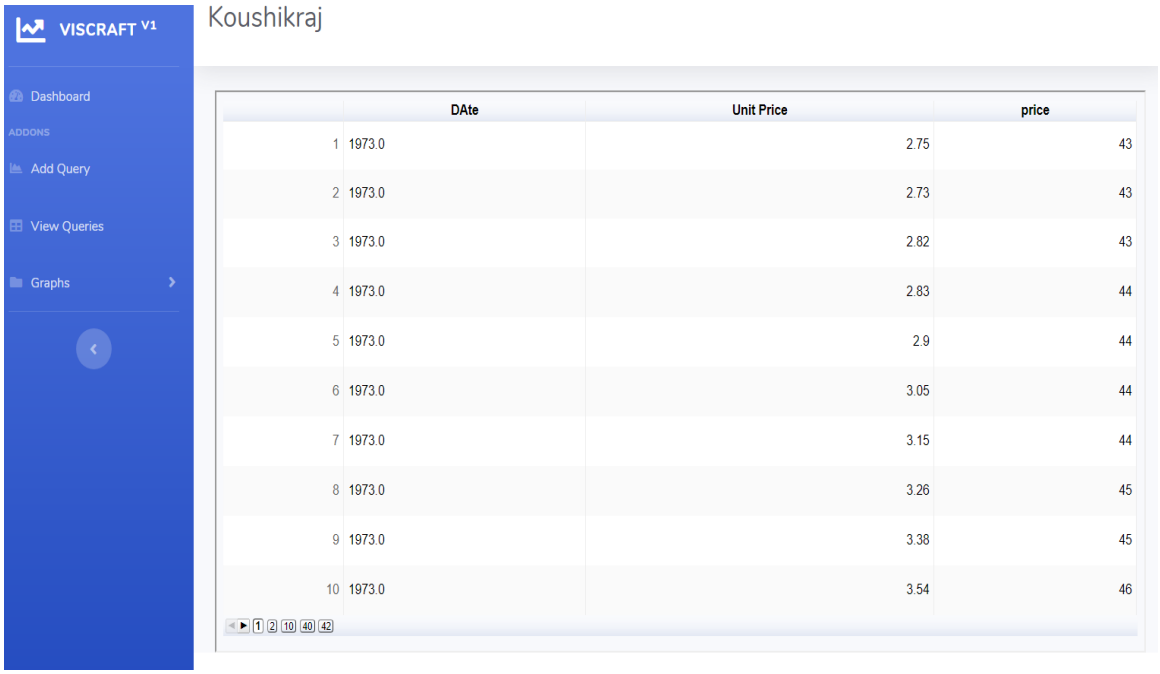
```
        final Runnable script = new Runnable() {
             public void run() {
              try {

              String command = "C:\\Users\\91984\\Anaconda3\\pythonw.exe
F:\\Dessertation\\Time_series\\fits\\lstm.py";
String command = "C:\\Users\\91984\\Anaconda3\\pythonw.exe
F:\\Dessertation\\Time_series\\fits\\lstm2.py";

                                    Process p = Runtime.getRuntime().exec(command  );
                               } catch (IOException e) {
                                    // TODO Auto-generated catch block
                                    e.printStackTrace();
                          }
         };
         ScheduledFuture beeperHandleAtFixedRate =
scheduler.scheduleAtFixedRate(script, 1, 35,TimeUnit.MINUTES);
```

4.5.7 Type of Visualizations Implemented

Table: Set of data displayed in a sequential form of rows and column.



	DAte	Unit Price	price
1	1973.0	2.75	43
2	1973.0	2.73	43
3	1973.0	2.82	43
4	1973.0	2.83	44
5	1973.0	2.9	44
6	1973.0	3.05	44
7	1973.0	3.15	44
8	1973.0	3.26	45
9	1973.0	3.38	45
10	1973.0	3.54	46

Figure 12:Table View of Data

Bar Graph: It is a representation of data in the form of horizontal bars, the height of the bars (y-axis) represent the value and x-axis denotes the discrete categories. It is commonly used to compare data among different groups.

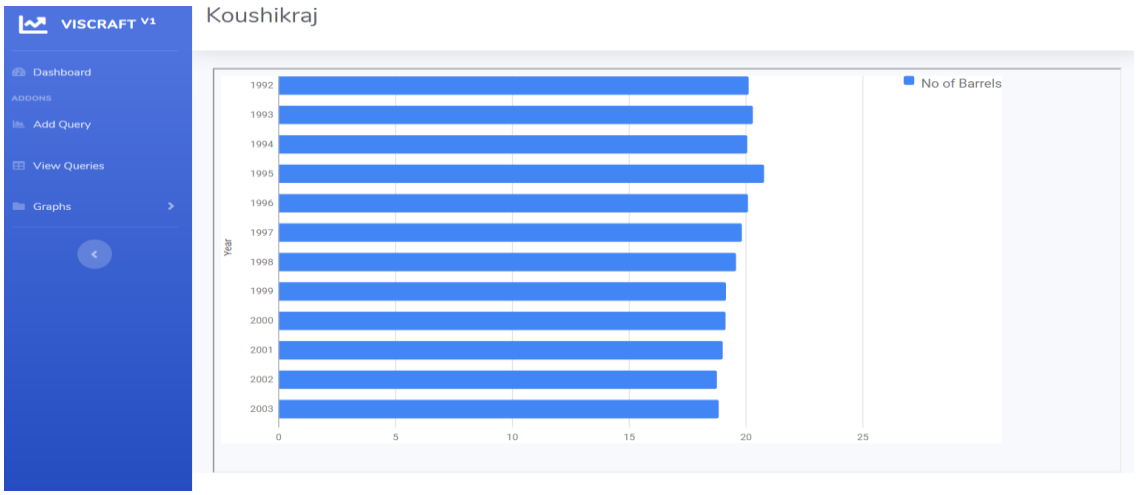


Figure 13:Bar Chart

Line Chart: A line chart displays data's as a series of points connected by a straight line. It is a one of basic chart for comparing different sets of continues data.

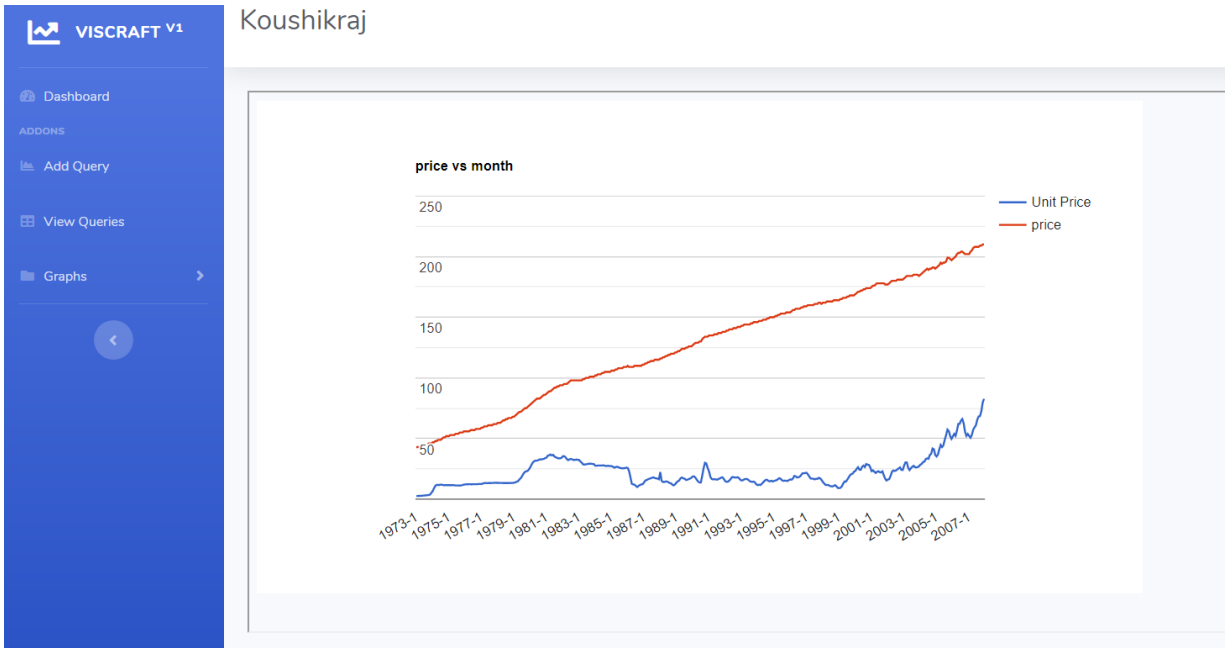


Figure 14:Line Chart

Area Chart: An area chart is based on the line chart, where the region between the line and the axis are filled.

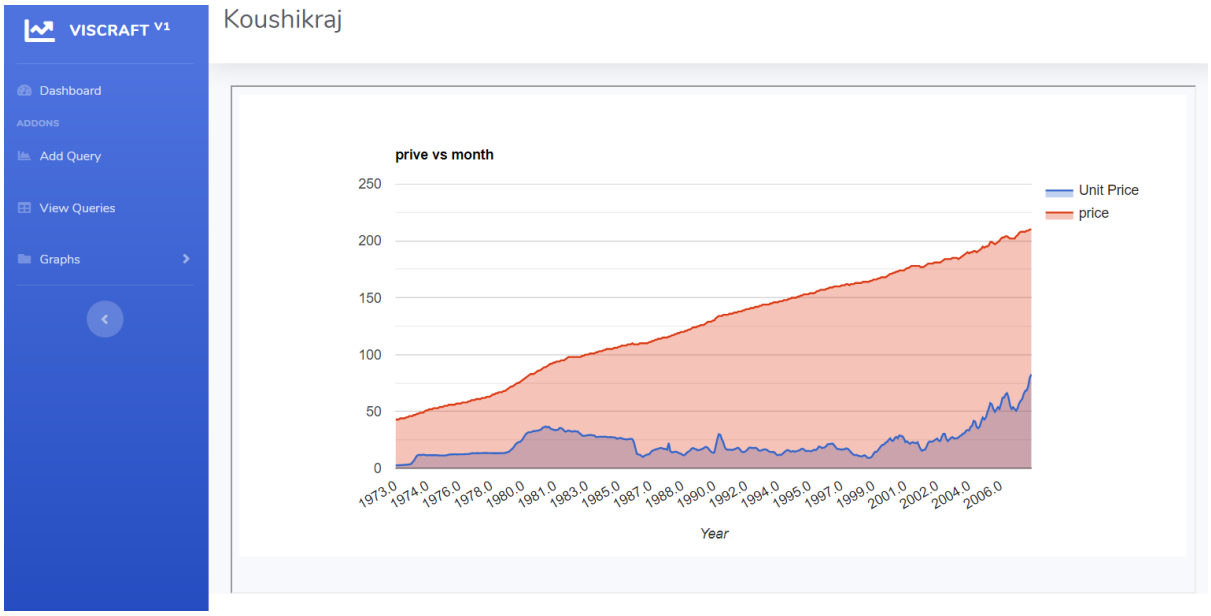


Figure 15 Area Chart

Scatter Plot: Visualization of data in the form of plotted points using Cartesian coordinates.

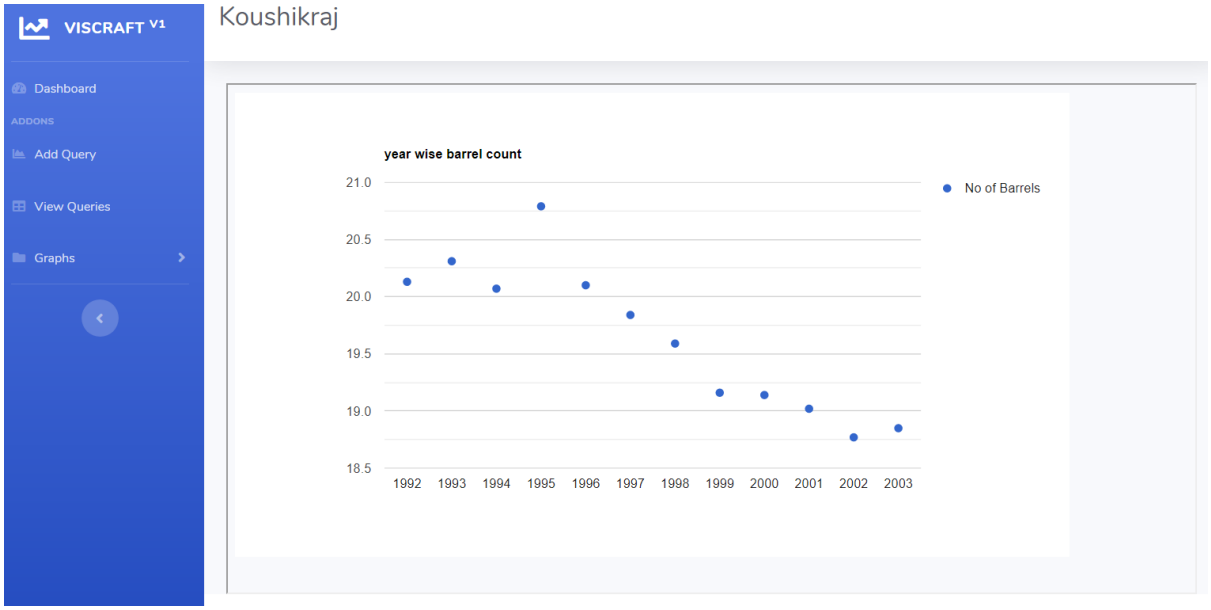


Figure 16:Scatter plot

Pie Chart: It is a statistical graph in circular form, divided into segments based on the percentage or proportions between the categories in the data.

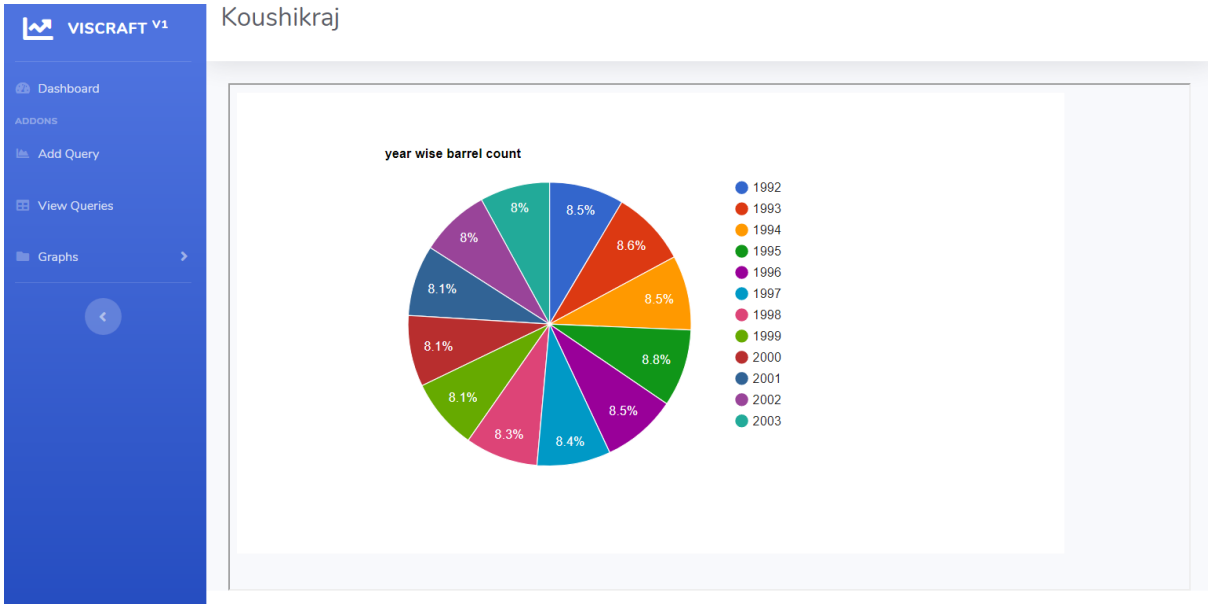
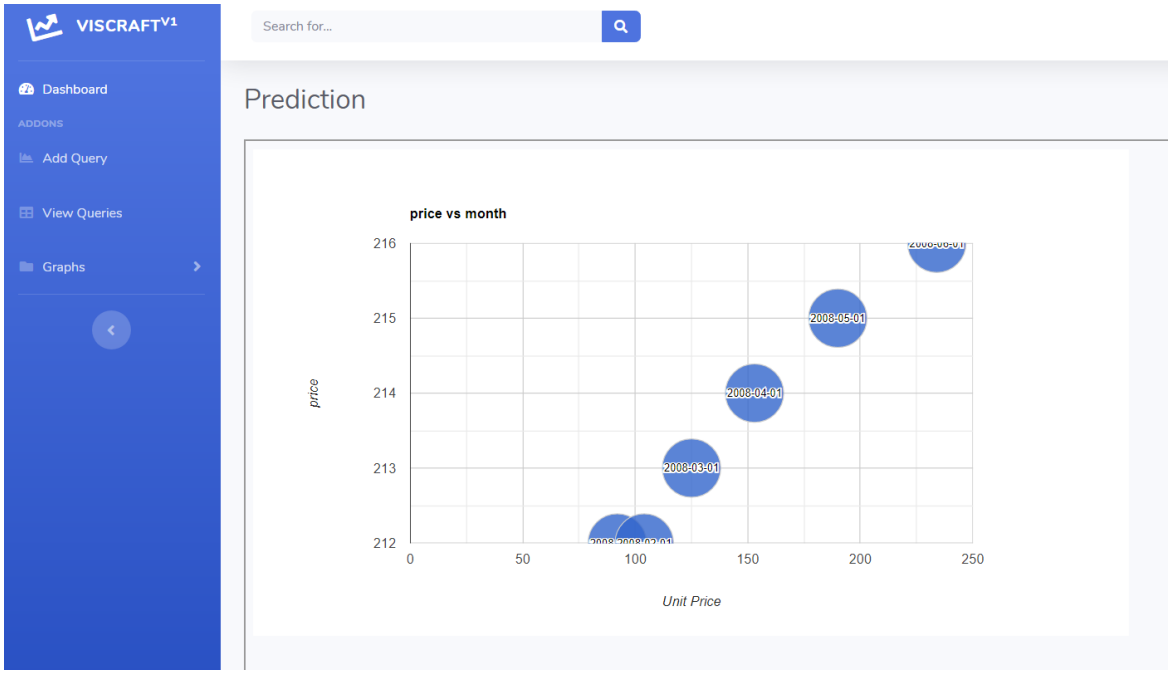


Figure 17:Pie Chart

Bubble Chart: It represent data in 3 dimension, along the X-Y-Z axis .X and Y corresponds to the coordinate and Z axis denotes the radius of the circle.



Histogram: It is a similar to bar chat, but groups the data into ranges. It is useful for visualizing continues data than discrete data. Taller the bar is the more data falls in that group.

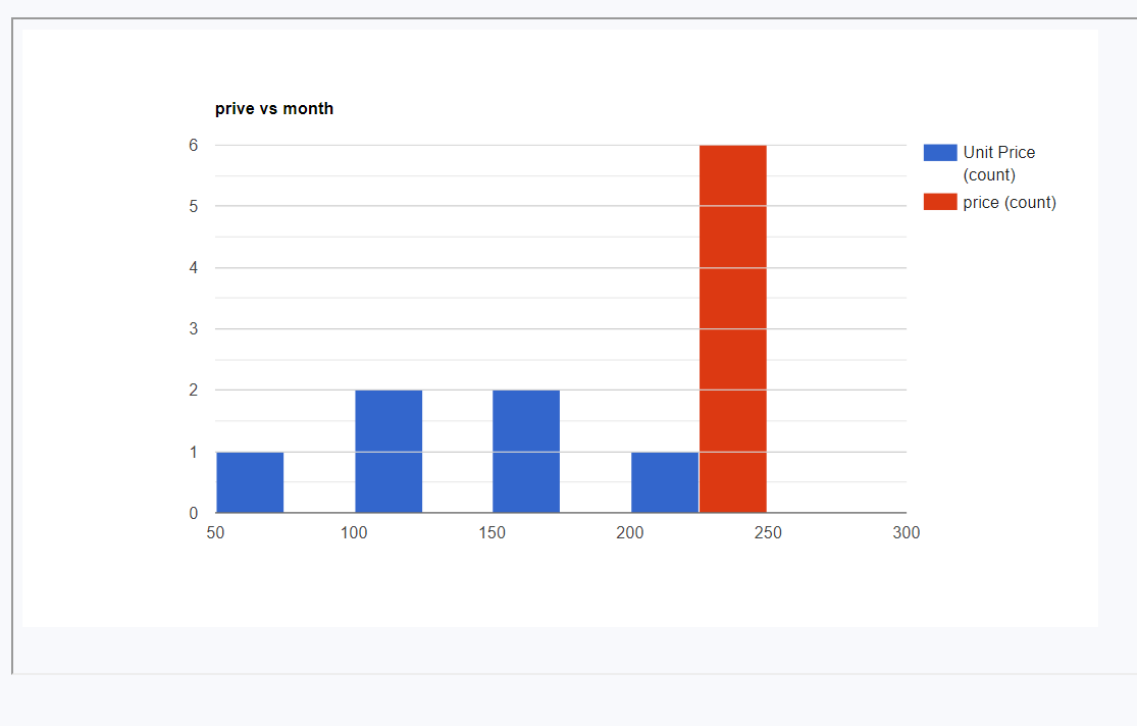


Figure 18:Histrogram

Column Chart: Column charts are similar to bar charts with only one exception, that is the bars are positioned vertically rather than horizontally.

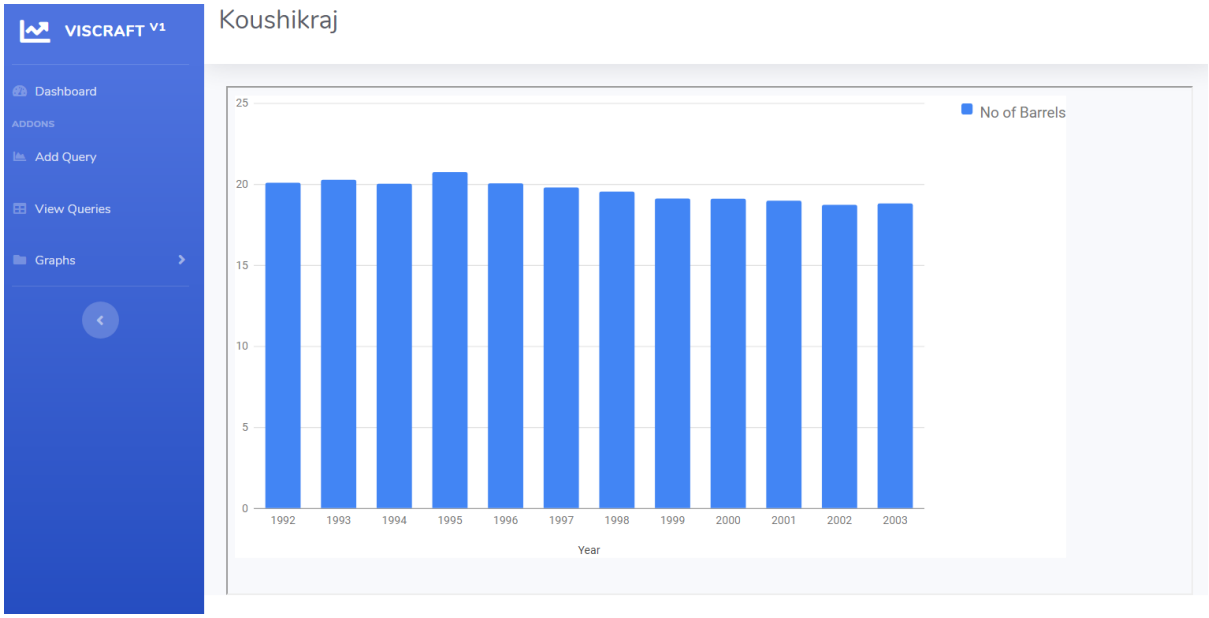


Figure 19:Column Chart

Stepped Area Chart: A stepped area chart is another type of area chart in which horizontal and vertical line connects the data points, creating a step like structure.

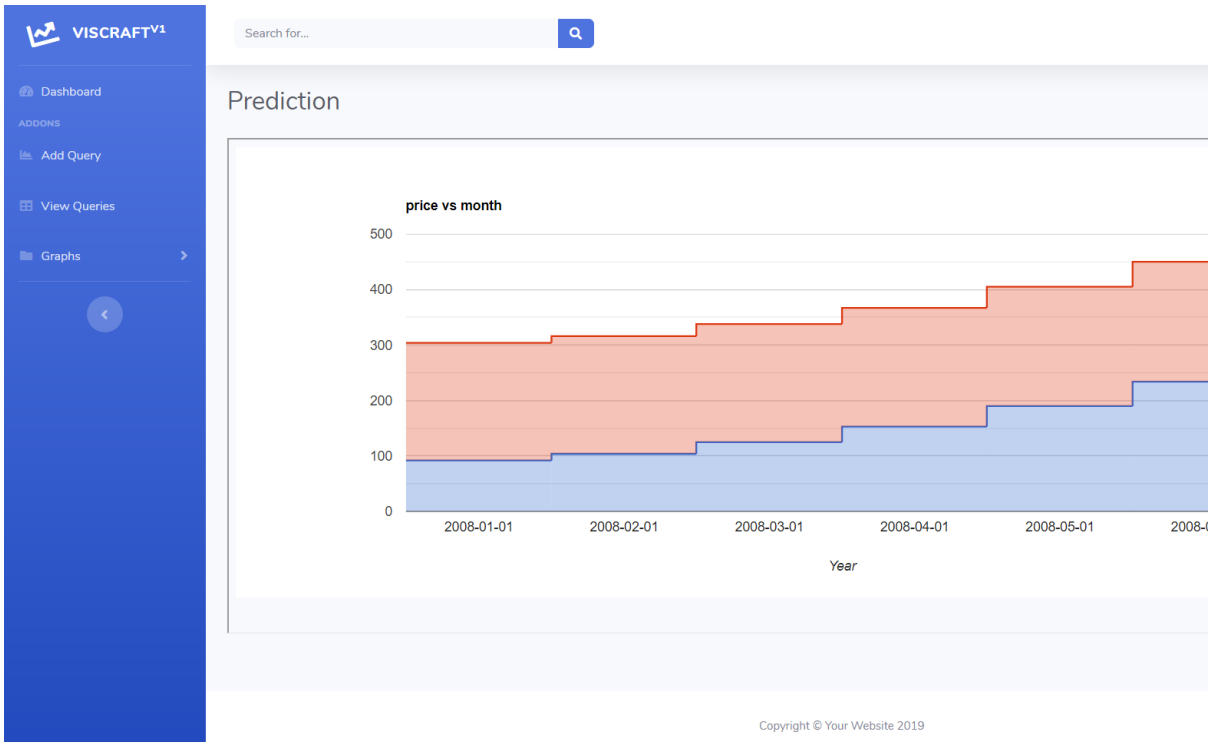


Figure 20:Stepped Area Chart

Donut Chart: It is another version of pie chart with a hollow at the centre of the circle.

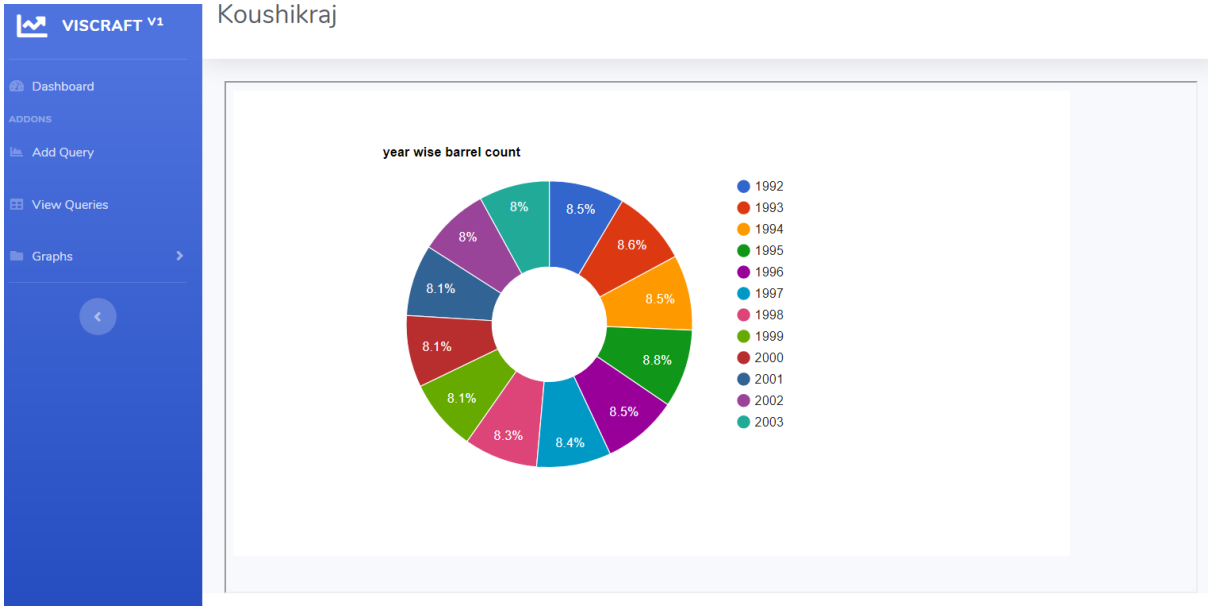


Figure 21:Donut Chart

CHAPTER 5: Data Analysis AND Discussions

5.1 The User Evaluation

The Testing of the artefact was done with the help of users. The user tests the artefact, after which they help in evaluating the artefact by answer the questionnaire prepared by the researcher.

The focus group of this user evaluation are developers and Analyst, who had abundant experience in data analysing, visualization, software development. The users where told to create a visualization using the artefact as a tool, the details of the case studies are given in the Appendix in detail. The user tool was hosted online for a short span of time and the URL along with the case study shared with users. The user where given the liberty of accessing it via mobile or laptop. Once the user interaction was finished, they were asked to fill out the feedback questionnaire on google forms. There were totally 11 Questions ,10 of which are multiple choice and the last question is a descriptive type. In accordance to the GDPR Norms, no personal information's where captured in the questionnaire nor in any form.

5.1.1 Questionnaire

A questionnaire is considered to be a research instrument with a set of questions, in-order to collect information from the users. A research questionnaire is a combination of both close and open-ended questions, which was developed in 1838 by the Statistical Society of London.

The research questionnaire used for this user evaluation can be categorized into different sections quality, design, usability, complacency and future scope. Quality based questions is used to evaluate the efficiency of the artefact in various aspect of response time and understandability. Design Based question are used to evaluate the user experience of the

product. Usability based question are centred around to find out whether the artefact will be useful now or in the near future as a whole or as an interface. Complacency question are concerned about the satisfaction of the user. Finally, the future scope questions take suggestions on what can be improved and implemented in future.

Table 11: Questionnaire for User Evaluation

ID	Questions	Category
1	How easier was it to create visualization for the first time in the artefact?	Quality
2	How fast was the response time of generating and switching between visualizations?	Quality
3	Is enough information provided in the form of titles, examples, legends, pop up, alerts?	Design
4	Was the UI appealing and less clustered with only relevant information on the pages?	Design
5	Where you able to navigate between the pages seamlessly and undo/redo actions done?	Design
6	will you consider using this module either as standalone or as an interface in the future?	Usability
7	What was the most valuable and usable function?	Usability
8	Do you think generating visualization using this tool, saves time and human efforts (by skipping data formatting or coding or doing analysis part of the visualization)?	Complacency

9	Was forecasting of time series data up-to your expectations?	Complacency
10	What feature do you think need to be upgraded in the future?	Future scope
11	Any new feature needs to be added in the future?	Future scope

The responses of the users were recorded, and the analysis of the feedback are as follows.

Question 1:

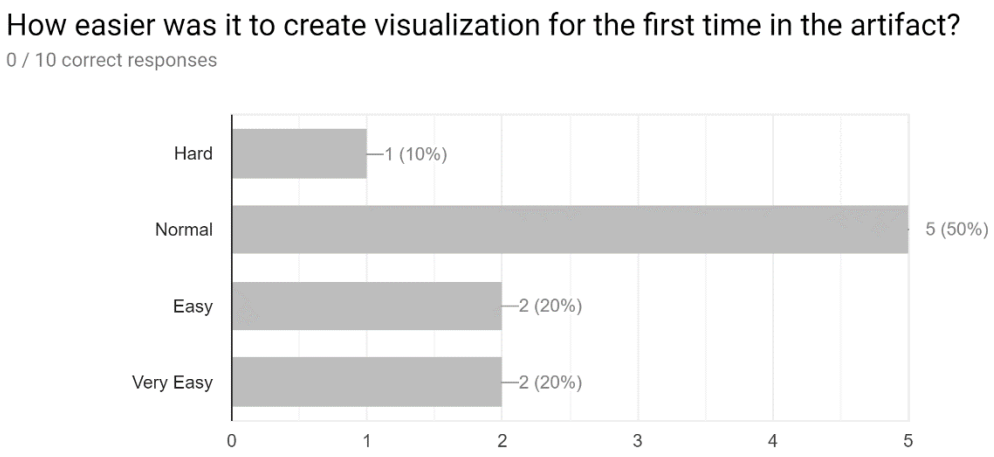


Figure 22: Question 1 from the User Evaluation Form(source: Author)

The first question evaluates the main role of artefact, which is visualization of a data. The user creates a query from the data present and visualize it in various format (i.e. bar chart, column chart, Pie chat, etc.). Most of the participants were able to generate visualization simple. Another major part of the uses found it easy to create the visualization.

Question 2:

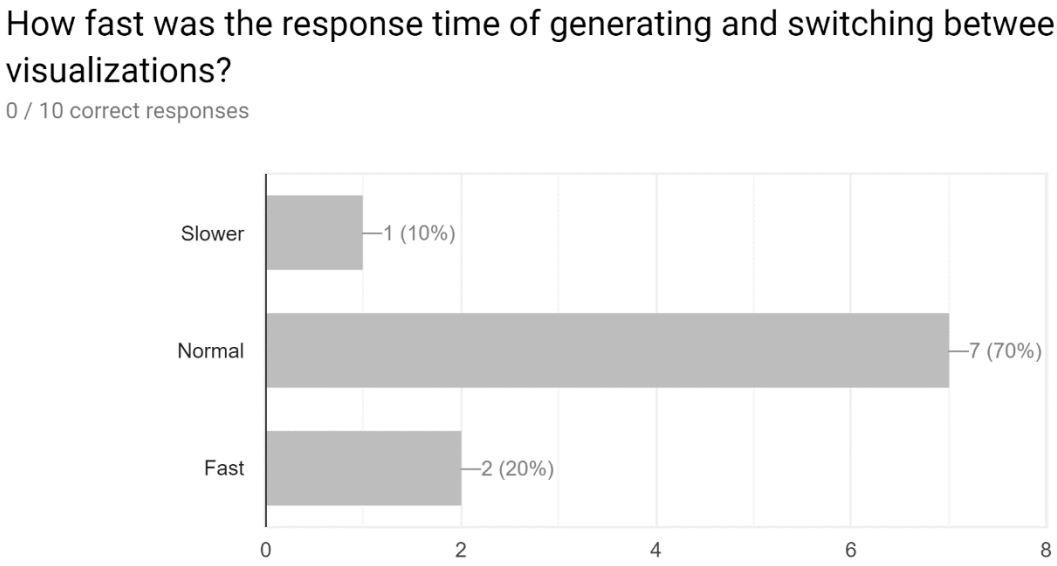


Figure 23: Question 2 from the User Evaluation Form(source: Author)

The second question focused on the response time of website in generating and switching between visualizations. The response time can be affected by many factors like internet speed, server capacity, server load and the device capacity. 77 % of the users found it normal and 11 % of the users found it faster than normal websites.

Question 3:

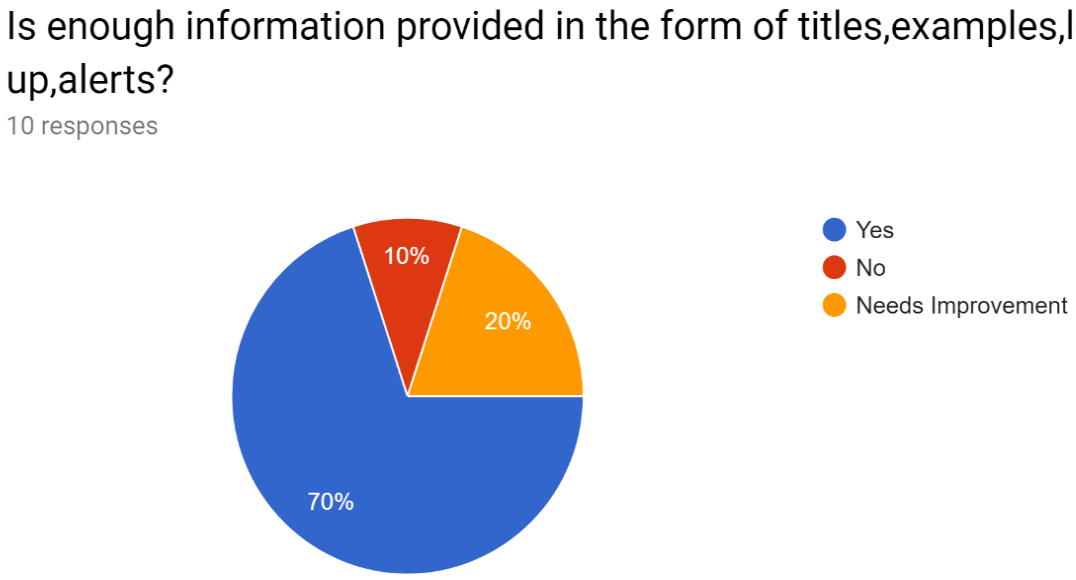


Figure 24: Question 3 from the User Evaluation Form(source: Author)

This question focus on the design part of the artefact, being self-explanatory is one of the critical aspects of a website. That being said almost 70% of the participants found it adequate enough to use the website without out any training or handbook and 22 % of them think there is a space for improvement.

Question 4:

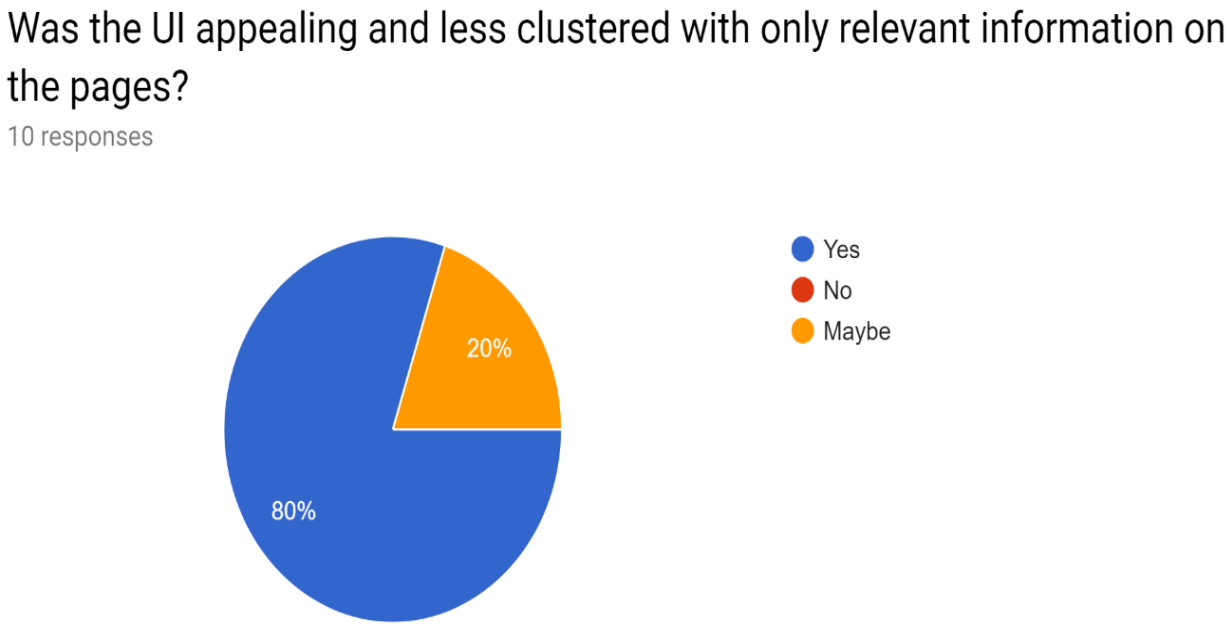


Figure 25: Question 4 from the User Evaluation Form (source: Author)

The Forth question focus on the presentation of the website, which keeps the user more engaged and less irritated or frustrated. More than 80% agrees that the UI is more appealing and presentable and remaining people still agrees but suggest more improvement.

Question 5:

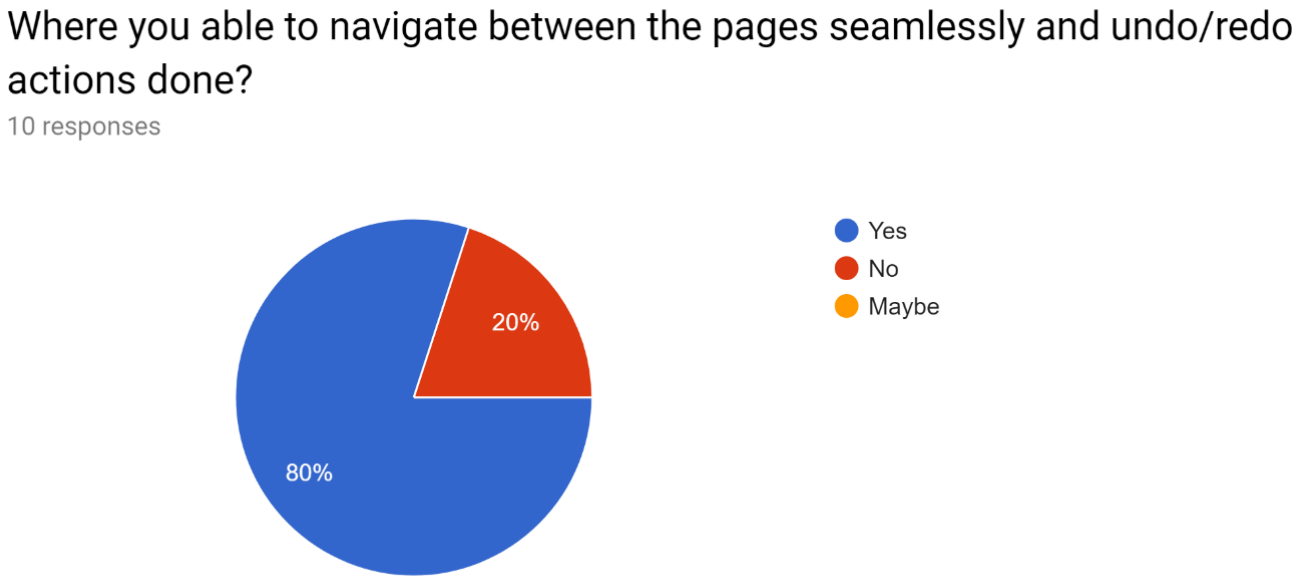


Figure 26: Question 5from the User Evaluation Form (source: Author)

This question focus on the accessibility of the artefact. A good software should always give the user the sense of control over it, the more the restriction the more user gets irritated. Also, users always tend to do mistake like clicking wrongly, misspelling, etc, the option to redo or undo should be there for the user. 80 % of the users found the navigation seamless and felt the control over the product.

Question 6:

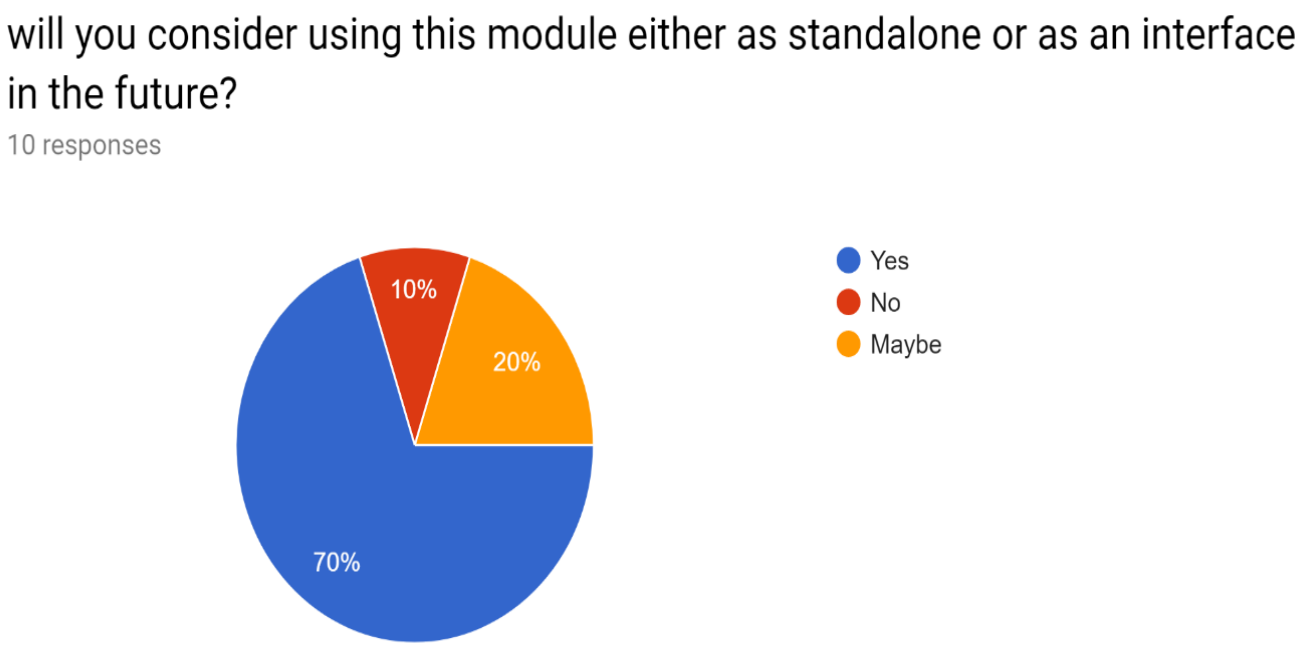


Figure 27: Question 6 from the User Evaluation Form (source: Author)

As the artefact were made in merely less than 3 months, the user was asked for a suggestion whether they would like to use a better version of this artefact in the future either as single system or as an interface to any of the existing system. 70% of the users gave a positive feedback, stating they are interested in the future use, 20% gave a neutral feedback and the remaining were not interested to use it in the future.

Question 7:

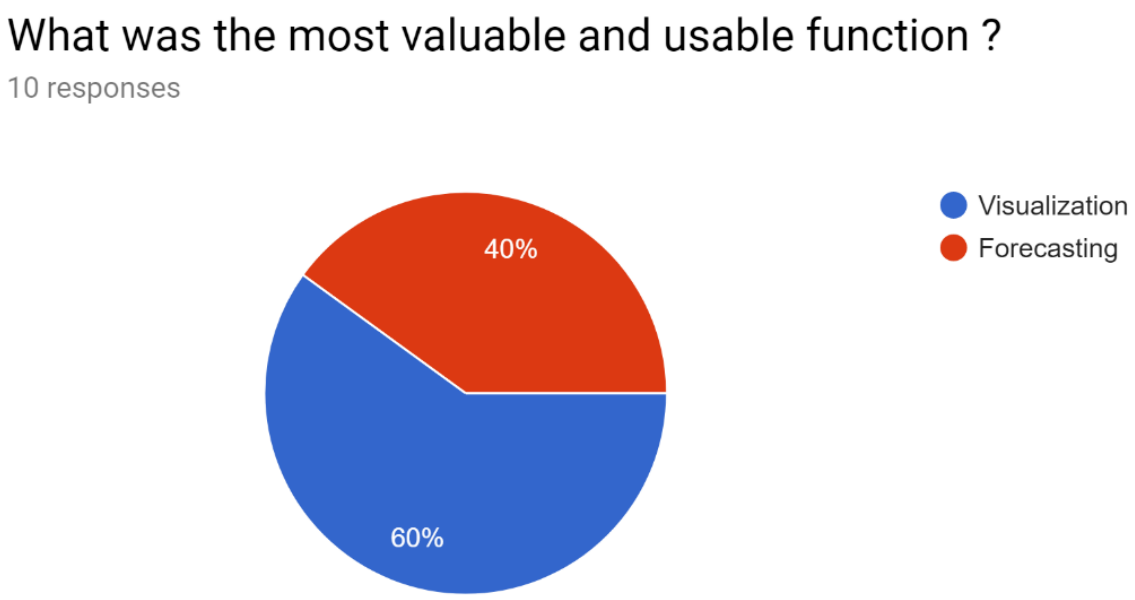


Figure 28: Question 7 from the User Evaluation Form (source: Author)

The artefact had two major functions visualization and forecasting. As per the users most of them were interested in automating visualization, then automating forecast in the ratio of 3:2.

Question 8:

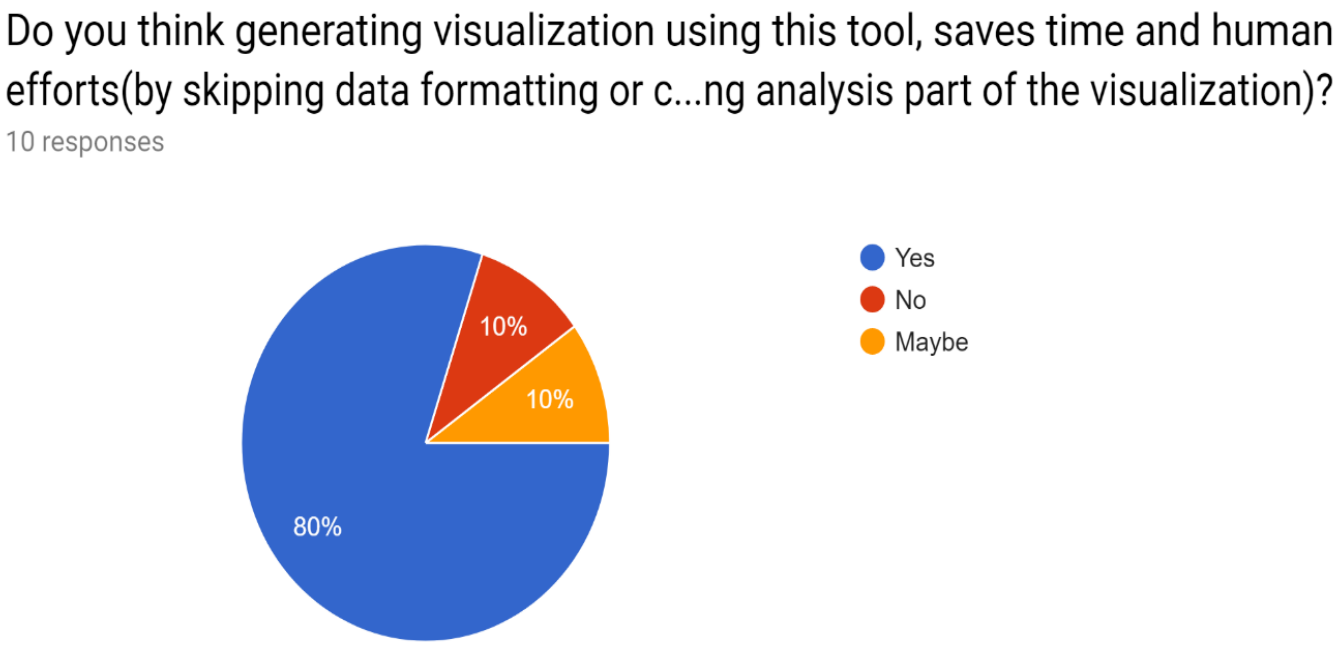


Figure 29: Question 8 from the User Evaluation Form (source: Author)

One of the main purposes of automating visualization is to save time and effort and avoid mistakes. Thus, this questions answers whether it is important and is it implemented correctly in this artefact. More than 80% thinks it is really saving human efforts and time by visualization easy and automatic and equal amount of users think neutral and others think automation doesn't fits the description of visualization.

Question 9:

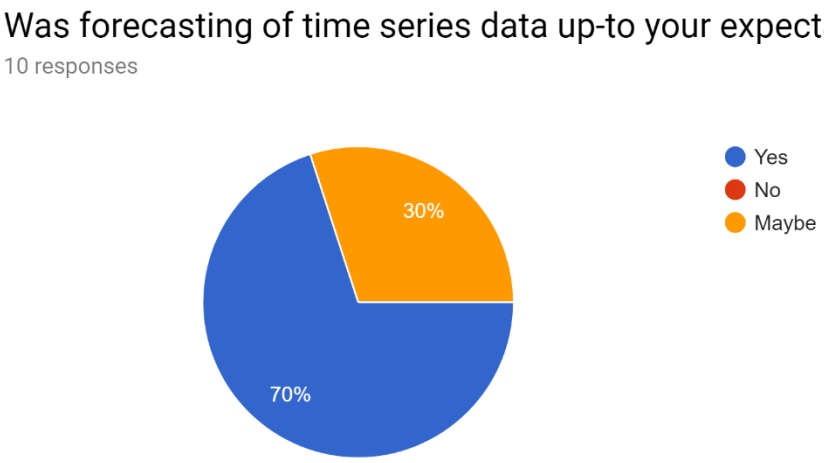


Figure 30: Question 9 from the User Evaluation Form (source: Author)

The Next main objective of the research is forecasting,70% of the user agree that the forecasting was up to their expectations.

Question 10:

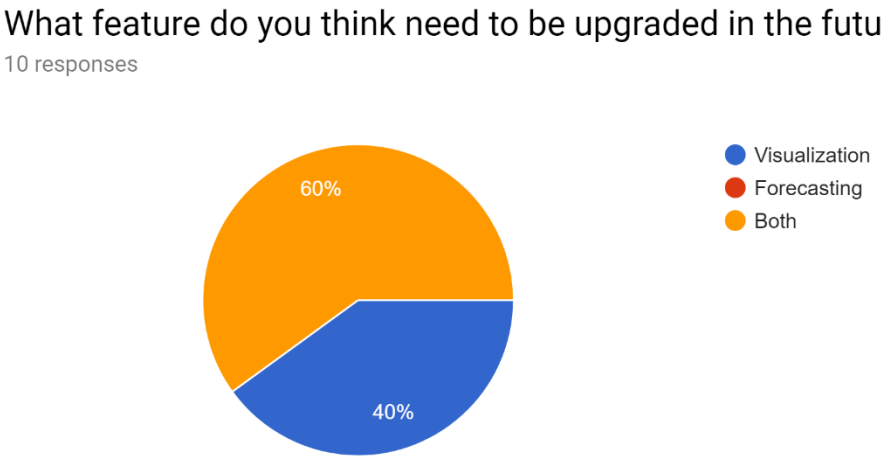


Figure 31: Question 10 from the User Evaluation Form (source: Author)

This question focus on the future of the research, most of the users want to see advance functions of both the main functions ,they choose both the requirements rather than any one, which means that there are more scope and need for a software which combines both analysis and visualization in the field.

Question 11:

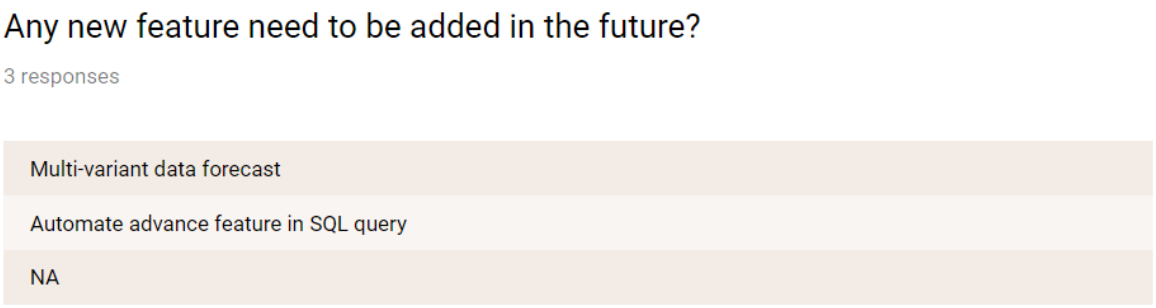


Figure 32: Question 11 from the User Evaluation Form (source: Author)

This is an open question, which also focus on future. A feature which is currently not present but needed most in the near future. Even though the researcher didn't get many responses, some of the responses which was collected are

- Multivariant data forecasting
- Advance feature of SQL

6. Research Discussions

In this chapter, the researcher is going to discuss about the requirement and objective that were stated in the first chapter and scrutinize the work done in order to achieve them. The objective of this research is to develop an easy and reliable system for forecasting and visualization; thus, the research question will also be critically evaluated against the proposed system to find out whether the requirements are satisfied. This chapter also focus on the recommendations for further improvement of the research, along with limitation of this study.

6.1 Research Conclusions

The main objective of conducting this research was to develop a system to automate visualization and forecasting, in order to help users with none or limited programming knowledge to visualize and analyse data. The questions asked in the user evaluation form were framed in a very orderly fashion for analysing a developed system. The focus group consist of developers and managers with less programming knowledge but in need for understanding various data. The questions are set in a group focusing on each and every feature in detail and also, they are started with basics and then went into details.

The first 36.36 % of the questions focus on the overall function and performance of the system. The researcher framed the question to know whether the research requirements are achieved. Especially the correctness and whether the system was able to save time and effort for generating visualization. More than 90 % of the users agrees that it is usefull and fullfilling the reqiurement of the research and approximatley 77 % of the users thinks the system has good response time and easy to operate with minimum or no training.Next 27 % of the questions were about the UI of the system which collected feedback whether the

design of the system is self-explanatory and presentable with enough information on it. 80% of the users found the UI is clear, pleasant and the website had information required for them to perform the given task without any hindrance. 70 % of the users agree that they are willing to use the advance version of the prototype in the future either as a completely separate system or as an interface with existing system. The last set of questions was to gather user's perceptiveness on which functionality needed more importance and what else they expect from the product.60 % of them gave importance to the upgrade of both the functionality and 40% of them choose visualization over prediction.

6.2Limitations of this Research

Visualization and prediction of data is a wide research application and there are countless ways to implement analysis and visualization, but currently our research scope is about developing a web-based prototype for automating visualization and time series prediction.

Some of the limitations of the research are.

- This application currently provides only 9 types of visualizations.
- To get more accurate LSTM prediction model, at least 35 years of month wise data are need.
- Multivariate Time series analysis cannot be done in this prototype.
- The visualizations are generic; thus, a user is not given suggestions on "what type of visualization is suitable for the data".

7. Conclusion

The researcher was successfully able to develop a system for automating visualization and prediction, which can be accessed via internet and gives full liberty to the user for creating new visualization and prediction on both mobile and desktop devices. The usage of Spring Boot made the application independent and as an interface for existing software's, considering the fact most of the web technologies are run using JAVA as their backend software. Usage of Hibernate ORM enables database reconfiguration with minimal or no configurations needed. Using deep learning techniques instead of machine learning and other linear techniques produce a better result, which was justified in the literature review and also better results were achieved with LSTM compared to other techniques like SARIMA, ARIMA, Fbphrophet, etc. Client-side rendering has many benefits over server side one of it is dynamic visualizations. Google charts are open source JavaScript's which are used in this prototype for visualization. Thus, combining all these technologies together the researcher was able to fulfil the requirements of the research. Also, user evaluation gave a positive feedback, at the same time suggesting about various grey areas that needs upgrading and future in-sights.

7.1 Future Works

The final prototype presented in this research was built in merely 3 months' time, the researcher agrees that there is a lot of scope for future developments. Following are some of the valuable recommendations by both the researcher and the users involved in the user evaluation.

- Enabling UI for more advance SQL functions.
- Multivariate data analysis.
- Update of Visualization Query.

References

1. Scacchi, Walt. (2002) Process Models in Software Engineering.
2. Adel Alshamrani and Abdullah Bahattab. (2015) A Comparison Between Three SDLC Models Waterfall Model, Spiral Model, and Incremental/Iterative Model. IJCSI International Journal of Computer Science Issues
3. Kamilya Smagulova. (2019) A survey on LSTM memristive neural network architectures and applications. The European Physical Journal Special Topics
4. Sommerville. (2011) SOFTWARE ENGINEERING. 9th edn. Pearson Education Limited
5. Kothari, C. R. (2004) Research methodology. Methods and techniques: New Age
6. International
7. Mark Saunders,Philip Lewis,Adrian Thornhill.(2015) Research Methods for Business Students.5th edn. Pearson Education Limited
8. Ragab, M., Arisha, A. (2017) Research Methodology in Business: A Starter s Guide. Vol. 5, No. 1.Management and Organizational Studies.
9. De Vaus, D. A. (2002) Surveys in social research.5th edn. Psychology Press.
10. Gerhard Greeff and Ranjan Ghoshal. (2004) Practical E-Manufacturing and Supply Chain Management.
11. Nason, GP. (2006) Stationary and non-stationary time series. in H Mader & SC Coles (eds). Statistics in Volcanology. Geological Society of London, pp. 129 - 142.
12. Ratnadip Adhikari, and R. K. Agrawal. (2013) An Introductory Study on Time Series Modeling and Forecasting. LAP Lambert Academic Publishing.

13. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. (2015) Time Series Analysis: Forecasting and Control. 5th edn Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712.
14. Juang, W. C., Huang, S. J., Huang, F. D., Cheng, P. W., & Wann, S. R. (2017). Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. BMJ open, 7(11).
15. Shengwei Wang,Juan Feng,Gang Liua. (2013) Application of seasonal time series model in the precipitation forecast. Mathematical and Computer Modelling, Pages 677-683.
16. Rana, O. et al. (2000) PaDDMAS: parallel and distributed data mining application suite , in Proceedings 14th International Parallel and Distributed Processing Symposium. IPDPS 2000. Proceedings 14th International Parallel and Distributed Processing Symposium. IPDPS 2000, pp. 387–392.
17. Roshdy, A. et al. (2018) Generic Data Visualization Platform , in 2018 22nd International Conference Information Visualisation (IV). 2018 22nd International Conference Information Visualisation (IV), pp. 56–57.
18. James, J. A., Moh, T.-S. and Edwards, C. A. (2016) Web-Based Visualization of Marine Environmental Data: Performance Analysis of a MatPlotLib Implementation , in 2016 International Conference on Collaboration Technologies and Systems (CTS). 2016 International Conference on Collaboration Technologies and Systems (CTS), Orlando, FL, USA: IEEE, pp. 288–293.
19. Zhu, N. (2017) Data Visualization with D3 4.x Cookbook. Birmingham, Packt Publishing.
20. Franzblau, L. E. and Chung, K. C. (2012) Graphs, Tables, and Figures in Scientific Publications: The Good, the Bad, and How Not to Be the Latter , The Journal of Hand Surgery, 37(3), pp. 591–596.
21. Kirk, A. (2016) Data Visualisation: A Handbook for Data Driven Design. Thousand Oaks.
22. Valipour, M. (2015) Long-term runoff study using SARIMA and ARIMA models in the United States. Meteorological Applications, 22(3), 592–598.

23. Tianhong Pan and Shaoyuan Li. (2006) Time Series Prediction Based on Lazy Learning. 6th World Congress on Intelligent Control and Automation, Dalian, pp. 6039-6042.

24. Q. Xiao and Y. Si. (2017) Time series prediction using graph model. 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, pp. 1358-1361.

25. Koulouzis, S., Zudilova-Seinstra, E. and Belloum, A. (2010) Data transport between visualization web services for medical image analysis , Procedia Computer Science. (ICCS 2010), 1(1), pp. 1727–1736.

26. Abeel, T., Peer, Y. V. de and Saeys, Y. (2009) Java-ML: A Machine Learning Library , Journal of Machine Learning Research, 10(Apr), pp. 931–934.

27. X. Song, J. Huang and D. Song. (2019) Air Quality Prediction based on LSTM-Kalman Model. IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 695-699.

28. Bender, M. et al. (2000) A functional framework for Web-based information visualization systems. IEEE Transactions on Visualization and Computer Graphics, 6(1), pp. 8–23.

29. Data Cleaning: Overview and Emerging Challenges (no date) AMPLab - UC Berkeley. Available at: https://amplab.cs.berkeley.edu/publication/data-cleaning-overview-and-emerging-challenges/ (Accessed: 9 August 2019).

30. doi: 10.1109/MCG.2012.114.

31. Hinderman, B. (2016) Building Responsive Data Visualization for the Web. Indianapolis, Wiley.

32. Social Research Methods - Knowledge Base - Deduction & Induction (no date a). Available at: http://www.socialresearchmethods.net/kb/dedind.htm (Accessed: 10 August 2019).

33. Social Research Methods - Knowledge Base - Deduction & Induction (no date b). Available at: http://www.socialresearchmethods.net/kb/dedind.htm (Accessed: 10 August 2019).

34. Wojciechowski, J. et al. (2004) MVC model, struts framework and file upload issues in

web applications based on J2EE platform , in Proceedings of the International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Proceedings of the International Conference Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004., pp. 342–345.

35. Scott MurrayAugust. (2017) Interactive Data Visualization for the Web: An Introduction to Designing with, Part 3, Edition 2, O'Reilly Media, Inc.

36. Slutsky, D. J. (2014) The Effective Use of Graphs , Journal of Wrist Surgery, 3(2), pp. 67–68.

37. Zhang, D., Wei, Z. and Yang, Y. (2013) Research on Lightweight MVC Framework Based on Spring MVC and Mybatis , in 2013 Sixth International Symposium on Computational Intelligence and Design. 2013 Sixth International Symposium on Computational Intelligence and Design, pp. 350–353.

38. Zhu, Y. (2012) Introducing Google Chart Tools and Google Maps API in Data Visualization Courses , IEEE Computer Graphics and Applications, 32(6), pp. 6–9.

39. Liu, J. et al. (2018) A Survey of Scholarly Data Visualization , IEEE Access, 6, pp. 19205–19221. doi: 10.1109/ACCESS.2018.2815030.

40. Mwalongo, F. et al. (2016) State-of-the-Art Report in Web-based Visualization , Computer Graphics Forum, 35(3), pp. 553–575.

41. Kumar, V. and Khosla, C. (2018) Data Cleaning-A Thorough Analysis and Survey on Unstructured Data , in 2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence). 2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 305–309.

42. Ostruszka, M., Sakowicz, B. and Napieralski, A. (2007) Universal Web-Based Charts Generator based on J2EE Platform , in 2007 9th International Conference - The Experience of Designing and Applications of CAD Systems in Microelectronics. 2007 9th International Conference - The Experience of Designing and Applications of CAD Systems in Microelectronics, pp. 524–528.

43. R. Akase and Y. Okada. (2015) Web Based Time-Tunnel: An Interactive Multidimensional Data Visualization Tool Using Genetic Algorithm. 19th International Conference on Information Visualisation, Barcelona, 2015, pp. 95-100.

APPENDIX

Appendix A: Case Studies

Below are some of the case studies for the dissertation. Some case studies which are marked for developer needs knowledge on programming background . You are required to select one of the below case studies and implement it on the prototype.

Common Instructions to All the Case study is

- Go to the link “”
- Navigate to Add Query Page
- Create a Visualization Query With you case study number as title

Case Study 1

- Create a visualization for Oil data with the following field-> Date, Price and No of barrels and enable prediction.
- Then Navigate to List Page and then view all the visualizations generated for your Query.
- Then view all visualization generated for the predicted values.

Case Study 2

- Create a visualization for Oil data with the following field-> Date, Price and No of barrels and disable prediction.
- Then Navigate to List Page and then view all the visualizations generated for your Query.

Case Study 3 (For Developers)

- Create a visualization for USA motor and retail data with the following field-> year, retail inventory/sales ratio and motor inventory /sales ratio
- Using Developer module group them by year(SQL Functions to be used:Sum() of fields and group by year)
- Then Navigate to List Page and then view all the visualizations generated for your Query.

Appendix B: Information Form and Consent Sheet

INFORMATION SHEET FOR PARTICIPANTS

PROJECT TITLE: Generic Web Console for Visualization and Time Series Forecasting

You are being asked to take part in a research study on a generic web-based visualization and prediction tool, which automates both visualization and prediction as much as possible. Client-side visualization is generated using google charts and spring MVC framework as background and python for prediction of time series data using recursive neural network is implemented. The prototype can be accessed from via LAN when your own devices when connected to it like a normal website. None of the personal or device information will be collected via it. The research aims to make visualization and prediction hassle free and easily with more accessibility. I am a student of Dublin Business School pursuing my master's in science of Information Systems and Computing, student ID 10506917. My dissertation supervisor is Mr. Obinna Izima.

WHAT WILL HAPPEN

In this study, you will be asked to test the artefact which was developed for this dissertation and then a feedback will be collected via a survey for accessing the artefact.

TIME COMMITMENT

The study typically takes 5-10 minutes

PARTICIPANTS' RIGHTS

You may decide to stop being a part of the research study at any time without explanation required from you. You have the right to ask that any data you have supplied to that point be withdrawn / destroyed. You have the right to omit or refuse to answer or respond to any question that is asked of you. You have the right to have your questions about the procedures answered (unless answering these questions would interfere with the study's outcome. A full

de-briefing will be given after the study). If you have any questions as a result of reading this information sheet, you should ask the researcher before the study begins.

CONFIDENTIALITY/ANONYMITY

The data I collect does not contain any personal information. The collected information is solely for the purpose of the research and will not be shared with anyone or put up anywhere.

FOR FURTHER INFORMATION

I or / and Mr. Obinna Izima will be glad to answer your questions about this study at any time. You may contact my supervisor at obinna.izima@dbs.ie

Appendix C: Consent form

INFORMED CONSENT FORM

Generic Web Console for Visualization and Time Series Forecasting

PROJECT SUMMARY:

My research focus on automating visualization using client-side rendering and forecast time series data.in order to increase the accessibility and availability, the tool build as a web-based system to be used as an interface with another project or standalone. The research tends to reduce the work and time taken to understand data and produce visualization. Forecasting data not only gives insights but also a concrete data which can help in future decision and risk management.

By signing below, you are agreeing that: (1) you have read and understood the Participant Information Sheet, (2) questions about your participation in this study have been answered satisfactorily, (3) you are aware of the potential risks (if any), and (4) you are taking part in this research study voluntarily (without coercion).

_____ _____

Participant's signature Participant's Name (Printed)

_____ _____

Student Name (Printed) Student Name signature

_____

Date