Big Data & IoT

Data Visualization

2017

Oh Soon Kit

14-Jun-17

Table of Contents

[Chapter 1 : Introduction 5](#_Toc485220552)

[1.1 Project Management 5](#_Toc485220553)

[Chapter 2 : Methodology 6](#_Toc485220554)

[2.1 FlowChart 6](#_Toc485220555)

[2.1.1 Description 6](#_Toc485220556)

[Chapter 3 : Implementation Plan 7](#_Toc485220557)

[3.1 Screenshot 7](#_Toc485220558)

[3.1.1 Screenshot for Graph of AGV Mileage Vs Timestamp 7](#_Toc485220559)

[3.1.2 Screenshot for Graph of Board Temperature Vs Timestamp 7](#_Toc485220560)

[3.1.3 Screenshot for Graph of Laser Sensor Output Vs Timestamp 8](#_Toc485220561)

[3.1.3 Screenshot for Graph of Battery Output Vs Timestamp 8](#_Toc485220562)

[3.2 Code 9](#_Toc485220563)

[3.2.1 Code for Time Selection 9](#_Toc485220564)

[3.2.1.1 Description 9](#_Toc485220565)

[3.2.2 Code for Drag and Zoom 10](#_Toc485220566)

[**3.2.2**.1 Description 10](#_Toc485220567)

[3.2.3 Code for Graph Display 11](#_Toc485220568)

[3.2.3.1 Description Code 11](#_Toc485220569)

[3.2.4 Code for URL redirection 12](#_Toc485220570)

[3.2.4.1 Description 12](#_Toc485220571)

[3.2.5 Grid Function 13](#_Toc485220572)

[3.2.5.1 Description 13](#_Toc485220573)

[3.2.6 ToolTip 14](#_Toc485220574)

[3.2.6.1 Description 14](#_Toc485220575)

[3.2.7 Text for Title, x-axis and y-axis. 15](#_Toc485220576)

[3.2.7.1 Description 15](#_Toc485220577)

[3.3 Software Development Process 16](#_Toc485220578)

[3.3.1 Prototyping 16](#_Toc485220579)

[3.4 Challenges Faced 17](#_Toc485220580)

[3.4.1 Scalability 17](#_Toc485220581)

[3.4.2 Hardware 17](#_Toc485220582)

[3.4.3 Communication with Task distributor 17](#_Toc485220583)

[3.4.4 Evaluation 17](#_Toc485220584)

[3.5 Limitation 18](#_Toc485220585)

[3.5.1 Data Storing Issue 18](#_Toc485220586)

[3.5.2 Partial Real Time Graph 18](#_Toc485220587)

[3.6 Future Work 18](#_Toc485220588)

[3.6.1 Expand data storage 18](#_Toc485220589)

[3.6.2 Enhance Interface of data visualization 18](#_Toc485220590)

[3.6.3 Completion of real time streaming data 18](#_Toc485220591)

[Extra 19](#_Toc485220592)

[I. SAS Software 19](#_Toc485220593)

[II. Login Details 19](#_Toc485220594)

[III. Visualization 19](#_Toc485220595)

**List of Figure**

Figure 1 : Graph of AGV Mileage Versus Timestamp 7

Figure 2 : Graph of Board Temperature Versus Timestamp 7

Figure 3 : Graph of Laser Sensor Output Versus Timestamp 8

Figure 4 : Graph of Battery Current Versus Timestamp 8

Figure 5 : Code for Time Selection 9

Figure 6 : Code for Drag and Zoom Function 10

Figure 7: Code for Graph Display 11

Figure 8 : Code for URLs redirection 12

Figure 9 : Code for Grid Function 13

Figure 10 : Code for Tooltip 14

Figure 11 Code for Text for title, x-axis, and y-axis 15

Figure 12 : Prototyping Model (Intro-To-SDLC, 2017) 16

Figure 13 : Screen for import data 19

Figure 14 : Screen for Drag and Drop 20

Figure 15 : Screen for output 20

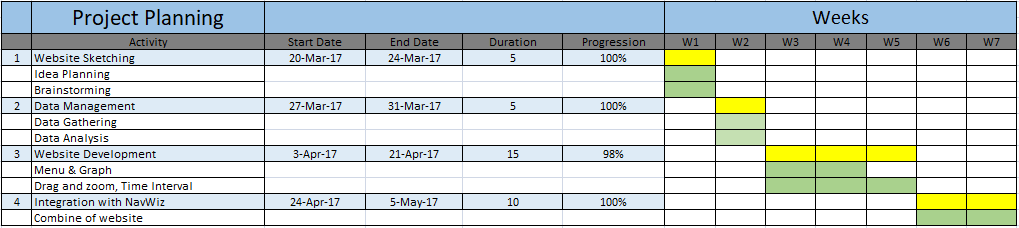
Figure 16 : Screen for Report 21

# Chapter 1 : Introduction

Data visualization is a method to help people to understand the meaning of data in visual context form. Aside from that, the primary goal of data visualization is to communicate information clearly and efficiently via statistical graphics, plots and information graphics. Numeric data may be encoded using dots, limes, or bars, to visually communicate a quantitative message. Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable.

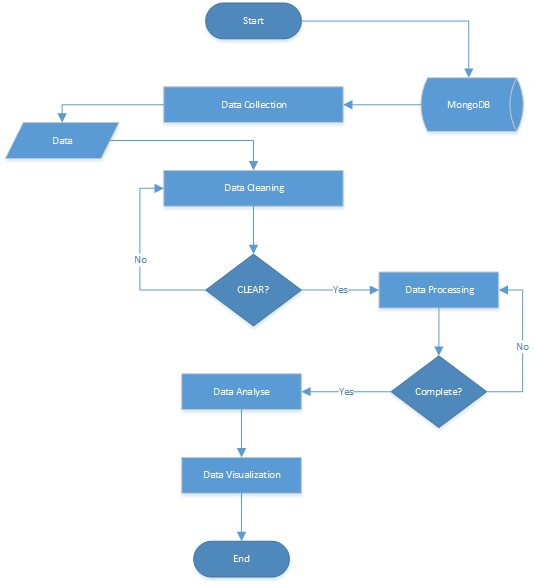
## 1.1 Project Management

*Gantt Chart*



# Chapter 2 : Methodology

## 2.1 FlowChart



## 2.1.1 Description

First step is to extract data from MongoDB, after the data is extracted from MongoDB then proceed to data cleaning step, Data cleaning step involved in eliminating the noise data (aka meaningless, corrupted, unstructured data and missing value). If the data cleaning part is done successfully, the data will be proceeding to data processing step else it will redo again. At the data processing step, the information will be classified to its according category. When the necessary step is done, then the data will proceed to data analyze step. In the data analyze step, the data will be analyzed and decided to present in an ideal way. After that, the analyzed data will be visualized in graph form.

# Chapter 3 : Implementation Plan

## 3.1 Screenshot

## 3.1.1 Screenshot for Graph of AGV Mileage Vs Timestamp

## 2.png

Figure : Graph of AGV Mileage Versus Timestamp

## 

## 3.1.2 Screenshot for Graph of Board Temperature Vs Timestamp

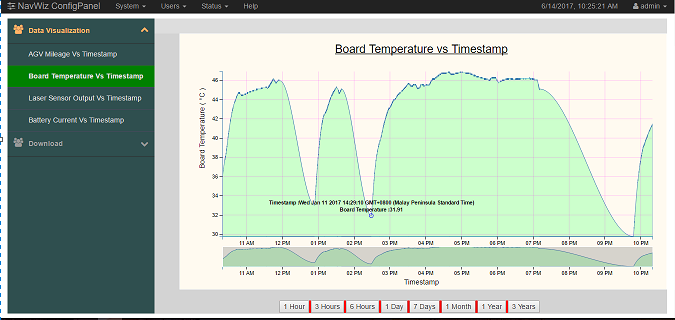
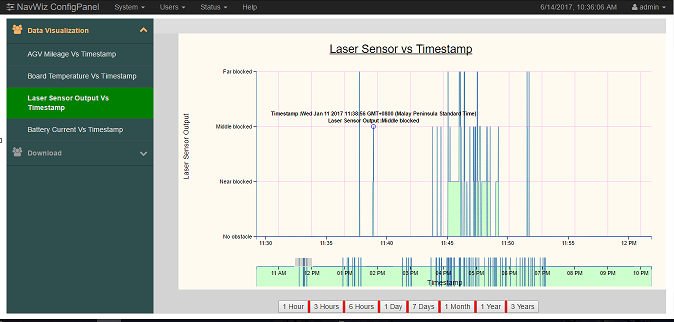


Figure : Graph of Board Temperature Versus Timestamp

## 3.1.3 Screenshot for Graph of Laser Sensor Output Vs Timestamp

 Figure : Graph of Laser Sensor Output Versus Timestamp

## 3.1.3 Screenshot for Graph of Battery Output Vs Timestamp

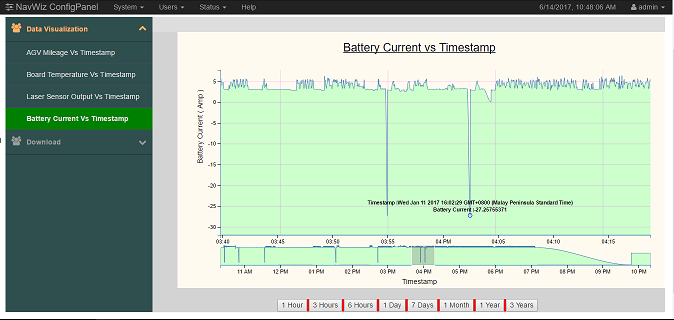


Figure : Graph of Battery Current Versus Timestamp

## 3.2 Code

## 3.2.1 Code for Time Selection

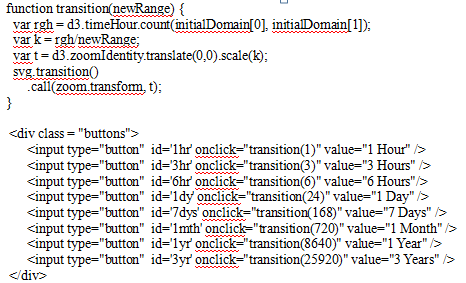


Figure : Code for Time Selection

## 3.2.1.1 Description

We are trying to find the scale factor k to apply to transform. The original range of rgh is hours. We find k dividing rgh by newRange and perform transformation on the graph. For the calculation of transition, it is measured by division of hour, therefore, all the ‘day’, ‘month’ and ‘year’ format must be converted to hour for transition function to measure. The formula is as listed below:

Day = number of day \* hour of the day (24hours)

Month = number of month \* day of the month (28/29/30/31 days) \* hour of the day

Year = number of year \* month of the year \* day of the month \* hour of the day

## 3.2.2 Code for Drag and Zoom

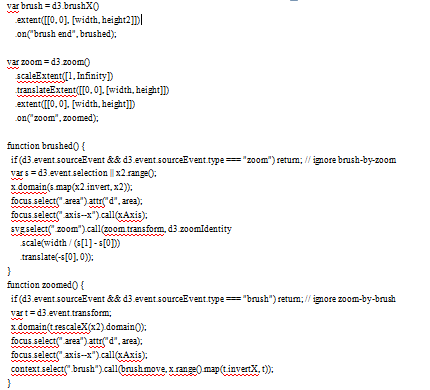


Figure : Code for Drag and Zoom Function

## **3.2.2**.1 Description

There are brushed and zoomed function that used to perform drag and zoom operation. For the brushed function, s is the selection on the x2.range and x-domain is mapped to the inversion of the selected scale of x2.range. Area and the x-axis of the graph are focused and called, the svg is zoomed and translated based on the scale (division of total width by the range of s). While for the zoomed function, x-domain is zoomed based on the scale of x2 domain and the graph will zoom based on the movement of brush.

## 3.2.3 Code for Graph Display

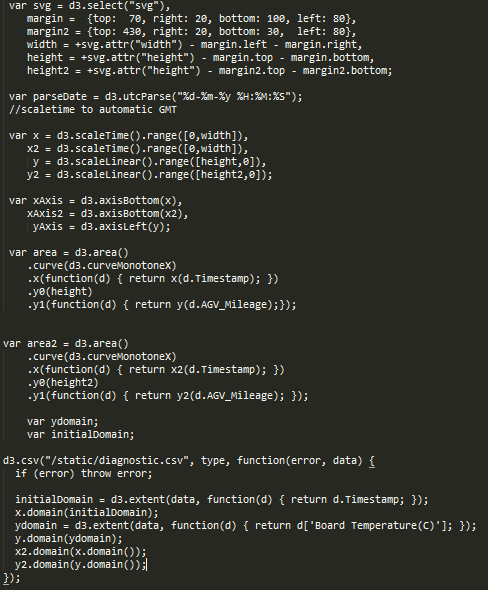


Figure : Code for Graph Display

## 3.2.3.1 Description Code

First, set margin, width and height for big and small graph, followed by set the timezone to GMT+0800, time scale for x-axis and linear scale for y-axis. D3.csv function is to read the csv file and domain function is to assign the desired column to x/y.domain.

\*\* the d3.csv here is to read the static csv file\*\*

\*\* for real time one please refer to the code

For instance : d3.csv(“{% url 'system:data' %}”) , type, function(error, data){

## 3.2.4 Code for URL redirection

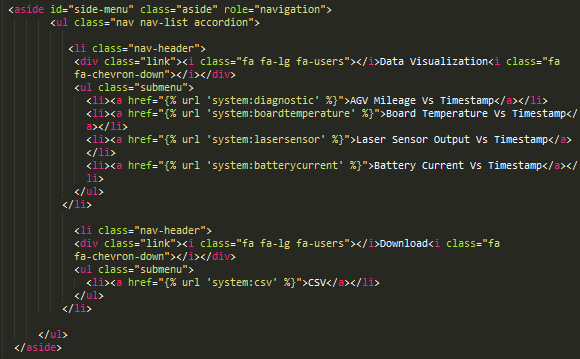


Figure : Code for URLs redirection

## 3.2.4.1 Description

The href function is used to link user to their selection of url to the specific page.

## 3.2.5 Grid Function

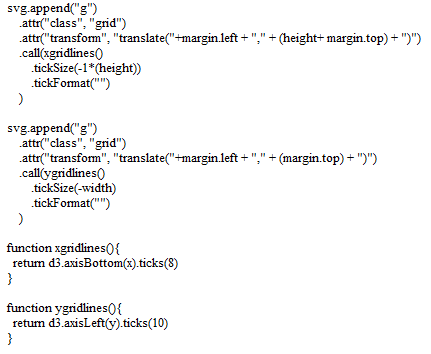


Figure : Code for Grid Function

## 3.2.5.1 Description

The svg append is to set the gridline for both x and y axis while the function x/ygridlines is to adjust the number of lines (ticks(numeric value)).

## 3.2.6 ToolTip

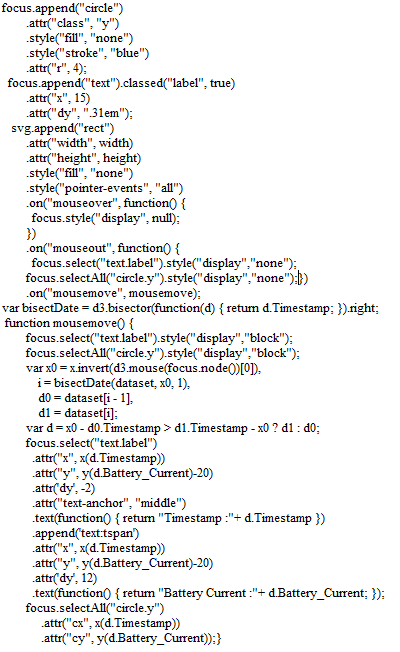


Figure : Code for Tooltip

## 3.2.6.1 Description

The bisectDate function is defined and cx, cy attributes of the circle are set to make the circle move. The mousemove function is to align the data and the timestamp, the focus append function is to set the attributes for the circle and text label, and the svg append is to execute the mousemove function.

## 3.2.7 Text for Title, x-axis and y-axis.

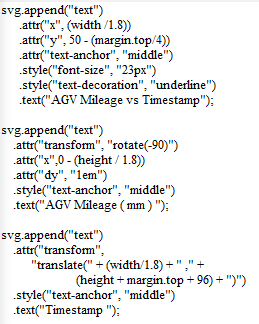


Figure Code for Text for title, x-axis, and y-axis

## 3.2.7.1 Description

The svg append text functions are to set the attributes for title, x-axis and y-axis. For example : the value of the text, the position of the text, font size of text and the style of the text.

## 3.3 Software Development Process

## 3.3.1 Prototyping

Figure : Prototyping Model

Prototyping model is used throughout the development of data visualization website. It is a model that is used to build a prototype which can be tested by the client and this prototype is then keep refining based on client’s requirement until an acceptable prototype is achieved.

There are few steps to be executed in designing a prototype. The first step is requirement gathering which the requirement gathered has to focus on the main function of the prototype while others requirement than main function can be ignored in this stage.

The next step is to design and develop the draft prototype. In this stage, User interface of prototype is first designed then followed by developing the prototype. The prototype should possess all the necessary requirements.

After the prototype is developed, the following step is the evaluation of the prototype. The prototype is to demonstrate to the client and other stakeholders that involved in the project. Client and stakeholders will give feedbacks for further improvement on the prototype.

The last step is to refine the prototype based on the feedbacks that collected from the client and stakeholders. If there is any feedback that is unachievable, developer will negotiate with the client and stakeholders until the feedbacks agreed by both parties. This cycle will be repeated until the prototype satisfies the requirement of the client.

In a nutshell, Prototyping model is quite popular nowadays and it is widely used by most of the software developer because of the interaction of developer and client to acquire better understanding on requirements of client. The feedback from the client helps to clarify what is needed for the product and the process of the prototype is visible to client that can help client in understanding the flow of prototype. This model reduces time and cost significantly due to the lessen work during maintenance phase as any missing or unambiguous function can be detected immediately.

## 3.4 Challenges Faced

## 3.4.1 Scalability

It is difficult to scale the visualization with both visual representations and analysis. For instance, there are some factors that are needed to be considered when scaling the visualization, such as continuous streaming of data input and pattern of the data.

## 3.4.2 Hardware

Powerful hardware are required to ensure the efficient process of high quality data (noiseless and have the minimal amount of missing values) and huge amount of data. For the huge amount of data, high processing hardware will greatly reduce the data loading time, hence it increased the efficiency. High resolution display is also required to ensure the better view of the visualization.

## 3.4.3 Communication with Task distributor

It is important to maintain a good communication between the developer and Task distributor. As bad communication between them will results in obtaining ambiguous information and eventually result in failure of the development. Aside from that, communicating with task distributor can also benefits the developer from knowing whether the assigned task is beyond their limit and made adjustment to the assigned task.

## 3.4.4 Evaluation

There are certain things to be considered such as the quality of the code, eliminate as much as possible for the hardcoded part, ensure common components of the code are reusable.

## 3.5 Limitation

There are few limitations remain unsolved at the moment due to insufficient time and resources.

## 3.5.1 Data Storing Issue

Currently the server can only store up to 1 day‘s data due the limited storing space for the data. The expected output will be storing the data up to years and display on the graph for visualization.

## 3.5.2 Partial Real Time Graph

Real Time Graph for back end can only run for 1 day’s data and for front end is currently facing issue to load the data.

## 3.6 Future Work

## 3.6.1 Expand data storage

To feed the data continuously from mongodb to python carbon graphite up to years.

\*\*Note\*\* update mongodb to version 3.0 or above as the version provided data compressing function that can reduce the data size and increase efficiency.

## 3.6.2 Enhance Interface of data visualization

To make the interface of data visualization looks nicer than the existing interface and to eliminate the visible ‘bug’ so that can attract the ‘eyeball’ of the clients.

## 3.6.3 Completion of real time streaming data

For the existing version of the real time data streaming graph, it can only stream up to one day data but for the future version of the real time data streaming graph, it must stream up to years of data.

# Extra

## I. SAS Software

SAS is known as statistical analysis system. SAS can gather and manage data from various resources and perform statistical analysis on it.

## II. Visualization

* Step 1 : Import Data

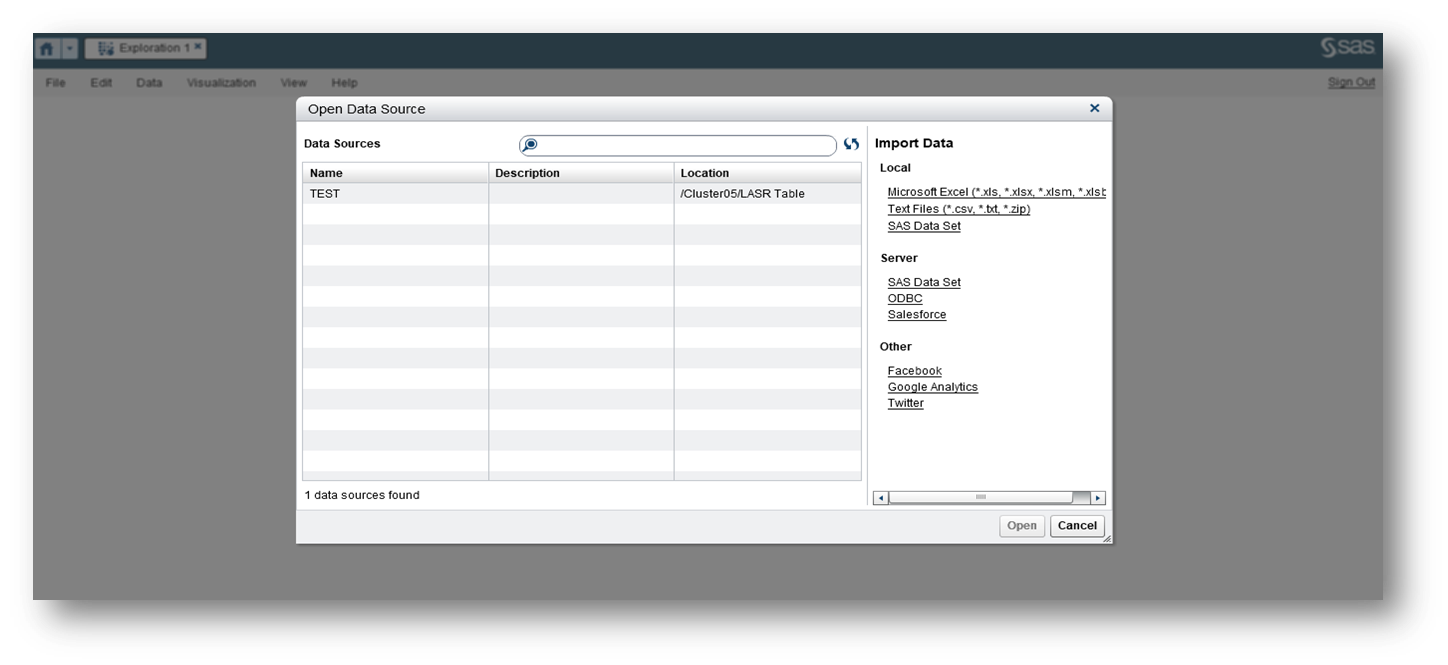


Figure : Screen for import data

First, Select ‘File’ from the SAS homepage and then choose ‘new visualization’ to

import the desired data to perform statistical analysis.

* Step 2 : Drag And Drop

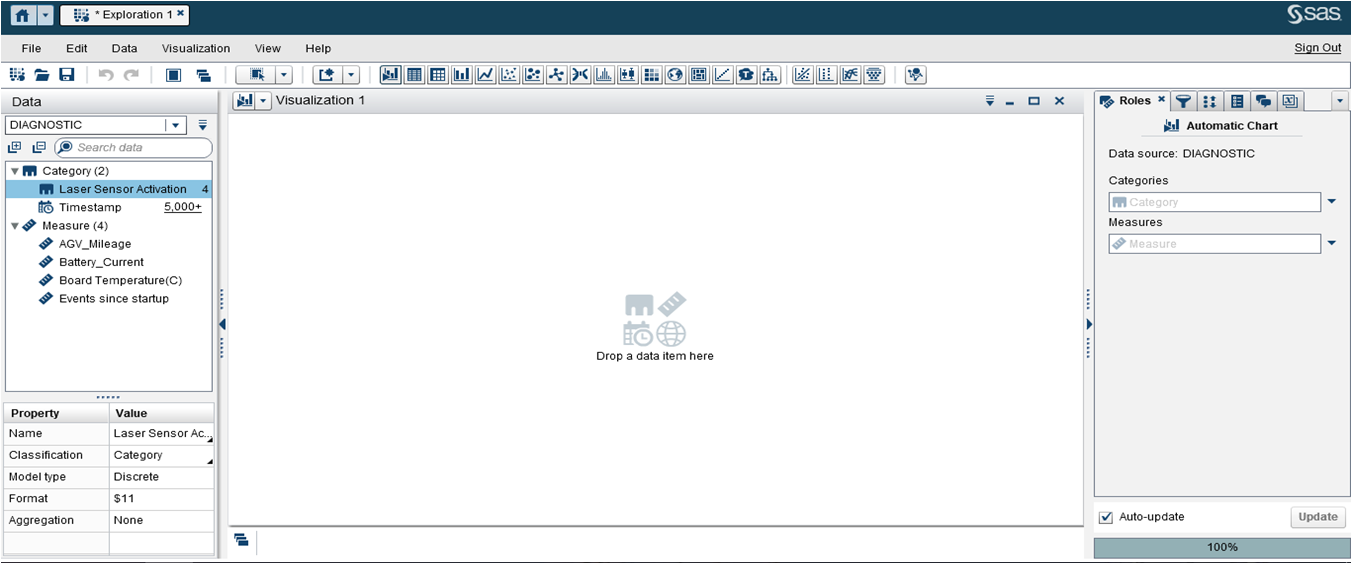


Figure : Screen for Drag and Drop

After imported the data, All the variables will be appeared on the left side of the page, Users are required to drag the variable and drop it on the x/y-axis. Categories represent x-axis while measures represent y-axis.

* Step 3 : Overview of output

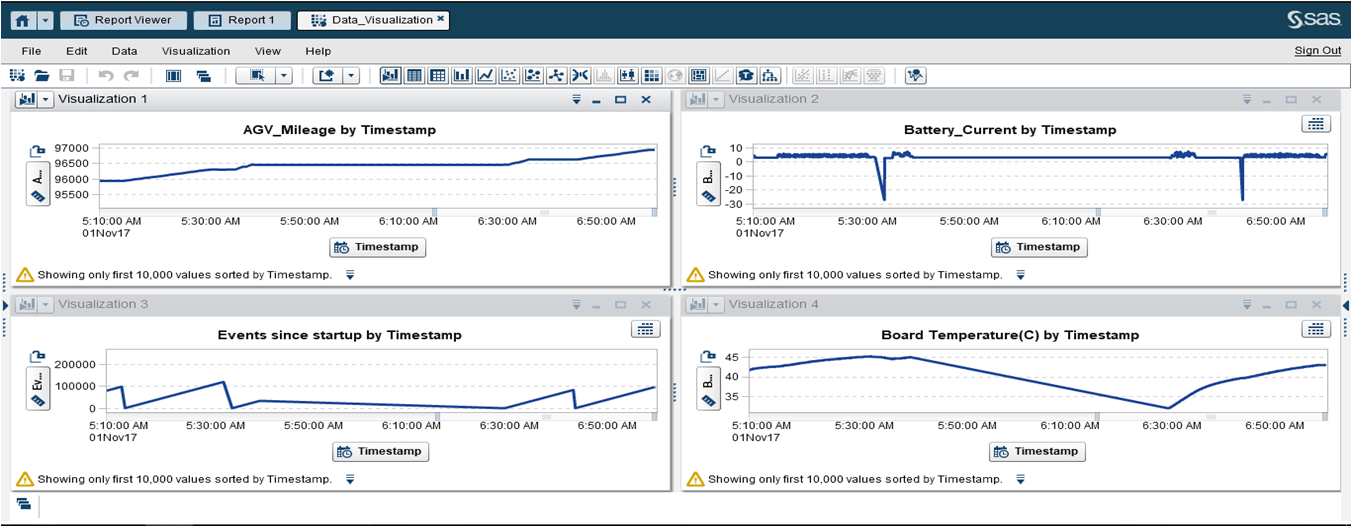


Figure : Screen for output

The output will be displayed on visualization box.

* Report Format

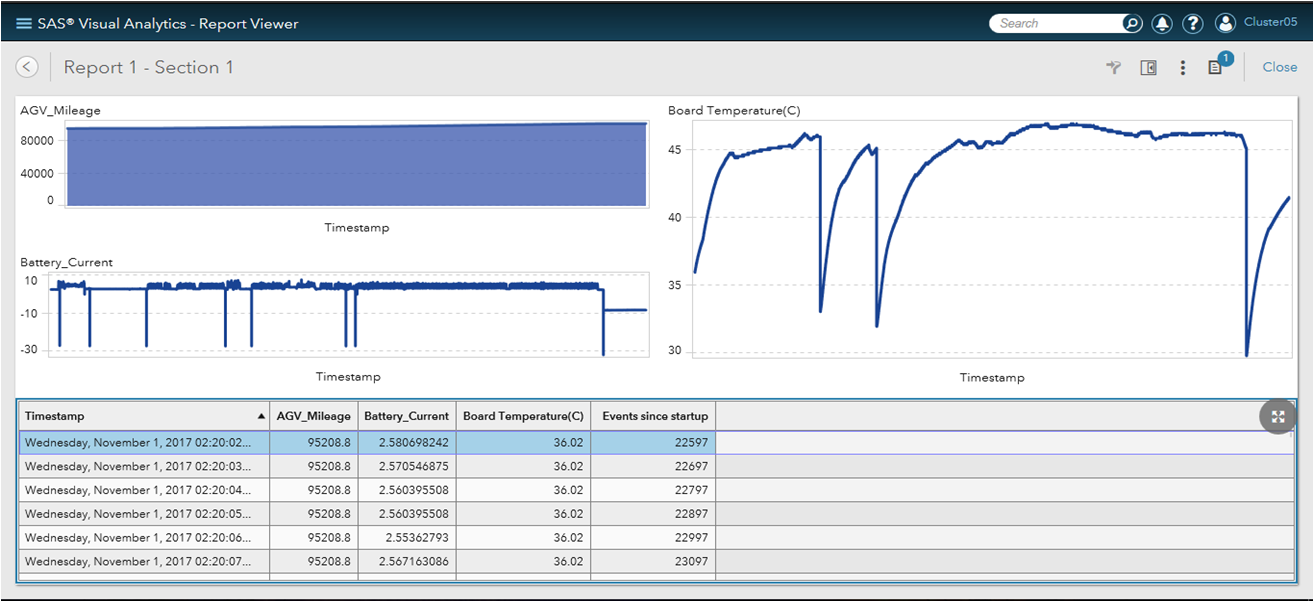


Figure : Screen for Report

Data Visualization can also be done in report format by selecting the Report Designer option on the select panel there. After finished plotting, users are required to click on the report viewer to view the full report. The report is created for the better view for the user.