

Visual Analytics

COMP 5048

Assignment 3 Final Report

GROUP 13

Team Members:

Bohua Zhang	450280418
Feng Wang	450518791
Jinyi Wang	460131494
Rinnie Aggarwal	450614587
Shailendra Pal	450598737
Xuhong Guo	450489321



THE UNIVERSITY OF
SYDNEY

School of Information Technologies
Faculty of Engineering & IT

ASSIGNMENT/PROJECT COVERSHEET - GROUP ASSESSMENT

Unit of Study: COMP5048 Visual Analytics

Assignment name: Assignment 3 Report

Tutorial time: Thursday 6:00pm-8:00pm Tutor name: Masahiro Takatsuka

DECLARATION

We the undersigned declare that we have read and understood the University of Sydney Academic Dishonesty and Plagiarism in Coursework Policy, and except where specifically acknowledged, the work contained in this assignment/project is our own work, and has not been copied from other sources or been previously submitted for award or assessment.

We understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy can lead to severe penalties as outlined under Chapter 8 of the University of Sydney By-Law 1999 (as amended). These penalties may be imposed in cases where any significant portion of my submitted work has been copied without proper acknowledgement from other sources, including published works, the internet, existing programs, the work of other students, or work previously submitted for other awards or assessments.

We realise that we may be asked to identify those portions of the work contributed by each of us and required to demonstrate our individual knowledge of the relevant material by answering oral questions or by undertaking supplementary work, either written or in the laboratory, in order to arrive at the final assessment mark.

Project team members				
Student name	Student ID	Participated	Agree to share	Signature
1. <u>Xuhong Guo</u>	<u>450489321</u>	<u>Yes / No</u>	<u>Yes/No</u>	<u>Xuhong Guo</u>
2. <u>Jinyi Wang</u>	<u>460131494</u>	<u>Yes / No</u>	<u>Yes / No</u>	<u>Jinyi Wang</u>
3. <u>Bahua Zhang</u>	<u>450280418</u>	<u>Yes / No</u>	<u>Yes / No</u>	<u>Bahua Zhang</u>
4. <u>Feng Wang</u>	<u>450518791</u>	<u>Yes / No</u>	<u>Yes / No</u>	<u>Feng Wang</u>
5. <u>Rinie Aggarwal</u>	<u>450614587</u>	<u>Yes / No</u>	<u>Yes / No</u>	<u>Rinie</u>
6. <u>Shailendra Pal</u>	<u>450598737</u>	<u>Yes / No</u>	<u>Yes / No</u>	<u>Shailendra</u>
7.		<u>Yes / No</u>	<u>Yes / No</u>	
8.		<u>Yes / No</u>	<u>Yes / No</u>	
9.		<u>Yes / No</u>	<u>Yes / No</u>	
10.		<u>Yes / No</u>	<u>Yes / No</u>	

SIT Building, J12
The University of Sydney
NSW 2006 Australia

T +61 2 9351 3423
F +61 2 9351 3838
E sit.info@sydney.edu.au
sydney.edu.au/it

ABN 15 211 513 464
CRICOS 00026A

Name	ID		
Bohua Zhang	450280418	Q1(visualization) -> version 1 (part) Q2(visualization+answer)->version 1 Q2(visualization+answer)->version 2 Introduction (part)	Bohua Zhang
Feng Wang	450518791	Q1(visualization) -> version 1 (part) Q3(visualization) -> version 1 (part) Introduction(part)	Feng Wang
Jinyi Wang	460131494	Q1(answer)->version 2 Q3(visualization) -> version 1 (part) Q5(answer)->version 1	Jinyi Wang
Rinnie Aggarwal	450614587	Q1(answer) -> version 1 Q3(visualization) -> version 1 (part) Q4(answer) -> version 2 Conclusion (part) Introduction (part)	Rinnie Aggarwal
Shailendra Pal	450598737	Q1(visualization) -> version 1 (part) Q4(answer) -> version 1 Q5(answer) -> version 2 Conclusion (part)	Shailendra Pal
Xuhong Guo	450489321	Q1(visualization)->version 1 (part) Q1(visualization)->version 2 Q3(visualization)->version 2 Introduction (part)	Xuhong Guo

Abstract

This report provides the solutions for the visualization issues in COMP 5048 Visual Analytic Assignment 3. The content lists the question description, the method used to obtain the solutions and the final results of the visualization tasks. Key parts of the solutions include visualization image, Key Goal Indicator (KGI), Key Performance Indicator (KPI), and explanation of each task.

I . Introduction

In this assignment, we are asked to visualize the specified dataset and analyze some key questions based on our visualization. The data for this assignment was collected from PDADB.net, which is an extensive online database of product data and specifications of smartphones, tablets, PDAs, PNAs, notebooks and other mobile devices. It contains the specifications of more than 10 thousand mobile devices around the world. The dataset contains the specifications of 2903 mobile devices introduced between 2003 and 2012, and for these devices, a category is assigned to each device to identify the cluster it belongs to. The data includes 13 continuously valued attributes of each device, such as CPU speed and RAM capacity, and all these data are normalized between 0 and 1.

There are five questions for this Assessment 3, each of which has been answered through following parts:

1. Show the visualization, and describe the visualization.
2. Show the various types of data in dataset have been mapped to visual variables.
3. Explain the answer to the question through the visualization.

The rest of the report is organized as follows:

Section I	Introduction
Section II	Lists of tasks
Section III	Describes the method used to obtain the solutions
Section IV	Solution for each question.
Section V	Conclusion

Task 1:

1.1 The description of the visualization developed

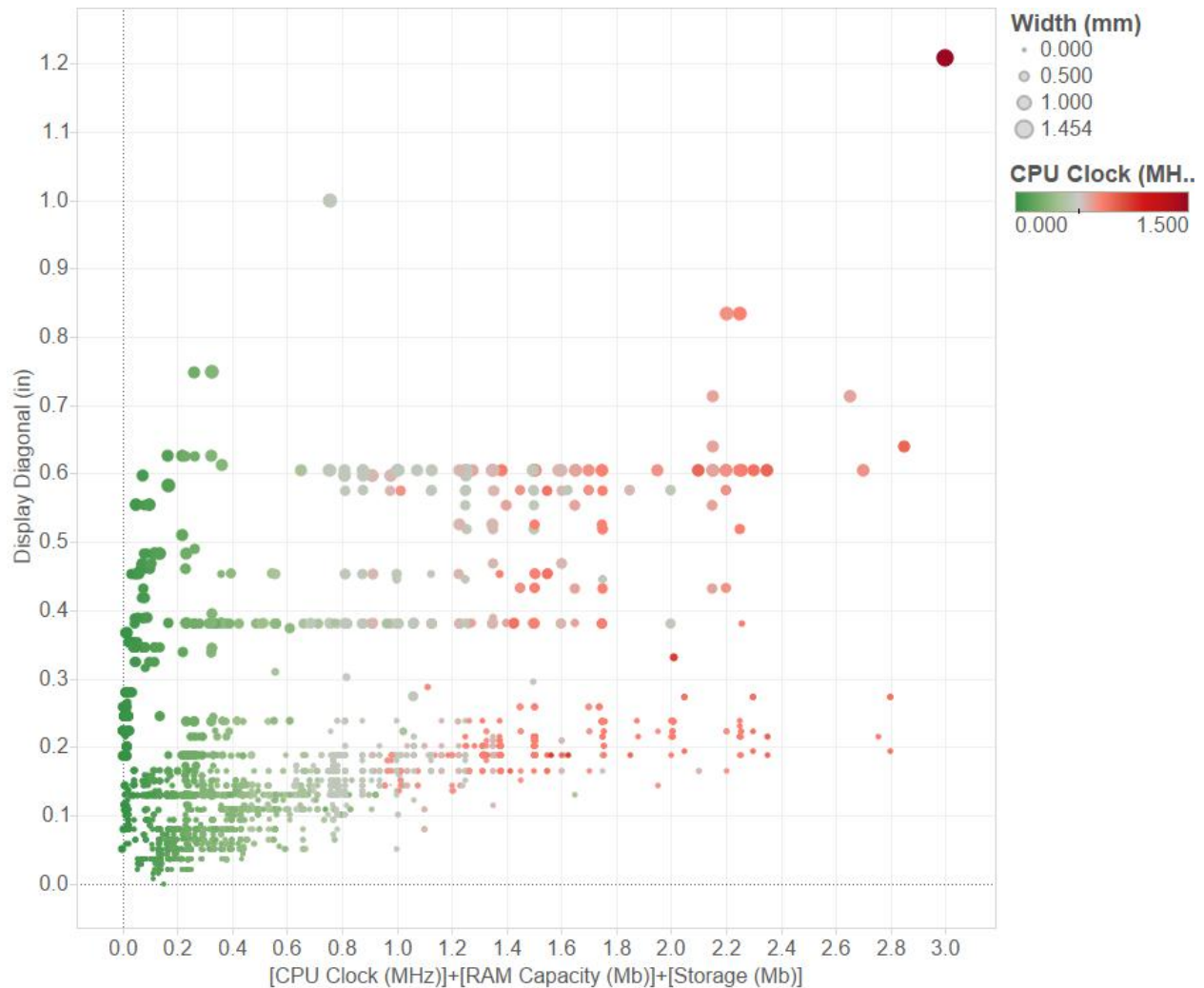


Figure 1. Visualization of market niches

Explanation:

- **X-axis:** the performance of the device. It was calculated by summing three attributes: CPU Clock, RAM capacity and Storage capacity. Performance of the device improves with increase in the performance value.
- **Y-axis:** the Display Diagonal. the higher is this value, the larger is the screen size.
- **Color of the dots:** CPU Clock which ranges from green to red. The Darker red implies high CPU Clock.
- **Size of the dots:** Width of the product. The wider the dots, the more is the product width.

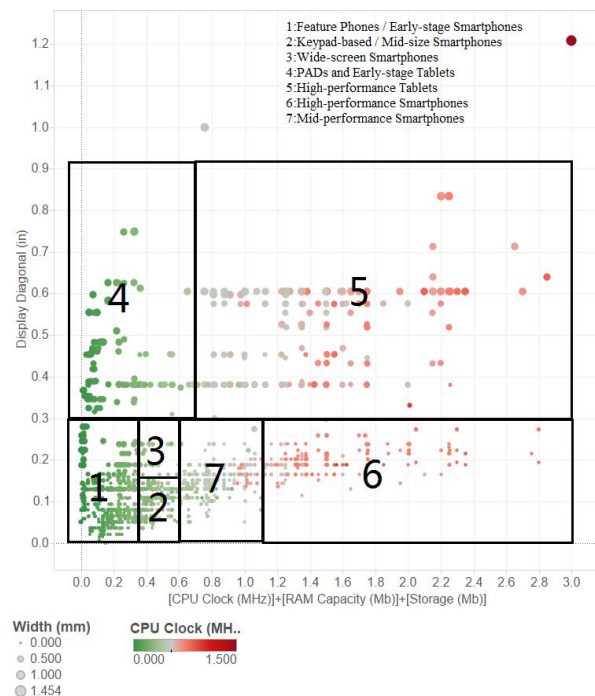
1.2 Describe how various types of data in the dataset have been mapped to visual variables.

KGI: Identify the seven market niches.

KPI:

- CPU Clock(MHz) - Interval
CPU Clock is an Interval variable. But in the visualization, this value has been transformed from Interval to Ordinal and is represented by Color.
- Width(mm) - Interval
Width of product is an Interval variable in the database. But in the Visualization, Interval value of Width is transferred to be size value which is Interval.
- Performance of the device - Interval
The performance is the sum of CPU Clock, RAM capacity and Storage capacity. It is Interval.
- Display Diagonal - Interval
This represents the screen size of the device.
- Model - Nominal
Model describes the name and model type of the device.

1.3 Derive answers to the question(s) through the visualization



Explanation:

Class 1: Class 1 is composed of all the products whose performance is represented as dark green color from 0 to 0.35 and display diagonal is below 0.3. This maps to Feature phones and Early-stage smartphones with small displays and low performances.

Class 2: Class 2 consists of the products whose performance is represented by light green color from 0.35 to 0.6 and display diagonal is between 0.5 and 0.15. This accounts for Keypad-based and Mid-size smartphones as their performance would be better than the Early-stage smartphones and the screen should be medium-sized.

Figure 2. Analysis of the visualization of market niches

Class 3: According to the figure, the products whose performance is represented by light green color, from 0.35 to 0.6 and the display diagonal is between 0.15 and 0.3 are grouped as Class 3. Since the Wide-size smartphones should have similar performance to Mid-size smartphones. But their display size should be larger than that of Mid-size smartphones.

Class 4: We choose the products whose performance is represented by green color below 0.7 and display diagonal over 0.3 as PADs and Early-stage tablets. For PADs and Early-stage tablets, the size of display should be larger than smartphone and some PADs would perform better than Early-stage smartphones.

Class 5: The products whose performance is represented by red color, over 0.7 and display diagonal over 0.3 can be marked as High-performance tablets. For the High-performance tablets, the size of display would be similar to other PADs and Early-stage tablets and their performance should be the best out of all these devices.

Class 6: We choose the products whose performance is represented by red color, over 1.1 and display diagonal below 0.3 as High-performance smartphones. Since their performance should be the best of all smartphones and the size of screen would be similar to other smartphones.

Class 7: The products which the products whose performance is represented by gray color, between 0.6 and 1.1 and display diagonal below 0.3 are marked as Class 7. Since for Mid-performance smartphones, the size of display should be similar to other smartphones, and their performance should be lower than High-performance smartphones but better than other smartphones.

Task 2:

2.1 The description of the visualization developed

Change of amount of devices

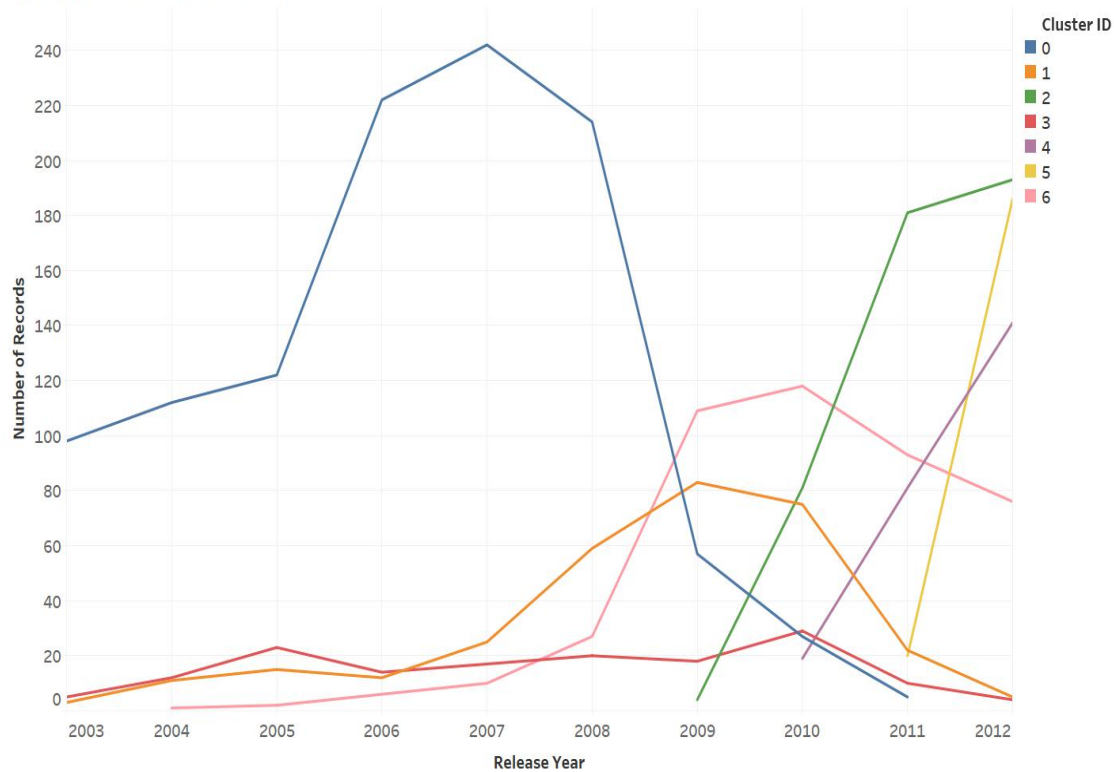


Figure 3. Change of amount of devices

products whose performance is represented by

- **X-axis:** Released year of products which use the whole number such as (2003, 2004, 2005...)
- **Y-axis:** Number of records representing number of products per year in each market.
- **Color:** Market niche (total 7 markets as per the assignment)

2.2 Various types of data in the dataset have been mapped to visual variables

KGI: Identify the change of the amounts of devices in each market niche

KPI:

- **Number of records** - Ratio. It counts the number of products in a particular market.
- **Release Year** - Ordinal
- **Model** - Nominal
- **Cluster ID** - Nominal

2.3 Derive answers to the question(s) through the visualisation

Explanation:

Considering each record corresponds to a unique model (aka device), the Number of Records is used to represent the number of devices in each market niche. The data used in visualization is shown in the table below.

Table 4. The change of amount of devices data

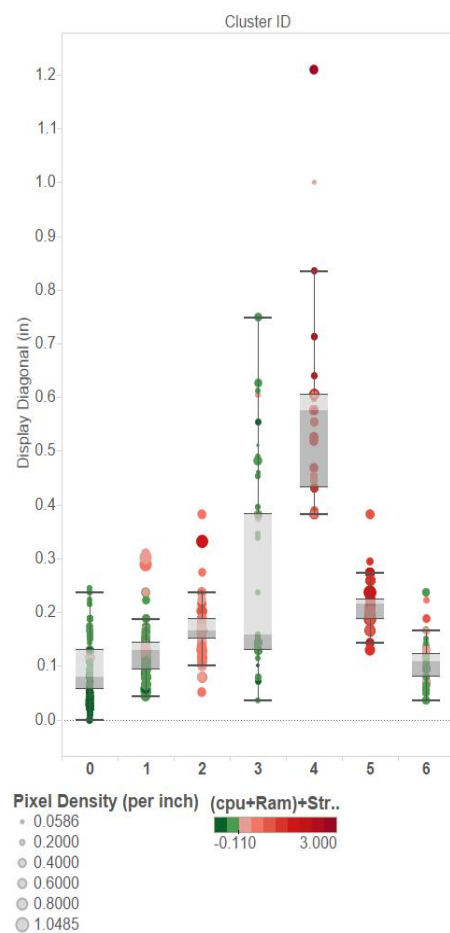
Amount of devices	Release of Year										
Cluster ID	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total
0	98	112	122	222	242	214	57	27	5		1099
1	3	11	15	12	25	59	83	75	22	5	310
2							4	81	181	193	459
3	5	12	23	14	17	20	18	29	10	4	152
4								19	81	141	241
5									20	186	206
6		1	2		10	27	109	118	93	76	436
Total	106	136	162	248	294	320	271	349	412	605	2903

The year range of the devices is from 2003 to 2012. However, some market niches suddenly appeared during these years. For instance, Cluster 2, which refers to keyboard-based/mid-size smartphones market niche, appeared after 2009. While some market niches vanished such as Cluster 0, which refers to feature phone/early-staged smartphone market niche, disappeared after 2012.

It can be deduced from the visualization that the number of models for Cluster 0, 1, 3, 6 reached their peaks and then started decreasing, while the ones in Cluster 2, 4, 5 kept increasing. These changes may reveal the trend of the customers' requirements and industry standards. For example, for Cluster 5, which refers to high performance tablets, the number of devices in 2012 is nearly ten times than that in 2011. This indicates that the customers became more and more interested in high performance tablets.

Task 3:

3.1 The description of the visualization developed



Explanation:

- **X-axis:** Cluster ID corresponding to each market niche.
- **Y-axis:** Display Diagonal.
- **Color:** Performance of product (sum of CPU, RAM and Storage).
- **Size:** Pixel density which represent the high definition.

Figure 4. Visualization of market leaders

3.2 Describe how various types of data in the dataset have been mapped to visual variables

KGI: Identify products which led the market in each market niche

KPI:

- Performance(CPU+RAM+Storage) - Interval
- Model - Nominal
- Display Diagonal - Interval
- Cluster ID - Nominal

3.3 Derive answers to the question(s) through the visualization

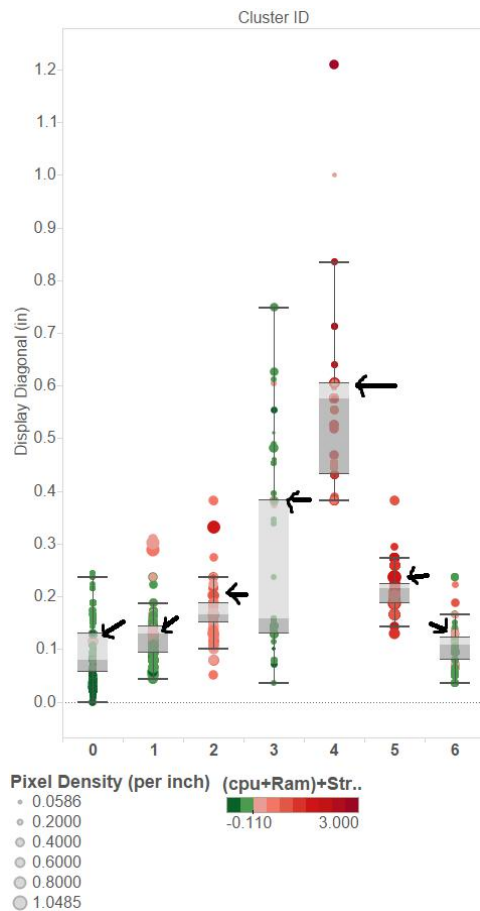


Figure 5. Analysis of visualization of market leaders

Explanation:

It is assumed that a product with high performance and larger number of pixels (high definition quality of a product) leads a market.

The box plot shows the majority of the products in each market.

The Color in the visualization indicates the performance of the product. Green means low performance and as it moves from green to darker shades of red, the performance increases.

The Size of the dots shows the pixel density

The point which is nearer to the majority, has a darker shade of red and is bigger in size, as compared to other points, leads the market. This is shown in the Figure 5 with the anchor sign.

Task 4:

From the task 1 to task 3, the visualizations showed that Display Diagonal, CPU Clock, RAM capacity, Storage capacity and Pixel Density are the main causes of those product leading the markets. As we know, CPU carries out the instructions of a computer program by performing the basic arithmetic, logical, control and input/output (I/O) operations.[2] Hence CPU, RAM and Storage play major role in defining the performance of a product.

From task 3, the visualization showed that display diagonal is an important characteristic of a product to lead in each market because the product must follow the majority.

The product which is leading the market belongs to the majority. However, it needs to have higher values of CPU Clock and Display Diagonal as compared to others. A high pixel density also indicates better visual quality for users.

Task 5:

From the visualization in task 3, it can be seen that in each market, the products are scattered, most of them form the majority, while some are separated from the majority.

These products, which aren't part of the majority, are assumed to be the ones who failed to catch up the market changes. This can be seen in Figure 6 below.

The box plot shows the majority in Display Diagonal, the color of the dots shows the majority of CPU Clock and the size shows the majority in pixel density in each market.

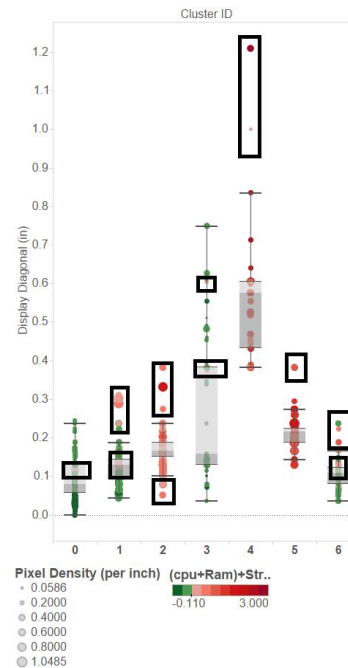


Figure 6. Analysis of box and whisker plot of failed products

V. Conclusion

It's a human tendency to look for clusters whenever there is a huge data set in front of us. So, to make the identification of clusters easier, we've used scatter and box plots which help us group data into clusters and make the differentiation easier.

Hence, we can conclude that with the help of visualizations, it is possible to reduce a very large data set into small sets of clusters/visualizations which would help us answer many simple to complex questions.

Reference

- [1] A. Carroll and G. Heiser, "An analysis of power consumption in a smartphone." in USENIX annual technical conference, vol. 14. Boston, MA, 2010.
- [2] M. H. Weik, "A third survey of domestic electronic digital computing systems," DTIC Document, Tech. Rep., 1961.