European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 08, Issue 02, 2021

An Interactive Dashboard For Information Visualization On The Air Pollution Index In Malaysia (Apimas)

Siti Nur Khadijah Juhair[1], Nor Aziah Daud[2], Nurulhuda Noordin[3], Fauziah Redzuan[4], Wan Adilah Wan Adnan[5], Noraidah Sahari Ashaari[6]

[1,2,3,4,5] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[6]*Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia*

[2]*e-mail: aziah@tmsk.uitm.edu.my*

Abstract: This research identifies the requirement used to visualize Air Pollution Index in Malaysia (ApiMAS). Air pollution is a worldwide issue in the area of environment and health, especially in developing countries, which calls for crucial part of governments to implement solutions to solve air quality issues. Most people are unaware and not alert of particle pollution levels that can affect their health. When pollution information is incompatible, the lack of information sharing causes difficulties to the people. This research focuses on the on-site of Kuala Lumpur. Data visualization is most suitable for researchers to analyze, discover, and monitor performance through graphs, charts, or pictographs that can represent the data for information virtualization. The Cross-Industry Standard Process for Data Mining (CRISP-DM) model which is used as the methodology has four phases including business understanding, data understanding, data preparation, and modelling. This research output is an interactive design dashboard for information visualization that is necessary for researchers, public users, and technologists. In conclusion, interactive dashboard ApiMAS may assist its users with prolific knowledge provided in the dashboard.

Keywords: Air pollution index; ApiMAS, data visualization, information visualization, interactive design, dashboard.

1. INTRODUCTION

The requirement of human health is clean air and continuous life. The increase cases of air pollution give out threat to human health. World Health Organization (WHO) [1] mentioned that in the valuation of illnesses due to air pollution, more than 2 million premature deaths that happened each year were contributed by the effects from outdoor and indoor air pollution caused by burning of solid fuels in the city. Statistics show that significant number of disease problems also emerged in developing countries.

According to the Department of Environment in Malaysia in 2019, the air pollutant index (API) system can trace and recognize many types of air pollutants that could be harmful to human health. As stated by WHO, in the year 2016, ninety-seven percent (97%) of

cities in low and middle-income level countries with more than 100,000 inhabitants did not meet the air quality requirements as recommended by the organization. However, for high-income level countries, the percentage was about forty-nine percent (49%).

Therefore, the crucial part is for the government to implement the solution regarding air quality status. Air pollution can affect the breathing system, cause sore eyes, and irritation to the skin. However, due to the industrial revolution and increment in the usage of transportation which causes more pollution, this situation leads to temperature change which eventually precipitate global warming. Thus, it is a significant responsibility to understand the status of air pollution to enable the decision making or construction of action plan to reduce air pollution. Another problem that arised was for the researchers to determine the main factors contributing to the air pollutions. Data collection is incompatible whereby lack of information sharing between agencies can cause difficulties for public to access any environmental information. Air pollution data are currently collected by relevant agencies, but the data are not made visible to the public users.

Data visualization, in general, reveals the effort to aid people in understanding the significance of data by placing it in a visual context [2] [3]. Patterns, trends, and correlation that could go under the radar in text-based data can be easily revealed and identified with data visualization. Data visualization is a presentation technique of quantitative information in an infographic arrangement which makes thoughtful information that is a lot easier to comprehend in order to make a better decision [4], [5], [6], [7]. It is the use of human natural talents to enrich data processing and organization proficiency. Its main goal is to distill large datasets into visual graphics to allow an easy understanding of complex relationships within the data [8]. In data visualization, a dashboard will be developed on uppermost of the page so that information can be presented clearly and efficiently with various types of graphs and maps [9]. In the process of visualization, many type of patterns can be identified. The patterns help in understanding comparisons, making connections, and drawing conclusion from the datasets [10]. According to [2], it assists a user's understanding of the data and makes it easier to analyses them. It is the transferal of non-concrete data using collaborative visual interfaces. It also focuses on the development of approaches for transmission information in intuitive ways. Information visualization is used to increase people's awareness of conceptual information to visualize interactive data.

This research presents an interactive dashboard for information visualization on the Air Pollution Index in Malaysia (ApiMAS). To accomplish the research, the Cross-Industry Standard Process for Data Mining (CRISP-DM) model is used as the methodology. This research aims to determine an interactive dashboard for information visualization on the ApiMAS. This interactive dashboard for information visualization on ApiMAS is hoped to assist its users with prolific knowledge provided in the dashboard.

2. MATERIAL METHOD

Dashboard is a platform to communicate critical information to the target audience in the way they can understand and delivers the need for the information. Dashboard can focus on displaying operational and analytical data that are typically aimed to help decision makers, and the most important part is to make target users understand to what and why something happened. Also, the same information on the dashboard can assist the users to implement appropriate changes. Data are visualized on a dashboard in many ways such as bar charts, line charts, and others that can be used to track and monitor through visual representations.

A. Review of Existing Dashboard

This section compares existing dashboard that relates to visualization on air pollution index such as AQI India, AirNow and Sciencemag.

A. AQI India is a website that presents the visualization of the output on the air quality index in India and displays it in a dashboard. It utilizes a table to represent the data of air quality of each state and relate with the parameter of air pollutants PM2.5 and PM10. Other information such as humidity, temperature, as well as noise data for the location are also displayed in the table. The table is one of the techniques of data visualization that is easy and clear to understand which allows the users to interpret, check and update the air pollution status in their areas. AQI India dashboard is shown in Figure 1.

Figure 1: AQI India – Dashboard for air quality status

AirNow is a dashboard that monitors the real-time air quality status in Kansas, United States. It also presents the visualization interactively in a map-based that covers the area of each state. It utilizes the indicator that can represent the level of air quality status for the user to understand the current level of air pollution in a simple manner. The data are also sorted according to zip codes and states. AirNow dashboard is shown in Figure 2.

Figure 2: AirNow – Dashboard for Air Quality Index in the United States

Sciencemag is a dashboard that visualizes the world in map based. It is interactive in terms of the graphics and methods to visualize the earth. It displays the air quality index of

the major cities in the world. The interactive map also shows the wind patterns and displays color-coded concentrations of PM2.5- airborne of particulate matter less than 2.5 microns in diameter. The Sciencemag dashboard is shown in Figure 3.

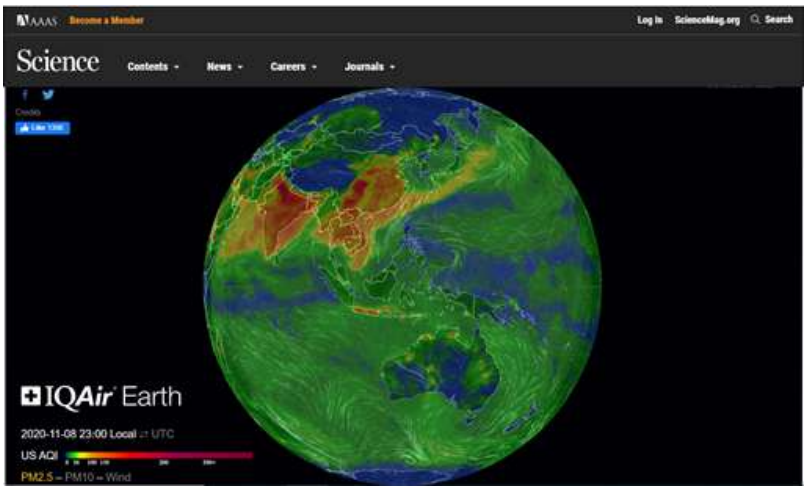

Figure 3: Sciencemag – Dashboard for Air Quality Index in AirVisual Earth

B. Review of Existing Dashboard

According to the Department of Environment in Malaysia (2019), air pollutant index (API) is recognized to deliver simple, and understandable information to users on air pollution. Malaysian air contaminants include ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), and particulate matter of less than 10 microns (PM10). The Malaysian API system closely follows the Pollutant Standard Index (PSI) developed by the United States Environmental Protection Agency (US-EPA)1.The air pollutant index scale and terms used in describing the air quality levels are shown in Figure 4.

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	*...air quality conditions are:*	*...as symbolized by this color:*
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Figure 4: Levels of Air Quality Index

C. Cross-Industry Standard Process for Data Mining (CRISP-DM)

This research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) model for data analytics techniques and visualization techniques [11]. The methodology is a crucial part to start the research in order to complete it in time and prevent project delays due

to the problems that can occur. According to [12], the data mining process helps to identify business issues in data mining activities, suggests suitable data processing and methods, as well as including documentation support in detail. This methodology carries out the data analytic process to analyze the data. The main reason this methodology is chosen is because it is based on the practical, and real-world experience of how people conduct data mining projects. Initially, this approach consists of six stages. However, this research is only carried out until the modeling phases due to this research necessity, which includes the creation of an interactive dashboard for data visualization to be revised in this part.

Business Understanding - It determines the understanding of the project's objectives and requirements needed to convert the information into a data mining problem statement, initial literature review and design to achieve the goals. It is vital to conduct preliminary studies on the air pollution index to analyze the data.

Data Understanding - This stage is defined as collection of raw data and to proceed with activities such as to identify data quality problems. Next is to discover the first perceptions into data or to detect interesting subsets to form a relationship for unseen information. According to [10], this phase starts with collecting data and activities to attain more understandable data, to identify data quality problems, to discover all possible outcomes in data, and to detect some exciting subsets of data. Air pollution index (API) datasets must be retrieved by sending emails to get permission to apply the data and setting appointment to interview the expert. The data are obtained from The Department of Environment (DOE) in Malaysia [13] and *Portal Data Terbuka Sektor Awam (MAMPU)*.

Data Preparation - Data preparation covers all actions to conduct the ultimate dataset from raw data, which is unstructured data. Tasks can include a table, record, and attribute selection as well as cleaning of data for modeling tools. Selected analysis techniques such as clean subsets of data, insert to the missing value or remove.

Modelling - Several modelling techniques are designed and applied for the same data-mining type. Parameters correlate to the optimal values. Some of the methods have specific requirements. This project utilizes descriptive analysis which is a fundamental part of the statistical method that can show exact interpretation for all parameters of the air pollutant index. This research focuses on basic descriptive such as mean, median, mod, and maximum based on average air pollution index starting from year 2015 until 2018. The cluster analysis is used to establish the characteristics of each air pollutant. It also creates a relationship from large samples and makes a summary out of it. This technique is also applied to identify air pollution index patterns from 2015 until 2018. Figure 5 shows the steps of the CRISP-DM process for ApiMAS.

D. CRISP-DM Method Modelling

Modelling the data refers to techniques used to communicate insights from data through visual representation. By plotting the various features in the air pollution index dataset, the purpose is to find any relationships and correlations between them and other datasets. Figure 6, Figure 7, and Figure 8 show the example of the data visualization.

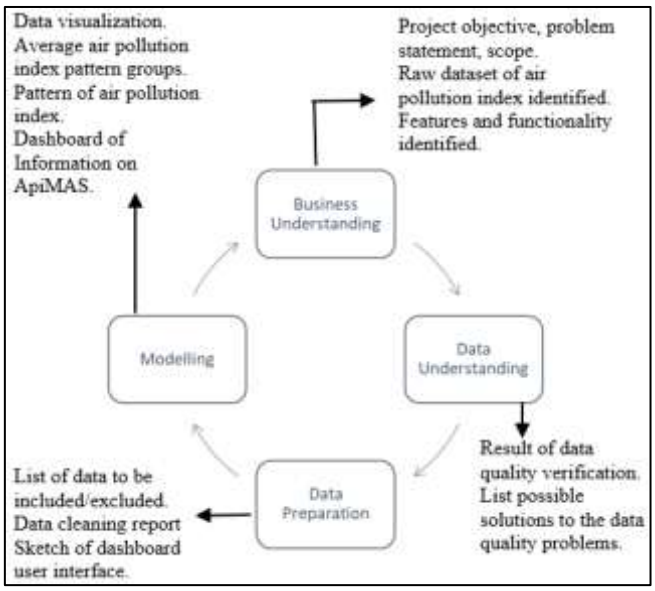

Figure 5: Cross Industry Standard Process for Data Mining (CRISP-DM) Model

Figure 6: Visualization of Air Pollution Index by Features in Data and Heat Map

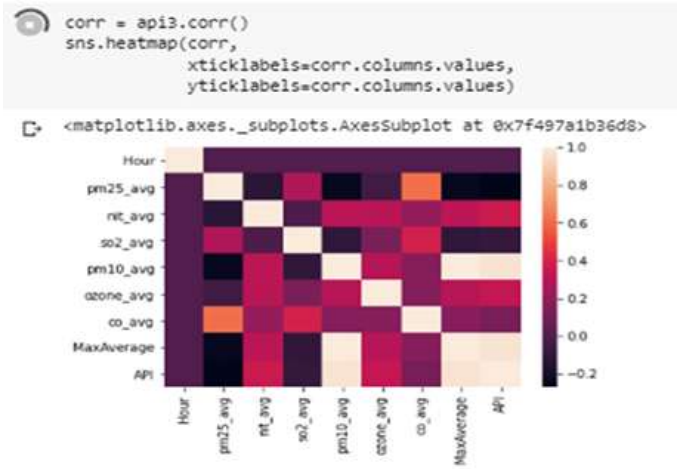

Figure 7: Visualization of Air Pollution Index by Features in Dataset and Heat Map Area in Kuala Lumpur

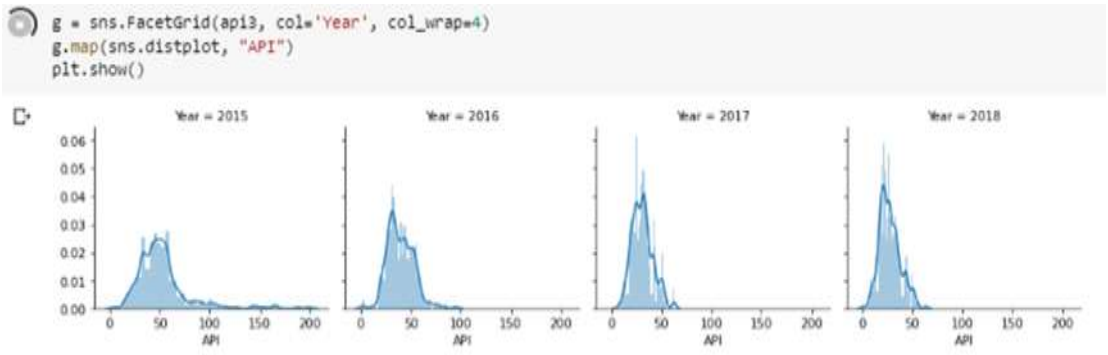

Figure 8: Visualization of Air Pollution Index by Features in Dataset

3. MATERIAL METHOD

The results of the research is a prototype of the ApiMAS interactive dashboard. The dashboard is able to offer supporting information that can compare and monitor varieties of data simultaneously [4]. Developing the dashboard is a crucial process to ensure that it aligns with all requirements. This section discusses the dashboard design that had been tested for user acceptance.

A. Interactive Dashboard ApiMAS

The dashboard presents the overall trend of Air Pollution Index (API) in Kuala Lumpur from 2015 until 2018. Based on Figure 9, the API shows a downward trend with a significant decrease in API between 2015 until 2018. The highest API was in 2015, where the value was 199 which shows unhealthy status. In 2016 to 2018, the API values were 95, 62 and 63 respectively, and the status of air pollution was moderate. Air pollution develops mainly from transportation (vehicles), industrial emissions, and open burning sources [14]. Among these sources, transportation contributes the most to air pollution. Malaysia had gone through lousy air pollution such as severe haze mostly caused by forest fires in Indonesia that released harmful pollutants in September until October 2015 [15]. This annual air pollution left tremendous impacts on the environment, socioeconomic and health issues that have raised concern among international political associate nations of the Association of Southeast Asian Nations (ASEAN).

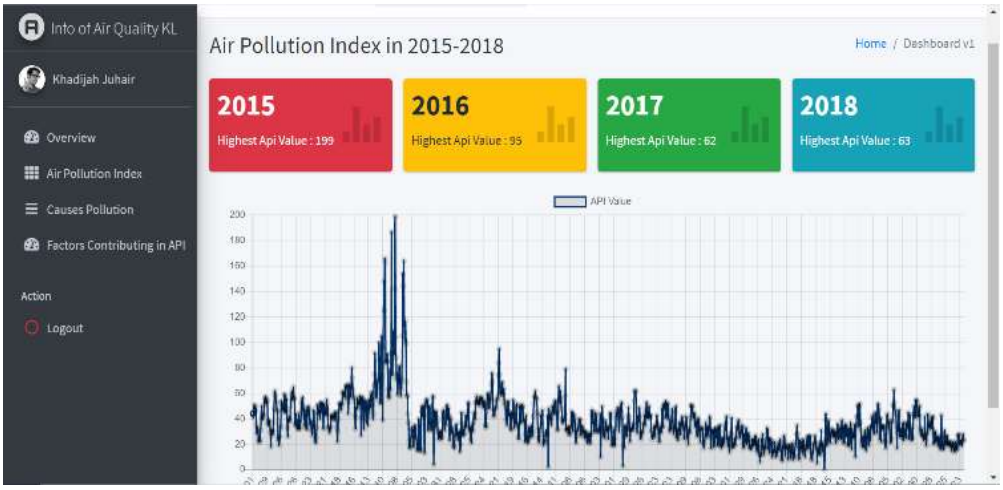

Figure 9: Trend of Air Pollution Index (API) in Kuala Lumpur

The ApiMAS dashboard also provides Map View of Industry in Kuala Lumpur. Figure 10 shows an overview of a map that marks the industrial area in Kuala Lumpur where users can find the details and longitude of the industry.

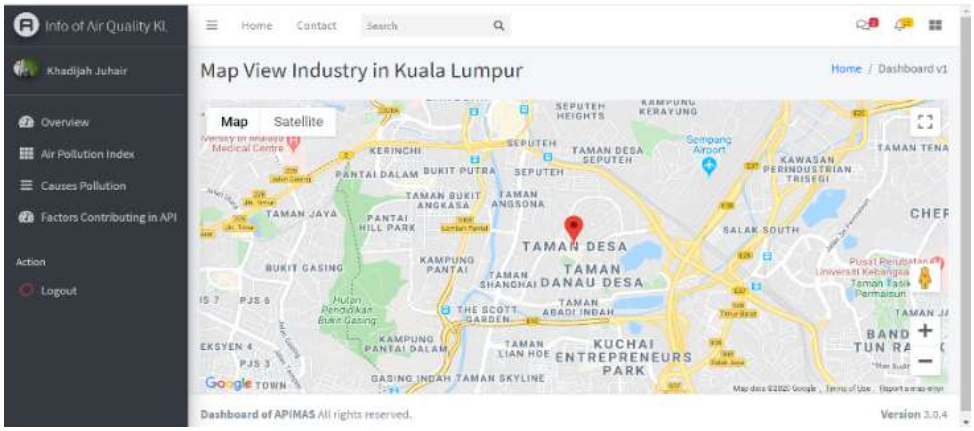

Figure 10: Map Page

On top of that, users can view the information on each of the graph such as in Figure 11 that shows the causes of pollution which are SO_2, particulate matter, NO_2 and CO, as well as the information on how this pollution can affect human health. While users hover over the chart, the information will appear on each of the bar graphs.

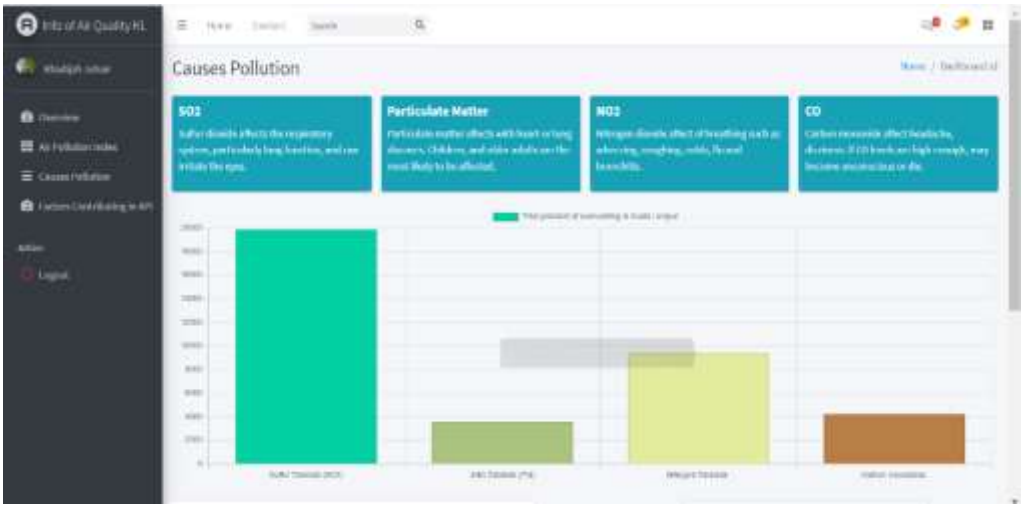

Figure 11: Causes of Pollution Page

Figure 12 to Figure 14 show the information on factors that contribute to air pollution. The users can also hover over the detail information on each of the bar graphs. Figure 12 shows the average rainfall as one of the air pollution factors between the years of 2013 to 2017. Meanwhile, Figure 13 shows vehicle usage as the main factor in year of 2015 to

2018 while Figure 14 shows another factor which is population growth. Different pages show the interactive graph and accurate information of population to users. The card information is updated from the data and appears with the population value as well as growth rate values.

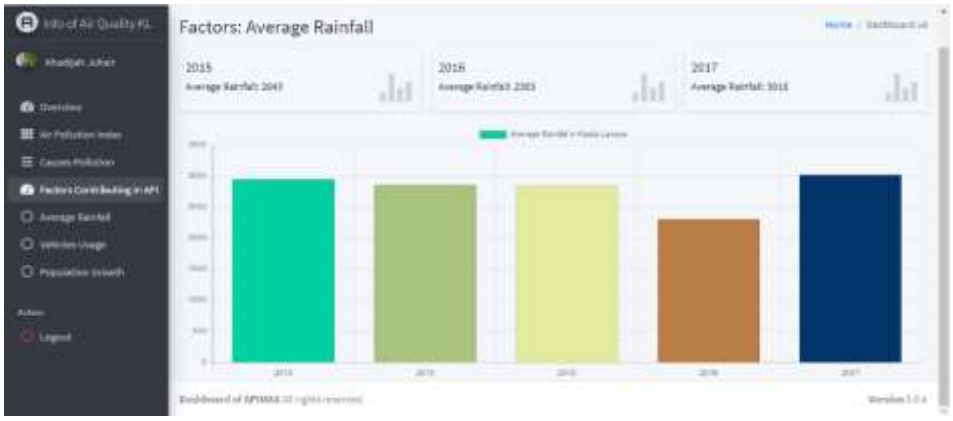

Figure 12: Factors: Average Rainfall Page

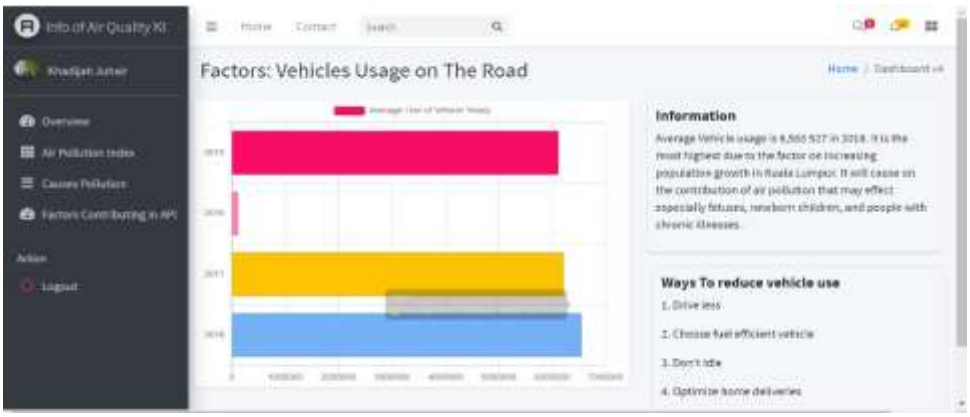

Figure 13: Factors: Vehicle Usage

Figure 14: Factors: Population Growth

4. CONCLUSION

The primary goal of this research is to design and develop interactive ApiMAS dashboard that enables users in identifying which factors cause most of the air pollution with the help of visualizations in the form of graphs. The information retrieved from the visualization can assist users in their decision-making. Understanding and awareness among the public users are more effective to help reduce the causes of air pollution. The technique applied to develop interactive dashboard ApiMAS is data analytic technique that follows the methodology of CRISP-DM model. All the chosen methods are briefly explained and figured. The development can also provide a good platform to convey air pollution information to public users. The user will be able to get hold of the data and information to be used as their reference. It also allows the users to be more aware about air pollution in the future. During the design phase, there are numerous elements and stages to be discovered and learned to provide new interpretation and perspectives that can be applied in responsive web application. As for future research, this area can be enhanced and expanded in depth.

Acknowledgements
The authors gratefully acknowledge the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia for the support and resources to this research.

REFERENCES

[1] World Health Organization (WHO), Air Quality Guidelines for particulate matter, ozone, nitrogen dioxide and sulphur dioxide, (2019), available at https://www.who.int/.
[2] Y. Li, S. Hou, Methods and Techniques in Data Visualization Model, International Conference on Computer Technology, Electronic and Communication, (2017).
[3] Izyan Izzati Kamsani, Norharyati Md Ariff, Noor Elaiza Abd Khalid, Survey of Star Glyph-Based Visualization Technique for Multivariate Data, Australian Journal of Basic and Applied Sciences, 9(26) Special 2015, Pages: 61-66, (2015).
[4] Nikhat Akhtar and Nazia Tabassum et al, "Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus)", Global Journal of Engineering and Technology Advances, 03(02), 028–050, (2020).
[5] Saidatul Rahah Hamidi and Anitawati Mohd Lokman, et al, "CancerMAS Dashboard: Data Visualization of Cancer Cases in Malaysia", Journal of Physics: Conference Series 1529 (2020).
[6] Wan Azzura Wan Ramli and Zainura Idrus et al, Data Visualization: Toward Visualization Awareness In Overlapped Dots Management Techniques, Journal of Theoretical and Applied Information Technology 30th June 2019. Vol.97. No. 12, (2019).
[7] M.Z.A.Razak, S.S.M.Fauzi, R.A.JM. Gining, M.N.F. Jamaluddin, Data visualisation of vehicle crash using interactive map and data dashboard, Indonesian Journal of Electrical Engineering and Computer Science Vol. 14, No. 3, June 2019, pp. 1405~1411, (2019).
[8] P. Firoj, Nikhat A and P. Dr. Yusuf. "A Close-Up View About Spark in Big Data Jurisdiction", International Journal of Engineering Research and Application (IJERA), ISSN: 2248-9622, 8(1), (Part -I1), 26-41, (2018).
[9] Y. Xin and B. Claude et al, Data Visualization in Conceptual Design: Developing a Prototype to Support Decision, Conference: Conference: 12th International Conference on Modelling, Optimization and Simulation, At Toulouse, France, (2018).

[10] R. Julie and K. Piotr, Using Design Patterns to Find Greater Meaning to Your Data. O'reilly.(2016).
[11] D. Arce and F. Lima et al, Discovering behavioural patterns among air pollutants: A data mining approach, Spanish: University of Azuay, (2018).
[12] P. Cahpman and J. Clinton et al, CRISP-DM 1.0: Step by step data mining guide. Denmark: NCR Systems Engineering Copenhagen. (2000).
[13] Department of Environment Malaysia (DOE), Air Pollutant Index, (2019), available at https://www.doe.gov.my/
[14] N. Shaadan and L.N. Nazeri et al, Data Visualization Of Temporal Ozone Pollution Between Urban And Sub-Urban Locations In Selangor Malaysia, Journal of Fundamental and Applied Sciences, ISSN 1112-9867, Special Issue, (2017), available at http://jfas.info
[15] L. Tacconi. COMMENTARY: Preventing fires and haze in Southeast Asia, Nature Climate Change, vol. 6, no. 7, pp. 640-643. (2016).