WALAILAK JOURNAL

The Role of Data Visualization and Analytics of Highway Accidents[†]

Jirapon SUNKPHO^{1,*} and Warit WIPULANUSAT²

¹Thammasat University AI Center, College of Innovation, Thammasat University, Bangkok 10200, Thailand ²Logistics and Business Analytics Center of Excellence, Faculty of Engineering and Technology, Walailak University, Nakhon Si Thammarat 80160, Thailand

(*Corresponding author's e-mail: jirapon@tu.ac.th)

Received: 17 June 2019, Revised: 28 February 2020, Accepted: 9 March 2020

Abstract

Thailand has been ranked as one of the most dangerous countries in terms of death from road accidents, representing ineffective road safety policies. The crucial mission of the Thai government is to provide safety and reduce accidents for road users on the highway system. This paper aims to explore the potential of using Business Intelligence (BI) in accident analysis. The availability of open accident data provides an opportunity for the BI, which can provide an advanced platform for conducting data visualization and analytics in both spatial and temporal dimensions in order to illustrate when and where the accidents occur. The accident data and provincial data were combined by using the Talend Data Integration tool. The combined data was then loaded into a MySQL database for data visualization using Tableau. The dashboard was designed and created by using Tableau as an analytical visualization tool to provide insights into highway accidents. This system is advised to be adopted by the Thai government, which can be used for data visualization and analytics to provide a mechanism to formulate strategy options and formulate appropriate contingency plans to improve the accident situation.

Keywords: Visualization, Data analytics, Business intelligence, Accidents, Thailand

Introduction

Accident, especially road traffic accident, is one of the leading causes of death fatality and personal injuries. According to the World Health Organization (WHO), nearly 1.3 million people die annually in highway-related accidents each year [1]. Road accidents are essential concerns in modern society, particularly in developing countries where the fatality rate is very high. Thailand has been categorized as a developing country and has been ranked as one of the most dangerous countries in terms of death from road accidents, representing ineffective road safety policies. Specific to motorcycle accidents, the mortality rate was the global highest road-based death rate, with a death rate of 27.3 per 100,000 population. A report by World Atlas revealed that the death rate of Thailand is number 2, caused by road accidents with a death rate of 36.2 per 100,000 population in 2018 [2]. Accidents result in direct damage to the vehicles and losses for fatal casualties, in addition to traffic congestion and travel delays [3].

Thailand is a festive kingdom celebrating numerous national holidays, such as the New Year and Songkran Festival. During these holidays, the relocated people were traveling to their hometowns for celebration with their families. Travel demand, increasing dramatically during these festivals, leads to the high potential of highway accidents [4,5]. It has been found that drunk driving is a significant cause of

[†]Presented at the 2nd International Symposium on Construction Innovation Research & PhD Symposium: July 18-19, 2019.

traffic accidents. The percentage of alcohol-intoxicated drivers was reported to be higher during New Year and Songkran holiday months; therefore, alcohol-related traffic accidents increased dramatically during these holidays [5]. The vital mission of the Thai government is to provide safety and reduce accidents for road users on highway systems [6]. In order to cope with this problem, the Government has tried several measures to improve the situation. These strategies include creating awareness to the road users, controlling driving under the influence of alcohol, and increasing patrolling during special-event periods.

There is limited study on the application of data visualization and analytics with large data sets of highway accidents [7]. This paper aims to fill this technological gap by developing an accident data analysis system that is an advanced tool to facilitate visualization and analytics for accident data. Visualization allows users to understand data through the meaningful presentation of charts and maps, whereas analytics conduct aggregations, computations, and data reductions. Data visualization and analytics provide decision makers with well-informed, data-driven, and actionable insights that can be used to develop strategies and programs for improving road traffic safety [8,9].

The availability of open accident data provides an opportunity for further analysis, which can be used for business intelligence. This paper involves performing business intelligence to Thailand Accident Database during the New Year holiday in Thailand from 2008 to 2015. In this study, the processes of acquiring data from the open data portal of Thailand, integrating this data with other data by using Talend Data Integration, and designing a dashboard with Tableau were discussed in the following section. The paper then explained the dashboard design and ended with key conclusions from the study.

Business intelligence for road accidents

To stay relevant in the highly competitive environment, organizations, more than ever, need to reap benefits from its data, which can be analyzed to gain insight that can efficiently respond to environmental conditions that organizations encounter. Business Intelligence (BI) can provide such capability by acquiring and integrating data that come from multiple sources of data available within and out of the organizations in order to gain such insight by performing data analysis and supporting the decisionmaking process. BI also helps organizations plan business strategies to optimize their performance, reduce costs, and increase profit. Several benefits of BI have been highlighted in previous studies, including its ability to transform data into meaningful information for decision making, to integrate business information across organizations that improve transparency and speed up the decisionmaking process, to recognize patterns or trends for improving business forecasting, and to detect fraud [10].

According to Reinschmidt and Francoise [11], a BI system is "an integrated set of tools, technologies, and programmed products that are used to collect, integrate, analyze, and make data available." Most BI systems comprise a set of tools that transform data into information to be used for decision making. They refer to any set of software platforms that support organizations in collecting, storing, analyzing, and retrieving valuable information and knowledge to support business operations. The critical components of business intelligence are data warehouse, extraction-transformation-load tools (ETL), online analytical processing (OLAP), data mining tools, reporting tools, and visualization tools [12].

In order to build a BI system, data was first gathered by an ETL process and stored in the data warehouse. The database was then designed to facilitate the analysis and mining of the data using OLAP and data mining tools. Finally, data visualization and reporting tools were used to present the results of the analysis to BI users. The results can come in different forms ranging from standard reports to interactive charts that allow BI users to drill down into each transaction's specific detail.

BI can be used to predict the number of highway accidents at specific locations and times. It can recommend strategies to decrease the frequency and consequence of such accidents because the BI provides a quality of accident data by the ETL process and a precision analysis via OLAP. In this study, the BI was used to analyze highway accidents in Thailand to gain more data insights using visualization and analytics.

Data integration

The road accident database during the New Year holiday was downloaded from the Thailand government open data portal (data.go.th). The data set contains 214,949 records of accidents during the New Year holiday (28 December - 3 January) from 2008 to 2015. Data were recorded in the Thai language and included each injured person's details from the accident that occurred during the long public holiday in Thailand. The information, including in the dataset, indicates the accident's time and location, the vehicle used, and the injured person's demographic information. The schema of the dataset is illustrated in **Table 1**.

Attribute	Туре	Value	Details		
Year	Polynomial	2008, 2009,, 2015	The year when accident occurs		
Province ID	Polynomial	10, 11,, 97	ID of the province according to Ministry of Interior		
Province name	Polynomial	Bangkok, Chiangmai, etc.	Name of the province in Thailand (in the Thai language)		
Hospital name	String	Hospital name	Name of the hospital admitting the injured patient		
Date of accident	Polynomial	28, 29, 30, 31, 1, 2, 3	Day in which accident occurs from 28 December to 3 January		
Time of accident	Polynomial	Time	Time when the accident occurs		
Sex	Binomial	Male/Female	Gender of the injured person		
Age	Integer	Number	Age of the injured person		
Location of accident	Polynomial	Urban, rural, highway	Location of the road where the accident occurs		
Type of vehicle	Polynomial	Passenger car, bus, motorcycle, bicvcle, rickshaw	Type of the vehicle		
Other vehicles	Polynomial	Passenger car, bus, motorcycle, bicycle, rickshaw	Vehicle of other parties during the accident		
Result of treatment	Polvnomial	Died, recovered, transferred	Treatment results		
Days at hospital	Integer	Duration	Number of days that patients spend in hospital		
Drunk driving	Polynomial	Drunk, not drunk, not known	Alcohol-related accidents		
Safety measure	Polynomial	Safety belt, helmet, none	Type of protection		
Delivery	Polynomial	Ambulance, relatives, friends, neighbors	The person taking the injured patient to the hospital		

Table 1 Attributes of the accident dataset.

Another dataset that was adopted to use is the Thai province (city) database. It was shared in the Thailand developer forum in the form of a structured query language (SQL) statement file from Github. This database contains information about the location of latitude and longitude and the province's name in English; therefore, it can be merged with the accident database. Thus, the data become more useful in illustrating this information to international audiences. Figure 1 illustrates the Entity-Relationship (ER) model of the province database. SQL is a language used for the retrieval and modification of database entries [13]. In this study, the SQL statement file was written and then executed in the MySQL database resulting in all the provinces in both Thai language and English language to be included in the database. Figure 2 also explains the development process of the visualization system for Thailand road accident data. The process includes extracting the data from the original data source, transforming the mapping

Walailak J Sci & Tech 2020; 17(12)

data from different data sources, loading the data into the data warehouse, and visualizing data from the dashboard.



Figure 1 ER diagram of the Thai province database.



Figure 2 Development flowchart.

Extract

The accident database was given as a 17 MB excel file. Since the data were mostly recorded in the Thai language, another database is needed to map the Thai language into English, especially the province name (city name). The Thai province database contains information about the province's name in the English language and the latitude and longitude information of each province, which can be used to display the result on the map when implementing the dashboard. As a result, both accident and province data were extracted from both the excel file and MySOL database using Talend Data Integration, which is the tool used to perform the data extraction process. It was also used to display the content received from the excel file and MySQL database. After successfully extracted the data from the excel file and MySQL database, the next process is to transform the data into a format that can be easily used and implemented when working on the dashboard in Tableau.

Transform

As discussed in the earlier section, data in the accident database was recorded in the Thai language; therefore, data will first be transformed from the Thai language to English. This step can be done by using the tMap and Expression Builder in the Talend Data Integration tool. In this case, there are 3 sources of data, which are accident data, provinces containing the English name of the province, and province th comprising latitude and longitude information of the province.

During the mapping process, several data were transformed using the Expression Builder provided in the Talend Data Integration tool, as illustrated in Figure 3. These data include the year, date of the

accident, time of the accident, sex and age of the injured person, result of treatment, number of days in the hospital, accident location, and type of vehicle.



Figure 3 Transforming data input.

As for illustration, Figure 4 presented the step that the type of vehicle used during the accident was transformed using TRIM and equals the function of the StringHandling object. The TRIM function removes spaces (if any) at the beginning and the end of the data. The equals compare 2 strings, whether they are the same word or phrase. Finally, the IF statement was used to check if 2 words are the same (equal). If that is the case, that particular word will be changed with an English equivalent description. For example, the statement below checks if the word is "insulutur" if so, change it to "Motorcycle" before loading to the database.

Expression Builder							
Expression							
Wrap Undo(Ctrl + Z)	Clear						
StringHandling.TRIM(rowl.typeOfvehocle).equals("จักรมานมนต์") ? "Motorcycle" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกัง/แก้กซ์") ? "Passenger Car" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้") ? "Van" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้") ? "Van" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้") ? "Tuk Tuk" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้") ? "Bus" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้จา") ? "Bus" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังแก้") ? "Truck" : (StringHandling.TRIM(rowl.typeOfvehocle).equals("จากกังจา") ? "Truck" : "Others")))))))							
	×						
+ - * / == < <= != >= > and or not ()							

Figure 4 Vehicle type transformation.

The process was repeated for the year, date of the accident, time of the accident, sex and age of the injured person, result of treatment, number of days in the hospital, and accident location. Moreover, the English names of the province and latitude and longitude information of the province were added to the accident information by using the information from the Thai province database discussed earlier.

Load

After preparing the data transformation, the next step is to load the transformed data into a new database that will be used by the visualization tool to provide a dashboard. **Figure 5** illustrates the process of loading, which transformed accident information into a MySQL database. After successfully loading the data into MySQL, there were 214,949 records in the table indicating all the data were loaded and prepared to perform the next operation. **Figure 6** presents partial information of accident information in the database.



Figure 5 Loading accident data to MySQL database.

year	dateOfAccident	timOfaccident	sex	age	resultOfTreatment	accident_location	daysAt
2008	31	12:01-13:00	Male	39	Recovered	Others	
2008	28	11:01-12:00	Female	11	Recovered	Urban Area	
2008	31	09:01-10:00	Male	32	Recovered	Urban Area	
2008	29	Not Know	Female	17	Recovered	Urban Area	
2008	30	05:01-06:00	Male	44	Recovered	Urban Area	
2008	31	22:01-23:00	Male	23	Recovered	Urban Area	
2008	2	01:01-02:00	Female	18	Recovered	Urban Area	
2008	1	11:01-12:00	Male	19	Recovered	Others	
2008	29	23:01-24:00	Male	44	Recovered	Others	
2008	29	20:01-21:00	Male	30	Recovered	Urban Area	
2008	30	23:01-24:00	Male	36	Recovered	Highway	

Figure 6 Partial information in the newly created table after the ETL process.

Dashboard design

Many transport agencies adopt Tableau as visual analytic platforms. For example, the National Highway Traffic Safety Administration uses Tableau as an analytical visualization tool to provide insights into highway accidents across the USA [8]. Tableau was used to design a dashboard on accident data described earlier in this paper. Several representations of the data were created and then merged into a dashboard.

http://wjst.wu.ac.th

Spatial view of accident occurrences

In order to illustrate how an accident occurs throughout the country, the spatial view of the crash location is illustrated by showing the number of accidents on the map. To create a map, the latitude and longitude of provinces were used as the column and row, while the number of accidents for males and females for each province was also used as values to show on the map. **Figure 7** illustrates the map of Thailand, showing the number of accidents in each province. Due to the image's size, those numbers were removed automatically by Tableau, and the proportion of the pie chart represented the actual numbers.

As can be seen from the map, the provinces in the northern and northeastern parts had a higher number of accidents. Highway agencies are interested in provinces with high accident rates and therefore prepare a proactive response to accidents in such provinces. The proactive response includes necessary medical equipment and other necessary support in preparation at high-crash locations [14].



Figure 7 Spatial view and visualization of accidents across Thailand.

Temporal view of accident occurrences

Accidents commonly occur in the temporal periodical pattern, meaning that crash frequencies may increase in a specific time more than in other periods [15]. Thus, it is beneficial to analyze how accidents spread throughout the day. **Figure 8** illustrates the temporal view of accident occurrences, showing the number of accidents and the time that accidents occur. As can be seen, most accidents mainly occur between 4 and 8 p.m., especially in the nighttime from 6 to 7 p.m. This might be because it is more challenging to drive at the beginning of the darkness. Also, due to the lower traffic density, drivers tend to drive at higher speeds with increased carelessness, which can lead to accidents, which can lead to severe injuries [16]. As a result, speed limit control should be rigorously enforced during these hours. The platform can also inform road users to be cautious during the time span that the accidents occur frequently. The decision maker can use this information to improve the situation, such as increasing the number of polices to patrol during that time of the day or having hospital staff and ambulances ready during that particular period. The response time is vital for a medical emergency because the decreased

response time has been found to be associated with an increased survival rate of accident victims [17]. Thus, there should be more training for the medical staff and better ambulances with self-help equipment.

Accidents involving motorcycles occurred more than any other type of vehicle. This is consistent with previous studies that motorcycles have the highest frequencies of accidents. This might be due to higher numbers of intoxicated motorcyclists who were involved in traffic accidents. Brain injury, the most significant injury for motorcyclists, is related to not wearing a helmet. This information is useful for decision makers to promote the helmet's use to increase public safety. In Thailand, around one-third of the drivers and passengers, the helmet wearing rate is relatively low [18]. This implies that the helmet law enforcement is still inefficient. According to Jiwattanakulpaisarn, *et al.* [19], the helmet use tends to increase with the perceived risk of apprehension for helmet law non-compliance. Therefore, a helmet use campaign (e.g. more police checkpoints and random scheduling of enforcement activities) should be implemented to reduce motorcyclists' accident rates, especially during the long holiday. Ultimately, it is essential to inculcate helmet-wearing behavior into a motorcyclist attitude not only as mere law compliance but also as a safety mindset.



Figure 8 Time of accident and type of vehicle used.



Figure 9 Graph showing the number of dead victims from accidents.

http://wjst.wu.ac.th

Creating a graph showing number of casualties

Thai government gives priority to the urgent issue of the reduction of accidents. The Government has enforced traffic law and implemented road safety policies to reduce the number of casualties from road accidents, especially during the holiday. Therefore, it is imperative to illustrate this information in the dashboard. **Figure 9** presents the graph showing the number of people who lost their lives during the accident. The dashboard has been designed to show the number of casualties for several criteria filtered by the result of treatment, used as the value to display on the graph. The result of treatment can be filtered to include only the number of the person who lost their lives during the accident that are categorized into patients who die at the scene after the rescue services arrive at the scene, or after the patients receive emergency care, and includes death at the hospital in different periods. Those victims who recovered from the accident were not included in this graph.

Creating a dashboard

Finally, the dashboard for illustrating accident information during the New Year holiday in Thailand was created. The 3 charts that were discussed earlier were included in the dashboard, and all the information can be filtered by year and province. These 2 parameters were used as filters because the Thai government wanted each province to reduce the number of accidents and casualties during the extended holiday period. **Figure 10** illustrates the dashboard to show this information.



Figure 10 Dashboard illustrating the visualization and analytics of accident occurrences.

Walailak J Sci & Tech 2020; 17(12)

http://wjst.wu.ac.th

It shows that most traffic accident victims are males, typically the majority of drivers [20]. Male drivers tend to drive faster than female drivers with an average speed higher than 90 km/h, exceeding the speed limit of intercity highways [16]. Also, drunk driving among male drivers is considerably higher than among female drivers, who are limited to alcohol consumption by Thai culture [5]. Therefore, male traffic accident victims are more likely to have a higher risk of injury than females.

Conclusions

Thailand has become the top country with the highest death rate in a road accident. This ranking exemplifies an unsafe traffic condition in Thailand. The increased travel demand during festivals leads to more frequencies of highway accidents. The Government by the Ministry of Public Health published the information on road accidents during the New Year holiday when accidents occurred the most. This information was used along with the database about Thailand's provinces to illustrate how an accident occurs in various parts of the country. This paper proposed a framework for conducting data visualization and analytics in both spatial and temporal dimensions in order to illustrate when and where the accidents occur. The accident data and provincial data were combined by using the Talend Data Integration tool and were transformed from the Thai language to the English language using the Expression Builder provided by Talend Data Integration Software. The combined data was then loaded into the MySQL database for data visualization using Tableau.

The dashboard to display the accident information was designed and created by using Tableau. The interactive dashboard was developed for the near-instant rendering of queries through charts and maps. The dashboard illustrates the accident's information on the map using latitude and longitude information from the provincial database and displaying the time of the accident and the number of casualties. The information in the dashboard can be filtered by province and year.

This study's findings are useful for improving data management, which can apply a database for accident analysis. By integrating all data, this platform can be used to retrieve, process, and graphically analyze the accident. It applies the diagnosis rules more realistically. As such, the reasoning is determined by identifying accident causes from a set of observations to establish the most likely policy-maker strategies. This system should be adopted by related departments, which can be used for visualization and analytics of accident data to provide a mechanism to formulate strategy options and formulate appropriate contingency plans to improve the accident situation. This paper has proposed the road safety policies that can be implemented to reduce fatal accidents for the related departments. The visualization of accidents in both spatial and temporal dimensions provides insight for drivers to raise accident awareness during the New Year holiday.

This data visual analytics framework employed structured data in a tabular form. The limitation of the framework is that it is unable to deal with non-structured data such as video and images. Future work should also involve the ability to manage non-structured data. In addition, future work should also integrate the geometric design of the highway (e.g. horizontal and vertical alignment, number of lanes, superelevation) to analyze in spatio-temporal dimension to locate crash hot spots.

References

- [1] World Health Organization. Global Status Report on Road Safety 2015. Geneva, Switzerland, 2015.
- [2] J Burton. Countries with the most car accidents, Available at
- https://www.worldatlas.com/articles/the-countries-with-the-most-car-accidents.html, accessed 2018.
 [3] W Wipulanusat and J Sunkpho. A Mobile platform for roadway incident documentation: A case study of intercity motorways. The Eastern Asia society for transportation studies volum 7. *In*: Proceedings of the 8th Eastern Asia society for transportation studies, Surabunya, Indonesia. 2009, p. 362.
- [4] B Chaichannawatik, K Kanitpong and T Limanond. Departure time choice (DTC) behavior for intercity travel during a long-holiday in Bangkok, Thailand. J. Adv. Transport. 2019; **2019**, 1-11.

- [5] M Seesen, P Siviroj, R Sapbamrer and S Morarit. High blood alcohol concentration associated with traumatic brain injury among traffic injury patients during New Year festivals in Thailand. *Traffic Inj. Prev.* 2019; **20**, 115-21.
- [6] W Wipulanusat, J Sunkpho, A Jotisankasa, S Chayanan and P Sangsuwan. The implementation of business process reengineering: A case study of roadway incident data management. *Eng. J.* 2013; 9, 1-10.
- [7] MJ Islam, A Sharma and H Rajan. A Cyberinfrastructure for big data transportation engineering. J. Big Data Anal. Transport. 2019; 1, 83-94.
- [8] Y Adu-Gyamfi. PU-enabled visual analytics framework for big transportation datasets. J. Big Data Anal. Transport. 2019; 1, 147-59.
- [9] VH Trieu. Getting value from Business Intelligence systems: A review and research agenda. *Decis. Support Syst.* 2017; **93**, 111-24.
- [10] E Rubin and A Rubin. The impact of Business Intelligence systems on stock return volatility. *Inform. Manag.* 2013; **50**, 67-75.
- [11] J Reinschmidt and A Francoise. Business Intelligence Certification Guide. IBM, California, 2000.
- [12] CM Olszak and E Ziemba. Critical success factors for implementing business intelligence systems in small and medium enterprises on the example of upper Silesia, Poland. *Interdiscipl. J. Inform. Knowl. Manag.* 2012; 7, 129-50.
- [13] JK Simandl, AJ Graettinger, RK Smith, S Jones and TE Barnett. Making use of big data to Eealuate the effectiveness of selective law enforcement in reducing crashes. *Transport. Res. Rec.* 2016; 2584, 8-15.
- [14] S Jomnonkwao, S Uttra and V Ratanavaraha. Forecasting road traffic deaths in Thailand: Applications of time-series, curve estimation, multiple linear regression, and path analysis models. *Sustainability* 2020; 12, 395-412.
- [15] X Fan, B He and P Brézillon. Context-aware big data analytics and visualization for city-wide traffic accidents. Volume LNCS 10257. *In*: Proceedings of the 10th International and Interdisciplinary Conference on Modeling and Using Context, Paris, France, 2017, 395-405.
- [16] K Kanitpong, P Jiwattanakulpaisarn and W Yaktawong. Speed management strategies and drivers' attitudes in Thailand. *IATSS Res.* 2013; 37, 39-48.
- [17] H Jaldell, P Lebnak and A Amornpetchsathaporn. Time is money, but how much? The monetary value of response time for Thai ambulance emergency services. *Value Health* 2014; **17**, 555-60.
- [18] J Kumphong, T Satiennam and W Satiennam. The determinants of motorcyclists helmet use: Urban arterial road in Khon Kaen City, Thailand. J. Saf. Res. 2018; 67, 93-7.
- [19] P Jiwattanakulpaisarn, K Kanitpong, S Ponboon, N Boontob, P Aniwattakulchai and S Samranjit. Does law enforcement awareness affect motorcycle helmet use? Evidence from urban cities in Thailand. *Glob. Health Promot.* 2013; 20, 14-24.
- [20] P Woratanarat, A Ingsathit, P Chatchaipan and P Suriyawongpaisal. Safety riding program and motorcycle-related injuries in Thailand. *Accid. Anal. Prev.* 2013; **58**, 115-21.