Journal of Information and Knowledge Management (JIKM) Vol 2 Special Issue (2023)

The Development of CrackDown: A Prototype of Data Visualisation System on Crime Rate Trends in Malaysia

Maisarah Ros Mahadi, Norisan Abd Karim, Jasber Kaur GS, and Fauziah Redzuan

College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Received Date: 20 August 2023
Acceptance Date: 15 September 2023
Published Date: 1 November 2023

Abstract. The use of a crime dashboard is important as it should be widely spread for citizens to understand and track crime trends in Malaysia. There is currently a lack of interactive dashboards about crime rates in Malaysia for Malaysian citizens. The Department of Statistics Malaysia (DOSM) only provided read-only files and out-of-date information available for everyone to use. The dashboard should integrate data visualization, forecasts, and educational materials to provide Malaysian residents with a better understanding of the situation. This article focuses on the steps involved in constructing a CrackDown interactive data visualization dashboard. This research will benefit its stakeholders such as Malaysian citizens, Royal Malaysia Police (RMP), and researchers. The aim is to create an interactive dashboard for Crackdown using a combination of Agile and OSEMN methodology, with visualization and relevant information. The design of the dashboard uses feature visualization, and related articles that are educational for the user. The research also applied the data design process which involved the extract, transform, and load (ETL) and helped in obtaining a good data quality. Based on the findings, despite a decrease in crime patterns from 2017 to 2022, there are still concerns among citizens. Selangor was the state with the highest crime cases followed by Kuala Lumpur, Johor, Sarawak, and Kedah. Meanwhile, the most frequent crimes in Malaysia are motorcycle theft, followed by stealing, home breaking, car theft, and unarmed bandits. From the result, Crackdown intends to enable Malaysian citizens to track crime rates in Malaysia based on different categories and raise awareness and knowledge of crime prevention and guidance for a more secure future.
Keywords: Crime, data visualization, forecast, knowledge management.

1 Introduction

Crime is an illegal act or behaviour for which a person may receive criminal consequences. It is an act that is harmful to a community, society, or the state as well as to some individuals. Depending on the seriousness of the offense, a crime might result in anything from a penalty to an extended prison sentence. Despite a recent decline in

crime, Malaysia's crime rate is still quite high when compared to other Southeast Asian nations. With high rates of crimes reported in major cities, crime is still a concern in Malaysia (Mulok et al., 2018). There is a lack of interactive data visualization available for users to track crimes rate and predict crime trends. Additionally, only a few features in the Malaysian dashboard crime which make it less functional and only includes data up to 2021, which is not the most recent year (DOSM, 2023).

Action must be taken to prevent crime and raise citizen understanding of the problems with crime in the country. The objectives that will be accomplished by this research is firstly to identify the data requirement for visualizing crime rates trends on the Crackdown. The second objective is to design a visualization dashboard for Crackdown and lastly to develop a visualization dashboard for Crackdown. In this research, dashboard of crime in Malaysia will be develop in order to display crime statistics and patterns in Malaysia. Additionally, the dashboard makes use to predict crime over the next few years. The public and many users can use this dashboard to monitor trends in crime. Using this dashboard, citizens may raise awareness and take early precautions based on dashboard predictions as well as be pre-pared for any criminal activity.

2 Literature Review

2.1 Big Data and Characteristics

Big data is a large and complex data sets collected from a variety of sources, such as social media and other digital technologies. Big Data has been developing in recent years, which has encouraged the development of brand-new, highly advanced techniques for gathering, storing, and analysing massive data sets (Xu, 2021). There are 5 characteristics of big data which are volume, variety, veracity, value and velocity. Volume describes the huge amount of data that to create, store, and process. Variety describes the various sorts of data that are produced, such as unstructured data, semi-structured data, and structured data (Dhar & Liu, 2018).

The term velocity describes the rate at which data is collected and must be processed in real-time to satisfy the needs of various applications and business processes. For businesses that rely on big data for decision-making and must verify that the data they utilise is correct and reliable, veracity is crucial. Factors including measurement error, missing data, and inaccurate data entry can affect the quality of data, and these problems might jeopardize the reliability of the outcomes of big data analysis (Wang, 2018). Value is the insights, information, and business value that may be obtained from the data if it is handled, processed, and evaluated correctly.

2.3 Data Visualisation

Data visualisation is significant for data analysis, which lets users to understand data relationships and make appropriate decisions (Wang et al., 2018). Dashboards system provide a creative and interesting way to learn new things, and their design is the most important factor since, without it, the data visualisation would be unable to fulfil its

objectives of quickly conveying messages, which will limit the knowledge that can be learned. The dashboard can use a variety of data visualisation techniques, including pie charts, bar graphs, and tree maps, which can then be used to track and monitor using visual representations. The purpose of the visualisation techniques is to make patterns, connections, and trends in data visible that might not be instantly seen from looking at original data.

A bar chart is a technique of data visualization in which data is displayed as a sequence of bars, each bar representing a point for a specific or group. Each bar's height refers to the number of the category or group it is meant to represent. The next technique is a pie chart, which shows data as a circular graphic with sections within it. The size of each slice is dependent on the value of each slice's respective category or group, which is represented by each slice. Pie charts are frequently used to show categorizable data and to illustrate the percentage of each category compared to the total (Wang et al., 2018). Pie charts can be used to present large and complicated data sets in big data visualization in a clear and understandable decision way.

2.4 Crime

A person may face criminal penalties for engaging in an illegal act or behaviour, which is referred to as a crime. It's an act that harms certain people as well as a community, society, or the state. The punishment for a crime might range from a fine to a lengthy prison term, depending on how serious the offence was. A person may commit a crime for a variety of reasons without thinking about the penalties. In Malaysia, Violence-related and property crimes are the two main categories of crime (Dass, 2019).

Table 1: Selected World Crime Index, 2018

Rank	Country	Crime Index	Safety Index
1	Venezuela	82.59	17.41
2	Papua New Guinea	79.95	20.05
3	Honduras	78.89	21.11
4	South Africa	75.71	24.29
5	Trinidad And Tobago	72.22	27.78
6	Brazil	70.55	29.45
7	El Salvador	69.72	30.28
8	Bangladesh	68.52	31.48
9	Namibia	68.22	31.78
10	Syria	67.74	32.26
11	Jamaica	66.87	33.13
12	Puerto Rico	64.94	35.06
13	Peru	63.91	36.09
14	Nigeria	63.12	36.88
15	Malaysia	63.05	36.95

Note. Adapted from Multi-Factor Crime in Malaysia, 1980–2013: Bounds Testing of Level Relationships and Granger Non-Causality Analysis, by Mulok, D., Kogid, M., Asid, R., & Lily, J, 2018. Copyright 2018 by Numbeo.

In 2018, Malaysia was noted as having one of the highest crime indices worldwide with an index score of 63.05, it was rated 15th and had one of the lowest security indexes in the world, 36.95 (Mulok et al., 2018). Comparing Malaysia to other nations, which were comparatively safer and had lower crime indices, it was found that Malaysia had the third highest crime and least secure index at the Asian level. In addition, Malaysia, which is considered the least secure country in Southeast Asia compared to

other ASEAN nations, placed top on the list of the region's worst crime indices at the regional level (Mulok et al., 2018).

2.5 Factors of Crime

The incidence of crime in Malaysia is driven by a variety of factors. Among the critical ones are unemployment and education opportunities. In Malaysia as well as other nations, unemployment may contribute to the prevalence of crime (Anser et al., 2020). People who can't find a job may suffer financially and feel like they can't make a decent living, which can result in feelings of hopelessness. Some people may commit crimes because of these emotions.

Education plays an important role in one's individual and community development. People with a lack of education may struggle to solve problems and make decisions, which puts them at risk of criminal temptation and unhealthy peer pressure. Individuals who do not have access to a good education may have fewer opportunities and are therefore more likely to commit criminal conduct (Zakaria et al., 2022). People may be more likely to commit crimes without fully understanding the potential consequences if they lack a proper education.

3 Prototype Design and Framework

The agile and OSEMN methodology combination displayed in Figure 1 was applied where the data analysis and visualization tools were used.

In the figure below, the methodology is illustrated for each phase. For OSEMN methodology, the data science life cycle is shown in a clear and simple manner (Hotz, 2023). The phases that are being presented follow a logical order that is representative of the overall life cycle of data science. OSEMN creates a taxonomy to help clarify how data science research progresses, which leads to a shared understanding. However, for the most of research in the actual world, OSEMN is insufficient as a framework. It lacks a lot of key details needed to carry out research. Another framework was added to the phase described in the above diagram to show the details of the research (Li et al., 2021). Combining OSEMN with an Agile framework is an additional detail built on top of the base framework. The Agile approach works because it can handle Big Data Analytics needs including the requirement for fast analytics that can be used to design and develop processes, and visualization that is combined with the need for insight collection.

Besides, Agile methodology combines sequential and iterative methodologies. It is iterative because it intends to build on the results of one iteration in subsequent iterations. An iterative process allows us to identify problems and respond to them very quickly. Because the work is given at varying stages during the research, it is sequential. Agile enables quickly resolving pressing problems, which saves both money and time. Next, users participate more actively and valuable input at every stage of the process. The agile methodology represents the dashboard system phase while the data pipeline

production is represented by the OSEMN methodology to create complementary data science research. Table 2 below describes the phases involved in this research.

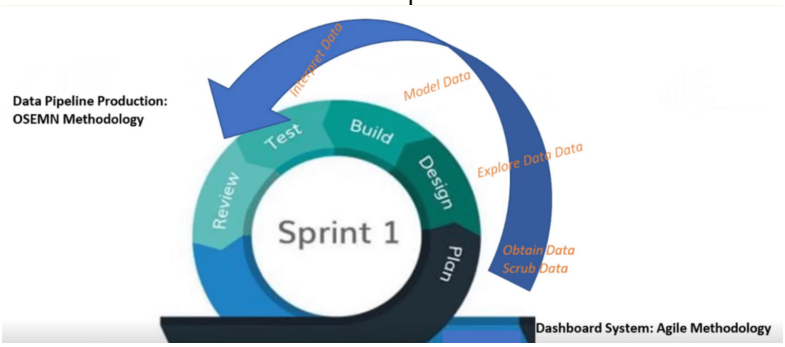

Figure 1 Shows Agile combined with OSEMN Model Methodology
Source: Census.gov

Table 2: Description of Methodology Phases

Agile	OSEMN	Description
Plan	Obtain Data (Identification of data requirement)	Collecting the crime-related dataset from the source, Royal Police Malaysia (RMP), and identifying data requirement
	Scrub Data (Data Cleaning)	Clean raw data by modifying and deleting data that is inaccurate, duplicated, or incomplete inside a dataset.
Design	Explore (Storyboard Design)	To better comprehend the data and its story, the dataset was explored the data using various techniques such as looking at the initial few rows and identifying similar columns. The storyboard also is built to show the usage of the CrackDown system.
Build	Model (Prototype of Data Visualization)	The cleaned dataset is used as model data to develop the CrackDown dashboard using Tableau.
Test	Interpret	The Crackdown dashboard system is tested using Google form feedback.
Review		CrackDown is reviewed based on user testing to be improved

3.1 Identification of data requirement

The complexity of the research was taken into account while developing the Crackdown dashboard. Notebook Jupyter is used for the research's level of complexity by using Phyton. Volume is one of the 5Vs qualities that make up big data, as was previously discussed. Naturally, the size of the raw data itself will come up first if big data is listed or usually mentioned. The Royal Malaysian Police (RMP) provided the raw datasets for this research. The datasets are in CSV format and span a 6-year period from

2017 to 2022. Additionally, the research's focus is on the Malaysia region, where each case's number is stated in terms of the state and kind of crime. Several steps must be completed in this area to guarantee the consistency of the data. States, crime types, the category of violent or property crimes, and the year are the attributes in the datasets. All 14 states in Malaysia are included in each dataset for each year's crime. The states include Selangor, Johor, Kedah, Kelantan, Melaka, Pahang, Kuala Lumpur, Pahang, Negeri Sembilan, Perak, Perlis, Sabah, Sarawak, and Terengganu.

3.2 CrackDown Dashboard Data Design

The ETL process is used in this research which can ensure the accuracy, completeness, and consistency of the data as well as prevent errors and duplication in the process (Wang et al. 2022). The process consists of extract, transform, and load (ETL) in which the data extraction process involves the the collection of data that was obtained from the Royal Police Malaysia (RMP) and preparation for big data transformation. It is retrieved in raw forms in an Excel file. The file goes through local storage and the selection of different data is done according to certain criteria to find the most suitable data related to the research and then transfers the data to the file.

Then, the transformed data for the transformation process occurs in the big data platform and tools stage. This process includes cleansing, correction, data integrity, and formatting the extracted datasets in the previous process. These processes are important to ensure a high quality and standardized data as well as easier for the data management. Lastly the data in the analytical data store which the data will be loaded into the database. The research has flexibility criteria which is important for the ETL process. It should be able to adapt to requirements and make it simple to include new data sources. However, the process in this research will be flexible but systematic according to the best and most suitable circumstances.

3.3 Data Cleaning

This section covers a certain critical part of the ETL process which is the data transformation stage using Jupyter Notebook. There are also storyboard diagrams that serve as the foundation for the dashboard's development. The diagrams serve as the basis for designing each and every system flow. Tableau is used to develop the interface and implement the features.

Based on the figure below, the list of column names that are common to all data frames can be identified by using a set.intersection() function. For merging several data frames, identifying the common columns is crucial because it indicates compatible data that can be joined based on those similar columns. This can ensure that the merge procedure is carried out appropriately and consistently. The function (len(common_columns) > 0) prints the common column names when identical columns are recognized, showing that the datasets can be combined using these columns.

```
# Get the common columns among all the DataFrames
common_columns = list(set.intersection(*[set(df.columns) for df in dfs]))

# Check if there are common columns
if len(common_columns) > 0:
    print("The datasets can be merged using the common columns:")
    print()
    print(common_columns)
else:
    print("The datasets do not have any common columns to merge.")

# Count the number of common columns
count = len(common_columns)
print()
# Print the count
print("Number of common columns:", count)
```

```
The datasets can be merged using the common columns:

['MELAKA', 'JENIS JENAYAH', 'PULAU PINANG', 'TERENGGANU', 'TAHUN', 'SABAH', 'KELANTAN', 'JUMLAH', 'PAHANG', 'SELANGOR', 'PERLI
S', 'KEDAH', 'SARAWAK', 'KUALA LUMPUR', 'PERAK', 'JOHOR', 'NEGERI SEMBILAN']

Number of common columns: 17
```

Figure 2 Finding Common Column

The following stage involves vertically combining many dataframes to create a merged dataframe (merged_df) by using pd.concat() function with the axis=0 argument. Reset-ting the index of the combined dataframe after concatenation with reset_in-dex(drop=True) ensures a new continuous index. The index col-umn is excluded from the saved file by specifying index=False. The data-frames can be combined vertically using this code, resulting in a dataset that is in one simple form.

```
# Concatenate the DataFrames vertically
merged_df = pd.concat(dfs, axis=0)

# Reset the index of the merged DataFrame
merged_df = merged_df.reset_index(drop=True)

# Save the merged data to a new file 'merged_file.csv'
merged_df.to_csv('merged_file.csv', index=False)

merged_df.head(30)
```

	TAHUN	JENIS JENAYAH	JOHOR	KEDAH	KELANTAN	KUALA LUMPUR	MELAKA	NEGERI SEMBILAN	PAHANG	PERAK	PERLIS	PULAU PINANG	SABAH	SARAWAK	SELANGOR	TERENGGANU	JUMLAH
0	2017	BUNUH	66	21	13.0	37.0	7.0	14.0	16.0	35.0	2.0	17.0	36	27.0	83	5.0	379.0
1	2017	ROGOL	196	119	114.0	132.0	69.0	91.0	163.0	95.0	30.0	80.0	211	150.0	321	64.0	1835.0
2	2017	GENG ROMPAK BERSENJATA	1	2	2.0	4.0	1.0	2.0	1.0	4.0	0.0	0.0	0	2.0	12	0.0	31.0
3	2017	GENG SAMUN TIDAK BERSENJATA	1220	333	141.0	1984.0	349.0	392.0	201.0	426.0	25.0	481.0	163	238.0	3747	79.0	9779.0
4	2017	ROMPAK BERSENJATA	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1.0	2	0.0	3.0
5	2017	SAMUN TIDAK BERSENJATA	481	157	78.0	1191.0	240.0	144.0	87.0	200.0	28.0	225.0	121	90.0	1197	76.0	4315.0
6	2017	KECEDERAAN	614	364	252.0	651.0	176.0	241.0	188.0	380.0	47.0	275.0	230	368.0	1108	130.0	5024.0
7	2017	CURI	2032	1079	547.0	3121.0	446.0	1121.0	707.0	1003.0	181.0	850.0	2518	1283.0	3951	365.0	19204.0
8	2017	KECURIAN KERETA	1097	378	722.0	1548.0	112.0	170.0	216.0	321.0	11.0	357.0	244	574.0	2586	146.0	8482.0
9	2017	KECURIAN MOTOSIKAL	4104	2846	1983.0	3328.0	1101.0	839.0	1173.0	1912.0	186.0	2322.0	666	2258.0	7945	914.0	31577.0
10	2017	KECURIAN KENDERAAN BERAT	274	81	98.0	291.0	27.0	64.0	56.0	81.0	5.0	65.0	116	75.0	844	22.0	2099.0
11	2017	PECAH RUMAH	1211	1350	570.0	1188.0	567.0	895.0	794.0	859.0	87.0	868.0	1908	1314.0	4133	456.0	16200.0
12	2018	BUNUH	43	9	6.0	33.0	13.0	9.0	8.0	33.0	1.0	21.0	37	17.0	91	2.0	323.0
13	2018	ROGOL	187	119	101.0	112.0	51.0	97.0	109.0	104.0	30.0	72.0	178	131.0	294	63.0	1648.0
14	2018	GENG ROMPAK BERSENJATA	2	2	4.0	3.0	1.0	0.0	2.0	1.0	0.0	0.0	1	3.0	11	1.0	31.0

Figure 3 Merge Data Frames

The steps to rename columns for the data frame 'merged_df' using the rename() method are shown in the figure above. First, the column 'TAHUN' is changed to 'YEAR' using the syntax 'TAHUN': 'YEAR'. Then, using 'JENIS JENAYAH': 'TYPE', the column 'JENIS JENAYAH' is renamed to 'TYPE'. The Data Frame merged_df is modified in place when the inplace=True argument is used to ensure that the rename is carried out. In order to verify the updated column names, print(merged_df.columns.values) is used to display the column names of the data frame. This step is done to ensure that all the datasets used have consistency and uniformity which is the English version.

```
# Rename the column
merged_df.rename(columns={'TAHUN': 'YEAR'}, inplace=True)
merged_df.rename(columns={'JENIS JENAYAH': 'TYPE'}, inplace=True)

#check columns name
print(merged_df.columns.values)
```

```
['YEAR' 'TYPE' 'JOHOR' 'KEDAH' 'KELANTAN' 'KUALA LUMPUR' 'MELAKA'
 'NEGERI SEMBILAN' 'PAHANG' 'PERAK' 'PERLIS' 'PULAU PINANG' 'SABAH'
 'SARAWAK' 'SELANGOR' 'TERENGGANU']
```

Figure 4 Rename Column Name

The figure below shows the installation and use of the GoogleTrans library for text translation. Firstly, the library using pip is installed and the 4.0.0-rc1 version specification. The code imports the Translator class from googletrans once the library has been installed. Text translation between languages is possible with the Translator class. It is used in this code to translate words from Malay to English. The language translation process uses the Translator class from the googletrans package which provides multilingual text processing and analysis in Python programs.

```
# Initialize the translator
translator = Translator(service_urls=['translate.google.com'])

# Function to translate a single string from Malay to English
def translate_text(text):
    translation = translator.translate(text, src='ms', dest='en')
    return translation.text

# Translate the 'Crime Type' column
merged_df['TYPE'] = merged_df['TYPE'].apply(translate_text)

# Display the translated DataFrame
print(merged_df.to_string(index=False))
```

YEAR	TYPE	JOHOR	KEDAH	KELANTAN	KUALA LUMPUR	MELAKA	NEGERI SEMBILAN	PAHANG	PERAK	PERLIS	PULAU PINANG	SABAH	SARAWAK	SELANGOR	TERENGGANU
2017	Kill	66	21	13	37	7	14	16	35	2	17	36	27	83	5
2017	Rape	196	119	114	132	69	91	163	95	30	80	211	150	321	64
2017	Armed robbery gang	1	2	2	4	1	2	1	4	0	0	0	2	12	0
2017	Unarmed robbery gang	1220	333	141	1984	349	392	201	426	25	481	163	238	3747	79
2017	Armed robbery	0	0	0	0	0	0	0	0	0	0	0	1	2	0
2017	Unarmed robbery	481	157	78	1191	240	144	87	200	28	225	121	90	1197	76

Figure 5 Translate Malay Data to English

In this step, all 14 columns of the states will be transformed into a single column name ‘STATE’. The purpose of this process is to enable the recognization of the state as an attribute and make it easier when performing the visualization of the data step by using Tableau as the tool. The dataframe merged_df is broken down in the figure above to convert the states into a single column. The id_vars parameter of the melt() function is set to ['YEAR', 'TYPE'], meaning that the 'YEAR' and 'TYPE' columns should continue to be used as identifier variables. One column labeled "STATE" is created by reshaping the remaining columns, which represent the states. A new column called "CASES" contains the values of these melted columns.

```
# Melt the DataFrame to convert states into a single column
transformed_df = merged_df.melt(id_vars=['YEAR', 'TYPE'], var_name='STATE', value_name='CASES')

# Save the transformed dataset to a new CSV file
transformed_df.to_csv('transformed_dataset.csv', index=False)

# Display transformed dataset
transformed_df.head(40)
```

	YEAR	TYPE	STATE	CASES
0	2017	KILL	JOHOR	66
1	2017	RAPE	JOHOR	196
2	2017	ARMED ROBBERY GANG	JOHOR	1

Figure 6 Change Structure of Data Frame

Lastly, categorize_crime classifies crime types based on specified lists of violent crimes and property crimes shown above. The function accepts a crime type as input and determines whether it fits any of the crime types in the lists of property crimes or crimes involving violence. It returns the category "VIOLENCE" if it matches a crime involving violence and "PROPERTY" if it matches a property crime. It returns 'Other' if none of the predefined categories apply. The function is then applied to the data frame transformed_df's 'TYPE' column using the apply() method, resulting in the creation of a new column called 'CATEGORY'.

```
# Define a function to categorize crime types
def categorize_crime(crime_type):
    violence_crimes = ['KILL', 'RAPE', 'ARMED ROBBERY GANG', 'UNARMED ROBBERY GANG', 'ARMED ROBBERY', 'UNARMED ROBBERY', 'INJURY']
    property_crimes = ['STEAL', 'CAR THEFT', 'MOTORCYCLE THEFT', 'THEFT OF HEAVY VEHICLES', 'BURGLARY']

    if crime_type in violence_crimes:
        return 'VIOLENCE'
    elif crime_type in property_crimes:
        return 'PROPERTY'
    else:
        return 'Other'

# Apply the function to create a new column
transformed_df['CATEGORY'] = transformed_df['TYPE'].apply(categorize_crime)

# Display the modified DataFrame
transformed_df.tail(30)
```

	YEAR	TYPE	STATE	CASES	CATEGORY
978	2020	INJURY	TERENGGANU	118	VIOLENCE
979	2020	STEAL	TERENGGANU	847	PROPERTY
980	2020	CAR THEFT	TERENGGANU	56	PROPERTY
981	2020	MOTORCYCLE THEFT	TERENGGANU	313	PROPERTY
982	2020	THEFT OF HEAVY VEHICLES	TERENGGANU	4	PROPERTY
983	2020	BURGLARY	TERENGGANU	601	PROPERTY
984	2021	KILL	TERENGGANU	4	VIOLENCE
985	2021	RAPE	TERENGGANU	68	VIOLENCE

Figure 7 Categorize Crimes

3.4 Storyboard Design

The storyboard helps in explaining the purpose of the Crackdown system and the importance of researching country crime trends. In addition, it demonstrates how users can examine the system to become more aware of criminality. Finally, it demonstrates how users may access and interact with various capabilities.

Figure 8 Storyboard

3.4 Prototype of Data Visualisation

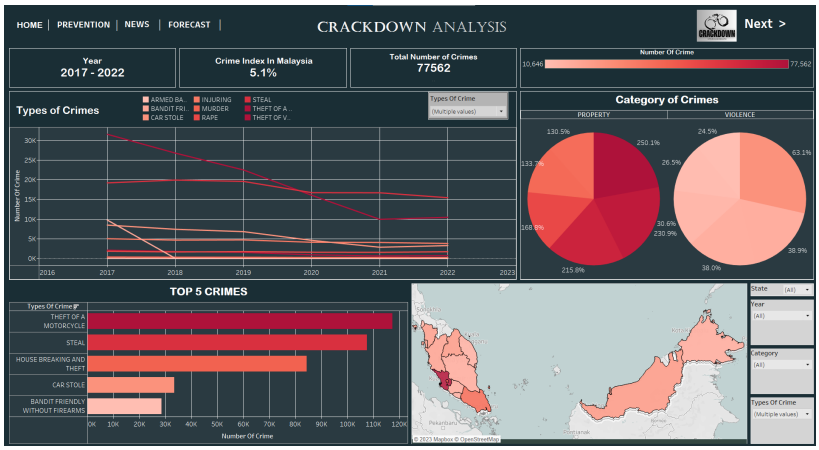

Figure 9 Main Dashboard

The description of the information and the functions of the main dashboard is described in the table 3 below:

Table 3: Main Dashboard Description

Function	Description
Filters	- In the Main dashboard page, there are 5 filters that can be manipulated by the user. - For the map crime visualization, the filters can be manipulated based on the state, year, types of crime, and category while the graph is by type of crimes
Navigation Button	This button is used to ensure the users know the current page and navigate easily between pages
Charts	Types of Crime: - The graphs show different kind of case numbers based on the types of crime and it is represented by a different colour for each crime type Category of Crime - This pie chart shows the average number of crimes for the two categories of crime and year and is represented by a different color for each year Top 5 Crimes: - The graph shows the highest 5 ranking types of crimes in Malaysia for the year 2017 to 2022 as it is very important for users to acknowledge. - The graph shows the total number of cases and each crime type is represented by a different colour. - The graph shows the highest 5 ranking types of crimes in Malaysia for the year 2017 to 2022 as it is very important for users to acknowledge. - According to the findings, the most common crimes are motorcycle theft (117,317 cases) followed by stealing (107,569 cases), home breaking (84,416 cases), car theft (33,431 cases), and unarmed bandits (28,485 cases). Crime Map - The crime map shows the number of crimes, year, state and category which in a map visualisation - All the 14 states in Malaysia include Selangor, Johor, Kedah, Kelantan, Melaka, Pahang, Kuala Lumpur, Pahang, Negeri Sembilan, Perak, Perlis, Sabah, Sarawak, and Terengganu. - The top 5 most common states with crimes were Selangor (114,332 cases), Kuala Lumpur (57,337 cases), Johor (48,503 cases), Sarawak (32,528 cases), and Kedah (31,431 cases) - The number of crimes can be clearly seen for people who prefer to analyze based on the map

4 Conclusion

The CrackDown dashboard, which uses data visualization, can inform Malaysian citizens about the country's crime statistics and encourage them to be aware of staying safe from criminal activity. It is because this research focuses on crime for each state, allowing citizens to track how crime is doing at the state level in their area. In the research methodology section, this paper only covers the plan to build phase in detail. Finding the data needed for the dashboard development is one of the goals of this research to make sure it is successful. By contacting the Royal Police Malaysia (RMP) and obtaining the necessary information, the objective is accomplished. Due to the sensitivity of crime data, this research uses historical crime data from the Royal Police Malaysia (RMP) where the cases for the consent of persons involved cannot be acquired. However, the dataset usage corresponded with all relevant legal and ethical standards. Users need to be aware of the possibility of bias based on unforeseen incidents, such as the underreporting of some crimes and the over-policing of specific districts. A review of the literature was conducted with the goal of developing the dashboard by locating articles or other publications. For the dashboard's interface, the acquired articles are examined to develop some designs for the dashboard during the designing step. Finally, the dashboard's development, it is made as a guide for developing the dashboard based on the prototype and design elements created for the previous goal. Based on the findings, the most commonly occurring crimes from 2017 to 2022 are motorcycle theft, stealing, home breaking, car theft, and unarmed bandits. Whereas, Selangor, Kuala Lumpur, Johor, Sarawak, and Kedah were the top five most common states with crimes in Malaysia. Following that, CrackDown may enlighten Malaysian citizens about the country's crime statistics and urge them to be aware of how to keep safe from criminals.

Acknowledgments

The authors would like to thank Universiti Teknologi MARA for the research support and valuable support.

References

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big Healthcare data: Preserving Security and Privacy. *Journal of Big Data*, *5*(1). Retrieved from https://doi.org/10.1186/s40537-017-0110-7

Anser, M.K., Yousaf, Z., Nassani, A.A. *et al.* Dynamic linkages between poverty, inequality crime, and social expenditures in a panel of 16 countries: two-step GMM estimates *Economic Structures* 9, 43 (2020). Retrieved from https://doi.org/10.1186/s40008-020-00220-6

Santhana Dass, R. A. (2019). Crime Trends and Patterns in Malaysia. Kyoto Review of Southeast Asia, Trendsetters, August 2019. Retrieved from https://kyotoreview.org/trendsetters/crime-trends-andpatterns-in-malaysia/

Department of Statistics Malaysia. (2023). Violent and property crime statistics for Malaysia, 2009-2018. https://open.dosm.gov.my/violent-property-crime

Dhar, V., & Liu, X. (2018). Big data: A review. Information & Management, 55(1), 1-14. https://doi.org/10.1016/j.im.2017.07.007

Hotz, N. (2023) *OSEMN data science life cycle*, *Data Science Process Alliance*. Available at: https://www.datascience-pm.com/osemn/ (Accessed: 13 August 2023).

Li, Y., Liu, X., & Zhang, J. (2021). Limitations of the OSEMN framework in real-world data science researchs. Journal of Ambient Intelligence and Humanized Computing, 12(3), 3471-3478.

Mulok, D., Kogid, M., Asid, R., & Lily, J. (2018). Multi-Factor Crime in Malaysia, 1980 – 2013: Bounds Testing of Level Relationships and Granger Non-Causality Analysis *Journal of International Business Economics and Entrepreneurship*, *3*(2), 55. https://doi.org/10.24191/jibe.v3i2.14433

Wang, X., Liu, X., & Wang, Y. (2018). A comprehensive review of big data quality challenges and solutions. Journal of Big Data, 5(1), 1-26. https://doi.org/10.1186/s40537-018-0093-3

Wang, J., Zhang, Y., & Liu, L. (2022). A scalable and flexible ETL process for big data integration. Journal of Ambient Intelligence and Humanized Computing, 11(8), 11923-11930.

Xu, B. (2021). Artificial intelligence teaching system and data processing method based on big data. *Complexity*, *2021*, 1-11. Retrieved from https://doi.org/10.1155/2021/9919401

Zakaria, E., Kamarudin, N. N., Mohamad, Z. S., Suzuki, M., Rathakrishnan, B., Bikar Singh, S. S., ... & Kamaluddin, M. R. (2022). The role of family life and the influence of peer pressure on delinquency: qualitative evidence from Malaysia. *International journal of environmental research and public health*, *19*(13), 7846. Retrieved from https://doi.org/10.3390/ijerph19137846