

Specialist Diploma in Data Science For Business

Data Visualisation Fundamentals (CBG1S03)

Submitted by

2073306I

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student No

Wong Poh Yeng

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Name

4 June 2021

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date of Submission

## Annex B– Declaration of Work of Originality

Specialist Diploma in Data Science For Business

Data Visualisation Fundamentals (CBG1S03)

AY2021/2022

Assignment

Submitted by: <2073306I> <Wong Poh Yeng>

Date: <4 June 2021>

“By submitting this work, I am declaring that I am the originator(s) of this work and that all other original sources used in this work has been appropriately acknowledged.

I understand that plagiarism is the act of taking and using the whole or any part of another person’s work and presenting it as my own without proper acknowledgement.

I also understand that plagiarism is an academic offence and that disciplinary action will be taken for plagiarism.”

Name and Signature of student: …………………………………

Wong Poh Yeng

Terrorism is a serious problem that affected many countries for many years. It has become a growing concern, with terrorists motivated by political, racial, economic, health, and other causes. Every year, thousands of people become the victims of terrorist attacks.

Terrorist organizations are the most dangerous security threat to any country. These organizations utilize terrorism to cause fear in any country at any time and any location.

My goal on this project is to provide users with an interactive tool that will enable them better understands the pattern of terrorist attacks. The areas that I will be focusing on are:

* Find out how the terrorism trend has changed over time.
* Determine the number of people that died as a result of these attacks.

**1. Analysis on data quality**

**Dataset**

The dataset is downloaded from Polymall. The original source of the dataset is retrieved from <https://www.start.umd.edu/gtd>. It contains information on domestic and international terrorist attacks from 1970 to 2017.

The dataset is very comprehensive and contains a lot of terrorism-related information. It has 181,691 terrorist attacks cases with 79 attributes.

Let's get started with the data quality check. I discovered that the dataset was incomplete because there was no data for the entire year of 1993.

I also discovered that there are a lot of irrelevant features. After removing the irrelevant columns, there are 15 features left. I used Excel and Tibco Spotfire to examine the data quality. The metadata and findings on data quality are shown in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SN | Variable Name | Description | Units | Data Types | Finding | |
| 1 | eventid | Incident ID follow 12 digits format-“yyyymmdd+0001” | 181,691 |  |  | |
| 2 | iyear | Year | 47 | Numeric | From 1970 to 2017 but year 1993 was missing | |
| 3 | country | Country | 204 | Numeric |  | |
| 4 | region | Region | 12 | Numeric |  | |
| 5 | provstate | Province/State |  | Character | 421 missing value | |
| 6 | city | city |  | Character | 434 missing value | |
| 7 | success | Successful Attack?  1="Yes" , 0="No, not successful” | 2 | Numeric |  | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | |
|  |  |  |  |  |  | | |
| SN | Variable Name | Description | Units | Data Types | Findings |
| 8 | suicide | Suicide Attack? 1="Yes" , 0="No, not suicide attack" | 2 | Numeric |  |
| 9 | attacktype1 | Attack Type eg explosion, assault, kidnap, etc. | 9 | Numeric |  |
| 10 | targtype1 | Target/Victim Type eg military, police, airport, etc. | 22 | Numeric |  |
| 11 | natlty | Nationality of the attacker | 216 | Numeric | 1,559 missing value |
| 12 | gname | Perpetrator Group Name |  | Character | 45 inconsistent value which show date format eg 01/9/2020 instead of group name |
| 13 | Nperps | Total Number of terrorists particpating in the incident |  | Numeric | 71,115 missing value - replace with -99 |
| 14 | weaptype1 | Weapon Type eg firearms, explosive etc | 13 | Numeric |  |
| 15 | nkill | Total Number of Fatalities |  | Numeric | 10,313 missing value, 6% of the total |
| 16 | nwound | Total Number of Injured |  | Numeric | 16,311 missing value, 9% of the total |
|  |  |  |  |  |  |

**2. Preprocessing of data**

I did some data cleaning and preprocessing before performing data visualization in Tableau. Some of the cleaning works done are listed below.

* 1. For categorical data, I replaced the missing value in provstate, city and gname with “Unknown”.
  2. For numerical data, I did the following with regards to the missing value.

|  |  |  |
| --- | --- | --- |
| SN | Variable Name | Methods to handle numerical missing value |
| 1 | nkill | Replace 10,313 missing value with “0”. |
| 2 | nwound | Replace 16,311 missing value with “0”. |
| 3 | Nperps | Replace 71,115 missing values with “-99”. |
| 4 | natlty | Replace 1,559 missing values with “-99”. |

Note:

I chose to replace the missing value on these two variables, “Nkill” and “nwound”, with "0" because it is insignificant. They account for only 6% and 9% of the total population, respectively.

In the dataset, I discovered missing data for the entire year 1993. Therefore, I chose to replace the missing values on Nperps and Natlty rather than delete them.

* 1. I used Excel to check for eventid duplication and I noticed that there were no duplication cases. However, the 12 digits will be of no use in data visualization. Therefore, I deleted this column and replaced it with a new column called “Number of Attack”. On each row, the observation “1” is placed to indicate that one case/incident of Attack.
  2. I noticed that the columns weaptype, region, country, attacktype, targtype, and natlty have been encoded with numerical values.

To decode these numerical values, I used Tibco's left-outer join to combine the original dataset's table with the "globalterrorismdb dimension" file downloaded from Polymall.

* 1. The final dataset consists 181,691 number of cases x 18 features.
  2. The final dataset would then be imported into Tableau. I would double-check the datatype to make sure the numerical variables are in numerical format and the text variables are in string format.

When plotting the graphs, I would ensure that the year variable is set to "discrete," and that the nation, region, and province variables are set to "geographic roles" when plotting the map.

**3. Dashboard Design and Consideration**

In order to determine how the terrorism trend has changed over time and the number of people who have died as a result of these attacks. I build these two dashboards (as show below).

|  |
| --- |
| **3.1 Dashboard 1** |
| **3.2 Dashboard 2** |

**Dashboard 1**

My goal with this dashboard is to present users with an interactive dashboard that allows them to drill down to discover the information they need about the number of terrorist attacks and deaths over time.

1. Line graphs were used to show the time series of the attacks and deaths by regions and terrorist groups.
2. Highlight Actions was configured so that when users select specific detail on one graph, the details on the other graphs are automatically highlighted.
3. Important data and filters are placed at the top left corners of dashboards to grab the user's attention and make it easier for them to find the information they need.
4. The dashboard size was set to "automatic," which meant it would automatically resize to fit into a desktop, laptop, or mobile phone.
5. Tooltips were configured to include all of the important information that users require.
6. In Graphs 2 and 3, the number of line colors was limited to five.

**Dashboard 2**

My goal with this dashboard is to provide users an interactive tool that allows them to drill down to find data on the number of people killed by region, target, weapon, and attacking technique.

1. In this dashboard, bar, stack bar, and bubble charts were used to compare different categories of data.
2. Drill down, filters, and important information were placed at the top of the dashboard to grab user's attention and make it easier for them to find the information they require.
3. Boarder was used to create a clean look and to separate different worksheets.
4. In Graph 1: Number of people killed by country and region – unnecessary grid lines were removed, the bar was colored with saturation, and the total value was placed at the top of each bar.
5. The same legend was used for all three graphs (Dashbaord 2: Graph 2, 3, and 4).
6. Tooltips were configured to include all of the important information that users require.

**4. Findings from the dashboard**

Now, let's take a closer look at each dashboard.

**Dashboard 1: Global Terrorism Trend (1970 to 2017)**

|  |
| --- |
| **Dashboard 1 - Graph 1**: Number of Attack & Killed over the year |
| According to the graph, the total number of people killed between 1970 and 2011 was lesser than 12,000 each year. However, the number has increased since 2012 and it peak in year 2014.  Following the peak in 2014, the number of terrorist activities and deaths fell for the third year in a row. In 2017, the total number of attacks fell to 8,283 and death fell to 23,653. |
| **Dashboard 1 - Graph 2:** Top Terrorism attack regions over the year |
| The graph above shows the top five terrorist attack regions for the year.  Prior to 2011, fewer people were killed in terrorist attacks (an average of less than 1,443 people per year), and the hardest hit regions were the Middle East and North Africa, followed by Sub-Saharan Africa and South Asia.  However, the number of people killed in terrorist attacks has steadily increased, reaching a peak in 2014, particularly in the Middle East and South Asia, where 19,321 people were killed, and Sub-Saharan Africa, where 12,662 people were killed.  In 2017, the Middle East and North Africa were the most dangerous terrorist attack hotspots, with 3,526 attacks and 10,213 deaths.  South Asia overtaken Sub-Saharan Africa as the second most dangerous region in 2017, killing 7,372 people. |
| **Dashboard 1 - Graph 3:** Top 5 Deadliest Terrorist Group over the year    According to the graph above, ISIL was the most dangerous terrorist group prior to 2014, followed by Boko Haram and the Taliban. In 2017, the Taliban has overtaken Boko Haram to become the second most dangerous terrorist group. |

**Dashboard 2 - Fatalities from terror attacks**

|  |
| --- |
| **Dashboard 2 - Graph 1:** Number of people killed by countries & regions |
| This graph shows the number of people killed by country and region. Iraq has suffered the most death, with 75,790 people killed, followed by Afghanistan, which has lost 37,862 people.  The most affected regions have been the Middle East & North Africa, and South Asia. |

|  |  |
| --- | --- |
|  | **Dashboard 2 - Graph 2:**  Most Fatal Attacking Method used  This bubble chart shows the most Fatal Attacking Methods used.  Per the graph, bombing is the most lethal attacking technique, killing 139K people, followed by Armed Assault, which killed 120K people. |
|  | **Dashboard 2 - Graph 3:**  Top Targets and Attacked Method used  According to the graphs above, terrorists' favorite targets are private citizens and property, follow military, and police.  Terrorists killed 124K people when they targeted private citizens and property, and the most fatal attacking technique they used was armed assault when they targeted private citizens and property. |
|  | **Dashboard 2 - Graph 4**  Top Weapons and Attack Method used  The graphs show the most effective weapons and attacking techniques used by terrorists. The most fatal weapons are explosives and firearms.  153,557 people were killed by explosives. |
| **Findings – Dashboard 1** *(To answer questions on how the terrorism trend change over time)*  According to the above analysis, there was 123,919 terrorist activities and, 330,458 people killed worldwide in the last 48 years. There is a rising trend in terrorist activities since 2012.    Between 1970 and 2011, there was lesser than 12,000 terrorist activities each year. The number has increased since 2012 and it peak in year 2014. Following the peak in 2014, the number of terrorist activities and deaths fell for the third year in a row. In 2017, the total number of attacks fell to 8,283 and death fell to 23,653.  In 2017, the Middle East and North Africa were the most dangerous terrorist attack hotspots, with 3,526 attacks and 10,213 deaths. South Asia was the second most dangerous region, with 7,372 deaths.  In 2014, ISIL was the most dangerous terrorist groups, followed by Boko Haram and the Taliban. In 2017, the Taliban has overtaken Boko Haram to become the second most dangerous terrorist group.  **Findings – Dashboard 2** *(to answer question on how many fatalities resulted of these attacks)*  These attacks killed a total of 330,458 people. Iraq is the worst-affected country, with 75,790 people killed.  Terrorists target private citizens and property, as well as military and police. There were 124,395 people killed when terrorists targeted private citizens and property.  Explosives and firearms are the deadly weapons used. There was 153,557 people were killed by explosives.  Bombing is the most fatal attacking technique with 139,262 people died. | |

Conclusion

From the above analysis, I managed to discover some of the trends and pattern of terrorism attacks. START may consider some of the solutions described below to address the global terrorist problem.

1. Adopt a strong domestic law to control the export of deadly weapons.
2. International cooperation should be strengthened in the fight against terrorism.
3. Tackling terrorism finance, put special focus on the top few active groups.