Mohamed Shah

 5982571

Assignment 2

Table of Contents

[Executive summary 4](#_Toc89632371)

[Data Understanding report 4](#_Toc89632372)

[Data Cleaning 5](#_Toc89632373)

[Description of your methods. 5](#_Toc89632374)

[1. Insights found by the data 6](#_Toc89632375)

[2. Neural Network results 8](#_Toc89632376)

[Your results and explanation of your results. 9](#_Toc89632377)

[1. Insights found by the data 9](#_Toc89632378)

[2. Neural Network results 18](#_Toc89632379)

[Your summary and expert judgement based on all your analysis. 20](#_Toc89632380)

[Answering the general questions 21](#_Toc89632381)

  
  
**Assignment Cover Sheet**

|  |  |
| --- | --- |
| **Subject Code: CSCI316** |  |
| **Subject Name: BIG DATA MINING TECHNIQUES AND IMPLEMENTATION** |  |
| **Submission Type: Online** |  |
| **Assignment Title: Assignment 2** |  |
| **Student Name: M.S.Saalik Shah** |  |
| **Student Number: 5982571** |  |
| **Student Phone/Mobile No. 0507815780** |  |
| **Student E-mail: msss950**@uowmail.edu.au |  |
| **Lecturer Name: Dr Farhad Oroumchian** |  |
| **Due Date: 5/12/2021** |  |
| **Date Submitted: 5/12/2021** |  |

|  |  |
| --- | --- |
| **PLAGIARISM:** The penalty for deliberate plagiarism is FAILURE in the subject. Plagiarism is cheating by using the written ideas or submitted work of someone else. UOWD has a strong policy against plagiarism.  The University of Wollongong in Dubai also endorses a policy of non-discriminatory language practice and presentation.  **PLEASE NOTE:**STUDENTS MUST RETAIN A COPY OF ANY WORK SUBMITTED | **DECLARATION:** I/We certify that this is entirely my/our own work, except where I/we have given fully-documented references to the work of others, and that the material contained in this document has not previously been submitted for assessment in any formal course of study. I/we understand the definition and consequences of plagiarism.  **Signature of Student:** |

|  |  |  |
| --- | --- | --- |
| |  | | --- | | **Optional Marks:** | | **Comments:** | |

https://my.uowdubai.ac.ae/images/scissors.gif

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Lecturer Assignment Receipt**(To be filled in by student and retained by Lecturer upon return of assignment) | | | **Subject:** | **Assignment Title:** | | **Student Name:** | **Student Number:** | | **Due Date:** | **Date Submitted:** | | **Signature of Student:** | | |

https://my.uowdubai.ac.ae/images/scissors.gif

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Student Assignment Receipt** (To be filled in and retained by Student upon submission of assignment) | | | **Subject:** | **Assignment Title:** | | **Student Name:** | **Student Number:** | | **Due Date:** | **Date Submitted:** | | **Signature of Lecturer** | | |

# Executive summary

The AllStatesandTerrortories dataset consists of the data related to military grade equipment with the associated Quantity, Acquisition value, NSN, Agencies, DEMIL code, DEMIL IC and Ship Date. At first, we create a report on the data and its importance with the data type and how can it be further transformed by adding more columns by splitting a particular column that has more information, like the ship date or remove any unnecessary columns like the State type attribute.

Then we clean the data by replacing null or NANs values, normalize attributes and standardize the target class to one hot encoding (binary values). Merging columns to get better results like Quantity and Acquisition value

Then we move on to the different methods used, this is divided into to which are Insights from data and the performance of the neural networks used. The insights from data describe what were the insights that we wanted to obtain and explaining the expected results while the performance of the neural network described the different types of networks used and the architecture implemented.

Then we move on to the next which is the results found, which is again divided into two, the first part explains the results of the insights from the data and the second part explains the results of the neural network.

Finally, we state the summary and expert judgment of the insights attained from the data and then answer the general questions asked.

# Data Understanding report

Based on the dataset we see that the data is about the **expenditure of military equipment’s** bought by various agencies in the different states in USA and given to the police force because they are not needed anymore. The data also shows the different types of items bought with the associated quantity and Acquisition value.

By splitting the Ship date to year, we can determine which year had the highest number of acqusition. The dtype is object (string). This can be divided to dates and time and further divide date to: day, month and year for better and more detailed analysis.

The attribute State consists of all the states that had acquired military graded equipment. The dtype is object (string).

The attribute agency consists of all the agencies of different states that had acquired military graded equipment. The dtype is object (string).

The attribute NSN consists of all the items code. The dtype is object (string).

The attribute Item Name consists of all the names of the military graded equipment. The dtype is object (string).

The attribute Quantity consists of all the number of the particular military graded equipment acquired. The dtype is object (int). The quantity can be multiplied with the acquisition value to get the total spending for the item and for the state.

The attribute UI consists of whether the military graded equipment acquired is an individual, kit or other types. The dtype is object (string).

The attribute Acquisition value consists of all the value of the particular military graded equipment acquired. The dtype is float.

The attribute DEMIL code consists of all the associated code for particular military graded equipment acquired that determines whether the equipment needs to be destroyed, license and so on. The dtype is object (string).

A : imposing licensing for certain destinations

B : Destroying to an extent that the equipment is used for scraping purpose

C : remove and dematerailze key points

D : destroy the item to prevent restoration

E : reserve for exclusive use

F : special demil instructions

Q : Destroying to an extent that the equipment is used for scraping purpose outside US

The attribute DEMIL IC consists of all the associated code for particular military graded equipment acquired that determines validity of an item. The dtype is float.

0.0 - Demil code reviewed and Demil code not equal

1.0 - Demil code reviewed and no Demil code change recommended

3.0 - Requires mutilation

4.0 - Demil code could not be validated

5.0 - Item reviewed and coded review/collaboration cycle

6.0 - Requires mutilation overseas

7.0 - Failed to update the DEMIL code

Among these data, NSN and UI seem less important when compared to the other attributes as they can give a better insight.

# Data Cleaning

Firstly, all the data values corresponding to the attributes should have no missing values and should be consistent. We see that **column DEMIL IC** has **some missing values**, so these missing values need to be filled, looking at the most occurring value (mode) we see that the value 1 occurs the most and hence the missing values were assigned to 1.

The **column State type** is not of any use since the data is about different states in the US and so all the values in State type is State, so this column is of no use.

The agency name also does not play an important role since we can focus on what states have been spending more money rather than the agencies. We can consider the **state as a normalized version of the agency**.

The **Ship Date column can be split** so that we can get a better understanding of the data, so one column will be called as Date and the other as Time, in the time. But **when feeding the neural network** with the dataset we will keep the Ship Date as it is since we it was split it did not have any difference at all, but for **data understanding** it was useful.

For a **better performance** of the neural networks the X class or feature class was **normalized**, and the y class or target class was **categorized** to convert to binary class matrix

The attribute Quantity and Acquisition value was multiplied to get **Total spending**, this was done for **Data Understanding.**

# Description of your methods.

## Insights found by the data

1. ***Top 5 Highest and Lowest expenditure among the states in USA?***

Ans: Here we find the states that have the highest and lowest expenditure during the period of 1980 - 2021

1. ***Top 5 years which had the highest expenditure and lowest expenditure between 1980 - 2021?***

Ans: Here we find the years that have the highest and lowest expenditure during the period of 1980 - 2021

1. ***Which decade had the highest expenditure and the lowest expenditure?***

Ans: Here is the system in Germany: 1E4 = 10000, 1E7 = 10 million, 1E8 = 100 million , 1E9 = 1000 million or 1 billion

Here we find the decade that have the highest and lowest expenditure acquired during the period of 1980 - 2021

1. ***Top 10 items that were brought least and several times during this period (1980-2021)?***

Ans: Here we find the items that are acquired the most and least times during the period of 1980 - 2021

1. ***Top 10 items that were brought at the lowest and highest price during this period (1980-2021)***

Ans: Here we find the items that have the highest and lowest acquisition value during the period of 1980 - 2021

1. ***Top 10 items total purchased (Lowest and Highest) through the period (1980 - 2021)?***

Ans: Here we find the items in total (Quantity \* Acquisition value) that was acquired having the highest and lowest expenditure during the period of 1980 - 2021

1. ***Most Shipped Date?***

Ans: Here we find the Date that where the highest number of items were shipped

1. ***Most Shipped year?***

Ans: Here we find the year that where the highest number of items were shipped

1. ***Highest acquisition item spent by a state in each decade?***

Ans: Here we find the item with the highest acquisition in each decade

(1980-1989, 1990-1999, 2000-2009, 2010-2019, 2020-2021)

1. ***Highest acquisition item spent by a state in each year?***

Ans: Here we find the item with the highest acquisition in each year

1. ***Top 5 Highest and Lowest expenditure among the states and agencies in USA?***

Ans: Here we find the states and agencies with the highest and lowest total spending

1. ***Top 5 Highest and Lowest expenditure among the DEMIL Code in USA?***

Ans: Here we find the DEMIL Code with the highest and lowest total spending

1. ***Top 5 Highest and Lowest expenditure among the DEMIL IC in***

***USA?***

Ans: Here we find the DEMIL IC with the highest and lowest total spending

## Neural Network results

For the Neural Networks, we did three types which were ANN (Multi-perceptron), CNN and LSTM

1. ANN (Multi-perceptron): - The Artificial Neural Network used was a multi-perceptron architecture which is a feed forward network that takes multiple inputs and feeds it or passes it to the hidden layers were the assigned relu activation function performs and sends to the output layer for classification.

In this particular architecture, there is an input node containing inputs = 5000 which then follows by two hidden layers of 1000 and 500 respectively and was assigned relu as the activation function (relu was used since it simple, easy, and outputted satisfying or excellent results and is widely used, hence it was chosen), finally it is passed to the output containing 7 outputs since our target class contains 7 possible categories. This output layer had softmax has the activation function since we have a multi-class classification. We have used binary\_crossentropy as the loss function but when we had used categorical\_ crossentropy it had given almost the same results but with better error rate.

1. CNN: - The Convolutional Neural Network used was a 1DConv architecture which converts the data to an image format that takes multiple inputs and feeds it or passes it to the hidden layers were the assigned relu activation function performs and sends to the output layer for classification.

In this particular architecture, there is an input node containing filters = 64 and a kernel\_size = 3 which then follows by five hidden layers where the first hidden layer consists of filters = 64 and a kernel\_size = 3, the filters were set to 64 to capture more features and it was assumed that this would be the optimal number of filters for our data (since it was not an image).

The second hidden layer was a dropout layer set to 0.5 since we want to prevent overfitting, the third hidden layer was a MaxPooling1D layer set to 2 since we wanted to choose the most present feature and not the average feature. The fourth hidden layer is a flatten layer where it outputs a single long feature. And the fifth hidden layer which is the last feature layer in this architecture with the activation function set to relu and inputs was set to 100. And then finally the output layer contains 7 outputs since our target class contains 7 possible categories. This output layer had softmax has the activation function since we have a multi-class classification. We have used categorical\_ crossentropy as the loss function but when we had used binary\_crossentropy it had given almost the same results.

1. LSTM: - The Long Short Term Memory Neural Network is a Recurrent neural network architecture which passes information as it moves forwards and decides whether to keep or loss the information to the hidden layers.

In this particular architecture, there is an input node containing inputs = 100 which then follows by seven hidden layers where the first layer is a dropout layer set to 0.2 since we want to prevent overfitting, then we add a second LSTM layer and set the inputs to 100 and then the third layer is a dropout layer set to 0.2, then comes the fourth layer which is a LSTM layer and the inputs is set to 1000 inputs, then the fifth layer comes which is a dropout layer set to 0.2 followed by the LSTM layer and set the inputs to 100 and then the last hidden layer which is a dropout layer set to 0.2, Lastly it is passed to the output layer containing 7 outputs since our target class contains 7 possible categories.

We are using two LSTM neural network with different loss functions, one loss function is with binary\_crossentropy and the other loss function is mean\_squared\_error.

# Your results and explanation of your results.

## Insights found by the data

1. ***Top 5 Highest and Lowest expenditure among the states in USA?***

Top 5 States with the highest expenditure from 1980 - 2021

1. TX (TEXAS) = $1.547442e+08 (approx) $154,744,200

2. CA (CALIFORNIA) = $1.545611e+08 (approx) $154,561,100

3. TN (TENNESSEE) = $1.343888e+08 (approx) $134,388,800

4. FL (FLORIDA) = $1.040935e+08 (approx) $104,093,500

5. AZ (ARIZONA) = $9.537915e+07 (approx) $95,379,150

Top 5 States with the lowest expenditure from 1980 - 2021

1. VI (VIRGIN ISLANDS OF THE U.S.) = $690.00 (approx)

2. GU(GUAM) = $632,211.95 (approx)

3. MP(MISSISSIPPI) = $1,414,950.00 (approx)

4. VT (VERMONT) = $1659755.11 (approx)

5. DE (DELAWARE) = $1712960.29 (approx)

1. ***Top 5 years which had the highest expenditure and lowest expenditure between 1980 - 2021?***

Top 5 years with the highest expenditiure

year

2014 3.851037e+08

2016 2.478768e+08

2013 1.804401e+08

2019 1.521905e+08

2020 1.499985e+08

Top 5 years with the lowest expenditiure

year

1992 7261.8

1991 10074.0

1990 23046.0

1980 65070.0

1. 66113.0

From the 2000 it started to increase and then from 2014 it sky rocketed where the expenditure was at it's peak then started falling but gained an increase in 2019 and then fell , 2014 had the highest expenditure around $380,000,000

1. ***Which decade had the highest expenditure and the lowest expenditure?***

Here is the system in Germany: 1E4 = 10000, 1E7 = 10 million, 1E8 = 100 million, 1E9 = 1000 million or 1 billion

year

1980 6.507000e+04

1990 2.037046e+07

2000 7.677947e+07

2020 2.987331e+08

2010 1.503786e+09

1980 had the lowest expenditure and 2010 (2010 - 2019) had the highest expenditure

1. ***Top 10 items that were brought least and several times during this period (1980-2021)?***

Top 10 Least number of items brought

Item Name

ABSORBENT PAD,FLUID,MEDICAL 1

ACCESSORY KIT,ANTENNA 1

ACCESSORY KIT,UMBRE 1

ACCUMULATOR,AIR,TUR 1

ACOUSTIC SUPPR KIT 1

ACTUATOR ASSY,POSIT 1

ACTUATOR,LINEAR 1

ADAPTER 8-32 THREAD 1

ADAPTER ANTENNA 1

ADAPTER ASSEMBLY,INTERFACE DEVICE 1

Top 10 highest number of items brought

Item Name

RIFLE,5.56 MILLIMETER 46843

RIFLE,7.62 MILLIMETER 11146

SIGHT,REFLEX 10899

PISTOL,CALIBER .45,AUTOMATIC 5664

TRUCK,UTILITY 2634

ILLUMINATOR,INFRARED 1954

NIGHT VISION GOGGLE 1861

IMAGE INTENSIFIER,NIGHT VISION 1822

VIEWER,NIGHT VISION 1517

SIGHT,THERMAL 1339

1. ***Top 10 items that were brought at the lowest and highest price during this period (1980-2021)?***

Top 10 lowest Acquisition Value of items

Table

Description automatically generated

Top 10 Highest Acquisition Value of items

Table

Description automatically generated

1. ***Top 10 items total purchased (Lowest and Highest) through the period (1980 - 2021)?***

top 10 lowest total purchased

Table

Description automatically generated

top 10 highest total purchased

Table

Description automatically generated

1. ***Most Shipped Date?***

Top 10 Most shipped Date

Ship Date

2012-01-17 1205

2017-09-18 1134

2013-09-19 828

2014-07-15 730

2012-04-26 599

2002-10-15 593

2011-11-14 553

2014-08-27 494

2014-11-13 480

* + 1. 477

1. ***Most Shipped year?***

Top 10

year

2012 15648

2021 11786

2011 11521

2020 10026

2014 9396

2013 6702

2017 5341

2002 5307

2010 4535

2016 4342

1. ***Highest acquisition item spent by a state in each decade?***

Table

Description automatically generated

1. ***Highest acquisition item spent by a state in each year?***

Table

Description automatically generated

Table

Description automatically generated

1. ***Top 5 Highest and Lowest expenditure among the states and agencies in***

***USA?***

Top 5 Highest

State Agency Name

AZ ARIZONA DEPT OF PUBLIC SAFETY 52594669.90

CA HIGHWAY PATROL 22371680.00

FL DEPT OF LAW ENF TALLAHASSEE 20952245.69

SC LAW ENFORCEMENT DIVISION 19855979.32

TX HOUSTON POLICE DEPT 10823252.81

Top 5 Lowest

State Agency Name

PA SHAMOKIN POLICE DEPT 28.00

GA BERLIN POLICE DEPT 58.71

NM CIMARRON POLICE DEPT 60.60

MA WORCESTER POLICE DEPT 79.11

KY JUNCTION CITY POLICE DEPT 81.09

1. ***Top 5 Highest and Lowest expenditure among the DEMIL Code in USA?***

Top 5 Highest

DEMIL Code Item Name

C MINE RESISTANT VEHICLE 7.242213e+08

TRUCK,UTILITY 1.813147e+08

AIRCRAFT, FIXED WING 9.849900e+07

Q UNMANNED VEHICLE,GROUND 7.248551e+07

D MINE RESISTANT VEHICLE 5.008489e+07

Top 5 Lowest

DEMIL Code Item Name

A COLLECTIVE MODULAR SUPPORT SYSTEM 0.0

GUARD,SHIN 0.0

METEOROLOGICAL STATION,AUTOMATIC 0.0

PORTABLE SHOWER 5 G 0.0

TOOL KIT,POWER,CORDLESS 0.0

1. ***Top 5 Highest and Lowest expenditure among the DEMIL IC in***

***USA?***

Top 5 Highest

DEMIL IC Item Name

1.0 MINE RESISTANT VEHICLE 7.735962e+08

TRUCK,UTILITY 1.803480e+08

AIRCRAFT, FIXED WING 9.873600e+07

3.0 UNMANNED VEHICLE,GROUND 7.248551e+07

1.0 HELICOPTER,UTILITY 3.875357e+07

Top 5 Lowest

DEMIL IC Item Name

1.0 COLLECTIVE MODULAR SUPPORT SYSTEM 0.0

GUARD,SHIN 0.0

METEOROLOGICAL STATION,AUTOMATIC 0.0

PORTABLE SHOWER 5 G 0.0

TOOL KIT,POWER,CORDLESS 0.0

## Neural Network results

The Three Neural Network used are: Artificial Neural Network (Multi-perceptron), Convolutional Neural Network and Long Short-Term Memory Neural Network

1. ANN (Multi-perceptron): -

**Unbalanced Dataset**: The accuracy was an excellent score, but the recall and precisions were very low and this is largely due to the dataset being unbalanced.

Time taken (Training): 12 minutes

Loss function (Training) - 0.0268

Training Accuracy – 96.75%

Loss function (Testing) - 0.0289

Testing Accuracy – 96.54%

**Balanced Dataset**: The accuracy, recall and precisions were excellent scores this was largely due to the dataset being balanced.

Time taken (Training): 4 minutes

Loss function (Training) - 0.065

Training Accuracy – 92.63%

Loss function (Testing) - 0.067

Testing Accuracy – 92.40%

1. CNN: -

**Unbalanced Dataset**: The accuracy was an excellent score, but the recall and precisions were very low and this is largely due to the dataset being unbalanced.

Time taken (Training): 6 minutes

Loss function (Training) - 0.065

Training Accuracy – 90.82%

Loss function (Testing) - 0.060

Testing Accuracy – 91.53%

**Balanced Dataset**: The accuracy, recall and precisions were excellent scores this was largely due to the dataset being balanced.

Time taken (Training): 4 minutes

Loss function (Training) - 0.014

Training Accuracy – 90.82%

Loss function (Testing) - 0.060

Testing Accuracy – 91.53%

1. LSTM: -
2. Loss function: - binary\_crossentropy

**Unbalanced Dataset**: The accuracy was an excellent score, but the recall and precisions were very low, and this is largely due to the dataset being unbalanced.

Time taken (Training): 44 minutes

Loss function (Training) - 1.256

Training Accuracy – 71.32%

Loss function (Testing) - 1.248

Testing Accuracy – 71.52%

**Balanced Dataset**: The accuracy, recall and precisions were excellent scores this was largely due to the dataset being balanced.

Time taken (Training): 24 minutes

Loss function (Training) - 0.328

Training Accuracy – 50.32%

Loss function (Testing) - 0.385

Testing Accuracy – 25.50%

1. Loss function: mean\_squared\_error

**Unbalanced Dataset**: The accuracy was an excellent score, but the recall and precisions were very low and this is largely due to the dataset being unbalanced.

Time taken (Training): 45 minutes

Loss function (Training) - 0.008

Training Accuracy – 96.59%

Loss function (Testing) – 0.010

Testing Accuracy – 95.39%

**Balanced Dataset**: The accuracy, recall and precisions were excellent scores this was largely due to the dataset being balanced.

Time taken (Training): 25 minutes

Loss function (Training) - 0.016

Training Accuracy – 93.85%

Loss function (Testing) - 0.015

Testing Accuracy – 92.78%

# Your summary and expert judgement based on all your analysis.

In the 2000’s there was a sudden upward trend, this is due to some items sold at a cheaper price and an excess supply of equipment and the additional equipment made available for purchase within the country.

The state of Texas, California and Florida had the top 5 highest spending of military equipment over this period (1980-2021), one thing to note is that there are many military bases located in these states and they are in prime location which helps to have a cheaper shipping cost.

Before 1995, we see that most purchased weapon was the RIFLE,7.62 MILLIMETER but then suddenly the quantity of this weapon falls and a new weapon RIFLE,5.56 MILLIMETER had started to have a higher preference and this could be largely due to new and improved standards of the weapons required by the US military, the factors could be the speed of the bullet, the weight of the weapon, the number of ammunitions and so on. Hence, we could see a change in the purchasing preferences of the US military that had started to give the police forces with very high military grade equipment.

Some items of the military were given for free to the police force such as METEOROLOGICAL STATION, AUTOMATIC, LENS,CAMERA,GENERAL PHOTOGRAPHIC, PLATFORM LIFT and a few more were having an Acquisition value of 0. This could largely be due them having a very low demand.

The purchase AIRCRAFT, FIXED WING was the equipment having the highest acquisition value. Also, another thing noticed is that the top 9 items with the highest acquisition value was Aircrafts which were AIRCRAFT, FIXED WING and COMMUN,EQUIP,SOMS-B.

The acquisition of MINE RESISTANT VEHICLE was acquired the highest in terms of its

(Quantity \* Acquisition Value) Total spending for an item. Followed by TRUCK,UTILITY and AIRCRAFT, FIXED WING.

The year which had the most items shipped was 2012, 2021 and 2011 with 15648, 11786 and 11521 in quantity respectively.

Most of the lowest valued equipment were marked as DEMIL Code = A and DEMIL IC = 1

Most of the lowest valued equipment were marked as DEMIL Code = D and DEMIL IC = 1

Providing the police force with such high-grade military equipment could help in situation where there is a sudden rescue force needed, so the flights and water tanks would of use in taking quick action to help civilians.

# Answering the general questions

1. *What is the best predictor model for prediction?*

The best predictor model out of the three predictors is ANN (Multi-perceptron). This is the best since it has the highest accuracy when compared to the other two and more importantly with excellent recall and precision and the loss function is very less, another point to note was it was time efficient. Overall making this predictor the best out of the three.

1. Is there a pattern in the data with respect to any or group of attributes?

When the Ship Date attribute was divided into date and time and the date further divided into year, month, and day. Then we had taken year to find patterns, we noticed that there were more items shipped during 2010-2014 when the spike on the acquisition of items started to rise. This seems true since when the spike happened the state agencies had acquired more equipment since there was a cheaper acquisition value and more supply.

Most of the military items were classified as items to be destroyed or DEMIL CODE = D since they seem to dangerous and not necessary for the police force.

During 2010-2014, when a spike had occurred in the number of equipment being shipped, we could see that during this period there was a huge supply and many state agencies were interested in these items and hence there were many acquisitions/purchases made. Hence 2014 had accounted to the highest total spending while 2012 was among the top 3.

On a general note, states like Texas are surrounded by many military bases, hence many agencies in Texas were able to acquire the equipment quickly as there were no issues of shipment.

Top 9 items having the highest acquisition value were Aircrafts. Top 3 items having quantity > 1 had highest total acquisition value were tanks.

Equipment such as Aircraft and tanks have DEMIL code set to A

Equipment such as RIFLE, 7.62 MILLIMETER and other RIFLES have DEMIL code set to D and DEMIL IC set to 1

1. What recommendation you can give based on your findings.

Based on the finding we can see that some weapons are too dangerous in a surrounding where there are so many civilians, having such high equipped military equipment seem unnecessary. For example, tanks, military grade guns seem unnecessary. This could give rise to many guns related activities within the US.

It is difficult to give a clear analysis about whether the police force should be given this military grade equipment since we do not have much data on what and how have the police been using these weapons, whether there was chaos that had taken place within the states that had higher military equipment acquisition or other reasons. It is also mentioned that most of the equipment are uncontrolled meaning not within the authority of the military representative but with the agencies and police department which make it quite difficult to judge on how these weapons would be used around civilians especially in populated areas where there are many minors. It would be a good idea to select the military equipment that would be adequate.

Even though the military equipment is used as a means of protection for officers and protection for civilians, it is better to consider safer options.

But looking at the context of our dataset we see the equipment such as office furniture, portable shower, screws, computer, GPS tracker and other equipment that cannot cause any harm to the people will be of great use.

1. Any insight or special cases, where you think we could consider.

Since Texas, California and Florida have the highest acquisition of military equipment we can see a large number of issues with the police force since they have access to this equipment and can be used against civilians.

The top 10 equipment with the highest acquired value were flights, and the highest acquired value with quantity > 1 were tanks since they were acquired in many quantities, and they were having a very high value.

During 2010-2014, when a spike had occurred in the number of equipment being shipped, we could see that during this period there was a huge supply and many state agencies were interested in these items and hence there were many acquisitions/purchases made. Hence 2014 had accounted to the highest total spending while 2012 was among the top 3.

On a general note, states like Texas are surrounded by many military bases, hence many agencies in Texas were able to acquire the equipment quickly as there were no issues of shipment

We also see that the equipment with the highest acquisition values have DEMIL code set to C which indicates the equipment key points need to be removed. The items with highest acquisition values were aircrafts and tanks and this military graded equipment might have custom features that are adapted to the war-type environment. Hence major key points should be removed.

Also, equipment such as Rifles have DEMIL code set to D and DEMIL IC set to 1.