Title:

Data Analysis On Terrorism Data Using Python

Project Plan:

Use Python Code to do the following:​

* Suggest multiple suitable hypothesis.​
* Clean the Data​
* Do distribution analysis​
* Represent our data in different graphical representation and with detailed analysis​

Working hours:

We expect a time of 35-40 manhours for the completion of the project

Distribution analysis:

Description of the data:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Year | Month | Day | Nkill | Nwound | Ransomamt |
| Count | 43915 | 43915 | 43915 | 41290 | 39313. | 2.730000e+02 |
| Mean | 2014.688079 | 6.442537 | 15.664716 | 3.110463 | 3.504006 | 4.066460e+06 |
| Std | 1.229350 | 3.390650 | 8.778536 | 13.343306 | 13.988988 | 2.180638e+07 |
| Min | 2012.000000 | 1.000000 | 1.000000 | 0.000000 | 0.000000 | -9.900000e+01 |
| 25% | 2014.000000 | 4.000000 | 8.000000 | 0.000000 | 0.000000 | -9.900000e+01 |
| 50% | 2015.000000 | 6.000000 | 15.000000 | 1.000000 | 0.000000 | 1.58604e+04 |
| 75% | 2016.000000 | 9.000000 | 23.000000 | 2.000000 | 3.000000 | 2.189562e+05 |
| Max | 2016.000000 | 12.000000 | 31.000000 | 1500.000000 | 1500.000000 | 2.000000e+08 |

**Correlation among the values**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Year | Month | Day | Nkill | Nwound | Ransomamt |
| Year | 1.000000 | -0.032845 | -0.002996 | -0.019340 | -0.025865 | -0.165084 |
| Month | -0.032845 | 1.000000 | 0.002112 | 0.005468 | -0.001070 | 0.077339 |
| Day | -0.002996 | 0.002112 | 1.000000 | -0.008516 | -0.001725 | 0.041875 |
| Nkill | -0.019340 | 0.005468 | -0.008516 | 1.0000000 | 0.364357 | 0.029284 |
| Nwound | -0.025865 | -0.001070 | -0.001725 | 0.364357 | 1.000000 | -0.038720 |
| Ransomamt | -0.165084 | 0.077339 | 0.041875 | 0.029284 | -0.038720 | 1.000000 |

**Content:**

1. Data Information:
   1. Data Info
   2. Data Describe
   3. Data Correlation
   4. Data Correlation Map
   5. Features of Data
2. Data Visualize & Analyze:
   1. US Terror Attacks Death and Injuries
   2. Death and Injuries at All Time
   3. Frequency of Terrorism Actions
   4. Frequency of Terrorist Actions in Customized Region
   5. Comparing People Who Died in Terrorist Attacks in The World and in The Region
3. Graphics and Analysis on Customized Data:
   1. Terrorist Attacks of a Particular Year and Their Locatios
   2. Total Number of People Killed in Terror Attack
   3. Types of Attacks That Cause Deaths
   4. Number of Killed in Terrorist Attacks by Countries

Relevent data:

The dataset has the following fields​

* iyear, imonth and iday: data of event​
* country and city​
* attacktype1\_txt: attack type​
* targtype1\_txt: target institution type​
* corp1: target institution​
* target1: target population type​
* natlty1\_txt: nationality of attacker​
* gname: name of attacking organization/terrorist group​
* weaptype1\_txt, weapsubtype1\_txt and weapdetail: describes the weapon of attack​
* nkill and nwound: number of people killed and wounded​
* propextent\_txt: extent of property damage​
* ransomamt: ransom amount (e.g. in case of hostage situations): values like 0 or negative indicate that either this was not a ransom case or the amount was unknown.

Possible approaches for data wrangling:

According to the dataset given we see that some of the fields in the given dataset are provided with NaN. So we need to take care of those fields and we must remove some of the unneccesary columns which may be irrelevant for our analysis.

1: In the column targertype1\_txt we see that some of the values are given as NaN. We can fill then with the most frequent target type (Private citizen) as per the given data set.

|  |  |
| --- | --- |
| Private citizens & property | 11780 |
| Military | 9419 |
| Police | 6691 |
| Government(general) | 3510 |
| Business | 3124 |
| Unknown | 2412 |

2: in the column ‘natlty1\_txt’ we see some values are missing . we can fill it up by the most frequent nationality of that country. For example

Afghanistan

|  |  |
| --- | --- |
| Nationality | Count |
| Afganistan | 5316 |

3: in the column ‘corp1’ we see that some samples have data as ‘Not Applicable’ we will replace them with NaN for our better computational efficiency.

Not Applicable count count in corp1: 737

4: in the column ‘gname’ we see that the maximum frequency of data lies in the unknown category . so we find an outlier . but we cannot remove it as it is difficult to predict the group name.

|  |  |
| --- | --- |
| Unknown | 12210 |
| Taliban | 4900 |
| Islamic State of Iraq and the Levant(ISIL) | 4273 |
| Al-Shabaab | 2350 |
| Boko Haram | 1921 |
| Maoists | 1124 |
| New People’s Army(NPA) | 1076 |

5: in the column weap\_detail is unneccesary as it is a mere decription of firearm type.

6.in the column propextent\_txt it will be useful to leave the NaN variables alone .

7: the column ransomamt is unneccesary as we see from the dataset that most of the terrorism didnt result to ransom amt.

no of nan values in ransomant : 43642

8: Now we will rename the columns as follows and drop ransomamt column

|  |  |
| --- | --- |
| From | To |
| iyear' | ‘year', |
| imonth' | 'month' |
| 'iday' | 'day' |
| 'country\_txt | 'country' |
| ‘attacktype1\_txt’ | 'attack1' |
| 'targtype1\_txt' | 'target0',' |
| natlty1\_txt' | 'nationality' |
| 'weaptype1\_txt' | ‘weapon' |
| 'weapsubtype1\_txt' | 'weapon\_sub' |
| 'propextent\_txt' | 'damage' |

Parameters:

1. To see distribution of global terrorismthroughout years.
2. To see the top 10 countries affected by global terrorism.
3. We see the distribution of targets attacked by the terrorist.
4. To see the distribution of the types of attack by the terrorist.
5. To display the count of each attack types for each country.
6. To display the more vulnerable terrorist group.
7. To display the distribution of weapons of attack.
8. To display the favored way of attack of the terrorists.
9. To display which targets are more vurnerable to terrorist attacks.
10. To display how each city of each country are vurnerable to terrorista attcks.
11. To see the distribution of the location of attacks and hence find the most vurnerable places of terrorism.
12. To generate the attack probabilities of each country .
13. To find the distribution of civilian casualitie of each country.
14. To see the distribution of domestic and international casualties for each country.
15. To see the distribution of casualties for each terrorism attacks.
16. To see the activities of the terrorist group in each year.
17. To see decrease or increase in terrorist attack of a country.
18. To find most active terrorist group.
19. To see which target types are most vulnerable to terrorist attack.
20. To see amount of casualties occurred by each terrorist group.
21. To see amount of casualties occurred in each country.

Additional objectives:

1. Try to predict some data from the data model we created.
2. Try to find a pattern how the known terrorist attacks are made and make a prediction.
3. Try to make a prediction of attack of a given year by generating a pattern of attack from the given data.
4. To see which countries will be at peace in future and free from terrorism .