

## Case Study on Data manipulation – Python

This case study is to get familiarized with Data manipulation using Pandas. The dataset involves records of terrorist attacks that took place around the world since 1970. Description of columns are given wherever necessary along with hints to answer the questions.

Answer the following questions using the dataset: terror\_gtd.csv

1. Which country had highest attacks since 1970? How many attacks happened in India?  
Hint: Use the country\_txt column.
2. How many attacks happened in India and upto 7 people were killed?  
Hint: Use the country\_txt and nkill column.  
nkill : Total Number of Fatalities
3. Extract any 4 cities and summary for attacks above  
Hint: Use country\_txt, nkill,city and summary columns
4. In a single terror incident in India, find out top 3 cities by number killed  
Hint: Use each row as a unique terror incident.  
Use country\_txt, nkill and city columns
5. In a single terror incident in India, find out top 5 cities by number killed and wounded  
Hint: Use each row as a unique terror incident. Use country\_txt, nkill, nwound and city columns  
nwound : Total Number of Injured
6. Label all the incidents where the number killed was more than 10 as severe. How many incidents were Severe?  
  
Hint: Use Aggregations and manipulations using apply and map  
#You can use lambda functions as well.  
Use the nkill column.

7. How many attacks were successful that were suicide attacks? Also, how many suicide attacks were not successful?

Hint: Use suicide and success columns

1 = "Yes" The incident was a suicide attack.

0 = "No" There is no indication that the incident was a suicide attack

1 = "Yes" The incident was successful.

0 = "No" The incident was not successful

8. Write a function to label an incident that was both successful and suicidal

Hint: We can use apply to use a function column wise where you can use columns success and suicide in the data.

9. Create a new category representing if the incident occurred in Afghanistan, Pakistan or India as one level of the category and all the other countries as another level. Label all other countries as ROW and new column which contains the new category as 'Local' in the data.

Hint: Create a new category representing if the incident occurred in Afghanistan, Pakistan or India as one level of the category and all the other countries as another level. Use the columns – country\_txt and create a new column called "Local" in the dataset.

10. How many incidents happened in Af-Pak-India vs ROW?

Hint: Use the newly created column "Local"