PYTHON PROJECT

REAL WORLD DATA ANALYSIS

REPORT FOR THE DATA ANALYSIS OF THE DRONE ATTACKS IN PAKISTAN

1) MUHAMMAD ATEEB

2) MUHAMMAD RESHAIL HUSSAIN

3) SYED IBBAD ASIF

22

TABLE OF CONTENT

INTRODUCTION

PURPOSE OF THE REPORT:

The aim of this thesis is to discover how to analyses data using jupyter with different data sets. The proposal of this report to analyses datasets of sales of product in the whole year by using Python´s libraries .Here we investigate data to utilizes logical techniques, procedures, calculations and frameworks to separate information Knowledge from organized and unstructured information which is identified with data mining and big data.

DATA SCIENCE:

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data and apply knowledge and actionable insights from data across a broad range of application domains. Data science is related to data mining, machine learning and big data

Data science is a "concept to unify statistics data analysis informatics and their related methods"

DATA SCIENTIST:

Data scientist are a new breed of analytical data expert who have the technical skills to solve complex problems – and the curiosity to explore what problems need to be solve

Why is Data Science Important?

Data creates magic. Industries need data to help them make careful decisions. Data Science churns raw data into meaningful insights. Therefore, industries need data science. A Data Scientist is a wizard who knows how to create magic using data. A skilled Data Scientist will know how to dig out meaningful information with whatever data he comes across. He helps the company in the right direction. The company requires strong data-driven decisions at which he’s an expert.

WHY DATA SCIENTIST ARE IMPORTANT

Data Scientists help the company to acquire customers by analyzing their needs. This allows the companies to tailor products best suited for the requirements of their potential customers. Data holds the key for companies to understand their clients. Therefore, the purpose of a Data Scientist here is to enable companies to recognize clients and help them deliver the needs of their customers.

(1)DATA ANALYSIS:

Is a process of inspecting, data cleaning, transforming, and interpreting data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effective.

1.1) THE PROCESS OF DATA ANALYSIS:

- DATA REQUIRMENTS
- DATA COLLECTION
- DATA PROCESSING
- DATA SORTING
- EXPLORATERY DATA
- ALGORITHMS
- RESULT

DATA REQUIRMENT:

The data is necessary as inputs to the analysis, which is specified based upon the requirements of those directing the analysis (or customers, who will use the finished product of the analysis).The general type of entity upon which the data will be collected is referred to as an experiment. Specific variables regarding a population) may be specified and obtained. Data may be numerical or categorical

EXAMPLES

- a person or population of people
- age and income
- a text label for numbers

DATA COLLECTION:

Data is collected from a variety of sources. The requirements may be communicated by analysis of the data; such as, Technology personnel information within an organization.[1] The data may also be collected from sensors in the environment, including traffic cameras, satellites, recording devices, etc. It may also be obtained through interviews, downloads from online sources, or reading documentation.

DATA PROCESSING:

Data, when initially obtained, must be processed or organized for analysis. For instance, these may involve placing data into rows and columns in a table format (*known as* structured datafor further analysis, often through the use of spreadsheet or statistical software.

DATA SORTING:

Once processed and organized, the data may be incomplete, contain duplicates, or contain errors. The need for *data cleaning* will arise from problems in the way that the datum are entered and stored. Data cleaning is the process of preventing and correcting these errors. Common tasks include record matching, identifying inaccuracy of data, and overall quality of existing data, reduplication, and column segmentation. Such data problems can also be identified through a variety of analytical techniques. For example; with financial information, the totals for particular variables may be compared against separately published numbers that are believed to be reliable. Unusual amounts, above or below predetermined thresholds, may also be reviewed. There are several types of data cleaning that are dependent upon the type of data in the set; this could be phone numbers, email addresses, employers, or other values. Quantitative data methods for outlier detection, can be used to get rid of data that appears to have a higher likelihood of being input incorrectly. Textual data spell checkers can be used to lessen the amount of miss-typed words. However, it is harder to tell if the word themselves are correct.

EXPLORATORY DATA:

Once the datasets are cleaned, they can then be analyzed. Analysts may apply a variety of techniques, referred to as exploratory data, to begin understanding the messages contained within the obtained data. The process of data exploration may result in additional data cleaning or additional requests for data; thus, the initialization of the *iterative phases* mentioned in the lead paragraph of this section. such as, the average or median, can be generated to aid in understanding the data visualization is also a technique used, in which the analyst is able to examine the data in a graphical format in order to obtain additional insights, regarding the messages within the data.

ALGORITHMS:

Mathematical formulas or models or algorithms, may be applied to the data in order to identify relationships among the variables; for example, using correlation or causation. In general terms, models may be developed to evaluate a specific variable based on other variable(s) contained within the dataset, with some *residual error* depending on the implemented model's accuracy (*e.g.*, Data = Model + Error)

RESULTS (product):

A **data product** is a computer application that takes *data inputs* and generates *outputs*, feeding them back into the environment. It may be based on a model or algorithm. For instance, an application that analyzes data about customer purchase history, and uses the results to recommend other purchases the customer might enjoy.

What is Data Analytics?

As the process of analyzing raw data to find trends and answer questions, the definition of data analytics captures its broad scope of the field. However, it includes many techniques with many different goals.

1.2) ROLE OF PYTHON:

Python is the leading programming language platform in the current era and it’s the most advance of all the platform till now Python make it a perfect tool for processing complex data. Python can also easily penetrate patterns, correlate information in extensive sets, and provide better insights, in addition to other critical matrices in evaluating performance.

What the future is expected to bring in Data Science?

The future of data science is promising for those with the right skill set pursuing it as a career. It is set to revolutionize many sectors such as health care, transport, business, finance and manufacturing industries through Artificial Intelligence and automation.

1.3DATA SCIENCE IN FILED OF FINANCE:

Data Science has become very important in the Finance Industry, which is mostly used for Better Risk Management and Risk Analysis. Better analysis leads to better decisions which lead to an increase in profit for financial institutions.

Companies also analyze the trends in data through business intelligence tools. Accuracy in the detection of anomalies and fraud have improved with the use of Data Science. This has helped to reduce risks & scam, minimizing the losses and saving the reputation of the financial institution.

How can one imply data science in the finance sector?

Data science is tremendously efficient when it comes to the finance sector. It provides modern approaches to the companies. To be able to implement data science for finance, one must excel in the skills that are mentioned here: Statistics and Probability are the two most important mathematical concepts of Data Science. Descriptive statistics including mean, median, and mode, linear regression, hypothesis testing are some of the topics of statistics and probability. You must go with one programming language and master it to code in it. There are plenty of languages out there but Python is the most preferable language due to the libraries and modules it provides. Machine Learning and Deep Learning are two separate domains and the subsets of Data Science at the same time. These topics will help you to get far in data science

1.4) LIBRARIES:

1) PANDAS:

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

2) NUMPY:

Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

3) METPLOTLIB:

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter,

4) MATPLOT INLINE:

Matplotlib inline sets the backend of matplotlib to the 'inline' backend: With this backend, the **output of plotting commands** is displayed inline within frontends like the Jupyter notebook, directly below the code cell that produced it. The resulting plots will then also be stored in the notebook document.

5) SUBPROCESS:

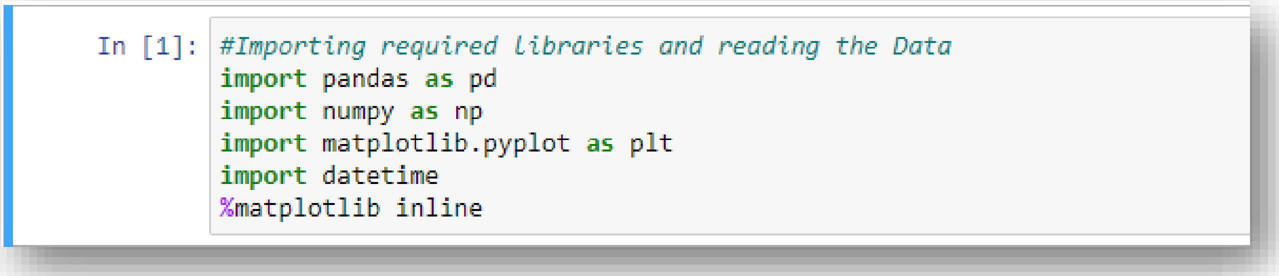
Sub process in Python is **a module used to run new codes and applications by creating new processes**. It lets you start new applications right from the Python program you are currently writing. ... You can also get exit codes and input, output, or error pipes using sub process in Python.

(2)CLEANING THE DATA:

WHAT IS DATA CLEANING

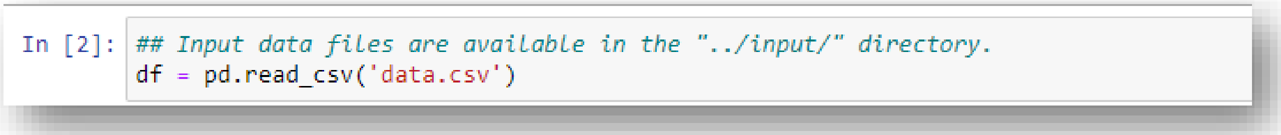
Data cleaning or cleansing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

2.1IMPORTING REQUIRED LIBRARIE



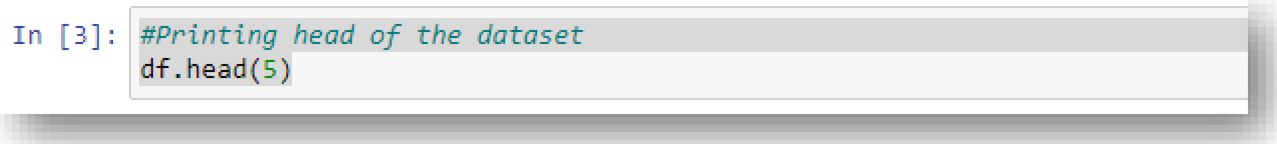
```
#Importing required libraries and reading the Data
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import datetime
%matplotlib inline
```

2.2READING CSV FILE

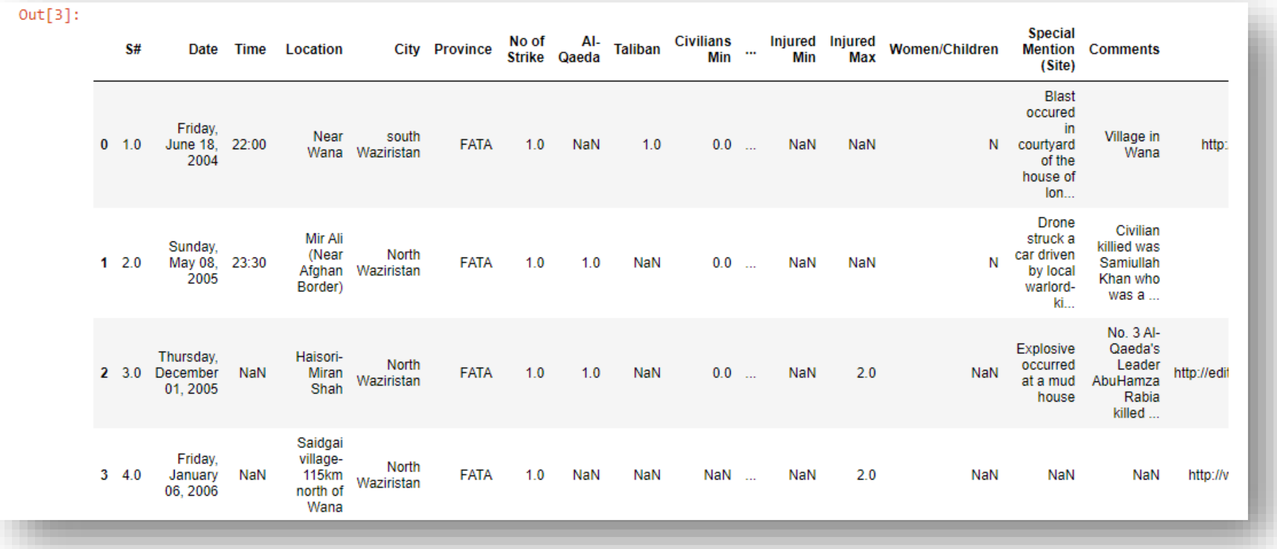


```
In [2]: ## Input data files are available in the "../input/" directory.
        df = pd.read_csv('data.csv')
```

2.3PRINTING HEAD OF THE DATA SET



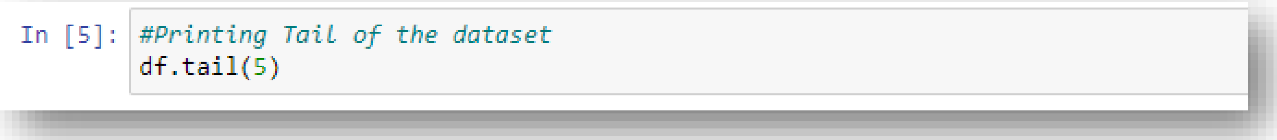
```
In [3]: #Printing head of the dataset
        df.head(5)
```



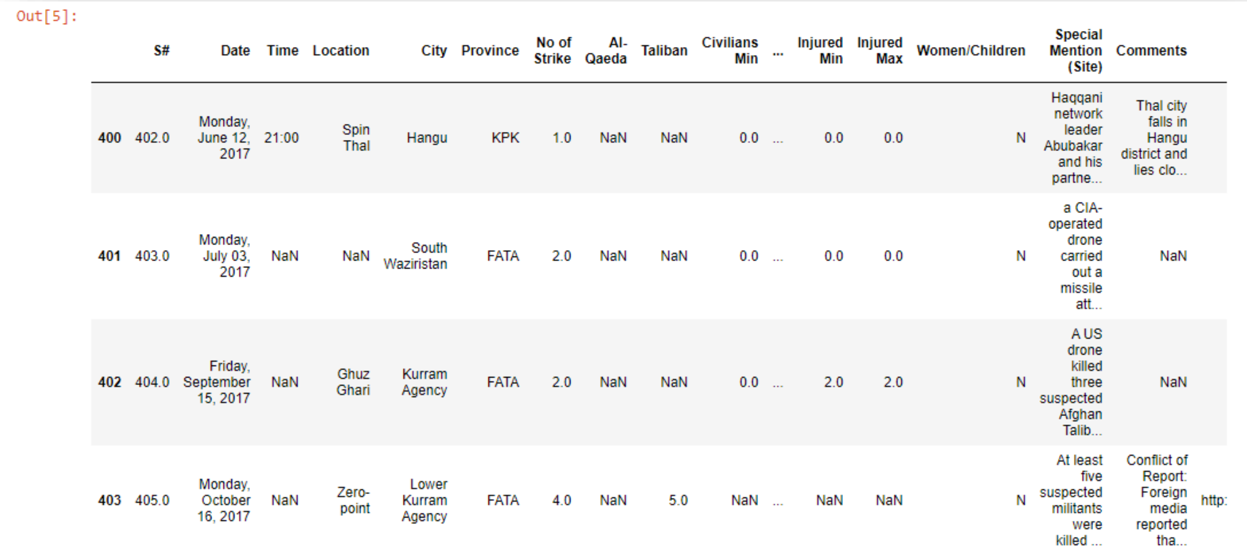
Out[3]:

	S#	Date	Time	Location	City	Province	No of Strike	Al-Qaeda	Taliban	Civilians Min	...	Injured Min	Injured Max	Women/Children	Special Mention (Site)	Comments	
0	1.0	Friday, June 18, 2004	22:00	Near Wana	south Waziristan	FATA	1.0	NaN	1.0	0.0	...	NaN	NaN	N	Blast occured in courtyard of the house of lon...	Village in Wana	http:
1	2.0	Sunday, May 08, 2005	23:30	Mir Ali (Near Afghan Border)	North Waziristan	FATA	1.0	1.0	NaN	0.0	...	NaN	NaN	N	Drone struck a car driven by local warlord-ki...	Civilian killied was Samiullah Khan who was a ...	
2	3.0	Thursday, December 01, 2005	NaN	Haisori-Miran Shah	North Waziristan	FATA	1.0	1.0	NaN	0.0	...	NaN	2.0	NaN	Explosive occurred at a mud house	No. 3 Al-Qaeda's Leader AbuHamza Rabia killed ...	http://edit
3	4.0	Friday, January 06, 2006	NaN	Saidgai village-115km north of Wana	North Waziristan	FATA	1.0	NaN	NaN	NaN	...	NaN	2.0	NaN	NaN	NaN	http://v

2.4PRINTING TAIL OF THE DATA SET



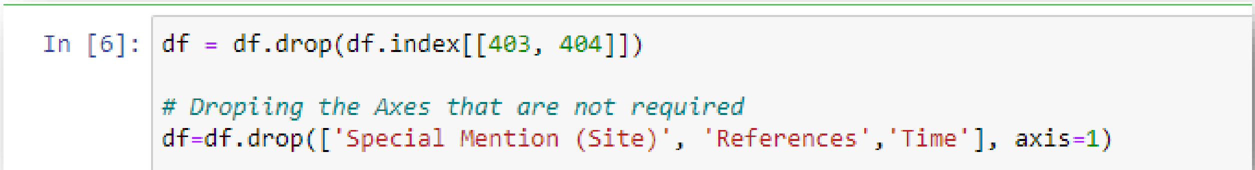
```
In [5]: #Printing Tail of the dataset
        df.tail(5)
```



Out[5]:

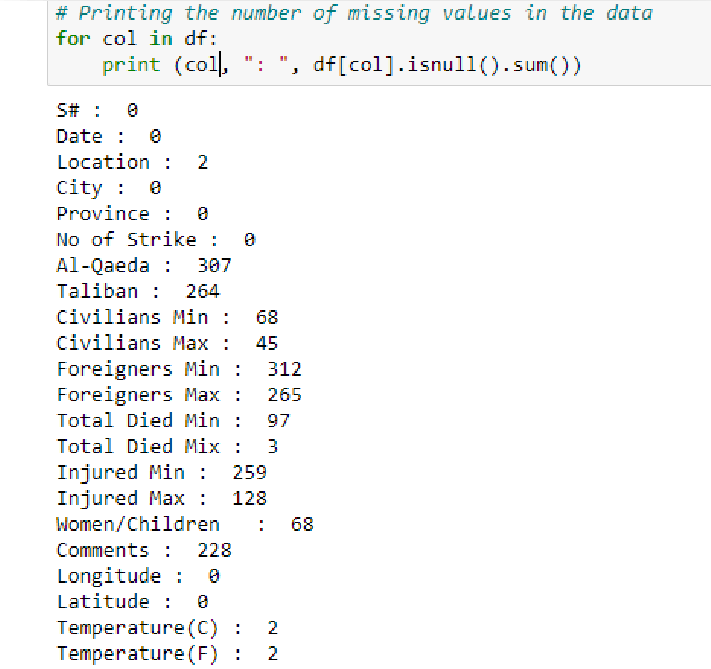
	S#	Date	Time	Location	City	Province	No of Strike	Al-Qaeda	Taliban	Civilians Min	...	Injured Min	Injured Max	Women/Children	Special Mention (Site)	Comments	
400	402.0	Monday, June 12, 2017	21:00	Spin Thal	Hangu	KPK	1.0	NaN	NaN	0.0	...	0.0	0.0	N	Haqqani network leader Abubakar and his partne...	Thal city falls in Hangu district and lies clo...	
401	403.0	Monday, July 03, 2017	NaN	NaN	South Waziristan	FATA	2.0	NaN	NaN	0.0	...	0.0	0.0	N	a CIA-operated drone carried out a missile att...	NaN	
402	404.0	Friday, September 15, 2017	NaN	Ghuz Ghari	Kurram Agency	FATA	2.0	NaN	NaN	0.0	...	2.0	2.0	N	A US drone killed three suspected Afghan Talib...	NaN	
403	405.0	Monday, October 16, 2017	NaN	Zero-point	Lower Kurram Agency	FATA	4.0	NaN	5.0	NaN	...	NaN	NaN	N	At least five suspected militants were killed ...	Conflict of Report: Foreign media reported tha...	http:

2.5DROPING EXCESIVE COLUMN



```
In [6]: df = df.drop(df.index[[403, 404]])

        # Dropiing the Axes that are not required
        df=df.drop(['Special Mention (Site)', 'References','Time'], axis=1)
```

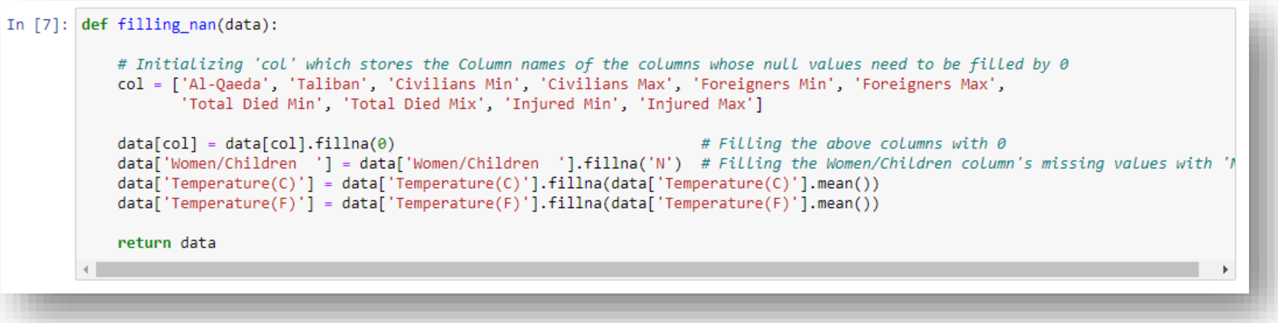
2.6FINDING OMITTED VALUES



```
# Printing the number of missing values in the data
for col in df:
    print (col, ": ", df[col].isnull().sum())
```

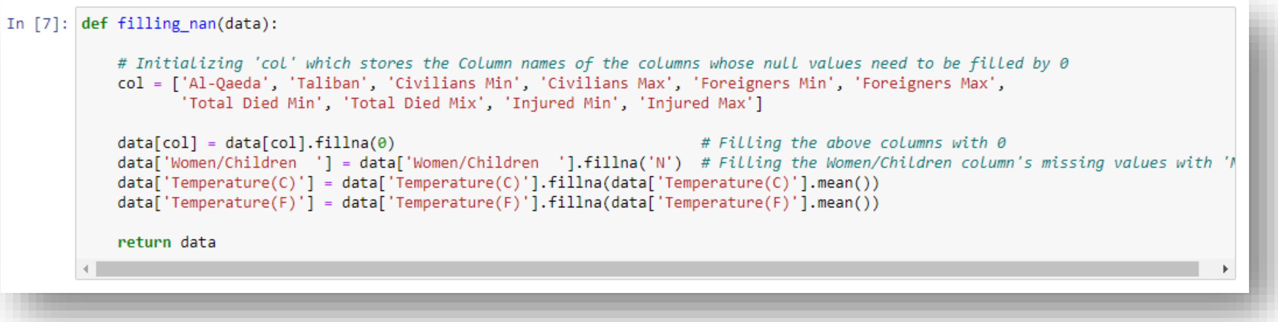
```
S# :  0
Date :  0
Location :  2
City :  0
Province :  0
No of Strike :  0
Al-Qaeda :  307
Taliban :  264
Civilians Min :  68
Civilians Max :  45
Foreigners Min :  312
Foreigners Max :  265
Total Died Min :  97
Total Died Mix :  3
Injured Min :  259
Injured Max :  128
Women/Children   :  68
Comments :  228
Longitude :  0
Latitude :  0
Temperature(C) :  2
Temperature(F) :  2
```

2.7FILLING OMITTED VALUES



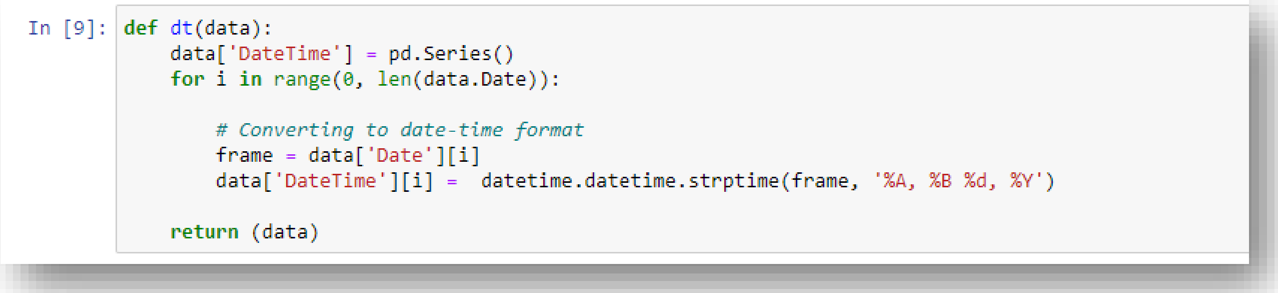
```
In [7]: def filling_nan(data):

            # Initializing 'col' which stores the Column names of the columns whose null values need to be filled by 0
            col = ['Al-Qaeda', 'Taliban', 'Civilians Min', 'Civilians Max', 'Foreigners Min', 'Foreigners Max',
                   'Total Died Min', 'Total Died Mix', 'Injured Min', 'Injured Max']

            data[col] = data[col].fillna(0)                                  # Filling the above columns with 0
            data['Women/Children  '] = data['Women/Children  '].fillna('N')  # Filling the Women/Children column's missing values with '
            data['Temperature(C)'] = data['Temperature(C)'].fillna(data['Temperature(C)'].mean())
            data['Temperature(F)'] = data['Temperature(F)'].fillna(data['Temperature(F)'].mean())

            return data
```

2.8RENOVATING OMITTED VALUES BY MEAN



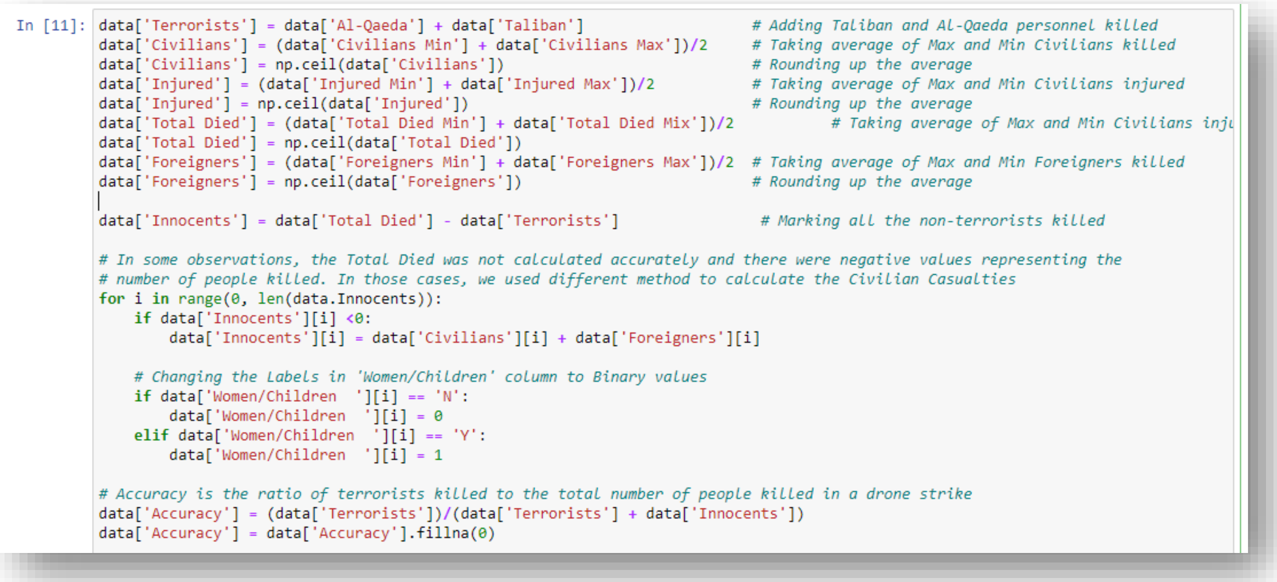
```
In [7]: def filling_nan(data):

            # Initializing 'col' which stores the Column names of the columns whose null values need to be filled by 0
            col = ['Al-Qaeda', 'Taliban', 'Civilians Min', 'Civilians Max', 'Foreigners Min', 'Foreigners Max',
                   'Total Died Min', 'Total Died Mix', 'Injured Min', 'Injured Max']

            data[col] = data[col].fillna(0)                                  # Filling the above columns with 0
            data['Women/Children  '] = data['Women/Children  '].fillna('N')  # Filling the Women/Children column's missing values with '
            data['Temperature(C)'] = data['Temperature(C)'].fillna(data['Temperature(C)'].mean())
            data['Temperature(F)'] = data['Temperature(F)'].fillna(data['Temperature(F)'].mean())

            return data
```

2.9CHANGING DATE AND TIME FORMATE

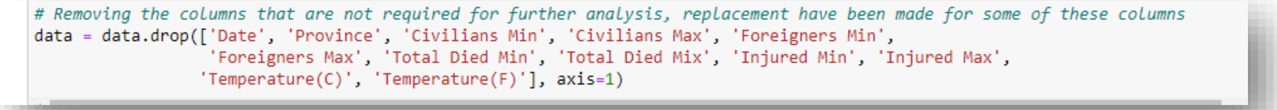


```
In [9]: def dt(data):
            data['DateTime'] = pd.Series()
            for i in range(0, len(data.Date)):

                # Converting to date-time format
                frame = data['Date'][i]
                data['DateTime'][i] =  datetime.datetime.strptime(frame, '%A, %B %d, %Y')

            return (data)
```

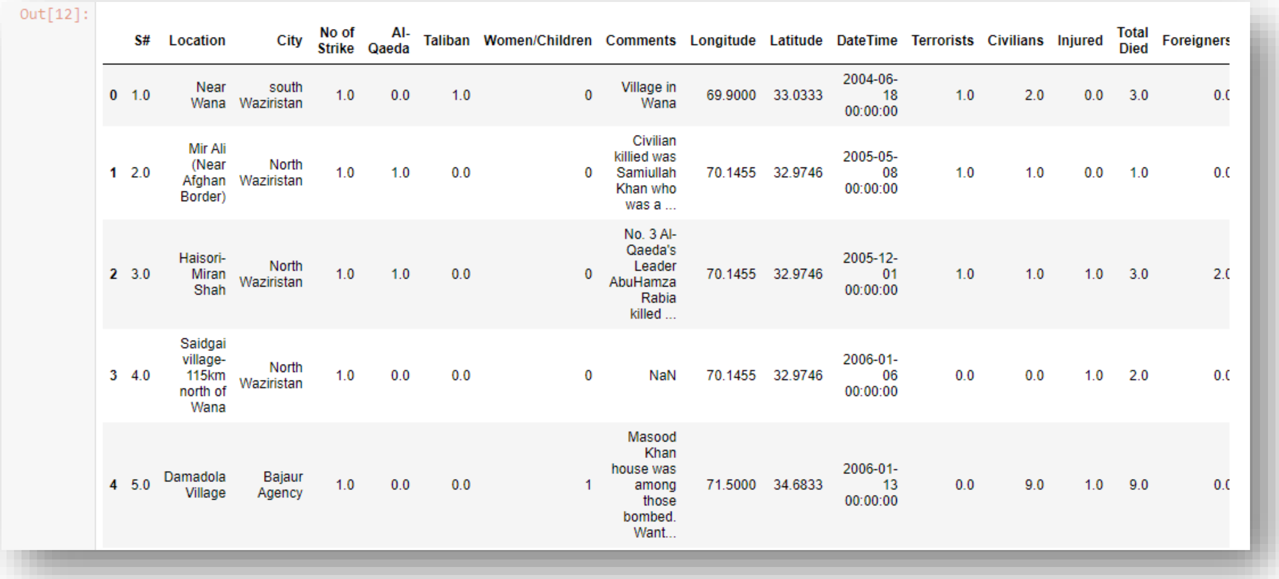
2.10COMBINING TERRORIST AND CIVILIAN DATA



```
In [11]: data['Terrorists'] = data['Al-Qaeda'] + data['Taliban']                   # Adding Taliban and Al-Qaeda personnel killed
         data['Civilians'] = (data['Civilians Min'] + data['Civilians Max'])/2     # Taking average of Max and Min Civilians killed
         data['Civilians'] = np.ceil(data['Civilians'])                            # Rounding up the average
         data['Injured'] = (data['Injured Min'] + data['Injured Max'])/2           # Taking average of Max and Min Civilians injured
         data['Injured'] = np.ceil(data['Injured'])                                # Rounding up the average
         data['Total Died'] = (data['Total Died Min'] + data['Total Died Mix'])/2            # Taking average of Max and Min Civilians inju
         data['Total Died'] = np.ceil(data['Total Died'])
         data['Foreigners'] = (data['Foreigners Min'] + data['Foreigners Max'])/2  # Taking average of Max and Min Foreigners killed
         data['Foreigners'] = np.ceil(data['Foreigners'])                          # Rounding up the average

         data['Innocents'] = data['Total Died'] - data['Terrorists']                # Marking all the non-terrorists killed

         # In some observations, the Total Died was not calculated accurately and there were negative values representing the
         # number of people killed. In those cases, we used different method to calculate the Civilian Casualties
         for i in range(0, len(data.Innocents)):
             if data['Innocents'][i] <0:
                 data['Innocents'][i] = data['Civilians'][i] + data['Foreigners'][i]

             # Changing the Labels in 'Women/Children' column to Binary values
             if data['Women/Children  '][i] == 'N':
                 data['Women/Children  '][i] = 0
             elif data['Women/Children  '][i] == 'Y':
                 data['Women/Children  '][i] = 1

         # Accuracy is the ratio of terrorists killed to the total number of people killed in a drone strike
         data['Accuracy'] = (data['Terrorists'])/(data['Terrorists'] + data['Innocents'])
         data['Accuracy'] = data['Accuracy'].fillna(0)
```



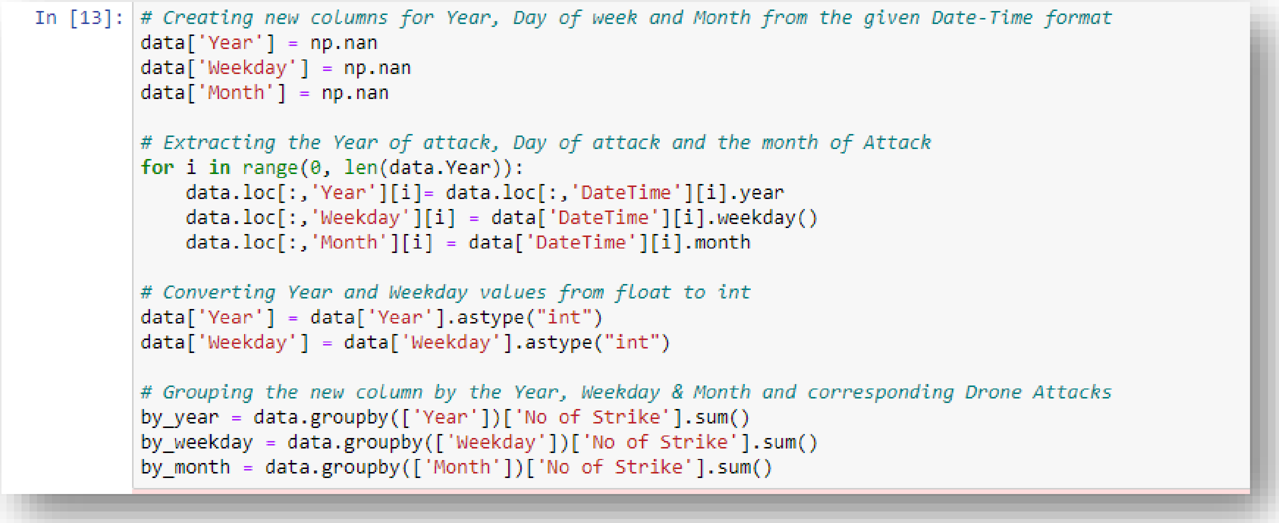
```
# Removing the columns that are not required for further analysis, replacement have been made for some of these columns
data = data.drop(['Date', 'Province', 'Civilians Min', 'Civilians Max', 'Foreigners Min',
                  'Foreigners Max', 'Total Died Min', 'Total Died Mix', 'Injured Min', 'Injured Max',
                 'Temperature(C)', 'Temperature(F)'], axis=1)
```



Out[12]:

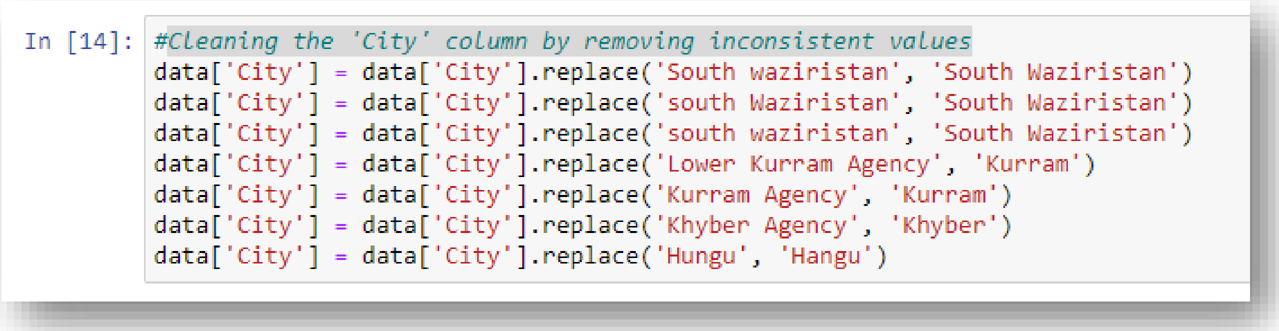
	S#	Location	City	No of Strike	Al-Qaeda	Taliban	Women/Children	Comments	Longitude	Latitude	DateTime	Terrorists	Civilians	Injured	Total Died	Foreigners
0	1.0	Near Wana	south Waziristan	1.0	0.0	1.0	0	Village in Wana	69.9000	33.0333	2004-06-18 00:00:00	1.0	2.0	0.0	3.0	0.
1	2.0	Mir Ali (Near Afghan Border)	North Waziristan	1.0	1.0	0.0	0	Civilian killied was Samiullah Khan who was a ...	70.1455	32.9746	2005-05-08 00:00:00	1.0	1.0	0.0	1.0	0.
2	3.0	Haisori-Miran Shah	North Waziristan	1.0	1.0	0.0	0	No. 3 Al-Qaeda's Leader AbuHamza Rabia killed ...	70.1455	32.9746	2005-12-01 00:00:00	1.0	1.0	1.0	3.0	2.
3	4.0	Saidgai village-115km north of Wana	North Waziristan	1.0	0.0	0.0	0	NaN	70.1455	32.9746	2006-01-06 00:00:00	0.0	0.0	1.0	2.0	0.
4	5.0	Damadola Village	Bajaur Agency	1.0	0.0	0.0	1	Masood Khan house was among those bombed. Want...	71.5000	34.6833	2006-01-13 00:00:00	0.0	9.0	1.0	9.0	0.

<u>2.11Creating new columns for Year, Months, Weeks and Days from the given Date-Time format</u>

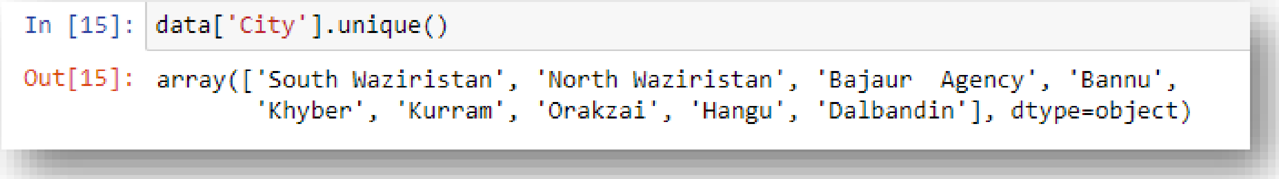


```
In [13]: # Creating new columns for Year, Day of week and Month from the given Date-Time format
         data['Year'] = np.nan
         data['Weekday'] = np.nan
         data['Month'] = np.nan

         # Extracting the Year of attack, Day of attack and the month of Attack
         for i in range(0, len(data.Year)):
             data.loc[:,'Year'][i]= data.loc[:,'DateTime'][i].year
             data.loc[:,'Weekday'][i] = data['DateTime'][i].weekday()
             data.loc[:,'Month'][i] = data['DateTime'][i].month

         # Converting Year and Weekday values from float to int
         data['Year'] = data['Year'].astype("int")
         data['Weekday'] = data['Weekday'].astype("int")

         # Grouping the new column by the Year, Weekday & Month and corresponding Drone Attacks
         by_year = data.groupby(['Year'])['No of Strike'].sum()
         by_weekday = data.groupby(['Weekday'])['No of Strike'].sum()
         by_month = data.groupby(['Month'])['No of Strike'].sum()
```

2.12Cleaning the 'City' column by removing inconsistent values

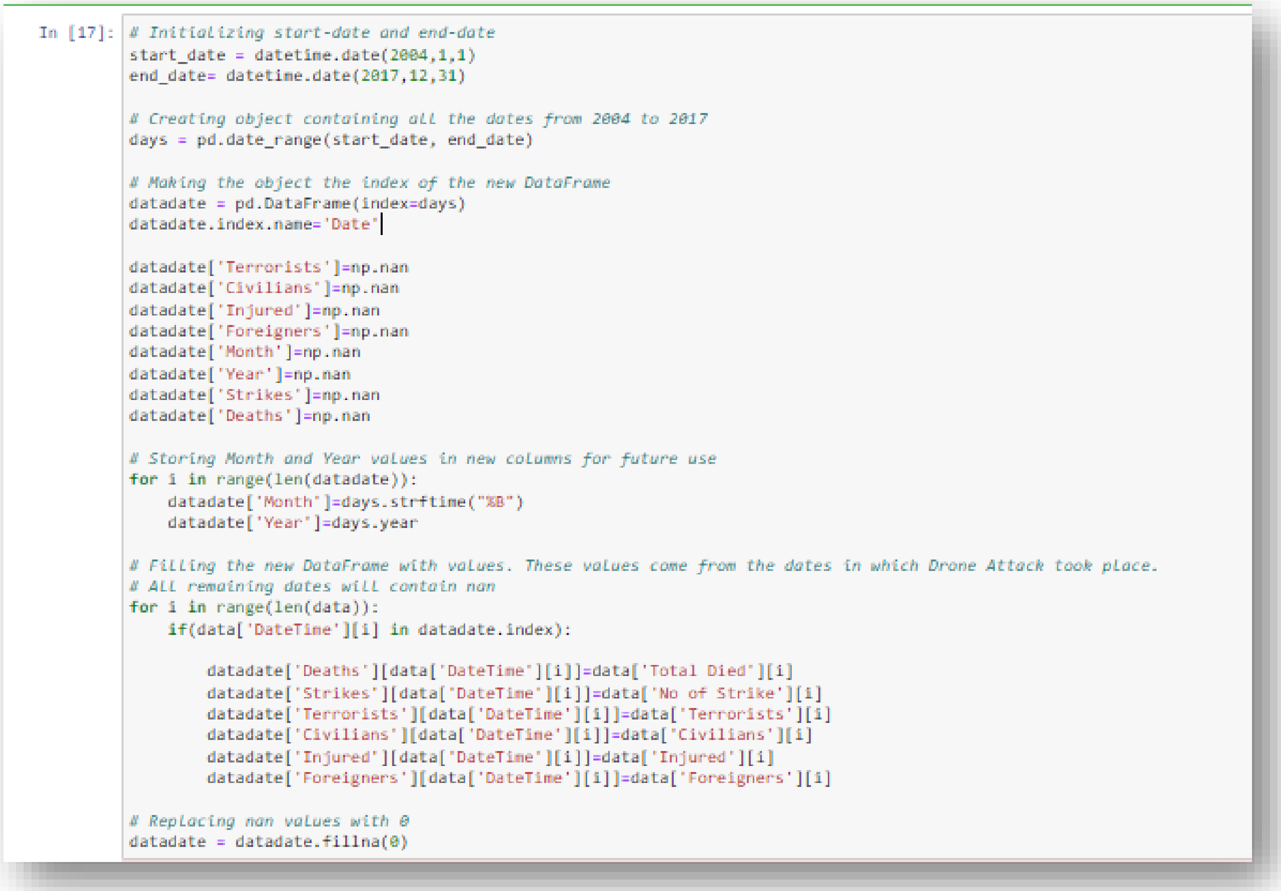


```
In [14]: #Cleaning the 'City' column by removing inconsistent values
         data['City'] = data['City'].replace('South waziristan', 'South Waziristan')
         data['City'] = data['City'].replace('south Waziristan', 'South Waziristan')
         data['City'] = data['City'].replace('south waziristan', 'South Waziristan')
         data['City'] = data['City'].replace('Lower Kurram Agency', 'Kurram')
         data['City'] = data['City'].replace('Kurram Agency', 'Kurram')
         data['City'] = data['City'].replace('Khyber Agency', 'Khyber')
         data['City'] = data['City'].replace('Hungu', 'Hangu')
```

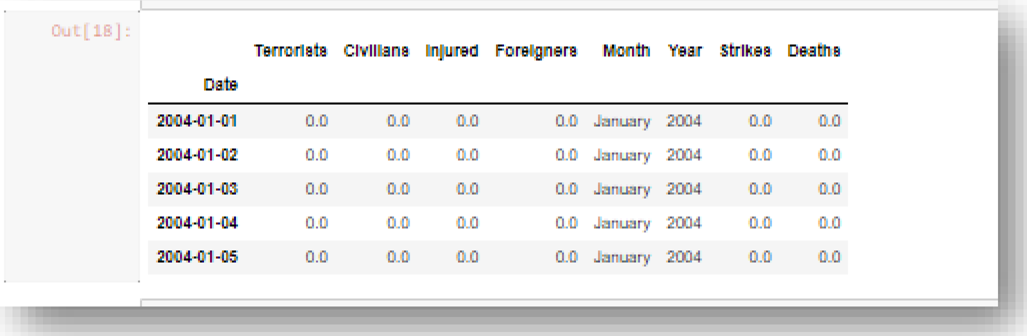


```
In [15]: data['City'].unique()

Out[15]: array(['South Waziristan', 'North Waziristan', 'Bajaur  Agency', 'Bannu',
                'Khyber', 'Kurram', 'Orakzai', 'Hangu', 'Dalbandin'], dtype=object)
```

2.13Creating Year-wise dataset



```
In [17]: # Initializing start-date and end-date
         start_date = datetime.date(2004,1,1)
         end_date= datetime.date(2017,12,31)

         # Creating object containing all the dates from 2004 to 2017
         days = pd.date_range(start_date, end_date)

         # Making the object the index of the new DataFrame
         datadate = pd.DataFrame(index=days)
         datadate.index.name='Date'

         datadate['Terrorists']=np.nan
         datadate['Civilians']=np.nan
         datadate['Injured']=np.nan
         datadate['Foreigners']=np.nan
         datadate['Month']=np.nan
         datadate['Year']=np.nan
         datadate['Strikes']=np.nan
         datadate['Deaths']=np.nan

         # Storing Month and Year values in new columns for future use
         for i in range(len(datadate)):
             datadate['Month']=days.strftime("%B")
             datadate['Year']=days.year

         # Filling the new DataFrame with values. These values come from the dates in which Drone Attack took place.
         # All remaining dates will contain nan
         for i in range(len(data)):
             if(data['DateTime'][i] in datadate.index):

                 datadate['Deaths'][data['DateTime'][i]]=data['Total Died'][i]
                 datadate['Strikes'][data['DateTime'][i]]=data['No of Strike'][i]
                 datadate['Terrorists'][data['DateTime'][i]]=data['Terrorists'][i]
                 datadate['Civilians'][data['DateTime'][i]]=data['Civilians'][i]
                 datadate['Injured'][data['DateTime'][i]]=data['Injured'][i]
                 datadate['Foreigners'][data['DateTime'][i]]=data['Foreigners'][i]

         # Replacing nan values with 0
         datadate = datadate.fillna(0)
```

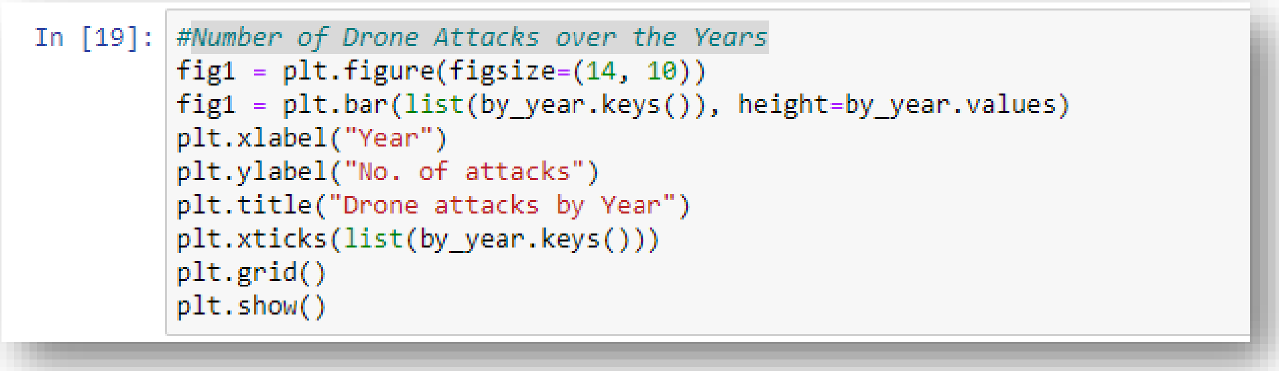


Out[18]:

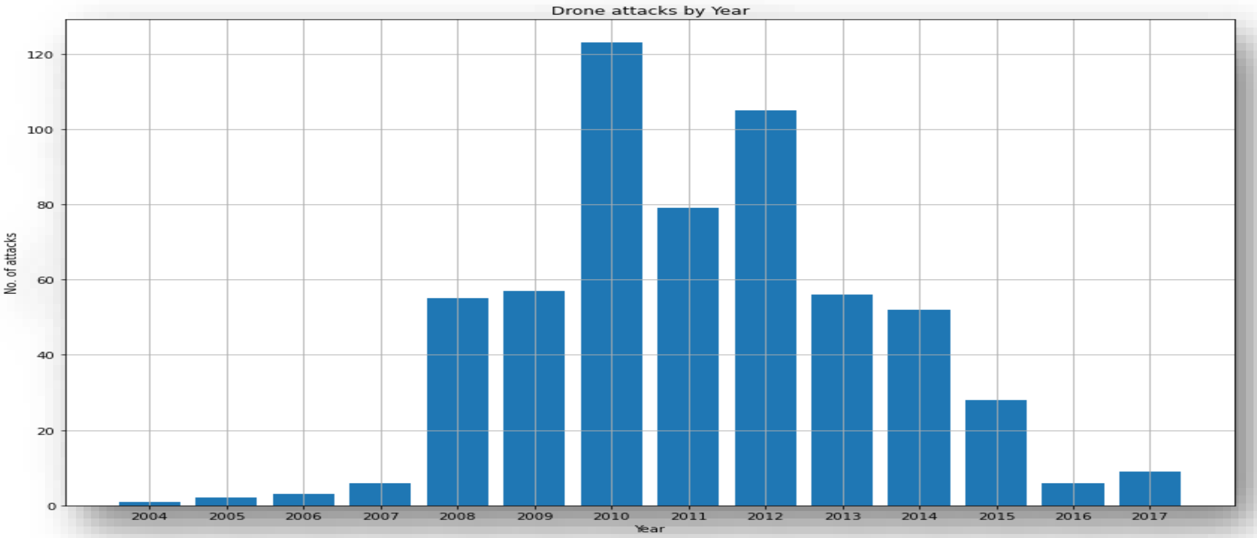
	Terrorists	Civilians	Injured	Foreigners	Month	Year	Strikes	Deaths
Date								
2004-01-01	0.0	0.0	0.0	0.0	January	2004	0.0	0.0
2004-01-02	0.0	0.0	0.0	0.0	January	2004	0.0	0.0
2004-01-03	0.0	0.0	0.0	0.0	January	2004	0.0	0.0
2004-01-04	0.0	0.0	0.0	0.0	January	2004	0.0	0.0
2004-01-05	0.0	0.0	0.0	0.0	January	2004	0.0	0.0

(3)EXPLORATORY DATA ANALYSIS:

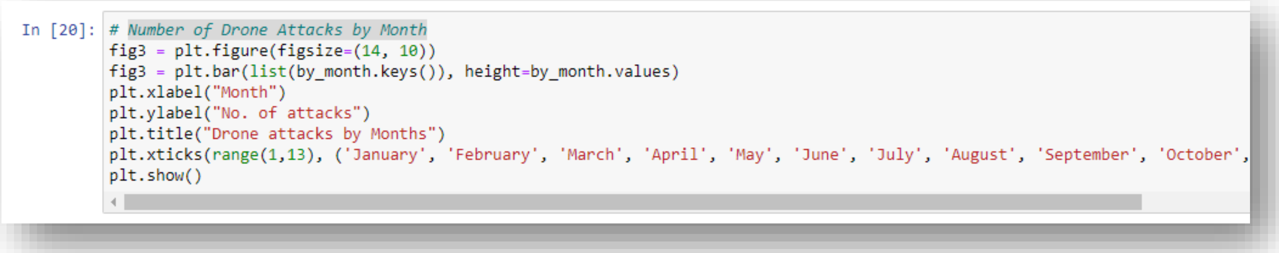
3.1Number of Drone Attacks over the Years



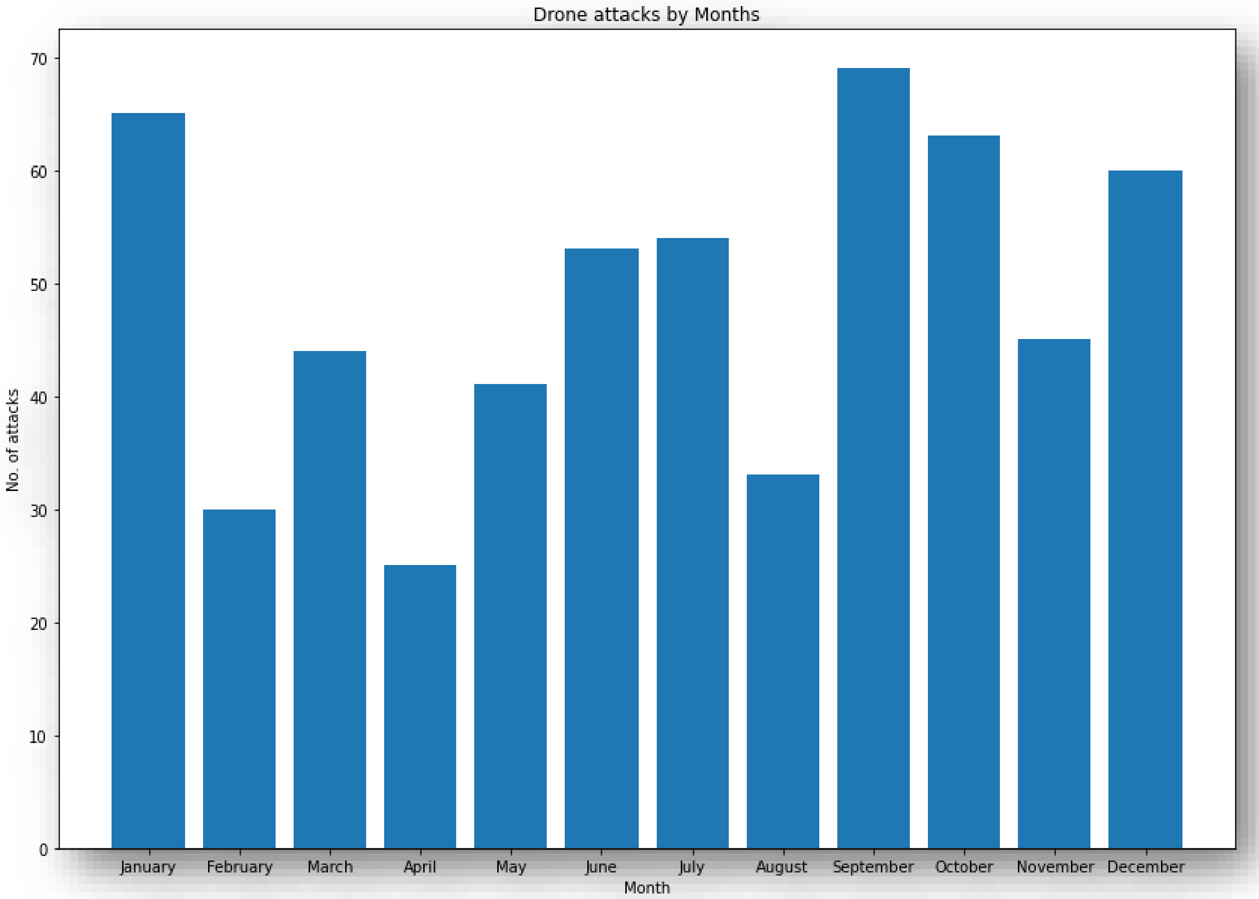
```
In [19]: #Number of Drone Attacks over the Years
         fig1 = plt.figure(figsize=(14, 10))
         fig1 = plt.bar(list(by_year.keys()), height=by_year.values)
         plt.xlabel("Year")
         plt.ylabel("No. of attacks")
         plt.title("Drone attacks by Year")
         plt.xticks(list(by_year.keys()))
         plt.grid()
         plt.show()
```



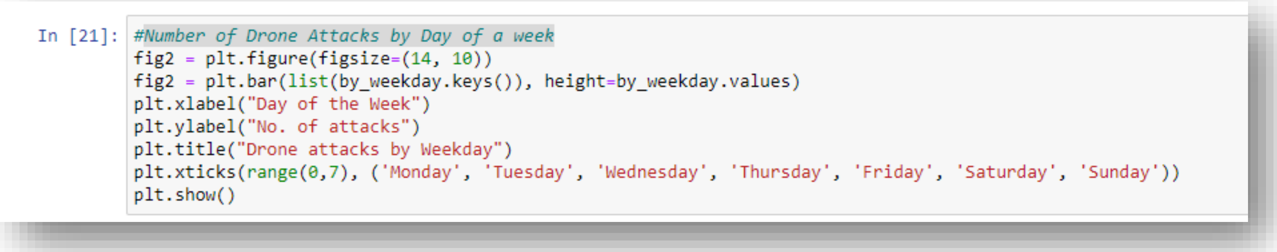
3.2Number of Drone Attacks by Month



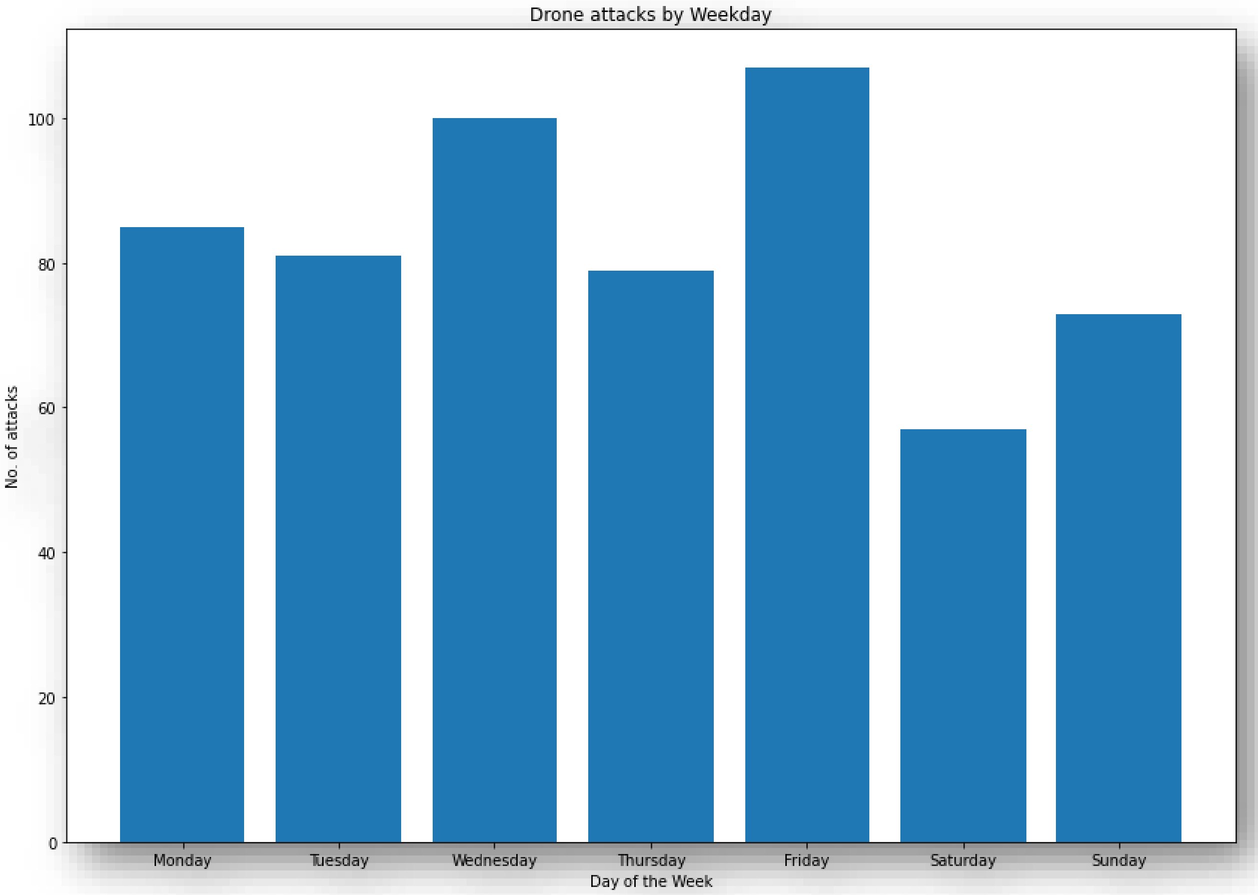
```
In [20]: # Number of Drone Attacks by Month
         fig3 = plt.figure(figsize=(14, 10))
         fig3 = plt.bar(list(by_month.keys()), height=by_month.values)
         plt.xlabel("Month")
         plt.ylabel("No. of attacks")
         plt.title("Drone attacks by Months")
         plt.xticks(range(1,13), ('January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October',
         plt.show()
```



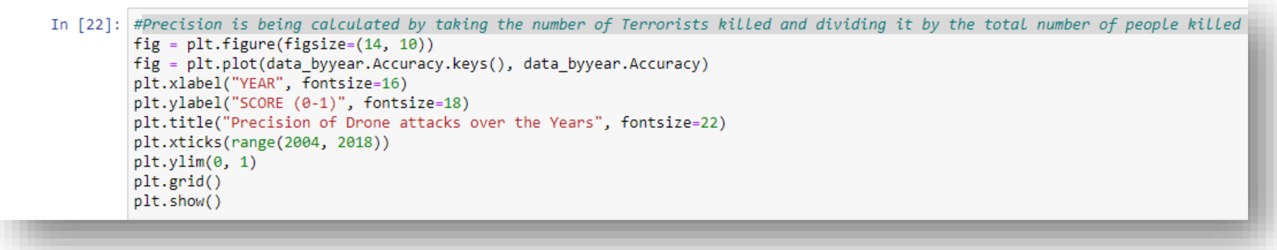
3.3Number of Drone Attacks by Day of a week



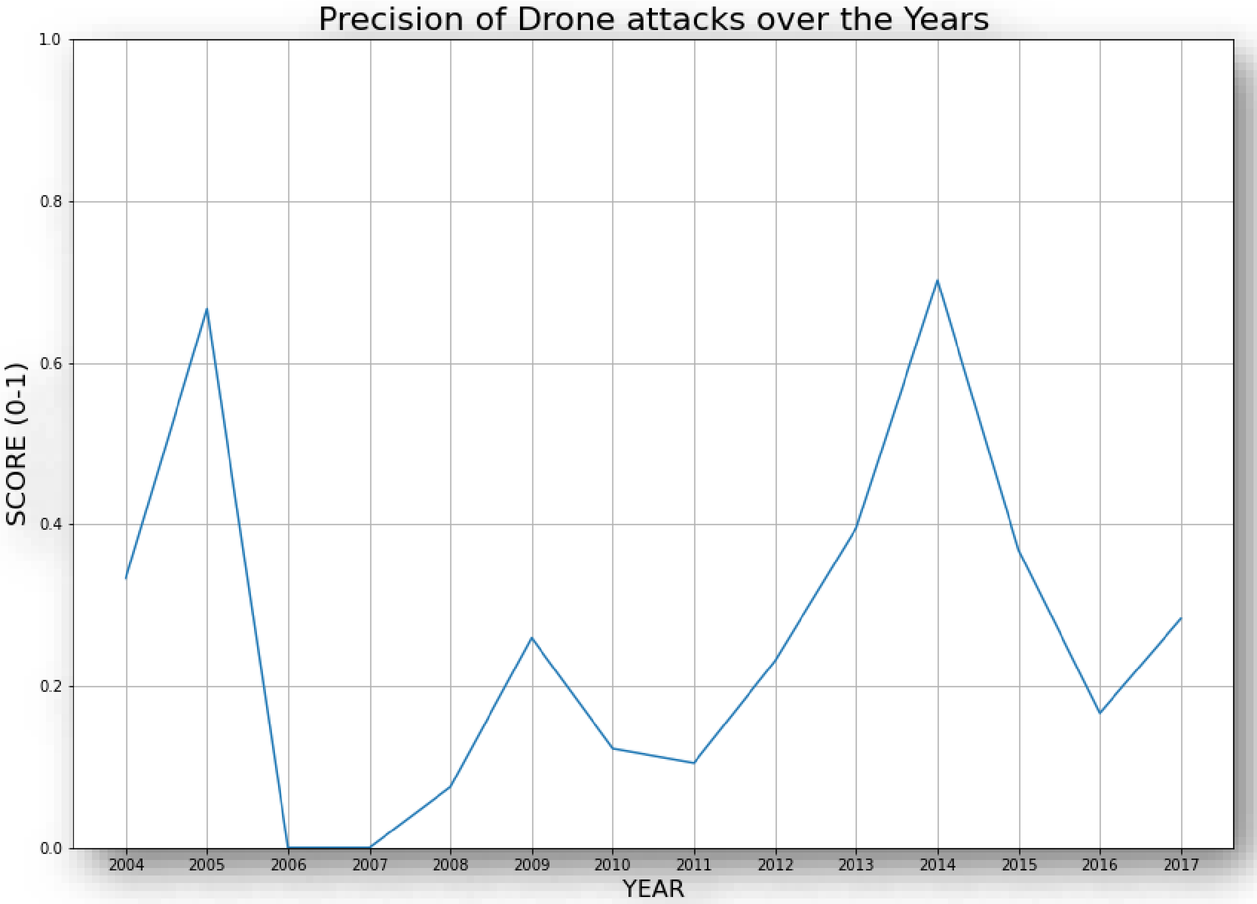
```
In [21]: #Number of Drone Attacks by Day of a week
         fig2 = plt.figure(figsize=(14, 10))
         fig2 = plt.bar(list(by_weekday.keys()), height=by_weekday.values)
         plt.xlabel("Day of the Week")
         plt.ylabel("No. of attacks")
         plt.title("Drone attacks by Weekday")
         plt.xticks(range(0,7), ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'))
         plt.show()
```



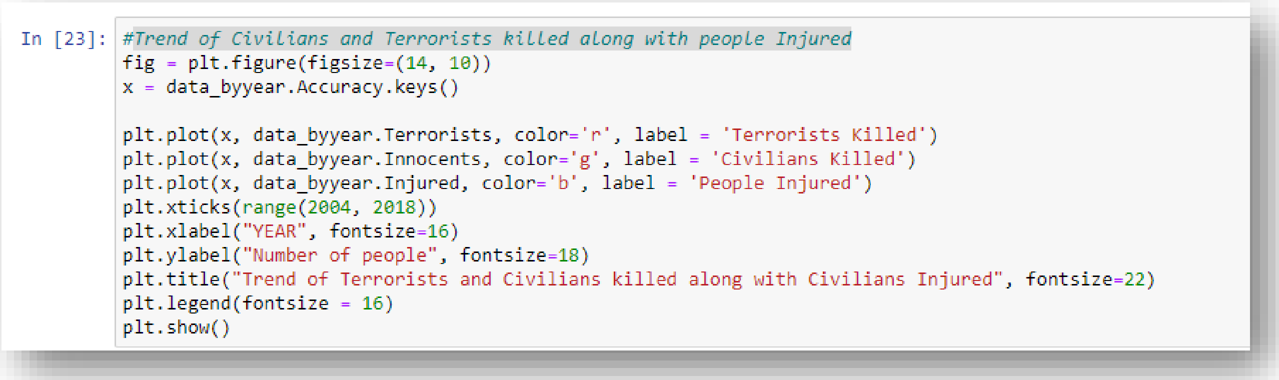
3.4Precision is being calculated by taking the number of Terrorists killed and dividing it by the total number of people killed (including 'Civilians' and 'Foreigners')



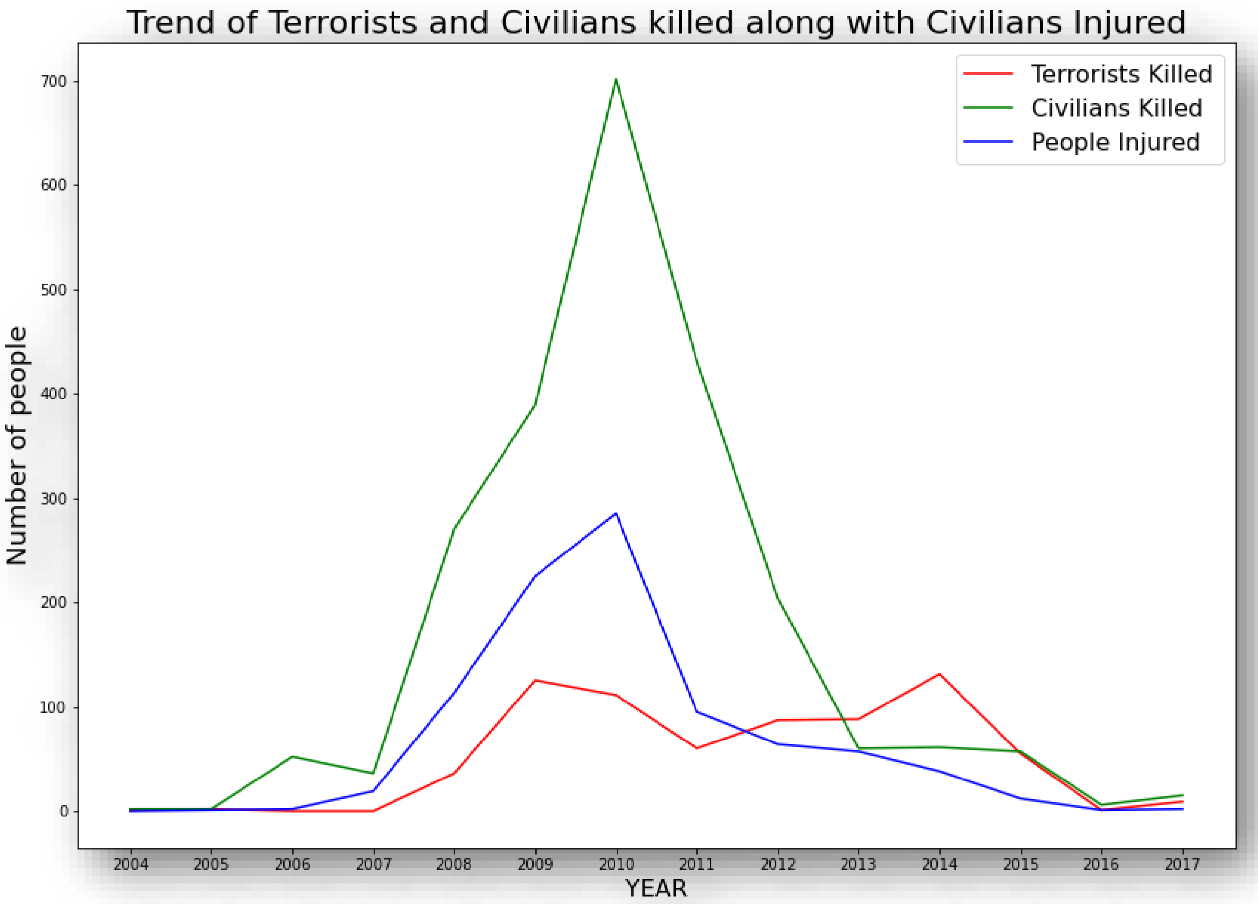
```
In [22]: #Precision is being calculated by taking the number of Terrorists killed and dividing it by the total number of people killed
         fig = plt.figure(figsize=(14, 10))
         fig = plt.plot(data_byyear.Accuracy.keys(), data_byyear.Accuracy)
         plt.xlabel("YEAR", fontsize=16)
         plt.ylabel("SCORE (0-1)", fontsize=18)
         plt.title("Precision of Drone attacks over the Years", fontsize=22)
         plt.xticks(range(2004, 2018))
         plt.ylim(0, 1)
         plt.grid()
         plt.show()
```



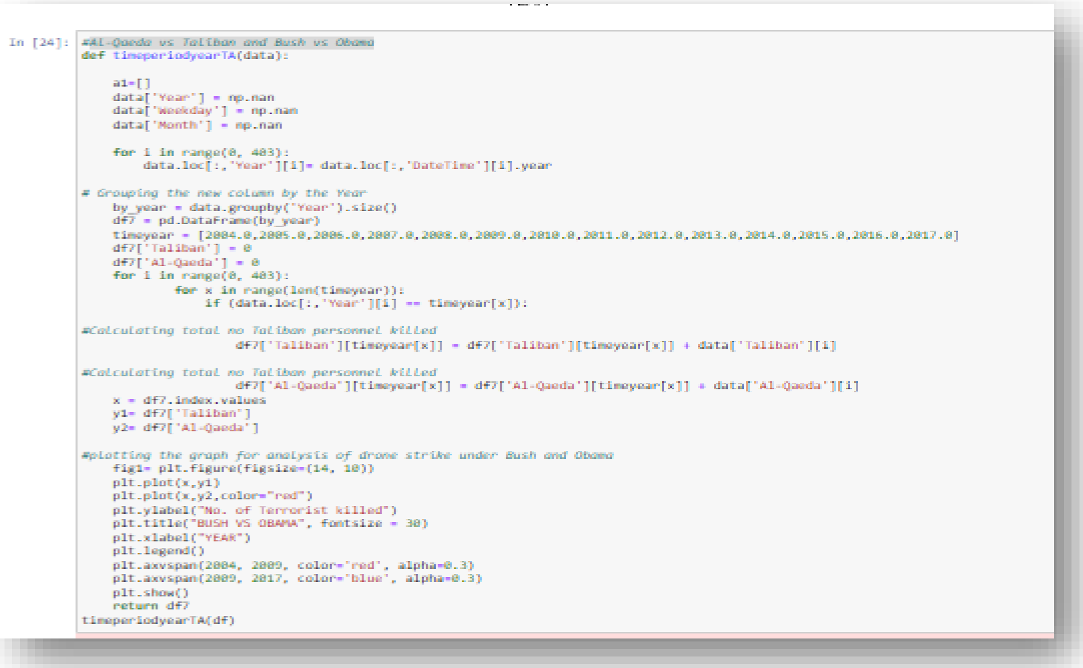
3.5Trend of Civilians and Terrorists killed along with people Injured



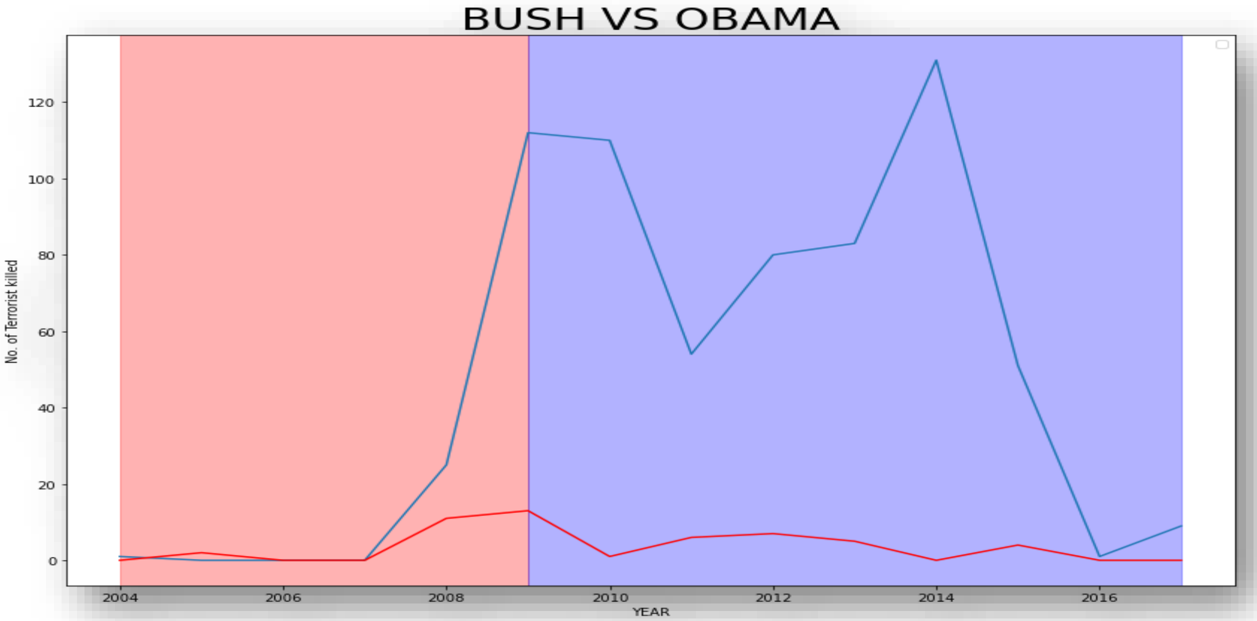
```
In [23]: #Trend of Civilians and Terrorists killed along with people Injured
         fig = plt.figure(figsize=(14, 10))
         x = data_byyear.Accuracy.keys()

         plt.plot(x, data_byyear.Terrorists, color='r', label = 'Terrorists Killed')
         plt.plot(x, data_byyear.Innocents, color='g', label = 'Civilians Killed')
         plt.plot(x, data_byyear.Injured, color='b', label = 'People Injured')
         plt.xticks(range(2004, 2018))
         plt.xlabel("YEAR", fontsize=16)
         plt.ylabel("Number of people", fontsize=18)
         plt.title("Trend of Terrorists and Civilians killed along with Civilians Injured", fontsize=22)
         plt.legend(fontsize = 16)
         plt.show()
```



3.6Al-Qaeda vs Taliban and Bush vs Obama



```
In [24]: #Al-Qaeda vs Taliban and Bush vs Obama
def timeperiodyearTA(data):

    a1=[]
    data['Year'] = np.nan
    data['Weekday'] = np.nan
    data['Month'] = np.nan

    for i in range(0, 483):
        data.loc[:,'Year'][i]= data.loc[:,'DateTime'][i].year

# Grouping the new column by the Year
    by_year = data.groupby('Year').size()
    df7 = pd.DataFrame(by_year)
    timeyear = [2004.0,2005.0,2006.0,2007.0,2008.0,2009.0,2010.0,2011.0,2012.0,2013.0,2014.0,2015.0,2016.0,2017.0]
    df7['Taliban'] = 0
    df7['Al-Qaeda'] = 0
    for i in range(0, 483):
            for x in range(len(timeyear)):
                if (data.loc[:,'Year'][i] == timeyear[x]):

#Calculating total no Taliban personnel killed
                    df7['Taliban'][timeyear[x]] = df7['Taliban'][timeyear[x]] + data['Taliban'][i]

#Calculating total no Taliban personnel killed
                    df7['Al-Qaeda'][timeyear[x]] = df7['Al-Qaeda'][timeyear[x]] + data['Al-Qaeda'][i]
    x = df7.index.values
    y1= df7['Taliban']
    y2= df7['Al-Qaeda']

#plotting the graph for analysis of drone strike under Bush and Obama
    fig1= plt.figure(figsize=(14, 10))
    plt.plot(x,y1)
    plt.plot(x,y2,color="red")
    plt.ylabel("No. of Terrorist killed")
    plt.title("BUSH VS OBAMA", fontsize = 30)
    plt.xlabel("YEAR")
    plt.legend()
    plt.axvspan(2004, 2009, color='red', alpha=0.3)
    plt.axvspan(2009, 2017, color='blue', alpha=0.3)
    plt.show()
    return df7
timeperiodyearTA(df)
```


BUSH VS OBAMA
No. of Terrorist killed
YEAR