**PYTHON PROJECT**

22

**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**

**TABLE OF CONTENT**

* **INTRODUCTION Pg#04**
* **DATA ANALYSIS Pg#05**
* **1.1)THE PROCESS OF DATA ANALYSIS Pg#06**
* **1.2)ROLE OF PYTHON Pg#09**
* **1.3DATA SCIENCE IN FILED OF FINANCE Pg#09**
* **1.4)LIBRARIES Pg#11**
* **CLEANING THE DATA Pg#12**
* **2.1IMPORTING REQUIRED LIBRARIE Pg#13**
* **2.2READING CSV FILE Pg#13**
* **2.3PRINTING HEAD OF THE DATA SET Pg#13**
* **2.4PRINTING TAIL OF THE DATA SET Pg#14**
* **2.5DROPING EXCESIVE COLUMN Pg#14**
* **2.6FINDING OMITTED VALUES Pg#15**
* **2.7FILLING OMITTED VALUES Pg#15**
* **2.8RENOVATING OMITTED VALUES BY MEAN Pg#16**
* **2.9CHANGING DATE AND TIME FORMATE Pg#16**
* **2.10COMBINING TERRORIST AND CIVILIAN DATA Pg#17**
* **2.11Creating new columns for Year, Months, Weeks and Days from the given Date-Time format Pg#18**
* **2.12Cleaning the 'City' column by removing inconsistent values Pg#19**
* **2.13Creating Year-wise dataset Pg#20**
* **EXPLORATORY DATA ANALYSIS Pg#21**
* **3.1Number of Drone Attacks over the Years Pg#21**
* **3.2Number of Drone Attacks by Month Pg#22**
* **3.3Number of Drone Attacks by Day of a week Pg#23**
* **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') Pg#24**
* **3.5Trend of Civilians and Terrorists killed along with people Injured Pg#25**
* **3.6Al-Qaeda vs Taliban and Bush vs Obama Pg#26**

**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.[]](https://en.wikipedia.org/wiki/Data_analysis#cite_note-18) 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.

## C:\Users\ibbad\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\7Y6DZU0F\FAQ_endocrinoologia[1].png

|  |
| --- |
| 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.

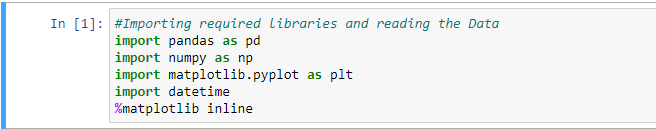
**5) SUBPROCESS:**

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.

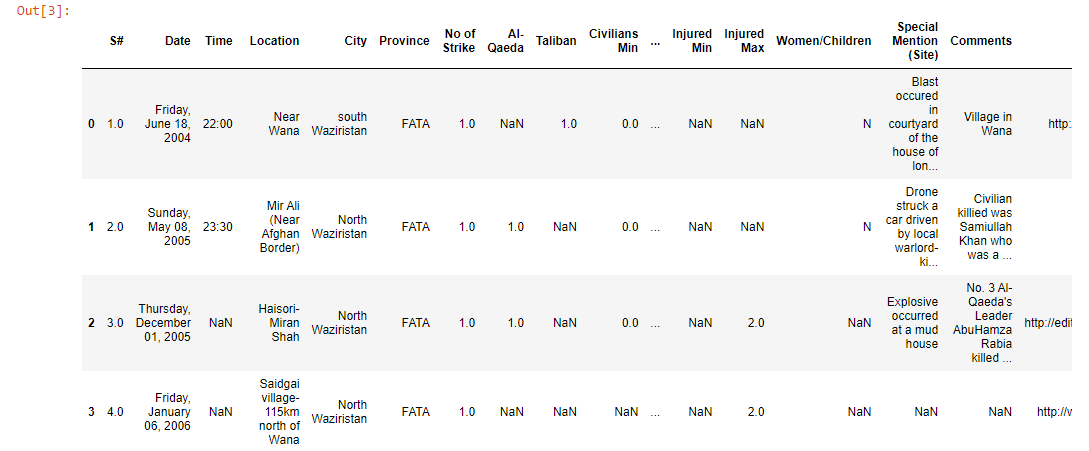
**2.1IMPORTING REQUIRED LIBRARIE**

**2.2READING CSV FILE**

****

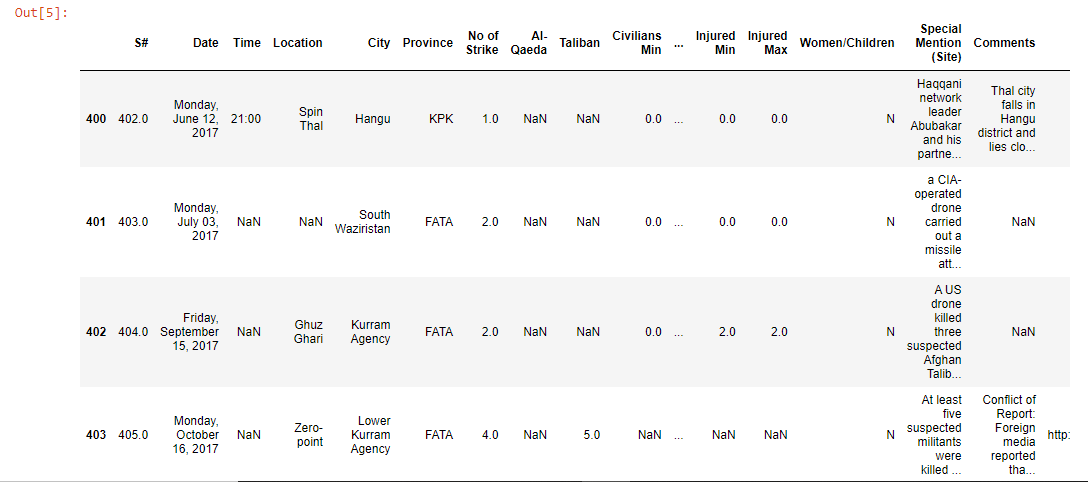
**2.3PRINTING HEAD OF THE DATA SET**

****

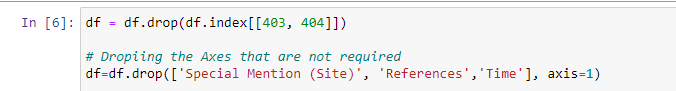
****

**2.4PRINTING TAIL OF THE DATA SET**

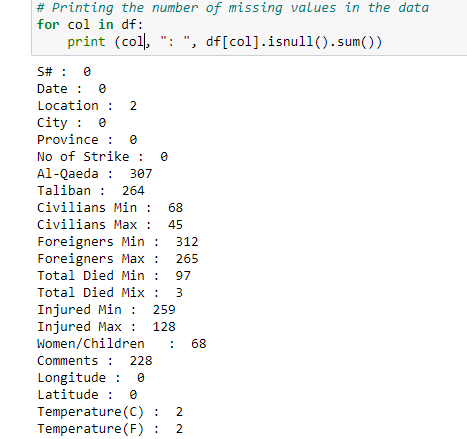
****

****

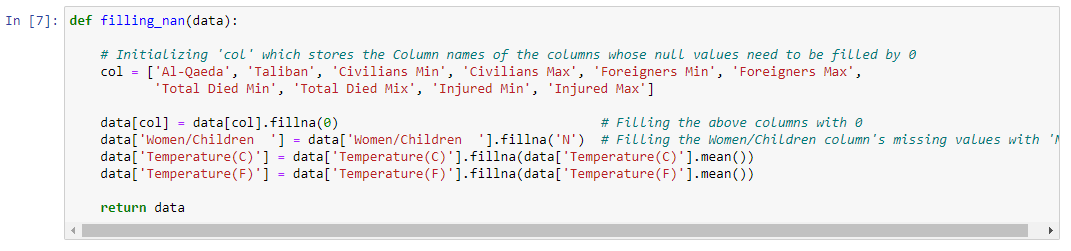
**2.5DROPING EXCESIVE COLUMN**

****

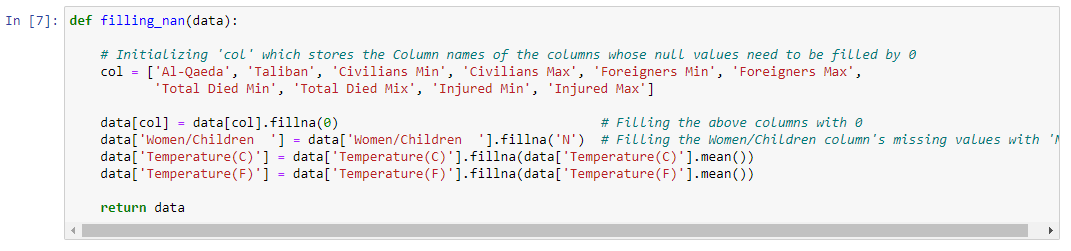
**2.6FINDING OMITTED VALUES**

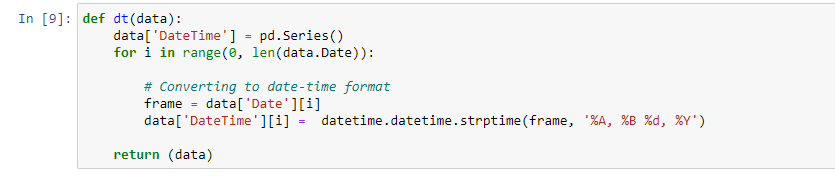
****

**2.7FILLING OMITTED VALUES**

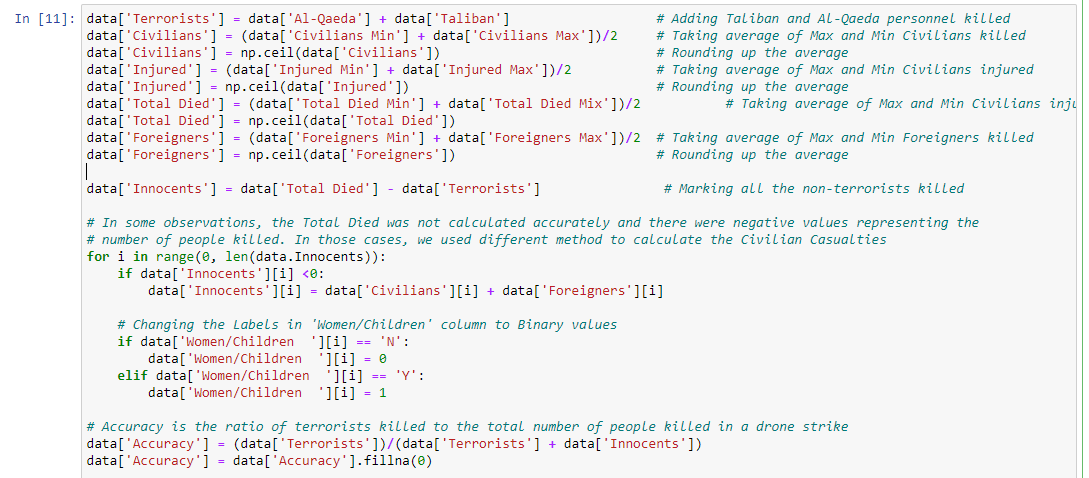
****

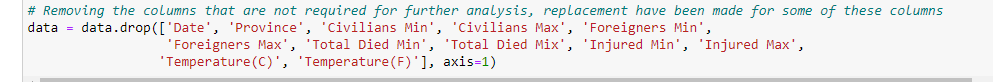
**2.8RENOVATING OMITTED VALUES BY MEAN**

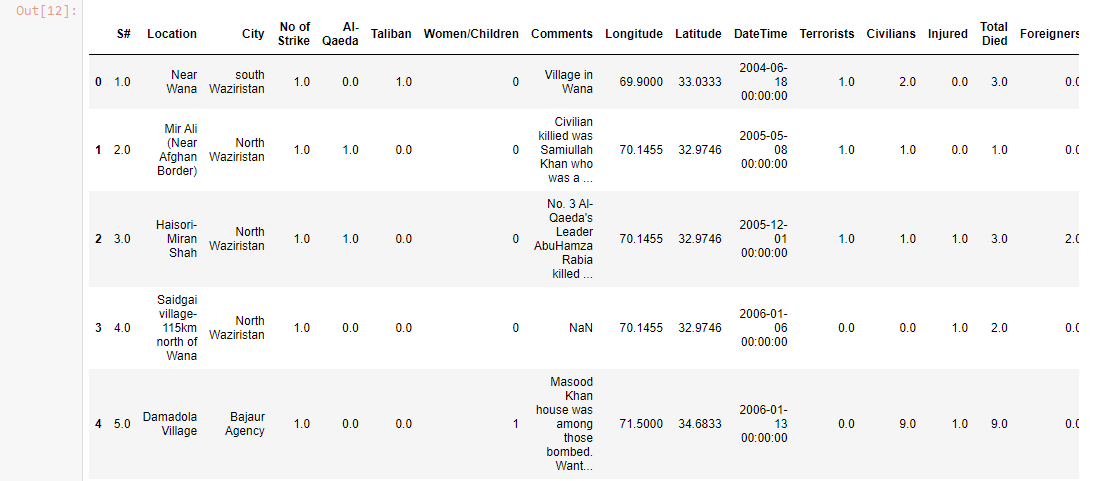
****

**2.9CHANGING DATE AND TIME FORMATE**

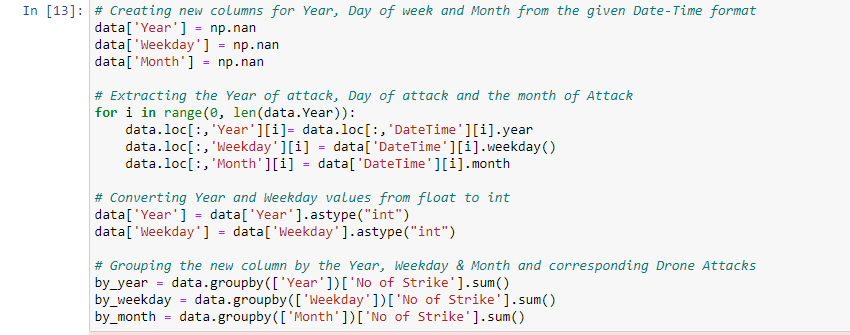
**2.10COMBINING TERRORIST AND CIVILIAN DATA**

****

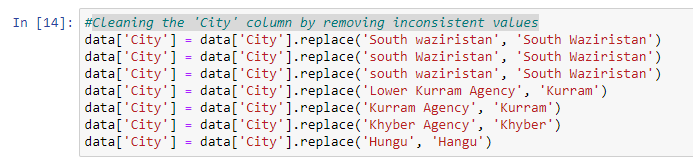
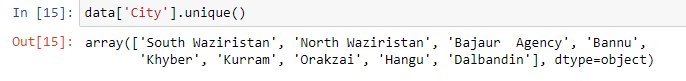
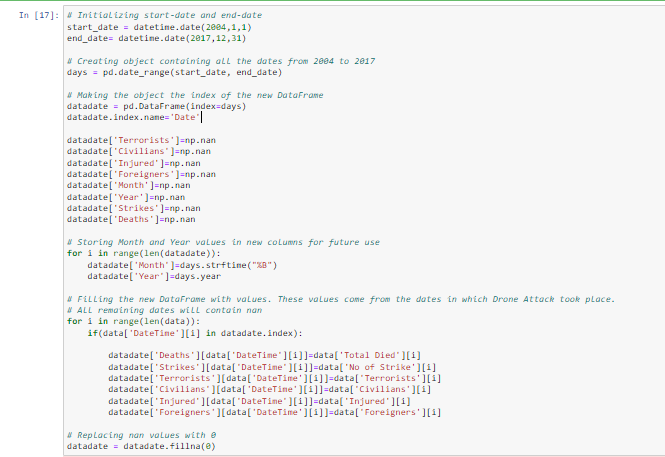
****

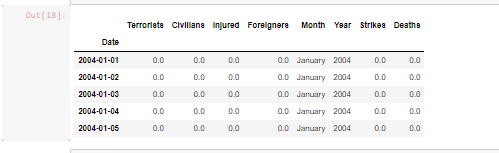
****

**2.11Creating new columns for Year, Months, Weeks and Days from the given Date-Time format**

****

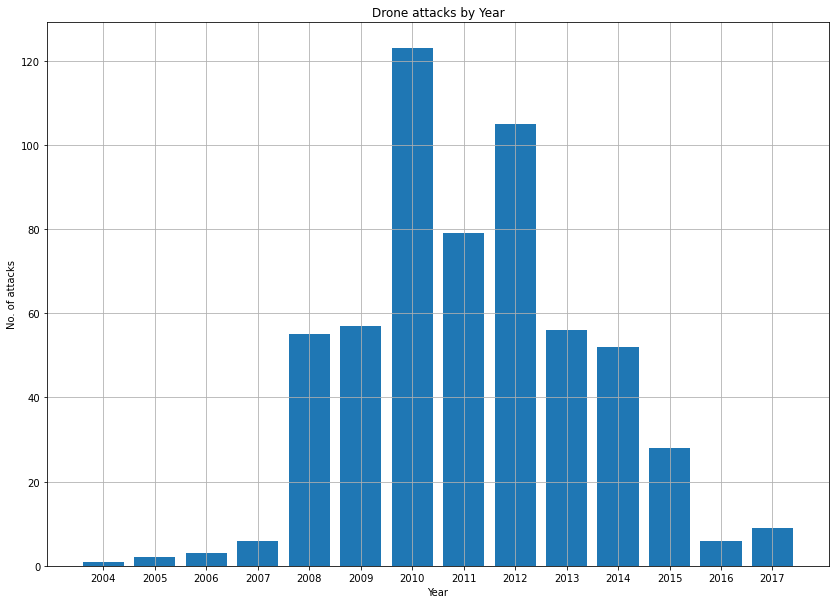
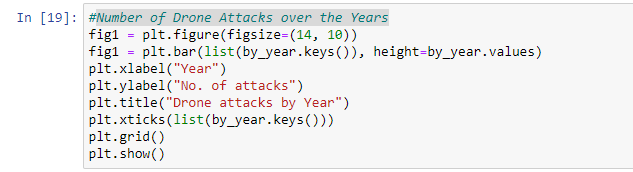
**2.12Cleaning the 'City' column by removing inconsistent values**

** 2.13Creating Year-wise dataset**

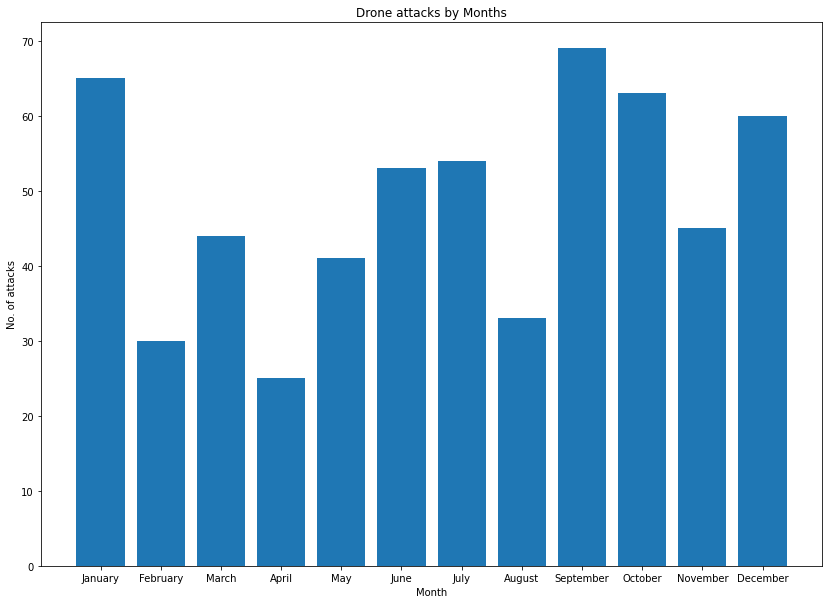
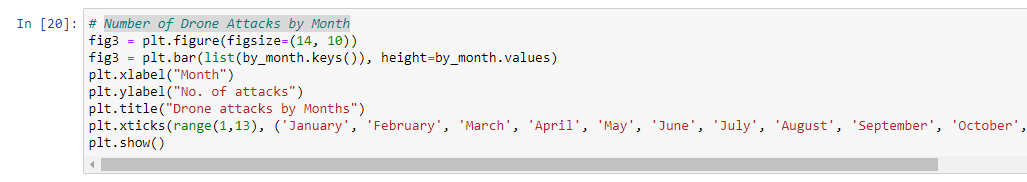
****

**(3)EXPLORATORY DATA ANALYSIS:**

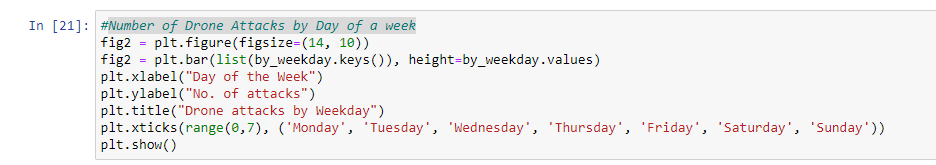
**3.1Number of Drone Attacks over the Years**

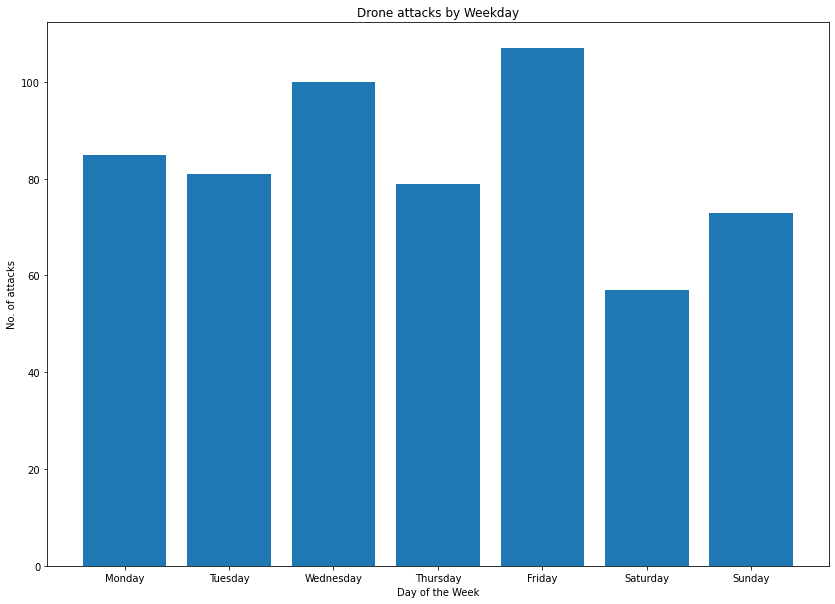
****

**3.2Number of Drone Attacks by Month**

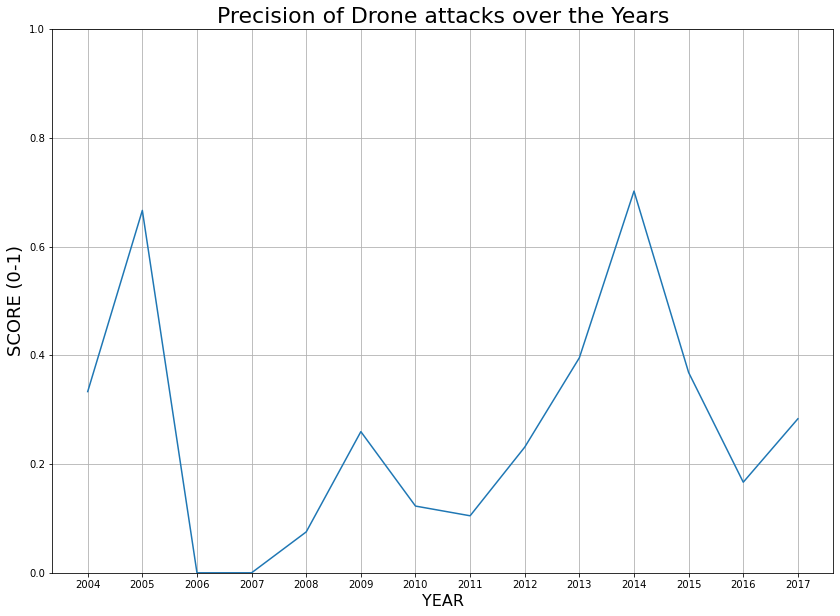
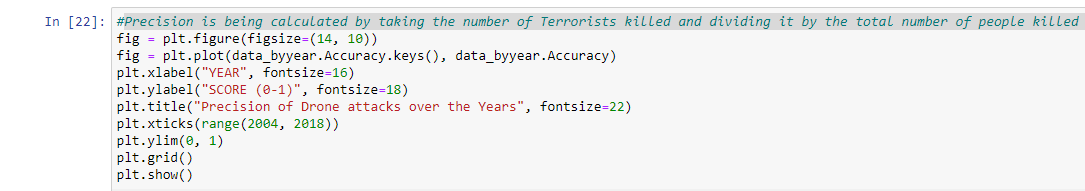
****

**3.3Number of Drone Attacks by Day of a week**

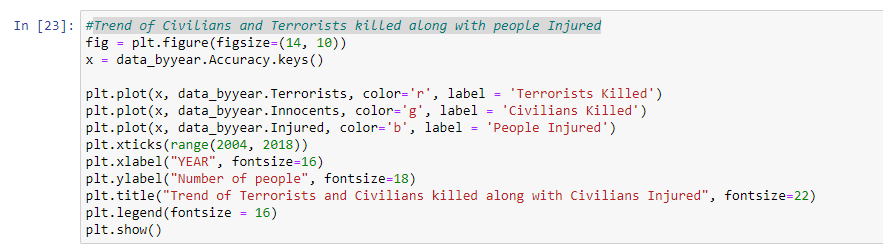
****

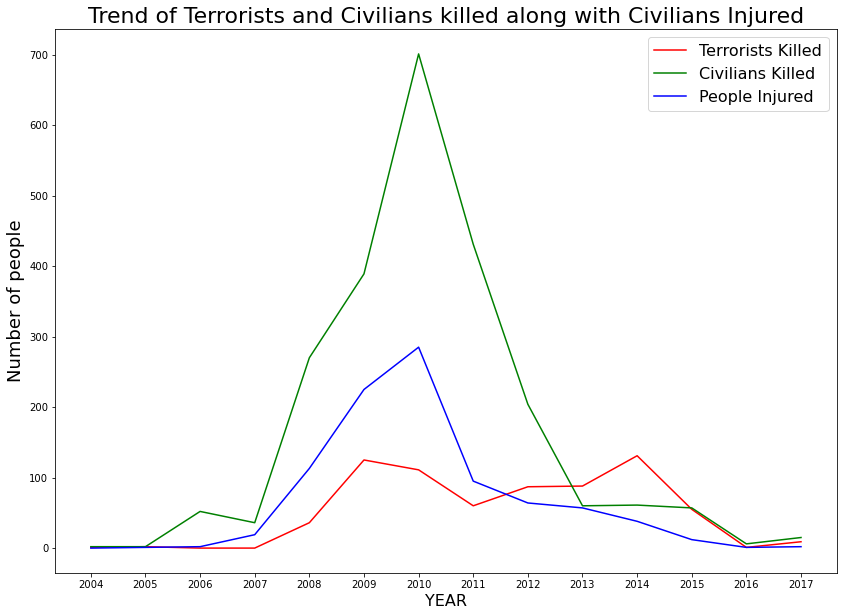
****

**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')**

****

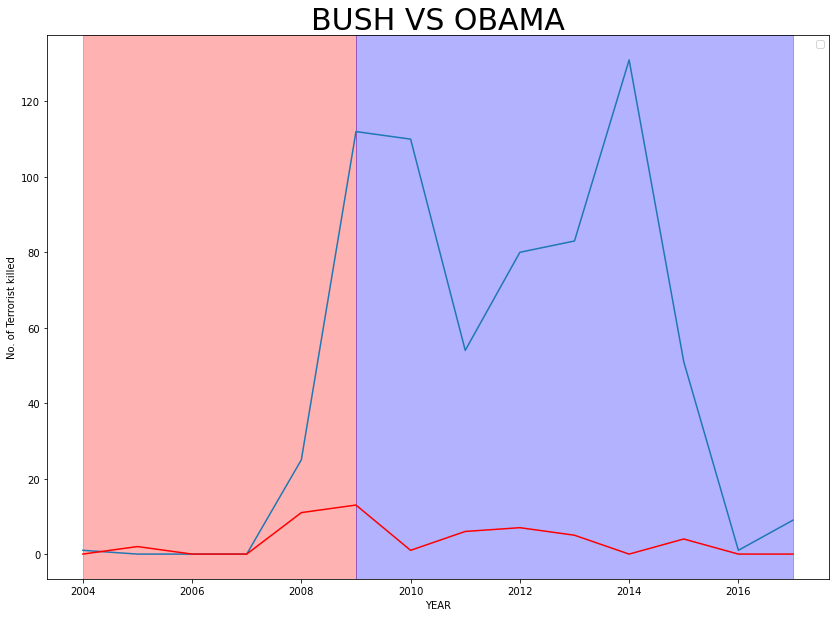
**3.5Trend of Civilians and Terrorists killed along with people Injured**

****

****

**3.6Al-Qaeda vs Taliban and Bush vs Obama**

****

****