Drone Attacks in Pakistan

(https://github.com/harshit0511/Drone-Attacks-in-Pakistan)

PROBLEM & BACKGROUND:

USA has been carrying out aerial attacks using Unmanned Aerial Vehicles (UAV) or drones inside the territory of Pakistan in order to eliminate high-value targets belonging to terrorist organizations like Al-Qaeda, Taliban, Haqqani-network etc. This has been going on since 2004 and US government has been claiming that these are 'precision strikes' where targets are identified using images captured by drones and eliminated by launching a missile from the same drone. It is important to know why the United States has initiated this plan and it's impact on Pakistani society.

One thing is common in these reports and this is that most of the drone attacks occur in Pakistan's northwest region, which is controlled by Pakistan's federal government and known as FATA or Federally Administered Tribal Areas. Drones are unmanned aircrafts which can spy with their high-range cameras and infrared rays. It can carry missiles as well and can launch the missile to the target precisely but it can only be precise at 'hitting' those targets. 'Choosing' the target precisely is something that makes the Drone Strike a success or a failure. These Drone strikes are carried out by CIA and controlled by the US Government.

There are many problems associated with drone attacks.The first one is that the United States is attacking a country that they are not actively engaged in a state of war with and it also a direct attack on Pakistan's sovereignty. Secondly, it is causing collateral damage on a larger scale. To kill one or two terrorists, the United States is possibly killing ten or more civilians including children, women and senior citizens. Malala Yousafzai, who was herself a victim of terrorism, appealed to Barack Obama to stop these Drone Strikes as they were doing more harm than good. With the data of every drone strike that has happened in Pakistan since 2004, we have investigated how accurate these strikes have been at their tasks.

DATA:

The data of every drone strike happened since 2004, was obtained from Pakistan Body Count (http://www.pakistanbodycount.org/). Their data was collected through media reports, hospitals and internet in an effort to show the world the intensity of Drone strikes in Pakistan. There are many sources collecting this data but we are relying on the data collected by an organisation which is working closely with the victims and first-responders of these attacks and validating their data from multiple sources. The data for Suicide Bombings was also available from the

same site. The Global Terrorism Database was obtained from the official website of Global Terrorism Database (GTD): http://start.umd.edu/gtd/. It contained information of all terrorist attacks that happened around the World since 1970. We filtered the data to attacks that took place in Pakistan and after 2004.

DATA ANALYSIS:

A quantitative analysis of the data has helped us answer some questions:

1. Precision of the Drone Strikes over the years
2. Difference in success of Drone strikes in Obama's Administration as compared to Bush's Administration
3. Comparison of Civilian deaths caused by Suicide Bombings and Drone Attacks
4. Correlation between Drone strikes and Terrorist Attacks
5. Number of days after which a Drone strike is followed by a Terrorist Attack

TARGET AND PREDICTOR VARIABLES:

We chose the target variable as the 'Number of Drone Strikes' as we wanted to build a model to forecast the Drone Strike frequency in coming years. The predictor variable in this case would be Time.

For analyzing relation between Drone strikes and Terrorist attacks, we used the 'Number of Drone Strike' and 'Number of terrorist attacks' in a day to calculate the correlation between the two.

PROBLEM STATEMENT:

The motivation of this project is to highlight the situation in Pakistan and if the use of drones as a military strategy by the United States has failed to help win the war on terror. The main aim of the project was to discover the impact of these drone strikes and how successful or unsuccessful have they been over the years at carrying out their tasks. The accomplishment of this aim was dependent on completion of certain objectives. These objectives were:

1. Compare the success of Bush administration to Obama administration in eliminating terrorists
2. Analysis of Terrorists and Civilians killed in the Drone strikes
3. Effect of Drone Strikes on Terrorist activities and the correlation between the two
4. Time-Series analysis and forecasting the Number of Drone strikes in coming years

TYPE OF MODEL:

We have used the time series model for Forecasting the Number of drone strikes. The reason for choosing this model is because time series enables for better understanding of time-dependent variables along with the ability to forecast future values. Time series analysis involves developing models that best captures or describes an observed time series in order to understand the underlying causes. Making forecasts is called extrapolation in the classical statistical handling of time series data.

Stationarity of Time Series:
A Time Series is said to be stationary if its statistical properties such as mean, variance remain constant over time.

Forecasting depends on the type of series used, in our case we deal with a series with significant dependence among values. Hence, we used statistical model like ARIMA to forecast the data.
The predictors depend on the parameters (p,d,q) of the ARIMA model

1 Number of AR (Auto-Regressive) terms (p)
2 Number of MA (Moving Average) terms (q)
3 Number of Differences (d): These are the number of nonseasonal differences (In our project we took the second order difference)

In the Jupyter Notebook 'Pakistan Drone Attacks.ipynb', you can see that we used AR (p) = 3 and MA (q) = 1 in the first model and AR (p) = 1, MA (q) = 0 and Differences (d) = 2 in the second one.

EVALUATION METRICS:

The forecasting was done for the years 2012 - 2017.

1. Augmented Dickey Fuller Test: Check stationarity of Time-Series
2. Durbin Watson Test: Check for serial correlation in the Series
3. Spearman's Rank Order Correlation: Check for correlation between two Time-Series
4. Mean Absolute Error (MAE) and Mean Forecast Error (MFE): Calculate the precision of the forecasts

Model 1:
Mean Absolute Error : 8.86424
Mean Forecast Error : 2.02236

Model 2:
Mean Absolute Error : 0.4928

Mean Forecast Error : 0.0097

ASSUMPTIONS AND LIMITATIONS:

The assumptions made for this project:

1. The data source is authentic and is true for all the attacks, be it Drone Attacks, Suicide Bombings or data of Terrorist Attacks.
2. Some of the Civilians marked in the dataset could have been family members of the Terrorists. Children or brothers of terrorists may or may not turn out to be terrorists in the future. Since we cannot know who was destined to become a terrorist and who was not, we consider all family members as 'Civilians'.

Limitation of the project merely outlines the potential to which the project could have been scaled. Some of these potentials are:

1. Sentiment Analysis: The column consisting of news clippings for each Drone Strike can be analyzed to categorize the attack as successful or unsuccessful.
2. Forecast of other variables: We implemented a Time-series model to forecast the No. of Drone Strikes but one can also forecast the Precision of strikes, number of terrorists killed, involvement of Women/Children killed etc.
3. More data: The scope of this project is only limited to the data we have, however, if we can get access to other datasets, we can possibly understand the planning that goes behind every drone strike.

PROBLEM IN SCOPE OF CLASS:

Data Pre-processing:

During the course of this project, we cleaned the data, created many new features and Dataframes which were then used in further analysis and Time-Series forecasting. There were some Dataframes which were used in multiple Jupyter Notebooks for analysis.

Data Visualization:

We visualized many values and tried to understand the underlying pattern in each of them. The Visualization helped us to get more information from the data and guided us in making decisions that were implemented later on in the project.

Time-Series Forecasting:

We used ARMA model to forecast the Number of Drone strikes that may happen in the coming years. The methods used to obtain stationarity in Time-Series, make meaningful predictions and

evaluate the predictions using appropriate metrics were similar to the ones learned from the lectures.

Significance Testing:

We tested correlation between two Time-Series by a lag of 1 day. The test was done based on the Null Hypothesis that the two Time-Series are not connected. The P-Value obtained was 0.00064 which is less than 0.05 suggesting a greater than 95% confidence in Rejecting the Null Hypothesis. Therefore, the correlation between the two Time-Series was significant enough that it cannot be ignored.

CHANGES FROM ORIGINAL PROPOSAL:

In the proposal, we planned to implement a Decision Tree for understanding what features are important and have an affect on the decisions. Unfortunately, the data was not appropriate for the task and we could not get desired results. Therefore, we chose Time-Series modelling to forecast the number of attacks in coming years.

Previously, we had thought of correlating the Drone Strikes data to the Suicide Bombings database but later on we came to a conclusion that Global Terrorism Database would be the most appropriate database in this case. Although a lot of Terrorist attacks that happened in Pakistan were Suicide Bombings but this was not representing all kinds of attacks and therefore we had to choose new database to correlate it with.