**Yemen Aerial Bombardments Analysis**

**Rio Atmadja**

**07 August 2020**

**Introduction**

The goal of this project to analyze and study the civilian casualties, injuries, fatalities, and overall damaged caused by the airstrikes conducted by the Saudi and its collations. Also, to predict civilian casualties and to classify the result of airstrike casualties as civilians or combatants. I obtained my dataset from the Yemen Data Project (<https://www.yemendataproject.org/data.html>) . It is a non-for-profit organization which produces data for the purpose of increasing transparency and promoting accountability. This data is freely available to use under a Creative Commons Attribution 4.0 International License.

**Yemen Data Project**

Yemen Data Project start collecting data on aerial bombardment in Yemen from March 26, 2015 to July 31, 2019. This dataset contains all air raids conducted by the military coalition led by Saudi Arabia and the United Arab Emirates, which also includes Bahrain, Egypt, Jordan, Sudan, Kuwait, Qatar until 2017, and Morocco until 2019. The coalition forces were backed by the United States and the United Kingdom amongst other western nations. The United States was responsible to provide intelligence and logistic support and the UK support the coalition in every practical way short of engaging in combat. Both countries have military personnel deployed at Saudi control center for coalition airstrikes.

**History of the conflict in Yemen**

The Yemeni crisis began with the 2011 to 2012 against President Ali Abdullah Saleh, who had led Yemen for more than 33 years. After, former President Ali Abdullah Saleh left office in early 2012 as part of agreement between the Yemeni government and opposition groups. Under the leadership of Abdrabbuh Mansur Hadi the former vice president, the country failed to unite the fractious political landscape of the country and fend off threats both from Al-Qaeda in the Arabian Peninsula and the Houthi militants that had been waging war in the north for years.

In 2014, Houthi fighters took over the capital Sana’a and forced Hadi to negotiate a unity government with other political factions. In January of 2015, Hadi resigned along with his ministers after his presidential palace and private residence came under attack from the militant group. Hadi fled to Aden, where he declared the country’s temporary capital and called on loyal government officials and members of the military to rally to him. Yemen was in total chaos in 2015, after Houthis established a new government in Sana’a and Hadi retreating with his supporters to Aden, and later to Saudi Arabia. The Arab League (Bahrain, Egypt, Jordan, Sudan, Kuwait, Qatar until 2017, and Morocco until 2019), led by the Saudis, began a bombing campaign and mobilization of various armed forces in the region for a possible invasion.

The 8 October 2016 attacked by the Saudi Arabian led coalition killed at least 140 people and injured more than 600 in Sana’a. This was one of the single worst death tolls in the two-year war.

**History of the Houthis**

According to Wikipedia, the Houthi movement officially called Ansar Allah or supporters of God and also the founder of the movement is from the Houthi tribe. Houthi is an Islamic political and armed movement that emerged from Sa’dah in northern Yemen in the 1990s. Under the leadership of Hussein Badreddin al-Houthi, the group become a big opposition to former Yemeni president Ali Abdullah Saleh. Hussein was killed in Sa’dah in 2004 along with a number of his guards by the Yemeni army. This event sparks the Houthi insurgency in Yemen led by his brother Abdul-Malik al-Houthi. In 2011, the Houthis took part in Yemeni Revolution by participating in the street protests and coordinating with other opposition groups. They joined the National Dialogue Conference in Yemen as part of the Gulf Cooperation Council. However, in November 2011 the Houthis reject the GCC deal’s because the deal did not fundamentally reform governance and the federalization only forced Yemen into poor and wealthy regions.

In late 2014, Houthis repaired their relationship with the former president Ali Abdullah Saleh, and with his help in 2014-2015, the Houthis was able to take over of the capital and much of the north and took over the government in Sanaa. By 2015, the Houthis have gained control of most of the northern part of Yemen’s territory and overthrow the current government of Abdrabbuh Mansur Hadi. The Houthis have launched repeated missile and drone attacks against Saudi cities or also known as proxy war between the Saudi and Iran.

**EDA Outcome**

From the Yemen Data Project, I have extracted important points:

1. **Yemen Military Infrastructures** 
   1. There are 3262 Forces in Yemen
   2. There are 1707 Military Site in Yemen
   3. There are 536 Moving Targets in Yemen
   4. There are 239 Weapon Storage in Yemen
   5. There are 223 Political Figures in Yemen
   6. There are 95 Political HQ in Yemen
   7. There are 83 Check Points in Yemen
2. **Airstrikes Frequency** 
   1. There were total of 14408 airstrikes conducted by the Saudi-Coalition forces.
   2. Summary airstrikes Frequency:
      1. There were 3481 totals of airstrikes conducted in the morning
      2. There were 5240 totals of airstrikes conducted in the afternoon
      3. There were 2992 totals of airstrikes conducted in the evening
      4. There were 2695 totals of airstrikes conducted in the night
3. **Location of Air defense**
   1. A screenshot of a cell phone

      Description automatically generated
   2. A screenshot of a cell phone

      Description automatically generated
4. **Number of Air defense guarding Yemen Infrastructures** 
   1. There are 67 Air defense Military Camp protecting Yemen airspace.
   2. There are 20 Air defense Military Site protecting Yemen airspace.
   3. There are 5 Air defense Military Site (Missiles Battalion) protecting Yemen airspace.
   4. There are 4 Air defense Battalion protecting Yemen airspace.
   5. There are 1 Air defense Military Site (assigned to protect Oil facility) protecting Yemen airspace.
   6. There are 1 Third Military region (Air defense battalion) protecting Yemen airspace.
   7. There are 1 170th Air defense Brigade protecting Yemen airspace.
   8. There are 1 Air defense Military Site located near the Safer Oil Facility protecting Yemen airspace.
   9. There are 1 49 Air defense Battalion protecting Yemen airspace.
   10. There are 1 Air defense Military Site protecting Yemen airspace.
   11. There are 1 180th Air defense Military Camp Radars protecting Yemen airspace.
   12. There are 1 Air defense Military Battalion protecting Yemen airspace.
   13. There are 1 Air defense site protecting Yemen airspace.
   14. There are 1 Air defense Batteries protecting Yemen airspace.
   15. There are 1 Air defense hill protecting Yemen airspace.
5. **Numbers of fatalities**
   1. Positive value skew of 17.74 indicates a pile up of lower scores of fatalities on the left of the distributions. Also, a positive value of kurtosis 497.40 indicates pointy and heavy tailed distributions (As you can see in Figure 1.5).
   2. The normal Q-Q plots for fatalities is deviating away from the diagonal, which suggests that fatalities numbers are much more clustered around the low end of the scale (As you can see in Figure 1.5), the fatalities distribution is positively skewed. Also, it can indicate that as the air raid progressed, the number of fatalities decreases. The skew occurs because the airstrikes were conducted in different parts of Yemen.
   3. In the distribution of fatalities, the highest numbers of fatalities are between 0 to 5 people, which matches the analysis from previous slides. Note: I did not include zero because it represents zero fatalities.
6. **Numbers of Injuries** 
   1. Positive value skew of 75.24 indicates a pile up of lower scores of fatalities on the left of the distributions. Also, a positive value of kurtosis 6677.76 indicates a pointy and heavy tailed distribution (As you can see in Figure 1.5.1).
   2. The normal Q-Q plots for injuries is deviating away from the diagonal, which suggests that injuries numbers are much more clustered around the low end of the scale (As you can see in Figure 1.5.1), the injuries distribution is positively skewed. Also, it can indicate that as the air raid progressed, the number of injuries decreases. The skew occurs because the airstrikes were conducted in different parts of Yemen.
   3. In the distribution of injuries, the highest numbers of injuries are between 0 to 5 people, which matches the analysis from previous slides. Note: I did not include zero because it represents zero injuries.
7. **Numbers of civilian casualties** 
   1. Positive value skew of 59.00 indicates a pile up of lower scores of civilian casualties on the left of the distributions. Also, a positive value of kurtosis 4694.13 indicates a pointy and heavy tailed distribution (As you can see in Figure 1.7.1).
   2. The normal Q-Q plots for fatalities is deviating away from the diagonal, which suggests that civilian casualties’ numbers are much more clustered around the low end of the scale (As you can see in Figure 1.7.2), the civilian casualties’ distribution is positively skewed. Also, it can indicates that as the air raid progressed, the number of civilian casualties decreases The skew occurs because the airstrikes were conducted in different parts of Yemen.
   3. In the distribution of fatalities, the highest numbers of civilian casualties are between 0 to 5 people, which matches the analysis from previous slides. Note: I did not include zero because it represents zero casualties.
   4. **Casualties from 2015 to 2019**
      1. There are 9089 numbers of civilian casualties 2015
      2. There are 5331 numbers of civilian casualties 2016
      3. There are 3366 numbers of civilian casualties 2017
      4. There are 2474 numbers of civilian casualties 2018
      5. There are 493 numbers of civilian casualties 2019
      6. There are 20753 Total Civilian Casualties
8. **Total Airstrikes by governorates**
   1. Abyan has a number of 502 airstrikes
   2. Aden has a number of 1768 of airstrikes
   3. Amran has a number of 3011 airstrikes
   4. Bayda has a number of 1593 of airstrikes
   5. Capital has a number of 15083 airstrikes
   6. Dhalie has a number of 943 airstrikes
   7. Dhamar has a number of 1520 of airstrikes
   8. Hadramawt has a number of 142 airstrikes
   9. Hajja has a number of 13686 airstrikes
   10. Hudaydah has a number of 17976 airstrikes
   11. Ibb has a number of 1966 of airstrikes
   12. Jawf has a number of 5746 airstrikes
   13. Lahj has a number of 3104 airstrikes
   14. Maharah has a number of 5 airstrikes
   15. Mahwit has a number of 604 airstrikes
   16. Marib has a number of 9236 airstrikes
   17. Raymah has a number of 134 airstrikes
   18. Saada has a number of 38744 airstrikes
       1. To identify the peak of civilian casualties in Saada, I summed up all the values up until a specific date, as a result we can see that there is a big jump in Civilian Casualties between 2018-08-08 to 2018-08-09 (Note: 2018-08-09 is the actual date of the incident). It turned out that on 2018-08-09, there were numbers of 130 civilian casualties, 51 fatalities, and 79 injuries.
   19. Sanaa has a number of 20232 airstrikes
       1. To identify the peak of civilian casualties in Sanna, I summed up all the values up until a specific date, as a result we can see that there is a big jump in Civilian Casualties between 2016-10-08 to 2016-10-09. It turned out that on 8 October 2016 attacked by the Saudi Arabian led coalition killed at least 140 people and injured more than 600 in Sana’a. This was one of the single worst death tolls in the two-year war.
   20. Shabwa has a number of 1752 of airstrikes
   21. Taiz has a number of 21038 airstrikes
   22. lahj has a number of 1 airstrike
9. **Damaged on Yemen Infrastructures** 
   1. There are 6684 in unknown infrastructures
   2. There are 6500 in military security target infrastructures
   3. There are 2572 in civilian infrastructures
   4. There are 1373 in infrastructure infrastructures
   5. There are 1060 in economic infrastructures
   6. There are 288 in political tribal infrastructures
   7. There are 240 in educational facility infrastructures
   8. There are 231 in economic infrastructures
   9. There are 200 in Pro-Houthi forces
   10. There are 510 in other infrastructures
10. **Total numbers of casualties, fatalities, and injuries by the jurisdictions** 
    1. Peak of casualties, fatalities, and injuries in Yemen occurred in October 08, 20168

**Analysis Outcome**

* **Interesting Facts**
  + Based on my timeseries analysis, I found that on October,08 2016 attacked by the Saudi Arabian led coalition killed at least 140 people and injured more than 600 in Sana’a. This was one of the single worst death tolls in the two-year war. The actual numbers, I got from my data set is 832 civilian casualties.
  + There were 13187 civilian casualties and only 6804 combatant casualties, as a result of aerial bombardments.
  + From casualties’ classifications, I learned that the positive coefficients increase the log-odds of the response and the probability. On the other hand, the negative coefficient decreases the log-odds of the response and decreases the probability.
* **Highly correlated variables** 
  + **Civilian Casualties and Injuries**
    - There is a strong Pearson Correlation of 𝜌 = 0.959477, between number of civilian injuries and civilian casualties, Also, 𝑅^2 = ρ = 0.920597 indicates that Injured variable accounts for 92.06% of variance in Civilian Casualties. Using Injured (Number of civilian injuries) to predict Civilian Casualties (Number of Civilian Casualties) reduces RMSE from 9.523292 to 0.029589. Moreover, a correlation of 0.959477 yields a reduction in RMSE by 10.19%.
    - The actual value of the correlation is 0.9595 which is highly correlated and the computed p-value for all simulation p-value < 0.05, therefore it is statistically significant. In other words, it is unlikely to have occurred by chance.
  + **Civilian Casualties and Fatalities** 
    - Correlation Test: The actual correlation is 0.714394 and the computed p-value < 0.05, therefore it is statistically significant. In other words, it is unlikely to have occurred by chance.
    - Difference in Means Test: The test yields that p-value < 0.05, therefore it is statistically significant. In other words, it is unlikely to have occurred by chance.
  + **Linear Regression Analysis: Predicting the number of civilian casualties**
* A screenshot of a cell phone

  Description automatically generated
  + **Classification Analysis: Predicting casualties as combatants or civilians**
* A screenshot of a cell phone

  Description automatically generated
* **Points missed during analysis**
  + I felt that, I missed some points in my analysis from due to labeling errors, for example because I might label the civilian infrastructures as military infrastructures (e.g. The house of Al-Abdul). As a result, my classification predictions are not as accurate.
  + I felt Sub-category variable can help me significantly in classifying the casualties as combatants or civilians because it has 3452 unique variables. I can create a function to properly filter the civilians or combatants, based on the governorate, districts, main category, and subcategory attributes.
* **Variables that can help in analysis** 
  + I felt like by placing the correct label on Main Categories and Target variables, I can get more accurate classification predictions. Because originally, the main categories variable has 294 unique variables and target variable has 3452 unique variables.
  + Applying the correct filters for Woman injured and Child injured variables might increase the classification accuracy, the reason is that the women and the children might have a combat role. For example, If I were to group by the Main Categories and Target variables together, the women and children might show up in military bases or at Pro-Houthis sites.
* **Wrong Assumptions** 
  + Make an assumption that women and children as civilians because in my dataset, they show up in military camps, sites, and Pro-Houthi sites.

**Challenges**

* The biggest challenge and the most time consuming are to label the Main category, Sub-Category, and Target columns. Because the columns contain some Arabic to English translations that I did not understand. Also, by applying the correct labels (e.g. Civilians or Combatants), it will help me to get more accurate predictions.

**References**

**[1]** Yemen Airstrikes. (2015, March). Retrieved June 18, 2020, from <https://www.yemendataproject.org/onewebmedia/Yemen%20Data%20Project%20Published%20Air%20raids%20Database%20from%2026%20March%202015%20to%2031%20July%202019.csv>.  
**[2]** Nevola, L. (n.d.). Data. Retrieved August 06, 2020, from <https://www.yemendataproject.org/data.html>**[3]** Saada Governorate. (2020, April 30). Retrieved August 06, 2020, from <https://en.wikipedia.org/wiki/Saada_Governorate>**[4]** Houthi movement. (2020, August 05). Retrieved August 06, 2020, from <https://en.wikipedia.org/wiki/Houthi_movement>**[5]** SAN10005 - The Al Kubra Funeral Hall Strike. (n.d.). Retrieved August 06, 2020, from <https://yemen.bellingcat.com/investigations/san10005-the-al-kubra-funeral-hall-strike>**[6]** Yemeni Crisis (2011–present). (2020, August 05). Retrieved August 07, 2020, from <https://en.wikipedia.org/wiki/Yemeni_Crisis_%282011%E2%80%93present%29>