# Analysis of The Casualties of The Vietnam War

##Executive Summary

This project is an analysis of the casualties and posthumous awards of the Vietnam war conflict. The question is to find any correlation between branch, rank, MOS, or location and the patterns of casualties.

## Tableau Dashboard

<https://public.tableau.com/app/profile/foster.wolff/viz/Capstone_17449048095410/Introduction>

## Table of Contents

\* [Executive Summary](#executive-summary)

\* [Tableau Dashboard](#Tableau-dashboard)

\* [Motivation](#motivation)

\* [Questions](#questions)

\* [Normalizing the Data](#normaling-the-data)

\* [Problems and Hurdles](#problems-and-hurdles)

\* [Technologies Used](#technologies-used)

\* [Sources](#sources)

## Motivation:

I chose this topic of analysis because I enjoy reading and researching history. Many of my family members were involved in the conflict, including my late grandpa Roy who was an air reconnaissance infantryman. The period of the Vietnam War had many large cultural shifts. Veterans of this conflict were treated harshly even though a large percentage had no choice in serving. In investigating this data, I hope to remember those who served and the sacrifices they made.

## Questions

1. Which military occupation experienced the most casualties?
2. What were the causes of recorded casualties?
3. Which time frames or specific dates had the largest number of losses?
4. What were some key demographics that correlated to losses?

## Normalizing the Data

To correctly analyze the data, it had to be scraped from the source page by page. The data source had no way to bulk download the tables. To accomplish this, I had to employ the use of C++, windows subprocesses, regular expressions, headless chrome browsers, and file organization. This first allowed me to repetitively scrape the host site and extract the bulk data in a (somewhat) efficient manner. After the bulk data was extracted and put into its respective folder, I then used python to parse and correctly structure the data into pandas dataframes and move those onto CSV files. After moving the data onto CSV files, I could then load them into a Postgres database.

## Problems and Hurdles

The main problem with this project was the need to create logic to scrape the site where the data was located. The process for correctly querying the site then updating the program to the next record was extensive. Another problem was the lack of quantitative data, although qualitative data seemed to make up for the disparity.

## Technologies Used

1. Python
2. SQL
3. Tableau
4. C++

## Data Sources

1. Casualty Data - <https://aad.archives.gov/aad/fielded-search.jsp?dt=2513&cat=WR28&tf=F&bc=,sl>
2. Award Data - <https://aad.archives.gov/aad/fielded-search.jsp?dt=1457&cat=WR28&tf=F&bc=,sl>
3. Confirmation Questions - <https://www.vvmf.org/Wall-of-Faces/>