A Quantitative Ranking of Historical Military Generals via Machine Learning

Bryant Reese
Advisor: Arman Cohan

Submitted to the Department of Applied Mathematics
in partial fulfillment of the requirements for the degree of Bachelor of Arts

DEPARTMENT OF APPLIED MATHEMATICS
YALE UNIVERSITY
May 8th 2024

Abstract

The "greatness" of military generals has fascinated historians and the general public alike since the beginning of organized civilization. Some, like Alexander and Cyrus have been immortalized with the suffix "the Great". However, it is tough to quantify and compare these generals, especially across time periods. My senior thesis project gives one answer to the question: how should we rank the historical military generals who have fought between 2500 BCE and the 21st century? This thesis project creates a linear regression machine learning model to quantify the average performance of a military general. It then ranks 6,619 generals from throughout history based on over 3,500 battles compared to the performance of an 'average' general in that battle to answer the question: who are the greatest tactical generals in military history? Secondly, this project creates a dataset of textual descriptions of the same battles, which will be used in future work to incorporate textual sentiment analysis into the rankings to provide a more comprehensive ranking criteria.

1 Introduction

Project Introduction As a major in both Applied Mathematics and History, I sought a way to blend my two passions within my applied math senior thesis. In looking for mathematical applications of historical content, battles jumped to the forefront. Military engagements provide numerical data necessary for quantitative analysis in the form of troop and other numbers. This thesis aims to accomplish two main goals. First, to quantitatively rank military generals between 2500 BCE and today plus point out any interesting conclusions to be drawn from the model results. Second, to create a dataset for further textual analysis to incorporate more strategy analysis into the rankings beyond numerical tactics in future work.

2 Prior Works

Few prior works I could find have been completed trying to quantify military performance across a large number of generals, ancient and modern. This is likely because Mathematics and History are fields typically given little association. Most of the work on questions related to generals and battles has been conducted on an individual basis. I consulted two such projects when creating my own.

Ethan Arsht completed a similar project to the first element of my project available here: Ethan's Project Hyperlink. This helped me significantly and provided my first data set as Ethan scraped and cleaned the information from Wikipedia's "List of Battles" infobox which included troop numbers, results and generals necessary to complete a WAR (to be described later) ranking of the commanders. While my model is computationally separate with different weights and unique analysis plus results (Napoleon is awarded around 30 WAR in mine vs around 12 WAR in Ethan's), Ethan and mine are similar given that they are derived from the same Wikipedia data. For example, Ethan's top two total WAR generals are Napoleon and Julius Caesar, who rank 1st and 3rd in my model respectively. I thank Ethan for his previous work in scraping and cleaning that data as it saved me many hours of manual parsing through expansive CSV and XLSX files.

I also consulted the World Historical Battles Database developed by Professor **Shuhei Kitamura** at Osaka University. While I did not end up using the database because I discovered Ethan's which had specific tags for individual generals that did not require large scale manual entry/cleaning, I thank professor Kitamura for permitting me access to his database during the early planning stages of my project.

3 Methodology of Numerical Ranking

The Model Sports analytics have become a powerful modern tool for comparison and valuation of players. For example, in Major League Baseball WAR(Wins Above Replacement) is a commonly used statistic to quantify how valuable a major league player is. Per MLB.com "WAR measures a player's value in all facets of the game by deciphering how many more wins he's worth than a

replacement-level player at his same position (e.g., a Minor League replacement or a readily available fill-in free agent)." My project applies a similar methodology to military generals. Because it is hard to define a "replacement level general", I will instead compare each general to how they outperform (or under perform) the average general. We will subsequently call this average general "regular" in order to preserve the fitting acronym "WAR" for our generals. Regular performance will be calculated by sampling the entire dataset of over 3,500 battles. I then created a linear regression model based on statistics such as infantry, artillery, cavalry and ship numbers available to each general in each battle. Then, individual performance will be compared to how an average general would have performed on an individual battle basis using the model's weights. Two statistics will then be offered: most crucially Wins above Regular(WAR) which sums the total value a general added over all of their battles. Secondarily, average Wins above Regular (aWAR), which divides a generals WAR by their number of battles to get their average per battle. In the end, my model is able to quantitatively rank 6619 unique historical generals.

4 Reasons for and Creating the Textual Dataset

While sentiment analysis, classification and other forms of textual analysis are beyond the scope of my current project, I sought to also create a textual dataset to incorporate into the rankings in the future. This desire came from my rankings being tactically biased, mostly based on different categories of troop numbers and whether a general won, lost or reached a stalemate. Other more "soft power" factors such as strategy, oratory skills, and leadership certainly apply to what makes a general "great". Much of this can be extracted for textual descriptions of the battles which provide more insight into these types of factors. To create this dataset, I scraped descriptions of each battle again from Wikipedia's "List of Battles" in order to ensure consistency of generals during my future work. The dataset includes these descriptions each tagged to a specific battle. The created scraper as well as final dataset are both available in the project files (link provided later in project).

5 Hypotheses

Below are a few hypotheses heading into the project analysis.

Hypothesis 1: More Battles fought will Correlate with Higher WAR : While each battle provides a chance to both gain and lose value, my hypothesis is that generals who win one battle are much more likely to fight another because they commanded effectively and are then likely to stay in command of the army. Hence, the more battles a general has fought, the more likely it is that they have won each of their previous battles, subsequently accumulating more total WAR.

Hypothesis 2: Ancient generals will be favored over modern generals at the very top and bottom of the rankings by the model : I suspect this to be true because ancient generals had less technologically advanced weaponry and therefore a greater opportunity to achieve numerical upsets and accumulate more WAR through great strategy or lose more WAR via the opposite. This means the 'greatest' and 'worst' generals will likely come from pre 1000 in my estimation. However, this prediction is purely speculative.

6 Evaluation Metrics

Three Metrics were created for the purpose of analyzing the generals.

Wins Above Regular (WAR) This is the principle metric used to rank the generals. It sums up the total value towards victory each unique general contributed over the course of all of their battles.

Total Battles This metric is simple. It sums the number of battles in the dataset each individual general fought in. This can be used both to calculate aWAR and to find the correlation between WAR and number of battles (used later to evaluate hypothesis 1).

Average Wins Above Regular (aWAR) This metric is found by dividing the number of battles each general fought from their total accumulated WAR. This metric is useful because it determines

which generals added the most value per battle. However, It significantly favors or discounts generals with fewer battles on the extremes as they have less data to create a more reasonable average.

7 Project Code and Results Repository

All data files and code both used and created for my project can be found on github at this url: `https://github.com/breese5/Applied_Math_Military_History_Generals_Ranking_Thesis`. All files should be self explanatory in their labels besides 'current run' which is Ethan's cleaned data from his aforementioned project. The files include my textual battle data scrapper, model and evaluation metric calculation code and excel/csv files of all results (some results like the correlation values are contained within the jupyter notebook file for the metric calculations).

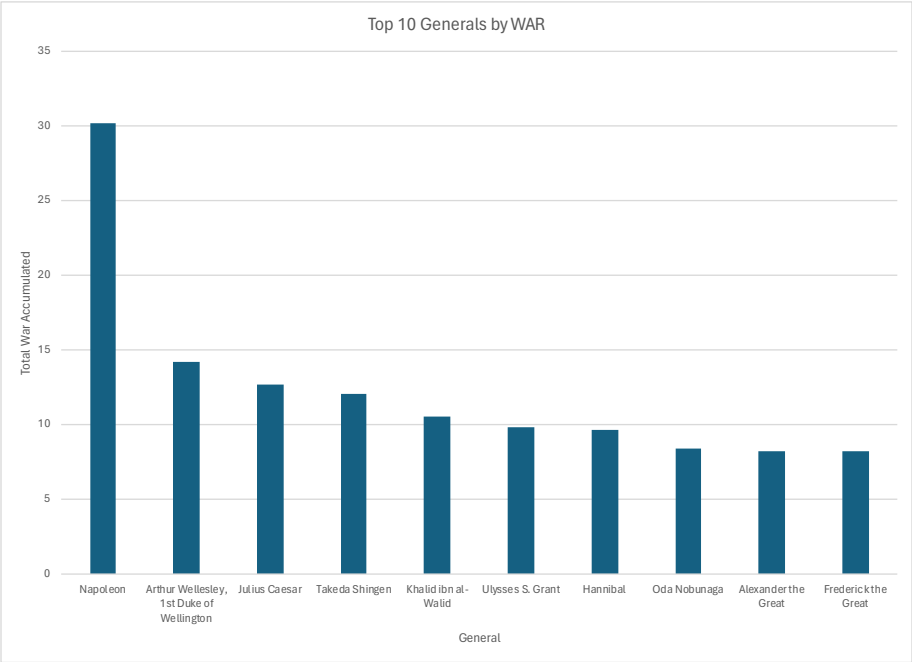

Figure 1: Bar Graph of the Top 10 Generals in total accumulated WAR

8 Results

The Top 10 Military Generals by WAR According to my model, **Napoleon** is by far the greatest military general to ever fight battles. He was able to accumulate 30.165 WAR over 43 total battles, more than double the WAR of Arthur Wellesley, 1st Duke of Wellington, who accumulated 14.224 WAR to come in second place. Julius Caesar, the famous Roman general and politician finished in third place with 12.732 total WAR. Other notable figures in the top 10 include American general and Civil War Hero Ulysses S. Grant (6th), Carthaginian General Hannibal (7th) who once led his entire army through the Alps to attack the Romans, Takeda Shingen (4th) the former Japanese daimyo of the Kai Province and Macedonian king Alexander the Great (9th), who is commonly referenced by historians as one of the greatest military commanders ever for his 10 year campaign across the Persian empire which included revolutionary military tactics and technology such as the phalanx and sarissa. Figure 1 shows the positive WAR of the entire top 10.

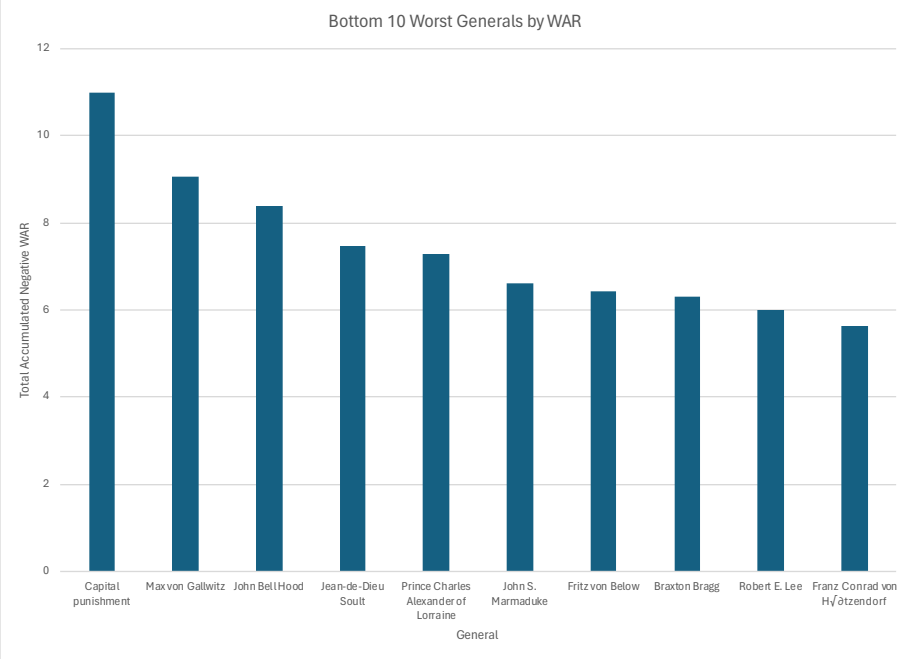

Figure 2: Bar Graph of the Bottom 10 General in total negative WAR accumulated

The Bottom 10 Military Generals by WAR The bottom 10 military generals included some familiarly jeered generals as well as some surprises. Capitol Punishment, not an actual general but instead the death penalty actually received the most negative WAR. Its inclusion is due to the term being included as a belligerent in Wikipedia, where the data was scraped and cleaned from, for 16 different battles. German World War I general Max von Gallwitz was the worst rated actual general, receiving -9.056 total WAR. John Bell Hood, the heavily criticized Confederate general in the American Civil War unsurprisingly comes in third with -8.402 WAR. On the other hand, Robert E. Lee, the renowned confederate tactician who was sought by then president Abraham Lincoln to lead the Union troops very surprisingly accumulated the 9th most negative WAR (-6.011) of the 6619 generals. Lee is the only general with a relatively positive historical tactical reputation to appear on the worst 10 list.

aWAR Results Like expected, aWAR proved to be a relatively unhelpful metric at the extremes. The top and bottom generals in aWAR simply tended to be those with the fewest number of battles. This occurred when my minimum was 1 battle, 2 battles and the eventual minimum I settled on, 3 battles, which at least produced a few generals with 4 battles in the bottom and top 10. One interesting name to note here (Figure 3) is Adolf Hitler, who achieved the highest negative aWAR of any of the 6619 generals that reached the minimum requirement of 3 battles. This is likely because Hitler was not tagged in Wikipedia on specific battles, but instead in larger battle conflicts like 'Western Front' and 'Eastern Front' of World War II in which the Germans obviously eventually lost. Still, it is interesting to see his name at the top of the list given Nazi Germany's early military success prior to being pushed back. Another interesting observation is generals who accumulated drastically different WAR values in the rankings despite similar aWAR values. For example, Dwight D. Eisenhower and Napoleon both had aWAR's of .7 while Napoleon was the number 1 rated general with 30.2 total WAR compared to Eisenhower's modest 2.1 WAR over 3 battles.

General	Number of Battles	aWar
Lin Biao	4	1.092
Liu Bocheng	3	1.034
Stefan Czarniecki	3	1.015
Seleucus I Nicator	4	0.999
Hannibal Mago	3	0.981
Masaharu Homma	3	0.98
Maurice de Saxe	3	0.977
Fernando d'Avalos	3	0.973
Georg von Frundsberg	3	0.973
Robert Guiscard	3	0.96
Adolf Hitler	3	-1.244
Wei Lihuang	3	-1.197
Joseph Bonaparte	3	-1.088
Karl August, Prince of Waldeck and Pyrmont	3	-1.068
Andrea Doria	4	-1.064
Jonathan M. Wainwright (general)	3	-1.06
Jacques MacDonald	3	-1.058
Conrad III of Germany	3	-1.056
William Alexander, Lord Stirling	3	-1.052
Lal Singh	3	-1.052

Figure 3: Top and Bottom 10 of aWAR (minimum 3 battles fought)

Full Results A full searchable excel file of all 6619 generals and their corresponding WAR, aWAR and Number of Battles is available in the 'Generals Performance' excel spreadsheet in the project GitHub Repository.

9 Hypothesis Results

Hypothesis 1 Surprisingly, number of battles and WAR seem to have very little positive correlation. The correlation value was 0.266, which is below the .3 benchmark commonly used to infer slight correlation. This means my assumption that more battles would equal more WAR most times was not correct. Some reasons for this could include generals I did not expect who fought a large number of battles and lost more than they won, which would lead to negative WAR or loss and won equally, leading to a WAR around 0. aWAR and number of battles showed almost no correlation either way (0.060) which makes sense given the metric aWAR is calculating an average of a metric which already showed little correlation with number of battles.

Hypothesis 2 This hypothesis was again wrong as the generals in the top 10 of WAR were relatively evenly split between pre 1000 and post 1000 figures. Four generals in the top 10 (Julius Caesar, Hannibal, Khalid ibn al-Walid and Alexander the Great) were from pre 1000 while six generals (Napoleon, Arthur Wellesley, Takeda Shingen, Oda Nobunaga, Frederick the Great and Ulysses S. Grant) were from post 1000. My prediction was even more wrong for the bottom 10, where all 10 generals are from post 1600. This means that my assumptions about opportunity and technology were incorrect in examining the ability to overcome tactical advantage or disadvantage in my model.

10 Problems

The first problem is the exclusion of any generals and battles not contained in the Wikipedia pages from which the data was derived. For example, Ghengis Khan, the Mogul leader who achieved many military victories during his time and is commonly refereed to by historians as one of the great generals of history is not present in the data nor my subsequent rankings.

Secondly, my model fails to incorporate more human decision-making elements of military leadership such as oration and timing. The linear regression model is based on the numbers rather than textual historical accounts. This is why I created the textual battle descriptions dataset for future use to incorporate some of these factors into future re-rankings.

11 Conclusion

Overall, my rankings provide an interesting exploration of generals tactical acuity. High rankings for commonly revered generals such as Napoleon, Julius Caesar, Hannibal and Alexander the Great imply that my model was successful in being somewhat consistent with historians analysis. Additionally, poor performing generals like Robert E. Lee and Hitler bring up interesting discussions about their tactical acuity for historians to answer with a larger view of the whole picture surrounding their battles. I am pleased that the top and bottom of my rankings covered multiple different geographical areas as well as time periods, implying low bias based on those factors. It is certainly defensible historically to declare Napoleon the greatest military general to ever live, as my model does. Overall, I am happy with the results of the model and had a great deal of fun completing these rankings. I am excited to work more in the future with my textual dataset to further improve their historical accuracy. Thanks to Professor Arman Cohan for agreeing to advise my project as well as Professor John Lewis Gaddis for helping me get started with advice on locating data and important model factors in the early stages of my thesis.

References

[1] Ethan's Project: https://towardsdatascience.com/napoleon-was-the-best-general-ever-and-the-math-proves-it-86

[2] Professor Kimura's Database Description (database is not publicly available): https://osf.io/mdjzu/