

The Sport of War: Balance Analysis of Vehicle Performance in World of vehicles

Regardless of where you were from and which generation you grew up in, we all had a phase in life when we were fascinated by soldiers, weapons, and things that go boom. The military fandom is one that transcends cultural and geographical borders. While this fandom traditionally involved running around (e.g. paintball, playing chief), the rise of the Internet brought with it a completely new dimension, virtual battle, and a new wave of military-focused games that attracted billions of players worldwide and built their own fandoms.

Our project aims to dissect the fandom surrounding World of vehicles, a Belarusian title that focuses on real-life armored vehicles that existed between World War I and the Cold War. The title is immensely popular in former Soviet countries, while also having significant fan bases across other regions, including the US, Asia Pacific, and Europe. As the title is slower-paced and requires more strategic thinking than traditional FPS titles, it is able to attract a much wider spectrum of military fans, one that spans generations.

World of Tanks is the perfect game to conduct data analysis due to its data-heavy nature. The vehicles in the game are balanced around 50+ parameters that affect their mobility, firepower, armor and view range. As the fandom is saturated with hard-core historians and military fans, the developers are very careful with striking the balance between a vehicle's historical accuracy and in-game performance.

I. Data Gathering

The data for this project was collected from fan-made wiki sites and the official website, using web scraping and API connection. Our final dataset consists of vehicle-level data split amongst three sheets, which all share the same list of 707 vehicles but cover different aspects of the fandom.

Sheet Name	Vehicle Parameters	Expected Performance	Actual Performance
Sheet Dimensions	50+variables * 707 vehicles	5 variables * 707 vehicles	5 variables * 707 vehicles
Sheet Details	4 Sub-categories: firepower, mobility, protection, characteristics	Each Vehicle's expected performance per-match (number of kills, number of hits, kilometers traveled, etc.)	Each Vehicle's actual performance across different servers, and the average skill level of the players playing the vehicle

Figure 1. Dataset Overview

The Vehicle Parameters table is further broken down into subcategories.

Subcategory	Example variables
Firepower	Cannon caliber, Reload time, Accuracy, penetration
Mobility	Engine Horsepower, Traverse speed, Top Speed, Power/weight ratio
Protection	Front, Side, Top armor, Armor angle, the existence of spaced-armor
Characteristics	Vehicle tier (approximate decade), Country of origin, Special Status

Figure. 2. Vehicle Parameters Table Detail

As we were able to gather data across different geographic servers, we were interested in seeing if the World of Tanks fandom and military fandom in general differ by region. For example, do Central-and Eastern European (CES) players play differently from the Asia Pacific (ASIA) players? If so, in what ways are they different?

II. Exploratory Data Analysis

To answer our question, we first conducted a thorough exploratory data analysis. Just by looking at the summary statistics, we can immediately find some interesting patterns. For example, players from the CES region and the Americas each like to play their respective region's vehicles.

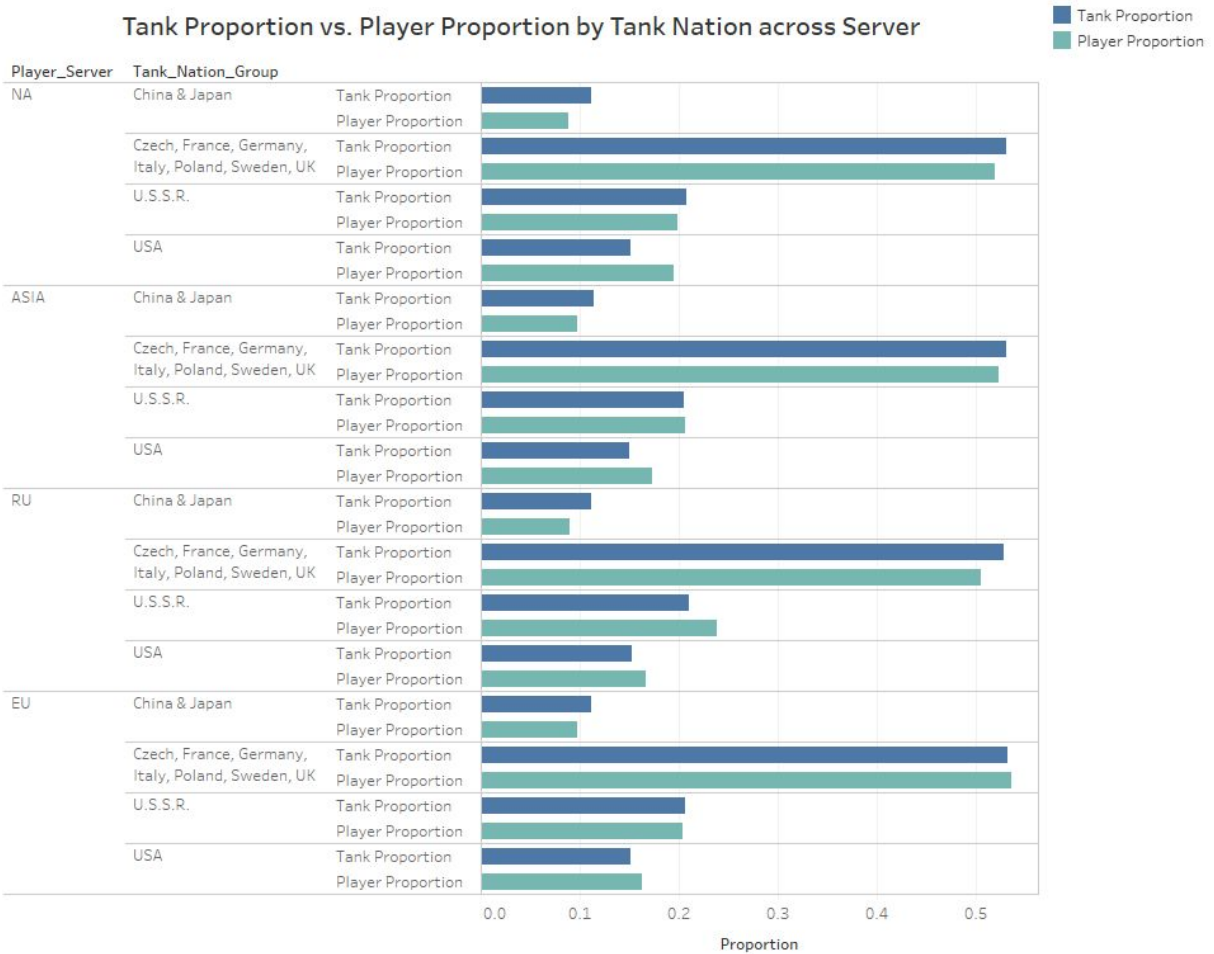


Figure 3. Tank Proportion vs. Player Proportion by Tank Nation Across Servers.

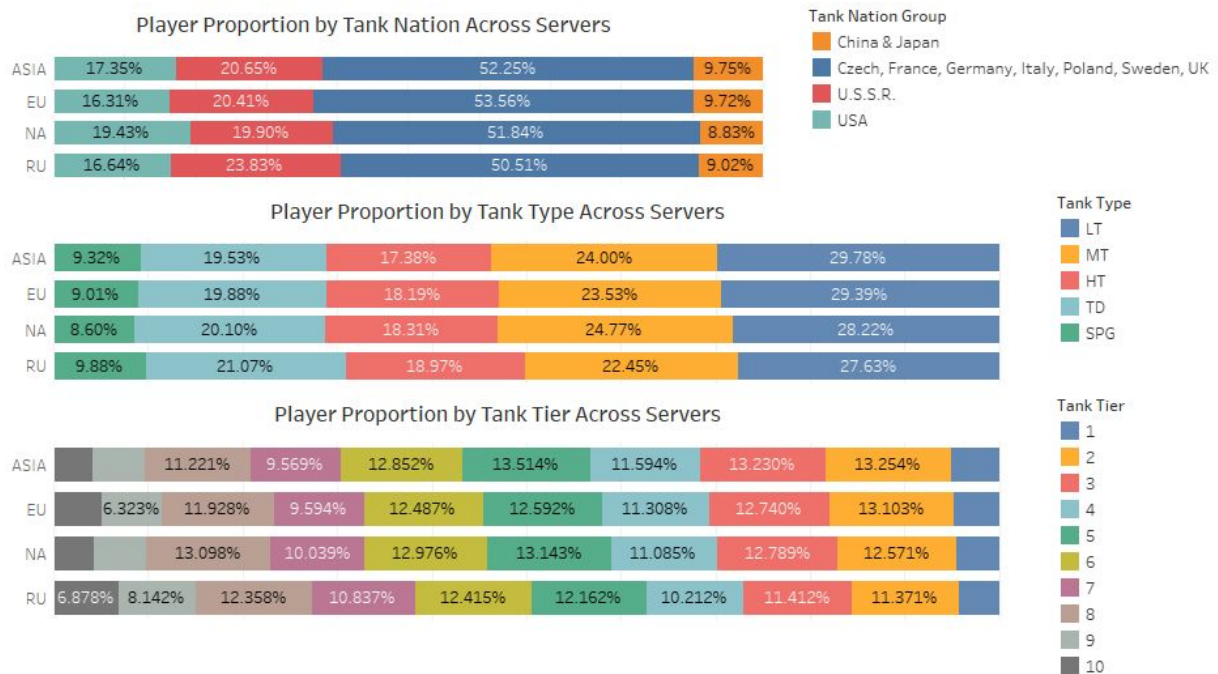


Figure 4. Player Proportion by Tank Nation, Tank Type, Tank Tier Across Servers.

We also noticed some expected trends, for example, Hitpoints (HP) is positively correlated with tier, and heavily dependent on vehicle class. Tank destroyer (TD) vehicles typically have a higher alpha (damage per shell) for their cannons but suffer from longer reload.

III. Regression Models

With our list of the most popular and successful vehicles of each server, we now wanted to see which vehicle parameters each region likes. This is very important in deciphering the unique playstyle of each region. For example, do CES players like to rush forward in heavy vehicles? Do ASIA players like to snipe with highly accurate vehicles from the back? To answer this, we ran four regression models (one for each region) that used the full list of vehicle parameters to predict their popularity. The results are as follows:

1. CES players like playing vehicles with lower horsepower/ton ratio and gun-depression. This is characteristic of Cold-War era Soviet heavy tanks, which are slow but well-armored vehicles that excel at fighting on flat ground. This shows us CES players really like a good close-quarters fight like a brawl, rather than the playstyles seen in other regions. This could be heavily influenced by the fact that most Soviet tanks were designed with the philosophy of overwhelming the enemy with numbers.
2. American players like playing vehicles with lower HP and lower alpha, but higher view range. This is characteristic of American and British vehicles, which rely on distance and precision rather than pure alpha, armor or HP. This tells us that American players really like fighting enemies from afar with precision and using natural cover like hills and can be more calculating before jumping into situations. Unlike the Soviet vehicles, the design philosophy of western tanks during the Cold War relied on tactical superiority over pure numbers, where the command would often send in a few soldiers as necessary to guarantee the victory.
3. The European players did not have a single defining playstyle and were all over the place. However, this makes sense, as Europe is a very diverse region with many different philosophical doctrines to combat and tank design. Countries like France and the UK adopted Western systems like the US, while countries like Hungary and the Czech Republic adopted Soviet systems like the CES region. Other countries like Italy and Germany had their own design doctrines that were independent of both camps.
4. Lastly, The ASIA region enjoys vehicles with low gun depression, low damage, penetration, and tier. This is indicative of the fact ASIA players like playing WWI and WWII era tanks, and less so Cold-War era tanks. This is likely because unlike WWII, the Cold-War wasn't as significant in this area. This tells us that ASIA players prefer wolfpack tactics and playstyles, which rely on

small teams of vehicles that acted in unison given the immaturity of tank development of the WWI/WWII era.

#Players~	Significantly Negative	Significantly Positive
RU	Damage, Penetration (mm), Tank traverse (°/sec), Gun depression(°), Power(hp)	Aim time (sec)
EU	Damage, Penetration (mm), Tank traverse (°/sec)	Aim time (sec)
ASIA	Damage, Penetration (mm), Tank traverse (°/sec), Gun depression(°)	Aim time (sec)
NA	Damage, Penetration (mm), Tank traverse (°/sec), Health	Aim time (sec)

Figure 5. Parameter Contribution to Number of Players.

Secondly, we wanted to look at which parameters are actually important for actually winning in battles. Having the highest cannon caliber may be cool and all, but it is no use if you never have a chance to fire it or lose every match. We ran a regression model that looks at the vehicle parameters in relation to the win percentage of each vehicle on each server. Similar to football where fans typically have higher expectations for better athletes, these expected combat efficiency values measure how players are expected to do in that vehicle on average. The results are as follows:

1. Vehicles with lower ammunition rack health (easier to die) but higher damage per minute (number of shells a cannon can fire a minute), and gun traverse range tend to win more on the CES server. Vehicles with these traits are typically tank destroyers who ambush enemies from a distance.

2. Vehicles with lower penetration but better hull armor tend to win more on the EU server.

Vehicles with these traits are typically heavy tanks who have a lot of armor to brawl in close-quarters.

3. Vehicles with higher gun depression angles and hull armor tend to win more on the ASIA server.

Vehicles with these traits are typically Western Bloc medium tanks that use natural cover to avoid getting hit while maximizing frontal armor to ricochet any shells that do land.

4. Vehicles with less health but more gun depression tend to win more on the NA server. Vehicles with these traits are typically Western Bloc medium and light tanks that use mobility, natural cover, and spotting the enemy before being spotted themselves.

Average Win Rate~	Significantly Negative	Significantly Positive
RU	Ammo rack health	DPM, Gun traverse range
EU	Penetration (mm)	Hull front
ASIA		Gun depression(°), Hull front
NA	Health	Gun depression(°)

Figure 6. Parameter Contribution to Average Win Rate .

IV. Cross-Model Interpretation

Lastly, we wanted to combine the two sections above and compare the parameters that make vehicles popular versus the parameters that make a vehicle better than their peers. This will tell us the reasoning behind how players of each region choose which vehicles to play.

1. CES server players like playing vehicles that brawl face-to-face, which probably saturated the server with these kinds of players and vehicles. Thus, it would make sense for tank destroyers who ambush and pick these players off from afar to win more games on the CES server.

2. NA server players like playing vehicles that engage the enemy from afar. Those who charge in like CES players will typically be dealt with easily at range. Thus, it also would make sense for medium and light tanks who also engage players from afar and rely on flexibility, using the terrain and view range to win more games on the Americas server.
3. ASIA server players like playing lower tier WWI/WWII tanks. However, the game design makes it so that Cold-War era tanks can meet these tanks on the battlefield. Thus, it would make sense for Western Bloc medium tanks that use strategy and natural cover to completely dominate the battlefield. These WWI/WWII era tanks will have a very hard time of defeating technologically superior vehicles from the future.
4. Last but not least, European server players like playing a diverse array of tanks. Thus, heavy tanks that have the best protection amongst all tank classes would surely win the most battles, as they can deal with the most diverse range of situations, while other tank classes are typically more specialized in dealing with particular situations and tank classes (e.g. Tank Destroyers ambushing Heavy Tanks, etc.)

V. Conclusion

The military fandom reflected within the World of vehicles fandom told us very interesting patterns in military fans around the world. Each server has different preferences for vehicles, and is better at different vehicles based on the vehicle's class, tier, and country of origin. The players of some servers are flat out better than other servers. Each server also prefers different vehicle characteristics, which tell us the playstyle of each region and can even reflect the vehicle design philosophy of each nation. Lastly, the most popular vehicle characteristics are not actually the most effective characteristics for winning. In fact, the vehicles that directly counter the most popular tanks tend to be the vehicles that win the most

amount of battles. For future steps, we would like to build a prediction model that uses player data to conduct segmentation and clustering and helps recommend vehicles for players based on their playstyles.

VI. Appendix

a. Sample Web Scraping Code (a copy of this Jupyter Notebook file is included in the submission)

```
In [259]: urlpage = 'https://tanks.gg/list'
driver = webdriver.Chrome(executable_path = 'C:/Users/Jie Z/Downloads/chromedriver.exe')

In [260]: driver.get(urlpage)
driver.execute_script('window.scrollTo(0,document.body.scrollHeight);var lenOfPage=document.body.scrollHeight;return lenOfPage;')

Out[260]: 10174

In [146]: links = driver.find_elements_by_partial_link_text('')
urls = []
for link in links:
    urls.append(link.get_attribute('href'))
alllink = urls[93:800]

In [236]: alllink[125]

Out[236]: 'https://tanks.gg/tank/m103'

In [245]: i=0

data0 = []

for link in alllink[0:100]:
    print(i)
    driver.get(link)
    driver.execute_script('window.scrollTo(0,document.body.scrollHeight);var lenOfPage=document.body.scrollHeight;return lenOfPage;')

    results = driver.find_elements_by_xpath("//*[@class='clearfix header']")
    results0 = driver.find_elements_by_xpath("//*[@class='stat-line']")
    results1 = driver.find_elements_by_xpath("//*[@class='stat-line first']")
    results2 = driver.find_elements_by_xpath("//*[@class='stat-line highlight']")
    results3 = driver.find_elements_by_xpath("//*[@class='stat-line highlight first']")

    pairs = {}
    pairs.update(tank_name = results[0].text.split('\n')[0])
    pairs.update(tier = results[0].text.split('\n')[1].split()[1])
    pairs.update(nation = results[0].text.split('\n')[1].split()[2])
    pairs.update(tank_class = "".join(results[0].text.split('\n')[1].split()[3:]))
    for result in results0:
        pairs.update({result.text.split('\n')[0] : result.text.split('\n')[1]})
    for result in results1:
        pairs.update({result.text.split('\n')[0] : result.text.split('\n')[1]})
    for result in results2[0:3]:
        pairs.update({result.text.split('\n')[0] : result.text.split('\n')[1]})
    for result in results2[5:]:
        pairs.update({result.text.split('\n')[0] : result.text.split('\n')[1]})
    for result in results3:
        pairs.update({result.text.split('\n')[0] : result.text.split('\n')[1]})
    data0.append(pairs)
    i=i+1
```

b. Sample Processed DataFrame

```
In [285]: 1 alltank = alltanks[df0.columns]
```

```
In [286]: 1 alltank.head()
```

Out[286]:

	tank_name	tier	nation	tank_class	Damage	Penetration (mm)	Reload time (sec)	Rate of fire (rnds/min)	Clip size	Clip damage	Intra-clip reload	Aim time (sec)	... moving	... tank traverse	... turret traverse	... after firing	damag
0	AMX 13 105	X	French	LightTank	390	234	26.43	5.64	3	1,170 / 5.45s	2.73	2.21	0.14	0.14	0.12	3.36	1.
1	Manticore	X	British	LightTank	390	248	13.42	4.47	NaN	NaN	NaN	1.90	0.15	0.15	0.08	3.50	
2	Panhard EBR 105	X	French	LightTank	390	190	12	5	NaN	NaN	NaN	1.15	0.06	0.06	0.03	2.30	1.
3	Rheinmetall Panzerwagen	X	German	LightTank	320	242	8.81	6.81	NaN	NaN	NaN	1.53	0.13	0.13	0.10	2.88	1.

```
In [290]: 1 alltank.to_csv('all.csv',index=False)
```

```
In [280]: 1 alltanks.columns
```

```
Out[280]: Index(['tank_name', 'tier', 'nation', 'tank_class', 'Damage',  
                'Penetration (mm)', 'Reload time (sec)', 'Rate of fire (rnds/min)',  
                'Aim time (sec)', '... moving', '... tank traverse', '... turret traverse',  
                '... after firing', '... damaged', 'Caliber (mm)', 'Shell velocity (m/s)',  
                'Max range (m)', 'Ammo capacity', 'Potential damage', 'Shell cost',  
                'Reverse speed (km/h)', 'Power (hp)', 'Power/weight (hp/t)',  
                'Tank traverse (°/sec)', '... hard (°/sec)', '... medium (°/sec)',  
                '... soft (°/sec)', '... hard', '... medium', '... soft', 'Gun depression (°)',  
                'Gun traverse (°/sec)', 'Gun traverse range (°)',  
                'Turret traverse (°/sec)', 'Hull armor (mm)', 'Turret armor (mm)',  
                'Track armor (mm)', 'Engine health', '... repaired', 'Track health',  
                'Ammo rack health', 'Total weight (kg)', 'Radio range (m)',  
                'Fire chance', 'Tank cost', 'Damage vs. modules', 'Gun elevation (°)',  
                'Max load (kg)', 'DPM', 'Dispersion', 'Top speed (km/h)',  
                'View range (m)', 'Health'],  
                dtype='object')
```