

Grand Theft Auto (GTA) Online Data Analysis

Saul Topete
Fall 2023

About GTA Online

- GTA Online is a game developed by Rockstar Games, where a player is put in an open-world environment and has the ability to complete different types of activities, purchase different items, or engage against or with other players.
- GTA Online was released in 2013, and has had various updates to provide players with more activities and items to interact with and purchase.
- One way a players progress in GTA Online is tracked is through reputation points (RP), which determine the rank a player is in the game.
 - E.g., a player with a lower rank will not be able to purchase certain items (weapons, clothing) that someone with a higher rank will be able to.
- Since all players at any rank are able to complete activities, it would make sense that higher level players should spend a greater in-game dollar amount than lower levels because of more money accumulated and more items to purchase?

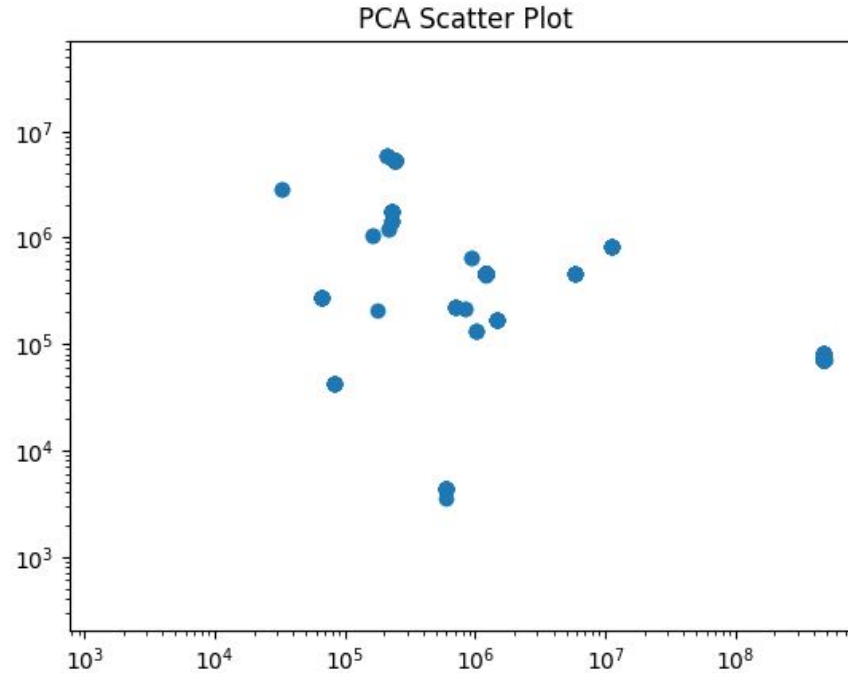
GTA Online Player's Ranks

- The rank of a player is an integer value anywhere between 1 - 8000.
 - 1 being the lowest rank, 8000 being the highest rank.
- All unique players on each console from dataset:
 - **PS4**: 2039
 - **Xbox**: 4451
 - **PC**: 1445
- Mean of all players' ranks for each console:
 - **PS4**: 53.862611579800365 ~ 54
 - **Xbox**: 51.482803613007405 ~ 51
 - **PC**: 70.29364013757417 ~ 70
- Median of all player's ranks for each console:
 - **PS4**: 40
 - **Xbox**: 38
 - **PC**: 52

GTA Online Player's Ranks (cont.)

- Active unique players on each console from dataset (active meaning the player has either 1 kill, death, or suicide tracked stat):
 - **PS4:** 526
 - **Xbox:** 1190
 - **PC:** 354
- Mean of active players' ranks for each console:
 - **PS4:** 51.00570342205323 ~ 51
 - **Xbox:** 56.15042016806723 ~ 56
 - **PC:** 67.4322033898305 ~ 67
- Median of active player's ranks for each console:
 - **PS4:** 38
 - **Xbox:** 40
 - **PC:** 50




PCA Testing



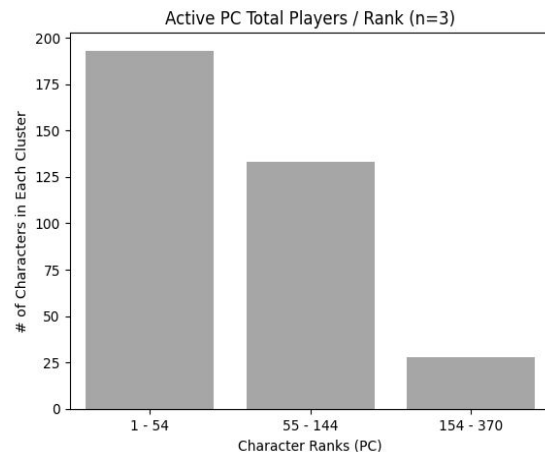
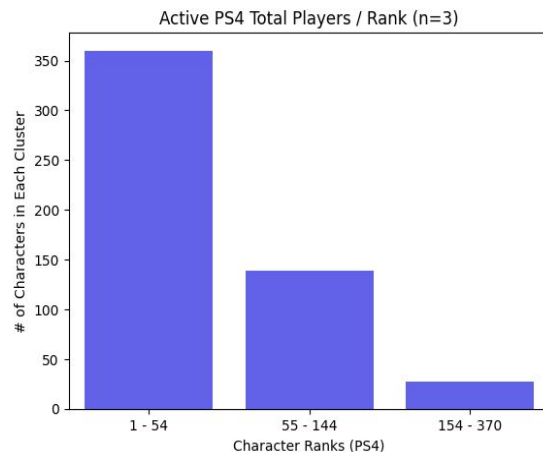
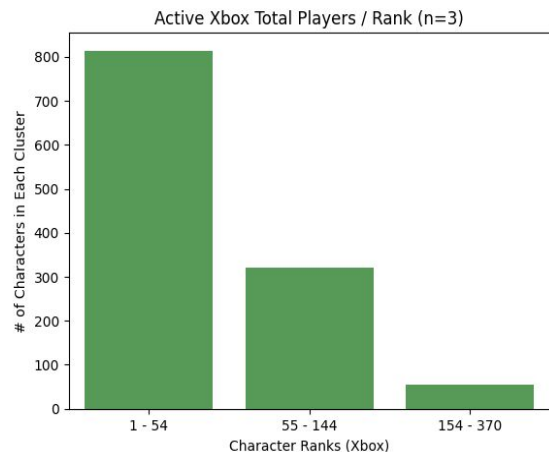
- Used to identify patterns to see if a specific group on playstation stands out, here we see a few outliers on the right side of the PCA plot.

GTA Online Player's Ranks (cont.)

- Ratio of active to total players
 - **PS4**: 0.25796
 - **Xbox**: 0.26735
 - **PC**: 0.24498
- ~25% of players across all consoles that play GTA Online are actually active and have at least 1 kill, death, or suicide.
- The mean rank of players increases on Xbox while decreasing on PC and PS4.
 - One reason could be the larger inactive player base on Xbox, since the game was given for free for Xbox Game Pass subscribers in early 2020 ([link](#)).
 - With the removal of many inactive lower ranks, the mean of active players can increase.

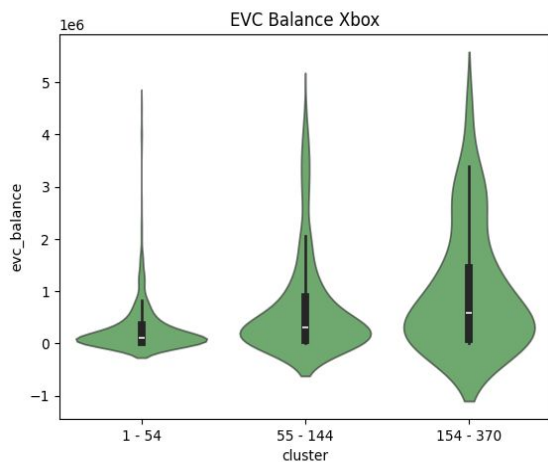
	Mean Rank All Players	Mean Rank Active Players
PS4	54	 51
Xbox	51	 56
PC	70	 67

Player Rank Clustering



- To determine low, mid, and high rank bins, the dataset was first separated into 3, with each one being PS4, Xbox, and PC.
- Once the dataset was separated by console, each console was put into a K-Means Machine Learning Algorithm with $n = 3$, to determine the low, mid, and high rank bins

Player In-Game Dollar Amount



Xbox

Rank 1-54:

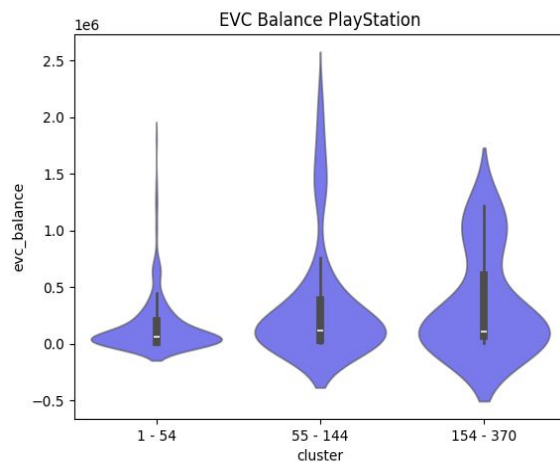
- Mean: \$271,053
- Median: \$108,349

Rank 55-144

- Mean: \$670,351 ↑
- Median: \$286,799 ↑

Rank 154-144

- Mean: \$1,040,882 ↑
- Median: \$628,341 ↑



PS4

Rank 1-54:

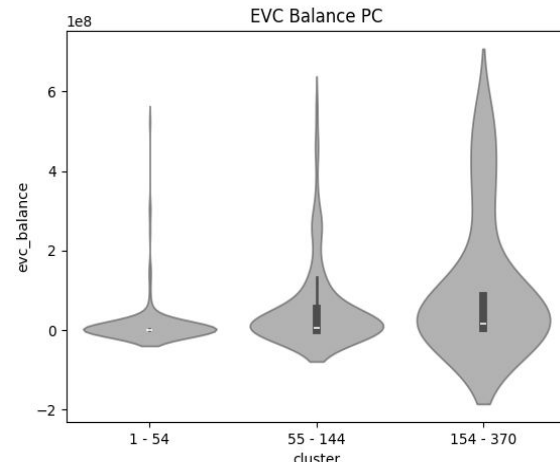
- Mean: \$152,263
- Median: \$55,840

Rank 55-144

- Mean: \$335,467 ↑
- Median: \$117,766 ↑

Rank 154-370

- Mean: \$349,928 ↑
- Median: \$102,755 ↓



PC

Rank 1-54:

- Mean: \$212,155
- Median: \$13,758,003

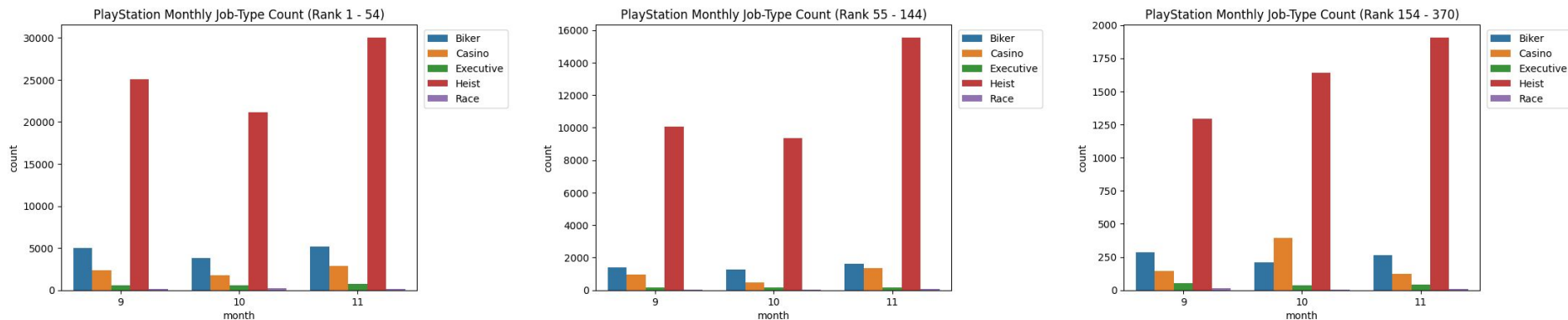
Rank 55-144

- Mean: \$55,686,620 ↑
- Median: \$4,276,826 ↓

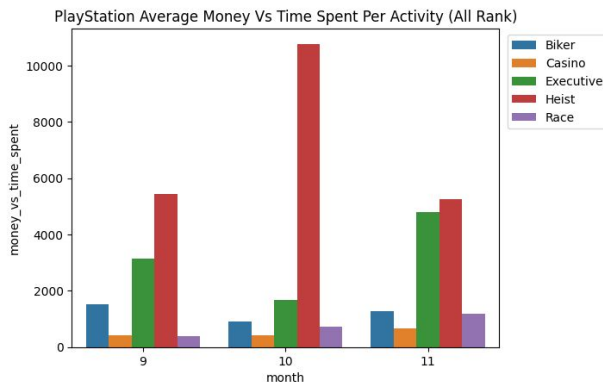
Rank 154-370

- Mean: \$79,481,553 ↑
- Median: \$15,238,969 ↑

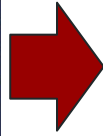
Count of Job-Types Completed By Players



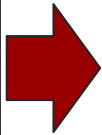
- This is to visualize how players are able to make their in-game earnings, seeing the frequency of which job being completed the most
- The graph to the right shows the ratio of money vs time spent to get an idea of why a job may be completed more than another.



GTA Online Weekly Events from Sept 2020 - Nov 2020



...

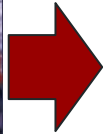


GTA Online Weekly Events from Sept 2020 - Nov 2020

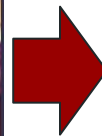


Weekly events in GTA Online provide discounts for purchasable items, and special bonuses for specific activities such as races, missions, versus modes, etc.

...

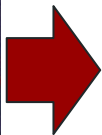
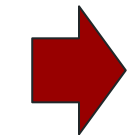


GTA Online Weekly Events Cont. + Question



Weekly events in GTA Online provide discounts for purchasable items, and special bonuses for specific activities such as races, missions, versus modes, etc.

My question is to see if these weekly events affect the amount of what is purchased by players, at different ranks in 3 clusters: low, mid, & high ranks.



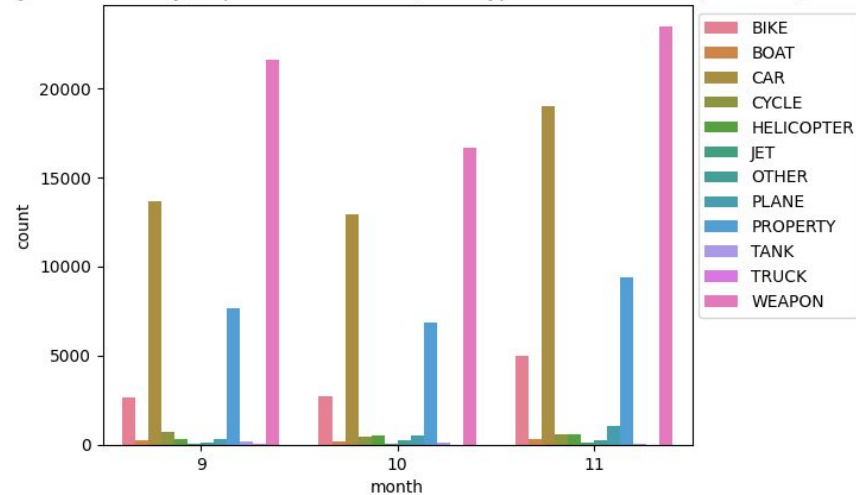
Hypothesis

- Here, I will only be analyzing players who play on PlayStation 4 (PS4) consoles and are active (meaning they have either 1 kill, death, or suicide).
- Null Hypothesis (H_0): There is no difference in the amount a low level player spends on items compared to a higher level player in the top 3 most purchased item types.
- Alternate Hypothesis (H_1): There is a difference in the amount a low level player spends on items compared to a higher level player in the top 3 most purchased item types.

PS4 Category All Ranks Purchase Stats

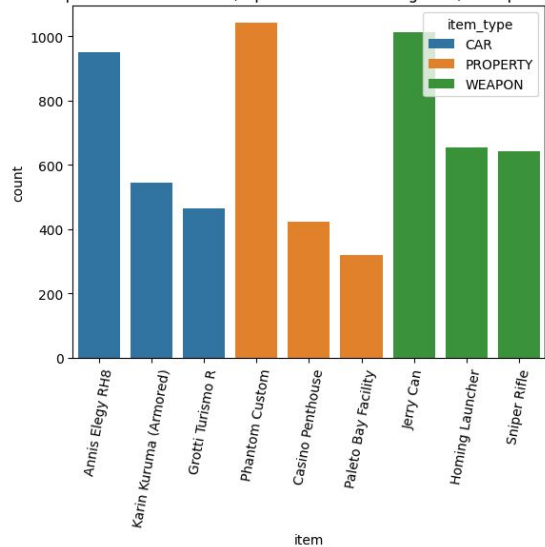
- Between all ranks, the most purchased items each month from September to November, are
 1. Weapons
 2. Cars
 3. Property

PlayStation Monthly (Sept 2020 - Nov 2020) Item-Type Purchase Counts (All Ranks)

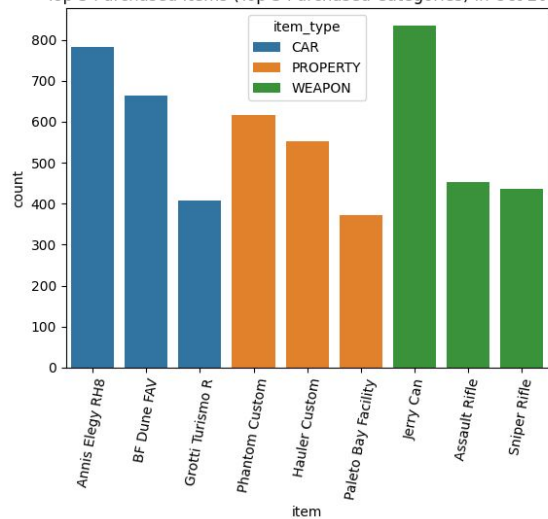


PS4 Top 3 Items in Each Category for All Levels

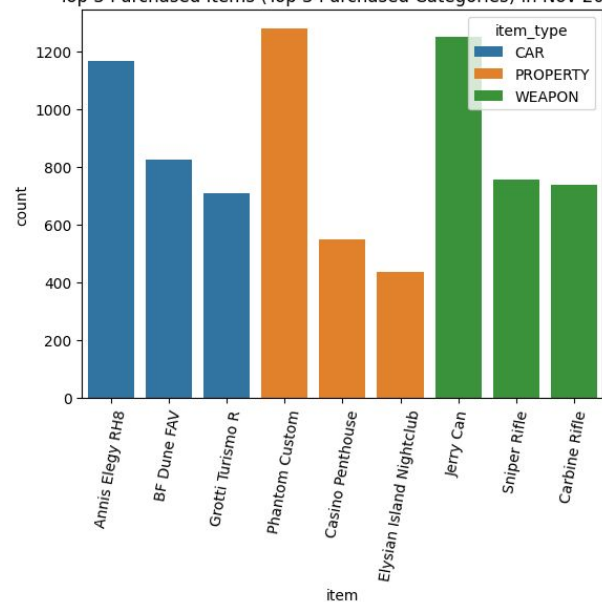
Top 3 Purchased Items (Top 3 Purchased Categories) in Sept 2020



Top 3 Purchased Items (Top 3 Purchased Categories) in Oct 2020



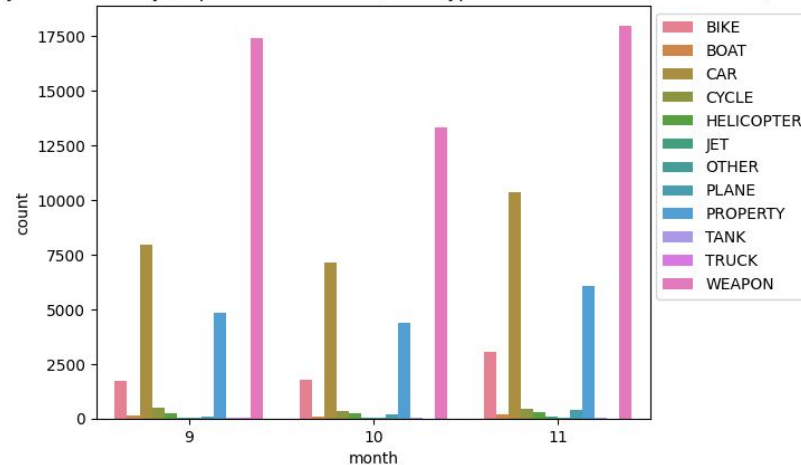
Top 3 Purchased Items (Top 3 Purchased Categories) in Nov 2020



PS4 Overall Categories Low Rank (1 - 54) Purchase Stats

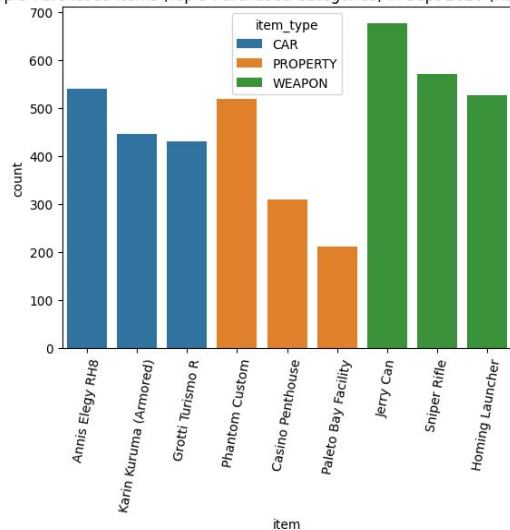
- Seeing lower level ranks, the most purchased items each month from September to November, are
 1. Weapons
 2. Cars
 3. Property

PlayStation Monthly (Sept 2020 - Nov 2020) Item-Type Purchase Counts (Rank 1 - 54)

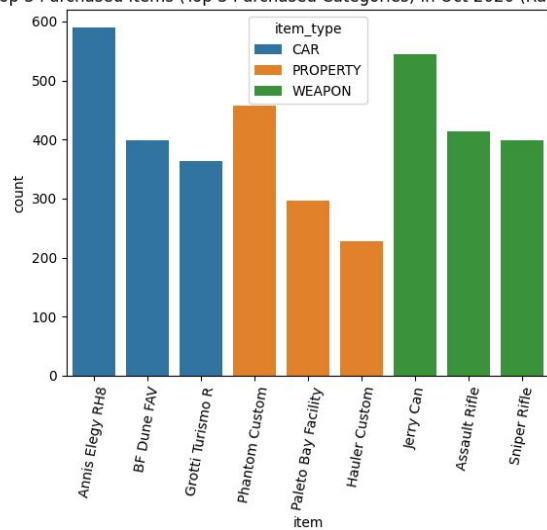


PS4 Overall Categories Low Rank (1 - 54) Purchase Stats Cont.

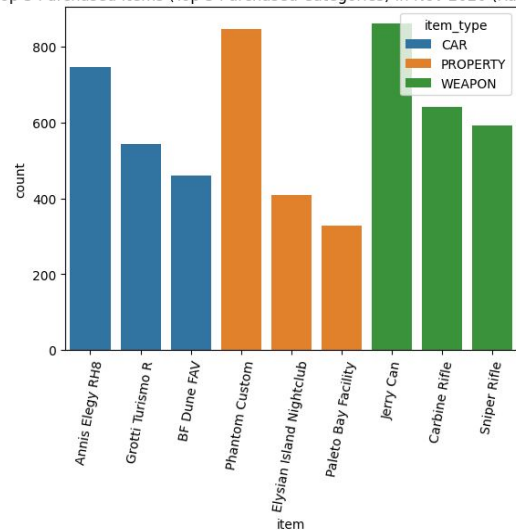
Top 3 Purchased Items (Top 3 Purchased Categories) in Sept 2020 (Rank 1 - 54)



Top 3 Purchased Items (Top 3 Purchased Categories) in Oct 2020 (Rank 1 - 54)



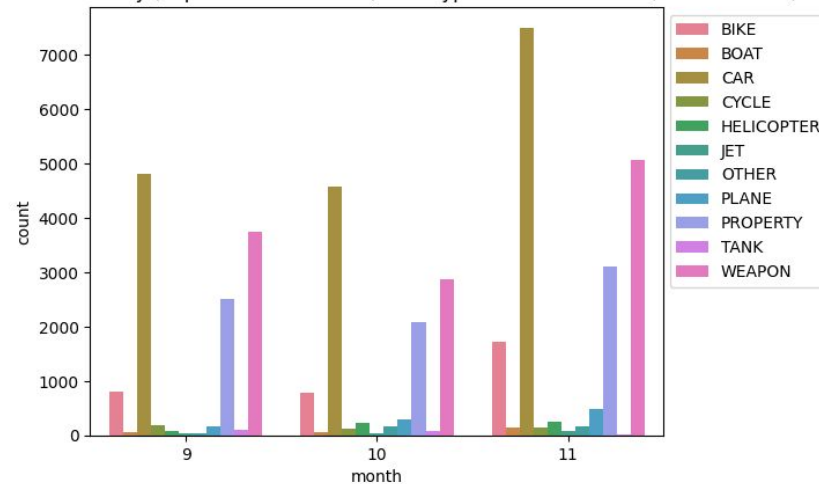
Top 3 Purchased Items (Top 3 Purchased Categories) in Nov 2020 (Rank 1 - 54)



PS4 Overall Categories Mid Rank (55 - 144) Purchase Stats

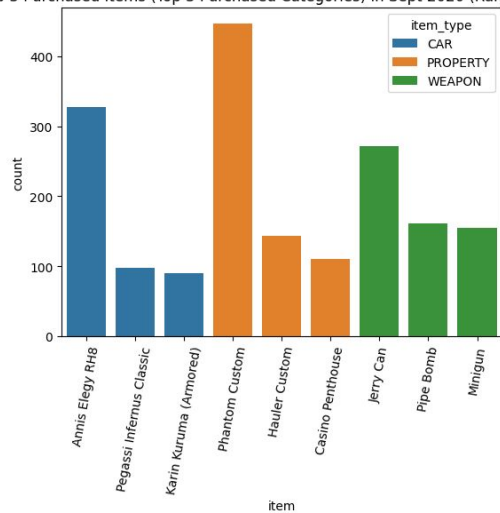
- Seeing mid level ranks, the most purchased items each month from September to November, are
 1. Weapons
 2. Cars
 3. Property

PlayStation Monthly (Sept 2020 - Nov 2020) Item-Type Purchase Counts (Rank 55 - 144)

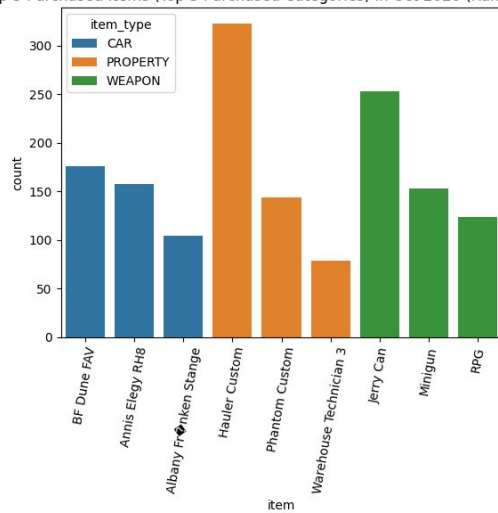


PS4 Overall Categories Mid Rank (55 - 144) Purchase Stats Cont.

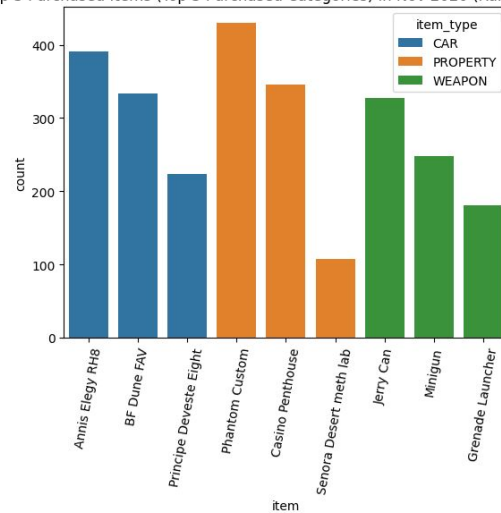
Top 3 Purchased Items (Top 3 Purchased Categories) in Sept 2020 (Rank 55 - 144)



Top 3 Purchased Items (Top 3 Purchased Categories) in Oct 2020 (Rank 55 - 144)

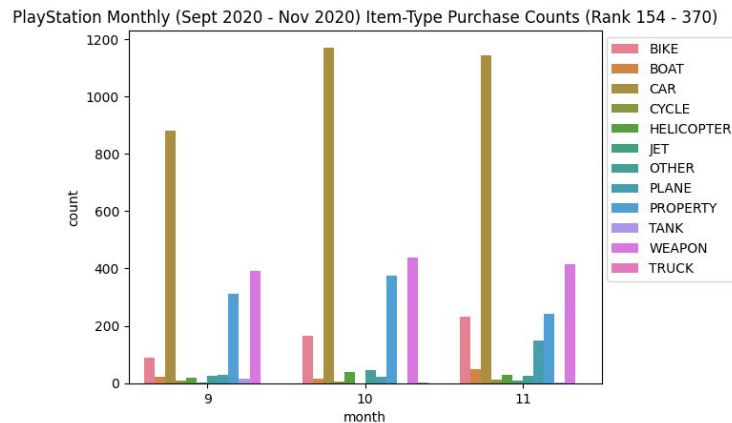


Top 3 Purchased Items (Top 3 Purchased Categories) in Nov 2020 (Rank 55 - 144)



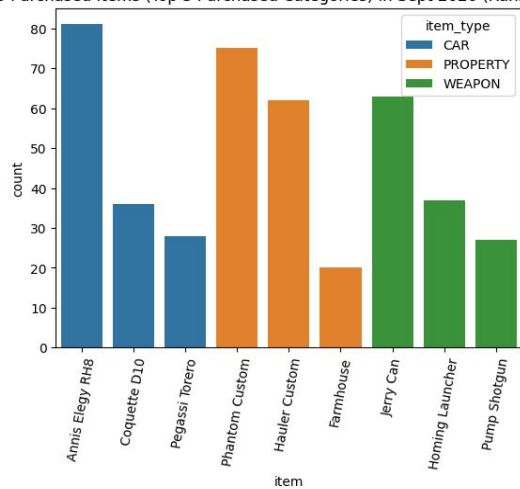
PS4 Overall Categories High Rank (154 - 370) Purchase Stats

- Seeing higher level ranks, the most purchased items each month from September to November, are
 1. Weapons
 2. Cars
 3. Property

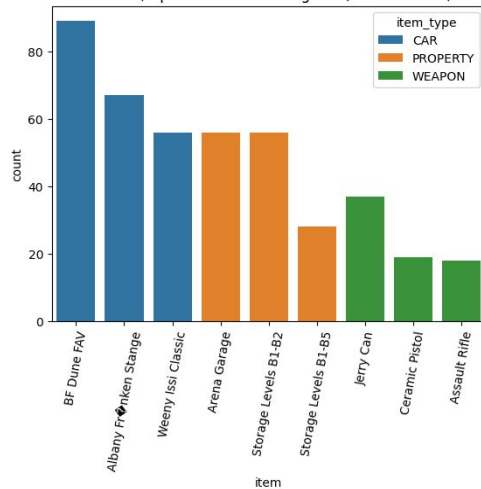


PS4 Overall Categories High Rank (154 - 370) Purchase Stats Cont.

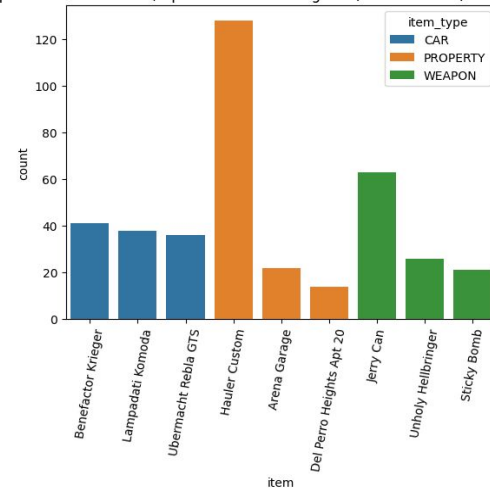
Top 3 Purchased Items (Top 3 Purchased Categories) in Sept 2020 (Rank 154 - 370)



Top 3 Purchased Items (Top 3 Purchased Categories) in Oct 2020 (Rank 154 - 370)



Top 3 Purchased Items (Top 3 Purchased Categories) in Nov 2020 (Rank 154 - 370)



ANOVA Test

- Performing an ANOVA test on the money spent on the top 3 categories and performing them on each cluster, we see the following:

ANOVA					
Dependent Variable: money_spent					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.679E+13	2	4.840E+13	600.279	<.001
Within Groups	3.577E+15	44374	8.062E+10		
Total	3.674E+15	44376			

- The test is significant with a $p < .001$.

Post-Hoc Test

Post Hoc Tests

Multiple Comparisons

Dependent Variable: money_spent

Tukey HSD

(I) cluster	(J) cluster	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.000000000	1.000000000	-84251.91 [*]	3020.69827	<.001	-91331.758	-77172.057
	2.000000000	-156665.6 [*]	6465.98048	<.001	-171820.41	-141510.74
1.000000000	.000000000	84251.907 [*]	3020.69827	<.001	77172.0570	91331.7578
	2.000000000	-72413.67 [*]	6746.32309	<.001	-88225.561	-56601.775
2.000000000	.000000000	156665.58 [*]	6465.98048	<.001	141510.743	171820.407
	1.000000000	72413.668 [*]	6746.32309	<.001	56601.7748	88225.5610

*. The mean difference is significant at the 0.05 level.

Post-Hoc Test Cont.

- Cluster 0 vs Cluster 1 (low rank vs mid rank)
 - The mean difference is -84251.91, with a $p < 0.001$, which is highly significant. ∴, lower rank purchase mean is lower than mid rank purchase mean.
- Cluster 0 vs Cluster 2 (low rank vs high rank)
 - The mean difference is -156665.6, with a $p < 0.001$, which is highly significant. ∴, lower rank purchase mean is lower than high rank purchase mean.
- Cluster 1 vs Cluster 2 (mid rank vs high rank)
 - The mean difference is -72413.67, with a $p < 0.001$, which is highly significant. ∴, mid rank purchase mean is lower than high rank purchase mean.

Post Hoc Tests

Multiple Comparisons						
Dependent Variable: money_spent						
Tukey HSD						
(I) cluster	(J) cluster	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.000000000	1.000000000	-84251.91*	3020.69827	<.001	-91331.758	-77172.057
	2.000000000	-156665.6*	6465.98048	<.001	-171820.41	-141510.74
1.000000000	.000000000	84251.907*	3020.69827	<.001	77172.0570	91331.7578
	2.000000000	-72413.67*	6746.32309	<.001	-88225.561	-56601.775
2.000000000	.000000000	156665.58*	6465.98048	<.001	141510.743	171820.407
	1.000000000	72413.668*	6746.32309	<.001	56601.7748	88225.5610

*. The mean difference is significant at the 0.05 level.

Conclusion

- \therefore , we can reject the null hypothesis and accept the alternate hypothesis for all cases (low vs mid, low vs high, mid vs high) that there is a difference in the amount a low level player spends on items compared to a higher level player in the top 3 most purchased item types.