



Not Enough Sugar For Milk Tea

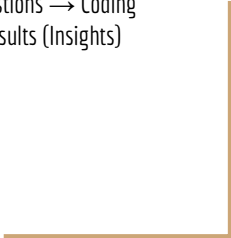
By: Hannah, Jenny, and Manjiri





Our Process

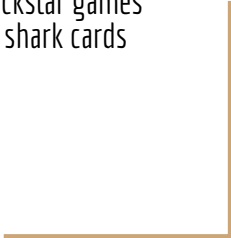
Explore Data Set → Research → Formulation of Questions → Coding
(Data Visualization, Data Cleaning) → Analysis of Results (Insights)
Suggestions for Improving Dataset





Business Aspect (Revenue)

From our research, revenue streams of Rockstar games
→ purchasing video games, and buying shark cards
(used to buy items),



PRE ANALYSIS CLEAN UP

- General player stats 'daily_playtime' column:
 - Dropped rows where values of 'daily_playtime' == 0.
- Item spend data 'item_sub_type' column:
 - Redundant strings were merged into one.
 - E.g. SPORT_CLASSIC, SPORTS CLASSIC, SPORT CLASSIC, SPORTS and SPORT were all categorized as just SPORT

QUESTIONS

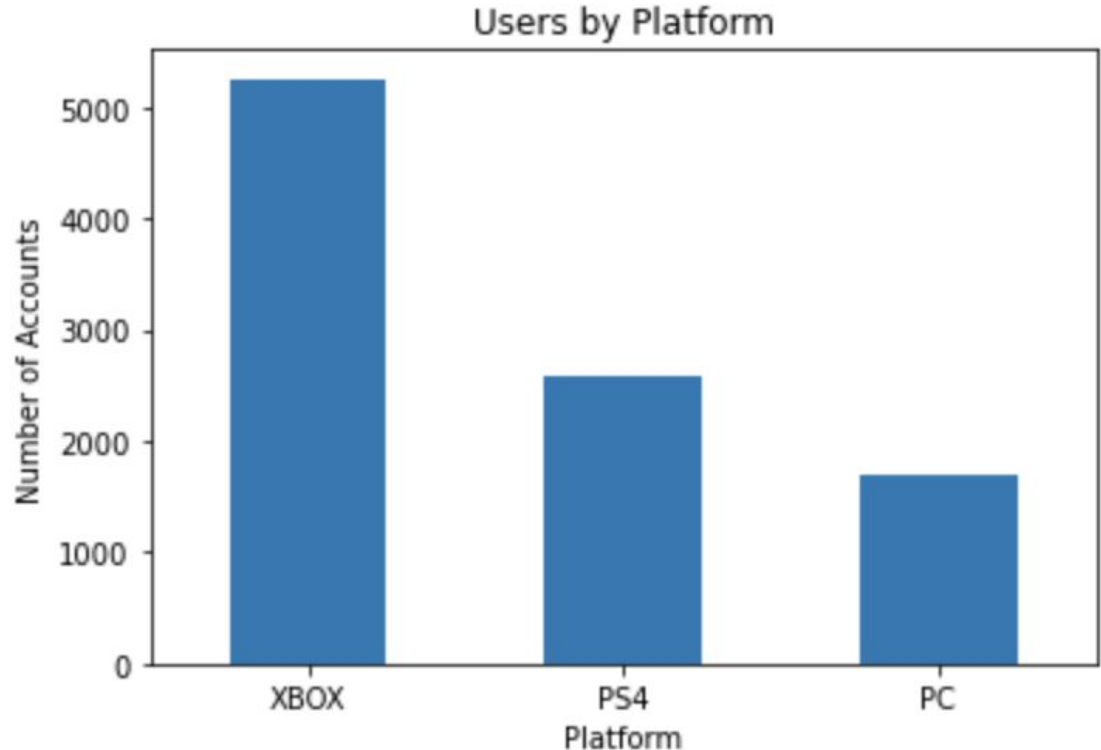
- How to increase video game sales? Through which platform? What features can be enhanced?
 - What is
 - How to attract more players?
 - How to keep players motivated? (player retention)
 - What is the correlation between success—rate of player and money spent and playing time
 - Which game mode is least played and what is the indication of this phenomenon ?
- Remaining questions (which couldn't be answered due to lack of dataset)

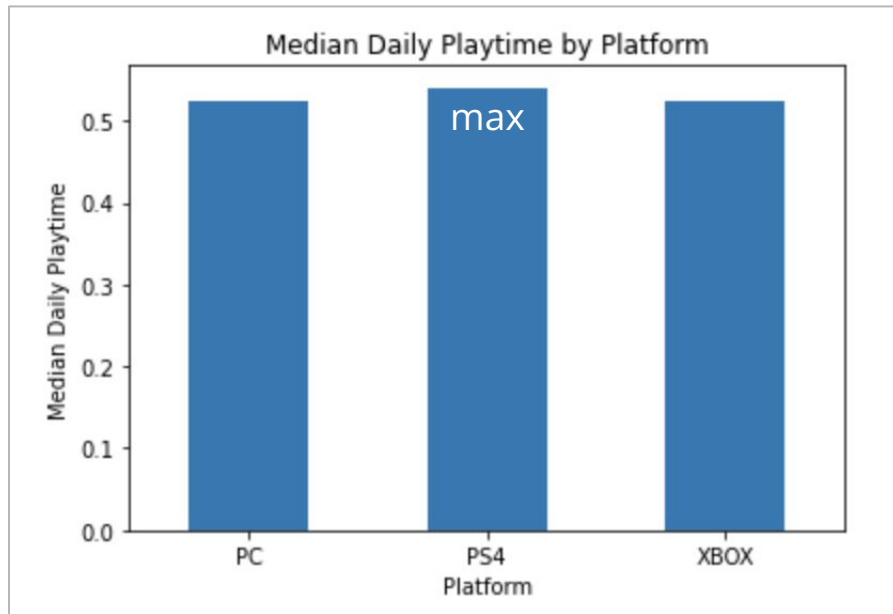
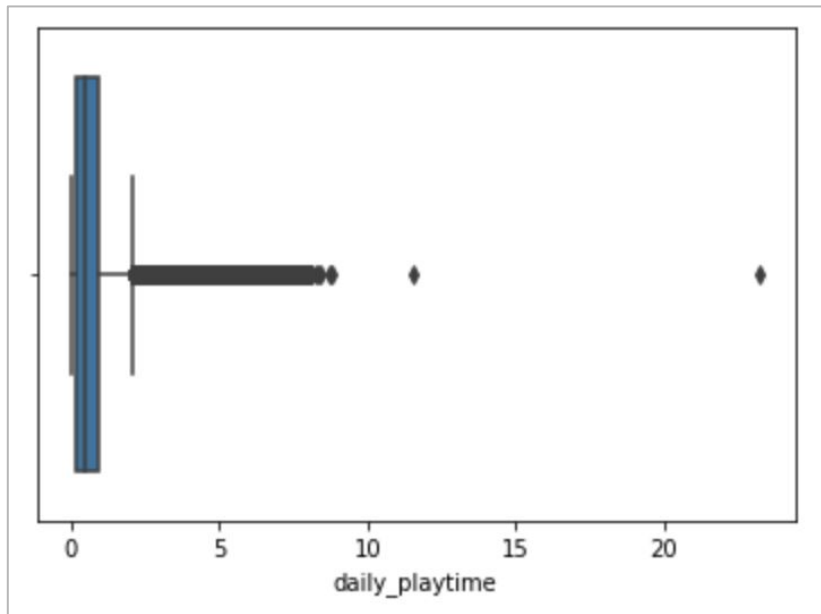
HOW TO INCREASE VIDEO GAME SALES? (*Platform approach*)

Seeing total number
of unique ids by
platform?



XBOX?



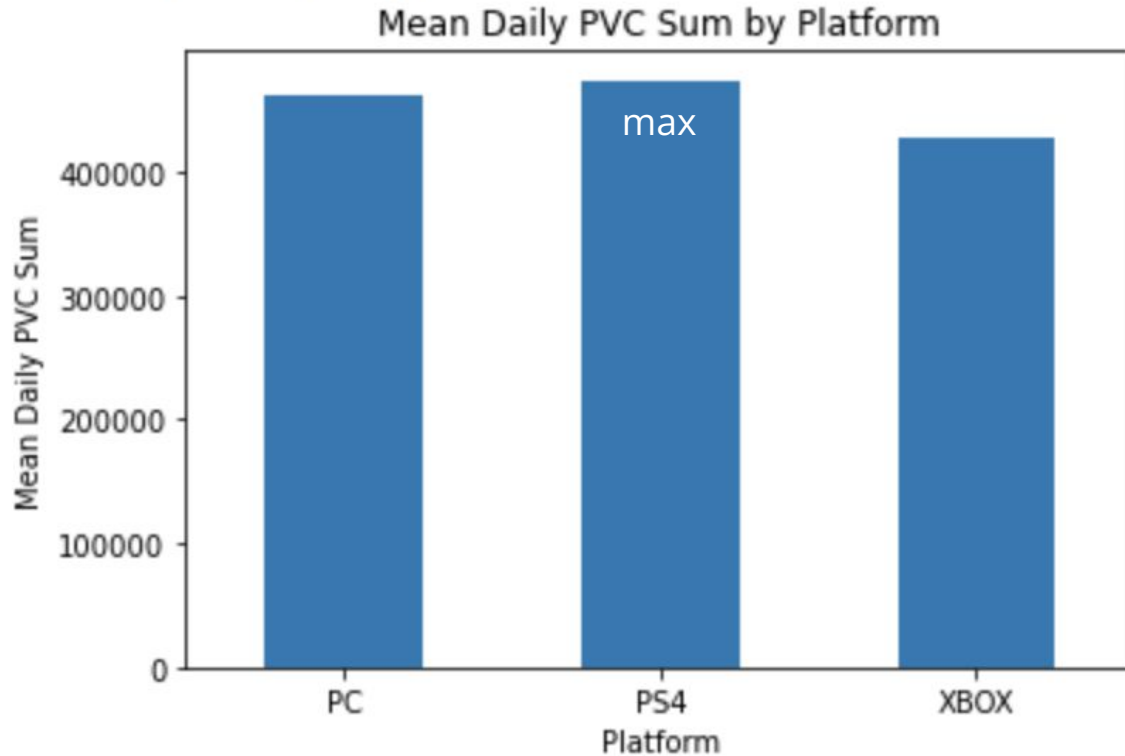


Multiple account id's repeat with different daily playtimes, aggregation was used to find average daily playtime for each account_id → Overall median was taken by platform as daily_platime has many outliers (according to box plot). Median is robust to outliers



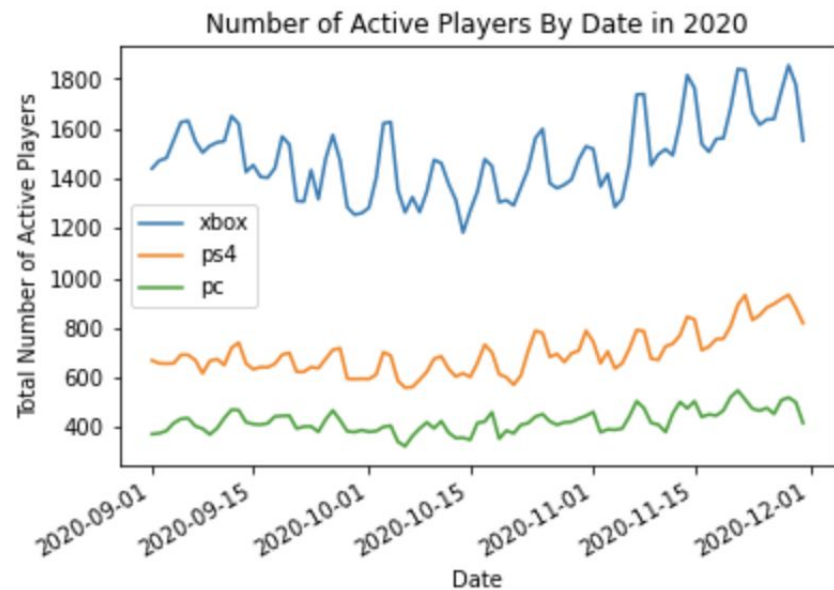
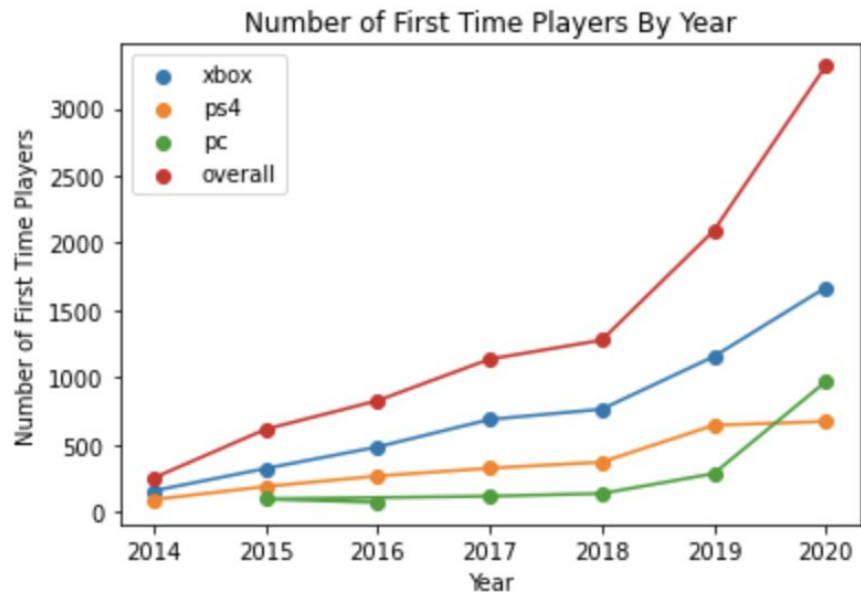
?

Let's Talk Money....



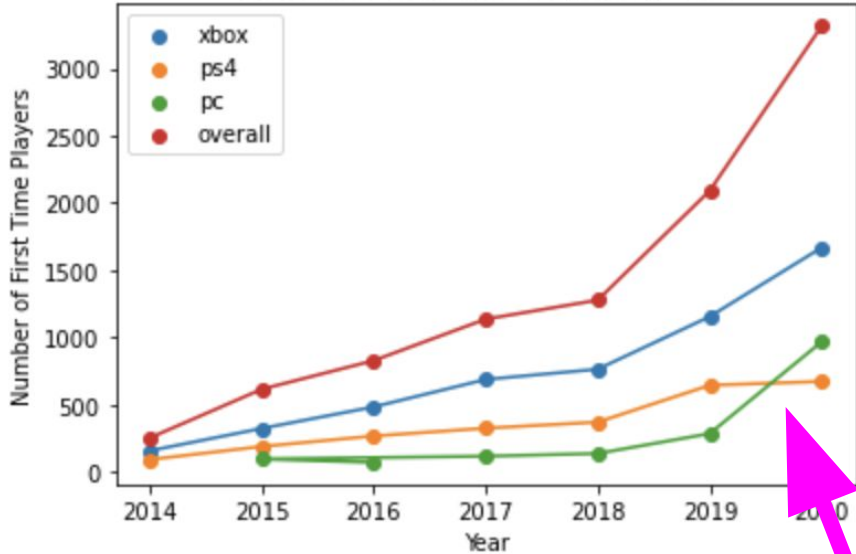
?

Growth At Platform Overtime? + 2020 Trends?

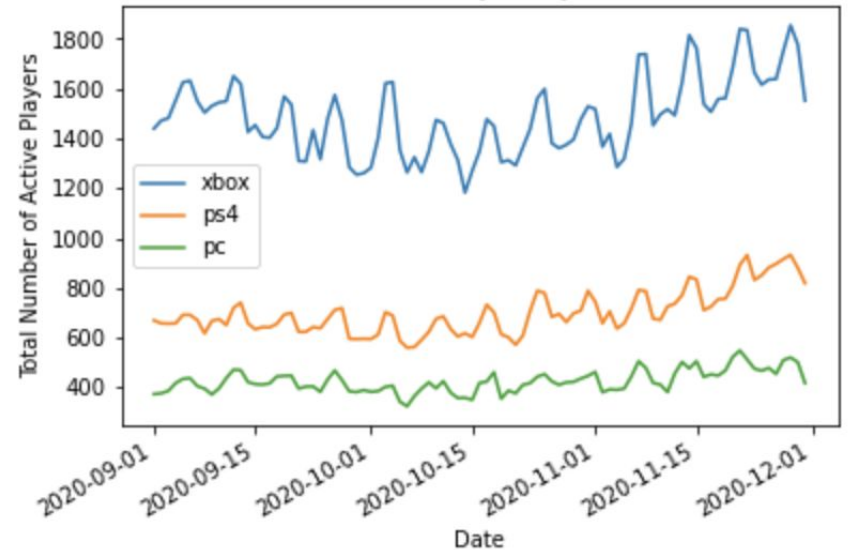


Growth At Platform Overtime? + 2020 Trends?

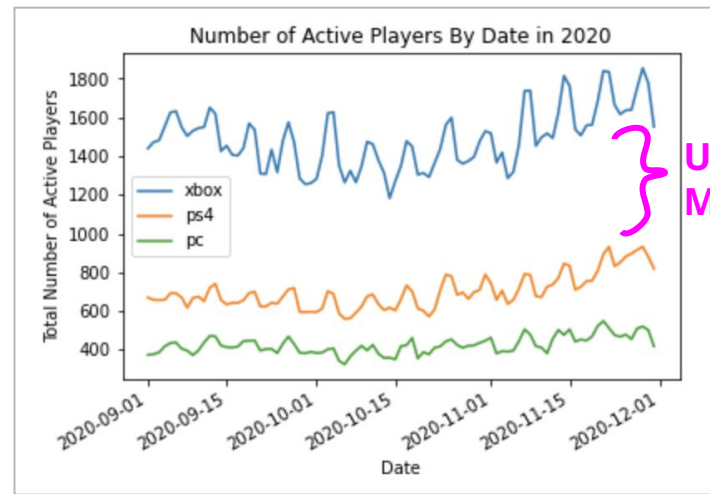
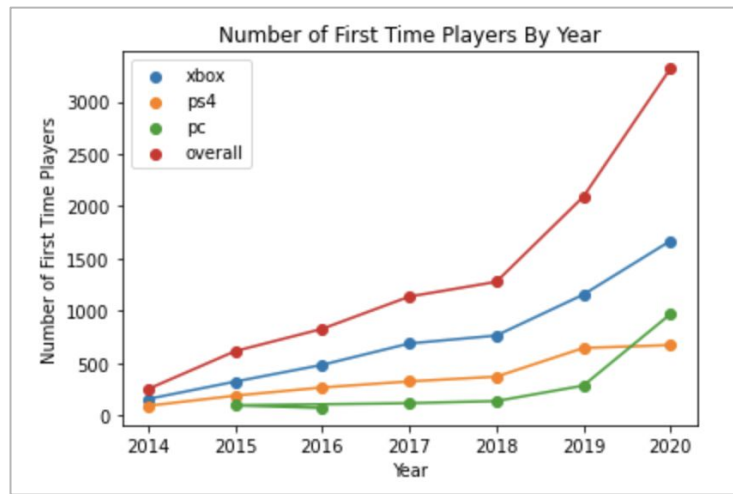
Number of First Time Players By Year



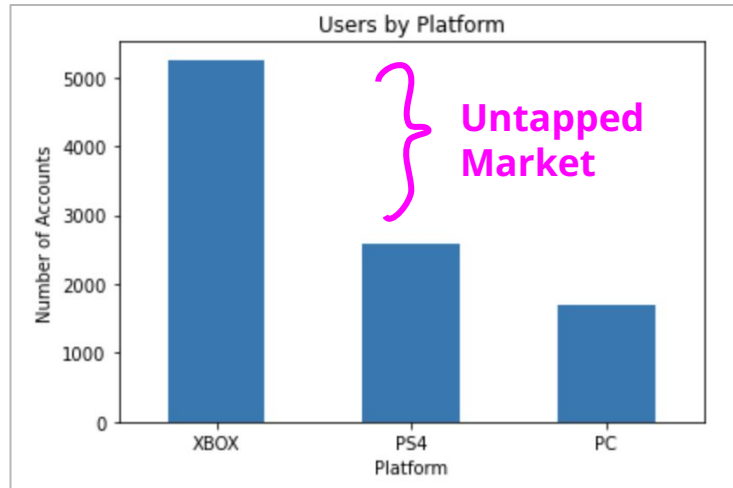
Number of Active Players By Date in 2020



AHHHHH! What's Going On Here???



Untapped Market



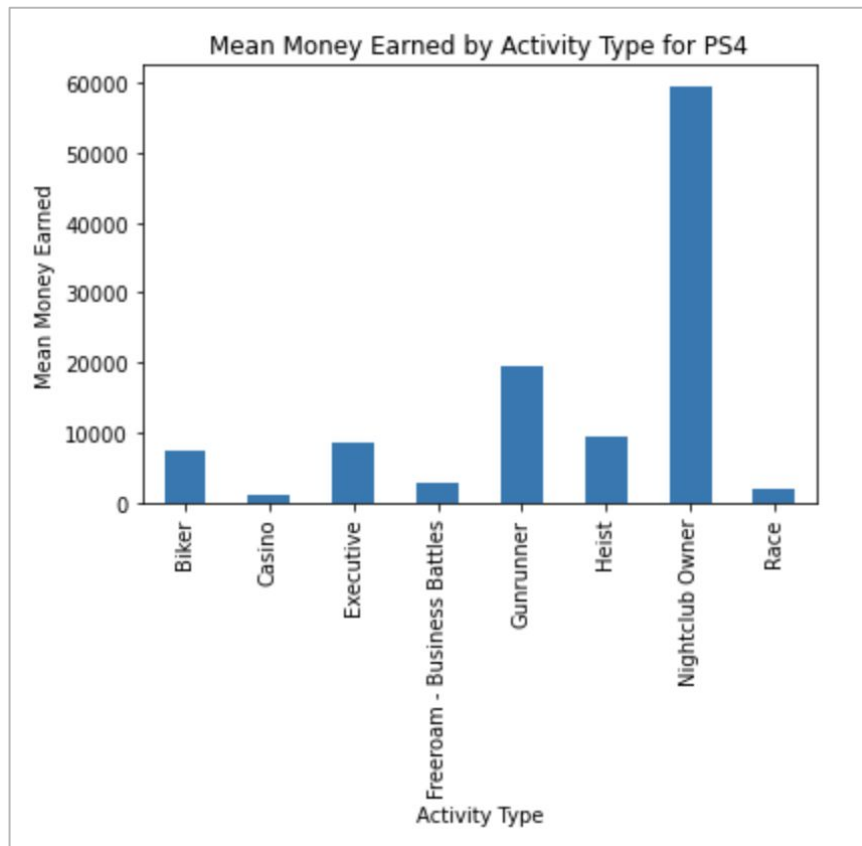
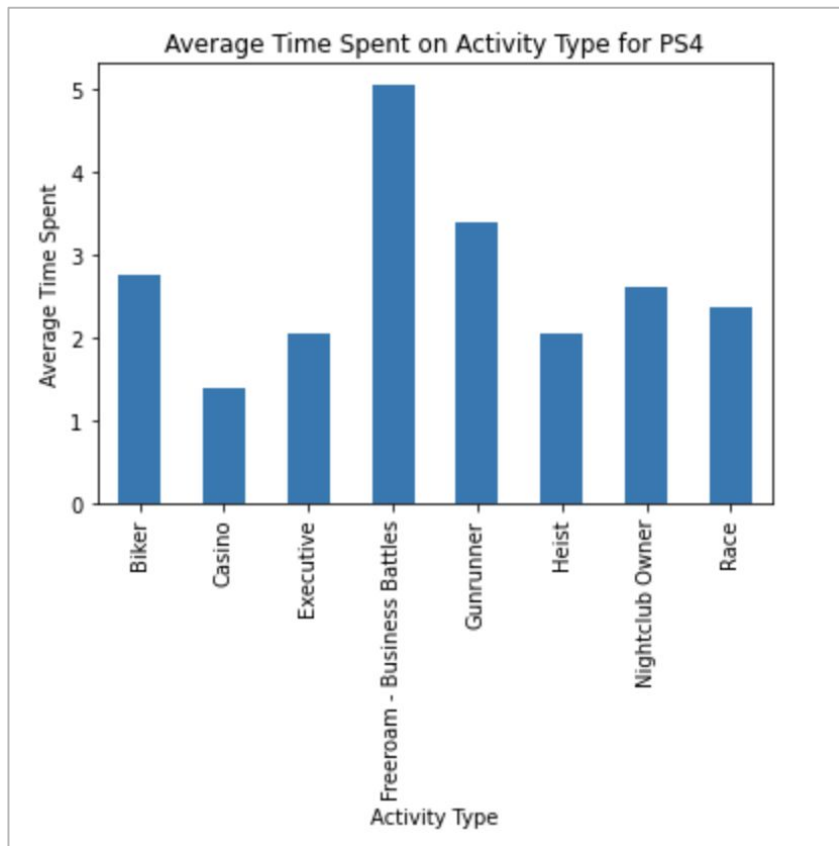
Untapped Market



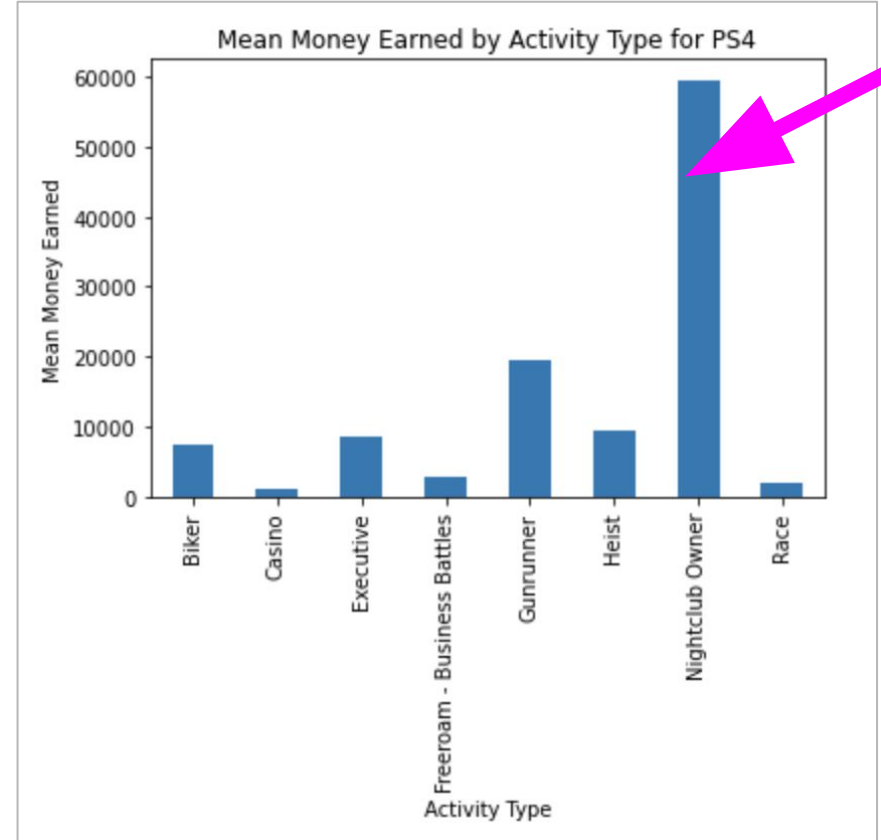
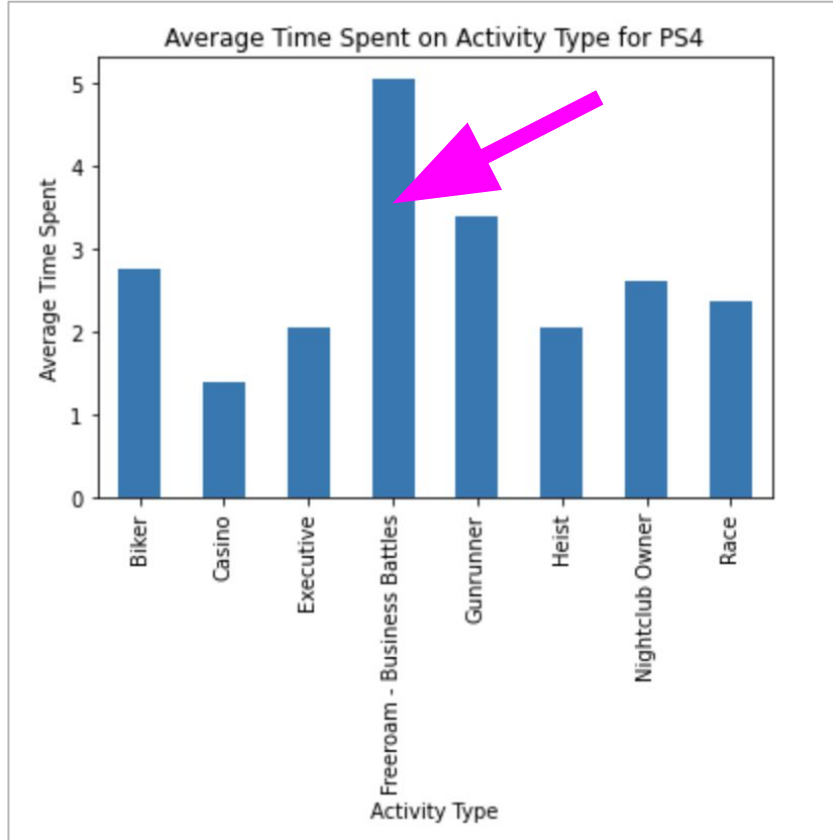
PS4

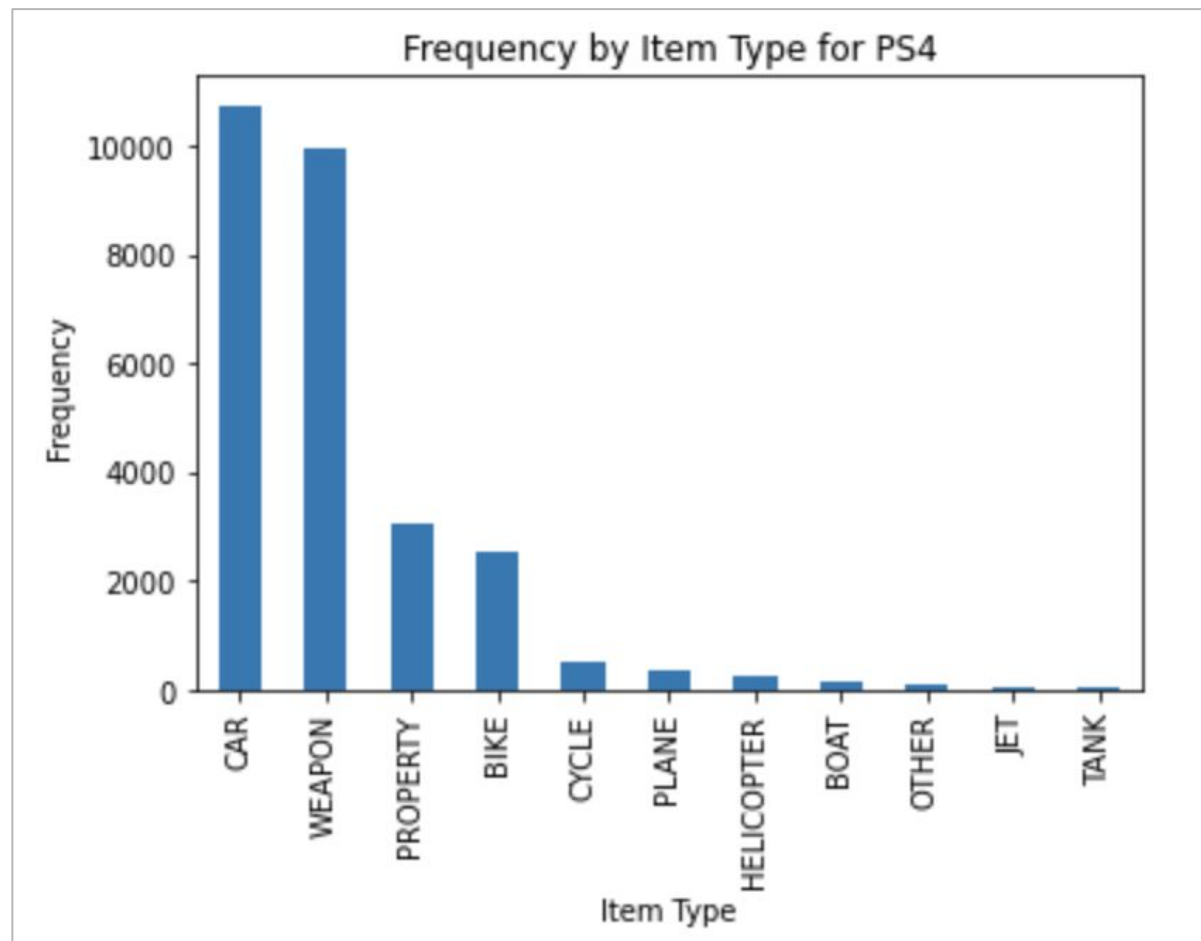
+Median daily playtime is highest
+PVC is highest

NOW WHAT? WHAT NEXT? HOW TO COLLABORATE WITH PS4?

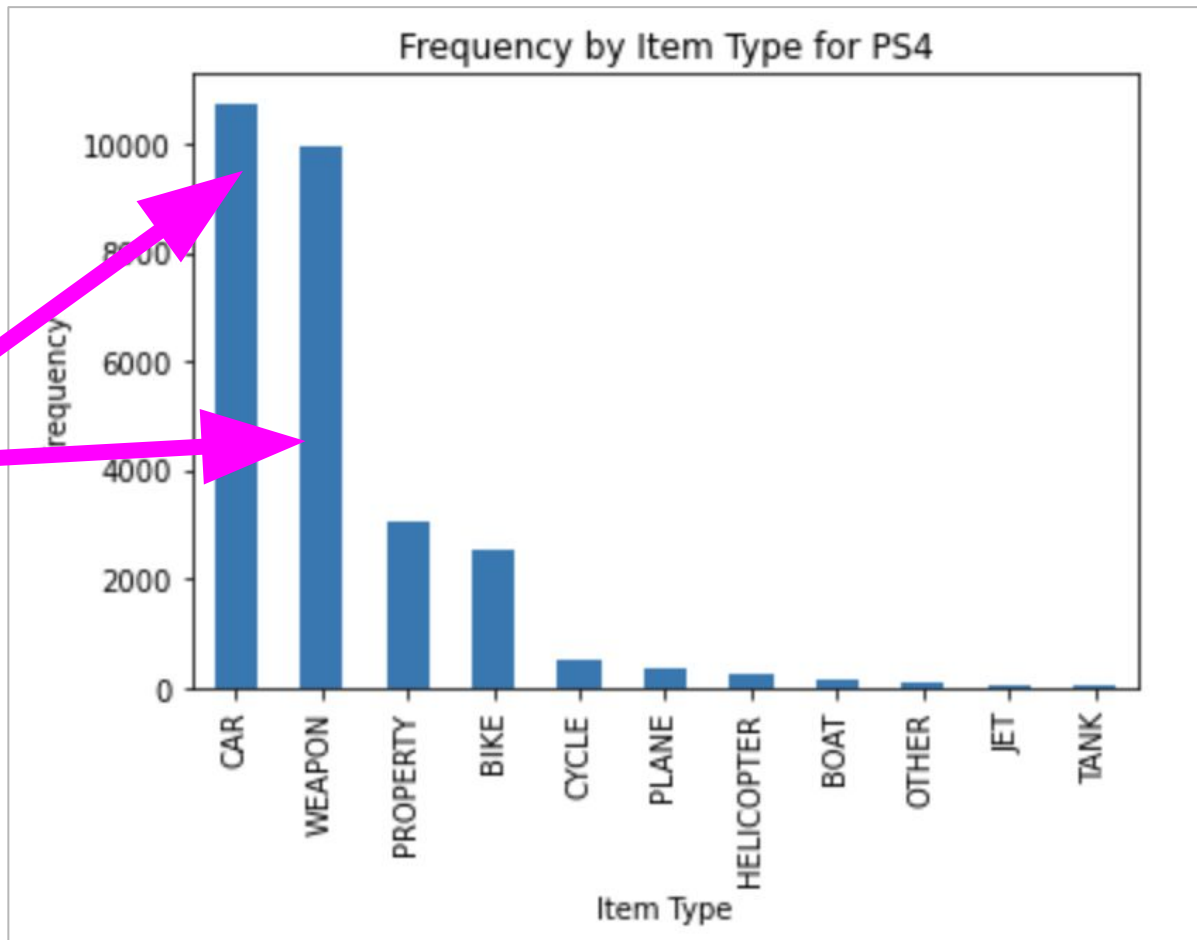


NOW WHAT? WHAT NEXT? HOW TO COLLABORATE WITH PS4?

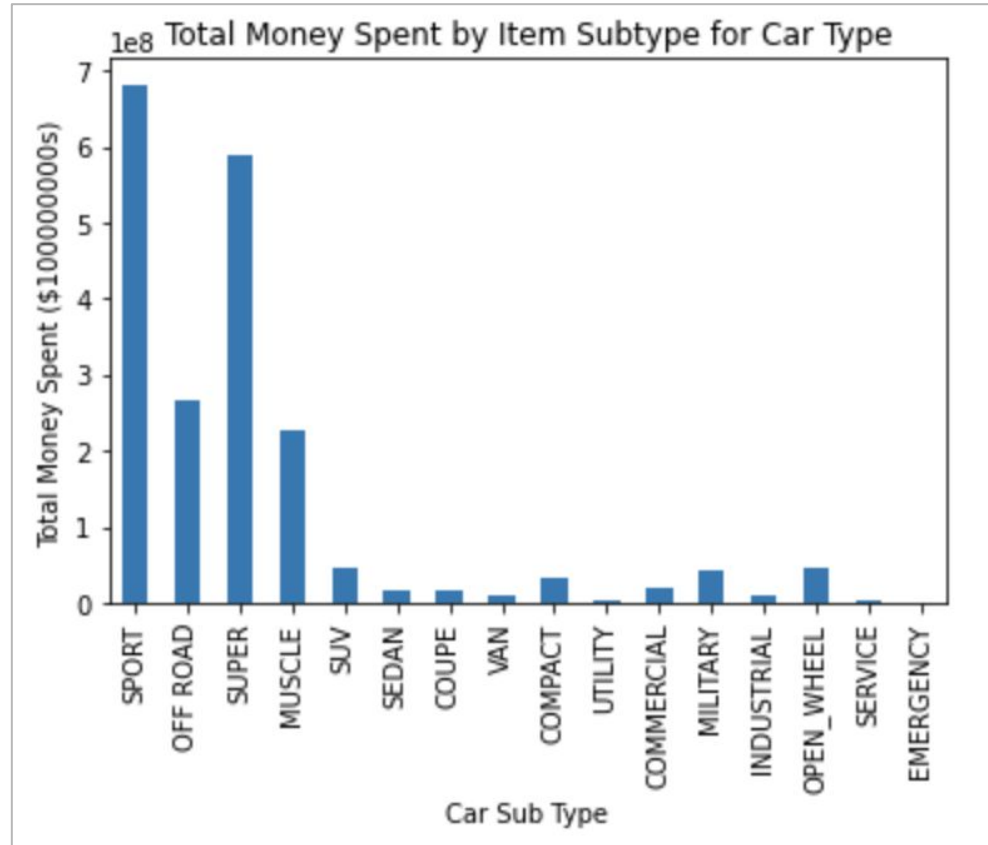




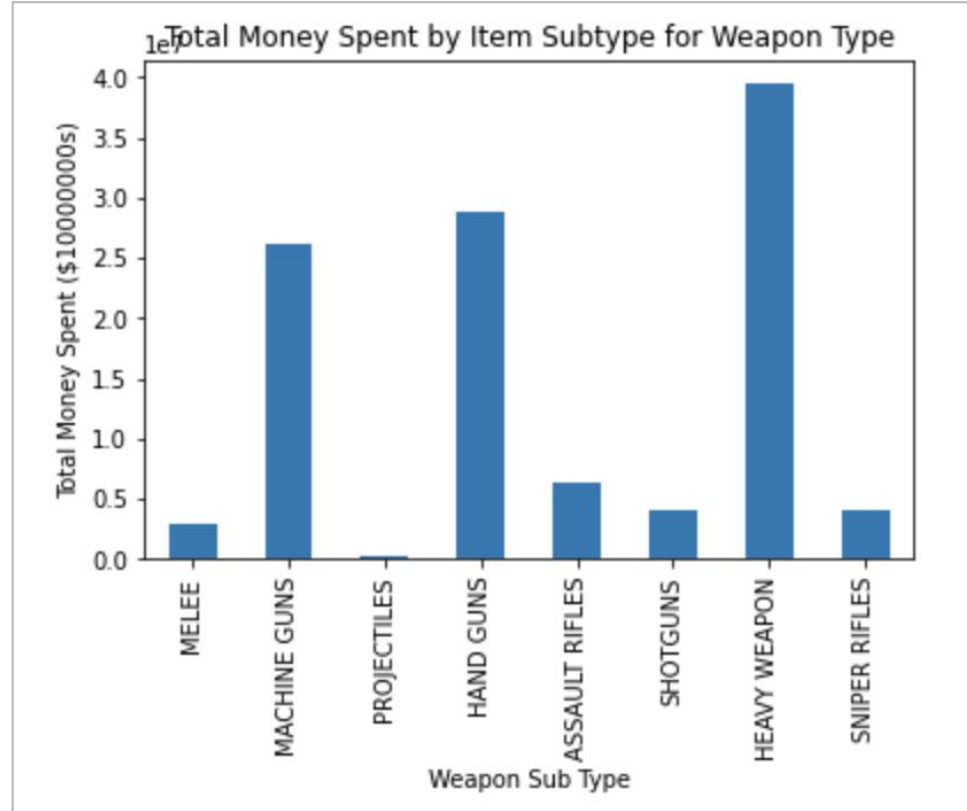
Let's
take a
closer
look



Total Money Spent Car Subtypes?



Total Money Spent on Weapon Subtypes?



INSIGHTS FOR Q1

How to increase video game sales?

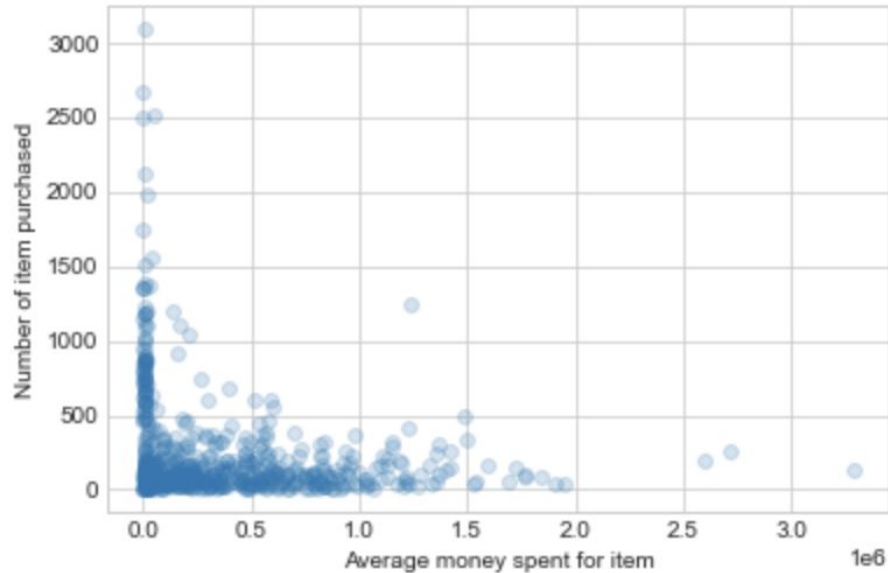
- Through which platform? PS4
- What features can be enhanced/upgraded?
 - Activities: Freeroam Business Battles (most players participate in), Nightclub Owners (player earns most money → increased satisfaction → retention)
 - Item: Sports Car, Super Car, Heavy Weapons, HandGuns and Machine Guns (most money spent on these → make expensive/rare versions of these items → incentivize people to buy Shark Cards)
- Some ethical concerns? Only focusing on PS4 (social class bias)
 - Accused of leaving out those who play GTA for free on PC
 - Stage 1: Focus on PS4, Stage 2: Move to PC



Predict player retention and money spent



Relationship between item price and number purchased

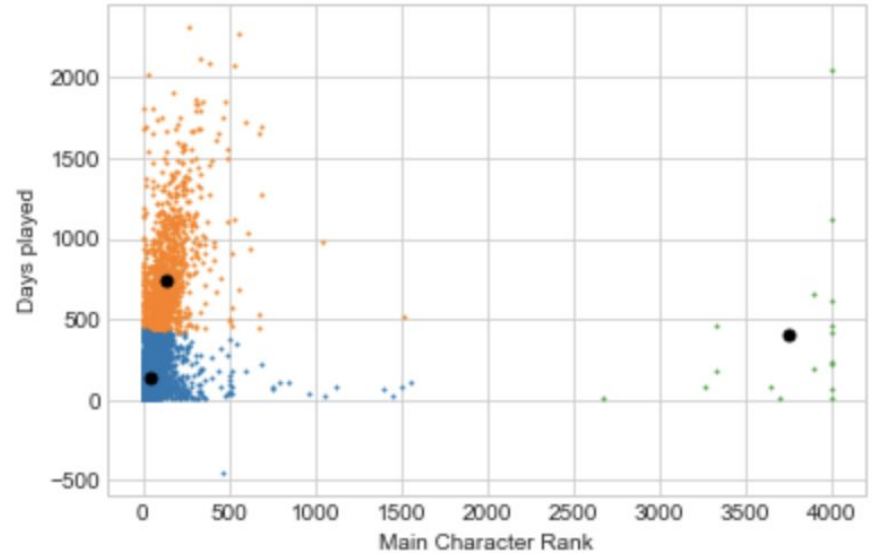
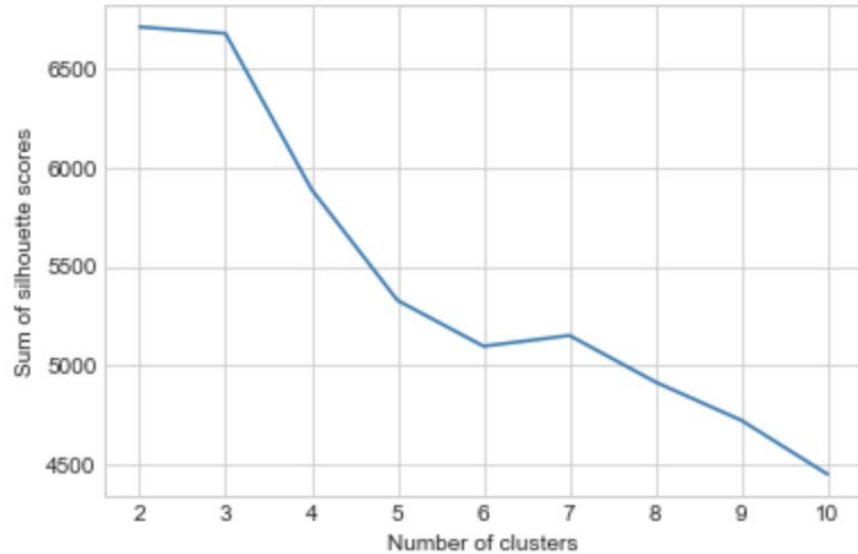


There were 1343 players who spend real money out of the 9527 - 14% (know using pvc_balance)

Currently only know about money spent to buy categorical items -> need more data on money spent on other features

Clusters of total days played and rank - k-means

Use days played and rank to predict daily time?



Rank as psychological factors (by group of ranks and does not matter after a certain point) -> need other incentives when at high rank (because this game has been around for a long time)

More societal structure in the game - high rank players have priorities and abilities to manage/train younger players if they choose to

Psychological factors that affect time spent

Predictors: average_success_rate, rp_earned, money_earned, time spent

Dependent factor: total time spent

Prediction: reinforcement and psychological value to motivate player and influence player retention rate

Multiple regression and split 80-20 train test: COD only 0.365

Favourite activity types

Nightclub Owner	18
Gunrunner	140
Casino	179
Freeroam - Business Battles	300
Race	840
Executive	1276
Biker	1663
Heist	4873

time_spent	
activity_type	
Casino	1.425201
Heist	2.002015
Executive	2.040761
Race	2.329375
Nightclub Owner	2.566882
Biker	2.702272
Gunrunner	3.201218
Freeroam - Business Battles	5.016076

money_earned	
activity_type	
Casino	1002.520258
Race	1959.233109
Freeroam - Business Battles	3326.435780
Biker	7301.648303
Heist	9105.173992
Executive	9414.413216
Gunrunner	27872.097889
Nightclub Owner	63068.974205

Interaction between players

GTA online vs. GTA: interaction vs storyline/mission

Favourite activity type is heist: thrilling and a lot of interactions

Need more interactions and mission with friends

Add competitions?

Research

Question :

- Which specific segment where the total time-played is the lowest ?
- On average, how much time are they spending playing this game mode ?
- What is the correlation of success rate, time-spent to money-earned and rp-earned ?

Process :

- Group the data by ['activity_type'] = 'Nightclub Owner'
- Process data
- Split 20-80 to train and test data, perform a linear regression analysis, $R^2 = 0.9687$

Analysis

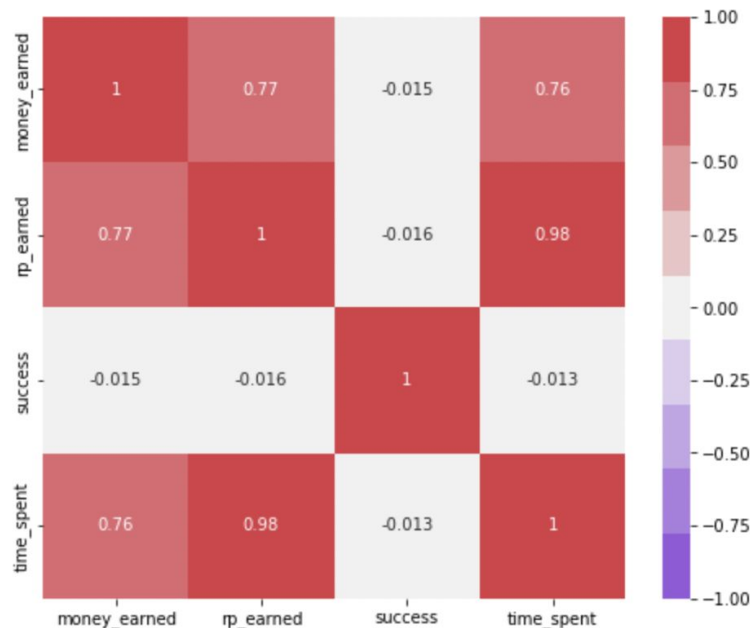
Counts of all careers

Heist	473976
Executive	126278
Biker	96970
Race	53395
Gunrunner	36540
Freeroam - Business Battles	28338
Casino	21453
Nightclub Owner	8668

Name: activity_type, dtype: int64

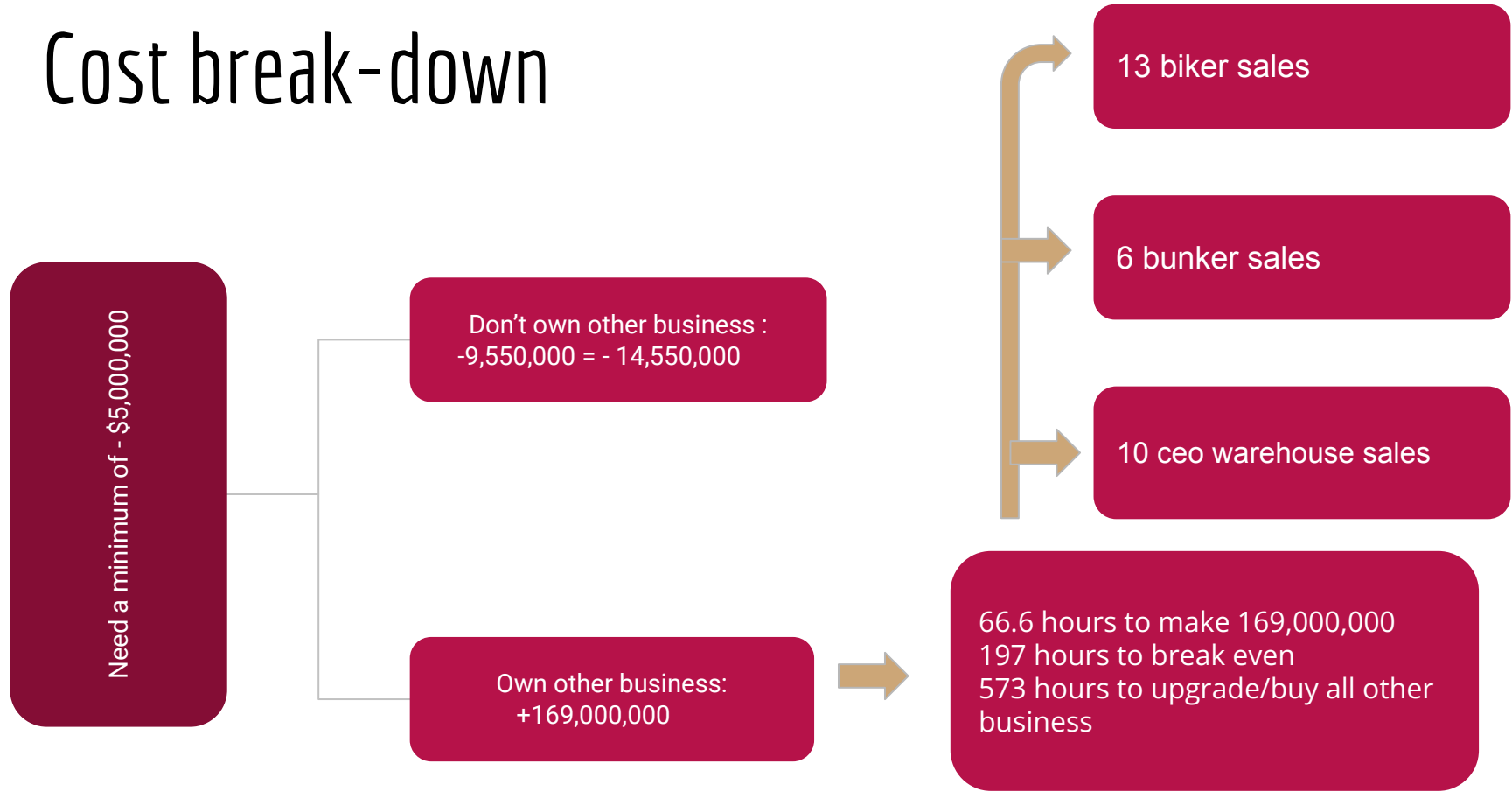
count	8668.000000
mean	2.566882
std	1.020809
min	0.010808
25%	1.853634
50%	2.480519
75%	3.183063
max	6.490422

Description of time spent
playing as a NightClub
Owner



Association Between the time spent vs the success rate vs
the money earned vs the rp earned

Cost break-down



Interpretation

While being a NightClub Owner does not necessary takes the amount of times presented, it is still least played by gamers on GTAO. There are multiple way to improve this, either by :

- + Increasing **incentives**
- + Promoting special events /**exclusivity** to the gamers
- + Introduce new business that creates **thrills** in the NightClubs.

-> Target hidden ad : scattering poster at random locations throughout the games to players can pick up on that.

-> Increase income-wise of being a Nightclub Owner

-> Collaboration with concert-event companies

Dataset improvement

- Total real money spent instead of only balance

The money **balance** does not reflect the amount of money the player put into the game -> need to record **transactions** in games .

- Coherence between time spent and daily spent in the two tables
- More details about types of heist - currently all combined into one, maybe that why it has the most frequency?
- Need to organize the activity-type/dataset in general according to each rank level.
- Add demographic of players

THANK YOU FOR LISTENING