COMBAT ELITE

DATA ANALYSIS REPORT

Sundas Taimoor

7/15/24

Graduate Data Analyst Challenge

Contents

COMBAT ELITE

INTRODUCTION

Combat Elite has 4-player matches which can be against other players and/or bots. There are 3 match types that are available in turns in a cycle to the player.

Combat Elite has products which can be bought with real money, known as in-app purchases (IAPs), as well as products that can be bought with the in-game virtual currency, Gold.

DATASET DESCRIPTION

1. data_daily_activity.csv containing the daily activity of the users. Each row represents a user who opened the game on that specific day. This data set has the following structure:

- activity_date - The calendar date when the user was active
- acquisition_date - The calendar date when the user was active in the game for the first time
- platform - The users' reported platform on each calendar date
- user_id - The unique user ID of each active user

2. data_matches.csv contains all the matches played during the analysis period per user. Each row represents a day of matches. This data set has the following structure:

- activity_date - The calendar date when the user was active
- user_id - The unique user ID of each active user
- match_type - The type of match that is played
- finish_position - The finish position of the players (1-4)
- bots - The number of bots in the match (0-3)
- n_matches - The total number of matches per user by calendar date, match type and
- finish position

3. data_in_app_purchases.csv contains all the in-app purchases made during the analysis period per user. Each row represents a transaction. This data set has the following structure:

- activity_date - The calendar date when the user was active
- user_id - The unique user ID of each active user
- purchase_number - The number of the purchase made by the specific user_id
- product_group - The group of the product purchased
- dollar_purchase_value - the revenue that purchased generated in dollars

4. data_virtual_purchases.csv contains all the purchases made using gold during the analysis period per user. Each row represents a day of transactions. This data set has

the following structure:

- activity_date - The calendar date when the user was active
- user_id - The unique user ID of each active user
- product_group - The group of the product purchased
- n_purchaes - The number of the purchase made per user by calendar date,
- product and product_group
- gold_spend - the total gold spent per user by calendar date, product and
- product_group

EXTRACT, LOAD AND TRASNFORM USING GOOGLE COLAB

The extraction process was simplified to downloading and unzipping the folder. The CSV files were loaded into 4 separate data frames in Google Colab and data was transformed after going through the following steps:

1. Dropping unnecessary columns. In this case the column 'unnamed' was dropped.
2. Datasets were check for structure, data types, and summary statistics for each data frame
3. Handling or dropping missing values

EXPLORATORY DATA ANALYSIS

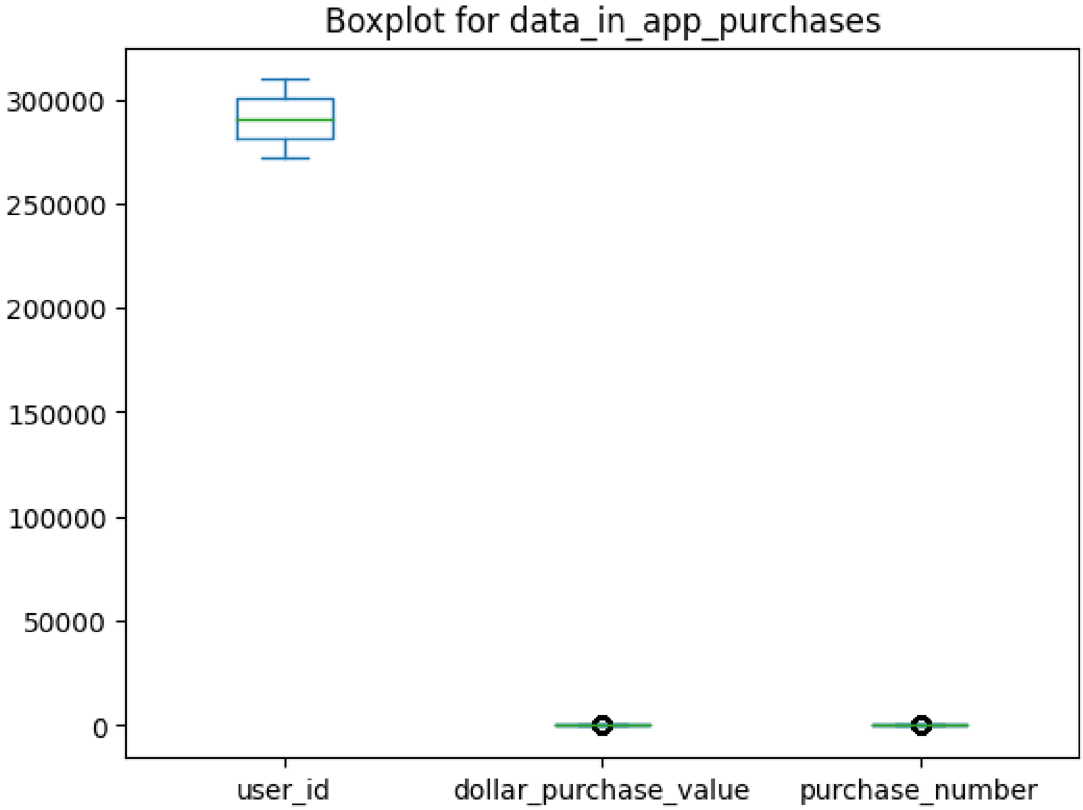
1. Dataframe *data_daily_activity*
- Analysing data_daily_activity yielded that there were no missing values.
- u*ser_id* happened to be in *float* and was changed into *int.*
- activity_date was changed to *date* from *object*
- *platform* was standardised

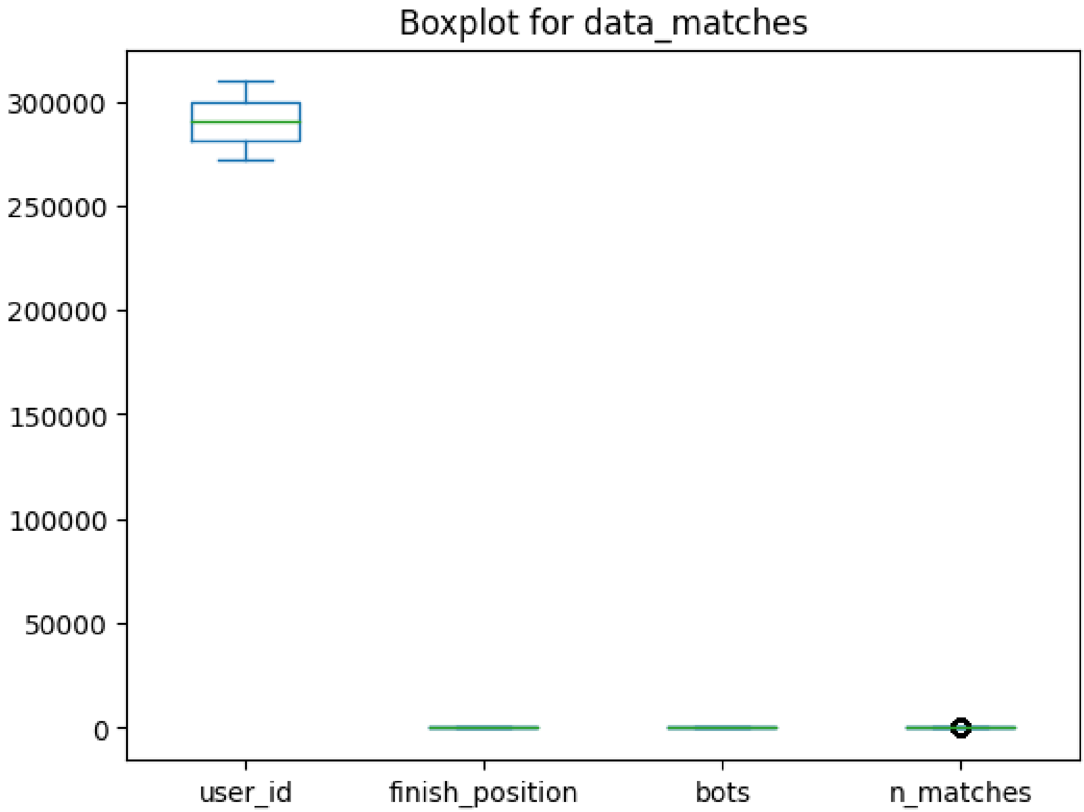
2. Dataframe data_in_app_purchases
- All entries were non-null hence no missing values.
- *activity_date* was changed to date from object
- 48 missing values were replaced with 0 as imputing them with Mean, Median or mode would change the meaning of data

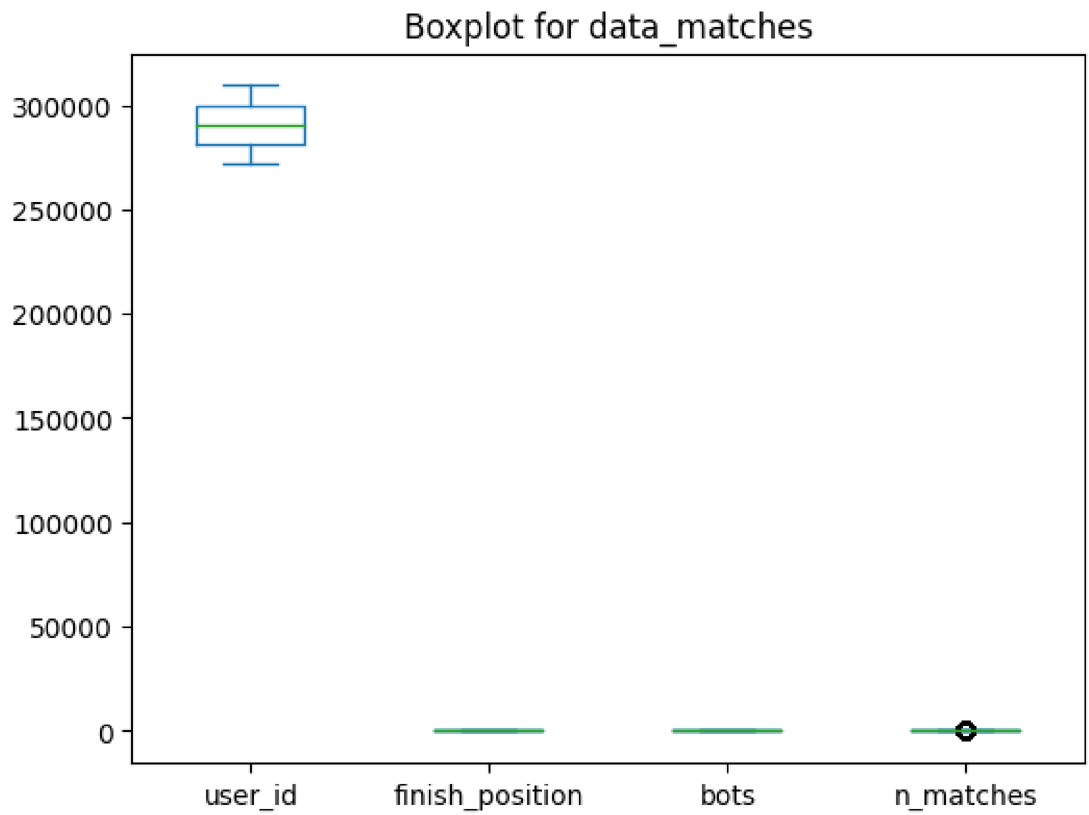
3. Dataframe data_matches
    - u*ser_id* happened to be in *float* and was changed into *int.*
    - activity_date was changed to *date* from *object*
    - finish_position had records with 'NA' which were imputed by the Median and data type was changed from object to int64.

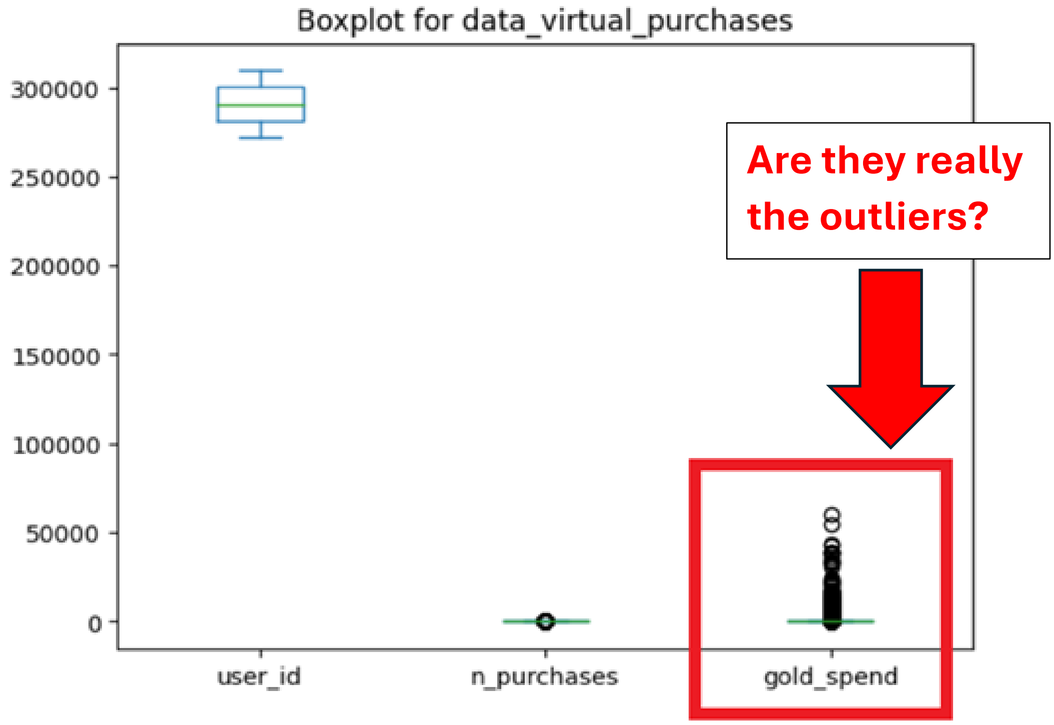
4. Dataframe data_virtual_purchases
    - activity_date was changed to *date* from *object*

BOX PLOT FOR OUTLIERS



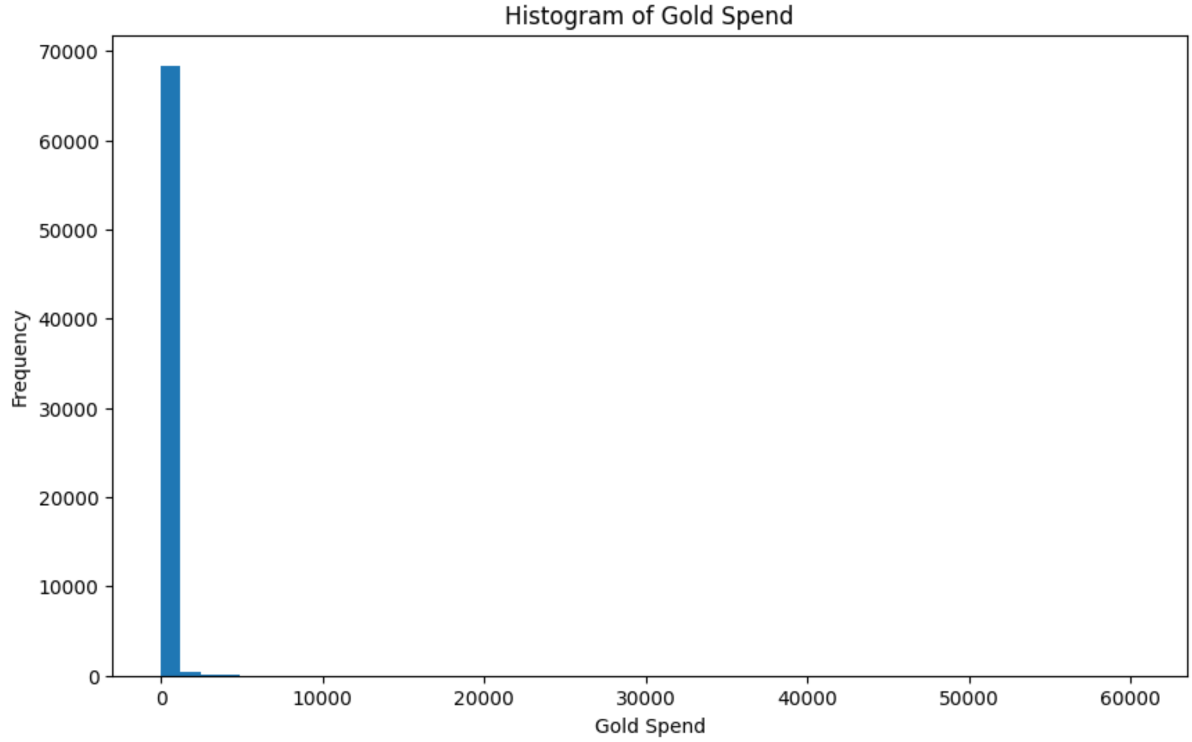


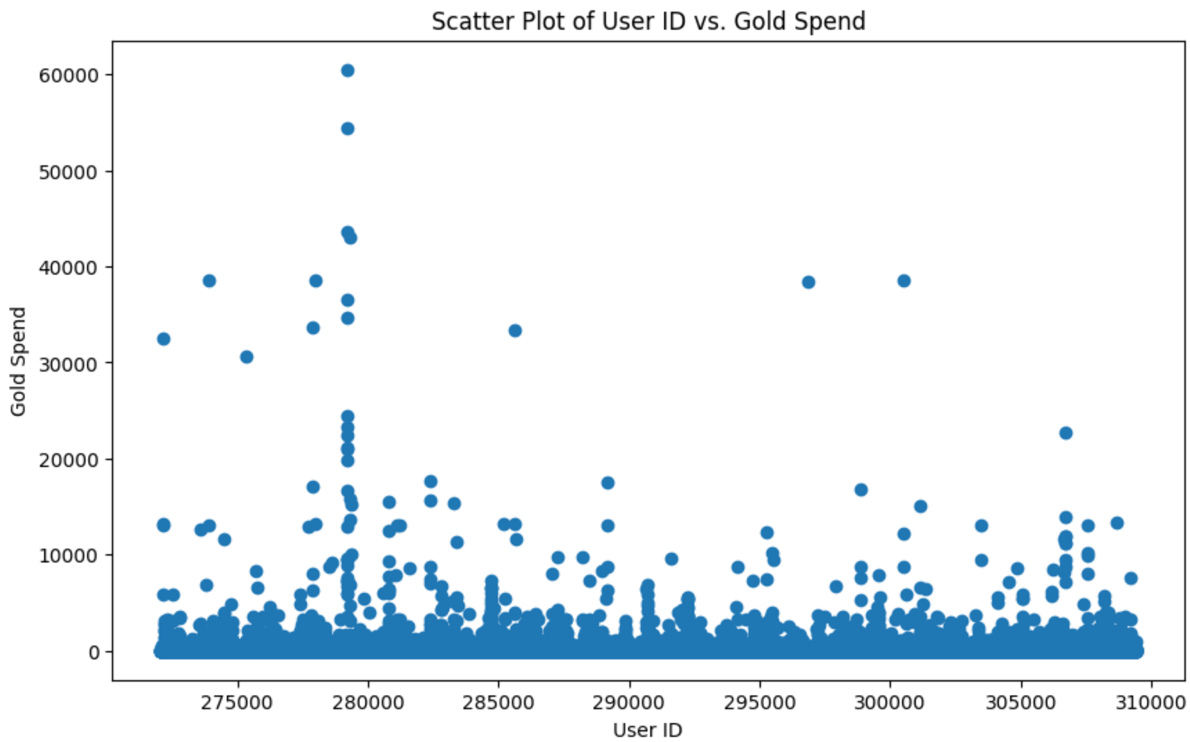

Boxplot for data_matches
300000
250000
200000
150000
100000
50000
0
user_id
finish_position
bots
n_matches

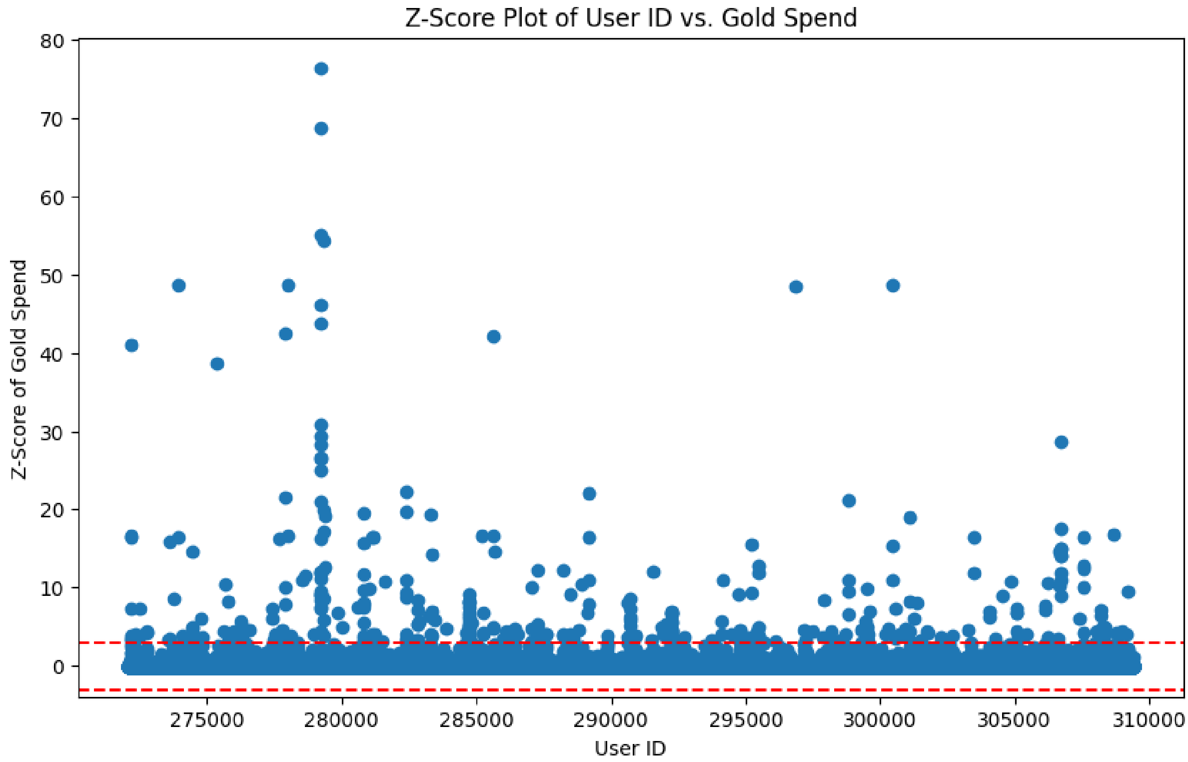

Boxplot for data_virtual_purchases
300000
250000
200000
150000
100000
50000
0
user_id
n_purchases
gold_spend
Are they really the outliers?

CONTEXT ANALYSIS ON golf_spend FROM data_vritual purchases

Let’s analyse gold_spend with the help of visualisations and Z-Score Plot.







Most users have Z-scores below 10, suggesting that most users' spending is within a reasonable range of the mean.

There are several notable outliers, particularly users with Z-scores exceeding 20, 40, and even 70. These users are spending significantly more gold compared to the average user.

ACTIONABLE INSIGHT

Users with extremely high Z-scores (outliers) could be particularly valuable for targeted marketing or special attention, as their spending habits are significantly different from the average user.

DATA MERGE USING HADOOP HIVE

Through SQL Union operation and relevant columns were chosen to merge the data sets. Here user_id acted as the primary key (PK).

I used Hive SQL for this purpose. Data was merged to have better insights,

HYPOTHESIS TESTING

Hypothesis1: Impact of match winning on purchases

H1: Users make a purchase after winning a match (finishing with finish_position of '1')

H0: No significant relation in purchases behaviour after winning.

Data Filtering: Identified days where users won matches (finish_position = 1).

Data Merging: Merged the winners' data with purchase data to find days when winners made purchases.

Proportion Calculation:

Calculated the proportion of winning days that lead to a purchase.

Calculated the overall proportion of match days that lead to a purchase.

Statistical Testing:

Conducted a z-test to compare the proportions of purchase days between winners and all players.

Description	**Value**
Total days where matches were played	177,117
Days with purchases (any day)	893
Proportion of any day leading to purchase	0.50%
Total winning days	125,819
Winning days with purchases	747
Proportion of winning day leading to purchase	0.59%
Z-Statistic	3.31
P-Value	0.00094

The p-value of 0.00094 from the z-test indicates a statistically significant difference between the proportions, suggesting that winning increases the likelihood of making a purchase.

ACTIONABLE INSIGHTS & RECOMMENDATIONS

Enhance Engagement and Rewards for Non-Purchasing Winners

Winning a match does appear to create a moment of heightened engagement, as evidenced by the slight increase in purchase probability. However, since most winners do not make a purchase, there is an opportunity to further capitalize on the positive emotions associated with winning.

Actionable Strategies:

Post-Win Engagement Boosts:

Implement a targeted communication strategy immediately after a win, such as congratulatory messages or personalized suggestions, to encourage interaction with in-game purchase options.

Provide time-sensitive offers or discounts that activate only after winning, to create a sense of urgency and exclusivity.

Hypothesis 2: Impact of Platform on Spending

Null Hypothesis (H0): There is no significant difference in the average in-app purchase spending between users on different platforms (iOS vs. Android).

Alternative Hypothesis (H1): There is a significant difference in the average in-app purchase spending between users on different platforms.

T-TEST RESULTS:

T-Statistic: -10.42

P-Value: 2.04e-25 (very close to zero)

INTERPRETATION:

T-Statistic: The large negative t-statistic suggests a significant difference between the two groups.

P-Value: The p-value is extremely low, far below the typical significance level of 0.05. Hence we can reject the null hypothesis.

ACTIONABLE INSIGHT:

Personalized Marketing Campaigns:

Android: Launch targeted advertising campaigns on Android devices that highlight the value of premium features or benefits that come with higher spending. Utilize the data on high spending to create more compelling, value-driven messages.

iOS: Focus marketing efforts on the affordability and incremental benefits of in-app purchases. Emphasize cost-effectiveness and the immediate enjoyment increase with small purchases.

User Experience Optimization:

Android: Design the user interface and in-app purchase process to seamlessly integrate with the preferences of Android users, who may appreciate a more straightforward, direct approach to accessing and using premium features.

iOS: Enhance the aesthetic and smoothness of the purchase process, as iOS users often value a polished and intuitive user interface. Simplify the transaction process to reduce friction and encourage impulse buying.

Hypothesis 3: Relationship Between Engagement and Virtual Currency Spending

Null Hypothesis (H0): There is no significant relationship between the frequency of user engagement (daily activity) and the amount of virtual currency spent.

Alternative Hypothesis (H1): There is a significant relationship between the frequency of user engagement (daily activity) and the amount of virtual currency spent.

Pearson Correlation Test:

Correlation Coefficient: 0.072

P-Value: 1.84e-25 (very close to zero)

Interpretation:

Correlation Coefficient: The value of 0.072 indicates a weak positive correlation between the frequency of user engagement (active days) and the amount of virtual currency spent. Although the relationship is positive, it is not strong.

P-Value: The p-value is extremely low, far below the typical significance level of 0.05.

Conclusion:

Since the p-value is much less than 0.05, we reject the null hypothesis (H0). There is a statistically significant relationship between the frequency of user engagement (daily activity) and the amount of virtual currency spent. However, the weak correlation coefficient suggests that while the relationship is statistically significant, it is not practically strong.

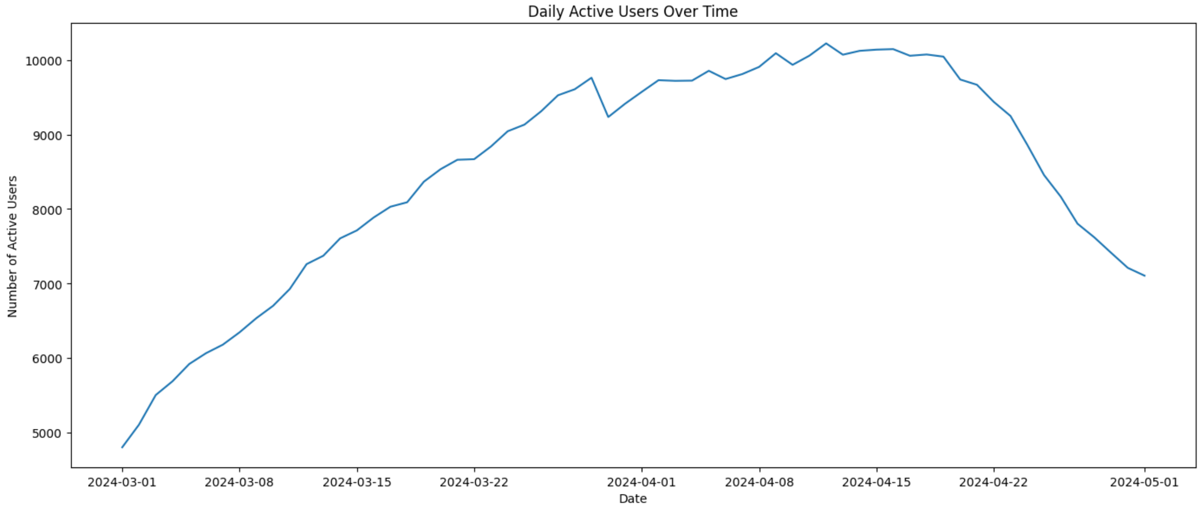
ACTIONABLE INSIGHTS:

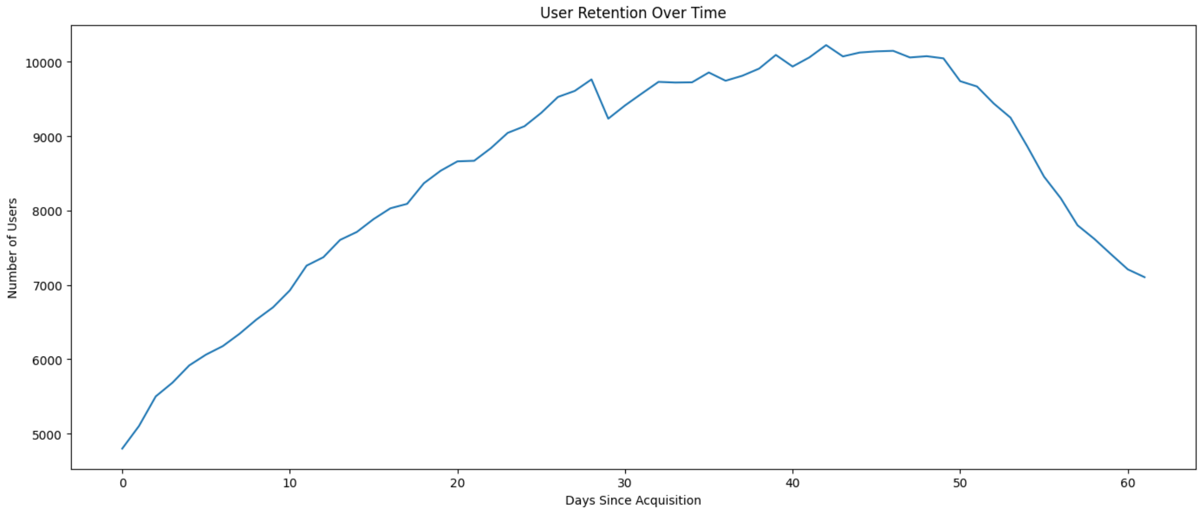
Enhance User Engagement:

Although the correlation is weak, the positive relationship suggests that increasing user engagement could still be beneficial. Focus on strategies that keep users coming back daily or engaging more deeply during each session.

Implement features like daily challenges, rewards for consecutive logins, or special events that encourage regular participation.

USER ENGAGEMENT ANALYSIS ON **data_daily_activity**





1. Daily Active Users (DAU):

The DAU trend shows the overall activity patterns over time. The graph would indicate if there are any specific spikes or drops, which could correspond to game updates, promotions, or external factors.

2. User Retention:

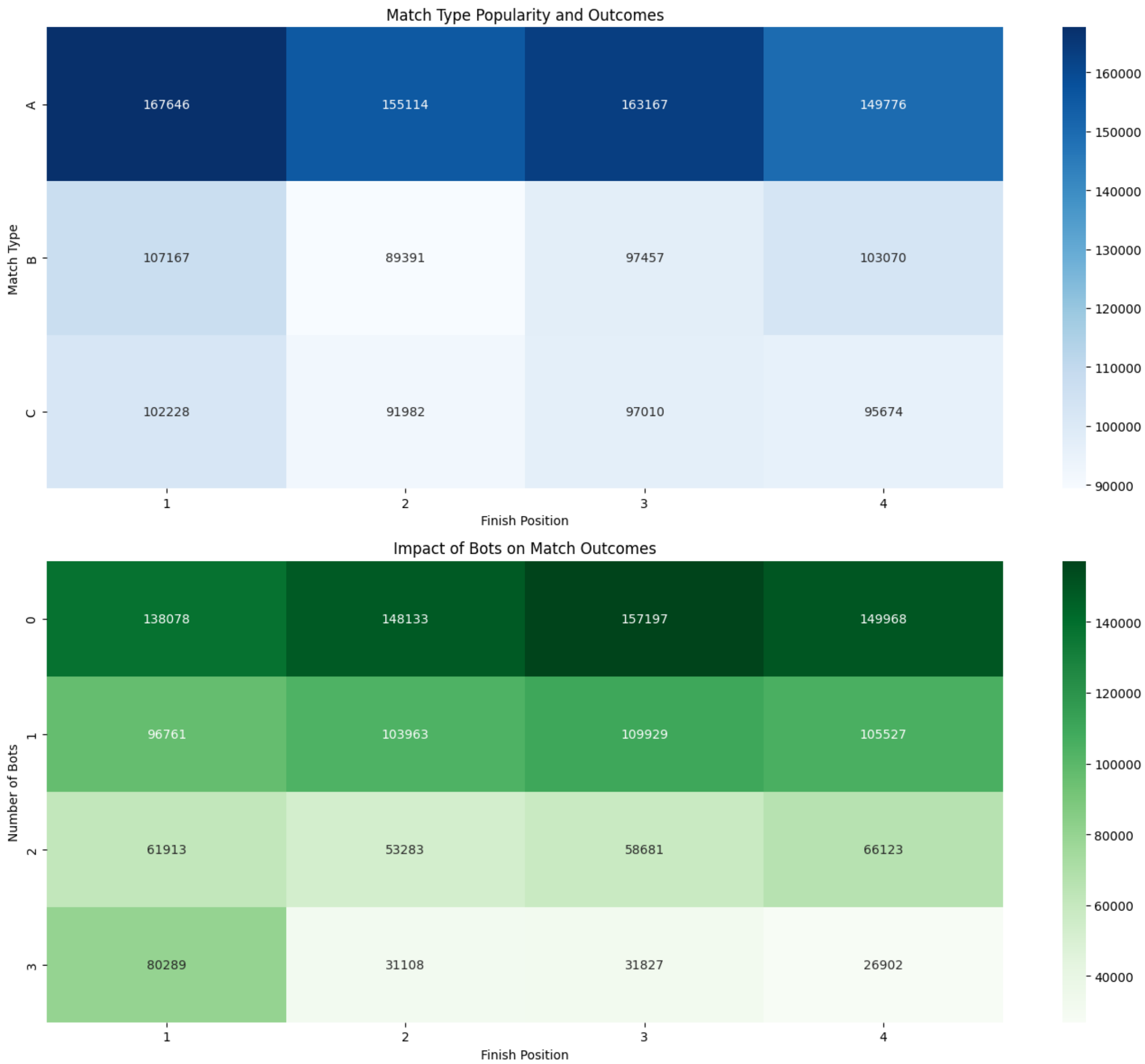
The retention curve helps to understand how well the game retains users over time after their first session. A typical retention curve should ideally show a steep drop initially, followed by a gradual flattening as loyal users continue to engage with the game.

ACTIONABLE INSIGHTS

Optimize for Android: Given the significant user base on Android, ensuring the game's performance and user experience on Android devices could be prioritized.

Investigate Retention Strategies: To improve retention, consider introducing more engaging content early in the game or implementing rewards that incentivize daily logins.

MATCH ANALYSIS ON data_matches



MATCH ANALYSIS INSIGHTS

1. Match Type Popularity and Outcomes:

The heatmap indicates the distribution of finish positions across different match types. This visualization helps identify which match types are more popular and how often players are achieving top positions in each type.

2. Impact of Bots on Match Outcomes:

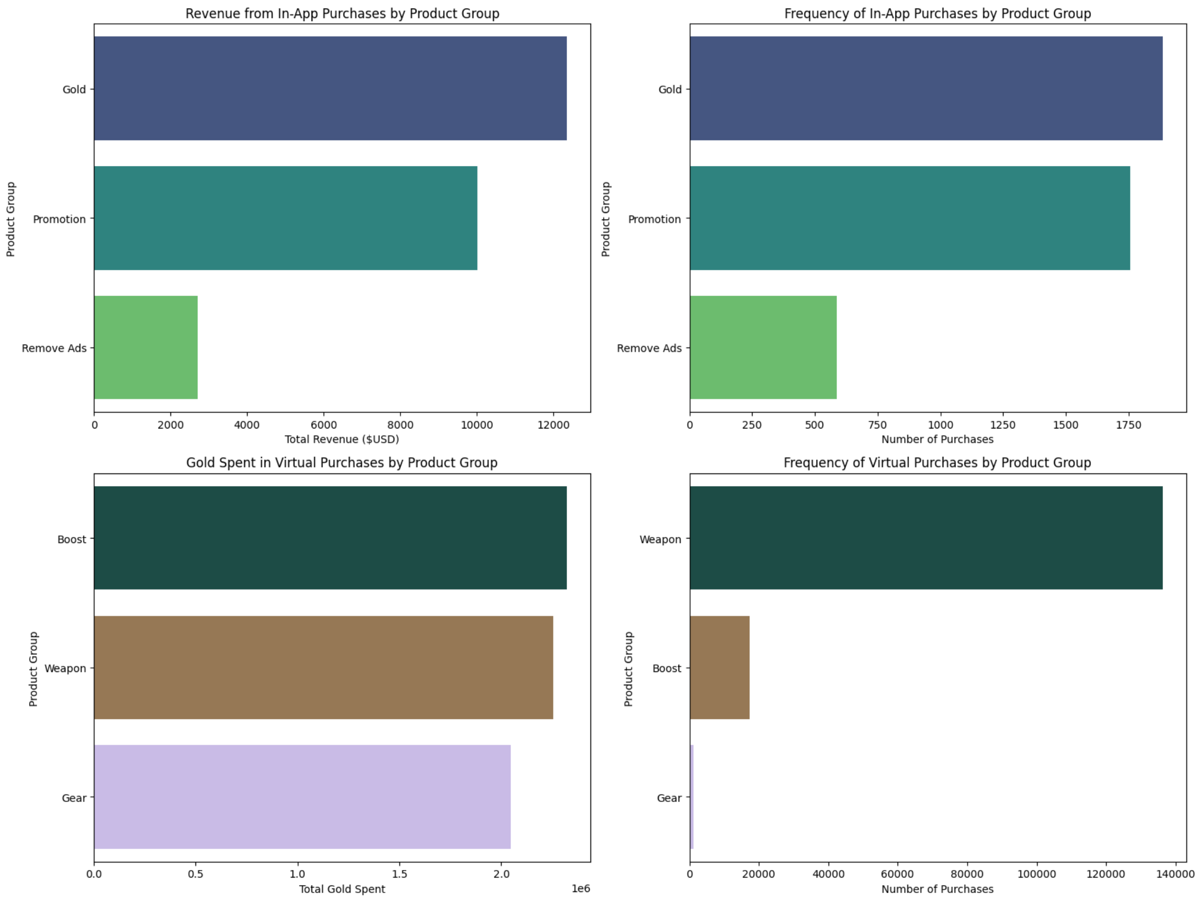
The presence of bots seems to have a distinguishable impact on the game outcomes. The heatmap provides insights into how different numbers of bots in matches affect player finish positions, potentially indicating whether the difficulty is perceived as fair or not by the players.

ACTIONABLE INSIGHTS

Match Type Balance: If certain match types show significantly different levels of engagement or success rates, consider rebalancing the game mechanics or providing tutorials to help players perform better in less popular types.

Bot Difficulty Adjustment: Analyze player feedback and match outcome data to adjust the bots' difficulty levels. Ensure that the game remains challenging yet fair, particularly in matches where player finish positions are consistently lower with higher numbers of bots.

PURCHASE ANALYSES ON IN-APP PURCHASES AND VIRTUAL PURCHASES



In-App Purchases (IAP):

Revenue Distribution: Certain product groups dominate in terms of revenue generation. This insight helps identify which items or features are most valuable to users, potentially guiding development focus or promotional efforts.

Purchase Frequency: The frequency of purchases by product group also shows which items are more popular, irrespective of the revenue they generate. This can help in understanding user preferences and could be used to design bundles or discounts to increase purchase rates.

Virtual Purchases (Using Gold):

Gold Expenditure: The distribution of gold spent by product group highlights where players are most willing to invest their in-game currency. High gold spend areas might be good candidates for introducing new related features or items.

Purchase Frequency: The frequency data for virtual purchases provides insights into the items' popularity and can help prioritize which virtual items to enhance or expand upon.

ACTIONABLE INSIGHTS

Enhance High Revenue Items: For both IAP and virtual purchases, focus on enhancing the features or visibility of high revenue and high frequency items. Consider seasonal promotions or limited-time offers to boost sales further.

Balance Offering: For items with high usage but low revenue, consider adjusting pricing strategies or bundling them with other popular items to increase their value proposition.

Feedback Incorporation: Use player feedback on these popular items to guide future updates or new item introductions, ensuring they align with player expectations and demands.

PLAYER SEGMENTATION THROUGH CLUSTER ANALYSIS

	total_iap_spend		total_virtual_spend		average_finish		total_matches		
	mean	median	mean	median	mean	median	mean	median	count
cluster									
0	0.516361	0.00	151.809558	45.0	2.515815	2.500000	54.085012	9.0	18268
1	0.658673	0.00	143.625398	40.0	1.154355	1.000000	2.007289	1.0	9741
2	1240.720000	1240.72	483263.000000	483263.0	2.249845	2.249845	5409.000000	5409.0	1
3	36.292909	0.00	8956.927273	100.0	2.404750	2.490974	1848.613636	1493.5	220

Cluster 0 (General Players):

In-App Spend: Very low average spend.

Virtual Spend: Moderate virtual currency usage.

Average Finish: Around 2.5, suggesting mid-level performance.

Total Matches: Regular players with moderate engagement.

Size: Largest group with 18,267 players.

Cluster 1 (Casual Players):

In-App Spend: Very low.

Virtual Spend: Low.

Average Finish: Typically high (1.0), indicating they often win or only play a few matches.

Total Matches: Very low, indicating these are casual or new players.

Size: Second largest group with 9,740 players.

Cluster 2 (The Whale):

In-App Spend: Extremely high.

Virtual Spend: Exceptionally high.

Average Finish: Around 2.25, decent performance.

Total Matches: Extremely active in the game.

Size: Only one player in this cluster, suggesting an outlier with heavy spending and engagement.

Cluster 3 (Highly Engaged and Spending Players):

In-App Spend: Moderate real money spending.

Virtual Spend: High usage of virtual currency.

Average Finish: Slightly below average performance.

Total Matches: Very high, indicating highly active players.

Size: Small, focused group of 222 players.

ACTIONABLE INISIGHTS

Recommendations Based on Player Segments

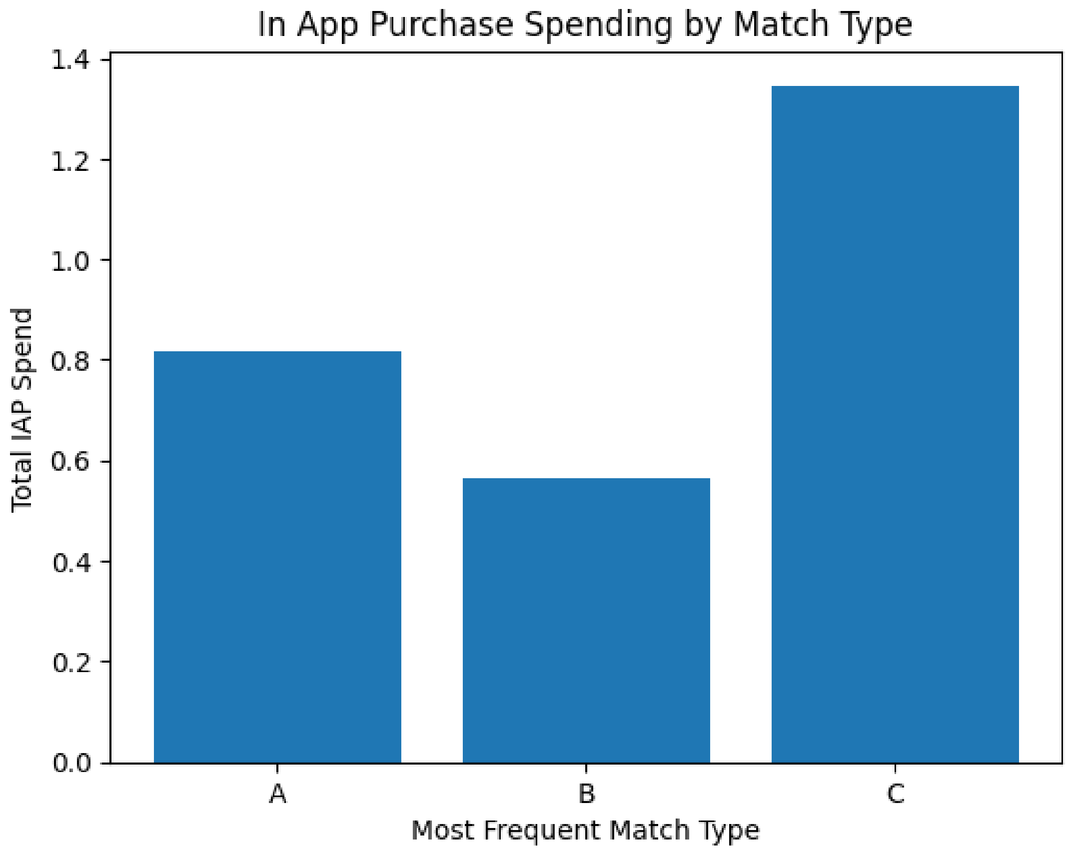
Cluster 0 (General Players): Focus on improving engagement by introducing new challenges or daily missions to encourage both spending and regular play.

Cluster 1 (Casual Players): Implement tutorials or welcome bonuses to help these players integrate into the game, potentially converting them to more regular users.

Cluster 2 (The Whale): Personalize offers and support to maintain and possibly enhance this player's engagement and satisfaction.

Cluster 3 (Highly Engaged and Spending Players): Offer loyalty rewards and exclusive content to retain these valuable players and encourage them to continue their engagement and spending.

IN APP PURCHASE SPENDING BY MATCH TYPE



ACTIONABLE INSIGHTS

Match Type C players tend to spend more both in terms of real money (IAP) and virtual currency compared to the other types. This could indicate a higher engagement or more invested player base in this match type.

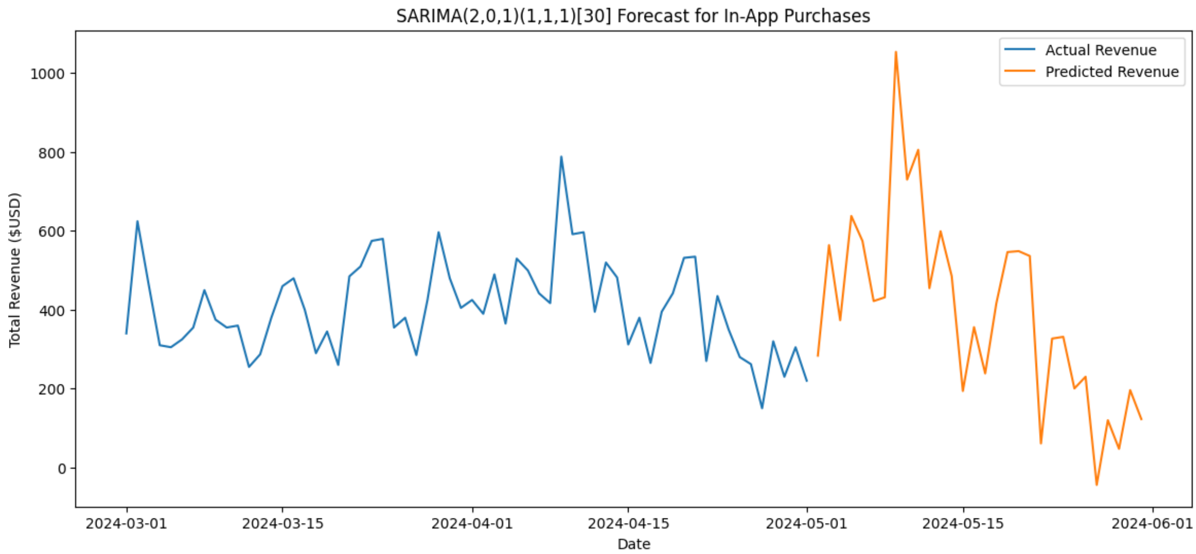
Match Type A and B show lower spending averages, with Match Type A slightly ahead in virtual spending but behind in IAP compared to Match Type B.

RECOMMENDATIONS

Focus on Match Type C: Given the higher spending in Match Type C, it might be beneficial to focus on enhancing this match type with more features, rewards, or related promotions to further boost engagement and spending.

Evaluate Match Types A and B: Investigate why these match types have lower spending. It could be useful to explore player feedback or game analytics to see if adjustments or improvements can increase their attractiveness and spending potential.

TIME-SERIES FORECASTING FOR IN-APP PURCHASES



Model Performance:

The predicted revenue (orange line) follows the overall trend of the actual revenue (blue line) quite closely until about mid-April 2024. After this point, the predicted values diverge more significantly from the actuals, particularly noticeable in the sharp rise and subsequent fall.

Trend and Seasonality:

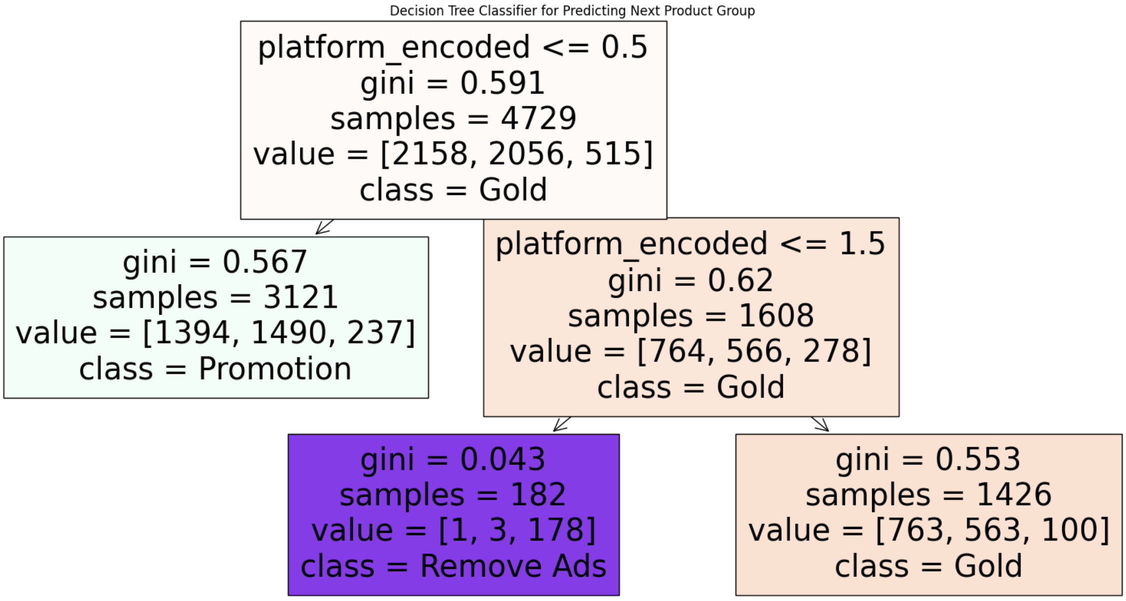
Both the actual and predicted lines show fluctuations that suggest seasonality in revenue, which the SARIMA(2,0,1)(1,1,1)[30] model seems to capture to an extent. The sharp increases and decreases could be tied to specific events or promotions within the app.

ACTIONABLE INISIGHTS

Targeted Promotions: Leverage the periods of high predicted revenue for scheduling marketing campaigns or special offers to maximize income.

Resource Allocation: Prepare for the expected high transaction periods by allocating adequate resources and support to handle increased user activity and transactions.

DECISION TREE REGRESSOR FOR PREDICTING THE PURCHASE



Interpretation:

The tree splits based on the platform_encoded feature, which indicates the platform used by the player.

The nodes show the predicted product group and the class distribution for each split, with color intensity representing the purity of the node.

ACTIONABLE INSIGHTS:

Platform Influence:

Different platforms influence purchasing behavior differently. Platform encoded as <= 0.5 has a different product preference compared to others.

Action: Tailor marketing and product placement strategies based on platform type. For example, platforms classified under <= 0.5 may respond better to promotions than to advertisements for removing ads.

Product Focus:

Gold is generally popular but the propensity to purchase Remove Ads significantly increases under certain conditions (Gini = 0.043).

Action: Investigate what specific platform or user conditions (like frequent app usage or annoyance by ads) lead to an increase in purchases of Remove Ads. Deploy targeted campaigns for Remove Ads in user segments showing high engagement or frustration with ads.

Optimize Campaigns:

The distribution in the left child of the root shows a competitive scenario between Gold and Promotion.

Action: Use A/B testing for promotional strategies on platforms corresponding to platform_encoded <= 0.5 to determine optimal balance between promoting in-app currency and other promotional offers.

Data-Driven Product Development:

Understanding which product groups are popular on which platforms can help guide product development and updates.

Action: Focus development resources on enhancing the features and availability of the most popular product groups for each platform segment.

CONCLUSION

The analysis underscores the complexity of user behavior in digital game platforms and highlights the importance of leveraging data analytics to inform business strategies. By understanding and responding to user preferences and behaviors, "Combat Elite" can enhance user engagement, increase revenue, and improve the overall user experience. This approach of continuous learning and adaptation is crucial in the fast-evolving digital entertainment landscape.