

# Case Study Report

## Data Analysis of a Slot Game's Player Activity

---

Presenter: Freddie (Haojie Chen)

October 26th, 2022

# Catalog

1. Summary
2. Player Engagement Analysis
3. Modeling

# 1. Summary

- Player Engagement

**Situation:** Overall, 41% of users re-visit in another day and 51% of users re-visit within 3 days.

**Problem:** From the target time period, the later the user installs the game, the less likely for the user to re-visit the game.

**Possible Reasons:** Quality of new users decrease overtime; Low bet yield rate discourage users

**Solutions:** Encourage users to spin more at a session due to the lag in the game experience;  
Encourage users to purchase more coins whose coin balance before bet are under 30k.

- Correlation between User Activity and Return Another Day

**Conclusion:** There exists moderate correlation (model accuracy rate 67.5%).

The main factors of influence are No. of Spins and No. of Sessions.

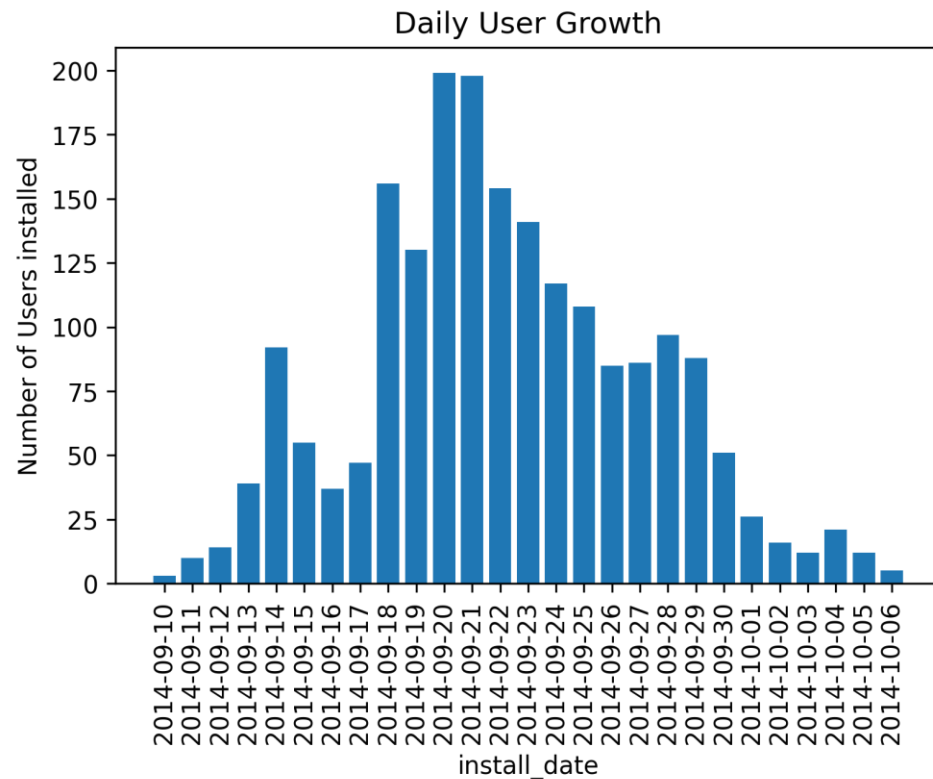
**Consideration:** Due to the limit of sample size and number of features, the model performance is not good enough.

More data points and features should be included to increase the model performance and generalization.

User characteristics change greatly over time, frequent training is necessary to ensure the accuracy of dealing with new data.

# 2. Player Engagement Analysis

## 2.1 User Growth



### Assumption:

- The data in this dataset is correct.
- The dataset includes all the users that installed the app from 2014.09.10 to 2014.10.06

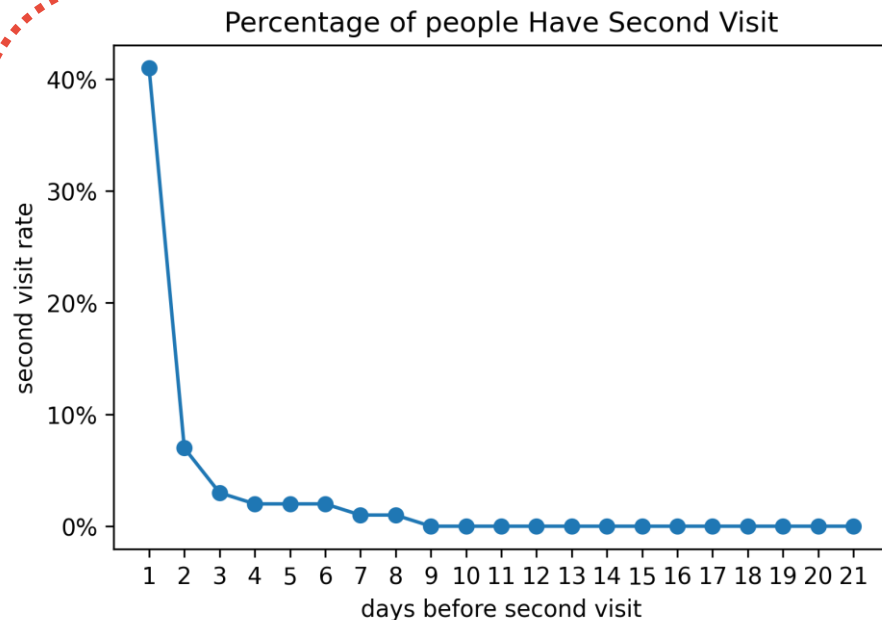
### Insight:

- This period can be divided into three stages: growth(09.10~09.17), boom(09.18~09.29), and decline(09.30~10.06).

# 2. Player Engagement Analysis

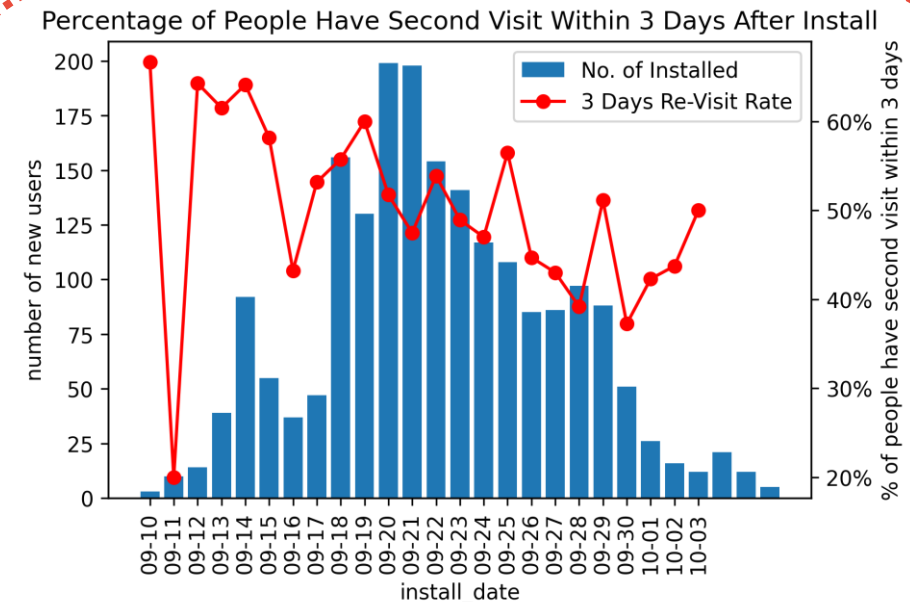
## 2.2 User Second Visit Analysis

### Relative Time Perspective



**60%** of all users have second visit  
**41%** of all users return another day  
**85%** of revisit users return in **3** days

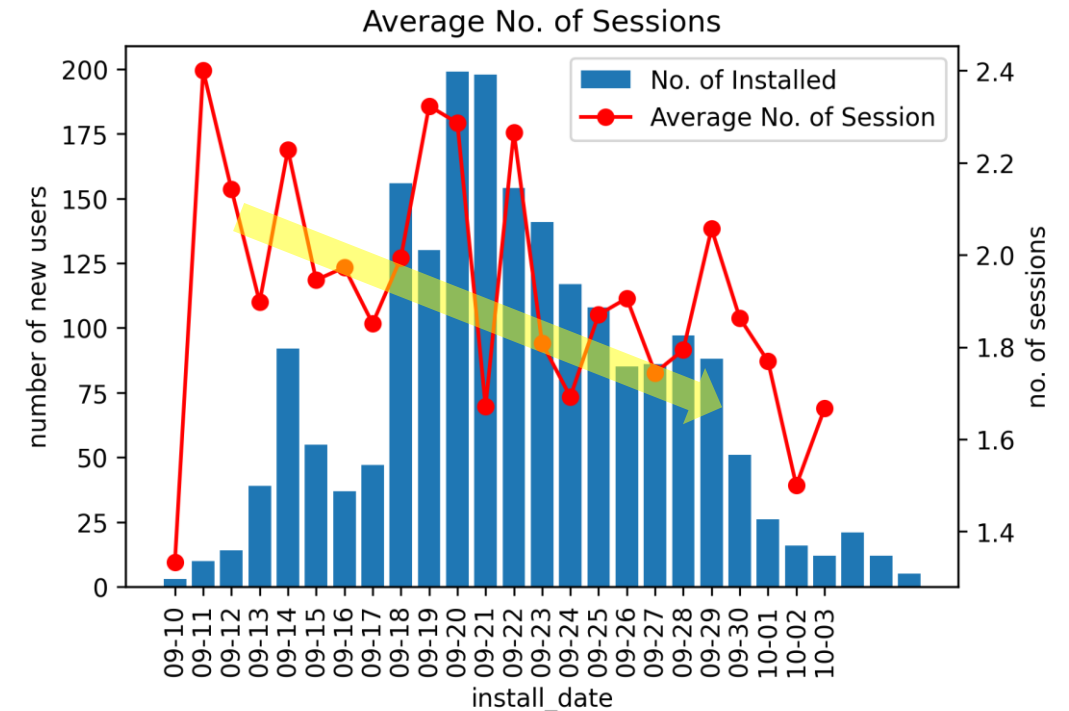
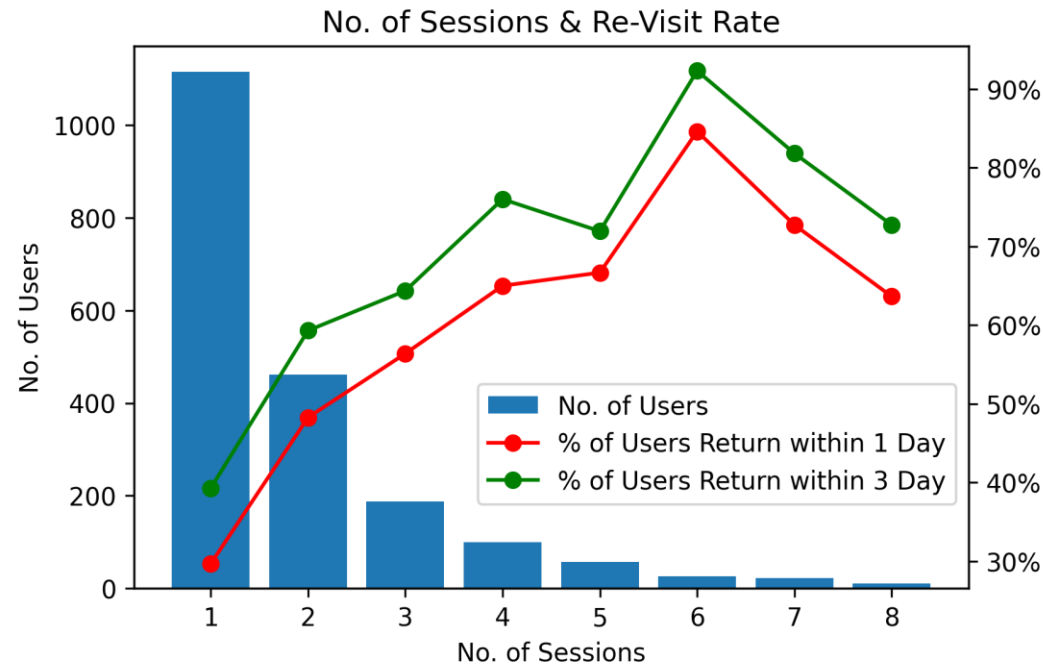
### Absolute Time Perspective



Percentage of revisit user in 3 days  
has a decreasing trend with time

# 2. Player Engagement Analysis

## 2.2.1 Decrease in Revisit Rate is caused by User Quality



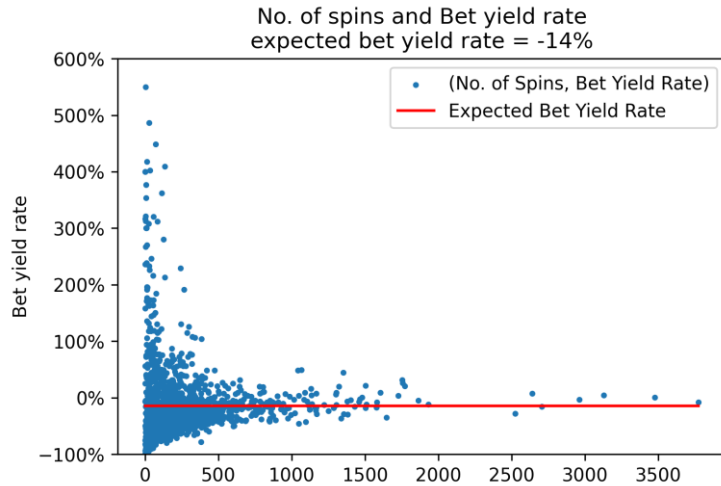
**Number of Session** can reflect user quality, higher No. of Sessions indicates higher Re-Visit Rate.

The decreasing trend in Avg. Number of Session explain retention decrease with time.

**Possible Solution: Analyze user acquisition channels to improve user quality**

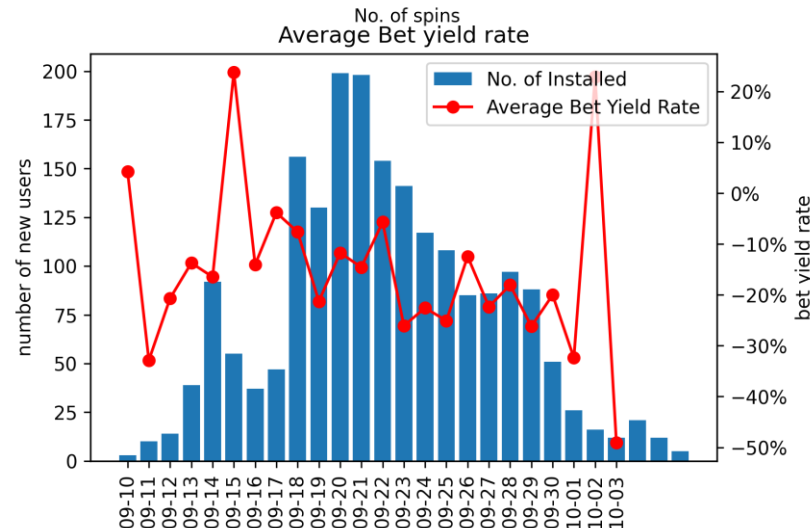
# 2. Player Engagement Analysis

## 2.2.2 Decrease in Revisit Rate is caused by low Yield Rate



$$[Bet\ Yield\ Rate]_i = \frac{[Total\ Coin\ Wins]_i - [Total\ Coin\ Bets]_i}{[Total\ Coin\ Bets]_i}$$

$$Expected\ Bet\ Yield\ Rate = \frac{\sum_i [No.\ of\ Spins]_i \times [Bet\ Yield\ Rate]_i}{\sum_i [No.\ of\ Spins]_i}$$

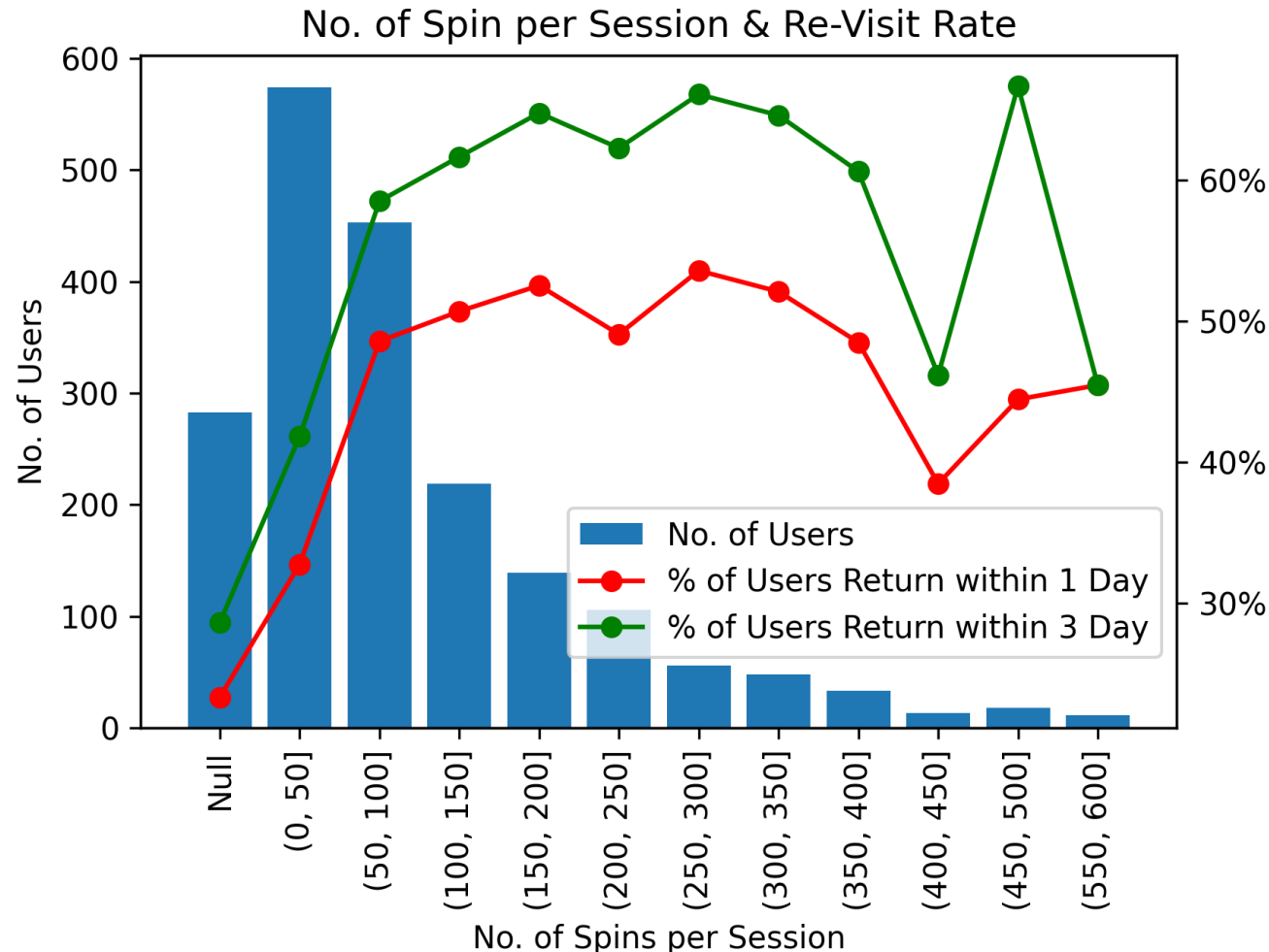


The expected bet yield rate is **-14%**, which might discourage users as the number of spins goes up.

Possible Solution: Give new users a higher yield rate or give old users bonuses

# 2. Player Engagement Analysis

## 2.3 No. of Spins per Session: Let users spin more at a session



**Insight:** With the increase of the No. of Spins per Session, the Re-Visit Rate first rises and then keep stable(noisy at end).

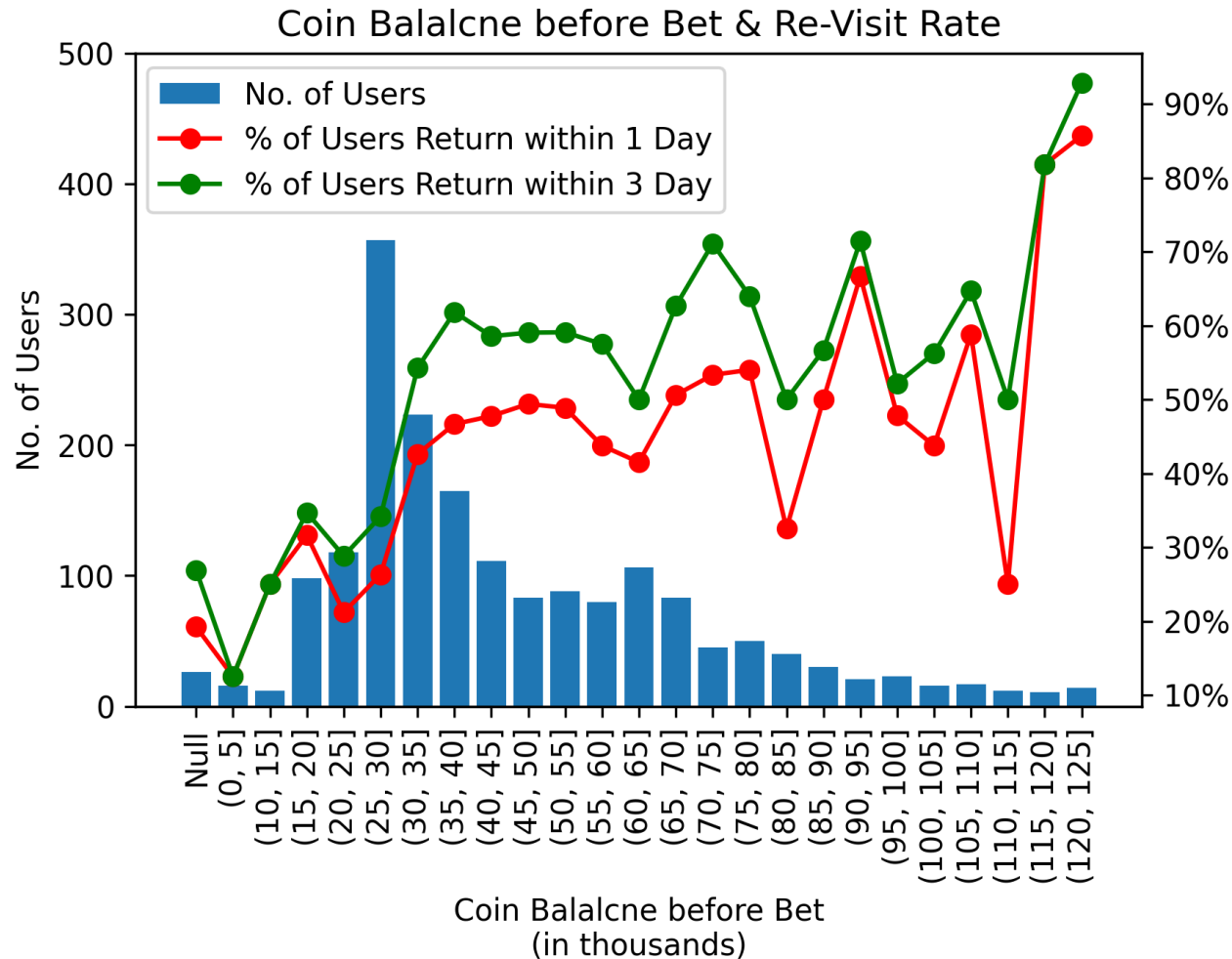
**Possible Reason:** Users has a lag in the game experience and needs a certain number of spins to be attracted by the game.

**Advice:** Encourage users with insufficient No. of Spins per Session to play more, which could enhance user experience and increase Re-Visit Rate .



# 2. Player Engagement Analysis

## 2.4 Coin Balance before Bet indicates Purchase Intention



**Assumption:** There are only two ways to get coins: **Purchase** and **Bet**.

$$\text{Coin Balance before Bet} = \text{Coin Balance} + \text{Coins Bet} - \text{Coins Win}$$

**Insight:** Coin Balance before Bet indicates the purchase intention. Low Purchase Group make up **32%** of all users and has low re-visit rate.

**Advice:** Stimulate the purchase intention can be helpful, methods such as provide coupon for the first time purchase.

# 3. Modeling

## 3.1 Feature Engineering and Data Cleaning

$$\text{Average Bet Amt. per Spin} = \frac{\text{Total Coin Bets}}{\text{No. of Spins}}$$

$$\text{Average Spins per Session} = \frac{\text{No. of Spins}}{\text{No. of Sessions}}$$

$$\text{Total Coin Earns} = \text{Total Coin Wins} - \text{Total Coin Bets}$$

$$\text{Bet Yield Rate} = \frac{\text{Total Coin Earns}}{\text{Total Coin Bets}}$$

$$\text{ROI} = \frac{\text{Total Coin Earns}}{\text{Coin Balance} - \text{Total Coin Earns}}$$

- exclude the data where install date = 2014-10-06 because label is unknown
- exclude feature "Current level" to avoid data leakage

# 3. Modeling

## 3.2 Logistic Regression

Confusion Matrix		Prediction Condition		
		come back another day	not come back another day	
Actual Condition	come back another day	TP 120	FN 697	Recall 14.7%
	not come back another day	FP 64	TN 1113	
		Precision 65.2%	Accuracy 61.8%	

**Model Performance:** Logistic Regression's screening ability of positive cases is insufficient.

$$\text{Precision} = \frac{TP}{TP + FP} = 65.2\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = 14.7\%$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = 61.8\%$$

The precision rate exceeds 60%, which indicates that more than half of the positives identified by the model are actual positives.

However, the recall rate is only 14%, which means most actual positives cannot be identified by the model.

# 3. Modeling

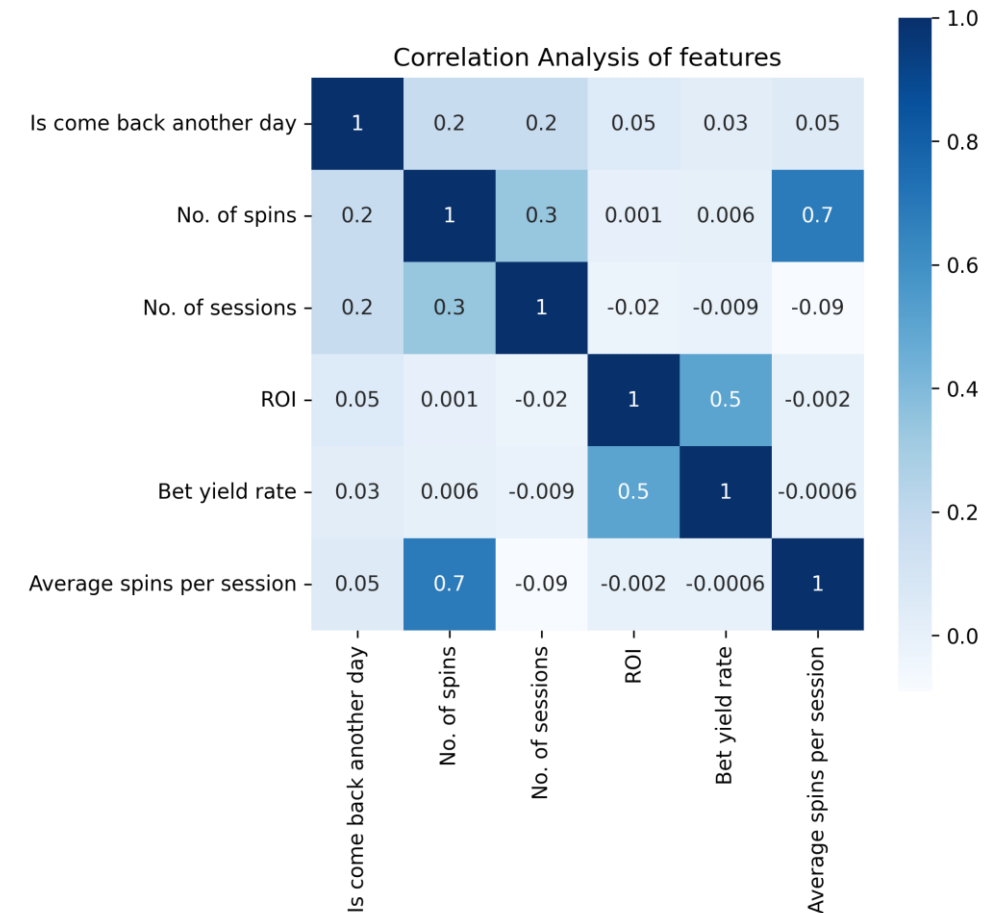
## 3.2 Logistic Regression

### Insights:

- No. of Spins and No. of Sessions are most influential features in the model.
- The correlation coefficient between No. of Spins and No. of Sessions is 0.3, indicating the existence of low correlation between these two features

### Feature Importance Rank

Feature	Coefficient
No. of spins	3.676256
No. of sessions	2.703804
ROI	0.810488
Bet yield rate	0.680844
Average spins per session	-0.457461



# 3. Modeling

## 3.3 Random Forest

Confusion Matrix		Prediction Condition		
		come back another day	not come back another day	
Actual Condition	come back another day	TP 378	FN 439	Recall 46.3%
	not come back another day	FP 209	TN 968	
		Precision 64.4%	Accuracy 67.5%	

**Model Performance:** Random forest has better screening ability and prediction precision for positive cases.

$$\text{Precision} = \frac{TP}{TP + FP} = 64.4\%$$

$$\text{Recall} = \frac{TP}{TP + FN} = 46.3\%$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} = 67.5\%$$

The precision rate exceeds 60%, which indicates that more than half of the positives identified by the model are actual positives.

The recall rate is 46%, which indicates that half of the actual positives can be identified by the model.

# 3. Modeling

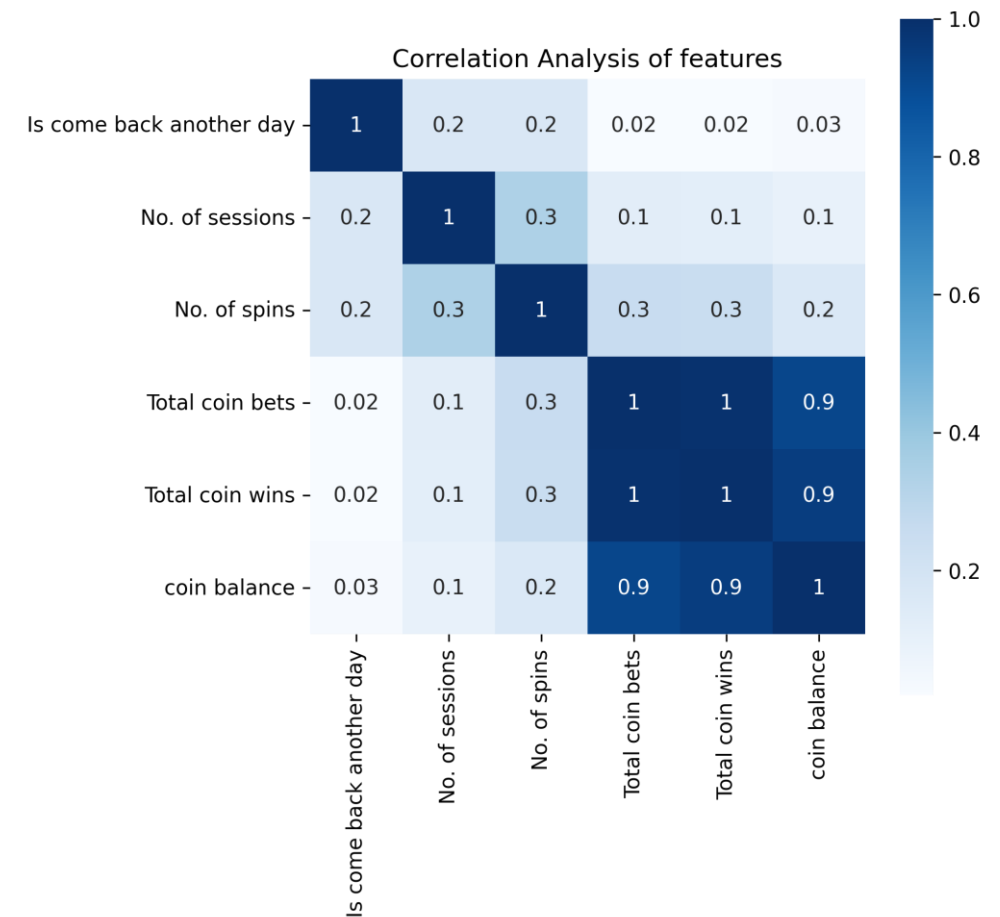
## 3.3 Random Forest

### Insights:

- No. of Sessions, No. of Spins, and Total coin bets are the feature that contributes the most to the model.
- No. of Sessions and No. of Spins have weak correlation with the target (is come back another day).
- Total coin bets, Total coin wins, and coin balance are highly correlated with each other.

### Feature Importance Rank

Feature	VIM
No. of sessions	0.244714
No. of spins	0.195443
Total coin bets	0.144737
Total coin wins	0.088844
coin balance	0.080992



# 3. Modeling

## 3.4 Conclusion and Consideration

**Conclusion:** There exists **moderate correlation** between user activity and the likelihood to come back to the game another day.

- Random forest can screen out around 50% users that come back to the game another day.
- Random forest has over 50% of sure that the prediction match the fact.

**Consideration:**

- Small sample size increase the chance of **overfitting**, even if I limit the model by regularization and pruning.
- The **number of features** limits the model performance, and more features may improve model performance.
- User characteristics change greatly over time, frequent training is necessary to ensure the accuracy of dealing with new data.