2022

Market Segmentation and Recency Frequency Monetary Value Analysis for a Freemium Mobile Game

Satvik Ajmera  
*Southern Methodist University, sajmera@smu.edu*

Taylor Bonar  
*Southern Methodist University, tbonar@smu.edu*

Dylan Scott  
*Southern Methodist University, drscott@smu.edu*

Carol Miu  
*carol@peoplefun.com*

Alana Manuel  
*alana.manuel@peoplefun.com*

Follow this and additional works at: [https://scholar.smu.edu/datasciencereview](https://scholar.smu.edu/datasciencereview)

Part of the Data Science Commons

**Recommended Citation**  
Ajmera, Satvik; Bonar, Taylor; Scott, Dylan; Miu, Carol; and Manuel, Alana (2022) "Market Segmentation and Recency Frequency Monetary Value Analysis for a Freemium Mobile Game," *SMU Data Science Review*. Vol. 6: No. 2, Article 1.  
Available at: [https://scholar.smu.edu/datasciencereview/vol6/iss2/1](https://scholar.smu.edu/datasciencereview/vol6/iss2/1)

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit [http://digitalrepository.smu.edu](http://digitalrepository.smu.edu).
Market Segmentation and Recency Frequency Monetary Value Analysis for a Freemium Mobile Game

Satvik Ajmera1, Taylor Bonar1, Dylan Scott1, Carol Miu2, Alana Manuel3
1 Master of Science in Data Science, Southern Methodist University, Dallas, TX 75275 USA
2 Chief Enterprise Officer
PeopleFun
Richardson, TX 75081 USA
3 Data Scientist
PeopleFun
Richardson, TX 75081 USA
{sajmera, tbonar, dscoott}@smu.edu
carol@peoplefun.com
alana.manuel@peoplefun.com

Abstract. Bricks ‘N Balls is a freemium game that relies on in-app purchases and ad monetization from users to be profitable at no upfront cost to the players. This study explores how in-game data analytics and purchase data can be used to segment players. Features taken into consideration for segmentation include past purchasing habits along with the players interactions within the missions. This study uses the Recency Frequency Monetary Value (RFM) framework to extract insights on player purchasing behavior to segment players into clusters and predict how much users will spend in the future.

1 Introduction

Bricks ‘n Balls (BnB) is a casual freemium mobile game that was founded by Chinese internet company called Cheetah Mobile in 2009. However, this game and several other of Cheetah Mobile’s applications were removed from the Google Play Store because of non-compliance of store policy and engagement of massive advertisement fraud [14]. As a result, a mobile technology company called AppLovin, purchased Bricks ‘n Balls and other software assets from Cheetah Mobile to be distributed among their various studios [15]. PeopleFun, a studio with AppLovin, has overhauled the entire mobile game to be in-compliance with mobile app store policy and privacy policy. Bricks ‘n Balls is currently featured on both the Apple App Store, and Google Play Store. Bricks n Balls has over 3 million players worldwide. As Bricks ‘n Balls is a freemium game, it generates revenue from users/players through in-app purchases and advertisement offers.

The mechanics in the game mirror Brick Breaker which was originally released in 1999. For each level, called a “Mission” a player must aim and shoot balls at bricks for each turn until all bricks are destroyed. The bricks move downward upon each turn. If a brick reaches the bottom before it is destroyed, the player fails the mission. There is no limit to how many times a player can retry a level. What differentiates Bricks ‘n Balls from Brick Breaker, analogously, a freemium game from a free-to-play game is
that *Bricks ‘n Balls* has several in-game “power-ups” or virtual products that help a player progress faster through the game. Freemium games sell in-game offers, and bundles for players to purchase. Specifically, *Bricks ‘n Balls* sells various packages of virtual currency called *Rubies* which can be used to purchase power-ups and level skips. Players can also earn rubies by watching short in-game advertisements. This study explores two components: market segmentation and purchase prediction.

For market segmentation, this research involves understanding the different segments of players and the similar characteristics between these player segments. From there, the research explores the use of different clustering algorithms (unsupervised learning) to learn the diverse types of purchasers. The second component will be to build a purchase prediction model to see how much money users will spend at the end of November.

The player of *Bricks ‘n Balls* can be identified by their actions within the game. Most of the player base is considered non-purchasers, players who will not make a purchase within the app. The players who do decide to make a purchase can be broken into two groups of frequent purchasers and one-time purchasers. Given the dataset timeframe, players identified as one-time purchasers may not stay in that category forever and other actions will be taken into consideration.

The problem facing PeopleFun and many other mobile gaming studios is when to make an offer and how frequently to maximize ROI (return on investment) and player satisfaction. This paper explores real world data from *Bricks ‘n Balls* to offer a data-driven solution to enhance the value of in-app purchases without interfering with a player's experience. This research was a rare opportunity as many mobile gaming studios are reluctant to lend their own data. Recency Frequency Monetary Value (RFM) analysis is a valuable tool that helps track the habits of purchasers. This analysis can easily be adapted to apply to other industries outside the mobile gaming realm.

A major motivation of this research is to understand the purchasing habits of players from a data perspective. Through market segmentation using clustering a better understanding can be achieved of how in-app purchase offers influence a player's experience and behavior.

## 2 Literature Review

### 2.1 Player Satisfaction and Player Behavior

Data is an important aspect of understanding a company’s product, however, it is not the only item for consideration. In the gaming industry much of a player’s journey encompasses an emotional response. Many decisions made throughout a player’s lifecycle revolve around a player’s goals. In *Bricks ‘n Balls* there are multiple game modes, each with their own goals. For *Bricks ‘n Balls* player’s gameplay styles can be broken into collection, completion, or escape/leisure. Understanding these motivations can aid the research done on player purchases and influence game items and designs that can enhance a user’s experience.
Some studies have reviewed potential theories on why players purchase additional items and content in video games. Rogers discusses the intrinsic motivation and enjoyment that video game players receive through the lens of the self-determination theory. His argument is that feedback, rules, and social elements of video games fulfills the SDT dimensions of competence autonomy and relatedness for video game players [11]. Rogers describes feedback as the “the fundamental interaction between the player and the game” [11]. The feedback in a video game allows a user to quickly reach the game’s goal. In Bricks N Balls, players’ goals within the game are to complete missions. The results of Roger’s study state that players who receive too much feedback cause an increase in cognitive load and lowers player satisfaction. If a player fails to complete a mission, Bricks N Balls delivers feedback in the form of an advertisement, or in-app purchase, to obtain virtual currency that can be used to buy items to help the player progress further. From before, players that receive too much feedback from offers results in cognitive overload, therefore, a motivation for this study is to create a recommendation system which optimizes offers sent to improve player satisfaction. Additionally, Rogers states that player satisfaction is dependent on players having autonomy subject to the flexible rules of within a game [11]. With casual freemium mobile games, it is difficult to provide autonomy and flexibility. However, Brick N Balls addresses these features of player satisfaction by giving players the option to freely explore and try the several different game modes, collect skins for personalization, or complete missions to complete a photo album.

There are several types of players who have different motivations to progress through a game. Some players are motivated to complete a game by collecting in-game items or collectables. In Bricks N Balls, these players tend to make in-app purchases to collect various ball skins or limited time special power-ups. Players can also be motivated by game completion achievements. These players look to complete every level and some of these players replay previous completed levels to further maximize their score. These players’ in-app purchases help them with second chances and boosting items to help boost their score. Lastly, casual players are motivated by playing the game at their own pace and to escape from daily tasks. These players’ in-app purchases allow them to increase their playtime and overall enjoyment.

After interviewing with a Senior User Experience Researcher at PeopleFun, the researcher discussed that loss-aversion is a principal factor when it comes to player satisfaction and motivation. In Bricks N Balls when a player fails a mission that is considered a loss and they will not be able to progress unless they pass the mission. According to Achievement Relocked: Loss Aversion and Game Design, “loss aversion” is seen as a sure outcome versus a gamble. Furthermore, loss-aversion is based on the circumstances presented to a player, where one is more appealing than the other. In Bricks N Balls, each turn can be seen as a gamble where a player can get closer to completing the mission. In fact, the several types of in-app purchase a player can make can be interpreted as a gamble. There is an opportunity cost for a player to purchase, for example, a power-up versus rubies. A power-up purchase may help a player achieve their goal of mission completion in the short-run. However, the power-up is only as good as the player is at using it. On the other hand, other in-app purchases are seen as a sure outcome. A player could have purchased rubies or a ball skin, because the items’ purpose is a known outcome to the player.
Gables are seen as an attractive offer when the sure outcome is not what the player wants. For example, if the scenario is purchasing the power-up which might help the player pass the level or fail no matter what the player is much more likely to purchase the power-up because the chance of winning aligns with their goals of not losing. This scenario aligns with the key principles of loss aversion as stated in chapter 1 of *Achievement Relocked: Loss Aversion and Game Design*:

1. “People prefer sure gains over a chance for a larger gain.”
2. “People will take a risk to avoid a sure loss.”
3. “People treat all small probabilities the same and just compare the gain or loss.”
4. “Both 100 percent and 0 percent are special.” (Engelstein. (2020))

This study aims to identify different purchasing groups within Bricks ‘n Balls player base and how their gameplay behavioral habits interact with in-app purchasing behavior utilizing live game data analytics. Further, to analyze the various market segmentations, the study will utilize RFM frameworks and extend with additional gameplay behavioral features using self-determination theory as a theoretical foundation. Finally, the study will investigate whether an offer recommendation system can be generalized to help players increase gameplay satisfaction and achieve gameplay goals.

### 2.2 Recency Frequency Monetary Value (RFM) Analysis

In-app games rely heavily on in-app purchases to drive revenue for a company in freemium games. However, when creating in-app offers to a customer base in the mobile game market, defining and organizing the customers is imperative to formulate a market strategy to not induce offer fatigue and lower satisfaction. One of the more popular solutions to data mine and identify valuable purchasing behaviors of customers is the RFM model (S. Q. Moghaddam, N. Abdolvand, and S.R. Harandi 2017, as cited by E. Ernawati, S. Baharin, and F. Kasmin 2020).

Traditional RFM analysis is a frequently used technique in market segmentation and market research focusing on three features to determine a customer’s value based on prior purchasing behavior [10]. Perisic and Pahor’s research emphasizes that defining the RFM framework and constructing new features is crucial to building a churn prediction model rather than the specific machine learning algorithm for freemium games [10]. In creating the RFM framework, the authors feature engineered frequency to include both their short-term and long-term frequency, where the short-term frequency aggregates the daily number of sessions completed weekly and the long-term frequency is the customer’s total lifetime value. The features examine three qualifiers that relate to the are used to quantify the likelihood that a customer would respond to a specific offer: (a) recency of a customer’s latest purchase, (b) frequency, or rate of a customer’s purchasing behavior, and (c) monetary, or the monetary amount a customer spends [10]. Additionally, the authors included a new feature called “Intensity” which is the length of the average session over an observation period. Finally, the monetary value is the overall customer lifetime value and recency is the number of days since currency was last spent over the observation period [10]. Lastly, RFM frameworks have
shown adaptability to include additional features to boost prediction power, such as, time since first purchase, and churn probability (Yeh, Yang, and Ting, 2009, as cited in Perisic and Pahor, 2021) [10]. This adaptability can be utilized to include additional information that a company has that can help refine information about a customer base as shown by Persic and Pahor.

2.3 Market Segmentation Techniques

When utilizing RFM analysis, a clustering method is typically employed to segment the population to segment players based on purchasing buying patterns such as k-means clustering. K-Means is a non-hierarchical, meaning existing clusters can be split and merged. The algorithm groups a dataset into the defined number of clusters, k, to create a better accuracy by minimizing the inertia, the within-cluster sum of squares criterion [9]. K-means can be simplified into four basic steps:

- Choose the initial centroids, such as, choose k samples from the dataset.
- Create new centroids via taking the mean value of all samples assigned to each previous centroid
- Compute the difference between the old and new centroids
- Loop steps 2-3 until the difference is less than a defined threshold (i.e., the centroids do not move significantly)

In practice, instead of manually computing and verifying the optimal value for k neighbors, cluster validity indices are used to identify the ideal number of clusters, such as the Elbow Method [5]. The Elbow Method is a visual method that varies the number of clusters and compares the inertia as a plotted line graph. Once the graph starts to become linear into the x-axis, the k value corresponding to this point is the optimal number of clusters. As shown in Figure 1, we can note the optimal cluster for this dataset is at k=7, and the “distortion”, average sum of squared distances to the centroids, is 26,333.181. Additionally, with the visualization in Figure 1, the time required to train the clustering model is also shown.
Another method for evaluating an optimal value for $k$ neighbors is utilizing a silhouette coefficient score. The silhouette coefficient is bound between values –1 for incorrect clustering, +1 for highly dense clustering, and scores around 0 indicate an overlapping of clusters [12]. This score is calculated for each sample by computing two scores, see Equation 1:

- **Intra-cluster Distance:** The mean distance for the sample and every other point within the same class which can be referred to as the partition from our clustering technique

- **Nearest Cluster Distance:** Mean distance between a sample and all other points in the next nearest cluster

\[
Equation 1 \\
S = \frac{(b - a)}{\max(a, b)}
\]

One advantage of using a silhouette coefficient compared to the Elbow Method is that we can visually see our potential cluster density and determine whether clusters are imbalanced as shown in Figure 2. This visualization is due to the score being bound between –1 and 1, we can easily gauge when clusters are dense and well-separated by the selected distance metric, such as Euclidean distances, when seeking compact and separated clusters [12].
A weakness of K-Means is that it is highly sensitive to outliers, and usually K-Medoids clustering is used in its place [8]. According to *The Elements of Statistical Learning*, this weakness is due to the K-Means algorithm utilizing squared Euclidean distance for its dissimilarity metric [6]. As K-Means is restricted to squared Euclidean distance, this means that the outliers, who have the largest distances from the centroid, result in having the highest influence [6].

K-Medoids is a related method to K-Means, but instead of utilizing a centroid, a representative object called a medoid is used. The medoid is different from a centroid where it is a data point with the least total distance to all other members of its cluster, which allows K-Medoids clustering to be robust to outliers, when compared to its counterpart [8]. According to *The Elements of Statistical Learning*, the K-Medoids clustering algorithm can be summarized as follows:

1. For a given cluster assignment \( C \) finds observation in the cluster minimizing total distance to other points in that cluster:

   \[
   i_k^* = \arg \min_{i \in C(i) = k} \sum_{i \in C(i) = k} D(x_i, m_k)
   \]

   Then \( m_k = x_{i_k^*} \) \( k = 1, 2, \ldots, K \) are the current estimates of the cluster centers

2. Given a set of cluster centers as a current set of cluster centers \( \{m_1, \ldots, m_k\} \), minimize the total error by assigning each observation to the closest (current) cluster center:

   \[
   C(i) = \arg \min_{k} D(x_i, m_k)
   \]

3. Repeat steps 1 and 2 until cluster assignments stabilize and no longer change [6] (p.516)
2.3 Player Spending Prediction Techniques

**Linear Regression**

Linear Regression is an extremely popular statistical model especially for comparing more advanced models. Part of its popularity is due to its simplicity. Linear regression models sit a straight line and many datasets fit that shape. When considering linear regression there are four assumptions that need to be met: linearity, homoscedasticity, independence, and normality. Linear regression runs quickly on large datasets but is not robust to some issues that also plague other models such as outliers.

**Support Vector Regression**

Support Vector Regression (SVR) is a supervised machine learning model that works well on yet to be seen data [1]. Given this research SVR can be used to predict purchases into the future. Like Support Vector Machines (SVM) SVR uses kernels to help draw a plane through the data fitting the targeted points. As mentioned in Support Vector Regression. In: Efficient Learning Machines “Although less popular than SVM, SVR has been proven to be an effective tool in real-value function estimation.” [1]. If our data is nonlinear then SVR can map our data into a higher dimension. This new dimension is called the kernel space. By mapping the data into a greater dimension SVR can achieve a higher accuracy by transforming nonlinear data into a higher dimension in which it may resemble something closer to a straight line. SVR is a very flexible model due to the kernel trick and robust to nonlinear data. SVR was used to predict the exchange rate of the EUR. [13]. Much like this research, that paper highlighted robust SVR was to predict unknown fluctuating outcomes such as exchange rates or how much someone is most likely to spend.

RMSE is used to judge the performance of the overall SVR model where the training data is historical data on the Euro exchange rate. This is like this study where SVR model is trained on past purchases to help predict future purchases. SVM also requires an understanding of Vapnik-Chervonenkis (VC) theory since SVM uses VC control of margins. SVM’s strengths are “kernels, sparse solution, VC control of margins and the number of support vectors” [1].

SVM uses a symmetrical loss function where a tube of minimal radius is fit to the estimation function having a symmetrical absolute value of errors on all sides. The points outside the estimate are ignored. The points outside the tube receive a penalty, however, the points within the tube do not.

**Multilayer Perceptron Regression (MLP)**

The Multilayer Perceptron is a supervised learning algorithm that is different from other learning algorithms by making use of hidden layers. The hidden layers consist of neurons that use the training data to learn different features for predictions. In fact, each hidden layer uses non-linearity through activation functions such as reLU, sigmoid or hyperbolic tangent. The hidden layers obscure the view of the training data. As a result, neural networks or MLPs are referred to as “black-box” models. Hyperparameter tuning is more challenging for MLPs compared to other algorithms due to the vast number of hyperparameters the modeler can adjust. This includes, but
is not limited to, the number of hidden layers, number of neurons, optimizers or the several activation functions need to be chosen to have useful model. [9].

**Random Forest Regression**
Random Forest (RF) is a popular supervised machine learning algorithm that can be used for both classification and regression. RF works by generating multiple trees of predictions to then arrive at a more accurate prediction. The multiple trees insulate themselves from error thorough volume. In other words, multiple models are fit until the averages of all predictions lead closer to the actual data. RFR trains multiple regression trees on bootstrapped samples of the data by averaging [4]. Random forest regression models consider many hyperparameters to run accuracy. Such parameters include number of decision trees, loss function (MAE or MSE), max depth of the tree, max features considered and max samples for each tree. Random Forest Regressor is robust to non-linear data making it a desirable alternative to linear regression.

3 Data

3.1 Data Preparation

The data was acquired in partnership with PeopleFun is over two months, October through November 2021. Each entry of the data consists of player events and their related features. An event is an interaction by a user within the Bricks ’n Balls game. For example, but not limited to, a unique recorded event can be a mission start, mission closed, mission failed, app closed, app opened, transaction and others. In addition, each event has associated features recorded relating to information for a given event (e.g., the number of rubies, power-ups, items used during a given level, reattempts of a level, etc.). Every event has a timestamp of when that player interacts with the game. Each event also contains a session ID which is a unique identifier to a single player in a game, but unique between sessions. Finally, each event has a unique user ID for the specific player that executed the event. Due to the nature of the data, this research does not have access to historical data to past player events. The downside of this is that the study will not be able to explore the lifetime value of players from their first event to the most recent event.

While investigating the data provided by PeopleFun some outliers were discovered which affected the results. Many users had abnormal amounts of spending within the Bricks ’n Balls app store. Though power-users are vital for the game and revenue they generate, we decided to remove the top 5% of total spenders from our dataset. This was to help normalize the data and since these users are consistent spenders within the app there is not much PeopleFun can do to increase the amount of spending these users do.
3.2 In-Depth Explanation on Bricks N Balls Gameplay

In Bricks ‘n Balls, there are several different game modes that can be played by users called **Endless, Gravity, Classic, 100 Ball, Survival, Interesting Trip, Treasure Hunt, Weekly Tournament, and Saga**. Each of these game modes has different game mechanics and goals. 76% of all missions completed were of the Saga game mode. In Saga, a player’s goal is to destroy bricks by aiming and shooting balls at them. For example, a player is equipped with 50-60 balls that are shot in succession per turn. Each brick is assigned a numerical value that will lower upon being hit by balls until it is destroyed. The challenge is to destroy every brick before they reach the bottom of the screen, otherwise, the player fails the mission. As a player completes more missions, the difficulty of the missions increases. There are over 4000 levels available for users to progress through. Since the data is a snapshot of the overall player data within two months, there is a wide range of players who have already progressed significantly versus others who have just begun the game. Looking at each unique player’s highest completed missions over the course of two months, the average completed level is 115 while the median completed mission is 44. Furthermore, there have only been 2,130 players who have completed over level 1000 of the 2.5 million players. Figure 2 is a kernel density plot of both purchasers and non-purchasers which shows the distribution of maximum mission (missionID) completed by each player.

![Figure 3. Maximum Completed Missions for Each Unique Player](image)

As the study involves building a purchase model, it is crucial to look at the specific levels when players are making in-app purchases. The visualization below shows us the missions at which players make an in-app purchase to further progress through the game. The transactions are grouped into bins of 30 missions. Over 200,000 transactions occurred between mission 446 and mission 550. Players are mostly purchasing earlier within the game and again later in the game. There may be some learning curve for
players to adjust to which may be why players have purchased more during the early missions where purchases sharply drop and rise significantly later. In fact, over 75% of all transactions occurred between mission 0 and mission 546, while 25% of all transactions occurred between mission 546 and mission 3547. Players’ missions may be useful features to be used for both clustering and modeling.

![Figure 4. Histogram of Player Transactions During Missions](image)

### 3.3 Exploratory Data Analysis (EDA)

Freemium game’s in-app purchases can be examined through Recency Frequency Monetary Value (RFM) analysis to explore questions centered around what users are purchasing, how often they are purchasing, and how much a user spends. In the player base for October and November 2021, the data shows 3,177,782 unique users with 1.98% of those users having at least one transaction in that period. Of the 1.98% of users who did have a purchase 56% of users were categorized as one-time purchasers who only had one transaction within the two-month dataset.

**Recency**

Initial data exploration shows that in the user’s transaction events for the period, it can cover anywhere between 1 day to 60 days (about 2 months) with the upper quartile reaching 39 days (Figure 4).

Recency is a key metric for identifying the different clusters of players within the dataset. Some players could be mistaken for one-time purchasers. However, if they made their purchase near the end of the dataset this does not suggest that they are strictly a one-time purchaser since there is not enough follow-up data to show no other purchase. Therefore, the habits of the players, what they are buying, what they are purchasing, and when they are purchasing are important metrics for an accurate model.
Inspecting the top 1 percent of users for transaction events, the data shows there is a spread of transaction events to a given session (Figure 5). However, the top spenders tend to mostly purchase Ruby bundles of 550, 100, and 1,150 (Figure 6).

Regarding the 56% of users, who were categorized as one-time purchasers, the products they purchased varied from some of the more frequent buyers. The top 3 selling products for this group were Remove Ads, Chest Number 1, and 550 Rubies. Chest Number 1 is the cheapest item of all three. Many of the more engaged players see value in removing ads, because this one-time purchase allows players progress through missions faster. Users can earn Rubies by completing missions. However, some of the higher missions can be more difficult to complete, resulting in users purchasing rubies. Rubies 550 was the most common quantity of Rubies purchased. The chest differs from other in-app products, because users are uncertain exactly what they will receive from the purchase. The low price point is a substantial value to those users who like to take chances. Understanding the different price points for in-app products that are purchased by players may improve that players’ purchase frequency.

Table 1. Aggregation of Top Purchasers Transaction Events and Sessions

<table>
<thead>
<tr>
<th>Anonymous User</th>
<th>Number of Transactions</th>
<th>Number of In-Game Sessions</th>
<th>Ratio of Transaction to Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A</td>
<td>608</td>
<td>61</td>
<td>9.967</td>
</tr>
<tr>
<td>User B</td>
<td>602</td>
<td>135</td>
<td>4.459</td>
</tr>
<tr>
<td>User C</td>
<td>547</td>
<td>117</td>
<td>4.675</td>
</tr>
<tr>
<td>User D</td>
<td>479</td>
<td>129</td>
<td>3.713</td>
</tr>
<tr>
<td>User E</td>
<td>391</td>
<td>119</td>
<td>2.047</td>
</tr>
<tr>
<td>User F</td>
<td>382</td>
<td>127</td>
<td>3.008</td>
</tr>
<tr>
<td>User G</td>
<td>381</td>
<td>31</td>
<td>12.290</td>
</tr>
<tr>
<td>User H</td>
<td>351</td>
<td>83</td>
<td>4.229</td>
</tr>
<tr>
<td>User I</td>
<td>349</td>
<td>116</td>
<td>3.009</td>
</tr>
<tr>
<td>User J</td>
<td>333</td>
<td>91</td>
<td>3.659</td>
</tr>
</tbody>
</table>
Monetary Value

As *Bricks ’n Balls* is not restricted to a single country for their player base, users can purchase in a variety of different currencies depending on the Apple Store or Google Play Store their account is registered to. To understand the monetary value analysis, all data must be first transcribed to a single currency type to understand the relationship of the user’s purchase transaction. Additionally, as *Bricks ’n Balls* was an acquired asset, the initial setup for the data analytics was in a currency’s lowest denomination and would need to be converted (e.g., USD was in pennies instead of dollars).

Furthermore, competition within the *Bricks ’n Balls* virtual store itself needs to be accounted for. The number of products offered as well as their pricing could impact who is purchasing, and what is being purchased. *Bricks ’n Balls* offers 50 assorted products for in-app purchase. Some of these products are offered seasonally, on a weekly rotation, or permanently offered within the in-game store. Looking at the price distribution between the various products, many of them are priced under $10.00 with two products priced under $1.00. A violin plot shows the distribution of prices among the *Bricks ’n Balls* products (Figure 7).
4 Methods

4.1 User Purchasing Clustering Dataset & Evaluation of Clustering Algorithms

To get an initial understanding of the player base engaging in the premium items available to purchase, the raw data needed to be transformed using the RFM framework. Each transaction event was taken from the month of October 2021 per user and calculated from anonymously collected game analytics. Each primary attribute of the RFM dataset was then also normalized with min-maxing for later clustering comparisons.

Table 2. RFM Feature Creation

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Data Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recency</td>
<td>Integer</td>
<td>Hours since last purchase</td>
</tr>
<tr>
<td>Frequency</td>
<td>Integer</td>
<td>Number of transaction events captured in analytics</td>
</tr>
<tr>
<td>Monetary Total</td>
<td>Integer</td>
<td>Total amount spent in USD pennies</td>
</tr>
<tr>
<td>s_Recency</td>
<td>Integer</td>
<td>Standard Scaling of Recency</td>
</tr>
<tr>
<td>s_Frequency</td>
<td>Integer</td>
<td>Standard Scaling of Frequency</td>
</tr>
<tr>
<td>s_Monetary_Total</td>
<td>Integer</td>
<td>Standard Scaling of Monetary_Total</td>
</tr>
<tr>
<td>m_Recency</td>
<td>Integer</td>
<td>Min-Max Scaling of Recency</td>
</tr>
</tbody>
</table>
Each set of features were then utilized in calculating an Elbow Curve, silhouette score, and silhouette visualization to determine the ideal number of clusters to set to and ideal scaling operation to utilize.

**RFM Cluster Evaluation**

Without scaling or normalizing the RFM dataset, the ideal number of clusters is set to 2, however, our distortion score is quite high as shown in Figure 9 for K-Means and slightly higher for K-Medoids in Figure 11. This high distortion score shows the importance of scaling the RFM dataset to improve the accuracy of clustering [7]. Additionally, when comparing Figure 10 shows there

<table>
<thead>
<tr>
<th>m Frequency</th>
<th>Integer</th>
<th>Min-Max Scaling of Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>m Monetary Total</td>
<td>Integer</td>
<td>Min-Max Scaling of Monetary Total</td>
</tr>
</tbody>
</table>
RFM Standardization Cluster Evaluation

With standardizing the RFM dataset, both K-Means and K-Medoids distortion score decreased as indicated in Figure 12 and Figure 14. However, there is disagreement between the silhouettes scores and the elbow plots on the optimal number of clusters. In both K-Means and K-Medoids, the suggested cluster size is 2 or 4. However, K-Means overall outperformed K-Medoids in both distortion and Silhouette scores.

![Fig 12. K-Means Elbow Curve](image1)

![Fig 14. K-Medoids Elbow Curve](image2)

![Fig 13. K-Means Silhouette Visualization](image3)

![Fig 15. K-Medoids Silhouette Visualization](image4)

RFM Min-Max Scaling Cluster Evaluation

Lastly, the min-maxing RFM dataset outperformed both the standardized and original RFM datasets in both K-Means and K-Medoids. The optimal cluster sizes are recommended to be 2 or 3. As both methods agree with 2 or 3 clusters,
4.2 Feature Creation for Prediction

Due to the large amount of data provided for this research, Google BigQuery was used to store and data-engineer new features from the existing data. These new features were used to further distinguish our users and purchase groups. The goal of these features is to highlight how users interact with the game and identify which habits lead to purchases within the app. Each new feature was created within Google BigQuery using SQL. Many of features take inspiration from RFM-LIR Feature Framework for Churn Prediction in the Mobile Games Market which explored how mobile games lose customers over time.

Running averages were created between mission completion and purchases to show the frequency of each metric. The running averages help show how often someone completes a mission as well as makes a purchase. With heavier use of the app, we could expect more frequent purchases to be made.

$Lags$ or $pools$ were also created for each week of October. Each pool represents a particular feature’s value during a week the user was interacting with the game. By separating out these weeks a pattern could be established over time showing when users are playing the game frequently and making frequent purchases. This helped with both outlier detection with binge users and seasonality detection over the different links. A breakdown of the pools can be seen in Table 3.

$Lifetime Session Duration$ was another feature created to indicate the amount of time a player has played within each pool (in minutes). Along with playing frequency, knowing how long a user has played will give a better sense of exposure to ads for the app store. This feature was created to help identify sessions that are too short or suspiciously long. Too short of sessions could be the result of a user accidentally opening the app and closing it. Too long of sessions could indicate mistakes of leaving the app open or even a developer test account from PeopleFun themselves. These are not sessions we would like to capture in our model as they have a low chance of resulting in a mission played or purchase made which adds noise to our models.

$Converted Product Amount$ shows the total United States Dollars (USD) a player used within each pool in the Bricks ’n Balls app store. Understanding who is spending money and how much they are spending will help establish a pattern between the users’ interactions with the app and their purchasing habits.
The Count Converted Product Amount shows the number of transactions a user has made during the pools. Tying this feature to Converted Product Amount yields a user’s price range for products. A user with a high count but low Converted Product Amount means they prefer cheaper items. This feature was included in the data to show how frequently or infrequently users have made purchases over the pools.

The Max Mission feature shows the most recent completed mission for each user during the pools. The intuition is that this adds more information on player engagement. The max completed mission during pools includes information on how quickly or slowly players are progressing during the pools. Additionally, it allows the model to learn about how far certain users have progressed. If a user has been stuck on a high level over the pools, they may be more likely to purchase in the future.

Bricks N Balls records every user’s interaction within the game including but not limited to when a user has opened an app or closed the app. Therefore, it was important to extract the number of sessions per user based on whether the user has mission completed, mission started, mission opened or mission closed for mode type Saga during the pool (missionCompleted, missionClosed, missionStarted, missionOpened). Saga was chosen because this game mode type is most frequently played by users. This feature is called Sum Sessions.

The Total Events feature is the actual number of events during the pool. The Average Number of Events is the total number of sessions (Sum Sessions) divided by the total number of events (Total Number of Events). These features indicate more information on player gaming habits over the pools.

Therefore, each of these seven features are there for each of the four pools which resulted in a total of 28 input features. The output feature is called the NovSumConvertedProductAmount which is defined as the total amount of dollars spent during November.

Table 3. Pool Date Breakdown

<table>
<thead>
<tr>
<th>Pool ID</th>
<th>October Date Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10/1/2021 – 10/8/2021 (8 Days)</td>
</tr>
<tr>
<td>2</td>
<td>10/9/2021 – 10/15/2021 (7 Days)</td>
</tr>
<tr>
<td>3</td>
<td>10/16/2021 – 10/22/2021 (7 Days)</td>
</tr>
<tr>
<td>4</td>
<td>10/23/2021 – 10/31/2021 (9 Days)</td>
</tr>
</tbody>
</table>

Table 4. Features Created for Modelling

<table>
<thead>
<tr>
<th>Input Features</th>
<th>Data Type</th>
<th>Description</th>
<th>Example (User B): For Pool 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count Converted</td>
<td>Integer</td>
<td>Number of Transactions Made During Each Pool</td>
<td>6</td>
</tr>
<tr>
<td>Product Amount</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum Converted Product Amount</td>
<td>Dollars</td>
<td>Total Dollars Spent During Each Pool</td>
<td>$241.38</td>
</tr>
<tr>
<td>Lifetime Session</td>
<td>Minutes</td>
<td>Minutes Played During Each Pool</td>
<td>10,555</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2 Supervised Learning Methods

Building a purchase prediction model required significant feature engineering to include user’s past behaviors as features. Each user’s input features consist of past recency, frequency, and monetary value for each pool. The goal of adding these features was for the models to use the past behavior to predict the target. The target is the total amount of dollars spent during the entire month of November. This resulted in a total of 42,148 users who have made at least one purchase during the month of October 2021. Looking at Table 6 below, only the top 5% of users have purchased over $88.03 and made over 16 purchases. There exists a single user that has made about 3,100 dollars and 378 transactions. Therefore, the data is significantly right skewed, and these top purchasers should be removed from the model. Our study removed this subset of users because these users are already purchasing significantly more than the overall player. The purchase prediction model will have more utility for predicting how much users will spend for those users that are less loyal. The top 5% of players consist of 2,108 users that were removed from the dataset. The data resulted in 40,040 observations for modeling proceeding forward. For modelling, we created a hold-out validation set of 10% or 4,004 users from the training data. Therefore, the training data consists of 36,036 users. The study used Sci-Kit-Learn’s ShuffleSplit with 10-fold cross-validation and test size of 20% to reduce bias and overfitting. All models were built and tested using Sklearn’s API.

Table 4. Percentiles of High Purchasers.

<table>
<thead>
<tr>
<th>Total Dollars Spent During October ($)</th>
<th>Number of Transactions Made in October</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>22.44</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>65.77</td>
</tr>
</tbody>
</table>
5 Results

5.1 Player Segmentation via RFM Analysis

Upon application of clustering techniques, the K-Means model with 3 clusters utilizing min-max scaled RFM data yielded the ideal Silhouette score, as shown in Table 7.

Table 6. Silhouette Scores of Clustering Algorithms

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Number of Clusters</th>
<th>Silhouette Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Means</td>
<td>3</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.558</td>
</tr>
<tr>
<td>K-Medoids</td>
<td>3</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.541</td>
</tr>
</tbody>
</table>

Exploring the cluster’s attributes of the original dataset, the clusters can be characterized by their features to gain further insight in their behavior. However, since our dataset is restricted to a single month of October 2021, these clusters should be approached with caution due to the establishment of player’s progression through the game.

5.2 Purchase Prediction Model Evaluation and Analysis

Table 7. Comparison of Model Performance Between Algorithms

<table>
<thead>
<tr>
<th>Learning Algorithm</th>
<th>Average Train RMSE over 10 folds</th>
<th>Validation RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>24.986</td>
<td>25.93</td>
</tr>
<tr>
<td>Support Vector Regression (SVR)</td>
<td>26.33</td>
<td>27.388</td>
</tr>
</tbody>
</table>
Our research can interpret each RMSE score as the average deviation between the predicted Sum Converted Amount in November to actual Sum Converted Amount in November (in dollars).

### 5.3 Hyper Parameters for Each Model

For the models below, a grid search was used for hyper-parameter turning to further minimize the RMSE for each machine learning algorithm.

<table>
<thead>
<tr>
<th>Table 8: Support Vector Regression</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>max_iter</td>
<td>10000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 9: Random Forest Regressor</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>max_depth</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>min_leaf_samples</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>n_estimators</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 10: Multilayer-Perceptron Regressor</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>hidden_layer_sizes</td>
<td>(256, 64)</td>
<td></td>
</tr>
<tr>
<td>learning_rate</td>
<td>adaptive</td>
<td></td>
</tr>
<tr>
<td>early_stopping</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>max_iter</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>activation</td>
<td>relu</td>
<td></td>
</tr>
<tr>
<td>solver</td>
<td>adam</td>
<td></td>
</tr>
<tr>
<td>batch_size</td>
<td>auto</td>
<td></td>
</tr>
</tbody>
</table>

### 5.4 Further Explanation on MLP’s Chosen Hyper-Parameters

Of the many models that were built, the MLP was the best performing model. The neural network uses two hidden layers. The first hidden layer consists of 256 neurons and the second hidden layer consists of 64 neurons where the activation function for both hidden layers is the Rectified Linear Unit (ReLU). After experimenting with several different hidden layer sizes, the hyper parameters used above performed well on the test set and better than the other machine learning algorithms on the validation set. The MLP may have performed better on this validation set, however, the MLP’s
drawback is the computational expensive and high bias. Furthermore, the MLP does not provide a significant reduction in RMSE over folds compared to Linear Regression.

5.6 Assessing Variable Importance

![Linear Regression Model Weights](image)

Figure 20. Linear Regression Model Weights

Looking at model performance, the results show that the Multilayer-Perceptron Regression performed the best on the validation set. However, linear regression, which may be considered a subpar model, performed fast and better than Support Vector and Random Forest Regression. These results may be surprising, but linear regression’s greatest strength compared to MLP Regression is interpretability. Linear regression clearly shows how each feature correlates with the target variable. The most influential features that are positively correlated with the target are the Sum Converted Product amounts for each pool. In fact, the highest correlated feature is the *Sum Converted Product Amount* for pool 4. In other words, the variable importance shows that the most recent amount spent during the last week of October is likely how much a unique user will spend in November. Furthermore, the *Max Mission* during pool 4 and pool 2 were negatively correlated with the target. The intuition for this feature’s negative correlation is the player’s game experience. It tends to be the case that most players in the early levels do not tend to make in-app purchases while players in much higher levels are much more likely to continue making purchases. Since the vast majority of players
6 Discussion

6.1 Importance

The importance of this research comes from the unique opportunity of using proprietary data from a mobile gaming studio. The video game industry values secrecy for intellectual property and data. Utilizing data can improve in-game experience, revenue, and reduce player churn rate.

Many gaming studios have adopted a freemium model which has proven to increase revenue in the long run. This research provides PeopleFun a deep insight into their player base, and their purchasing habits. Additionally, this research can be used to narrow down the number of in-app products offered. Furthermore, more personalized in-game products can be offered to players based on their buying habits. Predicting how much users spend will be a useful tool for PeopleFun to improve future revenue and resource management.

Churn prediction can be an issue in freemium games when users feel like they are being taken advantage of or are shown ads too much. With our research, adjustments can be made to Bricks ‘n Balls to give a better player experience without sacrificing revenue thus lowering churn and extending the game’s lifecycle.

6.2 Ethics

As the video game industry progresses so does their pay structure. One article, The changing face of desktop video game monetization: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010-2019, explored how these new pay structures have affected players. One ethical question of this research comes from the items offered in the Bricks ‘n Balls app store. Rubies can be purchased with real world currency and can be used within the app to purchase power-ups. One way of increasing purchases of rubies would be to increase the difficulty of a level to the point where a user would be forced to use a power up to complete the mission. If a user was out of the free power-ups they earn from simply playing the game, then they would be even more incentivized to purchase rubies for more power-ups. Intentional inflation of level difficulty could yield more sales but from an ethical perspective it would be a poor business practice. If a power up was needed then that level could be considered pay to win, a common term referring to video games where the more one pays, the better, they do.

The research conducted by Zendle, Ballou & Perales explored another ethical question that directly applies to the research conducted here [16]. Much like material products there is a target audience of mobile games. Since Bricks ‘n Balls is an E, For Everyone Game, their target audience is overly broad ranging from older children to senior adults. Advertising to children who are more susceptible to marketing and more likely to make a purchase has its own ethical ramifications. This ethical issue is mitigated by Bricks ‘n Balls by offering a completely free-to-play experience where no
in-app purchases are required. The in-game store allows players to have a choice of purchasing in-game products.

7 Conclusion

This paper explored the potential purchasing player base of an acquisitioned popular freemium mobile game and demonstrated the beginnings of purchase predictions. These contributions are fundamental building blocks to facilitating potential item offer and sales recommendations to an interested player base.

Furthermore, from the initial exploratory data analysis, the research concludes that a reduction of in-game store items is ideal to reduce choice overload. This reduction in the catalogue will allow PeopleFun to focus on a curated premium item for the player base. This game has been in the App Store for three years and the player base has been significantly established. The drawback of trying to predict how much users will spend in November is that the users we have built our purchase prediction model for are already habitual purchasers. Therefore, given the window of two months, it is difficult to see the variance of player purchasing behavior in the long run. Similar projects such as predicting customer lifetime value have access to at least six months of data for prediction.

In all, this research proved that clustering can be used successfully to segment players’ buying habits using the RFM framework. Additionally, this research showed it is possible to predict how much players will spend in the future. Certainly, the research is not perfect, but it does provide valuable insight into less explored areas like creating features from raw mobile gaming data into something that can be used for unsupervised and supervised learning.

References


