

# Data Exploration

## Data Set Overview

The table below lists each of the files available for analysis with a short description of what is found in each one.

| File Name      | Description  | Fields   |
|----------------|--|--|
| ad-clicks.csv  | <p>Ad Clicks Information</p> <p>A line is added to this file when a player clicks on an advertisement in the Flamingo app.</p>     | <p><b>timestamp</b>: when the click occurred.</p> <p><b>txID</b>: a unique id (within ad-clicks.log) for the click</p> <p><b>userSessionid</b>: the id of the user session for the user who made the click</p> <p><b>teamid</b>: the current team id of the user who made the click</p> <p><b>userid</b>: the user id of the user who made the click</p> <p><b>adID</b>: the id of the ad clicked on</p> <p><b>adCategory</b>: the category/type of ad clicked on</p>        |
| buy-clicks.csv | <p>In-App Purchase Information</p> <p>A line is added to this file when a player makes an in-app purchase in the Flamingo app.</p> | <p><b>timestamp</b>: when the purchase was made</p> <p><b>txID</b>: a unique id (within buy-clicks.log) for the purchase</p> <p><b>userSessionid</b>: the id of the user session for the user who made the purchase</p> <p><b>team</b>: the current team id of the user who made the purchase</p> <p><b>userid</b>: the user id of the user who made the purchase</p> <p><b>buyID</b>: the id of the item purchased</p> <p><b>price</b>: the price of the item purchased</p> |

|                             |   |  |
|-----------------------------|---|--|
|                             |   |  |
| <b>users.csv</b>            | <p>User Information</p> <p>This file contains a line for each user playing the game.</p>  | <p><b>timestamp</b>: when user first played the game</p> <p><b>id</b>: the user id assigned to the user</p> <p><b>nick</b>: the nickname chosen by the user</p> <p><b>twitter</b>: the twitter handle of the user</p> <p><b>dob</b>: the date of birth of the user</p> <p><b>country</b>: the two-letter country code where the user lives</p> |
| <b>team.csv</b>             | <p>Team Information</p> <p>This file contains a line for each team terminated in the game.</p>  | <p><b>teamid</b>: the id of the team</p> <p><b>name</b>: the name of the team</p> <p><b>teamCreationTime</b>: the timestamp when the team was created</p> <p><b>teamEndTime</b>: the timestamp when the last member left the team</p> <p><b>currentLevel</b>: the current level at which the team is playing</p>                               |
| <b>team-assignments.csv</b> | <p>Team Users Information</p> <p>A line is added to this file each time a user joins a team. A user can be in at most a single team at a time.</p>  | <p><b>time</b>: when the user joined the team</p> <p><b>team</b>: the id of the team</p> <p><b>userid</b>: the id of the user</p> <p><b>assignmentid</b>: a unique id for this assignment</p>  |
| <b>level-events.csv</b>     | <p>Team Level Information</p> <p>A line is added to this file each time a team starts or finishes a level in the game.</p>  | <p><b>time</b>: when the event occurred</p> <p><b>eventid</b>: a unique id for the event</p> <p><b>teamid</b>: the id of the team</p> <p><b>level</b>: the level started or completed</p> <p><b>eventType</b>: the type of event, either start or end</p>  |
| <b>user-session.csv</b>     | <p>User Session information</p> <p>Each line in this file describes a user session, which denotes when a user starts and stops playing the game. Additionally, when a team goes to the next level in the game, the session is</p> | <p><b>userSessionid</b>: a unique id for the session</p> <p><b>assignmentid</b>: the team assignment id for the user to the team</p> <p><b>startTimeStamp</b>: a timestamp</p>   |

|                        |  |  |
|------------------------|--|--|
|                        | ended for each user in the team and a new one started.   | denoting when the session started<br><br><b>endTimeStamp</b> : a timestamp denoting when the session ended<br><br><b>team_level</b> : the level of the team during this session<br><br><b>platformType</b> : the type of platform of the user during this session  |
| <b>game-clicks.csv</b> | <b>User Click Information</b><br><br>A line is added to this file each time a user performs a click in the game. | <b>time</b> : when the click occurred<br><br><b>clickid</b> : a unique id for the click<br><br><b>userid</b> : the id of the user performing the click<br><br><b>usersessionid</b> : the id of the session of the user when the click is performed<br><br><b>isHit</b> : denotes if the click was on a flamingo (value is 1) or missed the flamingo (value is 0)<br><br><b>teamId</b> : the id of the team of the user<br><br><b>teamLevel</b> : the current level of the team of the user |
|                        |  |  |

## Aggregation

This sections summarizes the key findings in the data.

From “buy-clicks.csv,” the following information can be calculated:

|  |       |
|--|-------|
| Amount spent buying items                        | 21407 |
| Number of unique items available to be purchased | 6     |

The following shows additional information about buy-clicks.csv:

- How many times items were bought: 2947
- Minimum price buyID value: 1.00
- Maximum price buyID Value: 20.00

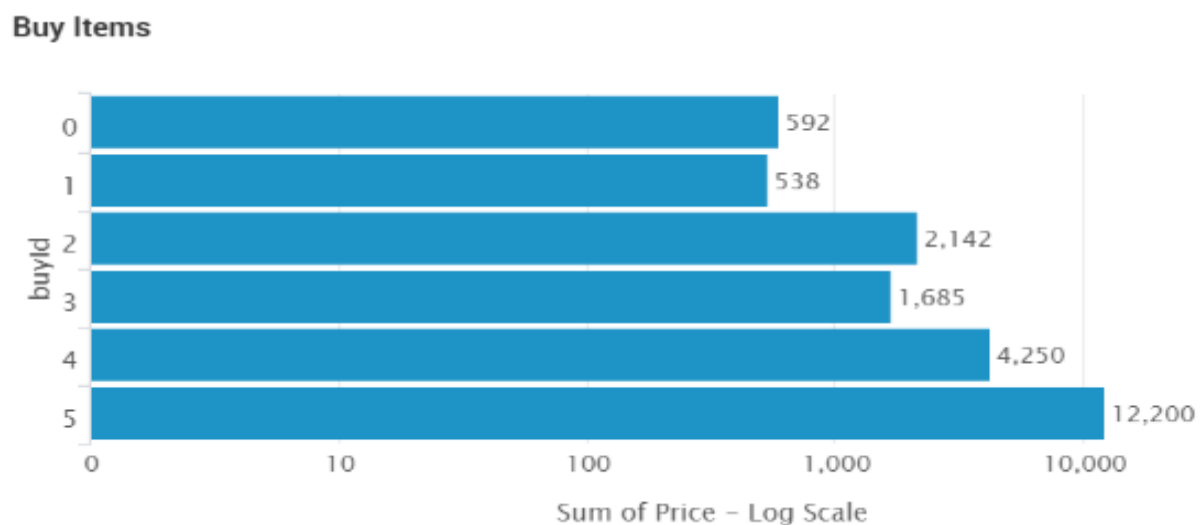
- Average price bought: 7.264

A histogram showing how many times each item is purchased:



A histogram showing how much money was made from each item:

The following Bar diagram shows the Revenues by Buy Items in a log scale.



## Summary – “buyID”

Item number 5 (price value is 20) is the most lucrative, but is not the most purchased.

The item most purchased is itemID 2 (price value is 3).

The Average buy is 7.26; much less than the maximum buyID value, which is 20.

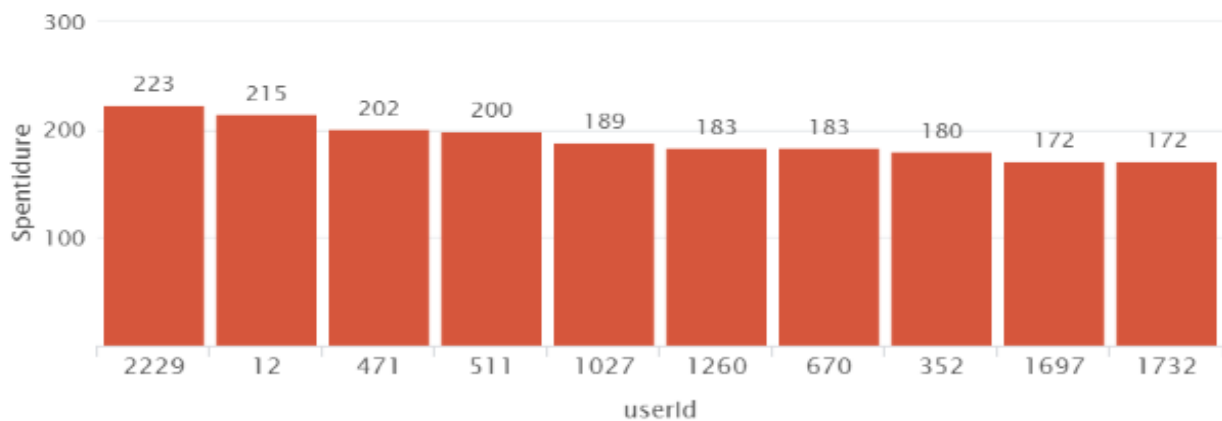
An increase in Average Buy value is achievable.

## Filtering

A histogram showing total amount of money spent by the top ten users (ranked by how much money they spent).

### 3. Filtering

Top 10 User Spent - Histogram



As one can see from the the above picture the UserId with most spending are:

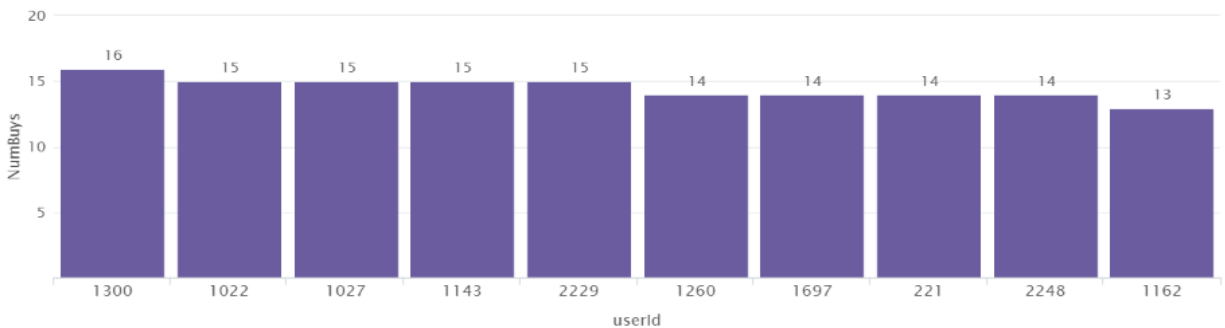
1. 2229
2. 12
3. 471

## Extra

Top 10 User Number of Buys - Histogram

### 3. Filtering Extra

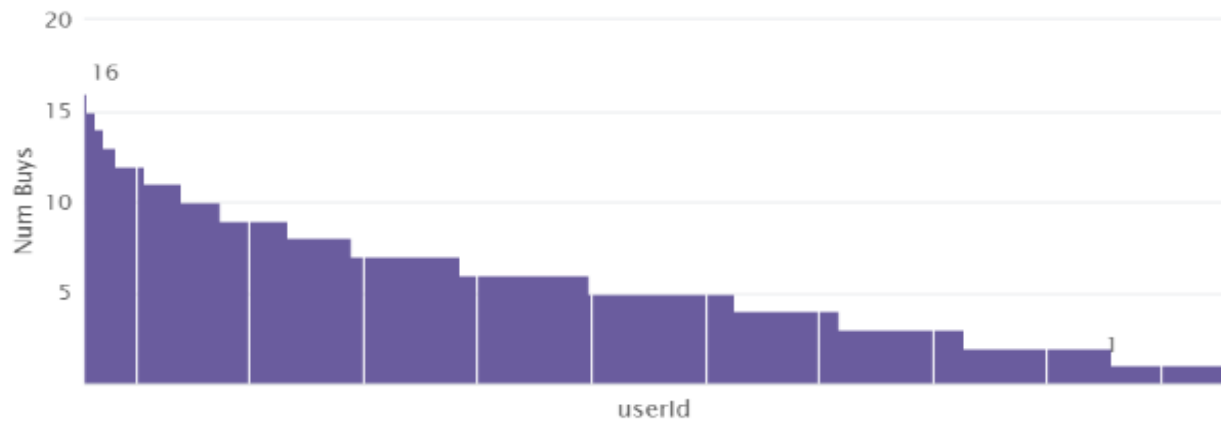
Top 10 User Number of Buys - Histogram



This means that the users having the most buys may be incentivized to buy higher priced items to increase revenue.

### 3- Filtering Extra

Histogram number of Buys by user



There is also a need to incentivize users that don't buy items. These should be converted to the Median, thus reducing the Standard Deviation.

**The Following Intermediate tables were used to populate the last table:**

#### Intermediate table to determine the Hit Ratio

3- Filtering - SessionID for Top 3 Spenders

#. Filtering - Intermediate - Know the Session ID

| userId | somaHit | contaHit | avgHit    | maxUSID | minUSID |
|--------|---------|----------|-----------|---------|---------|
| 12     | 92      | 704      | 13.068182 | 34557   | 7079    |
| 2229   | 61      | 526      | 11.596958 | 34357   | 5694    |
| 471    | 76      | 524      | 14.503817 | 34330   | 5863    |

This table was used to determine the Average Hit per Top 3 Spenders.

### Intermediate Table to determine the Platform

#### 3. Filtering - platFormType for Top 3 Spenders

| userid | platformType | CountuserSessionId |
|--------|--------------|--------------------|
| 12     | iphone       | 14                 |
| 2229   | iphone       | 14                 |
| 471    | iphone       | 14                 |

This table shows there is only one platform used by the Top 3 Spenders.

The following table shows the user id, platform, and hit-ratio percentage for the top three buying users:

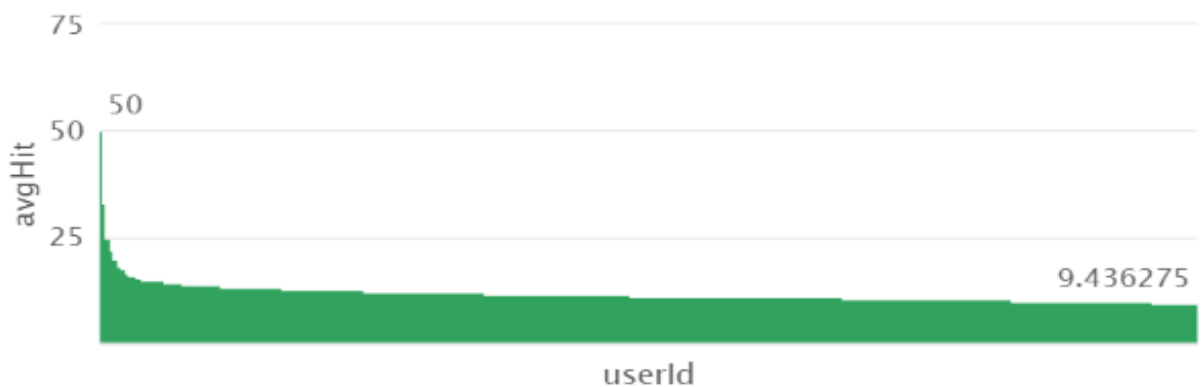
| Rank | User Id | Platform | Hit-Ratio (%) |
|------|---------|----------|---------------|
| 1    | 2229    | iphone   | 13,07         |
| 2    | 12      | iphone   | 11,6          |
| 3    | 471     | iphone   | 14,51         |

### Extra:

Hit Distribution

#### 3- Filtering Extra

##### Hit Accuracy Distribution



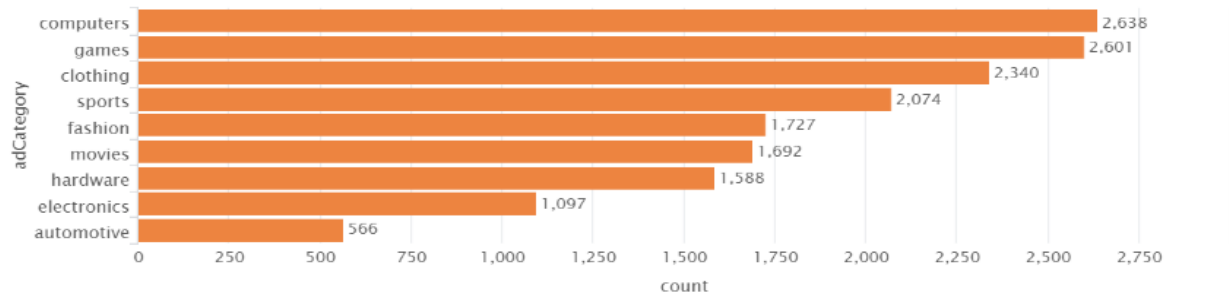
Conclusion: sell items to improve hit accuracy to the bellow average users.

## Category

There are 9 Ad Categories with the following Number of clicks:

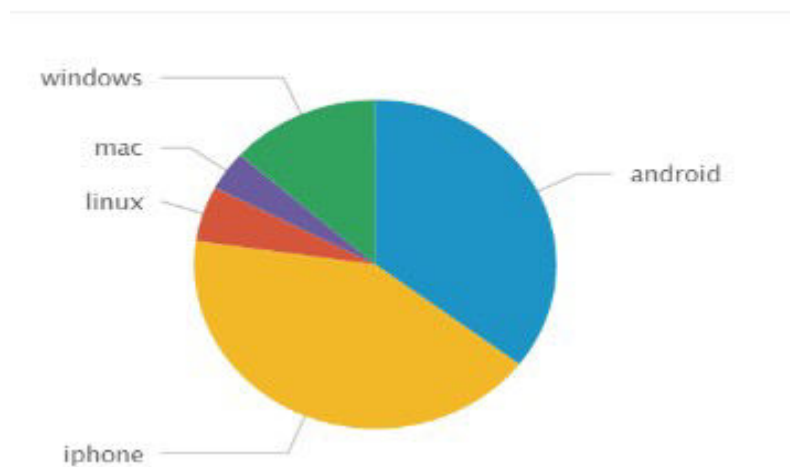
| Add Category Count for Revenue |                       |
|--------------------------------|-----------------------|
| adCategory                     | Count of Event Object |
| automotive                     | 566                   |
| clothing                       | 2340                  |
| computers                      | 2638                  |
| electronics                    | 1097                  |
| fashion                        | 1727                  |
| games                          | 2601                  |
| hardware                       | 1588                  |
| movies                         | 1692                  |
| sports                         | 2074                  |

ad Category Top



## Platform Type

The Iphone Platform has the most User Sessions



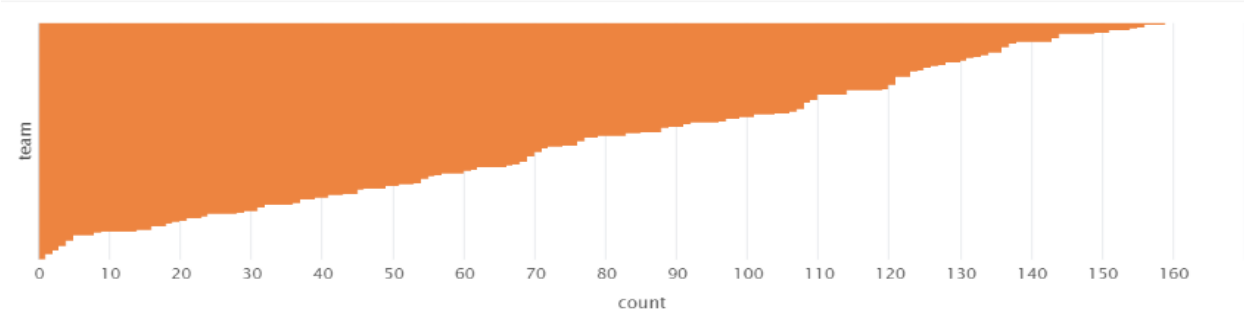


PlatformType

|   | platformType | Count of Event Object |
|---|--------------|-----------------------|
| 1 | iphone       | 3874                  |
| 2 | android      | 3274                  |
| 3 | windows      | 1240                  |
| 4 | linux        | 504                   |
| 5 | mac          | 358                   |

Team Assignment

TeamAssignment - Team Size



TeamAssignment - AVG Team Size KPI

Avg Team Size

78