

# Data Exploration

The objective of this document is to showcase , key findings from data exploration phase. Here the data into consideration is the data generated from the highly popular mobile game called “**Catch The Pink Flamingo**”. A short brief description about the game and its data structure , is followed. After which key finding and its analysis is documented.

## Catch The Pink Flamingo

The objective of the game is to catch as many Pink Flamingos as possible by following the missions provided by realtime prompts in the game and cover the map provided for each level. The levels get more complicated in mission speed and map complexity as the users move from level to level. Its a multi-user game where each user can be a member of a single team at any given time. Users are ranked based on their accuracy and speed and are categorized as “rising star”, “veteran”, “coach”, “social butterfly” and “hot flamingo”.Teams can be of size from one to thirty.Users are allowed in game purchases .Finally the game never ends, that is there will always be a more complicated next level.

## 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
<b>ad-clicks.csv</b>	An entry is added in this file when a player clicks on an advertisement.	<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>
<b>buyclicks.csv</b>	An entry is added to this file when a player makes an inapp purchase.	<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>	It contains a line for each user playing the game.	<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</p>

		<p>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>
--	--	--

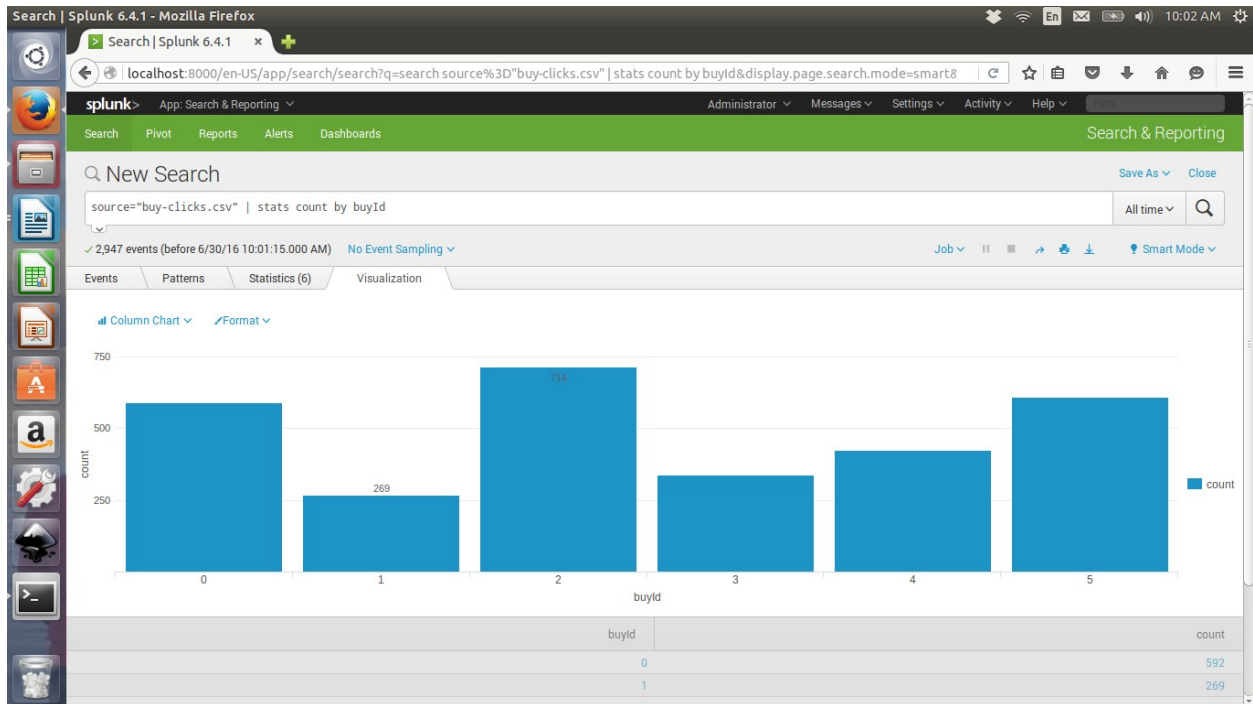
<b>team.csv</b>	This file contains a line for each team terminated in the game	<b>teamid</b> : the id of the team name: the name of the team  <b>teamCreationTime</b> : the timestamp when the team was created  <b>teamEndTime</b> : the timestamp when the last member left the team  <b>strength</b> : a measure of team strength, roughly corresponding to the success of a team  <b>currentLevel</b> : the current level of the team
<b>team-assignments.csv</b>	An entry 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.	<b>time</b> : when the user joined the team. team: the id of the team  <b>userid</b> : the id of the user  <b>assignmentid</b> : a unique id for this assignment
<b>level-events.csv</b>	An entry is added to this file each time a team starts or finishes a level in the game	<b>time</b> : when the event occurred. eventid: a unique id for the event  <b>teamid</b> : the id of the team  <b>level</b> : the level started or completed  <b>eventType</b> : the type of event, either start or end
<b>user-session.csv</b>	Each line in this file describes a user session, which denotes when a user starts and stops playing the game.	<b>timeStamp</b> : a timestamp denoting when the event occurred. userSessionId: a unique id for the session.  <b>userId</b> : the current user's ID.  <b>teamId</b> : the current user's team.  <b>assignmentId</b> : the team

		<p>assignment id for the user to the team.</p> <p><b>sessionType</b>: whether the event is the start or end of a session.</p> <p><b>teamLevel</b>: the level of the team during this session.</p> <p><b>platformType</b>: the type of platform of the user during this session.</p>
<b>game-clicks.csv</b>	An entry is added to this file each time a user performs a click in the game.	<p><b>time</b>: when the click occurred.  <b>clickid</b>: a unique id for the click.</p> <p><b>userid</b>: the id of the user performing the click.</p> <p><b>usersessionid</b>: the id of the session of the user when the click is performed.</p> <p><b>isHit</b>: denotes if the click was on a flamingo (value is 1) or missed the flamingo (value is 0)</p> <p><b>teamId</b>: the id of the team of the user</p> <p><b>teamLevel</b>: the current level of the team of the user</p>

## Aggregation

Amount spent buying items	21407.0
Unique items available to be purchased	6

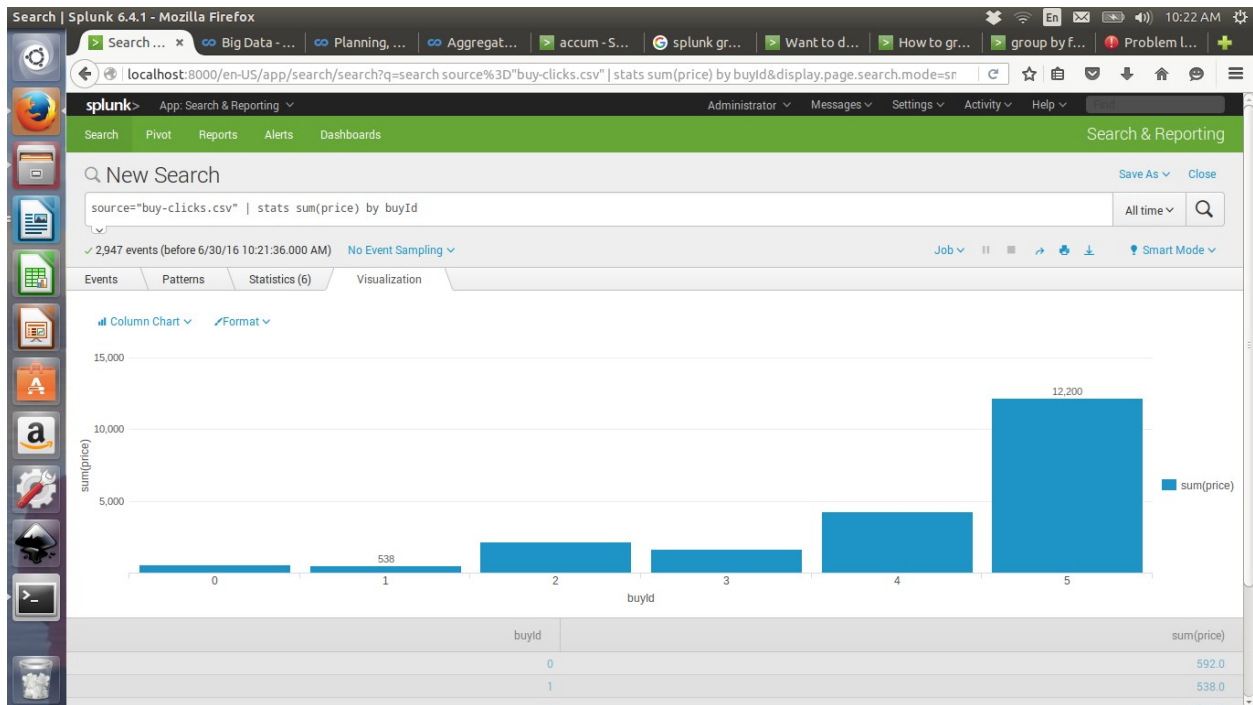
A histogram showing showing frequency of each item being purchased.



## Analysis

1. Item2 is purchased most often , 714 times.Followed by Item5 and Item0.

Below given histogram, depicts money made from each item:

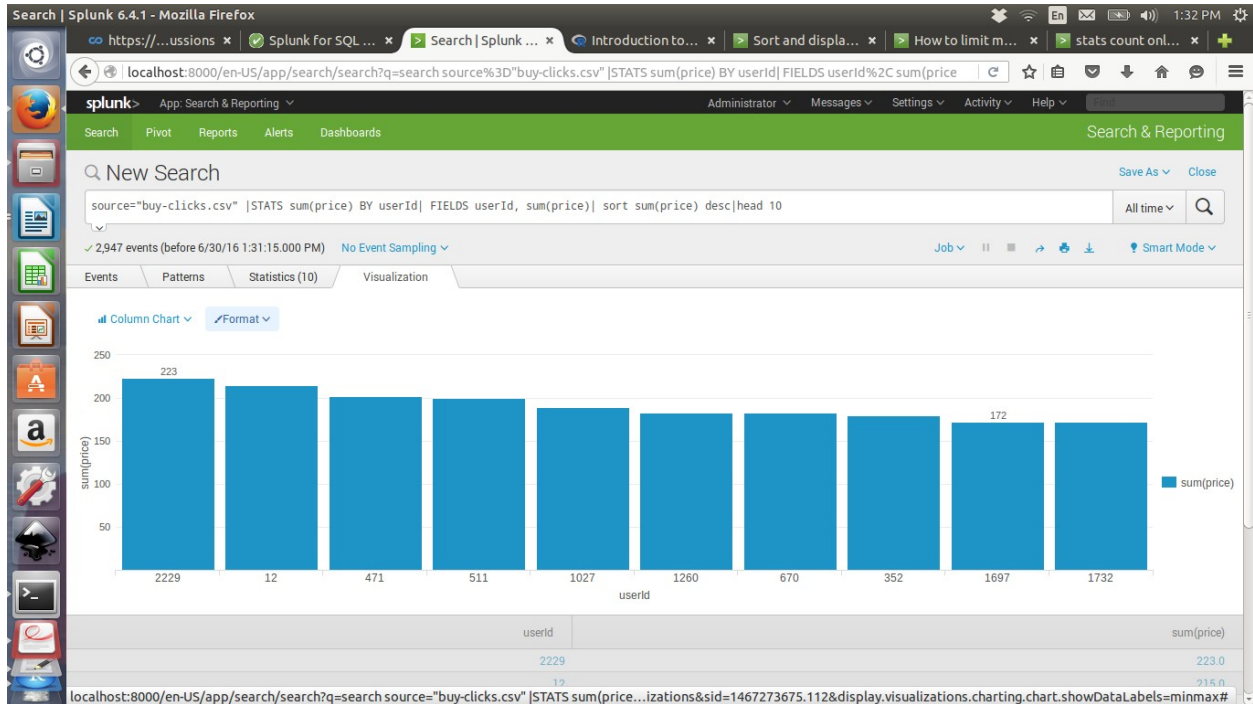


## Analysis

1. Item 5 makes the maximum money with the total of about 12,200.
2. Item 0 and Item 1 contribute the least, less than 3% of the total purchase done.
3. Though Item 2 is purchased most often, its price is just 3 as compared to item 5 which is 20. Hence while considering contribution in terms of total price amount, Item 5 stands out.

## Filtering

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



## Analysis :

1. The top buyer spends around 223.
2. Top 10<sup>th</sup> buyer spends around 172.
3. Doing further analysis , one can infer the variation among the top ten buyers is less as compared to variation across the data.



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	11.596
2	12	iphone	13.068
3	471	iphone	14.503

## Analysis

1. Hit-Ratio varies from 0 to 50%. with a mean of 11.09% and median at 10.95%.Three top users in Hit-Ratio have hit ratio more than 25% where as rest have below 25%.Hence can be interpreted as there is a large difference in hit ratio between top three users (in hit-ratio) and others.
2. 75% of times the platform used is iphone or android.Linux and mac platform is not much used for the game. Its recommened , that we further analyze the issue as to why there is such a huge variation based on the platform.
3. From the above table , one can conclude that top buyers in the game are not the ones with best hit ratio.No relationship between hit-Ratio and buying patterns.