

# Data Exploration

*Coursera Big Data Specialization - Week 1 Peer Graded Review Assignment*

## *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	Database of clicks on ads	<p>timestamp: when the click occurred.</p> <p>txId: a unique id (within ad-clicks.log) for the click</p> <p>userSessionid: the id of the user session for the user who made the click</p> <p>teamid: the current team id of the user who made the click</p> <p>userid: the user id of the user who made the click</p> <p>adId: the id of the ad clicked on</p> <p>adCategory: the category/type of ad clicked on</p>
buy-clicks.csv	Database of purchases.	<p>timestamp: when the purchase was made.</p> <p>txId: a unique id (within buy-clicks.log) for the purchase</p> <p>userSessionId: the id of the user session for the user who made the purchase</p>

		<p>team: the current team id of the user who made the purchase</p> <p>userId: the user id of the user who made the purchase</p> <p>buyId: the id of the item purchased</p> <p>price: the price of the item purchased</p>
game-clicks.csv	A record of each click a user performed during the game.	<p>timestamp: when the click occurred.</p> <p>clickId: a unique id for the click.</p> <p>userId: the id of the user performing the click.</p> <p>userSessionId: the id of the session of the user when the click is performed.</p> <p>isHit: denotes if the click was on a flamingo (value is 1) or missed the flamingo (value is 0)</p> <p>teamId: the id of the team of the user</p> <p>teamLevel: the current level of the team of the user</p>
level-events.csv	A record of each level event for a team. Level events are recorded when a team ends or begins a new level	<p>timestamp: when the event occurred.</p> <p>eventId: a unique id for the event</p> <p>teamId: the id of the team</p>

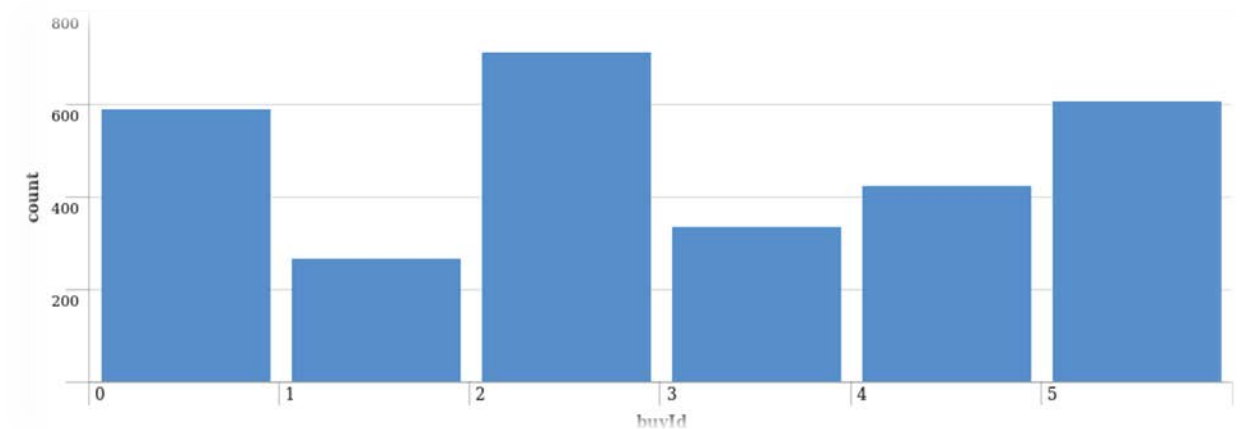
		<p>teamLevel: the level started or completed</p> <p>eventType: the type of event, either start or end</p>
team-assignments.csv	A record of each time a user joins a team.	<p>timestamp: when the user joined the team.</p> <p>team: the id of the team</p> <p>userId: the id of the user</p> <p>assignmentId: a unique id for this assignment</p>
team.csv	A record of each team in the game.	<p>teamId: the id of the team</p> <p>name: the name of the team</p> <p>teamCreationTime: the timestamp when the team was created</p> <p>teamEndTime: the timestamp when the last member left the team</p> <p>strength: a measure of team strength, roughly corresponding to the success of a team</p> <p>currentLevel: the current level of the team</p>
user-session.csv	<p>A record of each session a user plays.</p> <p>When a team levels up, each current user session ends and a</p>	<p>timestamp: a timestamp denoting when the event occurred.</p>

	<p>new session begins with the new level.</p>	<p>userSessionId: a unique id for the session.</p> <p>userId: the current user's ID.</p> <p>teamId: the current user's team.</p> <p>assignmentId: the team assignment id for the user to the team.</p> <p>sessionType: whether the event is the start or end of a session.</p> <p>teamLevel: the level of the team during this session.</p> <p>platformType: the type of platform of the user during this session.</p>
users.csv	Database of the game users	<p>timestamp: when user first played the game.</p> <p>userId: the user id assigned to the user.</p> <p>nick: the nickname chosen by the user.</p> <p>twitter: the twitter handle of the user.</p> <p>dob: the date of birth of the user.</p> <p>country: the two-letter country code where the user lives.</p>

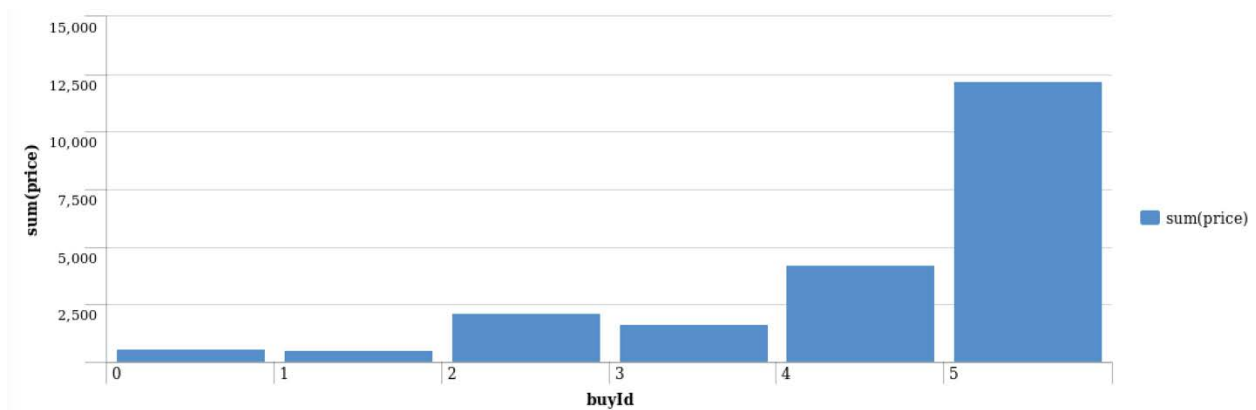
## Aggregation

Amount spent buying items	\$ 21407
Number of unique items available to be purchased	6

A histogram showing how many times each item is purchased:

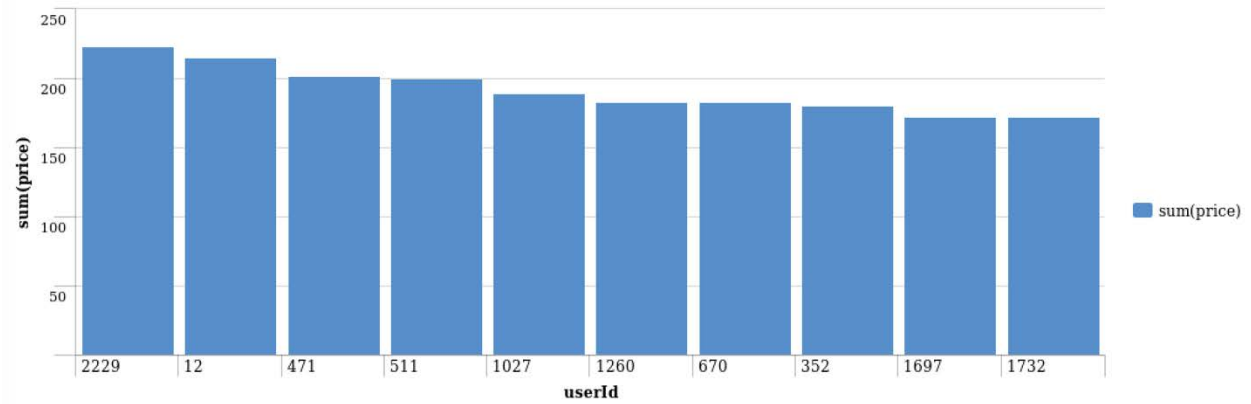


A histogram showing how much money was made from each item:



## Filtering

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



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	11.5%
2	12	iPhone	13%
3	471	iPhone	14.5%