**Data Exploration**

Cousera Big Data Specialization – Week 1 Peer Graded Review Assignment

**Data Set Overview**

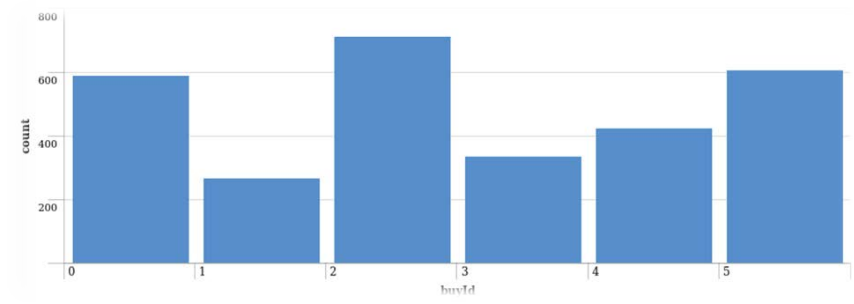
Here it lists all the files available for analysis with a brief description

|  |  |  |
| --- | --- | --- |
| **File Name** | **Description** | **Fields** |
| Ad-clicks.csv | Database of clicks on ads | **timestamp**: when the click occurred.  **txId**: a unique id (within adclicks.log) for the click.  **userSessionid**: the id of the user session for the user who made the click.  **teamid**: the current team id of the user who made the click.  **userid**: the user id of the user who made the click.  **adId**: the id of the ad clicked on.  **adCategory**: the category/type of ad clicked on. |
| Buy-clicks.csv | Database of purchases | **timestamp**: when the purchase was made.  **txId**: a unique id (within buyclicks.log) for the purchase.  **userSessionId**: the id of the user session for the user who made the purchase.  **team**: the current team id of the user  who made the purchase.  **userId**: the user id of the user who  made the purchase.  **buyId**: the id of the item purchased.  **price**: the price of the item purchased. |
| Game-clicks.csv | A record of each click a user performed during the game | **timestamp**: when the click occurred.  **clickId**: a unique id for the click.  **userId**: the id of the user performing the click.  **userSessionId**: the id of the session of the user when the click is performed.  **isHit**: denotes if the click was on a flamingo (value is 1) or missed the flamingo (value is 0)  **teamId**: the id of the team of the user **teamLevel**: the current level of the team of the user |
| 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 | **timestamp**: when the event occurred. **eventId**: a unique id for the event  **teamId**: the id of the team  **teamLevel**: the level started or completed **eventType**: the type of event, either start or end |
| Team-assignments.csv | A record of each time a user joins a team | **timestamp:** when the user joined the  team.  **team:** the id of the team  **userId:** the id of the user  **assignmentId:** a unique id for this  assignment |
| Team.csv | Record of each team in the game | **teamId**: the id of the team  **name**: the name of the team  **teamCreationTime:** the timestamp  when the team was created  **teamEndTime**: the timestamp when  the last member left the team  **strength:** a measure of team  **strength,** roughly corresponding to  the success of a team  **currentLevel:** the current level of the  team |
| User-session.csv | A record of each session a user plays.  When a team levels up, each current user session ends and a new session begins with the new level. | **timestamp**: a timestamp denoting  when the event occurred.  **userSessionId**: a unique id for the  session.  **userId**: the current user's ID.  **teamId**: the current user's team.  **assignmentId**: the team assignment  id for the user to the team.  **sessionType**: whether the event is  the start or end of a session.  **teamLevel**: the level of the team  during this session.  **platformType**: the type of platform of  the user during this session. |
| Users.csv | Database of the game users | **timestamp**: when user first played  the game.  **userId**: the user id assigned to the  user.  **nick**: the nickname chosen by the  user.  **twitter**: the twitter handle of the user.  **dob**: the date of birth of the user.  **country**: the two-letter country code  where the user lives. |

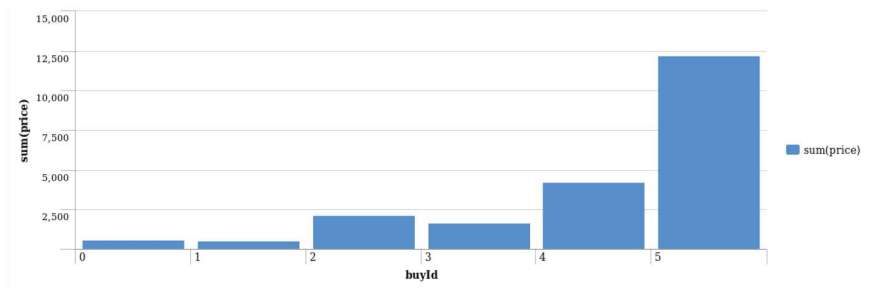
**Aggregation**

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

A histogram showing the quantity of purchases per time

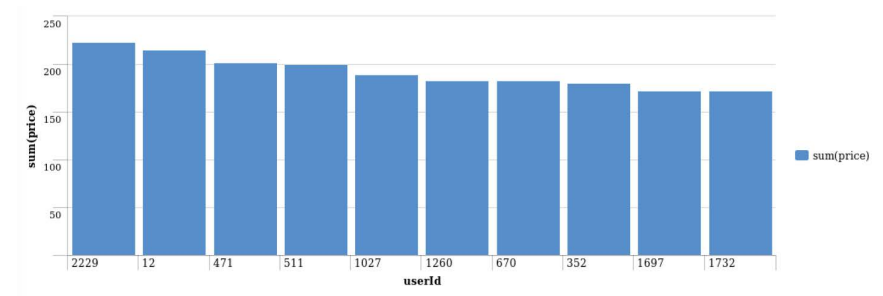


A histogram showing quantity of money made per item

****

**Filtering**

A histogram showing the total of money spent by the top ten users



The following table shows the top three buying uses ranking:

|  |  |  |  |
| --- | --- | --- | --- |
| **Rank** | **User Id** | **Platform** | **Hit-Ratio** |
| 1 | 229 | iphone | 11.5% |
| 2 | 12 | iphone | 13% |
| 3 | 471 | iphone | 14.5% |