Data Exploration

Data Set Overview

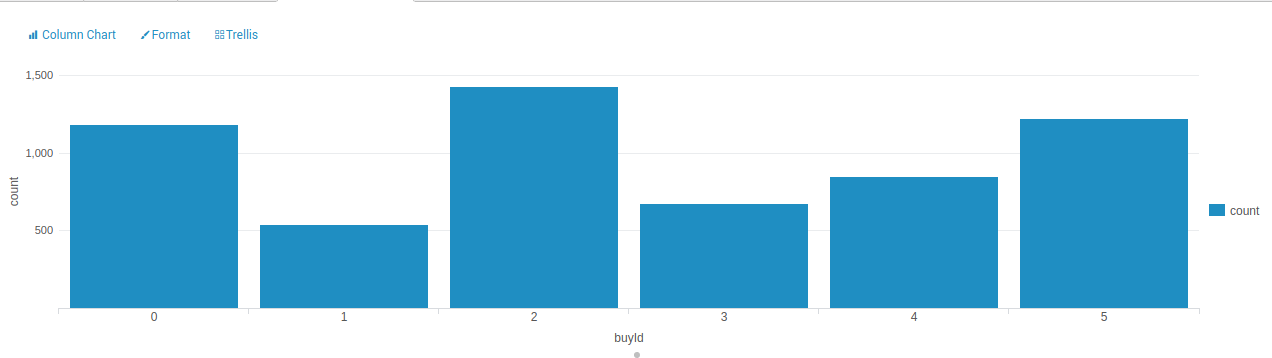
The table below lists each of the files available for analysis with a short description of what is found in each one.

|  |  |  |
| --- | --- | --- |
| **File Name** | **Description** | **Fields** |
| ad-clicks.csv | This table has accumulated all clicks of the users on an advertisement in the app | **timestamp**: the timestamp when the event occurs.  **txId**: a unique id for the click, its a primary key  **userSessionid**: the id of the user session, its a foreing key from the table User\_sessions  **teamid**: the current team id, its a foreing key from the table Team  **userid**: the user id of the user who made the click, its a foreing key from the table User  **adId**: the id of the ad clicked on. This id must be in other table wich have the id of the adds.  **adCategory**: the category/type of ad clicked on. This attribute is an enumerated type. |
| buy-clicks.csv | This table have one row for each purchase in the app | **timestamp**: the timestamp when the event occurs  **userSessionId**: the id of the user session for the user who made the purchase. its a foreing key from the table User\_sessions  **team**: the current team id of the user who made the purchase, its a foreing key from the table Team (teamId)  **userId**: the user id of the user who made the purchase, its a foreing key from the table User  **buyId**: the id of the item purchased, its a primary key  **price**: the price of the item purchased |
| users.csv | This table contains the players of the game | **timestamp**: when user first played the game.  **userId**: the user id assigned to the user. its a primary key. Numeric Id  **nick**: the nickname chosen by the user.  **twitter**: the twitter handle of the user.  **dob**: the date of birth of the user. In this format AAAA­MM­DD  **country**: the two­letter country code where the user lives. |
| team.csv | This table contains all the teams in the game. | **teamId**: the id of the team, its a primary key. Numeric Id  **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 |
| team-assignments.csv | Each row contains when a user join a team. When one user join a new team it indicates that leave the last team. | **timestamp**: when the user joined the team.  **team**: the id of the team, foreing key  **userId**: the id of the user, foreing key  **assignmentId**: a unique id for this assignment, primary key. |
| level-events.csv | This table contains all the level events, when a team start an event on when finish it. | **timestamp**: when the event occurred.  **eventId**: a unique id for the event, primary key  **teamId**: the id of the team, foreing key  **teamLevel**: the level started or completed. Its grater than 0  **eventType**: the type of event, either start or end. Enumerated value. |
| user-session.csv | 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 ended for  each user in the team and a new one started. | **timestamp**: a timestamp denoting when the event occurred.  **userSessionId**: a unique id for the session. Primary key  **userId**: the current user's ID. Foreing key  **teamId**: the current user's team. Foreing key  **assignmentId**: the team assignment id for the user to the team. Foreing key  **sessionType**: whether the event is the start or end of a session. Enumerated type  **teamLevel**: the level of the team during this session. Positive integer  **platformType**: the type of platform of the user during this session. |
| game-clicks.csv | A line is added to this file each time a user performs a click in the game. | **timestamp**: when the click occurred.  **clickId**: a unique id for the click. Primary key  **userId**: the id of the user performing the click. Foreing key  **userSessionId**: the id of the session of the user when the click is performed. Foreing key  **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, Foreing key  **teamLevel**: the current level of the team of the user |

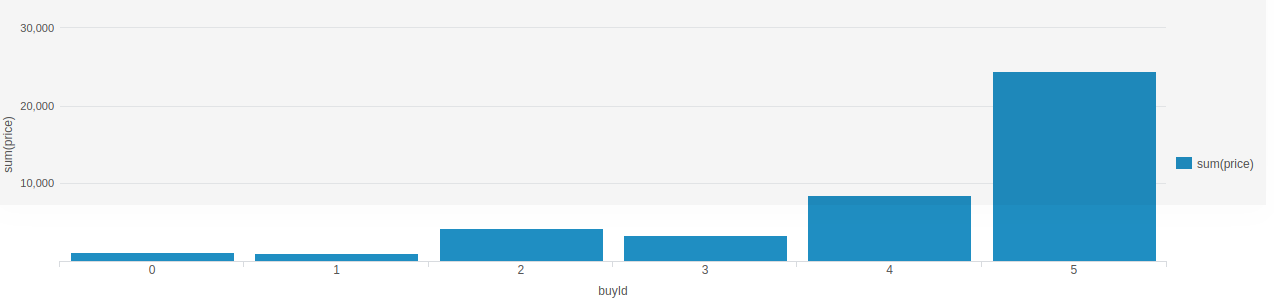
Aggregation

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

* A histogram showing how many times each item is purchased:

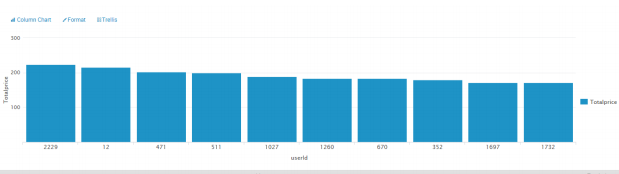
 the item 2 is the top shell item

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

The item 5 is the top of how much money spent the players

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 | 11.59% |
| 2 | 12 | iphone | 13.06% |
| 3 | 471 | iphone | 14.50% |