

Data Exploration using Splunk

Introduction

This data exploration is based on a mobile game called “**Catch The Pink Flamingo**”.

The objective of the game is to catch as many Pink Flamingos as possible by following the missions provided by real-time 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.

It's a **multi-user game** where the players have to catch Pink Flamingos that randomly pop up on a gridded world map based on missions that change in real-time. For the player or team to move to the next complexity level, they need to have at least one point in every map grid cell, i.e., cover the whole world map. An example mission would be “Catch the Flamingos on land with stars on their belly” in which the player should only click on flamingos that match the mission criteria, in this case, stars and being on land. If the player tags any other flamingo on the map, he/she or his/her team gets a negative point (-1) on that map location.

Importance of using splunk

Splunk can ingest and process data in real-time, allowing for immediate analysis and reaction to critical events and it can aggregate logs from various sources (e.g., servers, applications, network devices) into a centralized platform, making it easier to manage and analyze. Splunk can scale from small implementations to large, enterprise-wide deployments, accommodating growing data volumes and user bases. It indexes data, making it searchable and enabling fast retrieval of relevant information.

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
users.csv	This file contains a line for each user playing the game.	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.

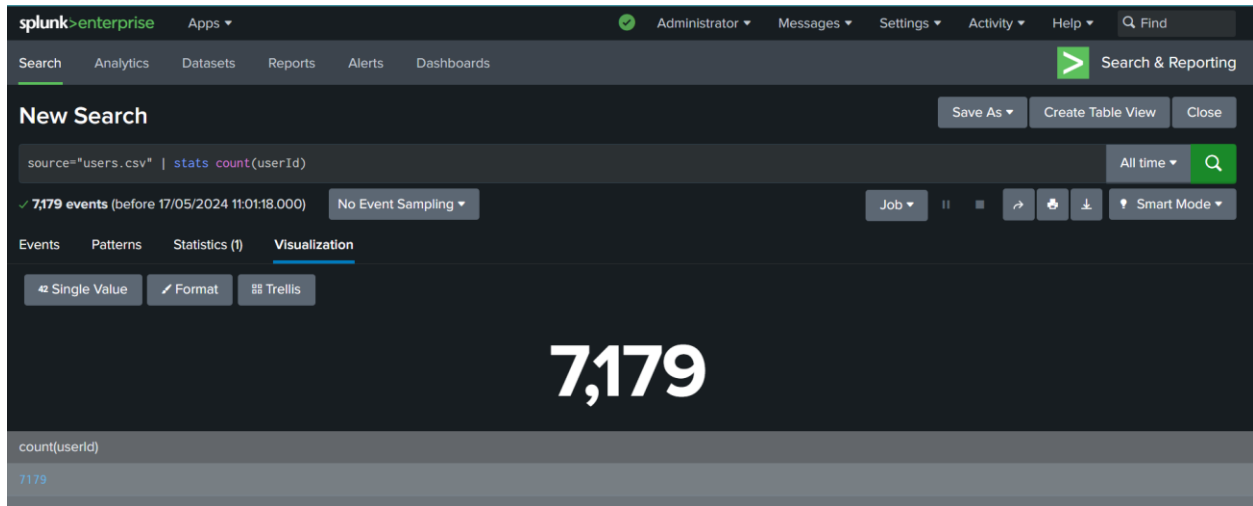
adclicks.csv	A line is added to this file when a player clicks on an advertisement in the Flamingo app.	<p>timestamp: when the click occurred.</p> <p>txId: a unique id (within ad-clicks.log) for the click.</p> <p>userSessionId: the id of the user session for the user who made the click.</p> <p>teamId: 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 a d clicked on.</p>
team.csv	This file contains a line for each team terminated 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>
team-assignments.csv	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>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>
levelevents.csv	A line is added to this file each time a team starts or finishes a level in the game	<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>
usersession.csv	Each line in this file describes a user session, which denotes when a user starts and stops playing the game. Additionally, when a te	<p>timestamp: a timestamp denoting when the event occurred.</p> <p>userSessionId: a unique id for the session.</p>

	am goes to the next level in the game, the session is ended for each user in the team and a new one started.	<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>
gameclicks.csv	A line is added to this file each time a user performs a click in 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>
buy-clicks.csv	A line is added to this file when a player makes an in-app purchase in the Flamingo app.	<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>

The data sets above are a generated dataset by the University of California San Diego.
Important:- Datasets are not containing missing values.

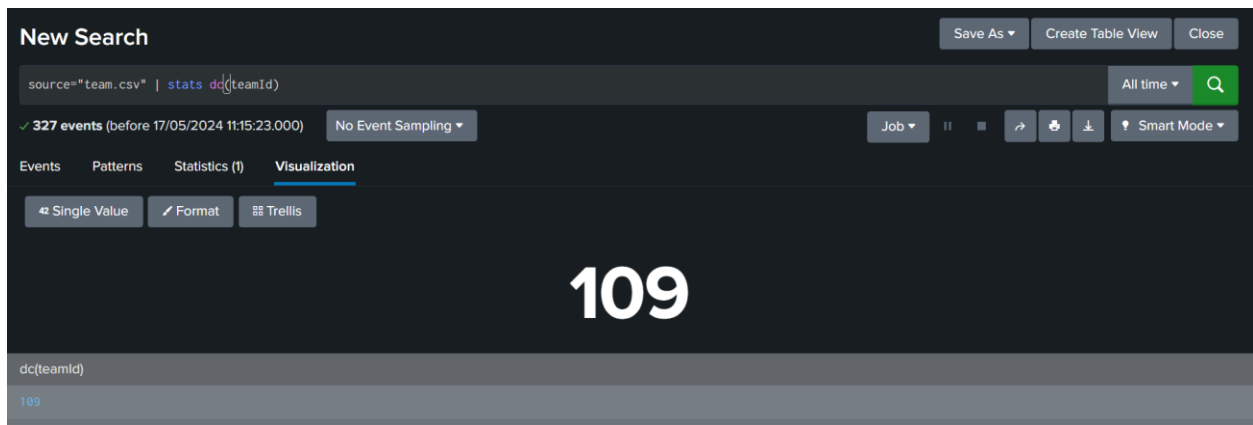
Data Exploration on 8 separate datasets based on the computer game “catch pink flamingo”, on Splunk.

1. Total number of players registered



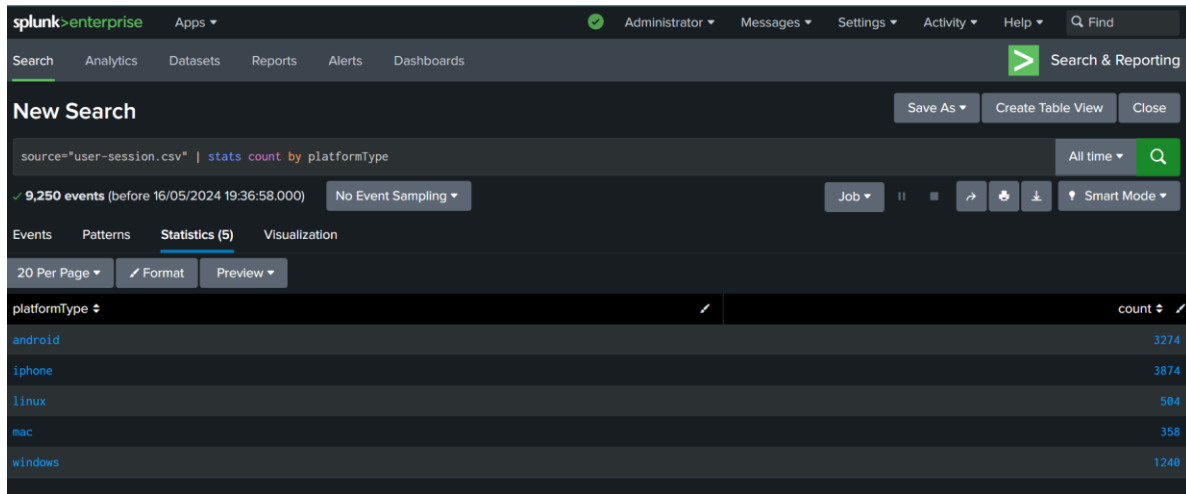
7179 players have already registered for the game.

2. Total number of teams registered



There are 109 teams registered.

3. Operating Systems that are used by the users

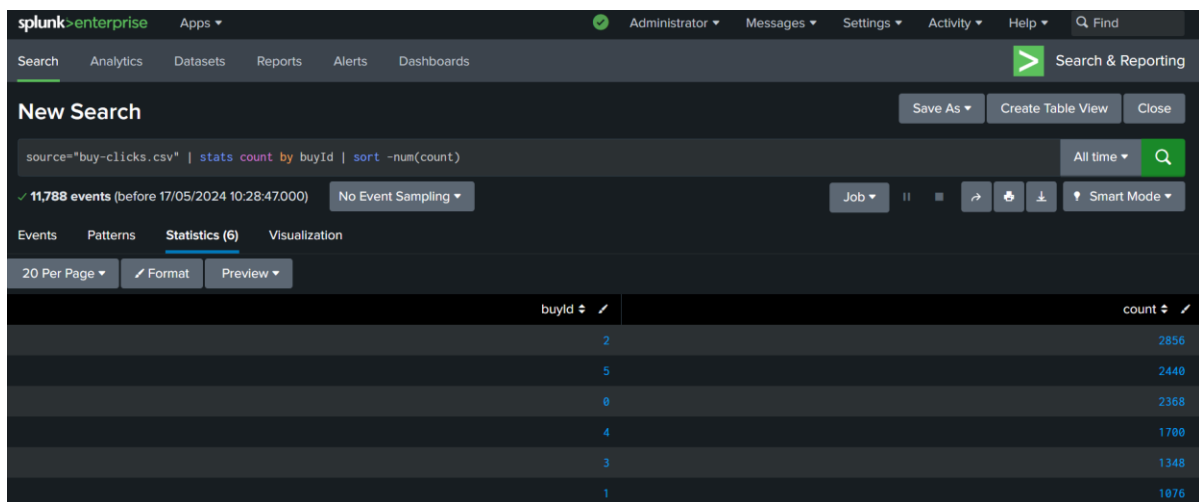


The screenshot shows the Splunk Enterprise interface with a search for "source='user-session.csv' | stats count by platformType". The search results are displayed in a table with two columns: platformType and count. The results are sorted by count in descending order.

platformType	count
android	3274
iphone	3874
linux	504
mac	358
windows	1240

Most of the players of the game have used iPhones and in second place there is Android. The lowest number of game players used Mac'.

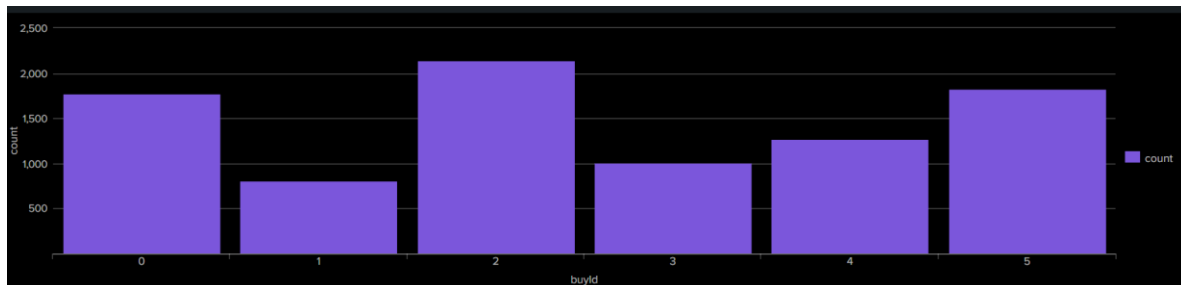
4. Most commonly purchased products



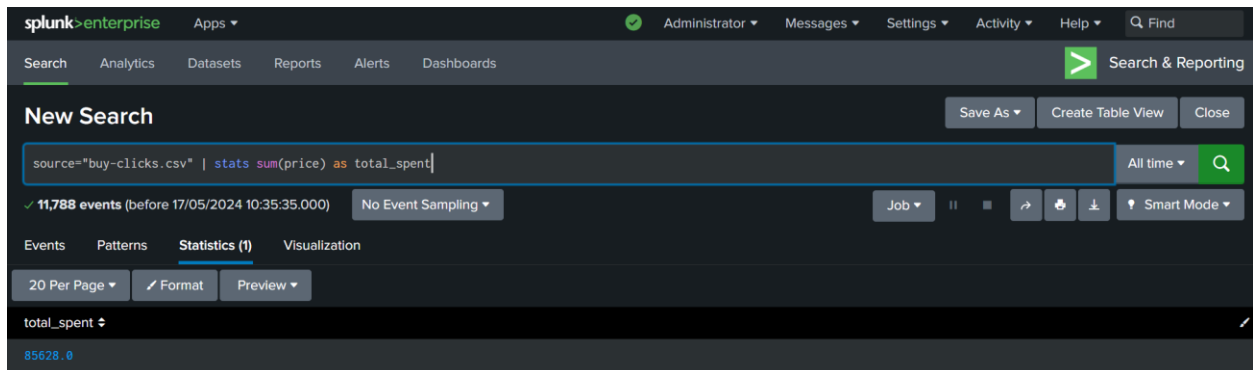
The screenshot shows the Splunk Enterprise interface with a search for "source='buy-clicks.csv' | stats count by buyId | sort -num(count)". The search results are displayed in a table with two columns: buyId and count. The results are sorted by count in descending order.

buyId	count
2	2856
5	2440
0	2368
4	1700
3	1348
1	1076

The most commonly purchased item is the item that has buyID = "2".

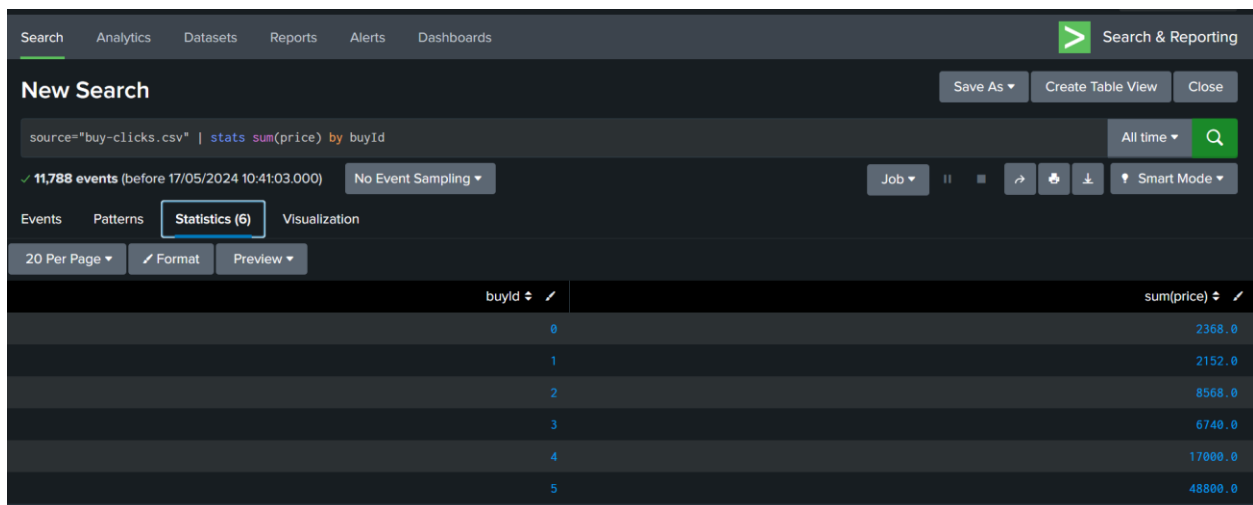


5. Total spent by the users on the products

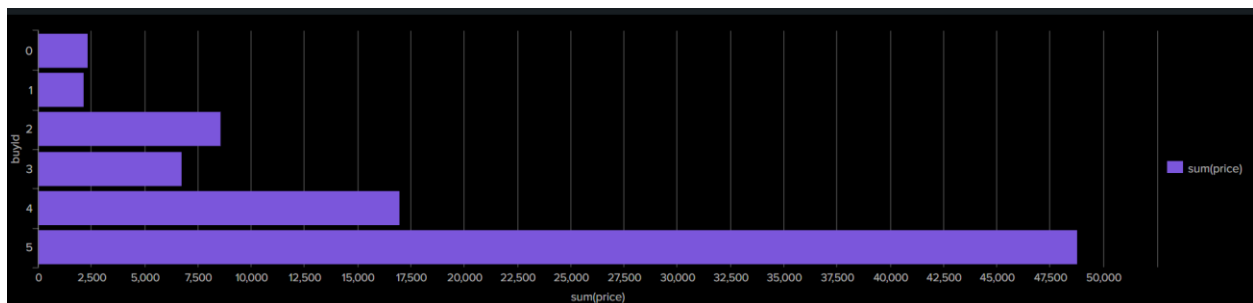


A total of \$85628 has been spent by the users on the products.

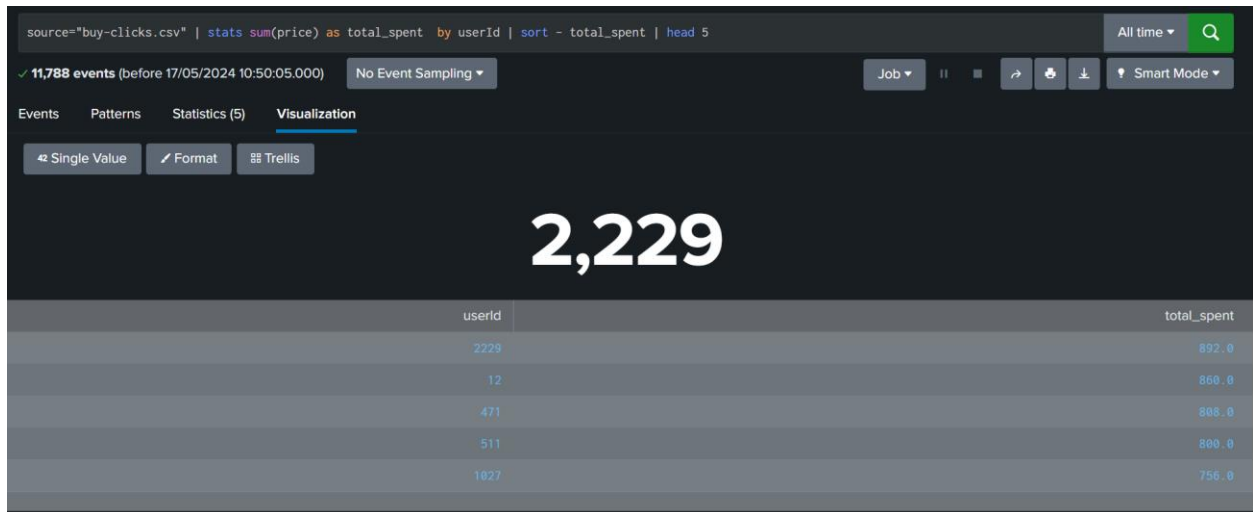
6. Amount of money was made from each item



The highest amount of money, \$48800 has been made on buyId "5".



7. The total amount of money spent by the top five users (ranked by how much money they spent).



The highest paid user is the userId "2229" and the amount spent is \$892.

8. Number of players from each country.

The screenshot shows a 'New Search' interface. The query bar contains: `source="users.csv" | stats count by country |`. Below the query bar, it indicates 7,179 events. The main table displays the count of players for each country.

country	count
AD	30
AE	30
AF	30
AG	27
AI	30
AL	30
AM	30
AN	30
AO	30
AQ	30

Important findings

- 7179 players have already registered for the game.
- There are 109 teams registered.
- Most of the players of the game have used iPhones and in second place there is Android.
- The lowest number of game players used Mac'.
- The most commonly purchased item is the item that has buyID = "2".
- A total of \$85628 has been spent by the users on the products.
- The highest amount of money, \$48800 has been made on buyId "5".
- The highest paid user is the userId "2229" and the amount spent is \$892.

Suggestions

- Since most players are using iPhones and Android devices, ensure the game is highly optimized for these platforms. Focus on improving user experience, performance, and compatibility for these operating systems.
- Run targeted marketing campaigns on iOS and Android platforms to attract more users, as these are the predominant platforms used by your players.
- Since the lowest number of players are using Mac, investigate potential issues or improvements needed for the Mac version of the game.
- Offer special incentives or promotions to encourage more Mac users to join and engage with the game.
- Consider offering bundles, discounts, or promotions on this item to boost sales further.
- The highest revenue-generating item is **buyID = 5**. Promote this item prominently within the game and explore opportunities to increase its sales further.
- Implement a VIP or loyalty program to reward high spenders with exclusive perks, discounts, or in-game rewards. This can enhance retention and encourage further spending.
- Collect feedback from users across all platforms to identify any issues or areas for improvement. Use this feedback to enhance gameplay, user interface, and overall experience.

Regularly update the game with new features, items, and improvements based on user feedback and analytics.