MKT 681 Sports Analytics
Hawks Mobile App Analytics Report

Team Hawkeye: Meng Cheng, Julie Wang, Carl Xi, Danielle (Danlei) Zhao, Jie Zhu
Feb 27, 2020

I. Exploratory Data Analysis

Our app usage data is collected in the time period 2/12/2019-1/23/2020. The total number of observations in our dataset is over 6 million.

From the app version data, we learned that there are a lot of users on different versions of the app, indicating that users are not finding the need to update their app. This means users use the app infrequently, or that they believe the core functionality is enough. The carrier information is missing from many records, and the top carrier is Comcast. The top 10 most used device models are all some generation of iPhone.

Based on user id there are 49,633 unique users. Around 25% of the users capture 80% of the total sessions/activities. Most sessions are started by people with no event plans and are not members. Most users are from within the US. A small portion is from India. The minimum app using time is 0 min, while the maximum is over 2000 minutes. This might be due to the fact that the app stops only when the app is actually closed out, instead of just the user exiting the app screen.

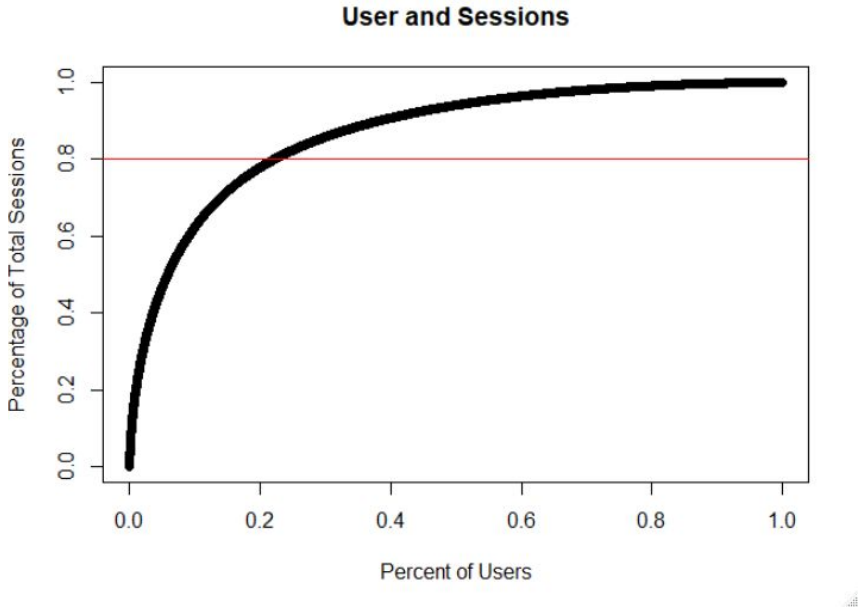

Figure 1: Around 25% of the users capture 80% of the total sessions/activities.

With some basic understanding of the variables, we would like to do some exploration of the events. We have our attendance data from two days, 2018-12-29 and 2019-01-19, with multiple records for scanned tickets. We also obtained rating and event name data from three days to do combined analysis.

We grouped events into 4 categories: basketball, show, wrestling, concert. There were 43 basketball events, 34 concert events, which were two main events in the Arena. Most events happened just once, but for basketball, about half events happen twice and Indiana pacers, Milwaukee Bucks, Philadelphia 76ers played 3 times during the year.

By analyzing factors including the distinct user count, sum session duration, time of visit(count of obs), attendance(scanned tickets), rating for each day of each event type, we found out basketball has the largest total attendance and the second-largest daily attendance. The attendance of shows and concerts is relatively small and it may be due to the limited space according to the event characteristics. Also, the LA Lakers is the most popular opponent team.

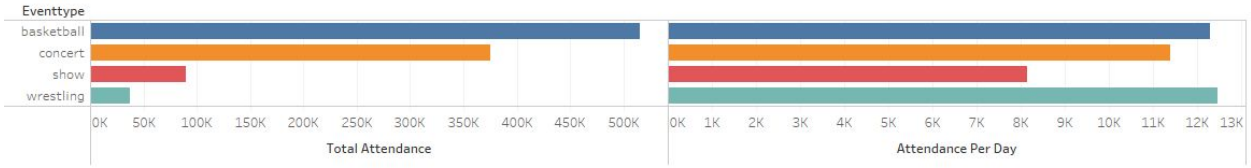

Figure 2: Comparing total attendance and attendance per day between event types.

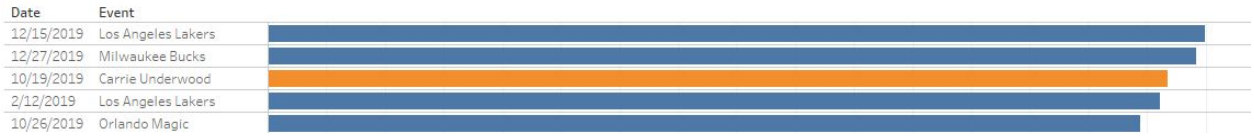

Figure 3: Top five events with the largest attendance.

When looking into user count, it's similar to when we see in the attendance data. Regarding the user count, basketball is extremely popular compared to other types.

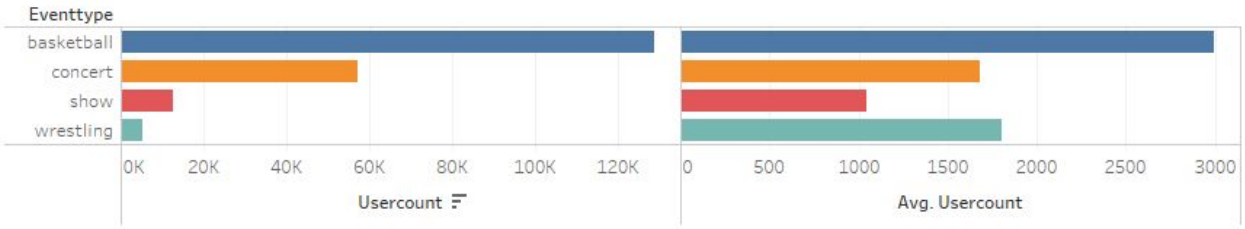

Figure 4: Comparing total user count and user count per day between event types.

Figure 5: Top five events with the largest user count.

Basketball fans contribute most to the session duration and have the longest daily session duration. Surprisingly, although there were only three wrestling events, the daily session duration is just about 15% less than the basketball event. It might because wrestling fans check the score and player information as often as basketball fans as it belongs to sports events.

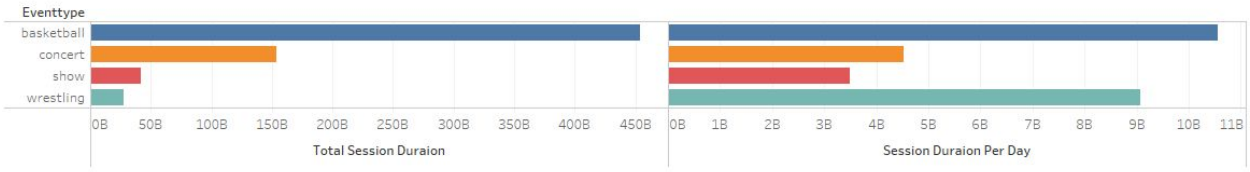

Figure 6: Comparing total session duration and session duration per day between event types.

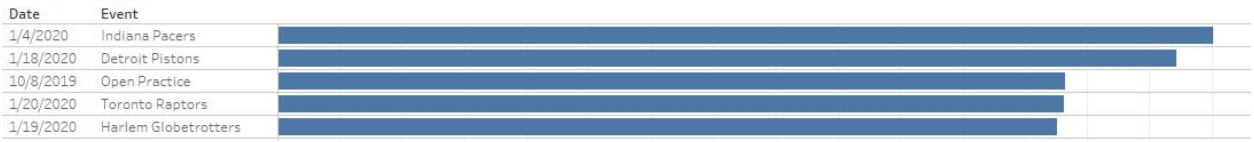

Figure 7: Top five events with the longest session duration.

Basketball fans have the highest rating for the app while concert audiences have the lowest rating. The reason behind might be that there are a lot of features in the app that are specially designed for Hawks game and people who just use the app to check location and transportation may feel messy and disturbing. Surprisingly, when we look into the top 5 events with the highest rating, Disney On Ice, the show and two concerts pop up which need more exploration about the specific event to see if there's any special interaction setting that makes the audience feel the app helpful.

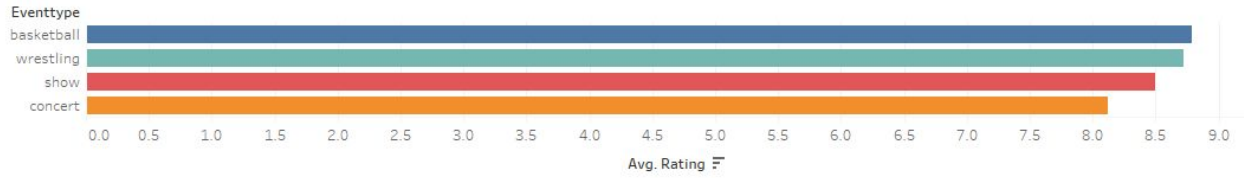

Figure 8: Comparing average app rating between event types.

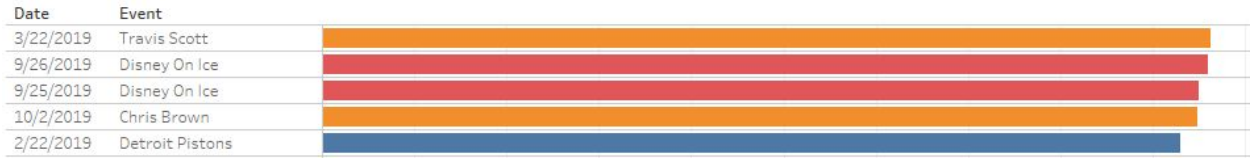

Figure 9: Top five events with the highest rating.

Figure 10: Top five events with the lowest rating.

By taking into account the proportion of people who use the app (distinct user count/attendance), average session duration per user/per visit, and comparison between event/event type, we found out that the wrestling fan has longest session time per user while concert-goers have the lowest, which is reasonable because the fandom for wrestling is usually stronger and concert-goers are likely watching the show. For basketball games, the New York Knicks game audience spent the longest time on the app on average.

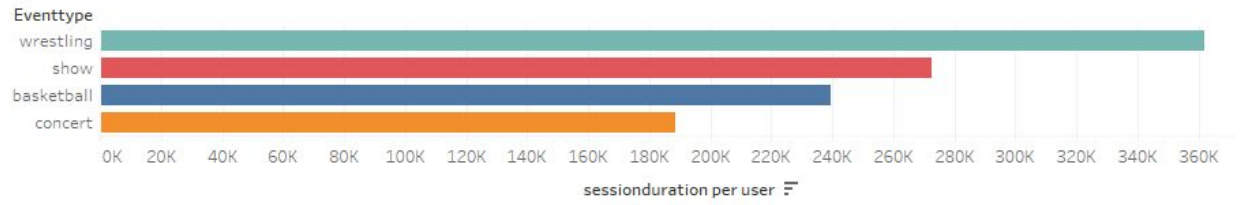

Figure 11: Comparing session duration per user between event types.

Date	Event
2/18/2019	Marvel Universe Live - ..
2/14/2019	New York Knicks
1/8/2020	Houston Rockets
12/27/2019	Milwaukee Bucks
12/28/2019	Premier Boxing Champi..

Figure 12: Top five events with the longest session duration per user.

Compared to fans of other events, basketball fans have the largest proportion of using the app. It is reasonable as Hawks fans are generally more familiar with the app and have a higher motivation to use it. When looking into the top five events with the largest app use proportion, we see not only basketball games like Chicago Bulls but also Marc Anthony concert and Impractical Jokers show, which may indicate the similarity between the audience of those events and need further exploration.

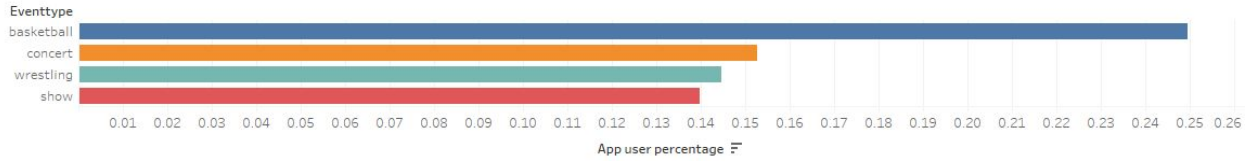

Figure 13: Comparing app user proportion between event types.

Date	Event
1/19/2020	Harlem Globetrotters
11/6/2019	Chicago Bulls
10/25/2019	Marc Anthony
10/9/2019	Orlando Magic
3/24/2019	Impractical Jokers

Figure 14: Top five events with the largest app use proportion.

The three most popular sections on the app are Hawks Home, Game Detail Shot Selection, and Hawks Home Menu according to the whole app dataset. We also dive deep into each event type. Basketball, concert and show audience show similar interest which wrestling audience really care about the scorecard and check it a lot. At the meanwhile, basketball fans have the largest user proportion checking stats_team and stats_players as team and player are the main elements of basketball and consist of the fandom.

	Top Three App Sections		
Event Type	No. 1	No. 2	No.3
Basketball/Concert/Show	hawks_home	hawks_home_menu	dock_menu
Wrestling	hawks_home	hawks_home_menu	hawks_scorecard_swipe

Figure 15: Top three App Sections by Event Type

We also tried to see the differences in membership percentage and event plans among fans of different event types. Surprisingly, wrestling fans have the highest membership percentage while basketball fans have the lowest membership percentage. It may suggest that a large proportion of basketball audiences are actually not die-hard fans but rather occasional watchers. Also, when looking into the top five events with the highest member proportion, we find the audience of Tyler the Creator and Bad Bunny are also into basketball, which may be a way to find Hawks' perspective customer.

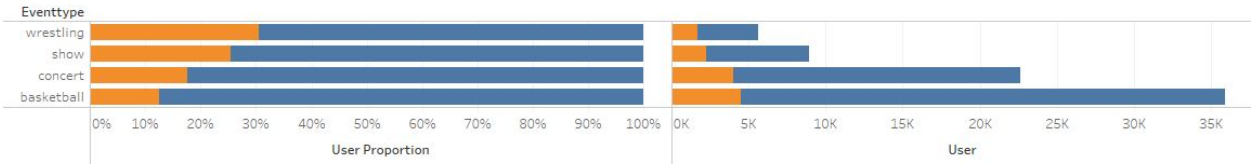

Figure 16: Count of members and percentage of members by event type

Figure 17: Percentage of Members by Event Type

The original dataset has three columns filled with JSON strings, which are *device identifiers*, *event parameters*, and *session properties*. After examining them carefully, we believe that the column, *event parameters,* contains information that is not provided by other columns. In the process of exploring event parameters, we find that there are cases in which multiple parameters are associated with one event name. Thus, we decide to use the *JSON* package within Python to break these event parameters into keys and values.

	eventname	eventparameters
0	Game_Detail_Stats	{""Type"": ""List"", ""Game"": ""20190225/ATLH...
1	Hawks_Home	{""Game_Day"": ""False""}
2	Arena_Home	{}

Figure 18: Data Before Breaking Event Parameters

	eventname	key	value
0	Game_Detail_Stats	Type	List
1	Game_Detail_Stats	Game	20190225/ATLHOU
2	Hawks_Home	Game_Day	False

Figure 19: Data After Breaking Event Parameters into Keys and Values

There are 117 event names in total. For the purpose of this analysis, we primarily looked at event names with high total session duration to find common user behaviors.

Using the event name *Hawks_Home_Menu* and key name *Menu_Name*, we are able to understand the primary use cases. It seems that most people use the Hawks app for ticket-related purposes. For parameter values, the majority app visits are associated with “My Tickets” and “Upgrade Your Ticket”, while the least use cases are “Get There”, “Team Roster” and “Hawks Shop”. Also, using the event name Hawks_Home and key name Game_Day, we find that 66% of the app visits are not on game day. Combining these two findings, we conclude that people primarily use Hawks app to purchase tickets and review ticket information and team stats before the game day.

My Tickets	345898
Upgrade Your Ticket	30317
Game Details	27939
Team Standings	11769
Team Stats	10857
Get There	5355
Team Roster	2688
HAWKS SHOP	1729
Game Schedule	1

Figure 20: Value Names and App Visits of Menu_Name in Hawks_Home_Menu

Using the event name *Game_Detail_Shot_Selection* and key name *Player_Name*, we are able to measure the popularity of players among basketball fans. It appears that Trae Young is the undisputable king of the arena. He has more app visits than the combination of the second and third popular players, who are De'Andre Hunter and Jabari Parker.

TRAE YOUNG	143020
DE'ANDRE HUNTER	56732
JABARI PARKER	56199
ALEX LEN	44871
DEANDRE' BEMBRY	44649
JOHN COLLINS	44565
KEVIN HUERTER	38383
VINCE CARTER	30989
DAMIAN JONES	24784
BRUNO FERNANDO	23171
CAM REDDISH	23122
BRANDON GOODWIN	16602

Figure 21: Value Names and App Visits of Player_Name in Game_Detail_Shot_Selection

Using the event name *Game_Detail_Scorebord_Minified* and key name *Game*, we are able to identify matches with most online attention. The top 3 viewed matches in the Hawks app are games against Los Angeles Clippers, Toronto Raptors, and Phoenix Suns.

20200122/LACATL	32547
20200120/TORATL	27566
20200114/PHXATL	14621
20200118/DETATL	13864
20200117/ATLSAS	12079
20200110/ATLWAS	10140
20191028/PHIATL	9747
20191024/ATLDET	9468
20191007/NOPATL	9443
20191026/ORLATL	8283

Figure 22: Value Names and App Visits of Game in Game_Detail_Scorebord_Minified

II. Business Insights

As we progressed through our EDA, several key business questions and observations began to surface. Un-updated app usage suggests a low return rate for non-core users, which could indicate that users are finding the core functionality enough for their needs or that users use the app infrequently when they are not at events and do not notice the need to update.

With our user base mostly following the 80-20 rule, we can further conduct segmentation to customize marketing campaigns to retain existing and attract new members and plan users. Some variables, such as gender and birth year, are currently not being collected by the app. Having these variables can really help us further segment our users. Carrier, country and device information can help us discover specific segment groups and their correlation with engagement (e.g. Comcast iPhone 11 users might be the most engaged with the app, which we can take advantage of with targeted campaigns). We also believe that carrier, country, and device information can help us infer the income bracket, and thus purchasing patterns of customer segments as well.

From segmenting our users by the events and matching app usage with the event dates, further patterns emerged. Concert-goers really do not like using our app, which could indicate that they only wanted arena-related information rather than being bombarded with Atlanta Hawks' information. This suggests that our app rating could drastically improve if we were to spin-off a smaller app dedicated to the arena's programming and facilities alone, and concentrate our current app to purely Hawks-focused to better serve our fans.

Our JSON analysis revealed the most popular games and players, further confirming some of our existing suspicions. Knowing that Trae Young is the most popular player can have fantastic implications on team branding, marketing campaigns and analysis of network effects that can make the team play better, more successful and more popular.

III. Next Steps

For further analysis, we would like to segment the users and see if we could recommend in the app certain tickets or events to them. More specifically, we would generate customer clusters by thinking about the difference between the population of those people that always come and those that can be persuaded to come who are price sensitive. To provide more personalized promotions, we would think about possible collaboration promotions based on devices, carriers, etc.